

# Supplementary notes on Mathematical Physics

Jari Kinaret

## EXPRESSING A FUNCTION AS A SUM OVER POLES

A function is called *meromorphic* if it is analytic except for a finite number of simple poles. For meromorphic functions there is a theorem, named after Gösta Mittag-Leffler, that states that if  $f(z)$  is meromorphic with poles  $z_j$  and residues  $R_j$ , and if  $z = 0$  is not a pole, and if  $f(z)$  does not diverge as  $|z| \rightarrow \infty$ , then

$$f(z) = f(0) + \sum_j R_j \left( \frac{1}{z - z_j} + \frac{1}{z_j} \right).$$

The proof is quite simple: define a function  $I(z) = \oint_C \frac{dw}{2\pi i} \frac{f(w)}{w(w-z)}$  where  $C$  is circle whose radius  $R$  approaches infinity. Then, firstly,  $I(z) = 0$  since on the integration contour the integrand is proportional to  $R^{-2}$  and the length of the contour is proportional to  $R$ . Secondly, by applying the residue theorem we have

$$I(z) = -\frac{1}{z}f(0) + \frac{1}{z}f(z) + \sum_j R_j \frac{1}{z_j(z - z_j)}$$

where the first term on the right hand side comes from the pole at  $w = 0$ , the second one from the pole at  $w = z$ , and the remaining ones from the poles of  $f(w)$ . The Mittag-Leffler result now follows immediately.

As an application, consider  $f(z) = \frac{1}{\sin z} - \frac{1}{z}$ . One might protest that this has an infinite number of simple poles at  $z = n\pi$ , but in any finite region of space we can approximate the  $f(z)$  with another function that only has a finite number of simple poles, and in the end take the size of the region very large. The end result is, however, what we get if we simply ignore the limitation to a finite number of poles. Then we have that  $f(0) = 0$ , the poles are at  $z_j = j\pi, j \neq 0$ , and the residues are  $R_j = \lim_{z \rightarrow j\pi} \frac{z - j\pi}{\sin z} = \lim_{z \rightarrow j\pi} \frac{1}{\cos z} = (-1)^j$  where we used l'Hospital's rule to evaluate the limit. Hence, we have that

$$\frac{1}{\sin z} = \frac{1}{z} + \sum_{j \neq 0} (-1)^j \left( \frac{1}{z - j\pi} + \frac{1}{j\pi} \right) = \sum_j \frac{(-1)^j}{z - j\pi}.$$

Occasionally this expansion can be used to simplify integrals. For instance, we have  $\coth z = \frac{1}{z} + 2 \sum_{j=1}^{\infty} \frac{z}{z^2 + (j\pi)^2}$ , and consequently

$$\lim_{\alpha \rightarrow 0^+} \int_0^{\infty} dz e^{-\alpha z} \sin z \coth z = \int_0^{\infty} dx \frac{\sin x}{x} + 2 \sum_{j=1}^{\infty} \int_0^{\infty} dz \frac{z \sin(kz)}{z^2 + (j\pi)^2} = \operatorname{sgn}(z) \left( \frac{\pi}{2} + \pi \sum_{j=1}^{\infty} e^{-j\pi|k|} \right)$$

where the final expression equals  $\frac{\pi}{2} \coth\left(\frac{1}{2}\pi k\right)$ . Note that the original integral is ill-defined for  $\alpha = 0$  (since for large arguments  $\coth z \approx 1$ ) but after substituting the expansion for the hyperbolic cotangent the limit of vanishing  $\alpha$  can be taken without problems.

## ANALYTIC CONTINUATION: RIEMANN ZETA FUNCTION $\zeta(s)$

Sometimes analytic continuation is a bit more involved than in the simple example of a geometric series. To illustrate this, let us consider the potentially profitable example of the Riemann zeta function.

The sum

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} \quad (1)$$

converges for  $\text{Re}(s) > 1$  and defines a function that is analytic in that region. For  $\text{Re}(s) \leq 1$  the sum diverges; however, it is possible to find an analytic continuation of  $\zeta(s)$  that agrees with (1) for  $\text{Re}(s) > 1$  and is analytic in the entire complex plane with exception of the point  $s = 1$  where it has a simple pole.

To begin with, let us recall that the Euler gamma function has the integral representation  $\Gamma(s) = \int_0^{\infty} dx x^{s-1} e^{-x}$  which yields, by change of variables,  $n^{-s}\Gamma(s) = \int_0^{\infty} dx x^{s-1} e^{-nx}$ . By summing over  $n$  we obtain

$$\zeta(s)\Gamma(s) = \int_0^{\infty} dx x^{s-1} \left[ \frac{1}{1-e^{-x}} - 1 \right] = \int_0^{\infty} dx x^{s-1} \frac{1}{e^x - 1}. \quad (2)$$

Due to the appearance of  $e^x$  in the dominator, this integral converges at the upper limit regardless of  $s$ , but it converges at the lower limit only for  $\text{Re}(s) > 1$ .

Let us now, for no apparent reason, consider the contour integral

$$I_C = \int_C dz \frac{(-z)^{s-1}}{e^z - 1}$$

where  $C$  is given in Fig. 1 and the branch of  $(-z)^{s-1}$  is so chosen that the integrand is real for  $z < 0$  — hence, there is a branch cut on the positive real axis. Writing  $(-1)^{s-1} = -e^{-is\pi}$  on the upper and  $(-1)^{s-1} = -e^{is\pi}$  on the lower branch of the contour (in agreement with the branch choice made above), and changing the integration limits on the upper branch so that the integral runs from 0 to  $+\infty$  instead of the other way, gives

$$I_C = -(e^{is\pi} - e^{-is\pi}) \int_0^{\infty} dx \frac{x^{s-1}}{e^x - 1} = -2i \sin(s\pi) \int_0^{\infty} dx \frac{x^{s-1}}{e^x - 1}. \quad (3)$$

