Math Methods

Shaojun Zan

2 April 2024

Contents

1 Fourier Transform

1.1 Preliminaries

Let $f(x)$ be a function defined for $x \in \mathbb{R}$. We define the **Fourier transform** $\hat{f}$ of f:

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

If the above integral exists for (almost) all $k \in \mathbb{R}$, then we may expand f as a **Fourier integral** by means of the **inverse Fourier transform** formula:

$$\mathcal{F}^{-1}\left[\hat{f}(k)\right] = f(x) = \int_{-\infty}^{\infty} \mathrm{d}k\, \hat{f}(k)\mathrm{e}^{ikx}. \tag{2}$$

Following is a list of properties of Fourier transforms:

- Linearity:

$$\mathcal{F}\left[\alpha f_1(x) + \beta f_2(x)\right] = \alpha\mathcal{F}\left[f_1(x)\right] + \beta\mathcal{F}\left[f_2(x)\right]. \tag{3}$$

- Scaling: for $\alpha \neq 0$,

$$\mathcal{F}\left[f(\alpha x)\right] = \frac{1}{|\alpha|}\hat{f}\left(\frac{k}{\alpha}\right). \tag{4}$$

  When $\alpha = -1$, we get the change of sign formula:

$$\mathcal{F}\left[f(-x)\right] = \hat{f}\left(-k\right). \tag{5}$$

- Fourier transforms preserve parity. It's the direct consequence of the change of sign formula. Furthermore, we can simplify the formulas when f has a definite parity. To see this, try

$$\begin{aligned}\int_{-\infty}^{\infty} \mathrm{d}x\, f(x)\mathrm{e}^{-ikx} &= \int_{0}^{\infty} \mathrm{d}x\left[f(x)\mathrm{e}^{-ikx} + f(-x)\mathrm{e}^{ikx}\right]\\ &\Rightarrow \begin{cases} f(x) \text{ is odd:} \quad \mathcal{F}\left[f(x)\right] = \dfrac{1}{2\pi i}\displaystyle\int_{-\infty}^{\infty} \mathrm{d}x\, f(x)\sin kx \\ f(x) \text{ is even:} \quad \mathcal{F}\left[f(x)\right] = \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{\infty} \mathrm{d}x\, f(x)\cos kx \end{cases}\end{aligned} \tag{6}$$

  This same approach applies to the inverse Fourier transform, but don't you mess up the minus signs in the process.

- Conjugation:

$$\mathcal{F}\left[f^*(x)\right] = \hat{f}^*(-k). \tag{7}$$

When $f(x) \in \mathbb{R}$, we get

$$\hat{f}(k) = \hat{f}^*(-k). \tag{8}$$

This implies f is real and even iff $\hat{f}$ is real and even. Similarly, f is real and odd iff $\hat{f}$ is purely imaginary and odd.

- Translation:

$$\mathcal{F}\left[f(x-a)\right] = \mathrm{e}^{-ika}\hat{f}(k). \tag{9}$$

Reversely, we write

$$\mathcal{F}\left[\mathrm{e}^{iax}f(x)\right] = \hat{f}(k-a). \tag{10}$$

- Derivative:

$$\mathcal{F}\left[\frac{\mathrm{d}f}{\mathrm{d}x}\right] = ik\hat{f}(k). \tag{11}$$

This applies to higher derivatives, where we mean

$$\mathcal{F}\left[\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^n f\right] = (ik)^n\hat{f}(k). \tag{12}$$

This property aids when we are dealing with differential equations with constant coefficients.

1.2 Absolutely Integrable Functions

To actually use the Fourier transforms, we need to answer 2 questions: for which functions $f(x)$ the integration formulas are properly defined, and how to compute them.

The second question is a shut-up-and-calculate thing, but the first one is a little more subtle. A natural requirement for the function is to be **absolutely integrable**, by which we mean

$$\int_{-\infty}^{\infty} \mathrm{d}x\, |f(x)| < \infty. \tag{13}$$

With this condition, we identify that f belongs to the space of absolutely integrable functions, denoted as L^1. For continuous functions, the condition $f \in L^1$ means, essentially, that $f \to 0$ sufficiently fast as $|x|$ grows (faster than $|x|^{-1}$, otherwise the integral would diverge).

Obviously, the condition $f \in L^1$ ensures that the Fourier transform integral converges for every k (as $\left|f(x)\mathrm{e}^{-ikx}\right| = |f(x)|$), so the Fourier transform $\hat{f}(k)$ is a well-defined function. In fact, we have a stronger statement:

Lemma

If $f \in L^1$, then $\hat{f}(k)$ is a **uniformly continuous function** of k.

By uniformly continuous, we mean that $f(x)$ is continuous, and the rate of this convergence depends on the distance between the two points only (but not on the position of the points, etc.)

Proof. We have

$$\left|\hat{f}(k_1) - \hat{f}(k_2)\right| = \frac{1}{2\pi}\left|\int_{-\infty}^{\infty} \mathrm{d}x\, f(x)\left(\mathrm{e}^{-ik_1x} - \mathrm{e}^{-ik_2x}\right)\right| \le \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}x\, |f(x)|\left|1 - \mathrm{e}^{-i(k_2-k_1)x}\right|.$$

What we need to do next is to show

$$\int_{-\infty}^{\infty} \mathrm{d}x\, |f(x)|\left|1 - \mathrm{e}^{-i(k_2-k_1)x}\right| \to 0 \quad \text{as} \quad k_2 - k_1 \to 0. \tag{14}$$

In order to estimate this integral, we note that $f \in L^1$, and $\left|1 - \mathrm{e}^{-i(k_2-k_1)x}\right| \le 2$ is uniformly bounded for all x, k_1, and k_2. Therefore, the contribution from large x is uniformly small, and we can approximate by an integral over a sufficiently large interval.

Now, on any fixed interval of integration, $\mathrm{e}^{-i(k_2-k_1)x} \to 1$ as $k_2 - k_1 \to 0$, uniformly with respect to x. Thus, the factor $\left|1 - \mathrm{e}^{-i(k_2-k_1)x}\right|$ goes uniformly to 0, and the integral over this interval becomes as small as we need when $k_2 - k_1$ gets small enough. □

This lemma illustrates the general principle: *the large scale behavior of the function translates into the small scale behavior of its Fourier transform.* Here, the large scale feature of f is the absolute integrability, and the small scale feature is the continuity of $\hat{f}$.

