Calculus (Winter) Notes

Camila Restrepo

Last updated February 10, 2024

Note: Theorem numbers come from the order they are presented in lecture, and do not correspond to any textbook or written course material.

Week 1

Classifying Critical Points

Theorem 1.1 2nd Derivative Test
Let $f \in C^2(\Omega)$ and let $a \in \Omega (\Omega \subseteq \mathbb{R}^n)$ be a critical point of f.

1. If $H_f(a)$ is positive definite then f has a local minimum at a.
2. If $H_f(a)$ is negative definite then f has a local maxiumum at a.
3. If $H_f(a)$ is indefinite then f has a saddle point at a.

Recall: Any symmetric $n \times n$ matrix A can be diagonalized, i.e., $\exists$ an orthonormal basis $u_1, u_2, \dots, u_n$ in $\mathbb{R}^n$ and real numbers $\lambda_1, \lambda_2, \dots, \lambda_n$ such that $Au_i = \lambda_i u_i \forall i = 1, 2, \dots, n$.

Proposition 1.2
Let Q be the quadratic form associated with an $n \times n$ symmetric matrix A. Then:

1. Q is positive $\iff$ all the eigenvalues of A are positive,
2. Q is negative $\iff$ all the eigenvalues of A are negative,
3. Q is indefinite $\iff$ A has both positive and negative eigenvalues.

Corollary 1.3
Let a be a critical point of a C^2 function $f : \Omega \to \mathbb{R}$. If $\det H_f(a) \neq 0$, then f has either a local minimum or a local mximum or a saddle point at a.

Definition of degenerate critical points:
A critical point a of a C^2 function f is called non-degenerate if $\det H_f(a) \neq 0$ and degenerate otherwise.

Example of a degenerate critical point:
When $f(x, y) = x^3$ then $(0, 0)$ is a degenerate critical point of f, and f has neither a local extremum at $(0, 0)$ nor a saddle point.

Definition of the principal minors of a matrix:
Let $A = (a_{ij})_{i,j=1}^n$ be an $n \times n$ matrix. Given $k = 1, 2, \ldots, n$, we will denote by A_k the $k \times k$ submatrix $A_k = (a_{ij})_{i,j=1}^k$.
The determinants $\det A_k$ are called the **principal minors of A**.

Proposition 1.4
Let A be a symmetric $n \times n$ matrix with $\det A \neq 0$. Then:

1. A is positive definite $\iff$ $\det A_k > 0 \forall k = 1, 2, \ldots, n$.
2. A is negative definite $\iff$ $(-1)^k \det A_k > 0 \forall k = 1, 2, \ldots, n$.
3. A is indefinite $\iff$ A is neither positive definite nor negative definite.

Corollary 1.5
Let $A = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}$. Then:

1. A is positive definite $\iff$ $\alpha > 0$ and $\alpha\gamma - \beta^2 > 0$
2. A is negative definite $\iff$ $\alpha < 0$ and $\alpha\gamma - \beta^2 > 0$
3. A is indefinite $\iff$ $\alpha\gamma - \beta^2 < 0$

Example of classifying a critical point:
We found that the function $f(x, y) = xye^{-x^2-y^2}$ has 5 critical points: $\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right)$, $\left(\pm\frac{1}{\sqrt{2}}, \mp\frac{1}{\sqrt{2}}\right)$, and $(0, 0)$, with an absolute maximum at $\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right)$ and an absolute minimum at $\left(\pm\frac{1}{\sqrt{2}}, \mp\frac{1}{\sqrt{2}}\right)$.

Investigate the nature of $(0,0)$,

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\left[y(1-2x^2)e^{-x^2-y^2}\right] = -4xye^{-x^2-y^2} - 2xy(1-2x^2)e^{-x^2-y^2}$$
$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left[x(1-2y^2)e^{-x^2-y^2}\right] = -4xye^{-x^2-y^2} - 2xy(1-2y^2)e^{-x^2-y^2}$$
$$\frac{\partial^2 f}{\partial x \partial y} = (1-2y^2)e^{-x^2-y^2} - 2x^2(1-2y^2)e^{-x^2-y^2}$$

So $H_f(0,0) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is indefinite $\implies$ f has a saddle point at (0,0).

Example of non-degenerate critical points:
Find and classify the critical points of $f : \mathbb{R}^3 \to \mathbb{R}$ where $f(x,y,z) = x^3 - y^3 + 3xy + z^2 - 2z$.
$\nabla f = (3x^2+3y, -3y^2+3x, 2z-2) = 0 \implies x^2 = -y, y^2 = x, z = 1 \implies x^2 = -y, x^4 = x, z = 1 \implies x^2 = -y, x = 0$ or $x = 1, z = 1 \implies (x,y,z) = (0,0,1), (1,-1,1)$.

So $(0,0,1)$ and $(1,-1,1)$ are the critical points. We have $H_f(x,y,z) = \begin{bmatrix} 6x & 3 & 0 \\ 3 & -6y & 0 \\ 0 & 0 & 2 \end{bmatrix}$,
so $H_f(0,0,1) = \begin{bmatrix} 0 & 3 & 0 \\ 3 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ is clearly indefinite since the first principal minor is 0 and $H_f(1,-1,1) = \begin{bmatrix} 6 & 3 & 0 \\ 3 & 6 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ is positive definite.
So we have non-degenerate critical points (as $\det H_f \neq 0$). Hence, $(0,0,1)$ is a saddle point; $(1,-1,1)$ is a local minimum.
But f has no global extrema because $f(x,0,0) = x^3$ can take arbitrarily positive and negative values.

Example of a degenerate critical point:
Let $f(x,y) = x^4 + y^4$ (with $(x,y) \in \mathbb{R}^2$).
$\nabla f = (4x^3, 4y^3) = 0 \iff (x,y) = (0,0)$.
$H_f(x,y) = \begin{bmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{bmatrix}$, $H_f(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.
So $(0,0)$ is a degenerate critical point and the 2nd derivative test does not apply.
However, f has a global minimum at $(0,0)$.

Week 2

Inverse Function Theorem and Implicit Function Theorem

Theorem 2.1
Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ is a continuous injective function. Then:

1. f is either strictly increasing or strictly decreasing.
2. $f(I)$ is an interval containing the same number of endpoints as I.
3. f is a homeomorphism of I onto $f(I)$.

Proof. 1. Let us first consider the case that $I = [a, b] (a < b)$. Since f is injective, either $f(a) < f(b)$ or $f(b) < f(a)$. Assume that $f(a) < f(b)$ (the other case can be done symmetrically). Let's show that f is strictly increasing on $[a, b]$, i.e., $f(x) < f(y)$ whenever $a \leq x < y \leq b$. We argue by contradiction, supposing that $f(x) > f(y)$ for some $a \leq x < y \leq b$.
Note that $f(y) > f(a)$, for otherwise $f(y) < f(a) < f(b)$ and by the Intermediate Value Theorem (IVT), $\exists \alpha \in (y, b)$ such that $f(\alpha) = f(a)$, contradicting the injectivity of f. Therefore $f(a) < f(y) < f(x)$ and so, again, by the IVT $\exists y' \in (a, x)$ such that $f(y') = f(y)$, again contradicting the injectivity of f.
Next, let I be any interval. Pick up any $a, b \in I$ with $a < b$. Suppose that $f(a) < f(b)$ (the case $f(a) > f(b)$ can be done symmetrically). By the previous paragraph, we know that f is strictly increasing on $[a, b]$. Now, if $x, y \in I$ and $x < y$, then with $\alpha = \min\{a, x\}, \beta = \max\{y, b\}$, we have $[a, b], [x, y] \subseteq [\alpha, \beta] \subseteq I$. Since f is strictly increasing on $[a, b]$, we must have (using the 1st paragraph again) $f(\alpha) < f(\beta)$ and f is strictly increasing on $[\alpha, \beta]$. Hence, we conclude that f is strictly increasing on I.

2. Since f is continuous, $J = f(I)$ is an interval. Suppose that f is strictly increasing. Note that the inverse function f^{-1} is then also strictly increasing.
Now, if I contains its left endpoint a, then $\forall x \in I$, $f(a) \leq f(x)$, so $f(a)$ is a left endpoint of J. Similarly, if I contains its right endpoint b, then $f(b)$ is the right endpoint of J. Applying the same argument with f^{-1} in place of f, we conclude if I contains its left (respectively, right) endpoint c, then $f^{-1}(c)$ is the left (respectively, right) endpoint of I. It follows that I and J contain the same number of endpoints.

3. If $I = [a, b]$, then f is a homeomorphism of I onto $f(I)$ because of our general result about continuous injective functions on compact sets.
Otherwise, it follows that $f|[a, b]$ is a homeomorphism onto $f([a, b])$ for any $a, b \in I$ with $a \leq b$. This implies that $f^{-1} : f(I) \to I$ is continuous (at any $y \in f(I)$).
Indeed, let $y \in f(I)$ and consider any sequence (y_n) in $f(I)$ with $y_n \to y$. Then the set $S = \{y\} \cup \{y_n : n \in N\}$ is compact, so it has both a smallest element $c = f(a)$ and a largest element $d = f(b)$. Assuming that f is strictly increasing we must have $a \leq b$, and $f([a, b]) = [c, d] \supseteq S$. Since $f|[a, b]$ is a homeomorphism onto $[c, d]$ (i.e., $(f|[a, b])^{-1} = f^{-1}|[c, d]$ is continuous), we obtain $f^{-1}(y_n) = (f^{-1}|[c, d])(y_n) \to (f^{-1}|[c, d])(y) = f^{-1}(y)$. It follows that f^{-1} is continuous at any $y \in f(I)$.

$\square$

Theorem 2.2
Let f be a bijection of a non-zero interval $I \subseteq \mathbb{R}$ onto an interval $J \subseteq \mathbb{R}$. If f is differentiable at $a \in I$, $f'(a) \neq 0$, and f^{-1} is continuous at $f(a)$ and $(f^{-1})'(f(a)) = \frac{1}{f'(a)}$

(Sketch). $\square$

Definition of a diffeomorphism:
Let f be a bijection of an open subset $U \subseteq \mathbb{R}^n$ onto an open subset $V \subseteq \mathbb{R}^n$. If both f and f^{-1} are differentiable (on U and V respectively), then f is called a **diffeomorphism** of U onto V. If both f and f^{-1} are C^k functions ($k = 1, 2, \ldots, \infty$), then f is called a **diffeomorphism of class** C^k.

Corollary 2.3
Let f be a differentiable homeomorphism of an open subset $U \subseteq \mathbb{R}$ onto an open subset $V \subseteq \mathbb{R}$. If $f'(a) \neq 0$ for all $a \in U$, then f is a diffeomorphism of U onto V. Moreover, if $f \in C^k(U)$, then f is a C^k diffeomorphism.