We now recognize the remaining integral as the one that appears in (2), and have the expression

$$\zeta(s) = \frac{1}{\Gamma(s)} \frac{1}{-2i \sin(s\pi)} \int_C dz \frac{(-z)^{s-1}}{e^z - 1} \quad (4)$$

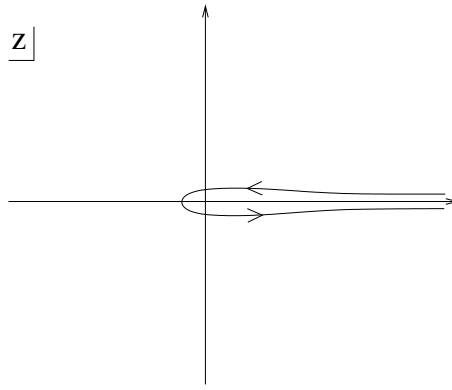


FIG. 1: The integration contour  $C$  used in the analytic continuation of the Riemann zeta function  $\zeta(s)$ .

which is analytic for  $s \neq 1$  and has a first order pole at  $s = 1$  (the problems for  $z$  near the origin disappear as a result of the loop nature of  $C$ ).[1] This can be further simplified using the relation  $\Gamma(s)\Gamma(1-s) = \pi/\sin(\pi s)$  which yields the final expression

$$\zeta(s) = -\frac{\Gamma(1-s)}{2\pi i} \int_C dz \frac{(-z)^{s-1}}{e^z - 1}. \quad (5)$$

Using this expression we can, *e.g.*, evaluate  $\zeta(-1)$ , which in the original sum representation would yield the meaningless expression  $1 + 2 + 3 + \dots$ . Inserting  $s = -1$  in (5) we have

$$\begin{aligned} \zeta(-1) &= -\frac{\Gamma(2)}{2\pi i} \int_C dz \frac{(-z)^{-2}}{e^z - 1} \\ &= -\frac{1}{2\pi i} \oint_{C'} dz \frac{(-z)^{-2}}{e^z - 1} \\ &= -\text{Res}_{z=0} \left( \frac{(-z)^{-2}}{e^z - 1} \right) \\ &= -\frac{1}{2!} \left[ \frac{d^2}{dz^2} \left( \frac{(-z)^{-2}}{e^z - 1} \right) \right]_{z=0} \\ &= -\frac{1}{12} \end{aligned}$$

where I used that (i)  $\Gamma(2) = 1! = 1$ , (ii) the contour  $C$  can be closed at infinity since the integrand vanishes for  $\text{Re}(z) \rightarrow +\infty$  (the resulting closed contour is called  $C'$ ), and (iii) the integrand has a third order pole at the origin.

Similarly, we obtain that  $\zeta(s) = 0$  for  $s = -2, -4, -6, \dots$  since the residue at  $z = 0$  vanishes by symmetry. These zeroes of the zeta function are known as trivial zeroes; there are other zeroes as well, and the most famous open problem in mathematics, the Riemann hypothesis (B. Riemann, 1859), speculates that all non-trivial zeroes of  $\zeta(s)$  occur on the line  $\text{Re}(s) = 1/2$ . Proving or disproving this hypothesis will bring you 1000000\$ from the

Clay Mathematical Institute (minus my share). Much work has been done, just to give a hint of how much I can quote that the  $(10^{22}+1)^{\text{st}}$  zero on the line is approximately given by

$$\frac{1}{2} + i1\,370\,919\,909\,931\,995\,308\,226.68016095 \pm i10^{-6}.$$

## EVALUATION OF SUMS

The standard mathematics curriculum focuses much on *integrals*, and pays relatively little attention to their discrete counterparts, *sums*. Sums occur quite frequently in physics, and while it is often justified to approximate them by integrals, it is nevertheless useful to know some tricks that allow either an exact or approximate summation of a series, or even may help us recover some of the effort that yielded a seemingly useless result, a divergent series.

### Sums over residues

One of the most common tricks in evaluating sums is to interpret them as sums over residues of a suitable function. This technique is widely used in field theory where such sums are known as Matsubara sums — in this context the technique was introduced by A.A. Abrikosov, L.P. Gorkov, and I.E. Dzyaloshinski in their invaluable book *Methods of Quantum Field Theory in Statistical Physics*, first published in 1961. As an example, consider the sum

$$S = \sum_{n=-\infty}^{\infty} (-1)^n f(n). \quad (6)$$

We start by evaluating the contour integral

$$I = \oint_C \frac{dz}{2\pi i} \frac{\pi}{\sin(\pi z)} f(z) \quad (7)$$

where  $C$  is a circle about the origin, with radius  $R \rightarrow \infty$ , and assume that  $|zf(z)| \rightarrow 0$  as  $|z| \rightarrow \infty$  so that the integral vanishes,  $I = 0$ . Using residue theorem we have

$$I = \sum_{\text{poles}} \text{Res} \left[ \frac{\pi}{\sin(\pi z)} f(z) \right] \quad (8)$$

where the poles arise either from zeroes of  $\sin(\pi z)$  or from the poles of  $f(z)$ . The former are given by  $z = n$ , and yield residues  $(-1)^n f(n)$  so that

$$0 = \sum_{n=-\infty}^{\infty} (-1)^n f(n) + \sum_{\text{poles of } f(z)} \text{Res} \left[ \frac{\pi}{\sin(\pi z)} f(z) \right]. \quad (9)$$

For example, we have

$$\begin{aligned} \sum_{n=-\infty}^{\infty} \frac{(-1)^n}{(n+a)^2} &= - \sum_{\text{poles of } (z+a)^{-2}} \text{Res} \left[ \frac{\pi}{\sin(\pi z)} \frac{1}{(z+a)^2} \right] \\ &= - \frac{1}{1!} \left( \frac{d}{dz} \frac{\pi}{\sin(\pi z)} \right)_{z=-a} \\ &= \pi^2 \frac{\cos(\pi a)}{\sin^2(\pi a)}. \end{aligned}$$

The sum  $\sum_{n=-\infty}^{\infty} f(n)$  is evaluated similarly using  $\pi \cot(\pi z)$  instead of  $\pi/\sin(\pi z)$ .