Surprisingly, it turns out that the Fourier transform and its inverse formulas we've been using throughout these years are based on a theorem. As physicists, we've acquiesced that the conditions are satisfied even if they aren't. However, "it's time for you to get real and sort out who you really are".

Theorem

If $f \in L^1$, then we have

$$\hat{f}(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}x\, f(x)\mathrm{e}^{-ikx}. \tag{15}$$

When we also have $\hat{f} \in L^1$, then we get

$$f(x) = \int_{-\infty}^{\infty} \mathrm{d}k\, \hat{f}(k)\mathrm{e}^{ikx}. \tag{16}$$

Proof. What follows would be vertiginous yet pivotal. What we want to prove here is

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{ikx} \int_{-\infty}^{\infty} \mathrm{d}y\, \mathrm{e}^{-iky} f(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}y\, f(y) \left[\int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{ik(x-y)}\right]. \tag{17}$$

While it would be so easy to recognize the formula in square brackets as a Dirac delta, we must pretend that we have no information about it and prove it as pure mathematical pedants.

Observe the trick here. As $\hat{f} \in L^1$, we have

$$\int_{-\infty}^{\infty} \mathrm{d}k\, \hat{f}(k)\mathrm{e}^{ikx} = \lim_{\epsilon\to 0}\int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{-(\epsilon k)^2} \hat{f}(k)\mathrm{e}^{ikx}. \tag{18}$$

Indeed, the factor we added here is bounded, so the integral is uniformly absolutely convergent for all ϵ. What we need to prove becomes

$$f(x) = \lim_{\epsilon\to 0}\int_{-\infty}^{\infty} \mathrm{d}y\, f(y) \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{-(\epsilon k)^2} \mathrm{e}^{ik(x-y)}\right].$$

We identify that as a Fourier transform of a Gaussian. Therefore,

$$\lim_{\epsilon\to 0}\int_{-\infty}^{\infty} \mathrm{d}y\, f(y) \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{-(\epsilon k)^2} \mathrm{e}^{ik(x-y)}\right] = \lim_{\epsilon\to 0}\frac{1}{2\epsilon\sqrt{\pi}}\int_{-\infty}^{\infty} \mathrm{d}y\, f(y) \exp\left[-\frac{(x-y)^2}{4\epsilon^2}\right].$$

We can split the integral into three parts. Take any $\delta > 0$, when $|y-x| > \delta$,

$$\left|\left(\int_{-\infty}^{x-\delta} + \int_{x+\delta}^{\infty}\right) \mathrm{d}y\, f(y) \exp\left[-\frac{(x-y)^2}{4\epsilon^2}\right]\right| \le \exp\left(-\frac{\delta^2}{4\epsilon^2}\right)\int_{-\infty}^{\infty} \mathrm{d}y\, |f(y)| \ll \epsilon.$$

This implies an approximation:

$$\begin{aligned}
\lim_{\epsilon\to 0}\frac{1}{2\epsilon\sqrt{\pi}}\int_{-\infty}^{\infty} \mathrm{d}y\, f(y) \exp\left[-\frac{(x-y)^2}{4\epsilon^2}\right] &\approx \lim_{\epsilon\to 0}\frac{1}{2\epsilon\sqrt{\pi}}\int_{x-\delta}^{x+\delta} \mathrm{d}y\, f(y) \exp\left[-\frac{(x-y)^2}{4\epsilon^2}\right] \\
&\approx f(x)\lim_{\epsilon\to 0}\frac{1}{2\epsilon\sqrt{\pi}}\int_{x-\delta}^{x+\delta} \mathrm{d}y \exp\left[-\frac{(x-y)^2}{4\epsilon^2}\right] \\
&= f(x).
\end{aligned}$$

□

The theorem establishes the relation between the function and its Fourier coefficients for a sufficiently large class of functions. However, it can still be restrictive. For example, the theorem requires that both f and $\hat{f}$ are continuous. In reality, many functions are not continuous, and we should discuss how the Fourier transform theory is extended to other classes of functions.

1.3 Smoothness & Decay Rate

By our theorem in the last session, if we have $f \in L^1$ and $\hat{f} \in L^1$, then

$$f(x) = \int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{ikx} \hat{f}(k). \tag{19}$$

We can differentiate this integral with respect to x as long as the resulting integral (absolutely) converges. This means that we have

$$f'(x) = \int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{ikx} ik\hat{f}(k) \quad \text{if} \quad \int_{-\infty}^{\infty} \mathrm{d}k\, |k| \left|\hat{f}(k)\right| < \infty. \tag{20}$$

Sometimes, the result holds even without this condition, but this condition should be identified as sufficient. A sufficient condition for the existence and continuity of the derivative $f'(x)$ is that the Fourier coefficients decay to 0 sufficiently fast:

$$\left|\hat{f}(k)\right| = \mathcal{O}\left(\frac{1}{|k|^{2+\delta}}\right) \quad \text{for some} \quad \delta > 0 \tag{21}$$

as $|k| \to \infty$. If we introduce o for higher order terms (different from $\mathcal{O}$), we have

$$\left|k\hat{f}(k)\right| \to 0 \quad \text{as} \quad |k| \to \infty \quad \Rightarrow \quad \left|\hat{f}(k)\right| = o\left(\frac{1}{|k|}\right). \tag{22}$$

More generally, if

$$\int_{-\infty}^{\infty} \mathrm{d}k\, |k|^n \left|\hat{f}(k)\right| < \infty, \tag{23}$$

then f has n continuous derivatives, and

$$\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^n f(x) = \int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{ikx} (ik)^n \hat{f}(k). \tag{24}$$

In particular, this holds if

$$\left|\hat{f}(k)\right| = \mathcal{O}\left(\frac{1}{|k|^{1+n+\delta}}\right) \quad \text{or} \quad \left|\hat{f}(k)\right| = o\left(\frac{1}{|k|^n}\right) \tag{25}$$

as $|k| \to \infty$, for some $\delta > 0$. The same argument apples to the formula of $\hat{f}(k)$:

$$\int_{-\infty}^{\infty} \mathrm{d}x\, |x|^n |f(x)| < \infty \quad \Rightarrow \quad \hat{f}^{(n)}(k) \text{ exists and is continuous.} \tag{26}$$

We've made all our arguments regarding decay properties and how they imply about differentiabilities. Now we want to do the converse.

