Mathematical Methods

Emre Özer

Spring 2019

Based on lectures by Dmitry Turaev.

Contents

1 Complex Analysis

We are interested in functions of complex variables and how they behave on the complex plane. We start by properties of complex numbers and the complex plane.

1.1 Complex Numbers and the Complex Plane

Any complex number $z \in \mathbb{C}$ can be written as a sum of real and imaginary parts, such that $z = x + iy$ where $x, y \in \mathbb{R}$. We can also write $z = re^{i\theta}$ where r is the *modulus* and θ is the *argument*.

Definition (Modulus and argument)**.** The modulus and argument of a complex number $z = x + iy$ is given by

$$r = |z| = \sqrt{x^2 + y^2}, \quad \theta = \arg z,$$

where $x = r\cos\theta$ and $y = r\sin\theta$.

Definition (Principal value of argument)**.** The principal value of the argument is the value of θ in the range $(-\pi, \pi]$.

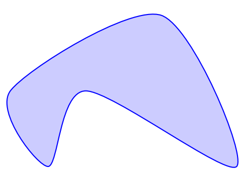
Definition (Open ball)**.** An n-dimensional open ball of *radius* r is the set of all points of distance less than r than a fixed point in n-dimensional space. Formally, an open ball around $\mathbf{x}$ with radius r is defined by

$$B_r(\mathbf{x}) = \{\mathbf{y} : |\mathbf{x} - \mathbf{y}| < r\}.$$

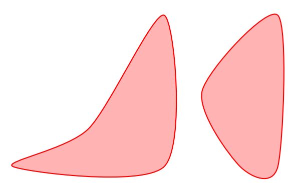
Definition (Open subset)**.** A subset $\mathcal{U} \subseteq \mathbb{C}$ is open if $\forall\ z \in \mathcal{U}$, there exists some $\varepsilon > 0$ such that the open ball $B_\varepsilon(z) \subseteq \mathcal{U}$.

Definition (Neighbourhood)**.** A neighbourhood of a point $z \in \mathbb{C}$ is an open subset $\mathcal{U} \subseteq \mathbb{C}$ containing z.

Definition (Path-connected subset)**.** A subset $\mathcal{U} \subseteq \mathbb{C}$ is path-connected if $\forall\ x, y \in \mathcal{U}$ there exists some $\gamma : [0,1] \to \mathcal{U}$ continuous such that $\gamma(0) = x$ and $\gamma(1) = y$.



Path-connected



Not path-connected

Definition (Domain)**.** A domain is a non-empty open path-connected subset of $\mathbb{C}$.

1.2 Differentiability of Complex Functions

Definition (Differentiable function)**.** A function $f : \mathcal{U} \to \mathbb{C}$ is differentiable at $z_0 \in \mathcal{U}$ if

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists. The derivative is given by the limit.

It is implicitly stated that the limit does not depend on which direction z approaches z_0, as there are infinitely many ways to do so.

Definition (Analytic/holomorphic function)**.** A function f is *analytic or holomorphic* at $z_0 \in \mathcal{U}$ if it is differentiable on a neighbourhood $B_\varepsilon(z_0)$ of z_0.

Definition (Entire function)**.** Functions that are analytic in $\mathbb{C}$ are called *entire functions*.

Proposition. Any complex function $f : \mathcal{U} \to \mathbb{C}$ can be written as a sum of two functions $u, v : \mathbb{R}^2 \to \mathbb{R}$ such that $f(z) = u(x,y) + iv(x,y)$ where $z = x + iy$.

Proof. We note that $u(x,y) = \mathrm{Re}(f(z))$ and $v(x,y) = \mathrm{Im}(f(z))$. Hence, we can explicitly state

$$u(x,y) = \frac{1}{2}\left(f(x+iy) + \overline{f}(x+iy)\right), \quad v(x,y) = \frac{1}{2i}\left(f(x+iy) - \overline{f}(x+iy)\right)$$

for any given function $f(z)$. Noting that for any given $z_0 \in \mathcal{U}$, we have $f(z_0) \in \mathbb{C}$. Hence, $u(x_0, y_0), v(x_0, y_0) \in \mathbb{R}$. This completes the proof. □

Proposition (Cauchy-Riemann conditions)**.** Let $f : \mathcal{U} \to \mathbb{C}$, and write $f(z) = u(x,y) + iv(x,y)$. Then, f is complex differentiable at z if and only if

$$u_x = v_y, \quad v_x = -u_y,$$

evaluated at $z = x + iy$. These are the *Cauchy-Riemann equations.* The derivative is then given by

$$f'(z) = u_x(x,y) + iv_x(x,y) = v_y(x,y) - iu_y(x,y).$$

Proof. We will show that the limit

$$f'(z) = \lim_{\delta \to 0} \frac{f(z+\delta) - f(z)}{\delta}$$

is independent of how $\delta \in \mathbb{C}$ approaches zero if and only if Cauchy-Riemann conditions are satisfied.

Let $z = z_0 + \delta$. Since δ can approach zero along any direction, we let $\delta = (\alpha + i\beta)\varepsilon$, for some $\alpha, \beta, \epsilon \in \mathbb{R}$. As $\varepsilon \to 0$, $\delta \to 0$ regardless of α and β (as long as both are not zero). Taking the limit yields

$$\begin{aligned}
f'(z) &= \lim_{\delta \to 0} \frac{f(z+\delta) - f(z)}{\delta} \\
&= \frac{1}{(\alpha + i\beta)} \lim_{\varepsilon \to 0} \frac{u(x + \alpha\varepsilon, y + \beta\varepsilon) + iv(x + \alpha\varepsilon, y + \beta\varepsilon) - u(x,y) - iv(x,y)}{\varepsilon} \\
&= \frac{1}{(\alpha + i\beta)} \lim_{\varepsilon \to 0} \frac{\alpha u_x + \beta u_y + i\alpha v_x + i\beta v_y + \mathcal{O}\left(\varepsilon^2\right)}{\varepsilon} \\
&= \frac{\alpha(u_x + iv_x) + i\beta(-iu_y + v_y)}{\alpha + i\beta}.
\end{aligned}$$

Where we assumed u and v were continuous. Now, we require the derivative to be independent of α and β, as they determine the direction of δ approaching zero. This can only be satisfied if

$$u_x + iv_x = v_y - iu_y,$$

which are the Cauchy-Riemann conditions. □

The same rules of differentiation (sum, product, chain) hold for all complex differentiable functions.

Example. $f(z) = \overline{z}$ is not analytic. We express $f(z)$ as $f(x+iy) = x - iy$, and look at the Cauchy-Riemann conditions.

$$u_x = 1 \neq -1 = v_y, \quad u_y = 0 = -v_x.$$

One of the conditions does not hold for any $z \in \mathbb{C}$, hence the function is not analytic at any point.

1.3 Contour Integration and Cauchy's Theorem

In real analysis, an integral is taken over a real interval. In complex analysis, we require a *path* instead. Since there are infinitely many paths joining two complex numbers on the complex plane, the integral might differ depending on the path we choose.

1.3.1 Contours and integrals

We start with some definitions.

Definition (Curve)**.** A curve $\gamma(t)$ is a continuous map $\gamma : [0,1] \to \mathbb{C}$.

Definition (Closed curve)**.** A curve $\gamma(t)$ is said to be closed if $\gamma(0) = \gamma(1)$.

Definition (Simple curve)**.** A curve is said to be simple if it does not intersect itself, except on the endpoints of a closed curve.

Definition (Piecewise smooth curve)**.** A curve γ is said to be piecewise smooth if each of its components has a bounded derivative which is continuous everywhere on $[0,1]$, except possibly at a finitely many isolated points.

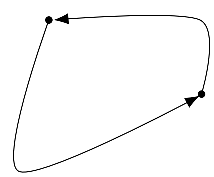
Definition (Contour)**.** A contour is a piecewise smooth curve. From now on, γ will refer both to the map and its image. So, the curve traversed in $\mathbb{C}$ will be referred to as γ. We denote the contour traversed in the opposite direction as $-\gamma$. Formally,

$$(-\gamma)(t) = \gamma(1-t).$$

We define the sum of two contours, $\gamma_1 + \gamma_2$, joined end-to-end with the condition $\gamma_1(1) = \gamma_2(0)$ as

$$(\gamma_1 + \gamma_2)(t) = \begin{cases} \gamma_1(2t) & t < \frac{1}{2}, \\ \gamma_2(2t-1) & t \geq \frac{1}{2}. \end{cases}$$

A piecewise contour:



Definition (Contour integral)**.** The contour integral over some γ of a function $f : \mathcal{U} \to \mathbb{C}$ over a domain containing γ is defined as the infinite sum (analogous to a Riemann sum over $\mathbb{R}$):

$$\int_\gamma f(z)\mathrm{d}z = \lim_{\Delta \to 0} \sum_k f(\gamma(t_k))\left(\gamma(t_{k+1}) - \gamma(t_k)\right),$$

where the interval $t \in [0,1]$ is split into discrete t_k with the condition

$$\Delta = \max_{i=0,\cdots,i=N} [t_{i+1} - t_i]$$

approaches zero.

We can rewrite this, using the definition of a Riemann sum, as an integral. First, we consider the γ term, which can be written as

$$\gamma(t_{k+1}) - \gamma(t_k) = \gamma'(t_k)\delta t_k + o(\delta t_k),$$

and as $\delta t_k \to 0$, t becomes a continuous variable so we get

$$\int_\gamma f(z)\mathrm{d}z = \int_0^1 f(\gamma(t))\gamma'(t)\mathrm{d}t\,.$$

Proposition. Properties of contour integration.

(i) For any two contours γ_1, γ_2, if $(\gamma_1 + \gamma_2)$ is also a contour then we have

$$\int_{\gamma_1+\gamma_2} f(z)\mathrm{d}z = \int_{\gamma_1} f(z)\mathrm{d}z + \int_{\gamma_2} f(z)\mathrm{d}z\,,$$

for all $f : \mathcal{U} \to \mathbb{C}$ such that $(\gamma_1 + \gamma_2) \in \mathcal{U}$.

Proof. We use the definition of the contour integral and the sum of contours to write

$$\begin{aligned}\int_{\gamma_1+\gamma_2} f(z)\mathrm{d}z &= \int_0^1 f\left((\gamma_1+\gamma_2)(t)\right)(\gamma_1+\gamma_2)'(t)\mathrm{d}t \\ &= \int_0^{1/2} f(\gamma_1(2t))\gamma_1'(2t)2\mathrm{d}t + \int_{1/2}^1 f(\gamma_2(2t-1))\gamma_2'(2t-1)2\mathrm{d}t\,.\end{aligned}$$

We now let $w = 2t$ for the first integral, and $w = 2t - 1$ for the second. This yields

$$\int_0^1 f(\gamma_1(w))\gamma_1'(w)\mathrm{d}w + \int_0^1 f(\gamma_2(w))\gamma_2'(w)\mathrm{d}w = \int_{\gamma_1} f(z)\mathrm{d}z + \int_{\gamma_2} f(z)\mathrm{d}z\,.$$

Hence, the proposition holds. □

(ii) For any contour γ, the integral over it traversed backwards, $-\gamma$, is given by

$$\int_\gamma f(z)\mathrm{d}z = -\int_{-\gamma} f(z)\mathrm{d}z\,.$$

Proof. We start with the usual definition of the contour integral,

$$\int_\gamma f(z)\mathrm{d}z = \int_0^1 f(\gamma(t))\gamma'(t)\mathrm{d}t\,.$$

Now, we consider the substitution $t = 1 - w$, which yields

$$\begin{aligned}\int_0^1 f(\gamma(t))\gamma'(t)\mathrm{d}t &= \int_1^0 f(\gamma(1-w))\left(-\gamma'(1-w)\right)(-\mathrm{d}w) \\ &= -\int_0^1 f(\gamma(1-w))\gamma'(1-w)\mathrm{d}w \\ &= -\int_{-\gamma} f(z)\mathrm{d}z\,,\end{aligned}$$

where we used $\gamma(1-t) = (-\gamma)(t)$. □

(iii) Let γ be a contour from a to b, i.e. $\gamma(0) = a$ and $\gamma(1) = b$ for any $a, b \in \mathbb{C}$. Then, we have

$$\int_\gamma f'(z)\mathrm{d}z = f(b) - f(a).$$

Proof. By definition, we have

$$\int_\gamma f'(z)\mathrm{d}z = \lim_{\Delta\to 0}\sum_k f'(\gamma(t_k))\left(\gamma(t_{k+1}) - \gamma(t_k)\right).$$

Also, by definition we have

$$f'(\gamma(t_i)) = \lim_{\Delta\gamma(t_i)\to 0}\frac{f(\gamma(t_i) + \Delta\gamma(t_i)) - f(\gamma(t_i))}{\Delta\gamma(t_i)},$$

assuming the limit exists for all t_i. Now, we can define $\Delta\gamma(t_i) = \gamma(t_{i+1}) - \gamma(t_i)$, noting that in the limit $\Delta \to 0$, we have $\Delta\gamma(t_i) \to 0 \; \forall \; t_i \in [0,1]$. Hence, the integral becomes

$$\lim_{\Delta\to 0} \sum_k \frac{f(\gamma(t_{k+1}) - f(\gamma(t_k))}{\gamma(t_{k+1}) - \gamma(t_k)} \left(\gamma(t_{k+1}) - \gamma(t_k)\right) = f(\gamma(1)) - f(\gamma(0))$$

since all except the first and the last terms of the sum cancel. □

(iv) (Integration by substitution) Let $f : \mathcal{U} \to \mathbb{C}$, and denote $u \in \mathcal{U}$. Let $\phi : \mathcal{D} \to \mathcal{U}$ be a bijection, and denote $z \in \mathcal{D}$. We then have

$$\int_{\gamma_u} f(u)\mathrm{d}u \;=\; \int_{\gamma_z} f(\phi(z))\phi'(z)\mathrm{d}z\,,$$

where $\gamma_u(t) = \phi(\gamma_z(t)) \; \forall \; t \in [0,1]$.

Proof. The proof is straightforward, working from the definition of the integral:

$$\begin{aligned}\int_{\gamma_u} f(u)\mathrm{d}u \;&=\; \int_0^1 f(\gamma_u(t))\gamma_u'(t)\mathrm{d}t \\ &=\; \int_0^1 f(\phi(\gamma_z(t)))\partial_t\phi(\gamma_z(t))\mathrm{d}t \\ &=\; \int_0^1 f(\phi(\gamma_z(t)))\phi'(\gamma_z(t))\gamma_z'(t)\mathrm{d}t \\ &=\; \int_{\gamma_z} f(\phi(z))\phi'(z)\mathrm{d}z\,.\end{aligned}$$

The requirement that ϕ is a bijection is important so that γ_z is a contour. □

(v) (Integration by parts) Let u and v be analytic functions, and let γ be a contour over the domains of u and v. Then, we have

$$\int_\gamma u'v\mathrm{d}z \;=\; uv|_{z=\gamma(1)} - uv|_{z=\gamma(0)} - \int_\gamma uv'\mathrm{d}z\,.$$

Proof. By the chain rule, we have $u'v = (uv)' - uv'$. So,

$$\begin{aligned}\int_\gamma u'v\mathrm{d}z \;&=\; \int_\gamma (uv)' - uv'\mathrm{d}z \\ &=\; uv|_{z=\gamma(1)} - uv|_{z=\gamma(0)} - \int_\gamma uv'\mathrm{d}z\,,\end{aligned}$$

where proposition (iii) was used. □

(vi) If γ has length L and $|f(z)|$ is bounded on γ by some $M \in \mathbb{R}$, we have

$$\left|\int_\gamma f(z)\mathrm{d}z\right| \leq ML.$$

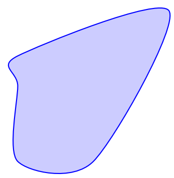
Proof. We have

$$\left|\int_\gamma f(z)\mathrm{d}z\right| \leq \int_\gamma |f(z)||\mathrm{d}z\,| \leq \int_\gamma M|\mathrm{d}z\,| = ML.$$

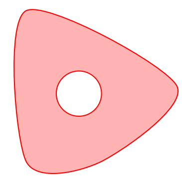
Some properties related to absolute values of integrals were assumed. □

1.3.2 Cauchy's theorem

Definition (Simply connected domain)**.** A domain $\mathcal{D} \subseteq \mathbb{C}$ is simply connected if every closed curve in $\mathcal{D}$ encloses only points in the domain. (i.e. it doesn't have holes.)



simply connected



multiply connected

Theorem (Cauchy's theorem)**.** If $f(z)$ is analytic in a simply connected domain $\mathcal{D}$, then for *every closed contour* γ in $\mathcal{D}$,

$$\oint_\gamma f(z)\mathrm{d}z = 0.$$

Proof. We will use Green's theorem, which states that for smooth real functions P and Q, we have

$$\oint_{\partial S} (P\mathrm{d}x + Q\mathrm{d}y) = \iint_S (\partial_x Q - \partial_y P)\,\mathrm{d}x\,\mathrm{d}y\,.$$

Letting $f(z) = u(x,y) + iv(x,y)$, we have

$$\begin{aligned}\oint_\gamma f(z)\mathrm{d}z &= \oint_\gamma (u(x,y) + iv(x,y))\,(\mathrm{d}x + i\mathrm{d}y) \\ &= \oint_\gamma (u+iv)\mathrm{d}x + (-v+iu)\mathrm{d}y \\ &= \iint_S (\partial_x(u+iv) - \partial_y(-v+iu))\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_S (-v_x + iu_x - u_y - iv_y)\,\mathrm{d}x\,\mathrm{d}y \\ &= 0\end{aligned}$$

due to Cauchy-Riemann conditions. We require u and v to have continuous partial derivatives in S, for Green's theorem to hold. □

Note. *A more general proof that doesn't require u and v to have continuous partial derivatives exists.*

Proposition (Contour deformation)**.** Let γ_1 and γ_2 be contours from some a to b in $\mathbb{C}$. Then, for all f that is analytic on γ_1 and γ_2, and on the region bounded by them, we have

$$\int_{\gamma_1} f(z)\mathrm{d}z = \int_{\gamma_2} f(z)\mathrm{d}z\,.$$

Proof. If the contours do not intersect, then $\gamma_1 + (-\gamma_2)$ is a closed contour and we have

$$0 = \oint_{\gamma_1-\gamma_2} f(z)\mathrm{d}z = \int_{\gamma_1} f(z)\mathrm{d}z - \int_{\gamma_2} f(z)\mathrm{d}z\,.$$

If the contours intersect, then consider every individual closed contour they form and the same reasoning applies. □

1.4 Cauchy's Integral Formula

So far, we have considered simply connected domains. However, we can generalise Cauchy's theorem to domains which are not simply connected.