The sums that occur frequently in field theory are of the form

$$\sum_{n=-\infty}^{\infty} F\left(n \frac{2\pi i}{\beta}\right) = \frac{\beta}{2\pi i} \oint_C dz n_B(z) F(z)$$

where  $n_B(z) = [e^{\beta z} - 1]^{-1}$  is the Bose function. What kind of sums can be evaluated similarly using the Fermi function  $n_F(z) = [e^{\beta z} + 1]^{-1}$ ?

### Poisson summation formula

If a series cannot be summed exactly, it is often sufficient to include few of its largest terms. This works well as long as there are only a few terms that dominate, but fails if the number of comparable terms is large. In that case it is often favorable to carry out a transformation known as Poisson summation formula. This basically amounts to a Fourier transform and relies on the equality

$$\sum_{m=-\infty}^{\infty} f(2\pi n) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} d\tau e^{-ik\tau} f(\tau).$$

Often it turns out that if the sum on the left is “flat”, the one on the right is “peaked” — *cf.* the fact that the Fourier transform of a constant (extremely flat) is a delta function (extremely peaked). Sometimes it is useful to divide a sum into two parts, one flat and the other peaked, and Fourier transform the flat part; this is *e.g.* done in condensed matter physics to obtain the energy of an ionic crystal (*Ewald summation*).

### Asymptotic series

Let us imagine a situation — rather a common one — that a physicist has encountered a complicated differential equation, and as a reflex solved it using the power series method.

To his amazement, he finds a closed expression for the coefficients, but that's where his luck runs out: the series turns out to diverge in the relevant parameter range. What can he do to rescue as much of his work as possible?

He should consider a method to determine “a value” (rather than the sum) of the series. There are quite a few such methods known as summation techniques, but the one that is most commonly used in physics is known as Borel summation. In brief, it is given by

$$S = \sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} \frac{n!}{n!} a_n = \sum_{n=0}^{\infty} \frac{\int_0^{\infty} dt e^{-t} t^n}{n!} a_n \rightarrow \int_0^{\infty} dt e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} a_n = S_B$$

where the new sum converges better than the original one. As an example,

$$S(q) = \sum_{n=0}^{\infty} (2n+1)!! q^n = \sum_{n=0}^{\infty} \frac{\int_0^{\infty} dt e^{-t} t^{2n+1}}{2^n n!} q \rightarrow \int_0^{\infty} dt e^{-t} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{t^2 q}{2}\right)^n = \int_0^{\infty} dt e^{-t} t e^{\frac{t^2 q}{2}}$$

which converges for  $q < 0$  whereas the original sum does not converge for any  $q \neq 0$ . Here  $(2n+1)!! = 1 \cdot 3 \cdot 5 \dots \cdot (2n+1) = (2n+1)!/(2^n n!)$ . The final expression, the Borel sum  $S_B(q)$ , may be an acceptable solution of the initial problem for  $q < 0$ . [2]

The Borel sum is usually an analytic continuation of the original sum to a larger domain. The original sum, in turn, is often an asymptotic series for the integral expression  $S_B(q)$ . The series  $\sum_{n=0}^N a_n (x-x_0)^n$  is an asymptotic series for the function  $f(x)$  if

$$\left| f(x) - \sum_{n=0}^N a_n (x-x_0)^n \right| \ll (x-x_0)^N \text{ as } x \rightarrow x_0 \text{ for fixed } N.$$

An asymptotic series often diverges in the limit  $N \rightarrow \infty$  but they are nevertheless quite useful if terminated properly. As a rule of thumb, truncating an asymptotic series just before the term whose magnitude is smallest yields a good approximation for  $f(x)$ .

As an example consider  $f(x) = \int_0^{\infty} dt e^{-t^2}/(1+xt)$ . This is an example of a so-called Stieltjes integral  $\int_0^{\infty} dt \rho(t)/(1+xt)$  where  $\rho(t) \geq 0$  and  $a_n = \int_0^{\infty} dt t^n \rho(t)$  are finite. In general, a Stieltjes integral converges for a positive  $x$  and diverges for a negative  $x$ , and has an asymptotic expansion of the form  $f(x) \sim \sum_{n=0}^{\infty} a_n (-x)^n$  as  $x \rightarrow 0^+$ . In our case

$$a_n = \int_0^{\infty} dt e^{-t^2} t^n = \frac{1}{2} \int_0^{\infty} du u^{\frac{n-1}{2}} e^{-u} = \frac{1}{2} \Gamma\left(\frac{n+1}{2}\right)$$

so that

$$f(x) \sim \frac{1}{2} \sum_{n=0}^{\infty} \Gamma\left(\frac{n+1}{2}\right) (-x)^n.$$

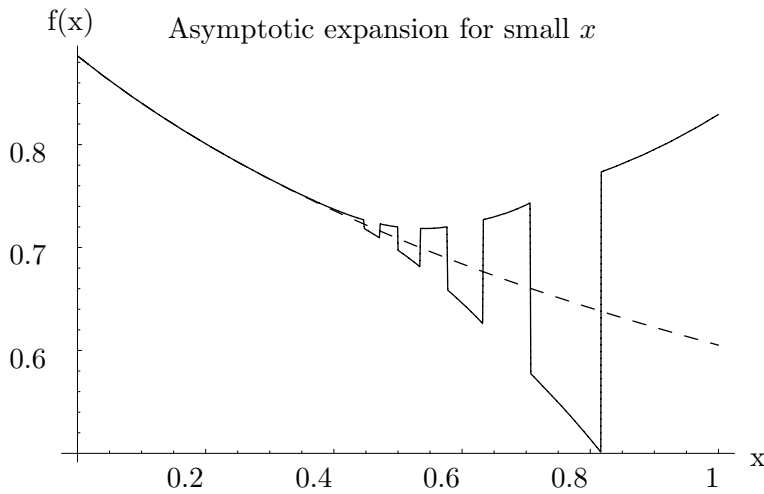


FIG. 2: Asymptotic approximation *vs.* numerical integration for the  $f(x) = \int_0^\infty dt e^{-t^2}/(1+xt)$ .