Theorem (Riemann-Lebesgue)

If $f \in L^1$, then $\hat{f}(k) \to 0$ as $|k| \to \infty$.

Proof. We know that $f \in L^1$, and that means for sufficiently large R the tails of the integral is negligible. Therefore,

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

Then, we can split the integral into intervals of length $2\pi/k$. As $N \to \infty$,

$$\begin{aligned} \frac{1}{2\pi} \int_{-R}^{R} \mathrm{d}x\, f(x) \mathrm{e}^{-ikx} &= \frac{1}{2\pi} \sum_{j=0}^{N-1} \int_{x_j}^{x_{j+1}} \mathrm{d}x\, f(x) \mathrm{e}^{-ikx} \\ &= \frac{1}{2\pi} \sum_{j=0}^{N-1} f(x_j) \int_{x_j}^{x_{j+1}} \mathrm{d}x\, \mathrm{e}^{-ikx} \\ &= 0. \end{aligned}$$

Thus, 0 is a good approximation to the integral above, and it gets better when the length of the intervals tends to 0, i.e. as $|k| \to \infty$. □

1.4 Dirac Delta Function

The idea of a Dirac delta function comes from a density over a point mass or a point charge, or anything else localized at a single point:

$$\delta(x) = \begin{cases} 0 & \text{if } x \neq 0, \\ \infty & \text{if } x = 0. \end{cases} \tag{27}$$

This is not specific enough, so we impose a normalization condition:

$$\int_{-\infty}^{\infty} \mathrm{d}x\, \delta(x) = 1. \tag{28}$$

We could approximate a delta function as a Gaussian, a top-hat function, or even a sinc function. Proceeding with the top-hat function, we get

$$\int_{-\infty}^{\infty} \mathrm{d}x\, f(x) \delta(x-a) = f(a). \tag{29}$$

This gives us an equivalent definition of the delta function: it is a linear operator which, for every continuous function f, returns a number $f(0)$.

The Dirac delta function has other properties that make it extra important:

- For any $\lambda \neq 0$,

$$\delta(\lambda x) = \frac{1}{|\lambda|}\delta(x). \tag{30}$$

- Let ϕ be an arbitrary smooth function with several simple roots. Then, we have

$$\delta[\phi(x)] = \sum_k \frac{1}{|\phi'(x_k)|}\delta(x - x_k). \tag{31}$$

Proof. Let's investigate the results of the integral

$$I = \int_{-\infty}^{\infty} \mathrm{d}x\, \delta[\phi(x)] f(x).$$

While we have no information about the behavior of $\delta[\phi(x)]$, we know $\delta(x)$. This alludes to the use of variable substitution. If we have $y = \phi(x)$, then $\mathrm{d}y = \phi'(x)\,\mathrm{d}x$, and

$$\begin{aligned} I &= \int_{x=-\infty}^{x=\infty} \frac{\mathrm{d}y}{\phi'(x)}\delta(y) f(x) \\ &= \int_{x=-\infty}^{x=\infty} \frac{\mathrm{d}y}{\phi'[\phi^{-1}(y)]}\delta(y) f[\phi^{-1}(y)] \\ &= \sum_{y_k=0} \frac{1}{|\phi'[\phi^{-1}(y_k)]|} f[\phi^{-1}(y_k)] \\ &= \sum_k \frac{1}{|\phi'(x_k)|} f(x_k). \end{aligned}$$

The choice of $f(x)$ is arbitrary, so what $\delta[\phi(x)]$ does to a function f is to return the values at the simple zeros of $\phi(x)$ with corresponding amplitudes:

$$\delta[\phi(x)] = \sum_k \frac{1}{|\phi'(x_k)|}\delta(x - x_k).$$

□

- Derivatives of the delta function:

By the technique of integration by parts, for any n-times continuously differentiable function f,

$$\int_{-\infty}^{\infty} \mathrm{d}x\, f(x)\delta^{(n)}(x) = (-1)^n f^{(n)}(0). \tag{32}$$

- Anti-derivatives of the delta function:

  The **heavyside function** follows immediately from the integral of the Dirac delta:

$$\int_{-\infty}^{x} \mathrm{d}y\, \delta(y-a) = \theta(x-a) = \begin{cases} 1 & \text{if } x > a, \\ 0 & \text{if } x < a. \end{cases} \tag{33}$$

- Dirac delta as a Fourier integral:

  From the Gaussian approximation,

$$\delta(x) \approx \frac{1}{\epsilon\sqrt{\pi}} \exp\left(-\frac{x^2}{\epsilon^2}\right), \tag{34}$$

  we know, as $\epsilon \to 0$,

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{ikx}. \tag{35}$$

  We can identify $\delta(x)$ as the Fourier transform of a constant.

1.5 Beyond Absolute Integrability

While the class of functions $f \in L^1$ encapsulates a lot, we are still interested in Fourier transform and Fourier integral for functions which do not decay fast as infinity and, as a result, are not absolutely integrable. One of the approaches here is to consider the following generalization of the Fourier transform and Fourier integral:

$$\begin{aligned} \hat{f}(k) &= \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{-ikx} f(x) \mathrm{e}^{-(\epsilon x)^2} \\ f(x) &= \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{ikx} \hat{f}(k) \mathrm{e}^{-(\epsilon k)^2}. \end{aligned} \tag{36}$$

These integrals exist for any finite ϵ for a very large class of functions. An important class of functions for which this definition works is the class L^2 of the **square integrable functions**. We say $f \in L^2$ if

$$\int_{-\infty}^{\infty} \mathrm{d}x\, |f(x)|^2 < \infty. \tag{37}$$

This is a weaker condition that $f \in L^1$. For example, $f(x) = 1/x \in L^2$ but not in L^1.

Theorem

If $f \in L^2$, then

$$\hat{f}(k) = \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{-ikx} f(x) \mathrm{e}^{-(\epsilon x)^2} \tag{38}$$

exists and is well-defined.