Definition (Multiply connected domain)**.** A domain which is *not simply connected* is called *multiply connected.* A domain would be multiply connected even if there was a hole at a *single point* in the domain.

Note (Orientation of closed contours)**.** With closed contours, it doesn't matter where we start and end, as long as we traverse a full cycle. However, there is ambiguity in the direction of traversal. The usual direction is *anticlockwise*, or the "positive sense". We can also traverse *clockwise*, which would be "negative sense".

Proposition. Given a multiply connected domain, the integral over *any* closed contour γ_0 enclosing n "holes" (open connected regions enclosed by the domain where the function f with said domain would not be defined) can be expressed as

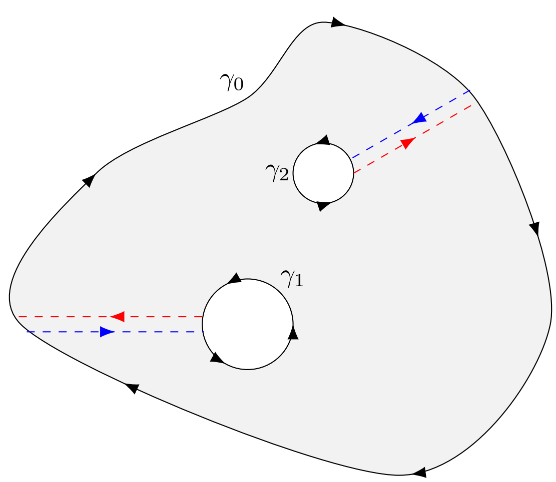
$$\oint_{\gamma_0} f(z)\,\mathrm{d}z = \sum_{i=1}^{n} \oint_{\gamma_i} f(z)\,\mathrm{d}z\,,$$

where each γ_i is a contour *enclosing* the i^{th} hole and the orientations of the contours are the same.

Proof. We can construct a closed contour which does not enclose any holes by combining the initial contour with contours that enclose each hole, where the two are oriented opposite to each other. In the limit, the path joining the contours does not contribute to the integral and can be omitted from the expression, in which case we would have

$$0 = \oint_{\gamma} f(z)\,\mathrm{d}z = \oint_{-\gamma_0} f(z)\,\mathrm{d}z + \sum_{i=1}^{n} \oint_{\gamma_i} f(z)\,\mathrm{d}z\,,$$

where we have used Cauchy's theorem. The minus sign in front of γ_0 denotes the opposite orientation. An example of this for a domain with two holes is shown below.



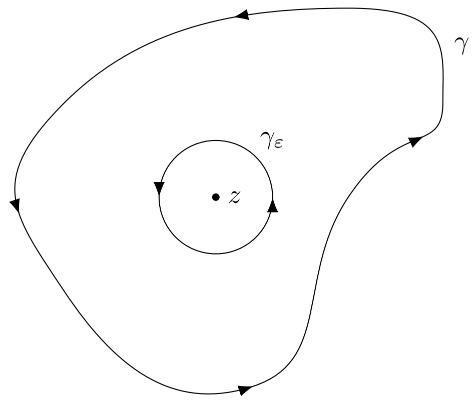
The function is analytic on the light gray region, and we take the limit as the red and blue dashed lines approach each other. The regions where the function is not analytic need not be circular.

We can use this to artificially create a singluar hole in a simply connected domain, and relate the value of the function at that point to an integral which encloses it. This is the idea behind Cauchy's integral formula.

Theorem (Cauchy's integral formula)**.** Suppose f is analytic on a simply connected domain $\mathcal{D} \subseteq \mathbb{C}$, and let $z \in \mathcal{D}$. Then

$$f(z) = \frac{1}{2\pi i} \oint_\gamma \frac{f(\xi)}{\xi - z} \, \mathrm{d}\xi \,,$$

for any *anticlockwise closed contour* $\gamma \in \mathcal{D}$ enclosing z.



Proof. We note that the integrand is undefined at $\xi = z$, hence its domain is not simply connected. Let $\gamma_\varepsilon = \partial B_\varepsilon(z)$ be a circle of radius ε centered at z. Then, by the previous proposition we have

$$\oint_\gamma \frac{f(\xi)}{\xi - z} \, \mathrm{d}\xi = \oint_{\gamma_\varepsilon} \frac{f(\xi)}{\xi - z} \, \mathrm{d}\xi \,.$$

We can evaluate the right hand side by letting $\xi = z + \varepsilon \mathrm{e}^{i\theta}$:

$$\begin{aligned} \oint_{\gamma_\varepsilon} \frac{f(\xi)}{\xi - z} \, \mathrm{d}\xi &= \int_0^{2\pi} \frac{f(z + \varepsilon \mathrm{e}^{i\theta})}{\varepsilon \mathrm{e}^{i\theta}} i\varepsilon \mathrm{e}^{i\theta} \, \mathrm{d}\theta \\ &= i \int_0^{2\pi} (f(z) + O(\varepsilon)) \, \mathrm{d}\theta \\ &\to 2\pi i f(z) \end{aligned}$$

in the limit $\varepsilon \to 0$. The result follows. □

Note. This result tells us that knowing f on a closed boundary γ, we can compute the value of f at *any point* enclosed by the boundary. One way of looking at this is to express f as $f = u + iv$. Since f is analytic, u and v are solutions to Laplace's equation. Stating the value of f on a closed boundary is equivalent to Dirichlet boundary conditions, which produces a unique solution - in this case the function f.

Corollary (Derivatives of f)**.** We can now compute the first derivative of f as:

$$f'(z) = \frac{1}{2\pi i} \oint_\gamma \frac{\partial}{\partial z} \left(\frac{f(\xi)}{\xi - z} \right) \mathrm{d}\xi = \frac{1}{2\pi i} \oint_\gamma \frac{f(\xi)}{(\xi - z)^2} \, \mathrm{d}\xi \,.$$

Similarly, the n^{th} derivative is given by

$$f^{(n)}(z) = \frac{n!}{2\pi i} \oint_\gamma \frac{f(\xi)}{(\xi - z)^{n+1}} \, \mathrm{d}\xi \,.$$

Hence at any point $z \in \mathcal{D}$ where f is analytic, *all* of its derivatives exist.

Theorem (Liouville's theorem)**.** Any bounded entire function is constant.

Proof. Let f be an entire function, and let $|f| \leq M$ for some $M \in \mathbb{R}$. Then, by Cauchy's integral formula we have for any $z \in \mathbb{C}$

$$f'(z) = \frac{1}{2\pi i} \oint_\gamma \frac{f(\xi)}{(\xi - z)^2} \, \mathrm{d}\xi \,.$$

Now, choose $\gamma = \partial B_r(z)$ to be a circle of radius r centered at z. Hence, we have $\xi = z + r\mathrm{e}^{i\theta}$ on γ. This gives

$$f'(z) = \frac{1}{2\pi i} \int_0^{2\pi} \frac{f(z + r\mathrm{e}^{i\theta})}{(r\mathrm{e}^{i\theta})^2} ir\mathrm{e}^{i\theta} \, \mathrm{d}\theta \; = \frac{1}{2\pi r} \int_0^{2\pi} \frac{f(z + r\mathrm{e}^{i\theta})}{\mathrm{e}^{i\theta}} \, \mathrm{d}\theta \,.$$

Now, we use the condition $|f| \leq M$:

$$\begin{aligned} |f'(z)| = \left| \frac{1}{2\pi r} \int_0^{2\pi} f(z + r\mathrm{e}^{i\theta}) \mathrm{e}^{-i\theta} \, \mathrm{d}\theta \right| &\leq \frac{1}{2\pi r} \int_0^{2\pi} |f(z + r\mathrm{e}^{i\theta})||\mathrm{e}^{-i\theta}|| \, \mathrm{d}\theta \,| \\ &\leq \frac{1}{2\pi r} \int_0^{2\pi} M| \, \mathrm{d}\theta \,| \\ &= \frac{M}{r}. \end{aligned}$$

Taking the limit as $r \to \infty$, we have that for all $z \in \mathbb{C}$

$$\lim_{r \to \infty} |f'(z)| = 0 \implies f'(z) = 0.$$

Since the first derivative is zero everywhere, the function f is constant. □

Proposition (Estimating derivatives)**.** The upper bound to $f^{(n)}$ at some point $z \in \mathcal{D}$ is given by

$$|f^{(n)}(z)| \leq n! \frac{\max_\gamma |f(\xi)|}{R^n}$$

where $\gamma = \partial B_R(z)$ and R can be as large as the distance to the nearest singularity.

Proof. By Cauchy's formula we have

$$\begin{aligned} f^{(n)}(z) &= \frac{n!}{2\pi i} \oint_\gamma \frac{f(\xi)}{(\xi - z)^{n+1}} \, \mathrm{d}\xi \\ \implies |f^{(n)}(z)| &\leq \frac{n!}{2\pi i} \oint_\gamma \left| \frac{f(\xi)}{(\xi - z)^{n+1}} \right| | \, \mathrm{d}\xi \,| \end{aligned}$$

We now let $\gamma = \partial B_R(z)$ for some $R \in \mathbb{R}$ such that $B_R(z)$ encloses only points in $\mathcal{D}$. So, we get

$$|f^{(n)}(z)| \leq \left| \frac{n!}{2\pi i} \right| \int_0^{2\pi} \frac{|f(\xi)|}{|(R\mathrm{e}^{i\theta})^{n+1}|} |R\mathrm{e}^{i\theta}|| \, \mathrm{d}\theta \,| \leq \frac{n! \max_\gamma |f(\xi)|}{2\pi R^{n+1}} 2\pi R = n! \frac{\max_\gamma |f(\xi)|}{R^n}.$$

Hence, the proposition holds. □

1.5 Taylor Series and Analytic Continuation

The Taylor series over $\mathbb{C}$ is defined similarly to the real case.

Definition (Taylor series)**.** If f is analytic at $z_0 \in \mathcal{D}$, then it has a Taylor series

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n,$$

where by definition (this can be proved as well) if the Taylor series converges, it converges to $f(z)$.

Proposition. The Taylor series of a function converges for all $z \in \mathcal{D}$ such that $|z - z_0| < R$ where R is the distance from z_0 to the nearest singularity.

Proof. By the previous proposition, we have

$$\left| \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n \right| \leq \max_{\gamma} |f(\xi)| \left| \frac{(z - z_0)^n}{R^n} \right|.$$

Since $\gamma \in \mathcal{D}$, the maximum is bounded for a given R. We then have, as $n \to \infty$ the terms in the sum approach zero if

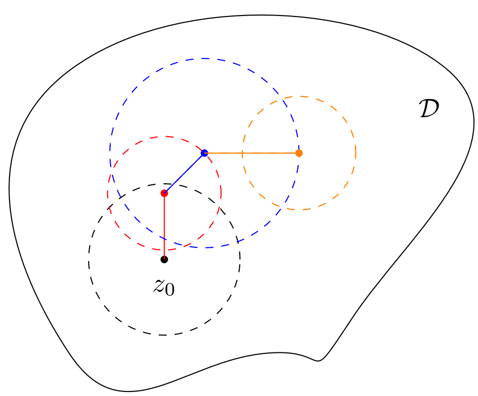
$$|z - z_0| < R.$$

Hence, we conclude the Taylor series converges over the open disk $B_R(z_0)$. $\square$

Note. We can prove that given a holomorphic function f, its Taylor series converges to itself by considering the limit $|f(z) - \sum_{n=0}^{N} f^{(n)}(z_0)(z - z_0)^n/n!|$ as $N \to \infty$. This would in turn prove that in the complex plane, analytic and holomorphic mean the same thing.

Proposition. If all derivatives $f^{(n)}(z_0)$ are zero at *some* $z_0 \in \mathcal{D}$, then $f \equiv 0$ in $\mathcal{D}$.

Proof. The Taylor series is identically zero in an open neighborhood of z_0. We can take the Taylor expansion at any point in the neighborhood, and the function will be identically zero everywhere in the new neighborhood. This can be repeated to cover the entire domain, as it is open.



We can expand the function over new regions, starting from z_0 and covering the entire domain. The function is identically zero everywhere.

Proposition. If for all n, $f^{(n)}(z_0) = g^{(n)}(z_0)$ at some point $z_0 \in \mathcal{D}$ then $f(z) \equiv g(z)$ in $\mathcal{D}$.

Proof. We have $f^{(n)}(z_0) - g^{(n)}(z_0) \equiv 0$ for all n. The result follows from the previous proposition.

Proposition. If $f \equiv g$ in some open subdomain of $\mathcal{D}$, then $f \equiv g$ everywhere in $\mathcal{D}$.

Proof. All the derivatives in the subdomain are equal, hence the result follows from the previous proposition.

Analytic continuation: We can use Taylor series to extend the domains of functions. This is known as *analytic continuation.* The procedure is as follows: suppose an analytic function is defined on an open interval which includes the point $a \in \mathbb{C}$. If we want to know the value of the function at some $b \in \mathbb{C}$ which is not in the interval, we draw a contour from a to b. Along the contour, we take the Taylor series of the function, and if we are lucky we get bigger regions where the function is defined (which we can as long as we don't hit singularities.)

The result of the analytic continuation can depend on the contour we draw. If there are singularities in the region bounded by two contours, then the function would not be analytic and our answer will not be unique.

Example. The logarithm of $x \in \mathbb{R}$ is defined as the function $f(x)$ that satisfies the following:

$$\mathrm{e}^{f(x)} = x.$$

We extend this definition to $\mathbb{C}$ by simply letting x be complex, such that

$$\mathrm{e}^{f(z)} = z.$$

Now, we can differentiate to obtain

$$f'(z) = \frac{1}{z} \implies \log(z) = \int_1^z \frac{\mathrm{d}z'}{z'}.$$

Now, we can choose any path from $z' = 1$ to $z' = z$ to obtain an expression for the logarithm. Note that our answer will not be unique because of the singularity at $z = 0$. This singularity is known as a *branch point*, and in order to have a single valued function we have to define *branch cuts.*

Definition (Branch point)**.** Branch points of $f(z)$ are points such that *any* neighborhood contains contours around which $f(z)$ varies continuously but does not return to its original value. The function is said to have a branch point at infinity if $f(1/z)$ has a branch point at 0.

For example, the logarithm has branch points at 0 and ∞. If we wish to make $\log z$ continuous and single-valued, we must prevent any curve from encircling the origin. We do this by introducing a branch cut from $-\infty$ to 0 on the real axis. No curve is allowed to cross this cut.

Once we've decided where our branch cut is, we can use it to fix the values of θ lying in the range $(-\pi, \pi]$ and we've defined a *branch* of $\log z$.