Proof. If $b = f(a) \in V$ (where $a \in U$), then there exists an open interval $I \subseteq U$ such that $a \in I$. Then $f(I)$ is another open interval and $f|I$ is a homeomorphism onto $f(I)$ (by the Inverse Function Theorem), and $f|I$ satisfies the assumptions of the above theorem. Hence, $(f|I)^{-1} = f^{-1}|f(I)$ is differentiable at b. But this means that f^{-1} is differentiable at b. Since $b \in V$ is artbitrary, f^{-1} is differentiable on V and so f is a diffeomorphism.
We also have $\left(f^{-1}\right)'(b) = \frac{1}{f^{-1}(a)} = \frac{1}{f'(f^{-1}(b))}$ for any $b = f(a) \in V$. Thus, $\left(f^{-1}\right)' = \frac{1}{f'} \circ f^{-1}$. That f^{-1} is C^k when f is C^k follows by induction on $k = 1, 2, \ldots$: When $k = 1$, then $\frac{1}{f'}$ is continuous (as $f \in C^1(U)$), and f^{-1} is continuous, so $\left(f^{-1}\right)' = \frac{1}{f'} \circ f^{-1}$ is continuous. Assuming that our claim is true for C^k functions, consider $f \in C^{k+1}(U)$. Then $f' \in C^k(U)$, and as $f \in C^k(U)$, $f^{-1} \in C^k(V)$ by induction. Hence, $\left(f^{-1}\right)' = \frac{1}{f'} \circ f^{-1}$ is a C^k function as the composition of two C^k functions. Therefore $f^{-1} \in C^k(V)$ $\square$

Corollary 2.4 Inverse Function Theorem in 1 variable
Let $I \subset \mathbb{R}$ be an open interval and $f : I \to \mathbb{R}$ a C^k function such that $f'(x) \neq 0$ for all $x \in I$. Then f is a C^k diffeomorphism of I onto $f(I)$.

Proof. By the IVT either $f'(x) > 0$ for all $x \in I$ (i.e., f is strictly increasing) or $f'(x) < 0$ for all $x \in I$ (i.e., f is strictly decreasing). Hence, f is injective and is a homeomorphism of I onto an open interval J. The assumption of the previous corollary are satisfied, hence the conclusion. □

Corollary 2.5 Inverse Function Theorem in 1 variable, local version
Let $U \in \mathbb{R}$ be open and $f : U \to \mathbb{R}$ be a C^k function. If $f'(a) \neq 0$ at some $a \in U$, then there exists an open interval I such that $a \in I \subseteq U$ and $f|I$ is a C^k diffeomorphism of I onto $f(I)$

How do these results generalize to functions of n variables?

Theorem 2.6
Let $\Omega \subseteq \mathbb{R}^n$ be an open set and let $f : \Omega \to \mathbb{R}^n$ be injective. Then $f(\Omega)$ is open and f is a homeomorphism of Ω onto $f(\Omega)$.

Proof. Omitted due to high difficulty. □

Lemma 2.7
If $T : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation then there exists a $c > 0$ such that for all $x \in \mathbb{R}^n$, $\|T(x)\| \geq C\|x\|$

Proof. Recall that T^{-1} is a Lipschitz function, i.e., there exists $M > 0$ such that $\|T^{-1}(x)\| \leq M\|x\|$ for all $x \in \mathbb{R}^n$. Hence, for all $x \in \mathbb{R}^n$, $\|x\| = \|T^{-1}(T(x))\| \leq M\|T(x)\|$, so $\|T(x)\| \geq \frac{1}{M}\|x\|$. □

Theorem 2.8
Let f be a bijection of an open subset $U \subseteq R^n$ onto an open subset $V \in \mathbb{R}^n$. If f is differentiable at $a \in U$, $\det(D_f(a)) \neq 0$, and f^{-1} is continuous at $b = f(a)$, then f^{-1} is differentiable at b and $D_{f^{-1}}(b) = (D_f(a))^{-1}$.

Proof. Let $T = D_f(a)$, $b = f(a)$. It suffices to show that

$$\lim_{y\to b} \frac{f^{-1}(y) - f^{-1}(b) - T^{-1}(y-b)}{\|y-b\|} = 0$$

But,

$$\frac{f^{-1}(y) - f^{-1}(b) - T^{-1}(y-b)}{\|y-b\|} = -T^{-1}\left(\frac{y-b-T\left(f^{-1}(y)-f^{-1}(b)\right)}{\|y-b\|}\right)$$

So it suffices to show that

$$\lim_{y\to b} \frac{y-b-T\left(f^{-1}(y)-f^{-1}(b)\right)}{\|y-b\|} = 0$$

and this will be done if we show that

$$\lim_{k\to\infty} \frac{y_k-b-T\left(f^{-1}(y_k)-f^{-1}(b)\right)}{\|y_k-b\|} = 0$$

For every sequence $(y_k) \in V\setminus\{b\}$ with $y_k - b$. Let $x_k = f^{-1}(y_k) \in U\setminus\{a\}$ (i.e., $y_k = f(x_k)$). Then $x_k \to f^{-1}(b) = a$ because f^{-1} is continuous at b. Thus we need to show that

$$\lim_{k\to\infty} \frac{f(x_k) - f(a) - T(x_k - a)}{\|f(x_k) - f(a)\|} =$$

$$\lim_{k\to\infty} \left[\frac{\|x_k - x\|}{\|f(x_k) - f(a)\|} \frac{f(x_k) - f(a) - T(x_k - a)}{\|x_k - a\|}\right] = \lim_{k\to\infty} A_k B_k = 0$$

Now, as $T = D_f(a)$, $\lim_{k\to\infty} B_k = 0$ (by the definition of the derivative). So to complete the proof it is enough to show that the sequence (A_k) is bounded. But

$$\frac{1}{A_k} = \|\frac{f(x_k) - f(a) - T(x_k - a)}{\|x_k - a\|} + T\left(\frac{x_k - a}{\|x_k - a\|}\right)\| =$$

$$\|B_k + T\left(\frac{x_k - a}{\|x_k - a\|}\right)\| \geq \|T\left(\frac{x_k - a}{\|x_k - a\|}\right)\| - \|B_k\|$$

and by the lemma, there exists a $c > 0$ such that $\|T\left(\frac{x_k-a}{\|x_k-a\|}\right)\| \geq c$ for all k. As $B_k \to 0$, there exists a k_0 such that for all $k > k_0$ $\frac{1}{A_k} \geq \frac{c}{2}$ and so for all $k \in \mathbb{N}$ $\frac{1}{A_k} \geq \min\left\{\frac{c}{2}, \frac{1}{A_1}, \frac{1}{A_2}, \ldots, \frac{1}{A_{k_0}}\right\} > 0$. Hence, (A_k) is bounded. $\square$

Corollary 2.9
Let f be a differentiable homeomorphism of an open subset $U \subseteq \mathbb{R}^n$ onto an open subset $V \subseteq \mathbb{R}^n$. If $\det(D_f(x)) \neq 0$ for all $x \in U$, then f is a diffeomorphism of U onto V. Moreover, if $f \in C^k(U)$ then f is a C^k diffeomorphism.

Proof. Clearly, the assumptions of the previous theorem are satisfied for each $a \in U$, so f^{-1} is differentiable at each $b = f(a)$, and f is thus a diffeomorphism of U onto V. □

Remark: The following example shows that the 1-dimensional Inverse Function Theorem cannot be generalized to n-dimensions.

Example of Polar Coordinate Mapping:
Let $f : (0, \infty) \times \mathbb{R}$ be given by $f(s,t)$

Theorem 2.10 Inverse Function Theorem (IFT)
Let $f : \Omega \to \mathbb{R}^n$ be a C^k function where $\Omega \subseteq \mathbb{R}^n$ is open (and $k = 1, 2, \ldots, \infty$). If $\det(D_f(a)) \neq 0$ for some $a \in \Omega$, then there exists an open set $U \in \Omega$ with $a \in U$ and an open set $V \subseteq \mathbb{R}^n$ with $f(a) \in V$ such that $f|U$ is a C^k diffeomorphism of U onto V.

Corollary 2.11 Open Mapping Theorem
Let $F : \Omega \to \mathbb{R}^n$ be C^1 function where $\Omega \subseteq \mathbb{R}^n$ is open. If $\det(D_f(x)) \neq 0$ for all $x \in \Omega$, then f is an open wrapping, i.e., for every open subset $W \subseteq \Omega$, $f(W)$ is open in $\mathbb{R}^n$.

Proof. Let $W \subseteq \Omega$ be open. To conclude that $f(W)$ is open, it suffices to show that for all $b \in f(W)$ there exists an open V such that $b \in V \subseteq f(W)$. But $b = f(a)$ for some $a \in W$ and $f|W$ and $a \in W$ satisfy the assumption of the IFT. Thus, there exists open $U \subseteq W$ and open $V \subseteq \mathbb{R}^n$ such that $a \in U$, $b \in V$ and $f(U) = (f|W)(U) = V$. Clearly, $b \in V \subseteq f(W)$. □

Corollary 2.12
Let $f : \Omega \to \mathbb{R}^n$ bw a C^k function where $\Omega \to \mathbb{R}^n$ is open. If f is injective and $\det(D_f(x)) \neq 0$ for all $x \in \Omega$, then $f(\Omega)$ is open and f is a C^k diffeomorphism of Ω onto $f(\Omega)$.

Proof. By a previous corollary, it suffices to show that $f(\Omega)$ is open and f is a homeomorphism of Ω onto $f(\Omega)$. But by the previous corollary, f is an open mapping, so, in particular, $f(\Omega)$ is open. Thus, it remains to prove that $f^{-1} : f(\Omega) \to \Omega$ is continuous. Recall that this will be true if for each open $U \subseteq R^n$, $\left(f^{-1}\right)^{-1}(U)$ is open relative to $f(\Omega)$, i.e., is open in $\mathbb{R}^n$ because $f(\Omega)$ is open. But $\left(f^{-1}\right)^{-1}(U) = \left(f^{-1}\right)^{-1}(U \cap \Omega) = f(U \cap \Omega)$ is indeed open in R^n by the Open Mapping Theorem. □

Example of determining a diffeomorphism:
The polar coordinate mapping $f(r,\theta) = (rcos\theta, rsin\theta)$ (considered on $(0,\infty) \times \mathbb{R}$), is an open mapping of $(0,\infty) \times \mathbb{R}$ onto $\mathbb{R}^2 \setminus \{(0,0)\}$ because $\det(D_f(r,\theta)) = r > 0$ for all $(r,\theta) \in (0,\infty) \times \mathbb{R}$.
Note that $\varphi = f|((0,\infty) \times (-\pi,\pi))$ is injective. Hence, by the last corollary φ is a C^∞ diffeomorphism on $(0,\infty) \times (-\pi,\pi)$ onto $\varphi((0,\infty) \times (-\pi,\pi)) = \mathbb{R}^2 \setminus ((-\infty,0] \times \mathbb{R})$.

$$D_{\varphi^{-1}}(rcos\theta, rsin\theta) = \begin{bmatrix} cos\theta & -rsin\theta \\ sin\theta & rcos\theta \end{bmatrix}^{-1} = \frac{1}{r}\begin{bmatrix} rcos\theta & rsin\theta \\ -sin\theta & cos\theta \end{bmatrix}$$

Similarly $\varphi|((0,\infty) \times (a,b))$, where $b - a = 2\pi$ is a c^∞ diffeomorphism on $(0,\infty) \times (a,b)$ onto $\mathbb{R}^2 \setminus \{r(cos\theta, sin\theta) : r \geq 0\}$.