The ratio of two consecutive terms is given by

$$x \frac{\Gamma\left(\frac{n+1}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} \approx x \frac{\left(\frac{n+1}{2} + \frac{1}{2}\right)^{\frac{n+1}{2}} e^{-\frac{n+1}{2} - \frac{1}{2}}}{\left(\frac{n+1}{2}\right)^{\frac{n+1}{2} - \frac{1}{2}} e^{-\frac{n+1}{2}}} \approx x \sqrt{n/2} \frac{1}{\sqrt{e}} \left(1 + \frac{1}{2} \frac{1}{n/2}\right)^{n/2} \approx x \sqrt{n/2}$$

where I approximated Gamma functions using the Stirling formula. This shows that (i) the complete sum diverges for all  $x \neq 0$  (the ratio of consecutive terms tends to infinity as  $n$  grows), and (ii) for a given  $x$  the smallest term is given by  $n_{\text{opt}} \approx 2/x^2$ , suggesting the approximation

$$f(x) \approx \frac{1}{2} \sum_{n=0}^{\lfloor 2/x^2 \rfloor} \Gamma\left(\frac{n+1}{2}\right) (-x)^n$$

which is plotted in Fig. 2 together with the numerical evaluation of the integral.[3]

### THE EASIEST FUNCTION OF MATHEMATICAL PHYSICS: $\delta(x)$

A remarkably useful concept in mathematical physics is the Dirac delta function,  $\delta(x)$ , which is not really a proper function — mathematicians call it a generalized function or a distribution. What makes  $\delta(x)$  useful is that for any continuous function  $f(x)$  we have

$$\int_{-\infty}^{\infty} dx \delta(x) f(x) = f(0). \quad (10)$$

Intuitively,  $\delta(x)$  is an extremely sharp peak at  $x = 0$  such that the area under the peak equals unity. As such, it often appears in simplified mathematical models — for



instance, the charge density of point charges with magnitudes  $q_i$  at positions  $\mathbf{r}_i$  is given by  $\rho(\mathbf{r}) = \sum_i q_i \delta^{(3)}(\mathbf{r} - \mathbf{r}_i)$  where  $\delta^{(3)}(\mathbf{r}) = \delta(x)\delta(y)\delta(z)$  is the three dimensional delta function.

More rigorously the delta function can be defined as a limit of a sequence of functions. Many choices of a sequence are possible and they are generally known as representations of the delta function. Among the more common representations are

$$f_n^{(1)}(x) = \frac{n}{\pi} \frac{1}{1 + (nx)^2} \quad (11)$$

$$f_n^{(2)}(x) = \sqrt{\frac{n}{\pi}} e^{-nx^2} \quad (12)$$

$$f_n^{(3)}(x) = \begin{cases} n, & |x| < \frac{1}{2n} \\ 0, & |x| > \frac{1}{2n} \end{cases}. \quad (13)$$

Despite the fact that  $\delta(x)$  is not a proper function, it often behaves like one. For instance, formal integration by parts yields  $\int_{-\infty}^{\infty} dx \delta'(x) f(x) = -f'(x)$  which can be verified by using any of the representations of the delta function ( $f(x)$  must satisfy suitable regularity requirements). The integral of the Dirac delta function is known as the Heaviside theta function, and is given by  $\Theta(x) = \int_{-\infty}^x dx' \delta(x')$ . For  $x < 0$  we have  $\Theta(x) = 0$  and for  $x > 0$   $\Theta(x) = 1$ . On the other hand, some innocent looking expressions like  $\int_{-\infty}^{\infty} dx [\delta(x)]^2 f(x)$  are meaningless for a general  $f(x)$ .

Perhaps the most common appearance of  $\delta(x)$  in physics is in connection with the strange looking formula

$$\lim_{\eta \rightarrow 0^+} \frac{1}{x + i\eta} = \mathcal{P} \left( \frac{1}{x} \right) - i\pi\delta(x) \quad (14)$$

which was proven using residues during the lecture on complex analysis. The equation is strange since in the limit  $\eta \rightarrow 0$  the left hand side looks real while the right hand side is manifestly complex. Furthermore, the right hand side contains the letter  $\mathcal{P}$  which denotes *Cauchy principal part*, and is a recipe of how to handle the singularity of  $\frac{1}{x}$  at  $x = 0$ :  $\mathcal{P} \int_{-\infty}^{\infty} dx \frac{f(x)}{x} \equiv \lim_{\epsilon \rightarrow 0^+} \left( \int_{-\infty}^{-\epsilon} dx \frac{f(x)}{x} + \int_{\epsilon}^{\infty} dx \frac{f(x)}{x} \right)$ . Thus  $\mathcal{P}$  tells us to regularize the divergent integral  $\int_{-\infty}^{\infty} dx \frac{f(x)}{x}$  by approaching the singularity at  $x = 0$  at an equal rate from the positive and negative sides so that contributions from above and below the singularity take out each other. The equality 14 is therefore only applicable inside an integral sign: for sufficiently well-behaved functions  $f(x)$ , the integral of  $f(x)$  times the left hand side of 14 equals the integral of  $f(x)$  times the right hand side of 14. Due to the presence of the delta function, the latter integral is usually easier to evaluate, wherein lies the usefulness of 14.

## APPROXIMATE EVALUATION OF INTEGRALS

[4] Often in physics we encounter a situation where an intermediate result of a calculation is given by a definite integral that is too complicated to evaluate exactly, but to proceed with the analysis we need an approximation of the integral. A general approximation is usually difficult to find, and often not even needed, whereas the asymptotic behavior of the integral in some limiting cases may be more tractable. An example is  $I(x) = \int_x^\infty dt e^{-t^4}$ . The obvious strategy of expanding the integrand in a Taylor series does not work (the integral of each term diverges), so we need to be a bit more creative.