1.6 Laurent Series, Singularities and Residues

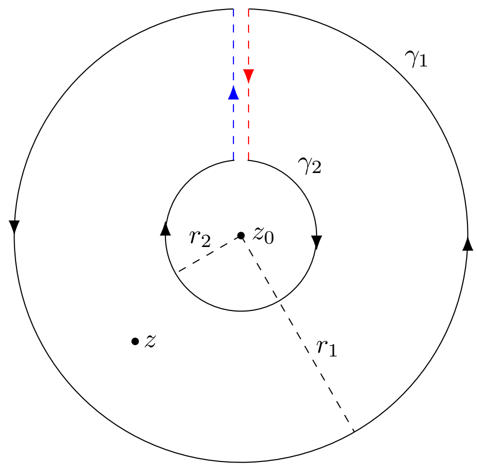
1.6.1 Laurent Series

We know that an analytic function can be expressed as a Taylor series. This implies the point around which we concern ourselves with is analytic. However, we can express a function around a singularity if we include negative powers of z in the sum.

Proposition (Laurent Series)**.** If f is analytic in an *annulus* $r_1 > |z - z_0| > r_2$, then it has a Laurent series

$$f(z) = \sum_{-\infty}^{\infty} c_n (z - z_0)^n.$$

This series converges to $f(z)$ inside the annulus.



Proof. We consider a annular contour and label the outer circle of radius r_1 as γ_1 and the inner circle of radius r_2 as γ_2. The singularity is denoted as z_0 (the behaviour of the function at any point inside γ_2 can be singular, it doesn't matter). Then, from Cauchy's integral formula we have

$$f(z) = \frac{1}{2\pi i} \oint_\gamma \frac{f(\xi)}{\xi - z} \, \mathrm{d}\xi \, .$$

We note that we have, for any z inside the region bounded by γ, $r_2 < |z - z_0| < r_1$. In the limit of cross-cuts approaching each other, we have $\gamma = \gamma_1 - \gamma_2$ where the minus sign is due to the orientation. So, the integral theorem applies:

$$f(z) = \frac{1}{2\pi i} \left\{ \oint_{\gamma_1} \frac{f(\xi)}{\xi - z} \, \mathrm{d}\xi - \oint_{\gamma_2} \frac{f(\xi)}{\xi - z} \, \mathrm{d}\xi \right\}$$

For the first term, we have

$$\begin{aligned} \oint_{\gamma_1} \frac{f(\xi)}{(\xi - z)} \, \mathrm{d}\xi &= \oint_{\gamma_1} \frac{f(\xi)}{(\xi - z_0) - (z - z_0)} \, \mathrm{d}\xi \\ &= \oint_{\gamma_1} \frac{f(\xi)}{\xi - z_0} \left(\frac{1}{1 - \frac{(z - z_0)}{\xi - z_0}} \right) \mathrm{d}\xi \\ &= \oint_{\gamma_1} \frac{f(\xi)}{\xi - z_0} \sum_{n=0}^{\infty} \left(\frac{z - z_0}{\xi - z_0} \right)^n \mathrm{d}\xi \\ &= \sum_{n=0}^{\infty} a_n (z - z_0)^n \end{aligned}$$

where

$$a_n = \oint_{\gamma_1} \frac{f(\xi)}{(\xi - z_0)^{n+1}} \, \mathrm{d}\xi \, .$$

The geometric series converges since $r_1 > |z - z_0| \Rightarrow |\xi| > |z|$.

Similary, for the second integral we have

$$\begin{aligned}
\oint_{\gamma_2} \frac{f(\xi)}{(\xi - z)}\,\mathrm{d}\xi &= \oint_{\gamma_2} \frac{f(\xi)}{(\xi - z_0) - (z - z_0)}\,\mathrm{d}\xi \\
&= \oint_{\gamma_2} \frac{-f(\xi)}{z - z_0}\left(\frac{1}{1 - \frac{(\xi - z_0)}{z - z_0}}\right)\mathrm{d}\xi \\
&= \oint_{\gamma_2} \frac{-f(\xi)}{z - z_0}\sum_{n=0}^{\infty}\left(\frac{\xi - z_0}{z - z_0}\right)^n \mathrm{d}\xi \\
&= -\sum_{n=0}^{\infty}(z - z_0)^{-(n+1)}\oint_{\gamma_2} f(\xi)(\xi - z_0)^n\,\mathrm{d}\xi \\
&= -\sum_{n=0}^{\infty} b_{n+1}(z - z_0)^{-(n+1)}
\end{aligned}$$

where

$$b_n = \oint_{\gamma_2} f(\xi)(\xi - z_0)^{n-1}\,\mathrm{d}\xi\,.$$

The geometric series converges since $r_2 < |z - z_0| \Rightarrow |\xi| < |z|$.

Combining the two results yields

$$f(z) = \frac{1}{2\pi i}\left\{\sum_{n=0}^{\infty} a_n(z - z_0)^n + \sum_{n=1}^{\infty} b_n(z - z_0)^{-n}\right\} \equiv \sum_{-\infty}^{\infty} c_n(z - z_0)^n$$

where we define c_n as

$$c_n = \begin{cases} a_n/2\pi i, & n \geq 0 \\ b_{-n}/2\pi i, & n < 0 \end{cases}.$$

Hence, we get the Laurent series. □

Example. $f(z) = \mathrm{e}^{1/z}$. From the Taylor series expansion of e^z, we have

$$\mathrm{e}^{1/z} = 1 + \frac{1}{z} + \frac{1}{2z^2} + \cdots + \frac{1}{n!z^n} + \cdots$$

1.6.2 Taylor series as a special case of Laurent series

If we have a function f analytic everywhere enclosed by the outer boundary of the annulus, we can make the inner contour arbitrarily small. This, in turn would imply that the coefficients b_n vanish. This is obvious since

$$(\xi - z_0)^n = (r_2^n \mathrm{e}^{in\varphi}),$$

which approaches zero as $r_2 \to 0$. When $n = 0$, we have

$$b_1 = \oint_{\gamma_2} f(\xi)\,\mathrm{d}\xi = \int_0^{2\pi} f(\xi) r_2 i \mathrm{e}^{i\varphi}\,\mathrm{d}\varphi\,,$$

which again tends to zero as $r_2 \to 0$. Hence, we are left only with $n \geq 0$ terms in the Laurent series:

$$f(z) = \sum_{n=0}^{\infty} c_n(z - z_0)^n.$$

Now, we are in a position to evaluate the coefficients c_n. By definition,

$$c_n = \frac{1}{2\pi i}\oint_{\gamma_1} \frac{f(\xi)}{(\xi - z_0)^{n+1}}\,\mathrm{d}\xi = \frac{f^{(n)}(z_0)}{n!},$$

by Cauchy's formula. Hence, we have

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n,$$

which is the Taylor series of f around z_0.

1.6.3 Singularities and zeros

Singularities and zeros are, in some sense, opposites of each other.

Definition (Zeros)**.** The *zeros* of an *analytic* function f are the points z_0 such that $f(z_0) = 0$. A zero is of *order* N if in the function's Taylor expansion the first non-zero coefficient is a_N.

Definition (Simple zero)**.** A zero of order 1 is called a *simple zero.*

Definition (Isolated singularity)**.** Let the function f be singular at a point z_0. This point is called an *isolated singularity* if there exists a neighborhood of z_0 such that the function f is analytic at every point in the neighborhood (except z_0).

If no such neighborhood exists, the singularity is *non-isolated.*

Example. $z = 0$ is a non-isolated singularity of $\log(z)$ since it is not analytic on any point on its branch cut.

Example. The function $f = 1/\sin(1/z)$ has a non-isolated singularity at $z = 0$, as there always exists singularities in any neighborhood of 0.

If we have an isolated singularity around a point, then we can always draw an annulus $0 < |z - z_0| < r$ within which the function is analytic and has a Laurent series. The different types of singularities are as follows

1. *Branch point singularity.* These are not isolated, and are not easy to work with.

2. *Poles.* Isolated singularity. If the negative power terms of the Laurent series terminate at some integer $N < 0$, the singularity is called a *pole.*

3. *Essential singularity.* If an isolated singularity is not a pole, then it is an *essential isolated singularity.*

4. *Non-isolated singularity.* If a singularity is not isolated, then it is non-isolated. These are difficult to work with.

Example. $f(z) = 1/z + 1/(z + 1)$ has poles at 0 and -1.

1.6.4 Residues

Definition (Residue)**.** The *residue* of a function f at an isolated singularity is the coefficient c_{-1} in its Laurent series around the singularity. We denote the residue of function f at $z = z_0$ as $\mathrm{Res}(f, z_0)$.

Proposition (Residue at simple poles)**.** The residue of a function at a simple pole is simply given by

$$\mathrm{Res}(f, z_0) = \lim_{z \to z_0} f(z)(z - z_0).$$

Proof. Writing f as its Laurent series, we have

$$f(z)(z-z_0) = f_{-1} + f_0(z-z_0) + f_1(z-z_0)^2 + \cdots$$

It is obvious that in the limit $z \to z_0$, the expression converges to the residue. □

Proposition (Residue at poles of order N)**.** The residue of a function at a pole of order N is given by

$$\operatorname{Res}(f, z_0) = \frac{1}{(N-1)!} \lim_{z \to z_0} \frac{\mathrm{d}^{N-1}}{\mathrm{d}z^{N-1}} (f(z)(z-z_0)^N).$$

Proof. It is easiest to derive this expression from the beginning. We consider $f(z)(z-z_0)^N$, as this would get rid of any singularities in the limit.

$$f(z)(z-z_0)^N = f_{-N} + f_{-N+1}(z-z_0) + \cdots + f_{-1}(z-z_0)^{N-1} + \cdots$$

Now, as we want the f_{-1} term, we can differentiate the expression $(N-1)$ times.

$$\frac{\mathrm{d}^{N-1}}{\mathrm{d}z^{N-1}} f(z)(z-z_0)^N = f_{-1}(N-1)! + f_0(z-z_0)N! + o(z-z_0).$$

Taking the limit as $z \to z_0$ yields

$$\lim_{z \to z_0} \frac{\mathrm{d}^{N-1}}{\mathrm{d}z^{N-1}} f(z)(z-z_0)^N = f_{-1}(N-1)!$$

and hence the proposition holds. □

1.7 The Calculus of Residues

Theorem (Residue theorem)**.** Given a function f analytic on a domain $\mathcal{D}$ except at a finite number of isolated singularities $z_1, z_2, \dots, z_n$, the integral over a contour $\gamma \in \mathcal{D}$ oriented anti-clockwise, enclosing z_i is

$$\oint_\gamma f(z)\,\mathrm{d}z = 2\pi i \sum_k \operatorname{Res}(f, z_k).$$

Proof. We take cuts from γ such that we exclude any singularities by circling around them. We denote this new contour as γ', and the contour around the singularity at z_i as $-\gamma_i$. Note that each $-\gamma_i$ is oriented clockwise. Then, the function f is analytic everywhere enclosed by γ'. By Cauchy's theorem, we have

$$0 = \oint_{\gamma'} f(z)\,\mathrm{d}z = \oint_\gamma f(z)\,\mathrm{d}z - \sum_k \oint_{\gamma_k} f(z)\,\mathrm{d}z\,,$$

where the minus sign is due to the orientation of γ_k. This implies

$$\oint_\gamma f(z)\,\mathrm{d}z = \sum_k \oint_{\gamma_k} f(z)\,\mathrm{d}z\,.$$

Now, we consider only the integral around a single γ_k. We let $\gamma_k = \partial B_r(z_k)$ for some r. Taking the Laurent series around z_k, we have

$$\begin{aligned}\oint_{\partial B} f(z)\,\mathrm{d}z &= \oint_{\partial B} \sum_{-\infty}^{\infty} f_n (z-z_k)^n \,\mathrm{d}z \\ &= \sum_{-\infty}^{\infty} f_n \oint_{\partial B} (z-z_0)^n \,\mathrm{d}z \\ &= \sum_{-\infty}^{\infty} f_n \int_0^{2\pi} (r\mathrm{e}^{i\varphi})^n ir\mathrm{e}^{i\varphi}\,\mathrm{d}\varphi \\ &= \sum_{-\infty}^{\infty} ir^{n+1} f_n \int_0^{2\pi} \mathrm{e}^{(n+1)i\varphi}\,\mathrm{d}\varphi \\ &= 2\pi i f_{-1}.\end{aligned}$$

The integral is non-zero only when $n = -1$, so the sum collapses. By taking the sum over all singularities, we obtain the residue theorem. □

1.7.1 Integration with residues

Let's use the residue theorem to calculate real integrals.

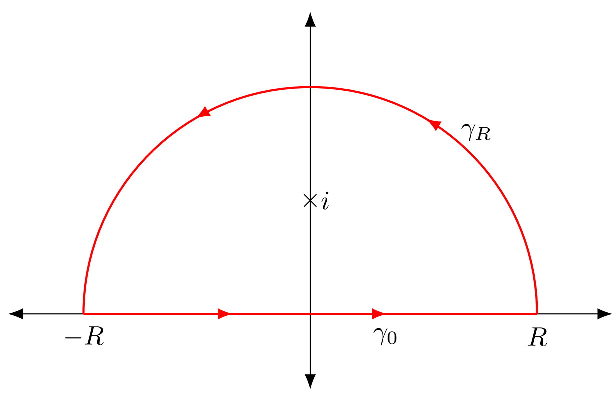
Example. Consider the integral

$$I = \int_0^{\infty} \frac{\mathrm{d}x}{1+x^2}$$

Instead of integrating along the real line, consider the integral

$$\oint_{\gamma} \frac{\mathrm{d}z}{1+z^2}$$

where we take the contour $\gamma = \gamma_R + \gamma_0$ as



We notice that, since we have an even integral, we have

$$2I = \int_{-\infty}^{\infty} \frac{\mathrm{d}x}{1+x^2} = \int_{\gamma_0} \frac{\mathrm{d}z}{1+z^2}$$

in the limit $R \to \infty$. Let's look at the integral over γ_R as $R \to \infty$. We have

$$\int_{\gamma_R} \frac{\mathrm{d}z}{1+z^2} \leq \pi R \max_{z\in\gamma_R} \left|\frac{1}{1+z^2}\right| = \pi R \cdot O(R^{-2}) = O(R^{-1}) \to 0.$$

Hence, we conclude

$$I = \frac{1}{2}\oint_\gamma \frac{\mathrm{d}z}{1+z^2} = \frac{\pi}{i} \cdot \mathrm{Res}(f, z=i)$$

since the only singularity we are enclosing is at $z = i$. This yields

$$I = \frac{\pi}{2}.$$

We see, from this example, that given an integral of the form $\int_{-\infty}^{\infty} f \,\mathrm{d}x$, we can choose a contour shaped like a semi-circle, taking the limit as the radius approaches infinity. The main issue given such contour is to determine whether the integral along the circular section approaches zero. We require the following:

$$\lim_{R\to\infty} R \cdot \max_\theta (f(R\mathrm{e}^{i\theta})) = 0.$$

There are a variety of different forms to consider. Let's look at integrals with branch cuts.