Definition of an implicit function:
Let $\Omega_n \subseteq \mathbb{R}^n$, $\Omega_m \subseteq \mathbb{R}^m$, $F : \Omega_n \times \Omega_m \to \mathbb{R}^m$, and $c \in \mathbb{R}^m$.
Consider the equation

$$F(x,y) = c \qquad (x \in \Omega_n, y \in \Omega_m)(*)$$

which we suppose needs to solved for y. If for every $x \in \Omega_n$ this equation has a solution, then by choosing for each $x \in \Omega_n$ a solution $y \in \Omega_m$ and calling it $f(x)$, we obtain a function $f : \Omega_n \to \Omega_m$ such that $F(x.f(x)) = c$ for all $x \in \Omega_n$. Any such function is called an **implicit function** defined by Eq. $(*)$.

Note: If for all $x \in \Omega_n$ there exists a unique $y \in \Omega_m$ such that $F(x,y) = c$, then Eq. $(*)$ defines a unique implicit function, but in general, implicit functions are not unique.

Example of:
Let $n = m = 1$, $\Omega_n = \Omega_m = [-1,1]$, $F(x,y) = x^2+y^2$, $c = 1$. Then the functions $f_\pm(x) = \pm\sqrt{1-x^2}$ are implicit functions defined by $(*)$ (i.e., eg. $x^2 + y^2 = 1$) and there are many other implicit functions.
If we replace Ω_m by $[0,1]$, then f_+ will be the unique implicit function defined by $(*)$ $(f_+(x) = \sqrt{1-x^2})$.

Question
Under what conditions does an implicit function exist; is unique; is it differentiable? If it is differentiable how can we obtain its derivative?

Note: Let $F : \Omega \to \mathbb{R}^m$ be a C^k function where $\Omega \subseteq \mathbb{R}n + m = \mathbb{R}^n \times \mathbb{R}^m$ is open. We will write the elements of $\mathbb{R}^n + m = \mathbb{R}^n \times \mathbb{R}^m$ as (x, y) where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Then

$$D_f(x,y) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(x,y) & \dots & \frac{\partial F_1}{\partial x_n}(x,y) & \frac{\partial F_1}{\partial y_1}(x,y) & \dots & \frac{\partial F_1}{\partial y_m}(x,y) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1}(x,y) & \dots & \frac{\partial F_m}{\partial x_n}(x,y) & \frac{\partial F_m}{\partial y_1}(x,y) & \dots & \frac{\partial F_m}{\partial y_m}(x,y) \end{bmatrix}$$

with the first $m \times n$ block will be named $\frac{\partial F}{\partial x}(x, y)$ and the second $m \times m$ block will be named $\frac{\partial F}{\partial y}(x, y)$.
Thus, we can write $D_F(x, y) = \begin{bmatrix} \frac{\partial F}{\partial x}(x,y) & \frac{\partial F}{\partial y}(x,y) \end{bmatrix}$

Theorem 2.13 Implicit Function Theorem (IPFT)
Let $F : \Omega \to \mathbb{R}^m$ be a C^k function where $\Omega \subseteq \mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$ is open. Suppose that for $(a, b) \in \Omega$ and $c \in \mathbb{R}^m$, $F(a, b) = c$ and $\det\left(\frac{\partial F}{\partial y}(a,b)\right) \neq 0$. Then there exist open sets $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ that satisfy:

1. $(a, b) \in U \times V$,
2. for all $x \in U$, there exists a unique $y \in V$ such that $F(x, y) = c$.

Moreover, the unique implicit function $f : U \to V$ defined by the equation $F(x, f(x)) = c$ $(x \in U, y \in V)$ is a C^k function.

Proof. Define $G : \Omega \to \mathbb{R}^{n+m}$ by $G(x,y) = (x, F(x.y))$. This is a C^k function, $G(a,b) = (a,c)$ and

$$D_G(x,y) = \begin{bmatrix} I_n & 0 \\ \frac{\partial F}{\partial x}(x,y) & \frac{\partial F}{\partial x}(x,y) \end{bmatrix}$$

Thus $\det(D_G(a,b)) = (\det I_n)\left(\det\left(\frac{\partial F}{\partial y}(a,b)\right)\right) \neq 0$.
Thus by the IFT, there exists an open subset $\Omega_1 \subseteq \Omega$ with $(a,b) \in \Omega_1$ and an open subset $\Omega \subseteq \mathbb{R}^{n+m}$ with $(a,c) = G(a,b) \in W$ such that $G|\Omega_1$ is a C^k diffeomorphism of Ω_1 onto W. Let $H = (G|\Omega_1)^{-1} : W \to \Omega_1$. Then $H(x,y) = (j(x,y), k(x,y))$ where $j : W \to \mathbb{R}^n$ and $k : W \to \mathbb{R}^m$ are C^k functions. Note that $(x,y) = G(H(x,y)) = (j(x,y), F(k(x,y)))$ for all $(x,y) \in W$. Hence, $j(x,y) = x$ and $F(k(x,y)) = y$ for all $(x,y) \in W$. Thus $H(x,y) = (x, k(x,y))$ and so for all $(x,y) \in W$,

$$(x, k(x,y)) \in \Omega_1 \text{ and } F(x, k(x,y)) = y$$

Note that we may assume that $\Omega_1 = U' \times V$ where $U' \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ are open. [Indeed, $(a,b) \in \Omega_1$ and Ω_1 is open, so there exists an $r > 0$ such that $B_r^{n+m}(a,b) \in \Omega_1$. But $B_r^{n+m}(a,b) \supseteq B_{\frac{r}{2}}^n(a) \times B_{\frac{r}{2}}^m(b)$. So we can take $U' = B_{\frac{r}{2}}^n(a)$, $V = B_{\frac{r}{2}}^m(b)$ and replace Ω_1 with $U' \times V$ and W with $G(U' \times V)$].
Moreover, since $(a,c) \in W$, we can find an open set U such that $a \in U \subseteq U'$ and $U \times \{c\} \subseteq W$. Then for all $x \in U$, $(x,c) \in W$ and so $F(x, k(x,c)) = c$. Thus when $f : U \to V$ is given by $f(x) = k(x,c)$, then f is an implicit funciton defined by the equation $F(x,y) = c$ (for $x \in U$, $y \in V$). It is clear that f is a C^k function.
It remains to confirm that for all $x \in U$ there exists a unique $y \in V$ such that $F(x,y) = c$. But if $y_1, y_2 \in V$ and $F(x,y_1) = c = f(x,y_2)$, then $G(x,y_1) = (x,c) = G(x,y_2)$, and so $y_1 = y_2$ as $G|U \times V$ is injective. $\square$

Week 3

IPFT Practice and Constraints

Corollary 3.1
With the assumptions and notation of the IPFT, let $S = \{(x,y) \in \Omega : F(x,y) = c\}$. Then $S \cap (U \times V) = \{(x,y) \in \mathbb{R}^{n+m} : x \in U \text{ and } y = f(x)\}$.

Remark: Note that when $m = 1$, then $\det\left(\frac{\partial F}{\partial y}\right) = \frac{\partial F}{\partial y}$. So if $\frac{\partial F}{\partial y}(a,b) \neq 0$ then the level set $S = \left\{(x,y) \in \mathbb{R}^{n+1} : F(x,y) = c\right\}$ in a neighbourhood of (a,b) is the graph of the implicit function.

Example of:
(IPFT, level set, and graph) Consider the level set $S = \left\{(x,y) \in \mathbb{R}^2 : x^3y^2 + y^3(x-1)^2 = 1\right\}$ of $F(x,y) = x^3y^2 + y^3(x-1)^2$.

1. Show that S is not the graph of any function $y = f(x)$, i.e., $S \neq \left\{(x,y) \in \mathbb{R}^2 : y = f(x)\right\}$.

2. Show that in a neighbourhood of $(1,1)$, S is the graph of a smooth function f and find the slope of the tangent line to the graph of f at $(1,1)$.

Solutions:

1. $(1,-1), (1,1) \in S$, so no such function exists.

2. $\frac{\partial F}{\partial y}(1,1) = 2x^3y + 3^2(x-1)^2\Big|_{x=1,y=1} = 2 \neq 0$. So by the IPFT (with $a = b = c = 1$) and the corollary there exist open sets $U, V \in \mathbb{R}$ with $(1,1) \in U \times V$ and a smooth function $f : U \to V$ such that $f(1) = 1$, $F(x, f(x)) = 1 = x^3 f(x)^2 + f(x)^3 (x-1)^2 = 1$ for all $x \in U$, and $S \cap (U \times V) = \{(x,y) : x \in U \text{ and } y = f(x)\}$.

The slope is $f^{-1}(1)$: Since $x^3 f(x)^2 + f(x)^3 (x-1)^2 = 1$ for all $x \in U$, so $0 = \frac{\mathrm{d}}{\mathrm{d}x}\left[x^3 f(x)^2 + f(x)^3 (x-1)^2\right] = 3x^2 f(x)^2 + 2x^3 f(x) f'(x) + 3f(x)^2 f'(x)(x-1)^2 + 2f(x)^3(x-1)$. When $x = 1$, $f(1) = 1$, and so $0 = 3 + 2f'(1)$. Thus $f'(1) = \frac{3}{2}$

Example of:
(Finding the derivative without the function) Consider the problem of solving the system of equations: $\begin{cases} xy^2 + xzu + yv^2 = 3 \\ u^3yz + 2xv - u^2v^2 = 2 \end{cases}$ (*). for u and v in terms of x, y, z near $x = y = z = u, v = 1$ and computing the partial $\frac{\partial u}{\partial z}, \frac{\partial v}{\partial z}$.

