DIFFERENTIATION AND INTEGRATION IN HIGHER DIMENSION

Introduction. For our discussion of differentiation and integration in higher dimension, for notational simplicity we first confine ourselves to the case of two real variables. For a real-valued function $f(x, y)$ of many real variables x, y, to perform differentiation we can always keep one variable fixed and regard the function as a function of only the remaining variable. In that case the derivatives we obtain are partial derivatives

$$\begin{aligned} \frac{\partial f}{\partial x}(a, b) = D_1 f(a, b) = \lim_{x \to a} \frac{f(x, b) - f(a, b)}{x - a}, \\ \frac{\partial f}{\partial y}(a, b) = D_2 f(a, b) = \lim_{y \to b} \frac{f(a, y) - f(a, b)}{y - b} \end{aligned}$$

without introducing any new theory. The geometric interpretation of the derivative in the one variable case as using a polynomial of degree at most 1 to approximate the function to an order higher than the first can be extended to the case of many variables. We are going to define the differentiation in the many variable case by using this geometric interpretation of derivative. After we introduce this definition of differentiation, we will do two results about it. One is to relate this differentiation to partial differentiation. Another is to discuss the question of commutativity of partial differentiation.

For differentiation in many variables, in the special case of a complex-valued function $w = f(z)$ of a complex variable $z = x + iy$ the definition of a derivative as the limit of the difference quotient

$$\lim_{z \to c} \frac{f(z) - f(c)}{z - c}$$

can be used to define the complex derivative $f'(c)$ of $w = f(z)$ at $z = c$. The existence of $f'(c)$ will be shown to be equivalent to the total differentiability of the real part $u(x, y)$ and the imaginary part $v(x, y)$ of f as functions of two real variables and a system of two equations involving the partial derivatives of $u(x, y)$ and $v(x, y)$ with respect to x and y evaluated at $z = c$ which are known as the Cauchy-Riemann equations.

For integration involving functions of many variables, we can also fix all the variables except one and consider integration with respect to the remaining variable without introducing any new theory. To handle integration in all the variables altogether, we will go through again the route of the Riemann sum, the upper sum, the lower sum, the upper integral, and the lower integral with the modification that the domain of integration will be approximated by sandwiching between two unions of closed parallelepiped defined by partitions of intervals on coordinate axes. This definition of integral for all the variables at the same time will be related to the result obtained by integrating with respect to one variable at a time by Fubini's theorem.

Finally the higher-dimensional analogue of the Fundamental Theorem of Calculus will be Stokes's theorem.

Differentiability in Many Variables. Let $a < b$ and $c < b$ be real numbers and let $f(x, y)$ be a function on $(a, b) \times (c, d)$. Let $(\xi, \eta) \in (a, b) \times (c, d)$.

Definition of Total Differentiability. The function $f(x, y)$ is said to be *totally differentiable* at (ξ, η) if there exists a polynomial $Ax + By + C$ of degree ≤ 1 with real coefficients in the variables x, y such that for any $\varepsilon > 0$ there exists some $\delta > 0$ with the property that

$$\left| \frac{f(x, y) - (A(x - \xi) + B(y - \eta) + f(\xi, \eta))}{\sqrt{(x - \xi)^2 + (y - \eta)^2}} \right| < \varepsilon$$

for

$$0 < \sqrt{(x - \xi)^2 + (y - \eta)^2} < \delta.$$

In other words, total differentiability of $f(x, y)$ at (ξ, η) means approximability of $f(x, y)$ at (ξ, η) by a real polynomial of degree ≤ 1 in x, y to an order higher than the first. Total differentiability of $f(x, y)$ is also known simply as differentiability. The qualifier "total" is added to distinguish it from partial differentiability when it is necessary to avoid ambiguity.

Specialization to Yield Partial Derivatives. When we specialize to the case of $y = b$, we conclude from

$$\left| \frac{f(x, y) - (A(x - \xi) + B(y - \eta) + f(\xi, \eta))}{\sqrt{(x - \xi)^2 + (y - \eta)^2}} \right| < \varepsilon$$

for

$$0 < \sqrt{(x-\xi)^2 + (y-\eta)^2} < \delta$$

that

$$\left| \frac{f(x,b) - (A(x-\xi) + f(\xi,\eta))}{x-\xi} \right| < \varepsilon$$

for $0 < |x-\xi| < \delta$. This implies that $\frac{\partial f}{\partial x}(\xi,\eta) = A$. Likewise the specialization to the case of $x = a$ yields $\frac{\partial f}{\partial y}(\xi,\eta) = B$. Thus the differentiability of $f(x,y)$ at (ξ,η) as a function of two real variables implies the existence of the two partial derivatives $\frac{\partial f}{\partial x}(\xi,\eta)$ and $\frac{\partial f}{\partial y}(\xi,\eta)$. The converse is not true. The existence of both partial derivatives $\frac{\partial f}{\partial x}(\xi,\eta)$ and $\frac{\partial f}{\partial y}(\xi,\eta)$ even for all points $(\xi,\eta) \in (a,b)\times(c,d)$ does not even imply the continuity of the function at a points of $(a,b)\times(c,d)$. An example for this statement is given in the homework assignment.