Proposition. Given a function f that is not singular at $z = 0$, we have

$$\int_0^\infty f(x)\,\mathrm{d}x = -\sum_{a_k} \mathrm{Res}(f(z)\log(z), a_k)$$

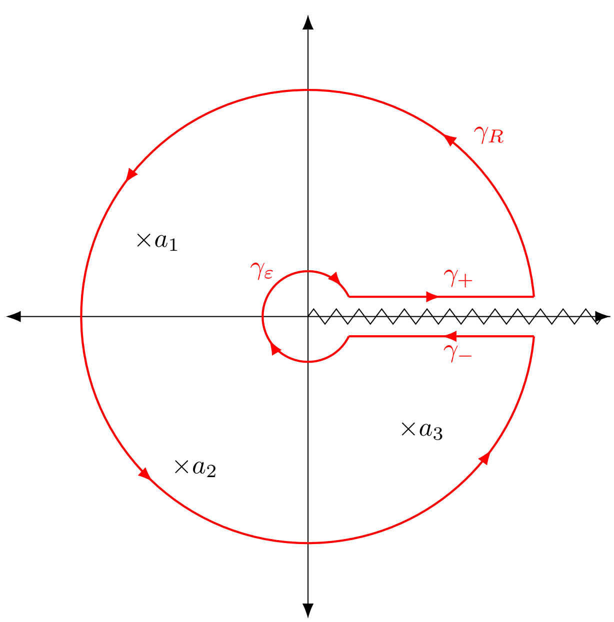
where a_k are singularities in all of the complex plane. The condition for convergence is

$$\lim_{R\to\infty} R\log(R) \max_\theta |f(R\mathrm{e}^{i\theta})| = 0.$$

Proof. Consider the integral

$$\oint_\gamma f(z)\log(z)\,\mathrm{d}z\,,$$

where the contour γ is the keyhole contour shaped as follows:



where we have used the positive real axis as a branch cut for the logarithm. Therefore, we define

$$\log(z) := \log(r) + i\varphi, \quad \varphi \in (0, 2\pi).$$

Now, our contour encloses all of the complex plane in the limit $R \to \infty$ and $\varepsilon \to 0$. Then, by the residue theorem

$$\oint_\gamma f(z)\log(z)\,\mathrm{d}z = 2\pi i \sum_{a_k} \operatorname{Res}(f(z)\log(z), a_k).$$

If $R\log(R)\max(|f|) \to 0$ as $R \to \infty$, the contribution from γ_R vanishes. If f is not singular at $z = 0$, we have $\varepsilon\log(\varepsilon) \to 0$ as $\varepsilon \to 0$. So, the integral over γ becomes

$$\begin{aligned}\oint_\gamma f(z)\log(z)\,\mathrm{d}z &= \int_{\gamma_+} f(z)\log(z)\,\mathrm{d}z + \int_{\gamma_-} f(z)\log(z)\,\mathrm{d}z \\ &= \int_0^\infty f(x)\log(x)\,\mathrm{d}x + \int_\infty^0 f(x)[\log(x) + 2\pi i]\,\mathrm{d}x \\ &= -2\pi i \int_0^\infty f(x)\,\mathrm{d}x = 2\pi i \sum_{a_k} \operatorname{Res}(f(z)\log(z), a_k).\end{aligned}$$

The result follows. □

Proposition.

$$\int_0^{2\pi} f(r\mathrm{e}^{i\varphi})\,\mathrm{d}\varphi = 2\pi \sum_{|a_k|<r} \operatorname{Res}\left(\frac{f(z)}{z}, a_k\right).$$

Proof. Let $z = r\mathrm{e}^{i\varphi}$, which implies $\mathrm{d}z = iz\,\mathrm{d}\varphi$. Hence,

$$\int_0^{2\pi} f(r\mathrm{e}^{i\varphi})\,\mathrm{d}\varphi \equiv \int_{\partial B_r(0)} \frac{f(z)}{iz}\,\mathrm{d}z = 2\pi \sum_{|a_k|<r} \operatorname{Res}\left(\frac{f(z)}{z}, a_k\right).$$

□

Proposition. Given a function $f(z)$, which is not singular at $f(z) = 0$, we have

$$\int_0^\infty f(x)\sqrt{x}\,\mathrm{d}x = \pi i \sum_{a_k} \operatorname{Res}(f(z)\sqrt{z}, a_k).$$

given that

$$\lim_{R\to\infty} R\sqrt{R}\max_\theta |f(R\mathrm{e}^{i\theta}) = 0.|$$

Proof. Consider the same keyhole contour with a branch cut along the positive real axis. We then have

$$\sqrt{z} := \sqrt{r}\mathrm{e}^{i\varphi/2}, \quad \varphi \in (0, 2\pi).$$

Since f is not singular at $z = 0$, the integral over γ_ε vanishes. Similarly, the integral over γ_R vanishes because of the condition stated above. Hence,

$$\begin{aligned}2\pi i \sum_{a_k} \operatorname{Res}(f(z)\sqrt{z}, a_k) = \oint_\gamma f(z)\sqrt{z}\,\mathrm{d}z &= \int_{\gamma_+} f(z)\sqrt{z}\,\mathrm{d}z + \int_{\gamma_-} f(z)\sqrt{z}\,\mathrm{d}z \\ &= \int_0^\infty f(x)\sqrt{x}\,\mathrm{d}x + \int_\infty^0 f(x)\sqrt{x}\mathrm{e}^{i\pi}\,\mathrm{d}x \\ &= 2\int_0^\infty f(x)\sqrt{x}\,\mathrm{d}x.\end{aligned}$$

The result follows. □

We can keep deriving new formulas using the keyhole contour with branch cuts. Finally, we need to look at the condition for convergence of a particular type of integral which comes up all the time when doing Fourier transforms.

1.7.2 Jordan's lemma

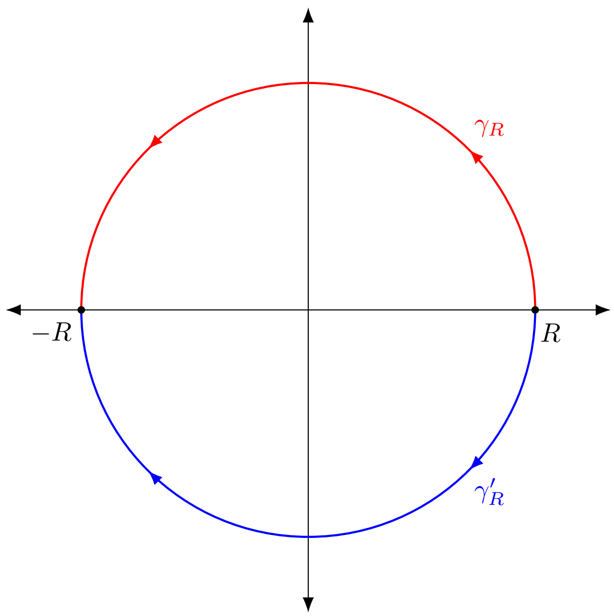
Lemma (Jordan's lemma)**.** Given an analytic function f which has a finite number of singularities, if $f(z) \to 0$ as $|z| \to \infty$ we have

$$\lim_{R\to\infty} \int_{\gamma_R} f(z)\mathrm{e}^{i\lambda z}\,\mathrm{d}z = 0, \quad \forall\, \lambda > 0,$$

and

$$\lim_{R\to\infty} \int_{\gamma_R'} f(z)\mathrm{e}^{i\lambda z}\,\mathrm{d}z = 0, \quad \forall\, \lambda < 0.$$

The contours γ_R and γ_R' are defined as semicircles of radius R above and below the real line (excluding the real line) respectively.



Proof. First, we note that in the interval $\theta \in [0, \pi/2]$, we have $\sin\theta \geq 2\theta/\pi$. *(This is obvious if you plot.)* Now, consider

$$\begin{aligned}
\left|\int_{\gamma_R} f(z)\,\mathrm{d}z\right| &= \left|\int_0^\pi f(R\mathrm{e}^{i\theta})\mathrm{e}^{i\lambda R\mathrm{e}^{i\theta}} iR\mathrm{e}^{i\theta}\,\mathrm{d}\theta\right| \\
&\leq R\max_\theta |f| \int_0^\pi \left|\mathrm{e}^{i\lambda R\cos\theta}\right| \cdot \left|\mathrm{e}^{-\lambda R\sin\theta}\right|\,\mathrm{d}\theta \\
&\leq 2R\max_\theta |f| \int_0^{\pi/2} \mathrm{e}^{-\lambda R\sin\theta}\,\mathrm{d}\theta \\
&\leq 2R\max_\theta |f| \int_0^{\pi/2} \mathrm{e}^{-\lambda R 2\theta/\pi}\,\mathrm{d}\theta \\
&= \frac{\pi}{\lambda}\max_\theta |f|(1 - \mathrm{e}^{-\lambda R}) = O(\max|f|).
\end{aligned}$$

Hence, if $f \to 0$ as $R \to \infty$, the integral over γ_R vanishes. Same for γ_R' if $\lambda < 0$.

2 Fourier Transforms

2.1 The Delta Function and its Properties

The delta function is not an actual function, but it is a very useful object to consider. There are many ways to define it, we choose the following:

Definition (Delta function)**.** The *delta function*, denoted $\delta(x)$ for $x \in \mathbb{R}$, is defined as a *linear functional* with the following property:

$$\int_{-\infty}^{\infty} \delta(x) f(x)\,\mathrm{d}x = f(0), \quad \forall\, f : \mathbb{R} \to \mathbb{C}, \tag{2.1}$$

as long as $f(0)$ is well defined.

We now consider its properties, all following from the definition.

Proposition. The delta function can be expressed as the limit $\varepsilon \to 0$ of the function

$$h(x) = \begin{cases} 1/\varepsilon, & -\varepsilon/2 < x < \varepsilon/2, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Let $f : \mathbb{R} \to \mathbb{R}$, and consider

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} f(x) h(x)\,\mathrm{d}x = \lim_{\varepsilon \to 0} \int_{\varepsilon/2}^{\varepsilon/2} \frac{1}{\varepsilon} f(x)\,\mathrm{d}x\,.$$

Assuming f is sufficiently well behaved in the neighborhood of $x = 0$, we can express it as

$$|f(x)| = |f(0) + f'(0)x + o(x)| \leq |f(0) + f'0(\varepsilon) + o(\varepsilon)|$$

since $|x|$ is bounded by ε. The integral reduces to

$$\lim_{\varepsilon \to 0} \int_{\varepsilon/2}^{\varepsilon/2} \frac{1}{\varepsilon} f(x)\,\mathrm{d}x = f(0)\frac{\varepsilon/2 + \varepsilon/2}{\varepsilon} + \lim_{\varepsilon \to 0} o(\varepsilon) = f(0).$$

This result holds for any f continuous at zero. □

Proposition. The delta function is the derivative of the step function,

$$\delta(x) = \frac{\mathrm{d}\Theta}{\mathrm{d}x}.$$

Proof. Consider the integral,

$$\int_{-\infty}^{\infty} \Theta(x) f(x)\,\mathrm{d}x = \int_{0}^{\infty} f(x)\,\mathrm{d}x = \lim_{\alpha \to \infty} F(\alpha) - F(0),$$

where $F' = f$. We can rewrite the integral by integration by parts as follows:

$$\begin{aligned} \int_{-\infty}^{\infty} \Theta(x) f(x)\,\mathrm{d}x &= F\Theta\big|_{0}^{\infty} - \int_{-\infty}^{\infty} \Theta' F\,\mathrm{d}x \\ &= \lim_{\alpha \to \infty} F(\alpha) - \int_{-\infty}^{\infty} \Theta' F\,\mathrm{d}x \\ \implies \int_{-\infty}^{\infty} \Theta' F\,\mathrm{d}x &= F(0) \\ \implies \Theta'(x) = \delta(x). \quad \square \end{aligned}$$

Proposition. We can shift the delta function, such that for any $a \in \mathbb{R}$ we have

$$\int_{-\infty}^{\infty} \delta(x-a) f(x) \, \mathrm{d}x = f(a),$$

assuming $f(a)$ is well defined.

Proof. Consider the substitution $u = x - a$,

$$\int_{-\infty}^{\infty} \delta(x-a) f(x) \, \mathrm{d}x = \int_{-\infty}^{\infty} \delta(u) f(u+a) \, \mathrm{d}u = f(a)$$

by definition. □

Proposition. We can scale the argument of $\delta(x)$ and obtain

$$\delta(\lambda x) = \frac{1}{|\lambda|} \delta(x).$$

Proof. Letting $u = \lambda x$,

$$\int_{-\infty}^{\infty} f(x) \delta(\lambda x) \, \mathrm{d}x = \int_{-\infty}^{\infty} \frac{\mathrm{d}u}{|\lambda|} f(u/\lambda) \delta(u) \, \mathrm{d}x = \frac{f(0)}{|\lambda|}.$$

The results follows. □

Corollary. The delta function is even.

Proof. Let $\lambda = -1$ and the results follows from the previous proposition.

Corollary.

$$\int_{0}^{\infty} f(x) \delta(x) \, \mathrm{d}x = \frac{f(0)}{2}.$$

Proof. As the delta function is even, we have

$$\frac{f(0)}{2} = \frac{1}{2} \int_{-\infty}^{\infty} \delta(x) f(x) \, \mathrm{d}x = \int_{0}^{\infty} \delta(x) f(x) \, \mathrm{d}x \,. \quad \square$$

Proposition. Consider a function $\varphi : \mathbb{R} \to \mathbb{R}$, and define $\mathcal{X} = \{x \in \mathbb{R} \,|\, \varphi(x) = 0\}$ such that φ is continuous at all x_n. Then, we have

$$\delta(\varphi(x)) = \sum_{x_n \in \mathcal{X}} \frac{\delta(x - x_n)}{|\varphi'(x_n)|}.$$

Proof. Consider the following integral:

$$\int_{-\infty}^{\infty} \delta(\varphi(x)) f(x) \, \mathrm{d}x \,.$$

We split the integral up as a sum of integrals around the zeros of φ and take the limit as the width of the ranges of integration approach to zero:

$$\int_{-\infty}^{\infty} \delta(\varphi(x)) f(x) \, \mathrm{d}x = \sum_{x_n \in \mathcal{X}} \lim_{\varepsilon \to 0} \int_{x_n - \varepsilon}^{x_n + \varepsilon} \delta(\varphi(x)) f(x) \, \mathrm{d}x \,.$$

Now, we note that in the neighborhood of each x_n, φ can be expressed as

$$\varphi(x) = \varphi(x_n) + \varphi'(x_n)(x - x_n) + o(x - x_n) = \varphi'(x_n)(x - x_n) + o(x - x_n)$$

where $|x_n - x| \leq \varepsilon$. The integral becomes

$$\begin{aligned}
\int_{-\infty}^{\infty} \delta(\varphi(x)) f(x)\,\mathrm{d}x &= \sum_{x_n \in \mathcal{X}} \lim_{\varepsilon \to 0} \int_{x_n - \varepsilon}^{x_n + \varepsilon} \delta(\varphi'(x_n)(x - x_n) + o(x - x_n)) f(x)\,\mathrm{d}x \\
&= \sum_{x_n \in \mathcal{X}} \lim_{\varepsilon \to 0} \int_{x_n - \varepsilon}^{x_n + \varepsilon} \delta(\varphi'(x_n)(x - x_n)) f(x)\,\mathrm{d}x \\
&= \sum_{x_n \in \mathcal{X}} \frac{f(x - x_n)}{|\varphi'(x_n)|} \\
\implies \delta(\varphi(x)) &= \sum_{x_n \in \mathcal{X}} \frac{\delta(x - x_n)}{|\varphi'(x_n)|},
\end{aligned}$$

where the previous two propositions were used. □

Proposition. The derivative of the delta function is another linear functional with the property

$$\int_{-\infty}^{\infty} \delta'(x) f(x)\,\mathrm{d}x = -f'(0).$$

Proof. Simply integrate by parts.

$$\int_{-\infty}^{\infty} \delta'(x) f(x)\,\mathrm{d}x = \delta(x) f(x)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \delta(x) f'(x)\,\mathrm{d}x = -f'(0). \quad \square$$

Proposition. The delta function can be given by the integral

$$\int_{-\infty}^{\infty} \mathrm{e}^{ikx}\,\mathrm{d}k = 2\pi\delta(x).$$

Proof. Simply compute, in the limit $\Omega \to \infty$, the integral

$$\int_{-\Omega}^{\Omega} \mathrm{e}^{ikx}\,\mathrm{d}k = \frac{\mathrm{e}^{ikx}}{ix}\Bigg|_{-\Omega}^{\Omega} = \frac{2\sin(\Omega x)}{x}.$$

Now, use the defining property of the delta function, given a function $f : \mathbb{R} \to \mathbb{R}$ and letting $u = \Omega x$,

$$\int_{-\infty}^{\infty} \frac{2\sin(\Omega x)}{x} f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{2\sin(u)}{u} f(u/\Omega)\,\mathrm{d}u\,.$$

We can express $f(u/\Omega)$ near $u = 0$ as

$$\begin{aligned}
f(u/\Omega) &= f(0) + f'(0)(u/\Omega) + o(u/\Omega), \\
\implies \lim_{\Omega \to \infty} f(u/\Omega) &= f(0).
\end{aligned}$$

The integral reduces to

$$\int_{-\infty}^{\infty} 2f(0)\frac{\sin(u)}{u}\,\mathrm{d}u = 2\pi f(0).$$

Hence, the proposition holds. □

2.2 Fourier Transforms and their Properties

We define the Fourier transform and Fourier integrals, and look at some of their key properties. Then, we look at their applications to linear differential equations with constant coefficients.

Definition (Fourier transform)**.** Given a function $f : \mathbb{R} \to \mathbb{C}$, its *Fourier transform*, denoted $\hat{f}(k)$, is defined as

$$\mathscr{F}[f(x)](k) = \hat{f}(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) \mathrm{e}^{-ikx} \, \mathrm{d}x \,, \tag{2.2}$$

given the integral exists.