Let $a = (1,1,1), b = (1,1), c = (3,2)$, and $F : \mathbb{R}^3 \to \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$F(x,y,z,u,v) = (xy^2 + xzu + yu^2, u^3yz + 2xv - u^2v^2).$$

Then $F(a,b) = c$, $\frac{\partial F}{\partial(u,v)} = \begin{bmatrix} xz & 2yv \\ 3u^2yz - 2uv^2 & 2x - 2u^2v \end{bmatrix}$.

$$\det\left(\frac{\partial F}{\partial(u,v)}(a,b)\right) = \det\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = -2 \neq 0.$$

Hence, by the IPFT there exists a smooth function $f(x,y,z) = (f_1(x,y,z), f_2(x,y,z))$ defined on a neighbourhood U of $u = (1,1,1)$ such that $F(x,y,z,f(x,y,z)) = (3,2) = c$ for all $(x,y,z) \in U$ and $f(1,1,1) = (1,1)$: $u = f_1(x,y,z), v = f_2(x,y,z)$ are the expressions of u and v in terms of x,y,z. To find $\frac{\partial u}{\partial z}$ and $\frac{\partial v}{\partial z}$ we differentiate Eqs(*) with respect to z, treating u and v as functions of x,y,z:

$$0 = \frac{\partial}{\partial z}\left(xy^2 + xzu + yv^2\right) = xu + xz\frac{\partial u}{\partial z} + 2yv\frac{\partial v}{\partial z}$$
$$0 = \frac{\partial}{\partial z}\left(u^3yz + 2xv - u^2v^2\right) = 3u^2\frac{\partial u}{\partial z}yz + u^3y + 2x\frac{\partial v}{\partial z} - 2u\frac{\partial u}{\partial z}v^2 - u^2 2v\frac{\partial v}{\partial z}$$

With $(x,y,z) = (1,1,1)$, $(u,v) = (1,1)$ we get

$$1 + \frac{\partial u}{\partial z} + 2\frac{\partial v}{\partial z} = 0, \frac{\partial u}{\partial z} + 1 = 0.$$

Hence, $\frac{\partial f_1}{\partial z} = \frac{\partial u}{\partial z}(1,1,1) = -1$, $\frac{\partial f_2}{\partial z} = \frac{\partial v}{partial z}(1,1,1) = 0.$

Proposition 3.2 Implicit Differentiation
Let $F : \Omega_n \times \Omega_m \to \mathbb{R}^m$ be a C^1 function where $\Omega_n \subset \mathbb{R}^n$ and $\Omega_m \subset \mathbb{R}^m$ are open and let $c \in \mathbb{R}^m$. If $f : \Omega_n \to \Omega_m$ is a differentiable function such that $F(x, f(x)) = c$ for all $x \in \Omega_n$, then

$$\frac{\partial F}{\partial y}(x, f(x)) D_f(x) = -\frac{\partial F}{\partial x}(x, f(x))$$

and

$$D_f(x) = -\left[\frac{\partial F}{\partial y}(x, f(x))\right]^{-1} \frac{\partial F}{\partial x}(x, f(x))$$

provided $\det\left(\frac{\partial F}{\partial y}(x, f(x))\right) \neq 0$.

Proof. Define $g : \Omega_n \to \Omega_n \times \Omega_m$ by $g(x) = (x, f(x))$. Then g is differentiable and

$$D_g(x) = \begin{bmatrix} I_n \\ D_f(x) \end{bmatrix}.$$

Since $(F \circ g)(x) = c$, the chain rule yields $0 = D_{F\circ g}(x) = D_F(g(x)) D_g(x) = \begin{bmatrix} \frac{\partial F}{\partial x}(g(x)) & \frac{\partial F}{\partial y}(g(x)) \end{bmatrix} \begin{bmatrix} I_n \\ D_f(x) \end{bmatrix} = \frac{\partial F}{\partial x}(x, f(x)) + \frac{\partial F}{\partial y}(x, f(x)) D_f(x)$.
Hence, the result. □

3.1 Constrained Extrema and Lagrange Multipliers

Let $\Omega \subseteq \mathbb{R}^n$ be open and $f, g_1, g_1, \ldots, g_m : \Omega \to \mathbb{R}$ be C^1 functions. Suppose that for some $c_1, c_2, \ldots, c_m \in \mathbb{R}$, $S = \{x \in \Omega : g_1(x) = c_1, g_2(x) = c_2, \ldots, g_m(x) = c_m\} \neq \emptyset$. The problem of finding the extreme values of f on the set S (i.e., the extrema of $f|S$) is referred to as the problem of of finding the extreme values of f subject to (or with) the constraints $g_1(x) = c_1, \ldots, g_m(x) = c_m$.

E.g., finding the extreme values of $f(x, y, z) = \sin(x + y)\cos(y + z)$ subject to the constraint $g(x, y, z) = x^2 + y^2 + z^2 = 1$ means finding the extreme values of f on the sphere $S_1(0, 0, 0) = \left\{(x, y, z) : x^2 + y^2 + z^2 = 1\right\}$.

Theorem 3.3
Let $f, g : \Omega \to \mathbb{R}$ be C^1 functions where $\Omega \subseteq \mathbb{R}^{n+1}$ and let $S = \{x \in \Omega : g(x) = c\}$ (where $c \in \mathbb{R}$). If $f|S$ attains an extreme value at some $s \in S$ where $\nabla g(s) \neq 0$, then there exists an $x \in \mathbb{R}$ (called a Lagrange multiplier) such that $\nabla f(s) = \lambda \nabla g(s)$.

Proof. Since $\nabla g(s) \neq 0$, $\frac{\partial g}{\partial x_i}(s) \neq 0$ for some $i = 1, 2, \ldots, n+1$. Let us first consider the case that $\frac{\partial g}{\partial x_{n+1}}(s) \neq 0$. Let $a = (s_1, \ldots, s_n)$, $b = s_{n+1}$ (so $s = (a, b)$). Then $g(a, b) = c$ and $\frac{\partial g}{\partial x_{n+1}}(a, b) \neq 0$. Hence by the IPFT there exist open sets $U \in \mathbb{R}^n$, $V \in \mathbb{R}$ such that $s = (a, b) \in U \times V$ and a C^1 function $\varphi : U \to V$ such that $\varphi(a) = b$ and $g(x, \varphi x) = c$ (i.e., $(x, \varphi(x)) \in S$) for all $x \in U$.
Define $\tilde{f} : U \to \mathbb{R}$ by $\tilde{f}(x) = f(x, \varphi(x))$. Clearly, $\tilde{f}$ is a C^1 function and $\tilde{f}$ has an extremum at $x = a$, so $\nabla \tilde{f}(a) = 0 = D_{\tilde{f}}(a)$. Note that $\tilde{f} = f \circ h$ where $h : U \to S \subseteq \mathbb{R}^{n+1}$ is given by $h(x) = (x, \varphi(x))$. Hence, by the Chain Rule

$$0 = D_{\tilde{f}}(a) = D_f(h(a)) D_h(a) = D_f(s) \begin{bmatrix} I_n \\ D_\varphi(a) \end{bmatrix}$$

or

$$0 = \frac{\partial f}{\partial x_i}(s) + \frac{\partial f}{\partial x_{n+1}} + \frac{\partial \varphi}{\partial x_i}(a) \, \forall i = 1, 2, \ldots, n.$$

But by the Implicit Differentiation Formula,

$$\begin{aligned} D_\varphi(a) &= -\left[\frac{\partial g}{\partial x_{n+1}}(a, \varphi a)\right]^{-1} \left[\frac{\partial g}{\partial x_1}(a, \varphi(a)), \ldots, \frac{\partial g}{\partial x_n}(a, \varphi a)\right] \\ &= -\left[\frac{\partial g}{\partial x_{n+1}}(s)\right]^{-1} \left[\frac{\partial g}{\partial x_1}(a, \varphi(a)), \ldots, \frac{\partial g}{\partial x_n}(s)\right] \end{aligned}$$

Therefore,

$$0 = \frac{\partial f}{\partial x_1}(s) - \frac{\partial f}{\partial x_{n+1}}(s) \left(\frac{\partial g}{\partial x_{n+1}}(s)\right)^{-1} \frac{\partial g}{\partial x_i}(s) \, \forall i = 1, 2, \ldots, n$$

Note that this equality also trivially holds when $i = n+1$. Thus, with $\lambda = \frac{\partial f}{\partial x_{n+1}}(s) \left(\frac{\partial g}{\partial x_{n+1}}(s)\right)^{-1}$ we obtain $\nabla f(s) = \lambda \nabla g(s)$.
If $\frac{\partial g}{\partial x_{n+1}}(s) = 0$, we can choose $p = 1, 2, \ldots, n$ such that $\frac{\partial g}{\partial x_p}(s) \neq 0$. Define a linear isomorphism $T : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ by $T(x_1, x_2, \ldots, x_{n+1}) = (x_1, x_2, \ldots, x_{p-1}, x_{n+1}, x_p, x_{p+1}, \ldots, x_n)$, and let $\Omega_* = T^{-1}(\Omega)$, $S_* = T^{-1}(S)$, $s_* = T^{-1}(s)$, $f_* = f \circ T : \Omega_* \to \mathbb{R}$, $g_* = g \circ T : \Omega_* \to \mathbb{R}$. Then $S_* = \{x \in \Omega_* : g_*(x) = c\}$ and $f_* | S_*$ has an extremum at s_*. Moreover, $\frac{\partial g_*}{\partial x_{n+1}}(s_*) = \frac{\partial g}{\partial x_p(s) \neq 0}$. So by the 1st part of the proof, there exists a $\lambda \in \mathbb{R}$ such that $\nabla f_*(s_*) = \lambda \nabla g_*(s_*)$. But

$$\frac{\partial f_*}{\partial x_i}(s*) = \begin{cases} \frac{\partial f}{\partial x_i}(s) \text{ for } i = 1, 2, \ldots, p-1 \\ \frac{\partial f}{\partial x_{i+1}}(s) \text{ for } i = p, p+1, \ldots, n \\ \frac{\partial f}{\partial x_p}(s) \text{ for } i = n+1 \end{cases}$$

and similarly for g_*. Hence, $\nabla f(s) = \lambda \nabla g(s)$. □

Example of Minimum distance with the Lagrange multiplier:
Find the minimum distance from the point $(1, 2, 0)$ to the surface $z^2 = x^2 + y^2, z \geq 0$, using the Lagrange multiplier.

The distance from $(1, 2, 0)$ to a point (x, y, z) is $d = \sqrt{(x-1)^2 + (y-2)^2 + z^2}$ and it suffices to minimize d^2, i.e., the function $f(x, y, z) = (x-1)^2 + (y-2)^2 + z^2$ on the set $\tilde{S} = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 - z^2 = 0, z \geq 0\}$. Recall that in Lecture 25 we solved this problem by eliminating z.