Total Differentiability from Continuous Partial Derivatives. On the other hand, when both partial derivatives $\frac{\partial f}{\partial x}(\xi,\eta)$ and $\frac{\partial f}{\partial y}(\xi,\eta)$ are continuous for all points $(\xi,\eta) \in (a,b)\times(c,d)$, the function $f(x,y)$ is totally differentiable at every point of $(a,b)\times(c,d)$. The continuity of $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ at (ξ,η) means that for any $\varepsilon > 0$ there exists some $\delta > 0$ such that

$$(*) \qquad |D_1 f(x,y) - D_1 f(\xi,\eta)| < \varepsilon \quad \text{for} \quad 0 < \sqrt{(x-\xi)^2 + (y-\eta)^2} < \delta.$$

Now choose arbitrarily x, y such that

$$0 < \sqrt{(x-\xi)^2 + (y-\eta)^2} < \delta.$$

First consider the case $x > \xi$ and $y > \eta$. We write

$$f(x,y) - f(\xi,\eta) = f(x,y) - f(x,\eta) + f(x,\eta) - f(\xi,\eta).$$

By applying the Mean-Value Theorem to the function $f(x,\eta) - f(\xi,\eta)$ of the variable x on the interval $[\xi, x]$, we conclude that there exists some $\sigma \in (\xi, x)$ such that

$$f(x,\eta) - f(\xi,\eta) = D_1 f(\sigma,\eta)(x-\xi).$$

By applying the Mean-Value Theorem to the function $f(x,y) - f(x,\eta)$ of the variable y on the interval $[\eta, y]$, we conclude that there exists some $\tau \in (\eta, y)$ such that

$$f(x,y) - f(x,\eta) = D_2 f(x,\tau)(y-\eta).$$

We obtain

$$\begin{aligned} f(x,y) - f(\xi,\eta) &= D_1 f(\sigma,\eta)(x-\xi) + D_2 f(x,\tau)(y-\eta) \\ &= D_1 f(\xi,\eta)(x-\xi) + D_2 f(\sigma,\eta)(y-\eta) \\ &\quad + \left(D_1 f(\sigma,\eta) - D_1 f(\xi,\eta)\right)(x-\xi) + \left(D_2 f(x,\tau) - D_2 f(\sigma,\eta)\right)(y-\eta) \end{aligned}$$

or

$$\begin{aligned} &f(x,y) - \left(f(\xi,\eta) + D_1 f(\sigma,\eta)(x-\xi) + D_2 f(x,\tau)(y-\eta)\right) \\ &\quad = \left(D_1 f(\sigma,\eta) - D_1 f(\xi,\eta)\right)(x-\xi) + \left(D_2 f(x,\tau) - D_2 f(\sigma,\eta)\right)(y-\eta). \end{aligned}$$

By (*)

$$|D_1 f(\sigma,\eta) - D_1 f(\xi,\eta)| < \varepsilon \quad \text{and} \quad |D_1 f(x,\tau) - D_1 f(\xi,\eta)| < \varepsilon.$$

Thus

$$\left| f(x,y) - \left(f(\xi,\eta) + D_1 f(\sigma,\eta)(x-\xi) + D_2 f(x,\tau)(y-\eta)\right)\right| < \varepsilon\,|x-\xi| + \varepsilon\,|y-\eta|\,,$$

which implies that

$$\left| \frac{f(x,y) - \left(f(\xi,\eta) + D_1 f(\sigma,\eta)(x-\xi) + D_2 f(x,\tau)(y-\eta)\right)}{\sqrt{(x-\xi)^2 + (y-\eta)^2}} \right| < 2\varepsilon.$$

The other three cases of $x < \xi, y > \eta$, $x > \xi, y < \eta$, $x < \xi, y < \eta$ and the cases when $x = \xi$ or $y = \eta$ can be handled analogously to enable us to conclude that

$$\left| \frac{f(x,y) - \left(f(\xi,\eta) + D_1 f(\sigma,\eta)(x-\xi) + D_2 f(x,\tau)(y-\eta)\right)}{\sqrt{(x-\xi)^2 + (y-\eta)^2}} \right| < 2\varepsilon$$

for $0 < \sqrt{(x-\xi)^2 + (y-\eta)^2} < \delta$. We can now conclude that $f(x,y)$ is differentiable at (ξ,η) as a function of two real variables x, y.

Complex Derivative. Let c be a complex number and $R > 0$. Let $w = f(z)$ be a complex-valued function of a complex variable z on $|z-c| < R$. The function $f(z)$ is said to be *complex differentiable* with complex derivative $f'(c)$ at $z = c$ if foe very $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\left| \frac{f(z) - f(c)}{z - c} - f'(c) \right| < \varepsilon$$

for $0 < |z - c| < \delta$. Let $c = a + ib$ and $z = x + iy$. We write $f(z)$ also as $f(x, y)$ and let $u(x, y)$ be the real part of $f(x, y)$ and let $v(x, y)$ be the imaginary part of $f(x, y)$. Specializing to the special case $y = b$, we obtain

$$\left|\frac{f(x,b) - f(a,b)}{x - a} - f'(c)\right| < \varepsilon$$

for $0 < |x - a| < \delta$, because $z - c = x - a$ when $y = b$. This means that $\frac{\partial f}{\partial x}(c) = f'(c)$. Here $\frac{\partial f}{\partial x}(c)$ means $\frac{\partial u}{\partial x}(a, b) + i\frac{\partial u}{\partial x}(a, b)$. Specializing to the special case $x = a$, we obtain

$$\left|\frac{f(x,b) - f(a,b)}{i(y - b)} - f'(c)\right| < \varepsilon$$

for $0 < |y - b| < \delta$, because $z - c = i(y - b)$ when $x = a$. This means that $\frac{1}{i}\frac{\partial f}{\partial y}(c) = f'(c)$. Here $\frac{\partial f}{\partial y}(c)$ means $\frac{\partial u}{\partial y}(a, b) + i\frac{\partial u}{\partial y}(a, b)$. Since both $\frac{\partial f}{\partial x}(c)$ and $\frac{1}{i}\frac{\partial f}{\partial x}(c)$ are both equal to $f'(c)$, we have the equation

$$(\sharp) \qquad \frac{\partial f}{\partial x}(c) = \frac{1}{i}\frac{\partial f}{\partial y}(c),$$

which is known as the Cauchy-Riemann equation. Taking the real and imaginary parts of both sides, we obtain

$$(\natural) \qquad \begin{cases} \frac{\partial u}{\partial x}(a, b) = \frac{\partial v}{\partial y}(a, b) \\ \\ \frac{\partial v}{\partial x}(a, b) = -\frac{\partial u}{\partial y}(a, b), \end{cases}$$

which is another form of the Cauchy-Riemann equation.