We will not prove this, but one should be able to prove the **Parseval's identity**:

$$\int_{-\infty}^{\infty} \mathrm{d}k\, \hat{f}(k)\hat{g}(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}x\, f(x)g(x), \tag{39}$$

where the stinky factor of $1/2\pi$ comes from our particular definition of Fourier transform and integral.

Roughly speaking, L^1 functions decay faster then $1/|x|$ as $|x| \to \infty$, while L^2 functions decay faster than $1/\sqrt{|x|}$. However, one can extend the theory even to functions which do not decay to zero at all. We will not pursue this in full generality, but we should meet some examples.

- $f(x) = 1$:

$$\hat{f}(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{-ikx} = \delta(k). \tag{40}$$

- $f(x) = \operatorname{sgn}(x)$:

  Noting that this function is real and odd, we know the function $\hat{f}$ should be purely imaginary and odd. It's hard to derive from scratch, so we begin by "noting" that $\hat{f}(k) = 1/k$ may satisfy the Fourier integral:

$$\begin{aligned} f(x) &= i\int_{-\infty}^{\infty} \mathrm{d}k\, \frac{\sin kx}{k} \\ &= i\,\mathrm{Im}\left\{\oint_C \mathrm{d}k\, \frac{\mathrm{e}^{ikx}}{k}\right\} \\ &= i\,\mathrm{Im}\left\{2\pi i\cdot \operatorname{sgn}(x)\frac{1}{2}\, \mathrm{e}^{ikx}\Big|_0\right\} \\ &= i\pi \operatorname{sgn}(x). \end{aligned}$$

  Because of the singularity of $1/k$ at $k = 0$, we cannot immediately claim that we've found the Fourier transform of $\operatorname{sgn}(x)$. Nevertheless, this expression does have its principal value convergent. Therefore, we say

$$\mathcal{F}\left[\operatorname{sgn}(x)\right] = \mathcal{P}\left(\frac{1}{k}\right)\cdot\frac{1}{i\pi} \tag{41}$$

  to stress the notion that whenever we take an integral including $\hat{f}$, we should take only the principal value of it. Namely, we can think of $\mathcal{P}\left(1/k\right)$ as a linear operator which, applied to a continuous function $h(k)$, returns the value of

$$\int_0^{\infty} \mathrm{d}k\, \frac{h(k) - h(-k)}{k}, \tag{42}$$

  if this integral converges.

- $f(x) = \theta(x)$:

$$\mathcal{F}\left[\theta(x)\right] = \mathcal{F}\left[\frac{1}{2}[1 + \operatorname{sgn}(x)]\right] = \frac{1}{2}\delta(k) + \mathcal{P}\left(\frac{1}{2\pi i k}\right). \tag{43}$$

We also know that $\theta'(x) = \delta(x)$, and this gives us

$$ik\hat{\theta}(k) = \delta(k) = \frac{1}{2\pi}.$$

Obviously, it is then tempting to write $\hat{\theta}(k) = \dfrac{1}{2\pi i k}$. However, we cannot simply multiply by $1/ik$, as it is not defined at $k = 0$. The message here is that we must be careful when dealing with functions which do not decay to zero and operators such as the delta function.

Behavior of $x\delta(x)$

We know that $x\delta(x) = 0\ \forall x$, so we say

$$A(x) = B(x) = B(x) + cx\delta(x) \tag{44}$$

for any finite c. However, if we divide by x, we must recognize that

$$\frac{A(x)}{x} = \frac{B(x)}{x} + c\delta(x) \tag{45}$$

is not necessarily true for arbitrary values of c.

As an example, we can see that

$$x\frac{\mathrm{d}}{\mathrm{d}x}\ln x = 1 = 1 + cx\delta(x)$$

is true for any choice of c, but

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln x = \frac{1}{x} + c\delta(x)$$

is true only for some special choice of c. To work out the value of c, integrate the above from $-\epsilon$ to ϵ for some small ϵ:

$$\int_{-\epsilon}^{\epsilon} \mathrm{d}\ln x = \int_{-\epsilon}^{\epsilon} \mathrm{d}x \left[\frac{1}{x} + c\delta(x)\right].$$

The integral of $1/x$ vanishes, as $1/x$ is an odd function of x (so wild). Therefore,

$$c = \ln(-1) = i(2n+1)\pi.$$

- Let $g \in L^1$, we want to find the Fourier transform of

$$f(x) = \int_{-\infty}^{x} \mathrm{d}y\, f(y). \tag{46}$$

If we were to define $I \equiv \int_{-\infty}^{\infty} \mathrm{d}x\, g(x) = 2\pi\hat{g}(0)$, we have $I = \lim_{x\to\infty} f(x)$. Therefore,

$$f(x) = I\theta(x) + h(x)$$

for some h such that $h(x) \to 0$ as $x \to 0$. The Fourier transform and the derivative give us

$$h'(x) = g(x) - I\delta(x) \quad \Rightarrow \quad \hat{h}(k) = \frac{\hat{g}(k) - \hat{g}(0)}{ik}.$$

Therefore,

$$\begin{aligned}\hat{f}(k) &= 2\pi\hat{g}(0)\left[\frac{1}{2}\delta(k) + \mathcal{P}\left(\frac{1}{2\pi ik}\right)\right] + \frac{\hat{g}(k) - \hat{g}(0)}{ik} \\ &= \pi\hat{g}(0)\delta(k) + \mathcal{P}\left[\frac{\hat{g}(k)}{ik}\right].\end{aligned} \tag{47}$$

1.6 Green's Function

Let us define the **convolution** of two functions f and g:

$$(f * g)(x) \equiv \int_{-\infty}^{\infty} \mathrm{d}y\, f(y)g(x-y) = \int_{-\infty}^{\infty} \mathrm{d}y\, f(x-y)g(y) \tag{48}$$

The importance of Fourier transform lies in the fact that the convoluted convolution becomes a simple product in the k space:

$$\mathcal{F}\left[\frac{1}{2\pi}(f * g)(x)\right] = \hat{f}(k)\hat{g}(k). \tag{49}$$