Definition (Fourier integral)**.** A function $f(x)$ may be represented in terms of its Fourier transform $\hat{f}(k)$, as

$$\mathscr{F}^{-1}[\hat{f}(k)](x) = f(x) = \int_{-\infty}^{\infty} \hat{f}(k) \mathrm{e}^{ikx} \, \mathrm{d}k \,. \tag{2.3}$$

The *Fourier inversion theorem* is an equivalent statement to the definition of the Fourier integral.

Theorem (Fourier inversion theorem)**.** Given a function $f : \mathbb{R} \to \mathbb{C}$ such that $\hat{f}(k)$ exists, then

$$f(x) = \mathscr{F}^{-1}\left[\mathscr{F}[f(x)](k)\right](x).$$

Proof. Using the definition of Fourier transform, we write

$$\begin{aligned} \mathscr{F}^{-1}[\hat{f}(k)](x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}k \int_{-\infty}^{\infty} \mathrm{d}y \, \mathrm{e}^{-iky} \mathrm{e}^{ikx} f(y) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}y \, f(y) \int_{-\infty}^{\infty} \mathrm{d}k \, \mathrm{e}^{ik(x-y)}. \end{aligned}$$

From the properties of the delta function, we have

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{ik(x-y)} \, \mathrm{d}k \; = \delta(x-y)$$

which implies

$$\int_{-\infty}^{\infty} \mathrm{d}y \, f(y) \int_{-\infty}^{\infty} \mathrm{d}k \, \mathrm{e}^{ik(x-y)} = \int_{-\infty}^{\infty} \mathrm{d}y \, \delta(x-y) f(y) = f(x). \quad \square$$

2.2.1 Properties of Fourier transforms

Proposition. Fourier transform is a linear operator.

Proof. Consider $(af(x) + bg(x))$ for some $f, g : \mathbb{R} \to \mathbb{C}$ and $a, b \in \mathbb{C}$. The Fourier transform is given by

$$\begin{aligned} \mathscr{F}[af(x) + bg(x)](k) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (af(x) + bg(x)) \mathrm{e}^{-ikx} \, \mathrm{d}x \\ &= \frac{a}{2\pi} \int_{-\infty}^{\infty} f(x) \mathrm{e}^{-ikx} \, \mathrm{d}x \, + \frac{b}{2\pi} \int_{-\infty}^{\infty} g(x) \mathrm{e}^{-ikx} \, \mathrm{d}x \\ &= a\hat{f}(k) + b\hat{g}(k). \quad \square \end{aligned}$$

Proposition. Given $g(x) = f(\lambda x)$ for some $\lambda \in \mathbb{R} \setminus \{0\}$, we have $\hat{g}(k) = \hat{f}(k/\lambda)/|\lambda|$.

Proof. Let $u = \lambda x$,

$$\begin{aligned}
\hat{g}(k) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(\lambda x)\mathrm{e}^{-ikx}\,\mathrm{d}x \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(u)\mathrm{e}^{-iu(k/\lambda)}\frac{1}{|\lambda|}\,\mathrm{d}u \\
&= \frac{\hat{f}(k/\lambda)}{|\lambda|}. \quad \square
\end{aligned}$$

Proposition. $\mathscr{F}[f(-x)](k) = \hat{f}(-k)$.

Proof. Let $\lambda = -1$ in the previous proposition and the result follows.

Proposition. Given $g(x) = f(x-a)$ for some $a \in \mathbb{R}$, $\hat{g}(k) = \mathrm{e}^{-ika}\hat{f}(k)$.

Proof. Let $u = x - a$,

$$\begin{aligned}
\hat{g}(k) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x-a)\mathrm{e}^{-ikx}\,\mathrm{d}x \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(u)\mathrm{e}^{-i(u+a)k}\,\mathrm{d}u \\
&= \frac{1}{2\pi}\mathrm{e}^{-ika}\int_{-\infty}^{\infty} f(u)\mathrm{e}^{-iku}\,\mathrm{d}u \\
&= \mathrm{e}^{-ika}\hat{f}(k). \quad \square
\end{aligned}$$

Proposition. Given $g(x) = \mathrm{e}^{ipx}f(x)$ for some $p \in \mathbb{R}$, $\hat{g}(k) = \hat{f}(k-p)$.

Proof.

$$\begin{aligned}
\hat{g}(k) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\mathrm{e}^{ipx}\mathrm{e}^{-ikx}\,\mathrm{d}x \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\mathrm{e}^{-ix(k-p)}\,\mathrm{d}x \\
&= \hat{f}(k-p). \quad \square
\end{aligned}$$

Proposition. Given $g(x) = f'(x)$, we have $\hat{g}(k) = ik\hat{f}(k)$

Proof. Simply integrate by parts.

$$\begin{aligned}
\hat{g}(k) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f'(x)\mathrm{e}^{-ikx}\,\mathrm{d}x \\
&= \frac{1}{2\pi}\left\{ f(x)\mathrm{e}^{-ikx}\big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} ik\mathrm{e}^{-ikx}f(x)\,\mathrm{d}x \right\} \\
&= ik\hat{f}(k),
\end{aligned}$$

assuming that $\lim_{|x|\to\infty} f(x) = 0$. $\square$

Corollary. Given $g(x) = f^{(n)}(x)$, $\hat{g}(k) = (ik)^n\hat{f}$.

Proof. By induction from previous result. There is a condition that $f^{(m)} \to 0$ as $x \to \pm\infty$, for all $m < n$.

Theorem (Plancherel's theorem)**.** For a given $f : \mathbb{R} \to \mathbb{C}$, provided $\hat{f}(k)$ exists, we have

$$\int_{-\infty}^{\infty} |f(x)|^2 \, \mathrm{d}x = 2\pi \int_{-\infty}^{\infty} |\hat{f}(k)|^2 \, \mathrm{d}k \,. \tag{2.4}$$

Proof. We write $f(x)$ as a Fourier integral *(or $\hat{f}(k)$ as a Fourier transform, it doesn't matter)*,

$$\begin{aligned} \int_{-\infty}^{\infty} |f(x)|^2 \, \mathrm{d}x &= \int_{-\infty}^{\infty} \mathrm{d}x \left| \int_{-\infty}^{\infty} \mathrm{d}k \, \hat{f} \mathrm{e}^{ikx} \right|^2 \\ &= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}k_1 \, \hat{f}(k_1) \mathrm{e}^{ik_1 x} \int_{-\infty}^{\infty} \mathrm{d}k_2 \, \mathrm{e}^{-ik_2 x} \hat{f}^*(k_2) \\ &= \int_{-\infty}^{\infty} \mathrm{d}k_1 \int_{-\infty}^{\infty} \mathrm{d}k_2 \, \hat{f}(k_1) \hat{f}^*(k_2) \int_{-\infty}^{\infty} \mathrm{d}x \, \mathrm{e}^{ix(k_1 - k_2)} \\ &= \int_{-\infty}^{\infty} \mathrm{d}k_1 \int_{-\infty}^{\infty} \mathrm{d}k_2 \, 2\pi\delta(k_1 - k_2) \hat{f}(k_1) \hat{f}^*(k_2) \\ &= 2\pi \int_{-\infty}^{\infty} |\hat{f}(k)|^2 \, \mathrm{d}k \,. \quad \square \end{aligned}$$

Corollary. Given two function $f, g : \mathbb{R} \to \mathbb{C}$, we have

$$\int_{-\infty}^{\infty} f(x) g^*(x) \, \mathrm{d}x = 2\pi \int_{-\infty}^{\infty} \hat{f}(k) \hat{g}^*(k) \, \mathrm{d}k \,.$$

Proof. Follows from Plancherel's theorem.

Definition (Orthogonal transformation)**.** An *orthogonal transformation* over a vector space, $T : V \to V$, is a *linear transformation* that preserves length and angles between elements of V.

We can think of the set of all square integrable functions $f : \mathbb{R} \to \mathbb{C}$ with the following definition for an inner product:

$$(f, g) = \int_{-\infty}^{\infty} f(x) g^*(x) \, \mathrm{d}x$$

as a vector space.

Corollary. Up to a factor of 2π *(which doesn't matter)*, Fourier transforms are *orthogonal transformations.* We have, for all square integrable functions f and g,

$$(f, g) = 0 \iff (\hat{f}, \hat{g}) = 0.$$

Note (Differentiation and convergence)**.** Differentiation is a *sharpening* operator. Even if f is square integrable, f' is not necessarily. This is easily seen from Plancherel's theorem:

$$\int_{-\infty}^{\infty} |f'(x)|^2 \, \mathrm{d}k = 2\pi \int_{-\infty}^{\infty} k^2 |\hat{f}(k)|^2 \, \mathrm{d}k \,.$$

The factor of k^2 slows down converges or makes the integral diverge. This *sharpening* is where the extra conditions on the Fourier transforms of derivatives of functions come from.

Proposition (Bound on $\hat{f}(k)$)**.** If there exists $N \in \mathbb{R}$ such that

$$\int_{-\infty}^{\infty} |f(x)| \, \mathrm{d}x \le N,$$

we have

$$|\hat{f}(k)| \le \frac{N}{2\pi}.$$

Proof. From the definition of Fourier transform, we have

$$\hat{f}(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\mathrm{e}^{-ikx}\,\mathrm{d}x \le \frac{1}{2\pi}\int_{-\infty}^{\infty} |f(x)||\mathrm{e}^{-ikx}|\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\infty}^{\infty} |f(x)|\,\mathrm{d}x\,.$$

The results follows. □

Proposition. Given an $N \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} |f^{(n)}(x)|\,\mathrm{d}x \le N \implies \hat{f}(k) \le \frac{N}{2\pi|k|^2}.$$

Proof. Define $g(x) = f^{(n)}(x)$, so we have $\hat{g}(k) = (ik)^n\hat{f}(k)$. So, we have

$$\hat{f}(k) = \frac{\hat{g}(k)}{(ik)^n} = \frac{1}{2\pi(ik)^n}\int_{-\infty}^{\infty} f^{(n)}(x)\mathrm{e}^{-ikx}\,\mathrm{d}x \le \frac{1}{2\pi|k|^n}\int_{-\infty}^{\infty} |f^n(x)|\,\mathrm{d}x = \frac{N}{2\pi|k|^2}. \quad \square$$

2.2.2 Applications to linear differential equations

The general form of the equations we are interested in is

$$a_n y^{(n)}(x) + a_{n-1}y^{(n-1)}(x) + \ldots + a_0 y(x) = f(x),$$

where $f(x)$ is known and a_i are constants.

We can take the Fourier transform with respect to x,

$$(a_n(ik)^n + a_{n-1}(ik)^{n-1} + \ldots + a_0)\hat{y}(k) = \hat{f}(k)$$

$$\implies \hat{y}(k) = \frac{\hat{f}(k)}{P(ik)},$$

where $P(ik)$ is the characteristic polynomial. After taking the inverse transform, we would have solved the differential equation! Only if life were that simple... there are some major issues with this. We only acquired a single solution, whereas we would expect to specify n initial or boundary conditions. What happened?

We assumed that the Fourier transform of $f(x)$ and, more importantly, $y(x)$ exist. Given such constraint on y, it turns out we have a unique solution. The solution is given by

$$y(x) = \int_{-\infty}^{\infty} \frac{\hat{f}(k)}{P(ik)}\mathrm{e}^{ikx}\,\mathrm{d}k\,, \tag{2.5}$$

where $\nexists\ k \in \mathbb{R}$ such that $P(ik) = 0$. This means $P(\lambda)$ has no purely imaginary zeros.

Theorem (Convolution theorem)**.** The Fourier transform of the convolution of two functions is the product of their Fourier transforms (up to a factor of 2π).

Proof. Define, for some $\hat{f}$ and $\hat{g}$, $h(x)$ such that

$$\hat{h}(k) = 2\pi\hat{f}(k)\hat{g}(k).$$

Then, $h(x)$ can be expressed as the Fourier integral:

$$\begin{aligned} h(x) &= 2\pi\int_{-\infty}^{\infty} \hat{f}(k)\hat{g}(k)\mathrm{e}^{ikx}\,\mathrm{d}k \\ &= \int_{-\infty}^{\infty}\mathrm{d}k\int_{-\infty}^{\infty}\mathrm{d}y\, f(y)\hat{g}(k)\mathrm{e}^{ik(x-y)}, \end{aligned}$$

letting $u = x - y$, we obtain

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathrm{d}k \int_{-\infty}^{\infty} \mathrm{d}y\, f(y)\hat{g}(k)\mathrm{e}^{ik(x-y)} &= \int_{-\infty}^{\infty} \mathrm{d}u\, f(x-u) \int_{-\infty}^{\infty} \mathrm{d}k\, \hat{g}(k)\mathrm{e}^{iku} \\
&= \int_{-\infty}^{\infty} \mathrm{d}u\, f(x-u)g(u) \\
&= f * g.
\end{aligned}$$

Hence, the theorem holds. □

We can now look at the solution to the differential equation as a convolution.

$$\begin{aligned}
\hat{y}(k) &= \hat{f}(k) \cdot \frac{1}{P(ik)} \\
\implies y(k) &= f * G
\end{aligned}$$

where $\hat{G}(k) = 1/(2\pi P(ik))$. So,

$$G(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\mathrm{e}^{ikx}\,\mathrm{d}k}{P(ik)} \tag{2.6}$$

is the *Green's function*. The solution is then given by

$$y(x) = \int_{-\infty}^{\infty} f(x)G(x-u)\,\mathrm{d}u\,. \tag{2.7}$$

The factor of 2π comes from the convolution. We could get rid of it by changing our definitions of Fourier transforms and Fourier integrals, or by including it in the definition of convolution *but it really doesn't matter.*

2.2.3 Green's functions for ODE's

But what really is a Green's function?

Definition (Green's function)**.** The Green's function, $G(x)$, of a given ODE is a solution to the equation

$$a_n y^{(n)}(x) + a_{n-1}y^{(n-1)}(x) + \ldots + a_0 y(x) = L[y(x)] = \delta(x)$$

where L denotes the differential operator.

We can check the equivalence of this definition with equation (2.7). We have, from the definition of the Green's function,

$$\begin{aligned}
L\left[G(x)\right] &= \delta(x) \\
\implies \delta(x-u) &= L\left[G(x-u)\right].
\end{aligned}$$

By the properties of the delta function, we can express any $f(x)$ as

$$f(x) = \int_{-\infty}^{\infty} f(u)\delta(x-u)\,\mathrm{d}u\,.$$

Combining the two, for an ODE of the form $L[y] = f$ yields

$$\begin{aligned}
f(x) &= \int_{-\infty}^{\infty} f(u)\delta(x-u)\,\mathrm{d}u \\
&= \int_{-\infty}^{\infty} f(u)L\left[G(x-u)\right]\,\mathrm{d}u\, = L[y(x)].
\end{aligned}$$

Since the operator L acts on x, this expression is equivalent to (2.7).

Example. Consider a square integrable solution to the equation

$$y'(x) + y(x) = \delta(x).$$

The solution will be the Green's function for this ODE. Let's solve it normally and compare with (2.6).

When $x \neq 0$, we have

$$y'(x) + y(x) = 0,$$

which implies

$$y(x) = \begin{cases} A\mathrm{e}^{-x} & x > 0, \\ B\mathrm{e}^{-x} & x < 0. \end{cases}$$

Now, to find the coefficients $A, B \in \mathbb{R}$, we integrate in the neighborhood of $x = 0$:

$$\int_{-\varepsilon}^{\varepsilon} y'(x)\,\mathrm{d}x = \int_{-\varepsilon}^{\varepsilon} \delta(x) - y(x)\,\mathrm{d}x \implies y(0^+) - y(0^-) = 1$$

in the limit $\varepsilon \to 0$. So, we have $A = B + 1$

As we are interested in square integrable solutions, we must have $B = 0$. Therefore, $A = 1$ and the solution is

$$y(x) = G(x) = \begin{cases} \mathrm{e}^{-x} & x > 0, \\ 0 & x < 0. \end{cases}$$

Now, let's solve by (2.6). We have $P(ik) = ik + 1$. There is a single, first order pole at $k = i$. The solution is

$$G(x) = \begin{cases} i \sum_{\mathrm{Im}>0} \mathrm{Res}(\mathrm{e}^{ikx}/(ik+1), k_j) & x > 0, \\ -i \sum_{\mathrm{Im}<0} \mathrm{Res}(\mathrm{e}^{ikx}/(ik+1), k_j) & x < 0. \end{cases}$$

Since there are no poles for $\mathrm{Im} < 0$, we have $G(x)$ for $x < 0$. The residue at $k = i$ gives

$$G(x) = \begin{cases} \mathrm{e}^{-x} & x > 0, \\ 0 & x < 0. \end{cases}$$

So, the two results agree, as expected.

2.2.4 Simple zeros of characteristic polynomial

What happens if $P(ik) = 0$ for some $k \in \mathbb{R}$? In this case, as we will show, there will no longer be a unique, square integrable solution but we can still make sense out of it.