Let

$$E_1(z) = \frac{f(z) - f(c)}{z - c} - f'(c).$$

We can rewrite it as

$$f(z) = f(c) + f'(c)(z - c) + E(z),$$

where $E(z) = E_1(z)(z - c)$. The statement that

$$\left|\frac{f(z) - f(c)}{z - c} - f'(c)\right| < \varepsilon$$

for $0 < |z - c| < \delta$ can be reformulated as $|E_1(z)| < \varepsilon$ for $0 < |z - c| < \delta$. This means that

$$f(z) = f(c) + f'(c)(z - c) + E(z),$$

with $\left|\frac{E(z)}{z-c}\right| < \varepsilon$ for $0 < |z - c| < \delta$. Let A be the real part of $f'(c)$ and B be the imaginary part of $f'(c)$. Taking the real parts of both sides of

$$f(z) = f(c) + f'(c)(z - c) + E(z),$$

we get

$$u(x, y) = u(a, b) + A(x - a) - B(y - b) + \operatorname{Re} E(z)$$

with $\left|\frac{\operatorname{Re} E(z)}{|z-c|}\right| < \varepsilon$ for $0 < |z - c| < \delta$. This means that $u(x, y)$ is differentiable at $(x, y) = (a, b)$ as a function of two real variables x, y. Likewise, taking the real parts of both sides of

$$f(z) = f(c) + f'(c)(z - c) + E(z),$$

we get

$$v(x, y) = u(a, b) + B(x - a) + C(y - b) + \operatorname{Im} E(z)$$

with $\left|\frac{\operatorname{Im} E(z)}{|z-c|}\right| < \varepsilon$ for $0 < |z - c| < \delta$. This means that $v(x, y)$ is differentiable at $(x, y) = (a, b)$ as a function of two real variables x, y.

To summarize, we have the following statement.

If the complex derivative $f'(c)$ of the complex-valued function $w = f(z)$ of a complex variable z exists at the point $z = c$, then (i) the real part $u(x, y)$ and the imaginary part $v(x, y)$ of $f(z)$ are differentiable at $(x, y) = (a, b)$ as functions of two real variables x, y and (ii) the Cauchy-Riemann equation ($\natural$) or its equivalent form ($\sharp$) is satisfied.

The converse of this statement is also true. Suppose (i) the real part $u(x, y)$ and the imaginary part $v(x, y)$ of $f(z)$ are differentiable at $(x, y) = (a, b)$ as functions of two real variables x, y and (ii) the Cauchy-Riemann equation ($\natural$) or its equivalent form ($\sharp$) is satisfied. We are going to prove that $f'(c)$ exists. Since the real part $u(x, y)$ and the imaginary part $v(x, y)$ of $f(z)$ are differentiable at $(x, y) = (a, b)$ as functions of two real variables x, y, it follows that

$$u(x,y) = u(a,b) + \frac{\partial u}{\partial x}(a,b)(x-a) + \frac{\partial u}{\partial y}(a,b)(y-b) + \hat{E}(x,y),$$
$$v(x,y) = v(a,b) + \frac{\partial v}{\partial x}(a,b)(x-a) + \frac{\partial v}{\partial y}(a,b)(y-b) + \hat{\hat{E}}(x,y),$$

with

$$\lim_{z\to c} \frac{\hat{E}(x,y)}{z-c} = 0 \quad \text{and} \quad \lim_{z\to c} \frac{\hat{\hat{E}}(x,y)}{z-c} = 0.$$

Let $A = \frac{\partial u}{\partial x}(a,b)$ and $B = \frac{\partial v}{\partial x}(a,b)$. From the Cauchy-Riemann equation ($\natural$) it follows that $A = \frac{\partial v}{\partial y}(a,b)$ and $B = -\frac{\partial u}{\partial y}(a,b)$. After we multiply the second equation of

$$u(x,y) = u(a,b) + A(x-a) - B(y-b) + \hat{E}(x,y),$$
$$v(x,y) = v(a,b) + B(x-a) + A(y-b) + \hat{\hat{E}}(x,y)$$

by i and adding it to the first equation, we obtain

$$f(z) = f(c) + (A+iB)(x-a) + (-B+iA)(y-b) + \hat{E}(x,y) + i\hat{\hat{E}}(x,y),$$

which is the same as

$$f(z) = f(c) + (A+iB)(z-c) + \hat{E}(x,y) + i\hat{\hat{E}}(x,y).$$

From

$$\lim_{z\to c} \frac{\hat{E}(x,y) + i\hat{\hat{E}}(x,y)}{z-c} = 0,$$

it follows that

$$\lim_{z\to c} \left(\frac{f(z)-f(c)}{z-c} - (A+iB) \right) = \lim_{z\to c} \frac{\hat{E}(x,y) + i\hat{\hat{E}}(x,y)}{z-c} = 0,$$

which means that $f'(c)$ exists.

We would like to give two interpretations of the Cauchy-Riemann equation. The first one concerns the definition of differentiability as approximability by polynomial of degree ≤ 1. For complex differentiability the approximation is by a polynomial $f(c) + f'(c)(z-c)$ of degree ≤ 1 in a complex variable z. The differentiability of $u(x,y)$ and $v(x,y)$ at $(x,y)=(a,b)$ means by polynomials $u(a,b)+\frac{\partial u}{\partial x}(a,b)(x-a)+\frac{\partial u}{\partial y}(a,b)(y-b)$ and $v(a,b)+\frac{\partial v}{\partial x}(a,b)(x-a)+\frac{\partial v}{\partial y}(a,b)(y-b)$ of degree ≤ 1 in the two real variables respectively. Putting together these two approximation we have the approximation of $u(x,y)+iv(x,y)$ by the polynomial

$$(u(a,b)+iv(a,b))+\left(\frac{\partial u}{\partial x}(a,b)+i\frac{\partial v}{\partial x}(a,b)\right)(x-a)+\left(\frac{\partial u}{\partial y}(a,b)+i\frac{\partial v}{\partial y}(a,b)\right)(y-b)$$