The Green's function is defined as a particular solution to a differential equation, where the free term happens to be a Dirac delta:

$$\hat{L}y(x) = f(x) \quad \Rightarrow \quad \hat{L}G(x) = \delta(x). \tag{50}$$

By some simple algebra, we can prove

$$y(x) = (G * f)(x). \tag{51}$$

Now, let's return to the linear ODE's with constant coefficients:

$$\hat{L} \equiv \sum_{n}^{N} a_n \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^n. \tag{52}$$

By the property of Fourier transforms where $\mathcal{F}\left[y^{(n)}\right] = (ik)^n \hat{y}$, we know

$$L(ik)\hat{G}(k) = \frac{1}{2\pi} \quad \Rightarrow \quad \hat{G}(k) = \frac{1}{2\pi}\frac{1}{L(ik)}. \tag{53}$$

However, the above manipulation is not well defined at $k = 0$, and this implies a notation of principal value:

$$\hat{G}(k) = \frac{1}{2\pi}\mathcal{P}\left[\frac{1}{L(ik)}\right], \tag{54}$$

where we allow $L(\lambda)$ to have simple purely imaginary roots.

From $L(ik)\hat{y}(k) = \hat{f}(k)$, we also know

$$\hat{y}(k) = \frac{\hat{f}(k)}{L(ik)} = 2\pi\hat{G}(k)\hat{f}(k). \tag{55}$$

In fact, let's investigate the simple case of $y' = f(x)$. Fourier transform gives us

$$ik\hat{y}(k) = \hat{f}(k) + ck\delta(k),$$

where the second term is uniformly zero for all k. Hence,

$$\hat{y}(k) = \mathcal{P}\left[\frac{\hat{f}(k)}{ik}\right] + C\delta(k), \tag{56}$$

where the first term is the partial solution, and the second term is the general solution of system under homogeneous conditions.

If we consider $y'' = f(x)$, note that the use of Fourier transforms becomes questionable, as the RHS of $\hat{y} = -\hat{f}/k^2$ is not integrable in a principal value sense. However, we can still solve it through the Green's function alone: with $B = C = 0$,

$$G(x) = Cx + B + x\theta(x) \quad \Rightarrow \quad y(x) = \int_0^\infty \mathrm{d}s\, s f(x-s) + Dx + E.$$

2 Lagrangian Mechanics

2.1 Euler-Lagrange Equation

We start with the following optimization problem: given a function $\mathcal{L}[y(x), y'(x), x]$, we want to know which function $y(x)$ will give a minimal or maximal value to the functional

$$\int_a^b \mathrm{d}x\, \mathcal{L}(y, y', x) \tag{57}$$

subject to the condition that the values of y at the endpoints of the integration interval $[a, b]$ are fixed. It was discovered by Euler and Lagrange that this problem is reduced to the solution of a certain second-order differential equation.

$$\begin{tikzcd}
 & y(x) \arrow[dr] & \\
\mathcal{L}(y, y', x) \arrow[ur] \arrow[rr] \arrow[dr] & & x \\
 & y'(x) \arrow[ur] &
\end{tikzcd}$$

Theorem

If a twice continuously differentiable function $y(x)$ is a minimizer (or maximizer) of

$$\int_a^b \mathrm{d}x\, \mathcal{L}(y, y', x),$$

then it must satisfy the equation

$$\frac{\partial \mathcal{L}}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial \mathcal{L}}{\partial y'} = 0. \tag{58}$$

Proof. If the function $y(x)$ really optimizes the functional, then any perturbation function of the function

$$F(\epsilon) = \int_a^b \mathrm{d}x\, \mathcal{L}\left(y + \epsilon\phi, y' + \epsilon\phi', x\right)$$

has an extremum at $\epsilon = 0$. This also means that $F'(\epsilon) = 0$ at $\epsilon = 0$. Therefore,

$$\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}F(\epsilon)\right|_{\epsilon=0} = \int_a^b \mathrm{d}x\, \phi(x)\left(\frac{\partial \mathcal{L}}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial \mathcal{L}}{\partial y'}\right) = 0$$

for any choice of $\phi(x)$ with $\phi(a) = \phi(b) = 0$. $\square$

One should pay attention that $\mathrm{d}/\mathrm{d}x$ is a total derivative, and this means that $\mathrm{d}\mathcal{L}/\mathrm{d}x \neq 0$ even when $\mathcal{L}$

doesn't involve x explicitly. The following theorem considers this situation and defines a new function called the Hamiltonian function.

Theorem

If the Lagrangian $\mathcal{L}(y, y')$ doesn't depend on x explicitly, then the **Hamiltonian function** (or energy function)

$$\mathcal{H}(y, y') \equiv y' \frac{\partial \mathcal{L}}{\partial y'} - \mathcal{L} \tag{59}$$

stays constant on any solution of the Euler-Lagrange equation.

Proof. To show that $\mathcal{H}$ is a constant, just take the total derivative with respect to x:

$$\frac{\mathrm{d}}{\mathrm{d}x} \mathcal{H}\left[y(x), y'(x)\right] = -\left(\frac{\partial \mathcal{L}}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial \mathcal{L}}{\partial y'}\right) y' = 0.$$

$\square$

The Hamiltonian function simplifies a lot of tedious calculations involved in the Lagrangian. If we were to extend all our definitions to a functional that involves n functions,

$$\int_a^b \mathrm{d}x \, \mathcal{L}(y_i, y_i', x), \tag{60}$$

we would get the result that

$$\frac{\partial \mathcal{L}}{\partial y_j} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial \mathcal{L}}{\partial y_j'} = 0, \qquad \forall\, j = 1, \ldots, n. \tag{61}$$

Similarly, if $\mathcal{L}$ doesn't involve x explicitly, we can find the Hamiltonian function

$$\mathcal{H} \equiv \left(\sum_j \frac{\partial \mathcal{L}}{\partial y_j'} y_j'\right) - \mathcal{L}. \tag{62}$$

2.2 Optimization with Constraints

Now, we consider the same problem as before, but this time with an additional constraint:

$$\int_a^b \mathrm{d}x\, G(y, y', x) = C, \tag{63}$$

where G is a given function and C is a given constant.