We will only consider simple zeros of P. In this case, we know that the Fourier integral for $y(x)$ will not converge in the *usual sense.* This is why define the *Cauchy principal value.*

Definition (Cauchy principal value)**.** For a function $f(x)$ with a single singularity at an *internal point* s of an interval $[a, b]$, the principal value is defined by the limit

$$\mathcal{P}\int_a^b f(x)\,\mathrm{d}x := \lim_{\varepsilon \to 0^+} \left[\int_a^{s-\varepsilon} f(x)\,\mathrm{d}x + \int_{s+\varepsilon}^b f(x)\,\mathrm{d}x \right].$$

Corollary. Let $h(x)$ we well-defined for all x. Then, for all even functions $g(x)$, we have

$$\mathcal{P}\int_{-\infty}^{\infty} \left(h(x) + \frac{g(x)}{x} \right) \mathrm{d}x = \int_{-\infty}^{\infty} h(x)\,\mathrm{d}x\,.$$

So, if we have an even function with an odd singularity, we just ignore it.

Let's proceed as we would if $P(ik)$ was non-singular and see when we arrive at inconsistencies *(hint - very soon)*. Taking the Fourier transform, we get

$$P(ik)\hat{y}(k) = \hat{f}(k).$$

Let $P(ik_j) = 0$ for some real $\{k_j\}$. Then, we have

$$P(ik_j)\hat{y}(k_j) = 0 \neq \hat{f}(k_j)$$

as there is no particular reason $\hat{f}$ would be zero. The function f is completely independent of our characteristic polynomial. So, we have arrived at an inconsistency. In order to resolve this, we must go to the homogeneous equation:

$$a_n y^{(n)} + \ldots + a_0 y = 0.$$

We know, since $P(ik_j) = 0$, that a solution of the homogeneous equation is

$$y(x) = c_j \mathrm{e}^{ik_j x}.$$

Hence, our differential equation becomes:

$$\sum_n a_n \frac{\mathrm{d}^n}{\mathrm{d}x^n}\left(y(x) + \sum_j c_j \mathrm{e}^{ik_j x}\right) = f(x).$$

Taking the Fourier transform yields

$$P(ik)\hat{y}(k) + \sum_j P(ik_j)c_j\delta(k-k_j) = P(ik)\hat{y}(k) + \sum_j P(ik)c_j\delta(k-k_j) = \hat{f}(k),$$

where for the second term we have $P(ik) = P(ik_j)$ because of the delta functions. A factor of 2π is implicitly included in c_j. So we have

$$\hat{y}(k) = \frac{\hat{f}(k)}{P(ik)} + \sum_j c_j\delta(k-k_j). \tag{2.8}$$

The constants c_j are arbitrary, so we don't need to worry about the minus sign. Now, we see that given m singularities, we have to choose m constants. We lose uniqueness, and our solutions are no longer square integrable. When we take the Fourier integral, we simply use the principal values. *(I genuinely don't understand why any of this works at all. Everything should diverge... but doesn't.)*

3 Calculus of Variations

Often in physics, we are interested in quantities that depend on the behaviours of *functions* over some interval. These quantities are called *functionals.*

3.1 Functionals

Definition (Functional)**.** A *functional* is a function that assigns a real number to a function $y : \mathbb{R} \to \mathbb{R}$. We denote functionals by square brackets: $F[y] \in \mathbb{R}$.

Note (Assigning numbers to functions.)**.** Generally, assigning a real number that depends on a function is done by integration at one point or another. *(Apart from evaluating the functions at certain points, nothing else comes to mind.)*

Example. Here are some examples of functionals:

- Length of a curve:
$$F\left[y\right] = \int_{x_0}^{x_1} \sqrt{1 + y'^2}\,\mathrm{d}x\,.$$
- The time it takes to traverse from some point to another, given a velocity field $v(x, y)$. *(This is an example of a functional that depends on multiple functions.)*
$$T\left[x, y\right] = \int_{s_0}^{s_1} \frac{\sqrt{x'^2 + y'^2}}{v(x, y)}\,\mathrm{d}x\,.$$
- As a more general example, let $F(x_1, x_2, x_3)$ be a continuous function of three variables. Then the expression
$$J\left[y\right] = \int_a^b F(y, y', x)\,\mathrm{d}x\,, \tag{3.1}$$
where $y(x)$ ranges over the set of all continuously differentiable functions in the interval $x \in [a, b]$ is a functional.

We will mainly consider functionals of the form (3.1). These functionals have a *localization property*, meaning that if we divide up y into sections and calculate $J[y]$ for each section, the sum of the values equals the functional for the entire curve.

Definition (Localization property)**.** A functional $J\left[y\right]$ is said to be *local* if, for any y in the domain of J we have
$$J\left[y\right] = \sum_i J_i\left[y_i\right]\,,$$
where each y_i denotes a section of the curve y. This can be generalized to functionals of multiple functions.

We are interested in *"stationary functions"* of functionals of the form (3.1). This seems like a very difficult problem - and it is. Instead of points at which the gradient of a function is zero, we are interested in functions for which the gradient of a functional is zero. One way of approaching this problem initially is the *method of finite differences.*

3.1.1 Method of finite differences

One of the most natural ways to approach the problem of finding stationary functions of functionals is to reduce it to a problem of classical analysis - i.e. finding stationary points.

Consider a functional of the form (3.1), with the condition $y(a) = y_a$ and $y(b) = y_b$. We start by dividing up the interval $x \in [a, b]$ into $n + 1$ sections using the points

$$x_0 = a, \quad x_1, x_2, \ldots, x_n, \quad x_{n+1} = b.$$

Now we replace (approximate) the function $y(x)$ by a set of lines with endpoints

$$(x_0, y_a), (x_1, y(x_1)), \ldots (x_{n+1}, y_b).$$

Hence, we approximate the functional $J[y]$ by the sum

$$J[y] \approx J(y_1, y_2, \ldots, y_n) = \sum_{i=0}^{n+1} F(y_i, \frac{y_i - y_{i-1}}{h}, x_i)h$$

where $y_i = y(x_i)$ and $h = x_i - x_{i-1}$.

Notice now that J is no longer a functional, but a function of n variables. To find stationary functions of $J[y]$, we find n stationary points of $J(y_1, \ldots, y_n)$. In the limit $n \to \infty$, we obtain the function $y(x)$ for which $J[y]$ is stationary.

Although this is an intuitive approach, it is not practical. We would rather not find an infinite number of stationary points. *I don't even know how any problem can be solved by this method, but apparently Euler made it work.*

3.1.2 Function spaces

Just as we would use n-dimensional Euclidian space to represent the arguments of functions of n variables, it is useful to consider a similar structure when dealing with functionals.

Definition (Function space)**.** A *function space* is a space whose elements are functions.

Note. Sometimes, the definition of a function space is given as a vector space, which implies the definition of a dot product between two functions. We can be more general.

Notice that the definition is very broad, and for most problems we need to specify the classes of functions we are interested in. For example, if we are working with a functional of the form

$$J[y] = \int_a^b F(y, y', x)\,\mathrm{d}x\,,$$

we would be interested all continuous functions f with continuous first derivatives on the interval $[a, b]$.

For our purposes, we will be interested in *normed linear spaces.* In a normed space, we can think about the distance between two elements x, y by considering $|x - y|$. We will consider the following:

- The space $\mathscr{C}(a, b)$, consisting of all continuous functions $y(x)$ in the interval $[a, b]$. Addition of functions and multiplication by real numbers are defined as usual, and the norm is defined as
$$|y| = \max_{x \in [a,b]} |y(x)|.$$
Two functions are regarded as close if one function lies within a strip of a small width around the other.

- The space $\mathscr{D}_n(a,b)$, consisting of all continuous functions $y(x)$ defined on the interval $[a,b]$ with continuous derivatives up to and including order n for some $n \in \mathbb{Z}$. Addition and multiplication by real numbers are defined as usual. The norm is defined as

$$|y| = \sum_{i=0}^{n} \max_{x\in[a,b]} |y^{(i)}(x)|.$$

Therefore, two functions in $\mathscr{D}_n$ are regarded as close together if the values of the functions themselves and of all their derivatives up to order n inclusive are close together.

Now, we are in a position to talk about continuity of functionals.

Definition (Functional continuity at a point)**.** The functional $J\,[y]$, defined over a normed linear space $\mathscr{L}$, is said to be *continuous at point* $y_0 \in \mathscr{L}$ if for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|J\,[y] - J\,[y_0]\,| < \varepsilon,$$

for all $y \in \mathscr{L}$ where $|y - y_0| < \delta$.

Note. In the definition of norm in $\mathscr{D}_n$, had we not included contributions from the derivatives, we could not have any continuous functionals that depend on derivatives of functions. This is obvious.

3.1.3 Variation of functionals and the functional derivative

The *variation* of a functional is analogous to the differential of a function. We will use variations to find stationary points. First, we give some definitions.

Definition (Continuous linear functional)**.** Given a normed linear space $\mathscr{L}$ and a functional $\varphi\,[h]$ defined over $\mathscr{L}$, $\varphi\,[h]$ is said to be a continuous linear functional if

1. $\varphi\,[\alpha h] = \alpha\varphi\,[h]$ for all $h \in \mathscr{L}$ and any $\alpha \in \mathbb{R}$.
2. $\varphi[h_1 + h_2] = \varphi[h_1] + \varphi[h_2]$ for any $h_1, h_2 \in \mathscr{L}$.
3. $\varphi[h]$ is continuous for all $h \in \mathscr{L}$.

Example. The integral

$$\varphi[h] = \int_a^b \left\{\alpha_0(x)h(x) + \ldots + a_n(x)h^{(n)}(x)\right\} \mathrm{d}x$$

where $\alpha_i(x) \in \mathscr{C}(a,b)$, defines a linear continuous functional in $\mathscr{D}_n(a,b)$.

Definition (First variation)**.** Given a functional J defined over $\mathscr{D}_n$, let $y, h \in \mathscr{D}_n$. If, for a fixed y, there exists a *continuous linear functional,* $\delta J[h]$ such that

$$\Delta J[h] \equiv J[y+h] - J[y] = \delta J[h] + o(\max|h|), \tag{3.2}$$

then δJ is the first variation of J at y. Considering an increment of the form ϵh, we can Taylor expand J about $\epsilon = 0$ and obtain

$$J[y+\epsilon h] = J[y] + \epsilon \left.\frac{\mathrm{d}J[y+\epsilon h]}{\mathrm{d}\epsilon}\right|_{\epsilon=0} + o(\epsilon),$$

from which it becomes clear that the first variation has the form

$$\delta J[h] = \left.\frac{\mathrm{d}J[y+\epsilon h]}{\mathrm{d}\epsilon}\right|_{\epsilon=0}. \tag{3.3}$$

Now, we are in a position to talk about the derivative of a functional at a given point x_0. There are different ways of approaching this definition intuitively, we will go over two of those.

Definition (Functional derivative 1)**.** Let J be a functional defined over $\mathscr{D}_n$ and $y \in \mathscr{D}_n$. Suppose we increment $y(x)$ by a function $h(x)$ defined as

$$h(x) = \begin{cases} h_0 & x_0 - \epsilon < x < x_0 + \epsilon, \\ 0 & \text{otherwise.} \end{cases}$$

We define $\Delta\sigma$ as the area between $y + h$ and h, *(or, equivalently, the area between h and the x axis)*. If the limit

$$\left.\frac{\delta J}{\delta y}\right|_{x_0} := \lim_{\epsilon \to 0} \frac{J[y + \epsilon h] - J[y]}{\Delta\sigma(\epsilon)} \tag{3.4}$$

exists, it is called the *functional derivative* of $J[y]$ at point $x = x_0$.

In the limit $\epsilon \to 0$, $\Delta\sigma \to h(x_0)\,\mathrm{d}x$ and rearranging gives

$$\left.\frac{\delta J}{\delta y}\right|_{x_0} h(x_0)\,\mathrm{d}x = \lim_{\epsilon \to 0} \frac{J[y + \epsilon h] - J[y]}{\epsilon} = \delta J[h].$$

This leads us to our second definition.

Definition (Functional derivative 2)**.** The functional derivative of $J[y(x)]$ is a function of x such that the variation $\delta J[h]$ is given by

$$\delta J[h] = \int \frac{\delta J}{\delta y}(x) h(x)\,\mathrm{d}x\,. \tag{3.5}$$

We can think of this as a generalization of (3.4) for any increment function $h(x)$.

3.2 Euler-Lagrange Equation

We consider functionals of the form (3.1). We want to find stationary points, so let's calculate the first variation.

We first define some boundary conditions, for all $y \in \mathscr{D}_1$ we require $y(a) = y_a$ and $y(b) = y_b$ for some $y_a, y_b \in \mathbb{R}$. This implies that our increment $h(x)$ satisfy $h(a) = h(b) = 0$. We require that for all $h \in \mathscr{D}_1$,

$$\begin{aligned} \delta J[h] &= \int_a^b \left.\frac{\mathrm{d}F(y + \epsilon h, y' + \epsilon h', x)}{\mathrm{d}\epsilon}\right|_{\epsilon=0} \mathrm{d}x \\ &= \int_a^b \partial_y F(y, y', x) h + \partial_{y'} F(y, y', x) h'\,\mathrm{d}x \equiv 0. \end{aligned}$$

Now, we integrate the second term by parts to obtain

$$\int_a^b \partial_{y'} F(y, y', x) h'\,\mathrm{d}x = \left[h \partial_{y'} F\right]_a^b - \int_a^b \frac{\mathrm{d}}{\mathrm{d}x}\left(\partial_{y'} F\right) h\,\mathrm{d}x\,.$$

Due to our boundary conditions, the first terms is identically zero. Hence, we obtain

$$\delta J[h] = \int_a^b h \left\{ \partial_y F(y, y', x) - \frac{\mathrm{d}}{\mathrm{d}x}\left(\partial_{y'} F\right) \right\} \mathrm{d}x \equiv 0, \tag{3.6}$$

$$\implies \frac{\delta J}{\delta y}(x) = \frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) \equiv 0. \tag{3.7}$$

This is the *Euler-Lagrange equation.* Notice that (3.6) is the statement that the first variation is identically zero, which implies that the functional derivative given by (3.7) is identically zero.

This allows us to easily generalize to functionals of n functions, in which case the Euler-Lagrange equation becomes

$$\frac{\partial F}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y_i'}\right) = 0 \quad \forall\, i.$$

3.2.1 First integrals

In general, the Euler-Lagrange equation gives us a second order differential equation to solve. However, under some special cases the equation reduces to a first order one, which simplifies the problem considerably. We look at those special cases.

1. If the integrand does not depend on y, so

   $$J[y] = \int_a^b F(y', x)\,\mathrm{d}x\,,$$

   the Euler-Lagrange equation becomes

   $$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) = 0$$

   which gives us the *first integral*

   $$\frac{\partial F}{\partial y'} = c$$

   for some $c \in \mathbb{R}$.

2. If the integrand doesn't depend on x,

   $$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) = y''\partial^2_{y'y'}F + y'\partial^2_{yy'}F$$

   The Euler-Lagrange equation reduces to

   $$\frac{\mathrm{d}}{\mathrm{d}x}\left(F - y'\partial_{y'}F\right) = 0$$

   which gives us the first integral

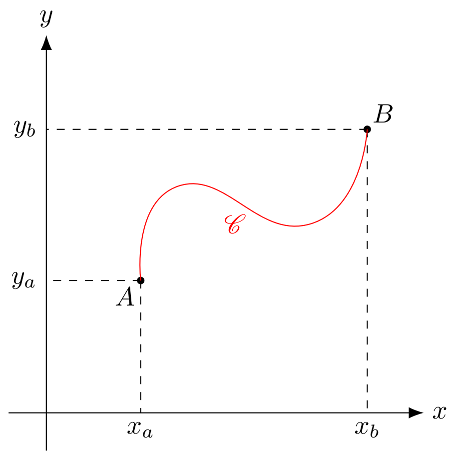
   $$F - y\partial_{y'}F = c.$$

3. If F does not depend on y', we simply solve for y:

   $$\partial_y F = 0.$$

In physics, such simplifications correspond to *conserved quantities.* In the Lagrangian formalism, the independent variable is time. So, if the time derivative of any expression is identically zero, as is the case with cases 1 and 2, we have a conserved quantity.

Example (Geodesics of a plane)**.** What is the shortest curve $\mathscr{C}$ joining two points A and B on a plane?