of degree ≤ 1 in the two real variables x,y with complex coefficients. To go from the differentiability of $u(x,y)$ and $v(x,y)$ at $(x,y)=(a,b)$ to the complex differentiability of $f(z)$ at $z=c$, what is needed to say that the polynomial

$$(u(a,b)+iv(a,b))+\left(\frac{\partial u}{\partial x}(a,b)+i\frac{\partial v}{\partial x}(a,b)\right)(x-a)+\left(\frac{\partial u}{\partial y}(a,b)+i\frac{\partial v}{\partial y}(a,b)\right)(y-b)$$

of degree ≤ 1 in the two real variables x,y with complex coefficients is a polynomial of degree ≤ 1 in the complex variable $z=x+iy$. This means that we need the identidy

$$\frac{\partial u}{\partial x}(a,b)+i\frac{\partial v}{\partial x}(a,b)=\frac{1}{i}\left(\frac{\partial u}{\partial y}(a,b)+i\frac{\partial v}{\partial y}(a,b)\right).$$

Its real and imaginary parts give precisely the Cauchy-Riemann equation ($\natural$).

Another interpretation of the Cauchy-Riemann equation is the condition for a $\mathbb{R}$-linear transformation T of the 2-dimensional $\mathbb{R}$-vector space $\mathbb{R}^2$ to itself to be a $\mathbb{C}$-linear transformation of the 1-dimensional $\mathbb{C}$-vector space $\mathbb{C}$ to itself. Let

$$J=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$$

which as the $\mathbb{R}$-linear transformation of the 2-dimensional $\mathbb{R}$-vector space $\mathbb{R}^2$ to itself satisfies $J^2=I_2$ and can serve as the scalar multiplication by i to make $\mathbb{R}^2$ a vector space over $\mathbb{C}$. Let

$$T=\begin{pmatrix}A & B\\ C & D\end{pmatrix}.$$

Then T commutes with J if and only if $A=D$ and $C=-B$, because

$$TJ=\begin{pmatrix}A & B\\ C & D\end{pmatrix}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}=\begin{pmatrix}B & -A\\ D & -C\end{pmatrix}$$

and

$$JT = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} -C & -D \\ A & B \end{pmatrix}.$$

When

$$T = \begin{pmatrix} \frac{\partial u}{\partial x}(a,b) & \frac{\partial u}{\partial y}(a,b) \\ \\ \frac{\partial v}{\partial x}(a,b) & \frac{\partial v}{\partial y}(a,b) \end{pmatrix},$$

the Cauchy-Riemann equation ($\natural$) is precisely the condition for the $\mathbb{R}$-linear transformation T of the 2-dimensional $\mathbb{R}$-vector space $\mathbb{R}^2$ to itself to be a $\mathbb{C}$-linear transformation of the 1-dimensional $\mathbb{C}$-vector space $\mathbb{C}$ to itself. The reason for this interpretation is that for a polynomial of degree ≤ 1 in the two real variables x, y with complex coefficients to be a polynomial of degree ≤ 1 in the complex variable $z = x + iy$ the obstruction is the homogeneous linear part which can be interpreted as defining a linear transformation of vector spaces.

Commutativity of Partial Differentiation. For a function $f(x, y)$ of two real variables x, y, the partial derivative of f with respect to the first variable x is denoted by $D_1 f$ and the partial derivative of f with respect to the second variable y is denoted by $D_2 f$. We use $D_2 D_1 f$ to denote the partial derivative of $D_1 f$ with respect to the second variable y. Likewise we use $D_1 D_2 f$ to denote the partial derivative of $D_2 f$ with respect to the first variable x.

Double Application of Mean-Value Theorem for Function of Two Variables. Suppose $\tilde{a} < a < b < \tilde{b}$ and $\tilde{c} < c < d < \tilde{d}$ are real numbers. Suppose f is a function on $\left(\tilde{a}, \tilde{b}\right) \times \left(\tilde{c}, \tilde{d}\right)$ such that $D_2 D_1 f$ exist at every point of $\left(\tilde{a}, \tilde{b}\right) \times \left(\tilde{c}, \tilde{d}\right)$. (Note that the existence of $D_2 D_1 f$ automatically implies the existence of $D_1 f$.) Then there exists some $(\xi, \eta) \in (a, b) \times (c, d)$ such that the alternate sum

$$f(a,c) - f(b,c) + f(b,d) - f(a,d)$$

of functional values at rectangle vertices is equal to the area $(b-a)(d-c)$ of the rectangle times the second-order derivative $D_2 D_1 f$ of f at (ξ, η).

The verification is as follows. Consider the difference

$$g(x) = f(x,d) - f(x,c)$$

of specializations of the function f on the two vertical sides. Then

$$g(b) - g(a) = f(b,d) - f(b,c) - (f(a,d) - f(a,c))$$

is equal to the alternate sum

$$f(a,c) - f(b,c) + f(b,d) - f(a,d)$$

of functional values at rectangle vertices By applying the Mean-Value Theorem to $g(x)$ on $[a,b]$, we obtain some $\xi \in (a,b)$ such that $g(b) - g(a) = g'(\xi)(b-a)$. Now $g'(\xi) = D_1 f(\xi, d) - D_1 f(\xi, c)$. We consider the function $h(y) = D_1 f(\xi, y)$ for $y \in [c,d]$. BY applying the Mean-Value Theorem to $h(x)$ on $[c,d]$, we obtain some $\eta \in (c,d)$ such that $h(d) - h(c) = h'(\eta)(d-c)$. Since $h'(\eta) = D_2 D_1 f(\xi, \eta)$, it follows that

$$\begin{aligned} g'(\xi) &= D_1 f(\xi, d) - D_1 f(\xi, c) \\ &= (d-c) D_2 D_1 f(\xi, \eta) \end{aligned}$$

and

$$\begin{aligned} &f(a,c) - f(b,c) + f(b,d) - f(a,d) \\ &\quad = g'(\xi)(b-a) \\ &\quad = (b-a)(d-c) D_2 D_1 f(\xi, \eta). \end{aligned}$$

Commutativity of Partial Differentiation. Suppose $\tilde{a} < \tilde{b}$ and $\tilde{c} < \tilde{d}$ are real numbers. Suppose f is a function on $\left(\tilde{a}, \tilde{b}\right) \times \left(\tilde{c}, \tilde{d}\right)$ such that $D_2 D_1 f$ exists at every point of $\left(\tilde{a}, \tilde{b}\right) \times \left(\tilde{c}, \tilde{d}\right)$. Suppose $(\sigma, \tau) \in \left(\tilde{a}, \tilde{b}\right) \times \left(\tilde{c}, \tilde{d}\right)$ and $D_2 D_1 f$ is continuous at (σ, τ) and $D_2 f$ exists at every point of $\left(\tilde{a}, \tilde{b}\right) \times \{\tau\}$. Then $D_1 D_2 f$ exists and is equal to $D_2 D_1 f$ at (σ, τ).