To solve this problem, we use the method of **Lagrange multipliers**:

Theorem

If y minimizes or maximizes the integral

$$\int_a^b \mathrm{d}x\, \mathcal{L}(y, y', x)$$

in the class of functions that satisfy the integral constraint

$$\int_a^b \mathrm{d}x\, G(y, y', x) = C,$$

then it must be a solution to the Euler-Lagrange equation for the unconstrained optimization problem with the Lagrangian

$$\mathcal{L} + \lambda G \tag{64}$$

with some constant λ (the Lagrange multiplier).

Proof. Similarly, we consider the perturbations $\epsilon\phi(x, \epsilon)$ to the optimized function y. At $\epsilon = 0$, we have $\phi_0(x) = \phi(x, 0)$ and

$$\int_a^b \mathrm{d}x\, \phi_0(x) \left(\frac{\partial \mathcal{L}}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial \mathcal{L}}{\partial y'} \right) = 0$$
$$\int_a^b \mathrm{d}x\, \phi_0(x) \left(\frac{\partial G}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial G}{\partial y'} \right) = 0.$$

Let $\phi_0(x) = \alpha\delta(x - s) + \beta\delta(x - q)$ for $s, q \in (a, b)$. If we fix some value of q and denote the corresponding expressions as the constant λ, we get

$$\frac{\partial \mathcal{L}}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial \mathcal{L}}{\partial y'} \bigg|_{x=s} + \lambda\, \frac{\partial G}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial G}{\partial y'} \bigg|_{x=s} = 0 \tag{65}$$

for any $s \in (a, b)$. This is the Euler-Lagrange equation for the constraint Lagrangian $\mathcal{L} + \lambda G$. $\square$

We can also consider extensions to the integral constraints.

- Multiple constraints:

  If we have multiple constraints on $\mathcal{L}$, like

$$\int_a^b \mathrm{d}x\, G_i(y, y', x) = C_i, \tag{66}$$

  then we identify the constraint Lagrangian as

$$\mathcal{L} + \underline{\lambda} \cdot \mathbf{G} \equiv \mathcal{L} + \sum_i \lambda_i G_i. \tag{67}$$

- Even more constraints:

  If we have a rule which $y(x)$ must satisfy for all x, like

$$G(y, y', x) = C, \qquad \text{for all } x \in (a, b), \tag{68}$$

  we can regard this as a partial case of an integral constraint:

$$\int_a^b \mathrm{d}x\, G(y, y', x)\delta(X - x) = C. \tag{69}$$

  Since the above must be true for all $X \in (a, b)$, we can think of infinitely many constraints. From the situation of multiple constraints and take the number of constraints to infinity, we get the constrained Lagrangian

$$\mathcal{L} + \int_a^b \mathrm{d}X\, \lambda(X) G(y, y', x)\delta(X - x) = \mathcal{L} + \lambda(x) G(y, y', x), \tag{70}$$

  where $\lambda(x)$ now becomes a function of x.

2.3 Geodesic Lines

In a general curved space, the distance between two points is

$$d(y, y + \Delta y) = \sqrt{\sum_{i,j} g_{ij} \Delta y_i \Delta y_j} + \mathcal{O}(\Delta y^2), \tag{71}$$

where the symmetric matrix g is called the **metric tensor**, or simply metric.

If we parameterize a curve in space with t, $0 \le t \le 1$,

$$S = \int_S \mathrm{d}s = \int_0^1 \mathrm{d}t \, \frac{\mathrm{d}s}{\mathrm{d}t} = \int_0^1 \mathrm{d}t \, \sqrt{\sum_{i,j} g_{ij} y'_i y'_j}.$$

Given any two points A and B, we can look for the curve of minimal length that connects them (a **geodesic line**). This implies the optimization problem

$$\mathcal{L}(\mathbf{y}, \mathbf{y}') = \sqrt{\sum_{i,j} g_{ij}(\mathbf{y}) y'_i y'_j} \tag{72}$$

with boundary conditions $y(0) = A$ and $y(1) = B$.

$\mathcal{L}$ can be interpreted as the speed with which we move along the geodesics, as it is $\mathrm{d}s/\mathrm{d}t$ by derivation. Therefore, different parameterizations are possible, and they each represent a different speed that we move along the curve. This also means that we can always find a parameter such that the speed of motion along the geodesic is constant. Therefore,

$$\mathcal{L} \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{1}{\mathcal{L}} \right) \sum_i g_{ik} y'_i + \sum_i \frac{\mathrm{d}}{\mathrm{d}s} g_{ik} y'_i = \frac{1}{2} \sum_{i,j} \frac{\partial g_{ij}}{\partial y_k} y'_i y'_j \quad \Rightarrow \quad \hat{\mathcal{L}} = \frac{1}{2} \mathcal{L}^2. \tag{73}$$

We identify the system Lagrangian as the kinetic energy (also, $\mathcal{H} = \hat{\mathcal{L}}$). **Free particles move, with a constant speed, along the geodesic lines.**

2.4 Lagrangian Mechanics

From Newton's 2nd law of motion, $m\ddot{\mathbf{x}} = \mathbf{F}$, if we define $\mathcal{L} = m\dot{\mathbf{x}}^2/2$, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}} - \frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \mathbf{F}. \tag{74}$$

This equation keeps its shape regardless of the frame of reference, and it is, therefore, easier to move between different coordinate systems.

2.4.1 Eradication of Inertial Forces in Lagrangian Formalism

No inertial forces appear in the Lagrangian formulation. To show this, let's consider the effect of a coordinate transformation. Suppose that $\mathbf{y}$ obeys

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{y}}} - \frac{\partial \mathcal{L}}{\partial \mathbf{y}} = \mathbf{F}.$$

Now let's transform to new coordinates $\mathbf{x}$:

$$y_i = \phi_i(\mathbf{x}). \tag{75}$$

Therefore,

$$\mathcal{L}(\mathbf{y}, \dot{\mathbf{y}}) \equiv \mathcal{L}\left[\phi_i(\mathbf{x}), \frac{\partial \phi_i}{\partial \mathbf{x}} \cdot \dot{\mathbf{x}}\right], \tag{76}$$

and so we can find

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}} - \frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \Phi\mathbf{F}, \tag{77}$$

where the matrix $\Phi_{ij} = \partial y_j/\partial x_i$. A direct corollary from here is that if the system is isolated, then the form of the equations remains the same after any coordinate transformations.