In particular, we found that f attains a global min value of $\tilde{S}$ but there does not exist a global max. Note also that $z = 0 \implies x^2 + y^2 = 0$, and $f(0, 0, 0) = 5$ while $f(0, 1, 1) = 3 < 5$. So f attains a global min on $S = \{(x, y, z) : x^2 + y^2 - z^2 = 0 \text{ and } z > 0\}$ and does not have a global max on S. We can apply our theorem to:

$\Omega = \{(x, y, z) \in \mathbb{R}^3 : z > 0\}$. $f : \Omega \to \mathbb{R}$, $f(x, y, z) = (x-1)^2 + (y-2)^2 + z^2$, $g : \Omega \to \mathbb{R}$, $g(x, y, z) = x^2 + y^2 - z^2$, and $S = \{(x, y, z) \in \Omega : g(x, y, z) = 0\} = \{(x, y, z) \in \Omega : x^2 + y^2 - z^2 = 0\}$ $(c = 0)$.

Note that $\nabla g(x, y, z) = (2x, 2y, -2z) \neq 0$ for all $(x, y, z) \in \Omega$, so by the theorem if a minimum occurs at $(x, y, z) \in S$ then $\nabla f(x, y, z) = (2(x-1), 2(y-2), 2z) = \lambda(2x, 2y, -2z)$ for some $\lambda \in \mathbb{R}$. So we need to solve the system:

$$\begin{cases} 2(x-1) = 2\lambda x \\ 2(y-2) = 2\lambda y \\ 2z = -2\lambda z \\ x^2 + y^2 - z^2 = 0 \end{cases} \implies \lambda = -1 \implies x = \frac{1}{2}, y = 1 \implies z = \sqrt{\frac{5}{4}}$$

So a minimum occurs at $\left(\frac{1}{2}, 1, \sqrt{\frac{5}{4}}\right)$ and the min distance is $d_{\min} = \sqrt{f\left(\frac{1}{2}, 1, \sqrt{\frac{5}{4}}\right)} = \sqrt{\frac{5}{2}}$.

Example of Maximum volume with the Lagrange multiplier:
Consider rectangular boxes $[-x, x] \times [-y, y] \times [-z, z]$ (x, y, z) incubed in the ellipsoid $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ (i.e., with vertices on the ellipsoid). Find the values of x, y, z which maximize the volume of such a box and the maximum volume.

Intuitively, it seems clear that the maximum exists. Can we confirm this mathematically?

Note that $\tilde{S} = \left\{(x, y, z) \in \mathbb{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1\right\}$ is compact, so by the EVT $f(x, y, z) = 8xyz$ attains its absolute maximum on $\tilde{S}$. It is clear that the maximum value is strictly positive, so (among other possibilities), it is attained at a point where $x, y, z > 0$. Hence, our problem has a solution.

Formally we work with the open set $\Omega = \{(x, y, z) : x, y, z > 0\}$ with the constraint function $g : \Omega \to \mathbb{R}$ given by $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$, and the function to maximize is $f(x, y, z) = 8xyz$. Note that $\nabla g(x, y, z) = \left(\frac{2x}{a^2}, \frac{2y}{b^2}, \frac{2z}{c^2}\right) \neq 0$ for all $(x, y, z) \in \Omega$. By the theorem, the max occurs at a point $(x, y, z) \in S$ where $\nabla f(x, y, z) = (8yz, 8xz, 8xy) = \lambda\left(\frac{2x}{a^2}, \frac{2y}{b^2}, \frac{2z}{c^2}\right)$ for some $\lambda \in \mathbb{R}$. So we need to solve the system:

$$\begin{cases} 8yz = \lambda\frac{2x}{a^2} \\ 8xz = \lambda\frac{2y}{b^2} \\ 8xy = \lambda\frac{2z}{c^2} \\ \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \end{cases} \implies \begin{matrix} 4xyz = \lambda\frac{x^2}{a^2} \\ 4xyz = \lambda\frac{y^2}{b^2} \\ 4xyz = \lambda\frac{z^2}{c^2} \end{matrix} \implies 2xyz = \lambda\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}\right) = \lambda$$

Given that $x, y, z > 0$, $\frac{1}{b} = \frac{x^2}{a^2} = \frac{y^2}{b^2} = \frac{z^2}{c^2} \implies x = \frac{a}{\sqrt{3}}, y = \frac{b}{\sqrt{3}}, z = \frac{c}{\sqrt{3}}$. The max volume is then $f\left(\frac{a}{\sqrt{3}}, \frac{b}{\sqrt{3}}, \frac{c}{\sqrt{3}}\right) = \frac{8abc}{3\sqrt{3}}$.

Week 4

Constraint Problems

Theorem 4.1 Lagrange multipliers for m constraints
Let $f, g_1, g_2, \ldots, g_m : \Omega \to \mathbb{R}$ be C^1 functions where $\Omega \subseteq \mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$ is open and let $S = \{x \in \Omega : g_1(x) = c_1, g_2(x) = c_2, \ldots, g_m(x) = c_m\}$ (with $c_1, c_2, \ldots, c_m \in \mathbb{R}$). If f attains an extreme value at some $s \in S$ where $\nabla g_1(s), \ldots, \nabla g_m(s)$ are linearly independent then there exists $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbb{R}$ (called Lagrange multipliers) such that $\nabla f = \lambda_1 \nabla g_1(s) + \lambda_2 \nabla g_2(s) + \cdots + \lambda_m \nabla g_m(s)$.

Proof. Let $g = (g_1, \ldots, g_m) : \Omega \to \mathbb{R}^m$. Since $\nabla g_1(s), \ldots, \nabla g_m(s)$ are linearly independent, the matrix $D_g(s) = \left[\frac{\partial g_i}{\partial x_j}(s)\right]_{i=1,j=1}^{m,m+n}$ has m linearly independent rows, so also m linearly independent columns. Let us consider the case that columns $n+1, n+2, \ldots, n+m$ are linearly independent.

Write (x, y) for the elements of $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$ and let $a = (s_1, \ldots, s_n), b = (s_{n+1}, \ldots, s_{n+m}), c = (c_1, \ldots, c_m)$. Clearly, $g(a, b) = c$, and with the notation used in the IPFT, $\frac{\partial g}{\partial y}(a, b) = \left[\frac{\partial g_i}{\partial x_j}(a, b)\right]_{i=1,j=n+1}^{m,n+m}$, so $\det\left(\frac{\partial g}{\partial y}(a, b)\right) \neq 0$. Therefore by the IPFT there exists open sets $U \subseteq \mathbb{R}^n, V \subseteq \mathbb{R}^m$ such that $s = (a, b)$ such that $U \times V$ and a C^1 function $\varphi : U \to V$ such that $\varphi(a) = b$ and $g(x, \varphi(x)) = c$ (i.e., $(x, \varphi(x)) \in S$) for all $x \in U$.

Define $\tilde{f} : U \to \mathbb{R}$ by $\tilde{f}(x) = f(x, \varphi(x))$. Clearly, $\tilde{f}$ is a C^1 function and it has an extremum at a, so $D_{\tilde{f}}(a) = 0 = \nabla \tilde{f}(a)$. Using the Chain Rule and the implicit differentiation formula for φ, we obtain:

$$0 = D_{\tilde{f}}(a) = D_f(s) \begin{bmatrix} I_n \\ D_\varphi(a) \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{\partial f}{\partial x}(s) & \frac{\partial f}{\partial y}(s) \end{bmatrix} \begin{bmatrix} I_n \\ D_\varphi(a) \end{bmatrix} =$$

$$= \frac{\partial f}{\partial x}(s) + \frac{\partial f}{\partial y}(s) D_\varphi(a) = \frac{\partial f}{\partial x}(s) - \frac{\partial f}{\partial y}(s) \left(\frac{\partial g}{\partial y}(s)\right)^{-1} \frac{\partial g}{\partial x}(s)$$

i.e.,

$$\frac{\partial f}{\partial x}(s) = \left(\frac{\partial f}{\partial y}(s) \left(\frac{\partial g}{\partial y}(s)\right)^{-1}\right) \frac{\partial g}{\partial x}(s)$$

Hence, with $\frac{\partial f}{\partial y}(s) \left(\frac{\partial g}{\partial y}(s)\right)^{-1} = [\lambda_1, \lambda_2, \ldots, \lambda_m]$,

$$\frac{\partial f}{\partial x_i}(s) = \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i}(s)\, \forall i = 1, 2, \ldots, n.$$

But this equality also holds for $i = n+1, \ldots, n+m$. Indeed,

$$\sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i}(s) = \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial y_{i-m}}(s) = \left[\frac{\partial f}{\partial y}(s) \left(\frac{\partial g}{\partial y}(s)\right)^{-1} \frac{\partial g}{\partial y}(s)\right] =$$

$$= \left[\frac{\partial f}{\partial y}(s)\, I_m\right] = \frac{\partial f}{\partial y_{i-m}}(s) = \frac{\partial f}{\partial x_i}(s).$$

Therefore $\nabla f(s) = \sum_{j=1}^{m} \lambda_j \nabla g_j(s)$. □

Example of two constraints:
THe planes $x + z = 4$ and $3x - y = 6$ intersect in a line L. Use the Lagrange multipliers to find a point on the line L that is closest to the origin.

From geometry the minimum distance exists (and no maximum exists). We will minimize the square of the distance from the origin to a point (x, y, z) on L.

Let $f, g_1, g_2 : \mathbb{R}^3 \to \mathbb{R}$ be given by $f(x, y, z) = x^2 + y^2 + z^2, g_1(x, y, z) = x + z, g_2(x, y, z) = 3x - y$. We look for the minimum of $f|L$, where $L = \{(x, y, z) : g_1(x, y, z) = 4, g_2(x, y, z) = 6\}$. We have

$$\nabla f(x, y, z) = (2x, 2y, 2z), \nabla g_1(x, y, z) = (1, 0, 1), \nabla g_2(x, y, z) = (3, -1, 0).$$

Clearly, $\nabla g_1(x, y, z)$ and $\nabla g_2(x, y, z)$ are linearly independent for all $(x, y, z) \in R^3$. So the minimum occurs when $\nabla f(x, y, z) = \lambda_1 \nabla g_1(x, y, z) + \lambda_2 \nabla g_2(x, y, z)$ for some $\lambda_1, \lambda_2 \in \mathbb{R}$ and we need to solve the system:

$$\begin{aligned} 2x &= \lambda - 1 + 3\lambda_2 \\ 2y &= -\lambda_2 \\ 2z &= \lambda_1 \qquad \implies 2x = 2z - 6y \implies x + 3y - z = 0 \\ x + z &= 4 \\ 3x - y &= 6 \end{aligned}$$

Solving the system, we get $(x, y, z) = (2, 0, 2)$.