Proof. The continuity of $D_2 D_1 f$ at (σ, τ) means that for any $\varepsilon > 0$ there exist $\delta > 0$ such that

$$(**) \quad |D_2 D_1 f(x,y) - D_2 D_1 f(\sigma, \tau)| < \varepsilon \quad \text{for} \quad |x - \sigma| < \delta \text{ and } |y - \tau| < \delta.$$

Choose $0 < \delta_1 < \delta$ and $0 < \delta_2 < \delta$ such that $\sigma + \delta_1 < \tilde{b}$ and $\tau + \delta_2 < \tilde{d}$. Let $a = \sigma$ and $c = \sigma$, $b = a + \delta_1$, and $d = c + \delta_2$. By the double application of the Mean-Value Theorem given above,

$$(\ddagger) \quad f(a,c) - f(b,c) + f(b,d) - f(a,d) = (b-a)(d-c) D_2 D_1 f(\xi, \eta)$$

for some $(\xi, \eta) \in (a, b) \times (c, d)$. By $(**)$

$$|D_2 D_1 f(\xi, \eta) - D_2 D_1 f(\sigma, \tau)| < \varepsilon$$

and by (‡)

$$\left| \frac{f(a,c) - f(b,c) + f(b,d) - f(a,d)}{(b-a)(d-c)} - D_2 D_1 f(\sigma, \tau) \right| < \varepsilon,$$

which we can rewrite as

$$\left| \frac{f(\sigma, \tau) + f(\sigma + \delta_1, \tau + \delta_2) - f(\sigma, \tau + \delta_2)}{\delta_1 \delta_2} - D_2 D_1 f(\sigma, \tau) \right| < \varepsilon,$$

or

$$\left| \frac{1}{\delta_1} \left(\frac{(f(\sigma + \delta_1, \tau + \delta_2) - f(\sigma, \tau + \delta_2)) - (f(\sigma + \delta_1, \tau) - f(\sigma, \tau))}{\delta_2} \right) - D_2 D_1 f(\sigma, \tau) \right| < \varepsilon.$$

Since $D_2 f$ exists at $(\sigma + \delta_1, \tau)$, we can let $\delta_2 \to 0$ in the difference quotient

$$\frac{(f(\sigma + \delta_1, \tau + \delta_2) - f(\sigma, \tau + \delta_2)) - (f(\sigma + \delta_1, \tau) - f(\sigma, \tau))}{\delta_2}$$

for the function $f(\sigma + \delta_1, y) - f(\sigma, y)$ of y and obtain

$$\left| \frac{1}{\delta_1} \left(D_2 f(\sigma + \delta_1, \tau) - D_2 f(\sigma, \tau) \right) - D_2 D_1 f(\sigma, \tau) \right| < \varepsilon.$$

Since this holds for $0 < \delta_1 < \delta$ with $\delta > 0$ chosen for an arbitrarily prescribed $\varepsilon > 0$, we conclude that

$$\lim_{\delta_1 \to 0+} \frac{D_2 f(\sigma + \delta_1, \tau) - D_2 f(\sigma, \tau)}{\delta_1} = D_2 D_1 f(\sigma, \tau).$$

Likewise, we can prove that

$$\lim_{\delta_1 \to 0-} \frac{D_2 f(\sigma + \delta_1, \tau) - D_2 f(\sigma, \tau)}{\delta_1} = D_2 D_1 f(\sigma, \tau)$$

by setting $a = \sigma - \delta_1$ and $b = \sigma$. Thus $D_1 D_2 f$ exists at (σ, τ) and is equal to $D_2 D_1 f(\sigma, \tau)$.

Integrals in Several Variables. Let E be a set in $\mathbb{R}^n$. Assume that E is bounded so that E is contained in the product

$$(a_1, b_1) \times \cdots \times (a_n, b_n)$$

for some real numbers $a_j < b_j$ for $1 \leq j \leq n$. Consider a partition P_k

$$a_k = x_0^{(k)} \leq x_0^{(k)} \leq \cdots \leq x_{m_j}^{(k)} = b_k$$

of $[a_k, b_k]$ for $1 \leq k \leq n$. Let $P = (P_1, \cdots, P_n)$. Let the $\underline{J}$ be the set of all multi-indices $j = (j_1, \cdots, j_n)$ with $0 \leq j_k < m_j$ such that the closed parallelepiped

$$I_j = \left[x_{j_1}^{(1)}, x_{j_1+1}^{(1)}\right] \times \cdots \times \left[x_{j_n}^{(n)}, x_{j_n+1}^{(n)}\right]$$

is contained in the set E and let $\overline{J}$ be the set of all multi-indices $j = (j_1, \cdots, j_n)$ $0 \leq j_k < m_j$ such that the closed parallelepiped

$$I_j = \left[x_{j_1}^{(1)}, x_{j_1+1}^{(1)}\right] \times \cdots \times \left[x_{j_n}^{(n)}, x_{j_n+1}^{(n)}\right]$$

which intersects E. Use $m(I_j)$ to denote the volume of I_j (also known as the measure of I_j). Let f be a nonnegative-valued function on E. Define the lower Riemann sum $L(f, E, P)$ to be