One way is to write  $I(x) = I(0) - \int_0^x dt e^{-t^4}$ , where the remaining function of  $x$  can be evaluated by Taylor expanding, and the constant  $I(0)$  can be evaluated, *e.g.*, by substituting  $u = t^4$ , which yields  $I(0) = (1/4) \int_0^\infty du u^{-3/4} e^{-u} = \Gamma(1/4)/4 = \Gamma(5/4)$  where  $\Gamma(x) = \int_0^\infty du u^{x-1} e^{-u}$  is the Euler gamma function (for positive integers,  $\Gamma(n+1) = n!$ ). The disadvantage of this approach is that for large  $x$  the remaining series converges very slowly. Therefore, for large  $x$  it is better to integrate by parts to obtain

$$I(x) = \int_x^\infty dt \left( -\frac{1}{4t^3} \right) \frac{\partial}{\partial t} \left( e^{-t^4} \right) = \frac{1}{4x^3} e^{-x^4} - \frac{3}{4} \int_x^\infty dt t^{-4} e^{-t^4} \quad (15)$$

where the remaining integral is much smaller than  $I(x)$  for large  $x$ :  $\int_x^\infty dt t^{-4} e^{-t^4} \leq x^{-4} I(x)$ . Hence,  $I(x) \approx \frac{1}{4x^3} e^{-x^4}$  for large  $x$ . If you need a more accurate approximation, you can proceed similarly with the remaining integral. Integration by parts can often be used to find out the asymptotic behavior of integrals.

However, sometimes the situation is slightly more subtle. One category of integrals that can be analyzed in a more powerful way originates from the Laplace integral of the form  $I(x) = \int_a^b dt f(t) e^{x\phi(t)}$  where  $x$  is regarded as a parameter and we are interested in the behavior of  $I(x)$  for large  $x$ . We can use integration by parts repeatedly, which results in a power series of the form  $e^{x\phi(c)} \sum_{n=1}^\infty A_n x^{-n}$  where  $c = a$  or  $c = b$ , depending of which of  $\phi(a)$  and  $\phi(b)$  is larger. More generally, if  $x$  is large, the main contribution to the integral comes from the vicinity of the point where  $\phi(t)$  is largest.

As an example, consider  $I(x) = \int_0^{10} dt (1+t)^{-1} e^{-xt}$ . For large  $x$  the maximum of the integrand occurs for  $t = 0$ , and only the vicinity of that point matters. We then Taylor expand the first factor of the integrand and obtain  $I(x) \approx \int_0^\epsilon dt (1-t+t^2-\dots) e^{-xt} \approx \int_0^\infty \sum_{n=0}^\infty (-t)^n e^{-xt} = \sum_{n=0}^\infty (-1)^n n! x^{-(n+1)}$  where I used repeatedly that only a small interval

near  $t = 0$  matters, so the range of integration could be modified in an arbitrary fashion for sufficiently large  $x$ . The resulting series is an asymptotic series for  $I(x)$  for large  $x$  — essentially, we have carried out Borel summation in reverse. Note that while the series diverges for all  $x \neq 0$ , it is nevertheless quite useful for approximating  $I(x)$  for large  $x$  provided that we keep only a finite number of terms.

There is a general result for Laplace integrals, known as Watson's lemma: if  $f(t) \sim t^\alpha \sum_{n=0}^{\infty} a_n t^{n\beta}$  as  $t \rightarrow 0^+$ , then

$$I(x) = \int_0^b dt f(t) e^{-xt} \sim \sum_{n=0}^{\infty} a_n \frac{\Gamma(\alpha + n\beta + 1)}{x^{\alpha + n\beta + 1}} \quad (16)$$

as  $x \rightarrow \infty$ .

Occasionally a straightforward application of Watson's lemma fails — for instance, the function  $f(t)$  may not have an expansion of the required form — and we have to be more careful. Then the method of choice is usually any of the methods that are referred to as method of steepest descent, method of stationary phase, or saddle point method; the different names refer to slightly different cases but the basic idea is always the same. Consider  $I(x) = \int_{-\infty}^{\infty} dt f(t) e^{x\phi(t)}$  where  $x \rightarrow \infty$ . As previously discussed, only the regions near the maxima of  $\phi(t)$  are important, so let us assume that  $\phi(t)$  assumes its largest value for  $t = t_0$ . Near this point we can write  $\phi(t) \approx \phi(t_0) + (1/2)\phi''(t_0)(t - t_0)^2$ , which yields

$$I(x) \approx e^{x\phi(t_0)} \int_{-\infty}^{\infty} dt f(t) e^{-\frac{1}{2}x|\phi''(t_0)|(t-t_0)^2} \approx e^{x\phi(t_0)} f(t_0) \sqrt{\frac{2\pi}{|x\phi''(t_0)|}} \quad (17)$$

where the corrections are higher order in  $1/x$ .

As an example, consider  $\int_0^{\infty} dt e^{-xt-1/t}$ . We cannot use Watson's lemma (why?), but we can use the saddle point method: the maximum of the integrand occurs at  $t_0 = \sqrt{x}$ , and near this point the exponent can be written as  $-2\sqrt{x} - x^{3/2}(t - t_0)^2$ , so a direct application of the above expression yields the approximation  $\sqrt{\pi}x^{-3/4}e^{-2\sqrt{x}}$ .

Applications of this method to oscillatory and complex integrals are known as method of stationary phase and saddle point method, respectively, and will be discussed at the lecture.

## FUNCTIONAL DERIVATIVES

A function can be regarded as a mapping of a number to a number,  $f : x \rightarrow y$  or  $y = f(x)$ . A functional, in turn, is a mapping of a function to a number,  $\mathcal{A} : f \rightarrow y$  or

$y = \mathcal{A}[f]$ . Examples of functionals are  $I[f] = \int_{-\infty}^{\infty} dx |f(x)|^2$  and  $I_x[f] = \int_{-\infty}^x dx' f(x')$  which are both functionals of  $f$  and the latter is also a function of  $x$ .