So the nearest point on L to the origin is $(2, 0, 2)$ and the minimum distance is $\sqrt{8} = 2\sqrt{2}$.

Remark: When doing the method of Lagrange multipliers, it is important to investigate the points where the gradients are linearly dependent separately.

4.1 The integral in $\mathbb{R}^n$

Remark: The regular method for computing the integral in $\mathbb{R}$ is by way of the antiderivative. But there is no analogue to the antiderivative in $\mathbb{R}^n$, so our method for finding the integral will also not be analogous to how it was in $\mathbb{R}$.

Definition of a rectangle:
A **rectangle** (rectangular box) R in $\mathbb{R}^n$ is the cartesian product of intervals:

$$R = \Pi_{i=1}^n [a_i, b_i] = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_b, b_n],$$

where $a_i < b_i$ for all $i = 1, 2, \ldots, n$.
The n-dimensional volume $v(R)$ of R is

$$v(R) = \Pi_{i=1}^n (b_i - a_i) = (b_1 - a_1) \ldots (b_n - a_n).$$

Definition of a partition:
Let R be a rectangle in $\mathbb{R}^n$. By a **partition** of R, we mean a finite collection $\mathcal{P}$ of subrectangles of R such that $\bigcup_{P\in\mathcal{P}} P = R$ and $R_1 \cap R_2 = \emptyset$ whenever $R_1, R_2 \in \mathcal{P}$ and $R_1 \neq R_2$.
The mesh (or norm) of of the partition $\mathcal{P}$ is the number $\|\mathcal{P}\| = \max\{\mathrm{diam}\,(P) : P \in \mathcal{P}\}$ where $\mathrm{diam}\,(P) = \max\{\|x-y\| : x, y \in P\}$ is the diameter of P (if $P = \Pi_{i=1}^n(\alpha_i, \beta_i)$, then $\mathrm{diam}\,(P) = \sqrt{\sum_{i=1}^n(\beta_i - \alpha_i)^2}$)

Definition of a refinement:
Let $\mathcal{P}$ and $\mathcal{Q}$ be partitions of a rectangle $R \subseteq \mathbb{R}^n$. We say that $\mathcal{Q}$ is a refinement of $\mathcal{P}$ (or is finer than $\mathcal{P}$) if for all $Q \in \mathcal{Q}$, there exists a $P \in \mathcal{P}$ such that $Q \subseteq P$.

Lemma 4.2
Let $\mathcal{P}$ and $\mathcal{Q}$ be partitions of a rectangle R. Then

1. $\mathcal{Q}$ is a refinement of $\mathcal{P}$ if and only if each $P \in \mathcal{P}$ is the union of those $Q \in \mathcal{Q}$ that are contained in P.
2. There exists a partition $\mathcal{T}$ of R which refines both $\mathcal{P}$ and $\mathcal{Q}$ (e.g., $\mathcal{T} = \{P \cap Q : P \in \mathcal{P}, Q \in \mathcal{Q},$ and $P \cap Q$ is a rectangle.$\}$)

Lemma 4.3
If $\mathcal{P}$ is a partition of a rectangle $R \subseteq \mathbb{R}^n$, then $v\,(R) = \sum_{P\in\mathcal{P}} v\,(P)$.

Definition of upper and lower sums:
Let $R \subset \mathbb{R}^n$ be a rectangle, $f : \mathbb{R} \to \mathbb{R}$ a bounded function, and $\mathcal{P}$ be a partition of R. Given $P \in \mathcal{P}$ let

$$m_P = \inf\{f\,(X) : x \in P\}\,, M_P = \sup\{f\,(x) : x \in P\}\,.$$

The **lower and upper (Darboux or Riemann) sums of f for $\mathcal{P}$** are the numbers

$$L_{\mathcal{P}}\,(f) = \sum_{P\in\mathcal{P}} m_P v\,(P) \text{ and } U_{\mathcal{P}}\,(f) = \sum_{p\in\mathcal{P}} M_P v\,(P)\,,$$

respectively. (where $v\,(P)$ is the volume of P.)

Remark: $v\,(R) \inf\{f\,(x) : x \in R\} \leq L_{\mathcal{P}}\,(f) \leq U_{\mathcal{P}}\,(f) \leq \sup\{f\,(x) : x \in R\}\,v\,(R)$

Lemma 4.4
If $\mathcal{Q}$ is a refinement of $\mathcal{P}$, then

$$L_{\mathcal{P}}(f) \leq L_{\mathcal{Q}}(f) \text{ and } U_{\mathcal{Q}}(f) \leq U_{\mathcal{P}}(f).$$

Proof. Each $P \in \mathcal{P}$ is the union of the subfamily $\mathcal{Q}_P \subseteq \mathcal{Q}$ where $\mathcal{Q}_{\mathcal{P}P} = \{Q \in \mathcal{Q} : Q \subseteq P\}$. Clearly, for all $Q \in \mathcal{Q}_P$,

$$m_P = \inf\{f(x) : x \in P\} \leq \inf\{f(x) : x \in Q\} = m_Q.$$

Hence,

$$\sum_{Q \in \mathcal{Q}_P} m_Q v(Q) \geq \sum_{Q \in \mathcal{Q}_P} m_p v(Q) = m_P v(P)$$

But $\mathcal{Q}_P \cap \mathcal{Q}_{P'} = \emptyset$ when $P \neq P'$ and $\bigcup_{p \in P} \mathcal{Q}_P = \mathcal{Q}$. Therefore,

$$L_Q(f) = \sum_{Q \in \mathcal{Q}} m_Q v(Q) = \sum_{P \in \mathcal{P}} \left(\sum_{Q \in \mathcal{Q}_P} m_Q v(Q) \right) \geq \sum_{P \in \mathcal{P}} m_P v(P) = L_{\mathcal{P}}(f)$$

Similarly for the upper sums. □

Corollary 4.5
For any two partitions $\mathcal{P}$ and $\mathcal{P}'$ of R,

$$L_{\mathcal{P}}(f) \leq U_{\mathcal{P}'}(f)$$

Proof. Let $\mathcal{Q}$ be a common refinement of $\mathcal{P}$ and $\mathcal{P}'$. Then

$$L_{\mathcal{P}}(f) \leq L_{\mathcal{Q}}(f) \leq U_{\mathcal{Q}}(f) \leq U_{\mathcal{P}'}(f)$$

□

Let $\mathbb{P}$ denote the collection of all partitions of the rectangle R.

Corollary 4.6

$$\sup\{L_{\mathcal{P}}(f) : \mathcal{P} \in \mathbb{P}\} \leq \inf\{U_{\mathcal{P}}(f) : \mathcal{P} \in \mathbb{P}\}.$$

Definition of lower and upper integrals:
The **lower and upper (Darboux/Riemann) integrals** of a bounded function $f : R \to \mathbb{R}$ are defined by

$$\int_{*R} f = \sup \{L_{\mathcal{P}}(f) : \mathcal{P} \in \mathbb{P}\} \, and \int_R^* f = \inf \{U_{\mathcal{P}}(f) : \mathcal{P} \in \mathbb{P}\},$$

respectively. If $\int_{*R} f = \int_R^* f$, then we say that f is (Darboux/Riemann) integrable over R. The number $\int_{*R} f = \int_R^* f$ is called the (Darboux/Riemann) integral of f pver R and is denoted by $\int_R f$ or $\int_R f(x)\,dx$ or $\int_R f(x_1, \ldots, x_n)\,dx_1 dx_2 \ldots dx_n$ or $\int\int \cdots \int_R f(x_1, \ldots, x_n)\,dx_1 dx_2 \ldots dx_n$.
In particular, when $n = 2$ (resp, $n = 3$) then

$$\int\int_R f(x, y)\,dxfy(\int\int\int_R f(x, y, z)\,dxdydz)$$

is called the double (respectively, triple) integral of f over R.

Example of lower and upper integrals:
When $f : R \to \mathbb{R}$ is constant, $f(x) = c$ for all $x \in R$ then $U_{\mathcal{P}}(f) = L_{\mathcal{P}}(f) = cv(R)$ for any $\mathcal{P} \in \mathbb{P}$, and so F is integrable over R and $\int_R f = cv(R)$.

Theorem 4.7 The Riemann condition
Let $R \subseteq R^n$ be a rectangle and $f : R \to \mathbb{R}$ a bounded function. Then f is integrable over R if and only if for all $\varepsilon > 0$, there exists a $\mathcal{P} \in \mathbb{P}$ such that

$$U_{\mathcal{P}}(f) - L_{\mathcal{P}}(f) < \varepsilon.$$

Proof. $\Longrightarrow$: By the definition of the supremum and infimum, there exist $\mathcal{P}', \mathcal{P}'' \in \mathbb{P}$ such that

$$-\frac{\varepsilon}{2} + \int_R f < L_{\mathcal{P}'}(f) \leq \int_R f \text{ and } \int_R f \leq U_{\mathcal{P}''}(f) < \frac{\varepsilon}{2} + \int_R f (*)$$

Choosing a common refinement $\mathcal{P}$ of $\mathcal{P}'$ and $\mathcal{P}''$, (*) will also hold with $\mathcal{P}'$ and $\mathcal{P}''$ replaced by $\mathcal{P}$. Hence,

$$U_{\mathcal{P}}(f) - L_{\mathcal{P}}(f) < \left(\frac{\varepsilon}{2} + \int_R f\right) - \left(-\frac{\varepsilon}{2} + \int_R f\right) = \varepsilon.$$

$\Longleftarrow$: Note that

$$0 \leq \int_R^* f - \int_{*R} f \leq U_{\mathcal{P}}(f) - L_{\mathcal{P}}(f) < \varepsilon.$$

Since ε is arbitrary, $\int_R^* f = \int_{*R}$, i.e., f is integrable. $\square$

Corollary 4.8
If $f : R \to \mathbb{R}$ is integrable over $R \subseteq \mathbb{R}^n$ and $S \subseteq R$ is a subrectangle, then $f|S$ is integrable over S.