$$\sum_{j \in \underline{J}} \left(\inf_{I_j} f\right) m(I_j).$$

Define the upper Riemann sum $U(f, E, P)$ to be

$$\sum_{j \in \overline{J}} \left(\sup_{I_j} f\right) m(I_j).$$

Define the upper Riemann integral $\overline{\int}_E f$ of f over E to be

$$\overline{\int}_E f = \inf_P U(f, E, P).$$

Define the lower Riemann integral $\underline{\int}_E f$ of f over E to be

$$\underline{\int}_E f = \sup_P U(f, E, P).$$

The nonnegative-valued function f is Riemann integrable over E if

$$\overline{\int}_E f = \underline{\int}_E f$$

and in such a case the integral $\int_E f$ of f over E is defined as the common value. Note that we have to restrict ourselves to nonnegative-valued functions first, because some parallelepiped is used in the Riemann upper sum but not in the Riemann lower sum and the omission of a negative number from the lower Riemann sum would have the effect of increasing it rather than decreasing it. To define the integral of a function which is not necessarily nonnegative, we can write it first as the difference of two nonnegative functions and then take the difference of the two integrals of nonnegative functions.

Fubini's Theorem. For a continuous function f on $[a,b]\times[c,d]$ (with $a<b$ and $c<d$ being real numbers), the integral

$$\int_{[a,b]\times[c,d]} f$$

exists and is equal to

$$\int_{[a,b]\times[c,d]} f = \int_a^b \left(\int_c^d f\right),$$

where the function

$$x \mapsto \int_c^d f(x,y)\,dy$$

is continuous in $x \in [a,b]$.

Proof. Before we do the proof, we make the following two remarks.

(1) The function $f(x,y)$ is uniformly continuous on $[a,b]\times[c,d]$ in the sense that for any $\varepsilon>0$ there exists some $\delta>0$ such that $|f(x,y)-f(\tilde{x},\tilde{y})|<\varepsilon$ for any two points (x,y) and $(\tilde{x},\tilde{y})$ of $[a,b]\times[c,d]$ with $\|(x,y)-(\tilde{x},\tilde{y})\|<\delta$, where $\|(x,y)-(\tilde{x},\tilde{y})\|$ means the distance between the two points (x,y) and $(\tilde{x},\tilde{y})$.

(2) Both the supremum and the infimum of the function $f(x,y)$ on $[a,b]\times[c,d]$ are achieved at some points of $[a,b]\times[c,d]$ in the sense that there exist $(x_1,y_1),(x_2,y_2)\in[a,b]\times[c,d]$ such that $f(x,y)\le f(x_1,y_1)$ for $(x,y)\in[a,b]\times[c,d]$ and $f(x,y)\ge f(x_2,y_2)$ for $(x,y)\in[a,b]\times[c,d]$.

Both come from the compactness of $[a,b]\times[c,d]$. More precisely, suppose the contrary of (1). Then for some $\varepsilon>0$ no $\delta=\frac{1}{n}$ for any $n\in\mathbb{N}$ satisfies the requirement so that there exist $(x_n,y_n),(\tilde{x}_n,\tilde{y}_n)\in[a,b]\times[c,d]$ with $\|(x_n,y_n)-(\tilde{x}_n,\tilde{y}_n)\|<\frac{1}{n}$ and $|f(x_n,y_n)-f(\tilde{x}_n,\tilde{y}_n)|\ge\varepsilon$. By going to several layers of subsequences, we have

$$x_{n_\nu}\to x^*,\quad y_{n_{\nu_\lambda}}\to y^*,\quad \tilde{x}_{n_{\nu_{\lambda_k}}}\to\tilde{x}^*,\quad \tilde{y}_{n_{\nu_{\lambda_{k_j}}}}\to\tilde{y}^*$$

as $\nu,\lambda,k,j\to\infty$ respectively, for some $x^*,\tilde{x}^*\in[a,b]$ and some $y^*,\tilde{y}^*\in[c,d]$. By passing to limit in

$$\left\|\left(x_{n_{\nu_{\lambda_{k_j}}}},y_{n_{\nu_{\lambda_{k_j}}}}\right)-\left(\tilde{x}_{n_{\nu_{\lambda_{k_j}}}},\tilde{y}_{n_{\nu_{\lambda_{k_j}}}}\right)\right\|<\frac{1}{n_{\nu_{\lambda_{k_j}}}}$$

and

$$\left|f\left(x_{n_{\nu_{\lambda_{k_j}}}},y_{n_{\nu_{\lambda_{k_j}}}}\right)-f\left(\tilde{x}_{n_{\nu_{\lambda_{k_j}}}},\tilde{y}_{n_{\nu_{\lambda_{k_j}}}}\right)\right|\ge\varepsilon$$

as $j\to\infty$ and using the continuity of f on $[a,b]\times[c,d]$, we obtain the contradiction that

$$\|(x^*,y^*)-(\tilde{x}^*,\tilde{y}^*)\|=0\quad\text{and}\quad|f(x^*,y^*)-f(\tilde{x}^*,\tilde{y}^*)|\ge\varepsilon.$$

For the verification of Remark (2) suppose the contrary. Then for $n\in\mathbb{N}$ there exist $(x_n,y_n)\in[a,b]\times[c,d]$ such that

$$-\frac{1}{n}+\sup_{[a,b]\times[c,d]}f\le f(x_n,y_n)\le\sup_{[a,b]\times[c,d]}f$$

from the definition of supremum. By going to a subsequence and then a subsequence of the subsequence, we have

$$x_{n_\nu}\to x^*,\quad y_{n_{\nu_k}}\to y^*$$

as $\nu,k\to\infty$ respectively, for some $x^*\in[a,b]$ and some $y^*\in[c,d]$. By By passing to limit in

$$-\frac{1}{n_{\nu_k}}+\sup_{[a,b]\times[c,d]}f\le f(x_{\nu_k},y_{\nu_k})\le\sup_{[a,b]\times[c,d]}f$$

as $k \to \infty$ and using the continuity of f on $[a,b] \times [c,d]$, we conclude that

$$\sup_{[a,b]\times[c,d]} f = f\left(x^*, y^*\right)$$

and the supremum of f on $[a,b] \times [c,d]$ is achieved at the point (x^*, y^*) of $[a,b] \times [c,d]$. Likewise, the infemum of f on $[a,b] \times [c,d]$ is achieved at some point of $[a,b] \times [c,d]$.