In calculus the derivative answers the question *How does the value of  $f(x)$  change if the value of  $x$  changes by a small amount?*, or formally

$$f(x + \delta x) = f(x) + \frac{df}{dx} \delta x + o(\delta x)$$

where  $o(x)$  denotes a quantity that is smaller than  $\delta x$ , *i.e.*,  $\lim_{x \rightarrow 0} \frac{o(x)}{x} = 0$ . By analogy the derivative of a functional should be defined as something like  $\mathcal{A}[f + \delta f] = \mathcal{A}[f] + \frac{\delta \mathcal{A}}{\delta f} \delta f + o(\delta f)$ . This definition has, however, one difficulty: there are many ways of changing a function and it is unclear how we should interpret the coefficient  $\frac{\delta \mathcal{A}}{\delta f}$ . A more careful definition is to write

$$\mathcal{A}[f + \delta f] = \mathcal{A}[f] + \mathcal{L}[\delta f] + o(\delta f) \tag{18}$$

where  $\mathcal{L}[\delta f]$  is a *linear* functional of  $\delta f$ . According to the Frechét-Riesz representation theorem from 1907, any linear functional can be written as an integral,

$$\mathcal{L}[\delta f] = \int dx \frac{\delta \mathcal{A}}{\delta f(x)} \delta f(x)$$

where  $\frac{\delta \mathcal{A}}{\delta f(x)}$  is some function of  $x$  — which function, depends on both  $\mathcal{A}$  and  $f$ . We call the coefficient  $\frac{\delta \mathcal{A}}{\delta f(x)}$  a functional derivative of  $\mathcal{A}$ : intuitively, it answers the question *How does the value of  $\mathcal{A}[f]$  change if  $f$  is changed slightly at point  $x$ ?* Note that it is not always possible to find a linear functional  $\mathcal{L}[\delta f]$  that satisfies the equation (18) — in such a case the functional is not differentiable — although right at the moment I cannot think of a physically relevant non-differentiable functional.

Finding the functional derivative of a functional is usually quite straightforward. Consider for instance the functional  $\mathcal{A}_n[f] = \int_{-\infty}^{\infty} dx [f(x)]^n$ : we have

$$\mathcal{A}_n[f + \delta f] - \mathcal{A}_n[f] = \int_{-\infty}^{\infty} dx ([f(x) + \delta f(x)]^n - [f(x)]^n) = \int_{-\infty}^{\infty} dx n [f(x)]^{n-1} \delta f(x) + o(\delta f)$$

so that  $\frac{\delta \mathcal{A}_n}{\delta f(x)} = n [f(x)]^{n-1}$ , *cf.* ordinary derivatives. A slightly more complicated case is the double integral  $\mathcal{A}[f] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dx' f(x) K(x, x') f(x')$  which yields  $\frac{\delta \mathcal{A}}{\delta f(x)} = \int_{-\infty}^{\infty} dx' [K(x, x') + K(x', x)] f(x')$ .

If the functional  $\mathcal{A}$  depends not only on  $f$  but also on some of its derivatives, integration by parts is usually necessary to cast the change  $\mathcal{A}[f + \delta f] - \mathcal{A}[f]$  into the integral form. As

an example, consider  $\mathcal{A}[f] = \int_{-\infty}^{\infty} dx |f'(x)|^2$  which yields

$$\mathcal{A}[f+\delta f] - \mathcal{A}[f] = \int_{-\infty}^{\infty} dx 2 \frac{df}{dx} \frac{d\delta f}{dx} + o(\delta f) = \int_{-\infty}^{\infty} dx \left[ 2 \frac{d}{dx} \left( \frac{df}{dx} \delta f(x) \right) - 2 \frac{d^2 f}{dx^2} \delta f(x) \right] + o(\delta f).$$

The substitution term must vanish if  $\mathcal{A}[f]$  is finite, and hence  $\frac{\delta \mathcal{A}_n}{\delta f(x)} = -2 \frac{d^2 f}{dx^2}$ .

If the functional  $\mathcal{A}[f]$  is not initially in an integral form, it is usually convenient to write it as such. Consider *e.g.*  $\mathcal{A}_y[f] = f(y)$  which can be written as  $\mathcal{A}_y[f] = \int_{-\infty}^{\infty} dx \delta(x-y) f(x)$  so that  $\frac{\delta \mathcal{A}_y}{\delta f(x)} = \delta(x-y)$ , which is the functional derivative equivalent of the familiar  $\frac{dx}{dx} = 1$ .

The usefulness of functional derivatives in physics stems from the fact that most major physical laws can be formulated as extremal principles. For instance, in classical mechanics the trajectory of a particle can be calculated in many ways, one of which is known as Hamilton's principle or the principle of minimal action: a particle follows the path  $\mathbf{r}(t)$  that minimizes the action  $S[\mathbf{r}(t)] = \int_{-\infty}^{\infty} dt L(t, \mathbf{r}(t), \dot{\mathbf{r}}(t))$  where  $L = T - V$  is the Lagrangian. Determining the physical path  $\mathbf{r}(t)$  from Hamilton's principle can be done using variational calculus which relies heavily on the use of functional derivatives; variational calculus will be covered in the final part of this course.

## SUFFICIENT CONDITIONS IN VARIATIONAL CALCULUS

This section follows the treatment found in I.M. Gelfond and S.V. Fomin: *Calculus of Variations*, Dover, 1963.