We parameterise the curve so that $\mathbf{r}(t) = (x(t), y(t))$ for $t \in [0, 1]$. The boundary conditions are $\mathbf{r}(0) = (x_a, y_a)$ and $\mathbf{r}(1) = (x_b, y_b)$. We want to minimize

$$\int_0^1 \mathrm{d}t \sqrt{\dot{x}^2 + \dot{y}^2}, \quad \text{so } F = \sqrt{\dot{x}^2 + \dot{y}^2}.$$

By the Euler-Lagrange equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial F}{\partial \dot{x}}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{x}}{\sqrt{\dot{x} + \dot{y}^2}}\right) = 0.$$

We have a first integral. Integrating yields

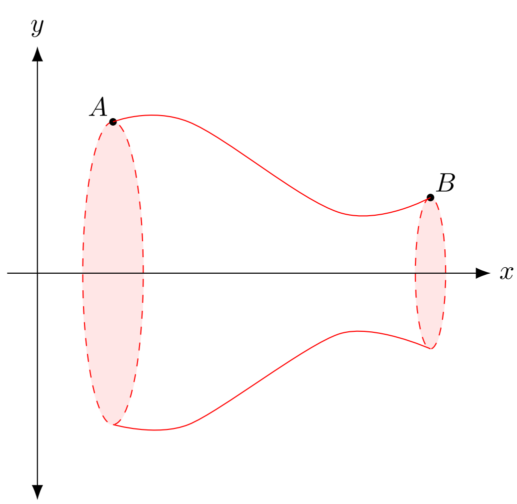
$$\frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}} = c_1, \quad \frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}} = c_2.$$

This implies

$$\dot{y} = \frac{c_2}{c_1}\dot{x},$$

which is the equation for a straight line. We set the constants such that the boundary conditions are matched.

Example (Minimum surface of revolution)**.** Among all curves joining two points $A = (x_a, y_a)$ and $B = (x_b, y_b)$, find the one which minimizes the surface area when rotated about the x-axis.



We would like to minimize

$$S = 2\pi \int_{x_a}^{x_b} y\sqrt{1+y'^2}\,\mathrm{d}x\,.$$

Notice that our integrand is independent of x, so we have the first integral:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(F - y'\partial_{y'}F\right) = 0$$

$$\implies y\sqrt{1+y'^2} - \frac{yy'^2}{\sqrt{1+y'^2}} = c_1$$

for some $c_1 \in \mathbb{R}$. Solving for y' yields

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \sqrt{\frac{y^2-c_1^2}{c_1^2}}.$$

Now, we separate variables and integrate:

$$x + c_2 = c_1 \int \frac{\mathrm{d}y}{\sqrt{y^2-c_1^2}}$$

$$\implies y = c_1 \cosh\left(\frac{x+c_2}{c_1}\right).$$

This describes *catenary.* The surface of revolution generated by rotating a catenary about the $x-$axis is a *catenoid.* The coefficients are found by solving for the boundary conditions.

3.2.2 Coordinate invariance

Suppose we use a different coordinate system such that we have $x = x(u,v)$ and $y = y(u,v)$ with the condition

$$\begin{vmatrix} \partial_u x & \partial_v x \\ \partial_u y & \partial_v y \end{vmatrix} \neq 0.$$

This requirement is simply the statement that the span of our new coordinates is not less than the span of the previous.

We now have the curve $y(x)$ correspond to some $v(u)$. Transforming our functional, we have

$$\begin{aligned} J[y] &= \int_a^b F(y,y',x)\,\mathrm{d}x \\ &= \int_{a^*}^{b^*} F\left(y(u,v), \frac{\partial_u y + v'\partial_v y}{\partial_u x + v'\partial_v x}, x(u,v)\right)(\partial_u x + v'\partial_v x)\,\mathrm{d}u \\ &\equiv \int_{a^*}^{b^*} F^*(v,v',u)\,\mathrm{d}u \;= J^*[v]. \end{aligned}$$

where we have defined

$$F^*(v,v',u) \equiv F\left(y(u,v), \frac{\partial_u y + v'\partial_v y}{\partial_u x + v'\partial_v x}, x(u,v)\right)(\partial_u x + v'\partial_v x).$$

We would like to show that

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) = \frac{\partial F}{\partial y} \iff \frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{\partial F^*}{\partial v'}\right) = \frac{\partial F^*}{\partial v}. \tag{3.8}$$

Consider the functional derivatives for some $y(x), h(x)$ and a corresponding $v(u), \eta(u)$:

$$\frac{\delta J}{\delta y} = \lim_{\epsilon \to 0} \frac{J[y + \epsilon h] - J[y]}{\Delta\sigma(\epsilon)} = \lim_{\epsilon \to 0} \frac{J^*[v + \epsilon\eta] - J^*[v]}{\Delta\sigma^*(\epsilon)} \left(\begin{vmatrix} \partial_u x & \partial_v x \\ \partial_u y & \partial_v y \end{vmatrix} \right)^{-1}$$

$$\implies \frac{\delta J}{\delta y} = \frac{\delta J^*}{\delta v} \left(\begin{vmatrix} \partial_u x & \partial_v x \\ \partial_u y & \partial_v y \end{vmatrix} \right)^{-1}$$

Since we required the Jacobian to be non-zero, (3.8) follows.

Whether or not a curve is stationary with respect to a functional of the form (3.1) is independent of the coordinate system.

3.3 Constrained Variations

What happens if we require the function y to satisfy additional conditions? We will be interested in functionals of the form (3.1), with an additional constraint by fixing the value of another functional of the form

$$K[y] = \int_a^b G(y, y', x)\,\mathrm{d}x = l.$$

for some function G and $l \in \mathbb{R}$. First, let's try to get an intuitive picture of the situation by considering functions defined over $\mathbb{R}^n$.

3.3.1 Lagrange multipliers

Consider a scalar field $f : \mathbb{R}^n \to \mathbb{R}$. If we wanted to find a stationary point $\mathbf{x}_0 \in \mathbb{R}^n$, we would solve

$$\mathrm{d}f = \nabla f \cdot \mathrm{d}\mathbf{x} \equiv 0$$
$$\implies \nabla f = \mathbf{0},$$

as the differential has to be zero for all displacements $\mathrm{d}\mathbf{x}$. This would be *unconstrained*, as we considered all values of $\mathbf{x}$. Suppose now we add a constraint such that we only consider $\mathbf{x} \in \mathbb{R}^n$ which satisfy $g(\mathbf{x}) = 0$ for some $g : \mathbb{R}^n \to \mathbb{R}$. Notice that this removes one degree of freedom, and would correspond to a surface in $\mathbb{R}^3$ or a curve in $\mathbb{R}^2$.

Now, our displacement $\mathrm{d}\mathbf{x}$ must lie on the "surface" defined by $g(\mathbf{x})$. As we still want $\mathrm{d}f$ to be identically zero, we require ∇f to be perpendicular to the surface. Since we know $\nabla g(\mathbf{x})$ is perpendicular to $g(\mathbf{x}) = 0$, we require

$$\nabla f = \lambda \nabla g, \quad g(\mathbf{x}) = 0.$$

This can be reduced to an unconstrained problem by considering a function of the form

$$\phi(\mathbf{x}, \lambda) \equiv f(\mathbf{x}) - \lambda g(\mathbf{x}).$$

Minimizing with respect to $\mathbf{x}$ gives the first condition, and λ gives the second.

Notice how this is almost identical to the problem we considered with functionals - we have the same type of constraint!

3.3.2 Generalization to functionals

Theorem. Given the functional

$$J[y] = \int_a^b F(y, y', x)\,\mathrm{d}x\,,$$

with constraints

$$y(a) = y_a, \quad y(b) = y_b, \quad K[y] = \int_a^b G(y, y', x)\,\mathrm{d}x \; = l \in \mathbb{R},$$

where $K[y]$ is another functional, let $J[y]$ have a stationary point for $y = y(x)$ subject to the constraints. Then, if $y(x)$ is not a stationary point of $K[y]$, there exists $\lambda \in \mathbb{R}$ such that $y(x)$ is a stationary point of the functional

$$\int_a^b (F + \lambda G)\,\mathrm{d}x\,.$$

Proof. Let $y = y(x)$ be a stationary point of J subject to the constraints. We choose two points $x_1, x_2 \in [a, b]$, where x_2 will remain fixed and x_1 will be arbitrary. Suppose we increment $y(x)$ by $h_1(x) + h_2(x)$ where h_1 is nonzero only in the neighborhood of x_1, and h_2 is nonzero only in the neighborhood of x_2. The first variation of $J[y]$ is

$$\delta J[h_1 + h_2] = \left.\frac{\delta J}{\delta y}\right|_{x_1} \sigma_1 + \left.\frac{\delta J}{\delta y}\right|_{x_2} \sigma_2, \tag{3.9}$$

where

$$\sigma_1 = \int_a^b h_1(x)\,\mathrm{d}x\,, \quad \sigma_2 = \int_a^b h_2(x)\,\mathrm{d}x\,.$$

We require that the varied curve $y^* = y + h_1 + h_2$ satisfy $K[y^*] = K[y]$. The first variation of K is

$$\delta K[h_1, h_2] = \left.\frac{\delta K}{\delta y}\right|_{x_1} \sigma_1 + \left.\frac{\delta K}{\delta y}\right|_{x_2} = 0. \tag{3.10}$$

Now, we choose x_2 to be a point such that

$$\left.\frac{\delta K}{\delta y}\right|_{x_2} \neq 0,$$

which exists since we have assumed $y(x)$ is not a stationary point of K. Now, we can write the condition (3.10) as

$$\sigma_2 = -\sigma_1 \left\{ \frac{\left.\frac{\delta K}{\delta y}\right|_{x_1}}{\left.\frac{\delta K}{\delta y}\right|_{x_2}} \right\}.$$

Now, we set

$$\lambda = -\frac{\left.\frac{\delta J}{\delta y}\right|_{x_2}}{\left.\frac{\delta K}{\delta y}\right|_{x_2}}.$$

Substituting into (3.9) yields

$$\delta J[h_1] = \left\{ \left.\frac{\delta J}{\delta y}\right|_{x_1} + \lambda \left.\frac{\delta K}{\delta y}\right|_{x_1} \right\} \sigma_1,$$

where now our first variation is only a functional of h_1. The constraint has removed one degree of freedom. As we require the first variation to be identically zero *(i.e. zero for all h_1)*, we have

$$\frac{\delta J}{\delta y} + \lambda \frac{\delta K}{\delta y} = \frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) + \lambda \left\{ \frac{\partial G}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial G}{\partial y'}\right) \right\} = 0.$$

This completes the proof. □

Example (Isoperimetric problem)**.** What closed curve of fixed length l encloses the maximum area?

We parameterize the curve by $\mathbf{r}(t) = (x(t), y(t))$, for $t \in [0,1]$. So,

$$A[y] = \int_0^1 y\dot{x}\,\mathrm{d}t\,, \quad L[y] = \int_0^1 \sqrt{\dot{x}^2 + \dot{y}^2}\,\mathrm{d}t\; = l.$$

Hence, we maximize the functional

$$A[y] + \lambda L[y] = \int_0^1 y\dot{x} + \lambda\sqrt{\dot{x}^2 + \dot{y}^2}\,\mathrm{d}t\,.$$

We obtain equations for maximization with respect to x and y:

$$\dot{x} = \lambda \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}}\right), \qquad \frac{\mathrm{d}}{\mathrm{d}t}\left(y + \frac{\lambda\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}\right) = 0.$$

Integrating with respect to t yields

$$x + c_1 = \frac{\lambda\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}}, \qquad -y - c_2 = \frac{\lambda\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}$$

for some constants c_1, c_2. This implies

$$\begin{aligned} &\dot{x}(x + c_1) + \dot{y}(y + c_2) = 0, \\ \implies &\frac{(x + c_1)^2}{2} + \frac{(y + c_2)^2}{2} = c_3. \end{aligned}$$

This is a circle of radius $\sqrt{c_3}$, centered at $(-c_1, -c_2)$. The constants can be found by the boundary conditions, as we have not stated any restrictions on the position of the curve c_1 and c_2 can be arbitrary.

3.4 Lagrangian Mechanics

We will look at how calculus of variations can be used to reformulate classical mechanics.

3.4.1 Generalized Coordinates

We start by defining a *configuration space*. This is a vector space containing *generalized coordinates* $\{q_1, q_2, \ldots, q_n\}$ that specify the configuration of the entire system. A vector on the configuration space contains all the information about a given system. In classical mechanics, this information is the positions of the particles.

We denote the configuration space by C. As a system evolves in time, it traces out a *curve* in C, and we would like to know which particular curve is traversed.

3.4.2 The Principle of Stationary Action

Definition (Lagrangian)**.** The Lagrangian is defined to be a function of the positions q, and velocities $\dot{q}$ of all particles in the system, given by

$$L(q, \dot{q}, t) = T(\dot{q}) - V(q, t) \tag{3.11}$$

where $T = \frac{1}{2}\sum_i m_i(\dot{q}_i)^2$ is the kinetic energy and V is the potential energy (which may be time-dependent). [1]

[1] We denote the set of generalized coordinates $q_1, q_2, \ldots, q_n$ by q, and similarly we denote the generalized velocities by $\dot{q}$.

We consider *all curves in* C with fixed endpoints

$$q(t_i) = q_{\text{initial}}, \quad q(t_f) = q_{\text{final}}.$$

Of all the possible paths, only one path is taken by the system. We assign a value to each path by defining the *action functional.*

Definition (Action)**.** The *action* is a functional of the generalized coordinates, given by

$$S[q] = \int_{t_i}^{t_f} L(q, \dot{q}, t)\,\mathrm{d}t\,. \tag{3.12}$$

We can now state the principle of stationary action:

Law (Principle of stationary action)**.** The path traversed by the system in the configuration space C is a *stationary point* of the action S.

It is easy to show that we can recover Newton's law from (3.12). The principle of stationary action implies that our Lagrangian is a solution of the Euler-Lagrange equation. Consider a system of one particle, and let the generalized coordinates be Cartesian coordinates $\{x, y, z\}$. The Lagrangian is

$$L(q, \dot{q}, t) = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - V(x, y, z, t).$$

The Euler Lagrange equation gives

$$m(\ddot{x} + \ddot{y} + \ddot{z}) = -\nabla V.$$

The right hand side is the force exerted on the particle, and so we recover Newton's law.

Note. This can be done in more generality, since assuming Cartesian coordinates is an arbitrary decision. We can, instead, define the *generalized momentum.*

Definition (Generalized momentum)**.** The generalized momentum corresponding to a set of generalized coordinates $q = q_1, \dots, q_n$ is defined as the partial derivative of the Lagrangian with respect to the generalized velocity $\dot{q}$:

$$p_k := \frac{\partial L}{\partial \dot{q}_k}. \tag{3.13}$$

With this definition, we can recover Newton's law for *any coordinate system.* This is not obvious due to fictitious forces.

3.4.3 Change of coordinates

We have already proved that the Euler-Lagrange equation is coordinate invariant in section 3.2.2. Here, we present another proof keeping the concept of generalized coordinates in mind.