2.4.2 Conservative Forces

Now consider the particular situation

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{y}}} - \frac{\partial \mathcal{L}}{\partial \mathbf{y}} = \mathbf{F} = -\boldsymbol{\nabla} U(\mathbf{y}).$$

In this case, we can modify our Lagrangian as follows:

$$\hat{\mathcal{L}} = \mathcal{L}(\mathbf{y}, \dot{\mathbf{y}}) - U(\mathbf{y}) \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \hat{\mathcal{L}}}{\partial \dot{\mathbf{y}}} - \frac{\partial \hat{\mathcal{L}}}{\partial \mathbf{y}} = 0. \tag{78}$$

2.5 Least Action Principle

Every isolated physical system is described by the Euler-Lagrange equations. As we know

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{y}}} - \frac{\partial \mathcal{L}}{\partial \mathbf{y}} = 0, \tag{79}$$

this gives a necessary condition for the trajectory $\mathbf{y}(t)$ to minimize the so-called **action functional**:

$$\int \mathrm{d}t\, \mathcal{L}(\mathbf{y}, \dot{\mathbf{y}}). \tag{80}$$

Therefore, the general principle above is also called the least action principle.

Let's meet some simple isolated systems.

2.5.1 Particle on a Line

If Newton's law reads

$$m\ddot{y} = f(y), \tag{81}$$

this equation coincides with the Euler-Lagrange equation with

$$\mathcal{L} = \frac{m\dot{y}^2}{2} + \int \mathrm{d}y\, f(y) = T - U. \tag{82}$$

2.5.2 Central Forces

If we have a force

$$\mathbf{F}(\mathbf{r}) = \hat{\mathbf{r}} g(r), \tag{83}$$

we identify the corresponding potential

$$U(\mathbf{r}) = -\int \mathrm{d}r\, g(r) \quad \text{such that} \quad \mathbf{F} = -\boldsymbol{\nabla} U. \tag{84}$$

Therefore, the system Lagrangian is

$$\mathcal{L} = \frac{m\dot{\mathbf{y}}^2}{2} + \int \mathrm{d}r\, g(r), \tag{85}$$

and the conserved energy is

$$\mathcal{H} = \frac{m\dot{\mathbf{y}}^2}{2} - \int \mathrm{d}r\, g(r). \tag{86}$$

Example: Particle in a Plane

Let's consider a particle in a plane. With plane polar coordinates, we can write

$$\mathcal{L} = \frac{m\dot{r}^2}{2} + \frac{mr^2\dot{\phi}^2}{2} - U(r). \tag{87}$$

By the Euler-Lagrange equation in ϕ and assuming rotational symmetry, we immediately get

$$\frac{\partial \mathcal{L}}{\partial \dot{\phi}} = mr^2\dot{\phi} = K = \text{const.} \tag{88}$$

This is the law of **conservation of angular momentum**, which is a direct consequence from the rotational symmetry.

The Euler-Lagrange equation in r gives us the equation of evolution of r:

$$m\ddot{r} = \frac{K^2}{mr^3} - U'(r). \tag{89}$$

Also, the conserved energy now reads

$$\mathcal{H} = \frac{m\dot{r}^2}{2} + \frac{K^2}{2mr^2} + U(r) = \text{const.} \tag{90}$$

We can solve for r from this equation of conservation of energy, and solve for ϕ from the angular momentum. The equations of motion for the particle in a central potential have enough symmetries to be completely solvable.

2.5.3 Two Interacting Particles

We consider the situation where the forces point to the position of the particles, like

$$\begin{aligned} \mathbf{F}_z &= (\mathbf{z} - \mathbf{y})g(\|\mathbf{z} - \mathbf{y}\|) \\ \mathbf{F}_y &= (\mathbf{y} - \mathbf{z})g(\|\mathbf{z} - \mathbf{y}\|). \end{aligned} \tag{91}$$

This is similar to the case of the central forces, if we define the distance r as the relative distance $r = \|\mathbf{z} - \mathbf{y}\|$. Therefore, the system Lagrangian can be worked out to be

$$\mathcal{L} = \frac{m_y\dot{\mathbf{y}}^2}{2} + \frac{m_z\dot{\mathbf{z}}^2}{2} - U(\|\mathbf{z} - \mathbf{y}\|), \tag{92}$$

so the system can be described by the Euler-Lagrange equations with $\mathbf{y}$ and $\mathbf{z}$.

This can be readily generalized to the situation with many particles, where we have

$$\mathcal{L} = \sum_j \frac{m_j \dot{\mathbf{r}}_j^2}{2} - \sum_{i,j} U_{ij}(\|\mathbf{r}_i - \mathbf{r}_j\|). \tag{93}$$

In this situation, the system can be described by the Euler-Lagrange equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}_j} - \frac{\partial \mathcal{L}}{\partial \mathbf{r}_j} = 0. \tag{94}$$

The law of conservation of linear momentum slips out once we find a correct cyclic coordinate.

2.6 Hamiltonian Systems

Let $\mathcal{L}(\mathbf{y}, \dot{\mathbf{y}})$ be the Lagrangian for some system. This system has conserved energy:

$$\mathcal{H} = \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{y}}} \cdot \dot{\mathbf{y}} - \mathcal{L}. \tag{95}$$

From the above, we define the **conjugate momentum** as

$$p_j = \frac{\partial \mathcal{L}}{\partial \dot{y}_j}, \qquad \mathbf{p} = \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{y}}}. \tag{96}$$

As an example, for a particle on a line, $\mathcal{L} = \dfrac{m\dot{y}^2}{2} + \displaystyle\int \mathrm{d}y\, f(y)$, so

$$p = \frac{\partial \mathcal{L}}{\partial \dot{y}} = m\dot{y}$$

is just the usual momentum. In general, from $\mathcal{L}(\mathbf{y}, \dot{\mathbf{y}})$ we identify $\mathbf{p}$ as functions of $\mathbf{y}$ and $\dot{\mathbf{y}}$. In principle, we can solve for $\dot{\mathbf{y}}$ with fixed $\mathbf{p}$, and this implies that we may use $\mathbf{y}$ and $\mathbf{p}$ as coordinates instead of $\mathbf{y}$ and $\dot{\mathbf{y}}$.