First of all, we prove that the function

$$x \mapsto \int_c^d f\left(x, y\right) dy$$

is continuous in $x \in [a,b]$. Take arbitrarily a positive number ε. By the continuity of f on $[a,b]\times[c,d]$ there exists some $\delta > 0$ such that $|f(x,y) - f\left(\tilde{x}, \tilde{y}\right)| < \varepsilon$ for any two points (x,y) and $(\tilde{x}, \tilde{y})$ of $[a,b]\times[c,d]$ with $\|(x,y) - (\tilde{x}, \tilde{y})\| < \delta$. Take a partition Q

$$c = y_0 \leq y_1 \leq \cdots \leq y_{m-1} \leq y_m = d$$

of $[c,d]$ with mesh

$$\max_{0 \leq k \leq m-1} \left(|y_{k+1} - y_k|\right) < \delta.$$

Let $G_x(y) = f(x,y)$. From

$$\sup_{[y_k, y_{k+1}]} G_x < \varepsilon + \inf_{[y_k, y_{k+1}]} G_x \quad \text{for } 0 \leq k \leq m-1 \text{ and } x \in [a,b]$$

and

$$U\left(G_x, Q\right) = \sum_{k=0}^{m-1} \left(\sup_{[y_k, y_{k+1}]} G_x\right) \left(y_{k+1} - y_k\right),$$
$$L\left(G_x, Q\right) = \sum_{k=0}^{m-1} \left(\inf_{[y_k, y_{k+1}]} G_x\right) \left(y_{k+1} - y_k\right),$$

it follows that

$$U\left(G_x, Q\right) - L\left(G_x, Q\right) < \varepsilon(d-c)$$

for $x \in [a,b]$, that is, the length of the interval with left end-point $L\left(G_x, Q\right)$ and right end-point $U\left(G_x, Q\right)$ is $< \varepsilon(d-c)$ for $x \in [a,b]$. Since

$$L\left(G_x, Q\right) \leq \int_c^d G_x \leq U\left(G_x, Q\right)$$

and

$$L\left(G_x, Q\right) \leq \sum_{k=0}^{m-1} G_x\left(y_k\right)\left(y_{k+1}-y_k\right) \leq U\left(G_x, Q\right),$$

as two points on an interval of length $< \varepsilon(d-c)$ the distance

$$\left|\int_c^d G_x - \sum_{k=0}^{m-1} G_x\left(y_k\right)\left(y_{k+1}-y_k\right)\right|$$

between $\int_c^d G_x$ and $\sum_{k=0}^{m-1} G_x\left(y_k\right)\left(y_{k+1}-y_k\right)$ is $< \varepsilon(d-c)$. For later use we single out this intermediate step and write it in the following form.

(‡) If the absolute-value of the difference of the values of a function at two points of $[c,d]$ with distance $< \delta$ is $< \varepsilon$, then for any partition of $[c,d]$ with mesh $< \delta$ the Riemann sum of the function for the partition which uses the value of the function at the left end-point of each interval in the partition differs from the integral over $[c,d]$ by no more than $\varepsilon(d-c)$.

It follows from

$$\left|\sum_{k=0}^{m-1} G_x\left(y_k\right)\left(y_{k+1}-y_k\right) - \sum_{k=0}^{m-1} G_{\tilde{x}}\left(y_k\right)\left(y_{k+1}-y_k\right)\right| < \varepsilon(d-c)$$

for $x, \tilde{x} \in [a,b]$ with $|x-\tilde{x}| < \delta$ that

$$(**) \qquad \left|\int_c^d G_x - \int_c^d G_{\tilde{x}}\right| < 3\varepsilon(d-c) \quad \text{for} \quad x, \tilde{x} \in [a,b] \quad \text{with} \quad |x-\tilde{x}| < \delta.$$

It means that the function

$$x \mapsto \int_c^d f\left(x,y\right) dy = \int_c^d G_x$$

is continuous on $x \in [a,b]$.

We are now going to prove that

$$\int_{[a,b]\times[c,d]} f = \int_a^b \left(\int_c^d f(x,y)dy\right) dx.$$

We take real numbers $\tilde{a} < a$, $b < \tilde{b}$, $\tilde{c} < c$, and $d < \tilde{d}$. We assume first that $f \geq 0$. For any given $\varepsilon > 0$, as above we have some $\delta > 0$ from the uniform continuity of f on $[a,b]\times[c,d]$ such that $|f(x,y)-f(\tilde{x},\tilde{y})|<\varepsilon$ for any two points (x,y) and $(\tilde{x},\tilde{y})$ of $[a,b]\times[c,d]$ with $\|(x,y)-(\tilde{x},\tilde{y})\|<\delta$. We take a partition P_1

$$\tilde{a}=x_0\leq x_1\leq\cdots\leq x_{n-1}\leq x_n=\tilde{b}$$

of $\left[\tilde{a},\tilde{b}\right]$ with mesh

$$\max_{0\leq j\leq n-1}\left(|x_{j+1}-x_j|\right)<\delta$$

such that both a and b belong to $\{x_j\}_{j=1}^{n-1}$. Let $a=x_{j_1}$ and $b=x_{j_2}$. We take a partition P_2

$$\tilde{c}=y_0\leq y_1\leq\cdots\leq y_{m-1}\leq y_m=\tilde{d}$$

of $\left[\tilde{c},\tilde{d}\right]$ with mesh

$$\max_{0\leq k\leq m-1}\left(|y_{k+1}-y_k|\right)<\delta$$

such that both c and d belong to $\{y_k\}_{k=1}^{m-1}$. Let $c=y_{k_1}$ and $d=y_{k_2}$. Denote by R_{jk} the closed rectangle $[x_j,x_{j+1}]\times[y_k,y_{k+1}]$ for $0\leq j\leq n-1$ and $0\leq k\leq m-1$. Denote by $|R_{jk}|$ is the area $(x_{j+1}-x_j)\times(y_{k+1}-y_k)$ of R_{jk}. Let

$$\bar{J}=\left\{\,(j,k)\,\middle|\,0\leq j\leq n-1\text{ and }0\leq k\leq m-1\text{ with }R_{jk}\cap[a,b]\times[c,d]\neq\emptyset\,\right\}$$

and

$$\underline{J}=\left\{\,(j,k)\,\middle|\,0\leq j\leq n-1\text{ and }0\leq k\leq m-1\text{ with }R_{jk}\subset[a,b]\times[c,d]\,\right\}.$$