Proof. Let $\varepsilon > 0$. By the theorem, there exists a partition $\mathcal{P} \in \mathbb{P}$ such that $U_{\mathcal{P}}(f) - L_{\mathcal{P}}(f) < \varepsilon$. But $\mathcal{P}$ has a refinement $\mathcal{Q}$ such that $\mathcal{Q}' = \{Q \in \mathcal{Q} : Q \subseteq S\}$ is a partition of S. Then

$$U_{\mathcal{Q}'}(f|S) - L_{\mathcal{Q}'}(f|S) = \sum_{Q \in \mathcal{Q}'} (M_Q - m_Q)\, v(Q) \leq \sum_{Q \in \mathcal{Q}} (M_Q - m_q)\, v(Q)$$

$$= U_{\mathcal{Q}}(f) - L_{\mathcal{Q}}(f) \leq U_{\mathcal{P}}(f) - L_{\mathcal{P}}(f) < \varepsilon$$

$\square$

Corollary 4.9
If $f : R \to \mathbb{R}$ is a continuous function on a rectangle $R \subseteq \mathbb{R}^n$ then f is integrable over R.

Proof. Since R is compact, f is bounded. Moreover, f is uniformly continuous. Thus, given $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \frac{\varepsilon}{v(R)}$ whenever $x, y \in \mathbb{R}$ and $\|x - y\| < \delta$.
Let $\mathcal{P}$ be any partition with $\|P\| < \delta$. Now, given $P \in \mathcal{P}$, by the EVT, $m_p = \inf\{f(x) : x \in P\} = f(x_P)$ and $M_p = \sup\{f(x) : x \in P\} = f(y_P)$ for some $x_P, y_P \in P$.
As diam $P \leq \|\mathcal{P}\| < \delta$, $M_P - m_P = f(y_P) - f(x_P) < \frac{\varepsilon}{v(R)}$. Hence,

$$U_{\mathcal{P}}(f) - L_{\mathcal{P}}(f) = \sum_{P \in \mathcal{P}} (M_P - m_P) v(P) < \sum_{P \in \mathcal{P}} \frac{\varepsilon}{v(R)} v(P) = \varepsilon$$

Therefore the Riemann condition is satisfied. □

Example of an integrable function:
Let $R = [0,1] \times [0,1]$, and $g : R \to \mathbb{R}$ be given by $g(x,y) = \begin{cases} 1 \text{ when } (x,y) = (\frac{1}{2}, \frac{1}{2}) \\ 0 \text{ otherwise} \end{cases}$.
Then g is integrable. Indeed given $\varepsilon > 0$, choose a partition $\mathcal{P}$ of R where the subrectangle $P \in \mathcal{P}$ with $(\frac{1}{2}, \frac{1}{2}) \in P$ has $v(P) < \varepsilon$. Then $L_{\mathcal{P}}(g) = 0$ while $U_{\mathcal{P}}(g) = 1 \cdot v(P) < \varepsilon$. So the Riemann condition is satisfied.

Theorem 4.10
Let $f : R \to \mathbb{R}$ be an integrable function where $R \subseteq R^n$ is a rectangle. Then for all $\varepsilon > 0$ there exists $\mathcal{P}_\varepsilon \in \mathbb{P}$ such that the following holds:
If $\mathcal{P} \in \mathbb{P}$ is a refinement of $\mathcal{P}_\varepsilon$ and for all $P \in \mathcal{P}$ a point $x_P \in P$ is chosen, then

$$\left| \sum_{P \in \mathcal{P}} f(x_P v(P)) - \int_R f \right| < \varepsilon (*)$$

Definition of a Riemann sum:
Given a partition $\mathcal{P}$ of R, a choice of points $x_P \in P$ for all $P \in \mathcal{P}$ and a function $f : R \to \mathbb{R}$, the sum

$$\sum_{P \in \mathcal{P}} f(x_P) v(P)$$

is called the Riemann sum corresponding to the partition $\mathcal{P}$ and the choice of points $x_P \in P$ for all $P \in \mathcal{P}$.

Week 5

Constructing the integral

Theorem 5.1
Let $f : R \to \mathbb{R}$ be integrable where $R \subseteq R^n$ is a rectangle. Then for all $\varepsilon > 0$, there exists a particular $\mathcal{P}_\varepsilon$ of R such that if $\mathcal{P}$ us a partition that is finer than $\mathcal{P}_\varepsilon$ and if for all $P \in \mathcal{P}$ a point $x_P \in P$ is chosen, then

$$\left| \sum_{P \in \mathcal{P}} f(x_P)\, v(P) - \int_R f \right| < \varepsilon$$

Proof. Proof omitted, I came late! $\square$

Theorem 5.2
Let $f : R \to \mathbb{R}$ be a bounded function where $R \subseteq \mathbb{R}^n$ is a rectangle. Then f is integrable over $R \iff$ there exists a number s with the following property:

$$\forall \varepsilon > 0 \exists \text{ a partition } \mathcal{P} \text{ of } R \text{ such that } \left| \sum_{P \in \mathcal{P}} f(x_P)\, v(P) \right| < \varepsilon$$

for any choice of points $x_P \in P$ for all $P \in \mathcal{P}$.

Proof. $\Longrightarrow$: V (last theorem with $s = \int_R f$ and $\mathcal{P} = \mathcal{P}_\varepsilon$)
$\Longleftarrow$: We will show that the Riemann condition holds. Our assumption ensures that for all $\varepsilon > 0$ there exists a partition $\mathcal{P}$ such that

$$s - \frac{\varepsilon}{4} < \sum_{P \in \mathcal{P}} f(x_P)\, v(P) < s + \frac{\varepsilon}{4}$$

for any choice of $x_P \in P$ for all $P \in \mathcal{P}$.
But from the definition of the supremum and infimum, we can choose $\xi_P, \eta_P \in P$ such that

$$m_P = \inf\{f(x) : x \in P\} \le f(\xi_P) < m_P + \frac{\varepsilon}{4v(R)}$$

$$\text{and } M_P = \sup\{f(x) : x \in P\} \ge f(\eta_P) > M_P - \frac{\varepsilon}{4v(R)}.$$

Then

$$L_{\mathcal{P}}(f) = \sum_{P \in \mathcal{P}} m_P v(P) > \sum_{P \in \mathcal{P}} \left(f\left(\xi_P - \frac{\varepsilon}{4v(R)} \right) \right) v(P) = \sum_{P \in \mathcal{P}} (f(\xi_P))\, v(P) - \frac{\varepsilon}{4} > s - \frac{\varepsilon}{2},$$

and similarly,

$$U_{\mathcal{P}}(f) = \sum_{P \in \mathcal{P}} M_P v(P) < \sum_{P \in \mathcal{P}} \left(f\left(\eta_P + \frac{\varepsilon}{4v(R)} \right) \right) v(P) = \sum_{P \in \mathcal{P}} (f(\eta_P))\, v(P) + \frac{\varepsilon}{4} < s - \frac{\varepsilon}{2}.$$

Consequently,

$$U_{\mathcal{P}}(f) - L_{\mathcal{P}}(f) < s + \frac{\varepsilon}{2} - \left(s - \frac{\varepsilon}{2}\right) = \varepsilon.$$

□

Remark: If the number s, as defined in the last theorem exists, then $s = \int_R f$. (exercise)

Definition of volume zero:
A subset $S \subseteq R^n$ is said to have n**-dimensional volume zero**, written $v(S) = 0$, if for all $\epsilon > 0$ there exist rectangles $R_1, R_2, \ldots, R_n$ such that $S \subseteq \bigcup_{i=1}^n R_i$ and $\sum_{i=1}^n v(R_i) < \epsilon$.

Example of:
- Every finite subset of R^n has the n-dimensional volume 0.
- The countable set $S = \mathbb{Q} \cap [0,1]$ does not have 1-dimensional volume 0. [Indeed, if $S \subseteq \bigcup_{i=1}^n R_i$ where R_i are closed interbals, then as $\bigcup_{i=1}^n R_i$ where R_i are closed intervals, then as $\bigcup_{i=1}^n R_i$ is closed, $[0,1] = \bar{S} \subseteq$

$\bigcup_{i=1}^n R_i$. Thus $\sum_{i=1}^n v(R_i) = \sum_{i=1}^n \text{length}(R_i) \geq 1$.]

- If $R \subset \mathbb{R}^n$ is a rectangle then ∂R has n-dimensional volume 0. Indeed, if $R = \Pi_{i=1}^n [a_i, b_i]$ then $\partial R = \bigcup_{i=1}^n (\{x \in R : x_i = a_i\} \cup \{x \in R : x_i = b_i\})$. But for any $\eta > 0$, $\{x \in R : x_i = a_i\} \subseteq R_i = [a_i, b_i] \times \cdots \times [a_{i-1}, b_{i-1}] \times [a_i - \eta_i, b_i] \times [a_{i+1}, b_{i+1}] \times [a_n, b_n]$ where $v(R_i) = \eta \Pi_{j=1, j\neq i}^n (b_j - a_j)$. Hence, $\{x \in R : x_i = a_i\}$ has volume zero. Similarly, $\{x \in R : x_i = b_i\}$ has volume zero, since the union of finitely many sets of zero volume is a set of zero volume, ∂R has volume zero.

Proposition 5.3
If $f : \Omega \to \mathbb{R}$ is continuous, where $\Omega \subseteq \mathbb{R}^n$ is compact, then $\text{graph}(f) = \left\{(x, y) \in \mathbb{R}^{n+1} : x \in \Omega \text{ and } y = f(x)\right\}$ has $(n+1)$-volume zero.
More generally, for any $k = 1, 2, \ldots, n+1$, the set $S = \{(x+1, \ldots, x_{n+1}) : (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{n+1}) \in \Omega \text{ and } x_k = f(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{n+1})\}$ has $(n+1)$-dimensional volume zero.

Proof. Note that Ω is contained in a rectangle $R \subseteq \mathbb{R}^n$. As f is uniformly continuous, given $\varepsilon > 0$, there exists an $s > 0$ such that $|f(x) - f(x')| < \frac{\varepsilon}{4v(R)}$ whenever $x, x' \in \Omega$ and $\|x - x'\| < \delta$.
Choose a partition $\mathcal{P}$ of R with $\|\mathcal{P}\| < \delta$ and let $\mathcal{P}_* = \{P \in \mathcal{P} : P \cap \Omega \neq \emptyset\}$. Note $\Omega \subseteq \bigcup_{P \in \mathcal{P}_*} P$. Given $P \in \mathcal{P}$, choose some $x_P \in P \cap \Omega$ and let $R_P = P \times [f(x_P) - \frac{\varepsilon}{4v(R)}, f(x_P) + \frac{\varepsilon}{4v(R)}]$ which is a rectangle in $\mathbb{R}^{n+1}$ with $v(R_P) = v(P) - \frac{\varepsilon}{2v(R)}$.
Note that if $(x, y) \in \text{graph}(f)$ then $x \in P$ for some $P \in \mathcal{P}_*$. Since $\|\mathcal{P}\| < \delta$, so $\|x = x_P\| < s$ and so $f(x) \in [f(x_P) - \frac{\varepsilon}{4v(R)}, f(x_P) + \frac{\varepsilon}{4v(R)}]$. It follows that $(x, y) \in R_P$. Consequently, $\text{graph}(f) \subseteq \bigcup_{P \in \mathcal{P}_*} R_P$. But

$$\sum_{P \in \mathcal{P}_*} v(R_P) = \sum_{P \in \mathcal{P}_*} v(P) \frac{\varepsilon}{2v(R)} = \frac{\varepsilon}{2} < \varepsilon.$$

As $\varepsilon > 0$ is arbitrary, $\text{graph}(f)$ has $(n+1)$-dimensional volume zero. $\square$

Theorem 5.4
Let $f : R \to \mathbb{R}$ be a bounded function where $R \subseteq R^n$ is a rectangle. If $D = \{x \in R : f \text{ is discontinuous at } x\}$ has n-dimensional volume zero, then f is integrable over R.