Consider a set of generalized coordinates

$$x_i : \quad (x_1, x_2, \dots, x_N),$$

and assume the Euler Lagrange equations hold in this set of coordinates, such that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}_i}\right) = \frac{\partial L}{\partial x_i} \quad \forall\ i. \tag{3.14}$$

Now, consider a new set of coordinates q_i, related to x_i by a transformation of the form:

$$q_i = q_i(x_1, x_2, \dots, x_N; t). \tag{3.15}$$

We are interested in "nice" transformations that we can invert, so we have

$$x_i = x_i(q_1, q_2, \dots, q_N; t). \tag{3.16}$$

Proposition (Coordinate invariance)**.** If equation (3.14) holds for the set of coordinates x_i, and if x_i and q_i are related by (3.16), then the Euler Lagrange equation also holds for q_i. That is,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i} \quad \forall\ i.$$

Proof. The Lagrangian in the transformed coordinates q is

$$L'(q, \dot{q}, t) = L(x(q,t), \dot{x}(q, \dot{q}, t), t),$$

where the prime indicates that L' is a different function than L. We would like to find an expression for the functional derivative in the q coordinates. First look at:

$$\frac{\partial L'}{\partial q^i} = \frac{\partial L}{\partial x^j}\frac{\partial x^j}{\partial q^i} + \frac{\partial L}{\partial \dot{x}^j}\frac{\partial \dot{x}^j}{\partial q^i}.$$

From (3.16), it follows that

$$\dot{x}^j = \frac{\mathrm{d}\dot{x}^j}{\mathrm{d}t} = \frac{\partial x^j}{\partial q^k}\dot{q}^k + \frac{\partial x^j}{\partial t} \implies \frac{\partial \dot{x}^j}{\partial q^i} = \frac{\partial^2 x^j}{\partial q^i \partial q^k}\dot{q}^k + \frac{\partial^2 x^j}{\partial q^i \partial t}.$$

So, we have

$$\frac{\partial L'}{\partial q^i} = \frac{\partial L}{\partial x^j}\frac{\partial x^j}{\partial q^i} + \frac{\partial L}{\partial \dot{x}^j}\left(\frac{\partial^2 x^j}{\partial q^i \partial q^k}\dot{q}^k + \frac{\partial^2 x^j}{\partial q^i \partial t}\right).$$

Now, look at:

$$\frac{\partial L'}{\partial \dot{q}^i} = \frac{\partial L}{\partial \dot{x}^j}\frac{\partial \dot{x}^j}{\partial \dot{q}^i} = \frac{\partial L}{\partial \dot{x}^j}\frac{\partial}{\partial \dot{q}^i}\left(\frac{\partial x^j}{\partial q^k}\dot{q}^k + \frac{\partial x^j}{\partial t}\right) = \frac{\partial L}{\partial \dot{x}^j}\frac{\partial x^j}{\partial q^i}.$$

Then, it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial L'}{\partial \dot{q}^i}\right] = \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial L}{\partial \dot{x}^j}\right]\frac{\partial x^j}{\partial q^i} + \frac{\partial L}{\partial \dot{x}^j}\left[\frac{\partial^2 x^j}{\partial q^i \partial q^k}\dot{q}^k + \frac{\partial^2 x^j}{\partial q^i \partial t}\right]$$

Putting the two terms together, we obtain a transformation rule for the functional derivative:

$$\frac{\delta L'}{\delta q^i} = \frac{\partial L'}{\partial q^i} - \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial L'}{\partial \dot{q}^i}\right] = \frac{\partial x^j}{\partial q^i}\left(\frac{\partial L}{\partial x^j} - \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial L}{\partial \dot{x}^j}\right]\right) = \frac{\partial x^j}{\partial q^i}\frac{\delta L}{\delta x^j}. \tag{3.17}$$

Firstly, we see that if the functional derivative is zero in x, it follows that it must be zero in q (since we assumed the Jacobian associated with $x \to q$ has non-zero determinant) so our proof is complete. But more importantly, we see that the functional derivative somehow transforms as a covariant tensor! Note that none of the two terms in the functional derivative transforms as a tensor, but their combination does.

4 Tensors

work in progress

5 Numerical Methods

5.1 Numerical Integration

We would like to numerically solve an integral of the form

$$\int_a^b f(x)\,\mathrm{d}x\,,$$

for some real function $f(x)$ defined over the domain $x \in [a, b]$. We are interested in obtaining the best accuracy while doing a minimum number of calculations.

We will first look at two specific methods: trapezium and Simpson's rule.

5.1.1 Trapezium rule

We approximate $f(x)$ by picking a set of points called *nodes*, and joining consecutive nodes by line segments. This is shown on Figure 5.1.

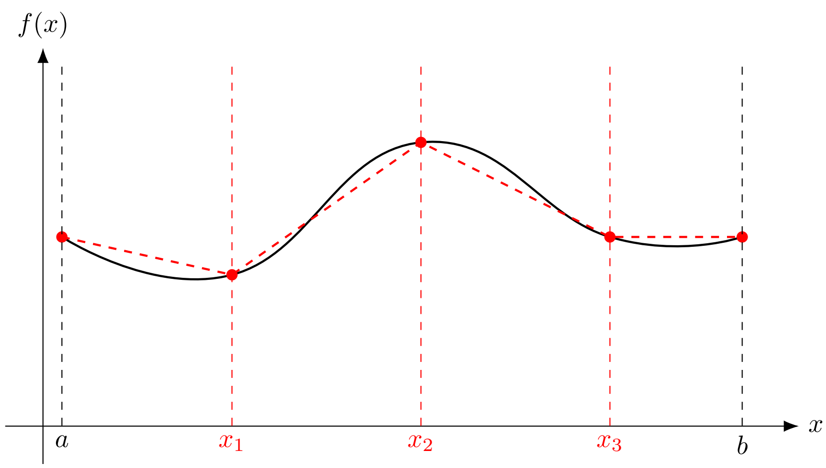


Figure 5.1

There are 5 nodes in total, with $x_0 = a$ and $x_4 = b$. The integral is then approximated by the area bounded by the trapeziums. The general expression for a set of n nodes $\{x_k\}$ is

$$\int_a^b f(x)\,\mathrm{d}x = \sum_{k=0}^{n-1} \int_{x_k}^{x_{k+1}} f(x)\,\mathrm{d}x \approx \sum_{k=0}^{n-1} \frac{f(x_k) + f(x_{k+1})}{2}(x_{k+1} - x_k).$$

We can simplify this by choosing *equidistant nodes*, so that we only need to calculate $(x_{k+1} - x_k)$ once. Let

$$h \equiv \frac{b-a}{n} \implies x_k = a + hk.$$

Denoting $f(x_k) \equiv f_k$, we have

$$\int_a^b f(x)\,\mathrm{d}x \approx \frac{h}{2}\sum_{k=0}^{n-1} f_k + f_{k+1} = h\left(\frac{f_0}{2} + f_1 + f_2 + \cdots + \frac{f_n}{2}\right). \tag{5.1}$$

This is the trapezium rule.

But we are not done yet. We need a method to estimate errors. In general, there is no way to do so - we need to quantify how different $f(x)$ is from a line on a given interval. However, we may make progress if the second derivative of f is bounded on the interval $[a, b]$.

Suppose that f has a <u>bounded</u> second derivative so that

$$|f''(x)| \leq M \quad \forall\, x \in [a,b],$$

for some $M \geq 0$. Now, we Taylor expand $f(x)$ around some $x^\star$.

$$f(x) = f(x^\star) + f'(x^\star)(x - x^\star) + \cdots + \frac{f^{(j)}(x^\star)}{j!}(x - x^\star)^j + \frac{f^{(j+1)}(\xi)}{(j+1)!}(x - x^\star)^{j+1},$$

where in the last term, $\xi \in [x^\star, x]$. The last term is the Lagrange remainder (see Appendix A.2). Now, we let $j = 1$ and $x^\star = x_k$ so we have

$$f(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{f''(\xi)}{2}(x - x_k)^2,$$

rearranging this expression and integrating yields

$$\begin{aligned} &\int_{x_k}^{x_{k+1}} f(x)\,\mathrm{d}x\ - f_k h - f'(x_k)\frac{h^2}{2} = \int_{x_k}^{x_{k+1}} \frac{f''(\xi)}{2}(x - x_k)^2\,\mathrm{d}x \\ \implies &\left| \int_{x_k}^{x_{k+1}} f(x)\,\mathrm{d}x\ - f_k h - f'(x_k)\frac{h^2}{2} \right| \leq \frac{M}{2} \int_{x_k}^{x_{k+1}} (x - x_k)^2\,\mathrm{d}x\ = \frac{Mh^3}{6}, \end{aligned} \tag{5.2}$$

where we note that since we do not know how ξ depends on x, we cannot evaluate $f''(\xi)$ hence we need an upper bound. Let's compute f_{k+1} by the same Taylor expansion:

$$\begin{aligned} f_{k+1} &= f_k + f'(x_k)h + \frac{f''(\xi)}{2}h^2 \\ \implies |f_{k+1} - f_k - f'(x_k)h| &\leq \frac{Mh^2}{2} \quad \text{(multiply by } h/2) \\ \iff \left| h\frac{(f_{k+1} + f_k)}{2} - hf_k - \frac{h^2}{2}f'(x_k) \right| &\leq \frac{Mh^3}{4} \end{aligned} \tag{5.3}$$

Comparing (5.3) and (5.2), we see that the error on our approximation is given by

$$\left| \frac{h(f_{k+1} - f_k)}{2} - \int_{x_k}^{x_{k+1}} f(x)\,\mathrm{d}x \right| \leq \frac{5}{12}Mh^3. \tag{5.4}$$

This is for a single line segment. To obtain the error bound over the whole interval $[a,b]$, we simply multiply by n:

$$\left| \int_a^b f(x)\,\mathrm{d}x\ - \sum_{k=0}^{n-1} \frac{h}{2}(f_{k+1} - f_k) \right| \leq \frac{5}{12}(b-a)Mh^2. \tag{5.5}$$

Hence, we see the error bounded on the order $O(h^2)$. As long as M and the interval length $(b-a)$ is not too large, trapezium can yield decent results. But we can do much better!

5.1.2 Simpson's rule

This is a much better method than trapezium, due to the error bound being on order $O(h^4)$. The rule is as follows:

$$\int_a^b f(x)\,\mathrm{d}x\ \approx \frac{h}{6}\left(f_0 + 4f_{1/2} + 2f_1 + 4f_{3/2} + \cdots + 4f_{n-1/2} + f_n\right), \tag{5.6}$$

where $f_{k+1/2} = f(a + (k+1/2)h)$. We will see where this method comes from, and how we can come up with even better ones.

5.1.3 General method

In general, it is not very helpful to think about integrals as areas under curves. That helps only for the trapezium rule. The better way to think about an integral is as related to the average value of f, since we have

$$\langle f \rangle = \frac{1}{(b-a)} \int_a^b f(x)\,\mathrm{d}x\,.$$

Generally, we may write

$$\begin{aligned} \langle f \rangle = \frac{1}{(b-a)} \int_a^b f(x)\,\mathrm{d}x \, &\approx \frac{1}{n} \sum_{k=0}^{n-1} \sum_{j=1}^{m} \alpha_j f(a + hk + \beta_j h), \\ \implies \int_a^b f(x)\,\mathrm{d}x \, &\approx h \sum_{k=0}^{n-1} \sum_{j=1}^{m} \alpha_j f(a + hk + \beta_j h), \end{aligned} \tag{5.7}$$

where β_j are the nodes on interval $[x_k, x_{k+1}]$ and α_j are weights of the nodes. Note the similarities to taking a weighted average!

The coefficients for the trapezium and Simpson's rules are:

$$\begin{aligned} &\text{Trapezium: } m = 2;\ \alpha_1 = \alpha_2 = \frac{1}{2};\ \beta_1 = 0,\ \beta_1 = 1. \\ &\text{Simpson's rule: } m = 3; \quad \alpha_1 = \frac{1}{6},\ \alpha_2 = \frac{4}{6},\ \alpha_3 = \frac{1}{6};\ \beta_1 = 0,\ \beta_2 = \frac{1}{2},\ \beta_3 = 1. \end{aligned}$$

How do we estimate errors in general? Suppose that we have a bound on the $(k+1)^{\text{st}}$ derivative of f,

$$|f^{(k+1)}(x)| \leq M.$$

Taylor expanding f about some point x_s up to the m^{th} term:

$$\begin{gathered} f(x) = f(x_s) + f'(x_s)(x - x_s) + \cdots + \frac{f^{(m)}(x_s)}{m!}(x - x_s)^m + \frac{f^{(m+1)}(\xi)}{(m+1)!}(x - x_s)^{m+1} \\ \implies |f(x) - P_m(x)| \leq M \frac{(x - x_s)^{m+1}}{(m+1)!}, \end{gathered}$$

where $P_m(x)$ is the first m terms in the Taylor expansion. Now, the idea is to choose α_j and β_j such that *for all* polynomials $g_k(x)$ of degree $k \leq m$, we have

$$\int_0^1 g_k(x)\,\mathrm{d}x \, = \sum_{j=1}^{m} \alpha_j g_k(\beta_j). \tag{5.8}$$

This means the error will be bounded on the order $O(h^{m+1})$.

Note. On equation (5.8), we are integrating from 0 to 1, but that does not matter. Since we require all polynomials to satisfy the condition, we can scale and shift as we would like.

Example (Trapezium rule)**.** Let's check the condition (5.8) for the trapezium rule.

$$\begin{aligned} k = 0 : \int_0^1 \mathrm{d}x \, &= 1 = \frac{1}{2}(1) + \frac{1}{2}(1), \\ k = 1 : \int_0^1 x\,\mathrm{d}x \, &= \frac{1}{2} = \frac{1}{2}(0) + \frac{1}{2}(1), \\ k = 2 : \int_0^1 x^2\,\mathrm{d}x \, &= \frac{1}{3} \neq \frac{1}{2}(0) + \frac{1}{2}(1) \end{aligned}$$

The condition holds up to $k = 1$, which implies error bound is on order $O(h^2)$.

Example (Simpson's rule)**.** Let's do the same procedure for Simpson's rule.

$$\begin{aligned}
k=0: \int_0^1 \mathrm{d}x &= 1 = \frac{1}{6}(1) + \frac{4}{6}(1) + \frac{1}{6}(1),\\
k=1: \int_0^1 x\,\mathrm{d}x &= \frac{1}{2} = \frac{1}{6}(0) + \frac{4}{6}\left(\frac{1}{2}\right) + \frac{1}{6}(1),\\
k=2: \int_0^1 x^2\,\mathrm{d}x &= \frac{1}{3} = \frac{1}{6}(0) + \frac{4}{6}\left(\frac{1}{4}\right) + \frac{1}{6}(1),\\
k=3: \int_0^1 x^3\,\mathrm{d}x &= \frac{1}{4} = \frac{1}{6}(0) + \frac{4}{6}\left(\frac{1}{8}\right) + \frac{1}{6}(1),\\
k=4: \int_0^1 x^4\,\mathrm{d}x &= \frac{1}{5} \neq \frac{1}{6}(0) + \frac{4}{6}\left(\frac{1}{16}\right) + \frac{1}{6}(1).
\end{aligned}$$

We see that the condition holds for Simpson's rule up to $k = 3$, which implies the error bound is on the order $O(h^4)$.

Generally, we have a set of m equations of the form:

$$\begin{aligned}
1 &= \alpha_1 + \cdots + \alpha_m\\
\frac{1}{2} &= \alpha_1\beta_1 + \cdots + \alpha_m\beta_m\\
&\vdots\\
\frac{1}{m} &= \alpha_1\beta_1^{m-1} + \cdots + \alpha_m\beta_m^{m-1}
\end{aligned}$$

Proposition. For *any* choice of nodes $\{\beta_i\}$, we can find weights $\{\alpha_i\}$ such that the error will be $O(h^m)$. This is equivalent to saying a solution exists for any choice of β_i. We assume $\beta_i \neq \beta_j$ if $i \neq j$.

Proof. Write the set of m equations in matrix form and show that the matrix of coefficients β is invertible. □

But we are not done yet! Note that we still have the freedom to choose the nodes $\{\beta_j\}$ as we wish. By choosing the right nodes, we can reduce error below $O(h^m)$. In fact, we may reduce it down to $O(h^{2m})$. This is apparent when we consider that we have $2m$ variables in total, so we may satisfy $2m$ equations. But how may we do this?

Let Q_m be a polynomial of degree m such that Q_m is orthogonal to all polynomials of degree $< m$ on the interval $[0, 1]$. So, we have

$$\int_0^1 Q_m x^s\,\mathrm{d}x = 0 \quad \forall\ 0 \leq s < m.$$

Now, let $\beta_1, \ldots, \beta_m$ be the *roots* of Q_m.

Proposition. For <u>this</u> choice of nodes, the error bound will be on order $O(h^{2m})$.

Proof. We let $g_{m+s}(x) = Q_m(x)x^s$. Then, the condition becomes

$$0 \equiv \int_0^1 Q_m(x)x^s\,\mathrm{d}x = \sum_{j=1}^{m} \alpha_j Q_m(\beta_j)\beta_j^s = 0,$$

where the last equality is due to the β_j being the roots of Q_m. □

A Appendix

A.1 Legendre Transform

The Legendre transform is a useful tool which we will use to switch between the Lagrangian and the Hamiltonian. We have already used it for thermodynamic potentials.

Suppose we have a differentiable function $f(x)$, and for some reason we are interested in expressing f as a function of its derivative, $p = \mathrm{d}f \,/\, \mathrm{d}x$. The simple choice $f^*(p) = f(x(p))$ won't work because it does not have nice properties. We require that

$$\frac{\mathrm{d}f}{\mathrm{d}x} = p \Longleftrightarrow \frac{\mathrm{d}f^*}{\mathrm{d}p} = x,$$

in other words, we want x and p to be conjugates under this transformation. Looking at the differentials, we see that

$$\mathrm{d}f = \frac{\mathrm{d}f}{\mathrm{d}x}\,\mathrm{d}x = p\,\mathrm{d}x\,,$$

and we require

$$\mathrm{d}f^* = x\,\mathrm{d}x\,.$$

So, we define

$$f^*(p) = x(p)p - f(x(p)).$$

This is very simplified, for a better and more general treatment refer to other sources.