Since the mesh of each of the two partitions P_1 and P_2 is $<\delta$, it follows that

$$\sum_{j\in\overline{J}-\underline{J}}|R_{jk}|\leq 2\delta\left((b-a)+2\delta\right)+2\delta\left((d-c)+2\delta\right).$$

Let $P=(P_1,P_2)$ and $E=[a,b]\times[c,d]$. Let A be the supremum of f on $[a,b]\times[c,d]$. Then the difference between the upper Riemann sum

$$U(f,E,P)=\sum_{(j,k)\in\overline{J}}\left(\sup_{R_{jk}}f\right)|R_{jk}|.$$

of f on E for $P=(P_1,P_2)$ and the lower Riemann sum

$$L(f,E,P)=\sum_{(j,k)\in\underline{J}}\left(\inf_{R_{jk}}f\right)|R_{jk}|.$$

of f on E for $P=(P_1,P_2)$ is no more than

$$2\delta\left((b-a)+2\delta\right)+2\delta\left((d-c)+2\delta\right).$$

times the supremum A of f on $[a,b]\times[c,d]$. Since

$$L(f,E,P)\leq\underline{\int}_E f\leq\overline{\int}_E f\leq U(f,E,P),$$

it follows that

$$\overline{\int}_E f-\underline{\int}_E f\leq 2A\delta\left((b-a)+(d-c)+4\delta\right).$$

Because δ can be chosen to be less than any prescribed positive number, we conclude that $\overline{\int}_E f=\underline{\int}_E f$ and $\int_E f$ exists. Since $\underline{J}$ consists of all (j,k) with $j_1\leq j<j_2$ and $k_1\leq k<k_2$, it follows that

$$L(f,E,P)\leq\sum_{\substack{j_1\leq j<j_2\\ k_1\leq k<k_2}} f\left(x_j,y_k\right)\left|R_{jk}\right|\leq U(f,E,P).$$

From

$$L(f,E,P)\leq\int_E f\leq U(f,E,P)$$

we know that the two numbers

$$\int_E f,\qquad\sum_{\substack{j_1\leq j<j_2\\ k_1\leq k<k_2}} f\left(x_j,y_k\right)\left|R_{jk}\right|$$

are both inside the closed interval with left end-point $L(f,E,P)$ and right end-point $U(f,E,P)$ whose length is no more than

$$2A\delta\left((b-a)+(d-c)+4\delta\right).$$

Hence

$$(\dagger\dagger)\qquad\left|\int_E f-\sum_{\substack{j_1\leq j<j_2\\ k_1\leq k<k_2}} f\left(x_j,y_k\right)\left|R_{jk}\right|\right|\leq 2A\delta\left((b-a)+(d-c)+4\delta\right).$$

For $x \in [a, b]$ let

$$H(x) = \int_c^d f(x, y)dy = \int_c^d G_x.$$

By (‡),

$$\left| H(x_j) - \sum_{k_1 \leq k < k_2} f(x_j, y_k)(y_{k+1} - y_k) \right| < \varepsilon(d - c)$$

for $j_1 \leq j < j_2$. By multiplying it by $(x_{j+1} - x_j)$ and then summing over $j_1 \leq j < j_2$, we obtain

$$\left| \sum_{j_1 \leq j < j_2} H(x_j)(x_{j+1} - x_j) - \sum_{k_1 \leq k < k_2} f(x_j, y_k) |R_{jk}| \right| < \varepsilon(b - a)(d - c).$$

By $(**)$ and (‡) (applied to $[a, b]$ instead of $[c, d]$ and to $3\varepsilon(d - c)$ instead of ε), we conclude that

$$\left| \int_a^b H(x)dx - \sum_{j_1 \leq j < j_2} H(x_j)(x_{j+1} - x_j) \right| < 3\varepsilon(b - a)(d - c).$$

Since

$$\int_a^b \left(\int_c^d f(x, y)dy \right) dx = \int_a^b H(x),$$

we have

$$\left| \int_a^b \left(\int_c^d f(x, y)dy \right) dx - \sum_{k_1 \leq k < k_2} f(x_j, y_k) |R_{jk}| \right| < 4\varepsilon(b - a)(d - c).$$

From (††) it follows that

$$\left| \int_E f - \int_a^b \left(\int_c^d f(x, y)dy \right) dx \right| \leq 4\varepsilon(b-a)(d-c) + 2A\delta\left((b - a) + (d - c) + 4\delta\right).$$

Since both ε and δ can be chosen to be less than any prescribed positive number, it follows that

$$\int_{[a,b]\times[c,d]} f = \int_a^b \left(\int_c^d f(x, y)dy \right) dx.$$

When f is not assumed to be nonnegative, we can write $f = f_1 - f_2$ with both f_1 and f_2 nonnegative and continuous on $[a, b] \times [c, d]$. From

$$\int_{[a,b]\times[c,d]} f_j = \int_a^b \left(\int_c^d f_j(x, y) dy \right) dx \quad \text{for} \quad j = 1, 2$$

and taking the difference of the two equations, we obtain

$$\int_{[a,b]\times[c,d]} f = \int_a^b \left(\int_c^d f(x, y) dy \right) dx.$$

This finishes the proof of Fubini's theorem.