5.1 Basic properties of integrals over rectangles

Theorem 5.5 *1
Let $f, g : R \to \mathbb{R}$ be integrable over the rectangle $R \subseteq \mathbb{R}^n$ and let $c \in \mathbb{R}$. Then:

1. cf is integrable over R, and $\int_R cf = c \int_R f$.
2. $f + g$ is integrable over R and $\int_R f + g = \int_R f + \int_R g$.
3. If $g \leq f$ on R, then $\int_R g \leq \int_R f$.
4. $|f|$ is integrable over R and $\left|\int_R f\right| \leq \int_R |f|$.

Proof. In the same order as before,

- We may assume that $c \neq 0$, let $\varepsilon > 0$. As f is integrable over R, there exists a partition $\mathcal{P}$ of R such that for any choice of $x_P \in P$ for all $P \in \mathcal{P}$,

$$\left|\sum_{P\in\mathcal{P}} f(x_P)\, v(P) - \int_R f\right| < \frac{\varepsilon}{c}.$$

But then

$$\left|\sum_{P\in\mathcal{P}} cf(x_P)\, v(P) - c\int_R f\right| > \varepsilon.$$

Hence, cf is integrable and $\int_R cf = c\int_R f$ by the 2nd theorem about Riemann sums.

- We again use Riemann sums. Given $\varepsilon > 0$ there exists a partition $\mathcal{P}'_\varepsilon$ (respectively, $\mathcal{P}''_\varepsilon$) such that for every partition $\mathcal{P}$ that is finer than $\mathcal{P}'_\varepsilon$ (respectively, $\mathcal{P}''_\varepsilon$) and for any choice of points $x_P \in P$ for all $P \in \mathcal{P}$,

$$\left|\sum_{P\in\mathcal{P}} f(x_P)\, v(P) - \int_R f\right| < \frac{\varepsilon}{2}(\text{respectively, } \left|\sum_{P\in\mathcal{P}} g(x_P)\, v(P) - \int_R g\right| < \frac{\varepsilon}{2})$$

Let $\mathcal{P}$ be a common refinement of $\mathcal{P}'_\varepsilon$ and $\mathcal{P}''_\varepsilon$. Then for any choice of points $x_P \in P$ for all $P \in \mathcal{P}$,

$$\left|\sum_{P\in\mathcal{P}} (f(x_P) + g(x_P))\, v(P) - \left(\int_R f + \int_R g\right)\right| \leq \left|\sum_{P\in\mathcal{P}} f(x_P)\, v(P) - \int_R f\right| + \left|\sum_{P\in\mathcal{P}} g(x_P)\, v(P) - \int_R g\right| <$$

$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

Hence, by the 2nd theorem about Riemann sums, $f + g$ is integrable over R and $\int_R f + g = \int_R f + \int_R g$.

- Clearly, $f - g \geq 0$ and so for any partition $\mathcal{P}$ of R, $L_{\mathcal{P}}(f - g) \geq 0$. Hence, $\int_R f - g \geq L_{\mathcal{P}}(f - g) \geq 0$ (we used the first two parts). Then again by these first two parts, $\int_R f - \int_R g = \int_R f - g \geq 0$, so $\int_R f \geq \int_R g$

$\square$

Proof.

- We will use the Riemann condition. Let $\mathcal{P}$ be a partition of R and given $P \in \mathcal{P}$, let
$$m_P = \inf\{f(x) : x \in P\} \quad M_P = \sup\{f(x) : x \in P\}$$
$$\bar{m_P} = \inf\{|f(x)| : x \in P\} \quad \bar{M_P} = \sup\{|f(x)| : x \in P\}.$$
Note that if $x, x' \in P$ then
$$||f(x)| - |f(x')|| \leq |f(x) - f(x')| \leq M_P - m_P.$$
Thus,
$$|f(x) \leq M_P - m_P + |f(x')||.$$
Hence, keeping x' fixed, $\bar{M_P} = \sup\{|f(x)| : x \in P\} \leq M_P - m_P + |f(x')|$ for all $x' \in P$, and so
$$\bar{M_P} - M_P + m_P \leq |f(x')|.$$
Hence, $\bar{M_P} - M_P + m_P \leq \inf\{|f(x')| : x' \in P\} = \bar{m_P}$, and so
$$\bar{M_P} - \bar{m_P} \leq M_P = m_P.$$
Therefore, $U_{\mathcal{P}}(|f|) - L_{\mathcal{P}}(|f|) = \sum_{P \in \mathcal{P}} \left(\bar{M_P} - \bar{m_P}\right) v(P) \leq \sum_{P \in \mathcal{P}} (M_P - m_P) v(P) = U_{\mathcal{P}}(f) - L_{\mathcal{P}}(f)$.
But by integrability of f and the Riemann condition, for any $\varepsilon > 0$, $\mathcal{P}$ can be chosen so that $U_{\mathcal{P}}(f) - L_{\mathcal{P}}(f) < \varepsilon$. Therefore the Riemann condition is also satisfied by $|f|$, so that $|f|$ is integrable over R.
Then as $-|f| \leq f \leq |f|$, $-\int_R |f| \leq \int_R f \leq \int_R |f|$ by the first two parts. Thus $\left|\int_R f\right| \leq \int_R |f|$.

□

Theorem 5.6 *2
Let $f : R \to \mathbb{R}$ be a bounded function where $R \subseteq \mathbb{R}^n$ is a rectangle. If $E = \{x \in R : f(x) \neq 0\}$ has n-dimensional volume zero then f is integrable over R and $\int_R f = 0$.

Corollary 5.7 *3
Let $f, g : R \to \mathbb{R}$ be bounded functions where $R \subseteq \mathbb{R}^n$ is a rectangle. If f is integrable over R and $\{x \in R : g(x) \neq f(x)\}$ has zero volume, then g is integrable over R and $\int_R f = \int_R g$.

Proof. By theorem *2, $g - f$ is integrable over R and $\int_R (g-f) = 0$. Hence, $g = g - f + f$ is integrable $\int_R g = \int_R (g-f) + \int_R f = \int_R f$. $\square$

Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n] = \Pi_{i=1}^n [a_i, b_i]$ be a rectangle and $f : R \to \mathbb{R}$ a bounded function. Given a permutation σ of $\{1, 2, \ldots, n\}$ and $x = (x_1, \ldots, x_n)$, $f\left(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}\right)$ is defined whenever $\left(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}\right) \in R$, i.e., whenever $x_{\sigma(i)} \in [a_i, b_i]$ for all $i = 1, 2, \ldots, n$, or equivalently whenever $x_i \in [a_{sigma^{-1}(i)}, b_{sigma^{-1}(i)}]$ i.e., $x \in \Pi_{i=1}^n [a_{\sigma^{-1}(i)}, b_{\sigma^{-1}(i)}] = R_\sigma$. Thus the formula,

$$f_\sigma(x_1, \ldots, x_n) = f\left(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\right)$$

defines a bounded function $f_\sigma : R_\sigma \to \mathbb{R}$. it is straightforward to see that we have a one-to-one correspondence between partitions of R and partitions of R_σ and that the corresponding lower and upper sums for f and f_σ have the same values. Hence,

Theorem 5.8
If $f : R \to \mathbb{R}$ is integrable over the rectangle $R = \Pi_{i=1}^n [a_i, b_i]$, then for any permutation σ of $\{1, 2, \ldots, n\}$, the function $f_\sigma : R_\sigma \to \mathbb{R}$ as defined above is integrable over R_σ and $\int_R f = \int_{R_\sigma} f_\sigma$, or

$$\int f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = \int f\left(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\right) dx_1 \ldots dx_n.$$

Example omitted due to sleepiness.

Let $w = (w_1, \ldots, w_n) \in \mathbb{R}^n$ be fixed. Clearly, if $R = \Pi_{i=1}^n [a_i, b_i]$ is a rectangle then $R - w = \{x - w : x \in R\} = \Pi_{i=1}^n [a_i - w_i, b_i - w_i]$ is another rectangle and if $f : R \to \mathbb{R}$ is a bounded function, then the function f_w given by $f_w(x) = f(x + w)$ is defined for $x \in R - w$. We have a one-to-one correspondence between partitions of R and partitions of $R - w$ and the corresponding lower and upper sums for f and f_w have the same values. Hence,

Theorem 5.9
If $f : R \to \mathbb{R}$ is integrable over the rectangle $R = \Pi_{i=1}^n [a_i, b_i]$ then for any $w \in R^n$ the function $f_w : R - w \to \mathbb{R}$ defined above is integrable over $R - w$ and

$$\int_R f = \int_{R-w} f_w,$$

or,

$$\int_R f(x)\, dx = \int_{R-w} f(x + w)\, dx$$

Suppose $\lambda \in (\mathbb{R} \backslash \{0\})^n = \{x \in \mathbb{R}^n : x_1, \ldots, x_n \neq 0\}$. Then given a rectangle $R = \Pi_{i=1}^{n}[a_i, b_i]$ the set $R_\lambda = \left\{\left(\frac{1}{\lambda_1}x_1, \dots, \frac{1}{\lambda_n}x_n,\right) : (x_1, \dots, x_n) \in R\right\} = \Pi_{i=1}^{n}\left[\min\left\{\frac{a_i}{\lambda_i}, \frac{b_i}{\lambda_i}\right\}, \max\left\{\frac{a_i}{\lambda_i}, \frac{b_i}{\lambda_i}\right\}\right]$ is another rectangle with $v\left(R_\lambda\right) = \left|\Pi_{i=1}^{n}\lambda_i^{-1}\right| v\left(R\right)$ and if $f : R \to \mathbb{R}$ is a bounded function, then the function f_λ, given by $f_\lambda\left(x_1, \dots, x_n\right) = f\left(\lambda x_1, \dots, \lambda_n x_n\right)$ is defined for $(x_1, \dots, x_n) \in R_\lambda$.