In the lectures we were rather vague about what kind of extrema — typically minima — we were looking for. In ordinary calculus we typically discuss local minima and global minima, and the former ones can be related to the points where the derivatives vanish. Similarly, the Euler-Lagrange equation, which corresponds to the vanishing of a functional derivative, yields candidates for local rather than global extrema. In the case of functions, however, it is not as easy to define what we mean by *local* as there are many possible norms we could use. The simplest measure for the difference between two functions is the maximum difference between them,  $\|y\|_0 = \max_{t_0 < t < t_1} |y(t)|$ . A function  $y(t)$  that gives the extremum of a functional relative to all functions  $z(t)$  that satisfy  $\|y(t) - z(t)\|_0 < \epsilon$  for some  $\epsilon > 0$  is called a *strong* extremum. Another norm is  $\|y\|_1 = \|y\|_0 + \max_{t_0 < t < t_1} |y'(t)|$  which takes into account even the difference. A weak extremum is defined as an extremum over all functions  $z(t)$  such that  $\|y(t) - z(t)\|_1 < \epsilon$ , which is a smaller domain than what is used for strong

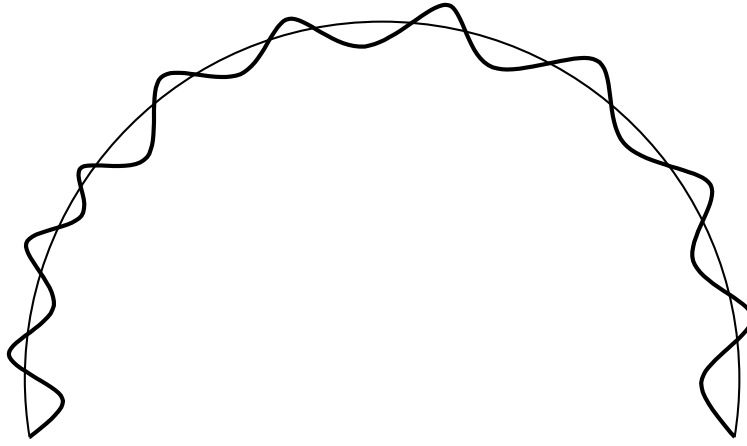


FIG. 3: Two curves that are close to each other in the strong sense (norm  $\|y\|_0$ ) but not in the weak sense (norm  $\|y\|_1$ ).

extrema, implying that weak extrema are easier to find than strong extrema. Typically we are interested in weak extrema. As an example of the difference between the two measures, consider the Figure 3 and note that while the two curves are in each other's neighborhood in terms of the strong norm, they differ significantly for instance in terms of the lengths. As is apparent from the definitions, each strong extremum is automatically a weak extremum as well, while not every weak extremum is a strong one. Hence, if we are interested in finding a strong extremum, we can begin with the requirements (necessary or sufficient) for weak extrema, and tighten them.

We have shown earlier that a necessary condition that the functional  $\mathcal{F} = \int_{t_0}^{t_1} dt F(t, y(t), y'(t))$  has a (weak) extremum for a particular function  $y_0(t)$  is that  $y_0$  satisfies the Euler-Lagrange equation

$$\frac{\partial F}{\partial y} - \frac{d}{dt} \frac{\partial F}{\partial y'} = 0$$

and the boundary conditions  $y(t_0) = a$  and  $y(t_1) = b$ . This is, however, just a necessary condition and not a sufficient one.

To derive a sufficient condition for a weak minimum we need to consider the second variation of  $\mathcal{F}$  and Taylor expand  $F(t, y + \delta y, y' + \delta y')$  to write

$$\begin{aligned} \mathcal{F}[y + \delta y] - \mathcal{F}[y] &= \int_{t_0}^{t_1} dt \left[ \frac{\partial F}{\partial y} - \frac{d}{dt} \frac{\partial F}{\partial y'} \right] \delta y(t) \\ &+ \frac{1}{2} \int_{t_0}^{t_1} dt \left[ \frac{\partial^2 F}{\partial y(t) \partial y(t)} [\delta y(t)]^2 + \frac{\partial^2 F}{\partial y(t) \partial y'(t)} \delta y(t) \delta y'(t) + \frac{\partial^2 F}{\partial y'(t) \partial y(t)} \delta y'(t) \delta y(t) + \frac{\partial^2 F}{\partial y'(t) \partial y'(t)} \delta y'(t) \delta y'(t) \right] + o((\delta y)^2). \end{aligned} \tag{19}$$

If  $y(t)$  satisfies the Euler-Lagrange equation, the first variation vanishes, and to second order in  $\delta y(t)$  we have

$$\mathcal{F}[y+\delta y]-\mathcal{F}[y]=\frac{1}{2}\int_{t_0}^{t_1}dt\left\{\left[\frac{\partial^2 F}{\partial y(t)\partial y(t)}-\frac{d}{dt}\frac{\partial^2 F}{\partial y'(t)\partial y(t)}\right](\delta y(t))^2+\left[\frac{\partial^2 F}{\partial y'(t)\partial y'(t)}\right](\delta y'(t))^2\right\}\quad (20)$$

after one integration by parts.

At first glance it would now appear the a minimum requires that the coefficients of the non-negative terms  $(\delta y(t))^2$  and  $(\delta y'(t))^2$  would be non-negative. That is, however, not quite correct. It is certainly necessary that the coefficient of  $(\delta y'(t))^2$  must be positive — if there is a point  $t'$  such that  $\frac{\partial^2 F}{\partial y'(t')\partial y'(t')} < 0$ , we can construct a function  $\delta(t)$  that is very small everywhere except near  $t'$ , and is strongly peaked at  $t = t'$ , which will force the second variation to be negative. Surprisingly enough, positivity of the coefficient of  $(\delta y(t))^2$  is not necessary.

To investigate the issue further we consider the integral  $\int_{t_0}^{t_1} dt [P(t)(h'(t))^2 + Q(t)(h(t))^2]$  where  $h(t)$  is a function that vanishes at the end points of the integral. In terms of  $h(t)$ , this can be viewed as variational problem whose Euler-Lagrange equation is given by

$$-\frac{d}{dt}[P(t)h'(t)] + Q(t)h(t) = 0. \quad (21)$$

Note that this auxiliary variational problem is closely related to the second variation we considered above, provided that  $P(t) = \frac{\partial^2 F}{\partial y'(t)\partial y'(t)}$  and  $Q(t) = \frac{\partial^2 F}{\partial y(t)\partial y(t)} - \frac{d}{dt}\frac{\partial^2 F}{\partial y'(t)\partial y(t)}$ . Hence, finding the conditions under which the second variation is positive definite so the solution of the Euler-Lagrange equation is a minimum of the original variational problem is equivalent to finding the conditions under which the auxiliary integral is positive definite.

The integral is positive definite if we can write it as  $\int_{t_0}^{t_1} dt R(t)\phi^2(t)$  where  $R(t) > 0$  and  $\phi(t)$  is an expression that is not identically zero unless  $h(t)$  is identically zero. We can accomplish this task if we add  $\int_{t_0}^{t_1} dt \frac{d}{dt}(w(t)h^2(t)) = 0$  to the integral and choose  $w(t)$  as the solution of  $P(t)(Q(t) + w'(t)) = w^2(t)$  since that allows us to write  $Ph'^2 + Qh^2 + \frac{d}{dx}(wh^2) = Ph'^2 + 2whh' + (Q+w')h^2 = P(h' + \frac{w}{P}h)^2$ . We now identify the expression in the parantheses as  $\phi(t)$  and conclude that the integral is positive semi-definite if  $P(t) > 0$ , and if  $P(t) > 0$ , the integral is zero only if  $h' + (w/P)h = 0$  for all  $t$  which requires that  $h(t)$  vanishes identically. Thus, if the differential equation for  $w(t)$  can be solved in the interval  $[t_0, t_1]$ , the positivity of the auxiliary integral is guaranteed if  $P(t) > 0$  for  $t \in [t_0, t_1]$ .

The equation for  $w(t)$  can be transformed to a simpler equation by setting  $w(t) = -\frac{u'(t)}{u(t)}P(t)$  (here we must have  $u(t) \neq 0$ ) which yields

$$-\frac{d}{dt}(P(t)u'(t)) + Q(t)u(t) = 0. \quad (22)$$

This is nothing but the Euler-Lagrange equation for the auxiliary minimization problem, equation (21). Now it is convenient to define a *conjugate point* so that a point  $z \neq t_0$  is a conjugate to the point  $t_0$  if the equation (21) has a solution that vanishes for  $t = t_0$  and  $t = z$  but is not identically zero. If there are no points  $z \in [t_0, t_1]$  that are conjugate to the point  $t_0$ , the solution of (21) is non-zero for all  $t \in [t_0, t_1]$ , the function  $w(t)$  is well defined, and the auxiliary integral is positive definite.

Thus, we have concluded that a sufficient condition for the second variation to be positive definite is that (i) the Euler-Lagrange equation is satisfied, (ii) the solution of the Euler-Lagrange equation satisfies  $P(t) = \frac{\partial^2 F}{\partial y'(t) \partial y'(t)} > 0$  for all  $t \in [t_0, t_1]$ , and (iii) there are no points conjugate to  $t_0$  in the interval  $[t_0, t_1]$ . It can be shown that these conditions are also necessary, that is, if the auxiliary integral is positive definite, then there cannot be a points conjugate to  $t_0$  in the interval  $[t_0, t_1]$ . The proof can be found *e.g.* in the book by Gelfond and Fomin, and we are not going to reproduce it here. Hence, we have the sufficient condition for a weak minimum:

*Suppose that the admissible curve  $y = y(t)$  satisfies*

1. *It is an extremal, that is, it satisfies the Euler-Lagrange equation.*

$$\frac{\partial F}{\partial y} - \frac{d}{dt} \frac{\partial F}{\partial y'} = 0.$$

2. *Along the curve  $y = y(t)$  the strengthened Legendre condition*

$$P(t) = \frac{1}{2} \frac{\partial^2 F}{\partial y' \partial y'} > 0$$

*is fulfilled*

3. *The interval  $[t_0, t_1]$  contains no points conjugate to  $t_0$  (the strengthened Jacobi condition).*

*Then the functional  $\int_{t_0}^{t_1} dt F(t, y, y')$  has a weak minimum for  $y = y(t)$ .*

Note that while we have a necessary condition (part one above), and a sufficient condition



(parts 1-3) above, we do not have a *necessary and sufficient* condition — there are weak minima that fail to satisfy the combination of points two and three above, analogous to the minimum of  $f(x) = x^4$  at  $x = 0$  which fails the second order test  $f''(x = 0) > 0$  but nevertheless is a minimum.

Most of the above arguments carry over to variational problems involving several functions  $\{y_i(t)\}$  as well. In particular, the condition  $P(t) > 0$  becomes a requirement that the corresponding matrix  $\frac{1}{2} \frac{\partial^2 F}{\partial y_i' \partial y_j'}$  must be positive semi-definite (may not have negative eigenvalues). The concept of a conjugate point is in multidimensional optimization problems related to the vanishing of a determinant composed of  $N$  different  $N$ -dimensional solutions of the Euler-Lagrange equations, etc..



- [1] More precisely: The integrand is an analytic function in the region that excludes the positive real axis and the zeroes of the denominator (which occur at  $z = i2\pi n$ ). Hence, the integration path can be modified (as long as we do not cross any of the singularities) without changing the value of the integral. In particular, it is not necessary for  $C$  to get very close to the origin. Since the origin is the only point that could give trouble, and we can avoid it regardless of  $s$ , the integral is well defined for all  $s$ . The pole at  $s = 1$  arises from the factors multiplying the contour integral.
- [2] If you peruse the above expressions with mathematical rigor, you will realize that the sin was committed when integration and summation were exchanged (the step indicated by  $\rightarrow$ ).
- [3] Actually, since  $n_{\text{opt}}$  diverges for  $x \rightarrow 0^+$ , at most 10 terms were included in the asymptotic expression. The neglected terms are so small that they do not affect the quality of the fit.
- [4] This section follows quite closely the discussion in the book by Bender and Orszag.