The energy considered as a function of $\mathbf{y}$ and $\mathbf{p}$ is called the **Hamilton function**, or the **Hamiltonian**. Like the Lagrangian, the Hamiltonian encodes all the information about the physics of the system.

Therefore, we get the Hamilton function:

$$\mathcal{H}(\mathbf{y}, \mathbf{p}) = \mathbf{p} \cdot \dot{\mathbf{y}} - \mathcal{L}. \tag{97}$$

This immediately gives us

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial \mathbf{y}} &= -\dot{\mathbf{p}} \\ \frac{\partial \mathcal{H}}{\partial \mathbf{p}} &= \dot{\mathbf{y}}. \end{aligned} \tag{98}$$

Systems of differential equations of this form are called Hamiltonian systems. These equations are obtained from the Euler-Lagrange equations just by a change of variables (from velocities to conjugate momenta), we may claim the same as we did in Lagrangian mechanics.

> The evolution of every isolated physical system can be described by a Hamiltonian system of differential equations.

To introduce the related theorems, we need the following concepts:

- The $(\mathbf{y}, \mathbf{p})$ space is called the **phase space** of the system.
- The volume of a region in the phase space is called the **phase volume**.
- The **time-t map** maps $(\mathbf{y}_0, \mathbf{p}_0)$ to $(\mathbf{y}_t, \mathbf{p}_t)$.

Theorem (Liouville)

The time-t maps of any Hamiltonian system

$$\begin{cases} \dot{y} = \dfrac{\partial \mathcal{H}}{\partial p} \\ \dot{p} = -\dfrac{\partial \mathcal{H}}{\partial y} \end{cases} \tag{99}$$

preserve the phase volume for all t.

Poincare Recurrence Theorem

Let the set (an energy level)

$$\Omega_C = \left\{ (\mathbf{y}, \mathbf{p}) \middle| \mathcal{H}(\mathbf{y}, \mathbf{p}) = C = \text{const.} \right\} \tag{100}$$

be bounded for some energy value C. Then, for a typical initial condition in Ω_C, the system returns arbitrarily close to its initial state infinitely many times.

To prove this second theorem, think about a ball of some volume in which the points never return. Therefore, all the images of the ball have a union of infinite volume, which is impossible to lie in the finite volume of Ω_C.

2.7 Symmetries & Conservation Laws

We've already known that if $\mathcal{L}(\mathbf{y}, \dot{\mathbf{y}})$ doesn't depend on time, then the system has conserved total energy:

$$\mathcal{H} = \mathbf{p} \cdot \dot{\mathbf{y}} - \mathcal{L} = \text{const.}$$

Noether's Theorem

There can be further conserved quantities. In general, conserved quantities are related to symmetries of the system. This relation between symmetry and conservation laws is known as **Noether's theorem**.

Cyclic Coordinates

We say that a system has a symmetry if we can introduce generalized coordinates $\mathbf{y}$ in such a way that the Lagrangian does not depend on some variable y_j. That is, $\partial\mathcal{L}/\partial y_j = 0$. The variable y_j may be referred to as a **cyclic coordinate**.

Even when y_j is cyclic, the Lagrangian $\mathcal{L}$ mat still depend on $\dot{y}_j$.

For a general Lagrangian $\mathcal{L}(\mathbf{x}, \dot{\mathbf{x}})$, suppose it has a symmetry. Then there exists a coordinate transformation $\mathbf{x} \to \mathbf{y}$ such that y_j is cyclic for some j. With the Euler-Lagrange equation, we know

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathcal{L}}{\partial\dot{y}_j} = \frac{\partial\mathcal{L}}{\partial y_j} = 0 \quad \Rightarrow \quad \frac{\partial\mathcal{L}}{\partial\dot{y}_j} = \text{const.} \tag{101}$$

This is the essence of Noether's theorem: if y_j is a cyclic variable, then its conjugate momentum $p_j = \partial\mathcal{L}/\partial\dot{y}_j$ stays constant along the trajectory of the system.

Due to the fact that y_j is cyclic, we can reduce the degree of freedom of the system by 1 and look for further symmetries. If we're auspicious enough, we may finally reduce the degrees of freedom to 1 and then solve the system completely with the last conservation law (energy conservation).

2.8 Holonomic Constraints

Holonomic constraints take the form

$$Q(\mathbf{x}) = 0 \tag{102}$$

for some smooth function Q. There is no explicit dependence on the velocities $\dot{x}_j$.

An obvious way to deal with this constraint is to solve for x_n:

$$x_n = g(x_1, \ldots, x_{n-1}). \tag{103}$$

This gives us the constraint on velocities:

$$\dot{x}_n = \sum_{j=1}^{n-1} \frac{\partial g}{\partial x_j} \dot{x}_j. \tag{104}$$

Then, we may reduce the system Lagrangian by the $n-1$ generalized coordinates. However, this approach can be problematic when it is not possible to represent the constraint as a single valued function of other coordinates.

To solve this problem, we can think of the holonomic constraint as being a result of an additional force that acts on the system and makes it to stay on the constraint surface described by $Q(\mathbf{x}) = 0$. It turns out that the force is orthogonal to the constraint surface:

$$\mathbf{F} = \lambda \boldsymbol{\nabla} Q. \tag{105}$$

This leads us to the constrained Lagrangian $\hat{\mathcal{L}}$:

$$\hat{\mathcal{L}}(\mathbf{x}, \dot{\mathbf{x}}, \lambda) = \mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}) + \lambda Q(\mathbf{x}). \tag{106}$$

We treat λ as a variable, so the Euler-Lagrange equation with λ reads $\frac{\partial \hat{\mathcal{L}}}{\partial \lambda} = 0$, and this gives our holonomic constraint.

If there are more than one holonomic constraints, we use

$$\hat{\mathcal{L}} = \mathcal{L} + \underline{\lambda} \cdot \mathbf{Q}. \tag{107}$$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \hat{\mathcal{L}}}{\partial \dot{\mathbf{x}}} - \frac{\partial \hat{\mathcal{L}}}{\partial \mathbf{x}} = 0 \quad &\Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}} - \frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \underline{\lambda} \cdot \mathbf{Q} \\ \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \hat{\mathcal{L}}}{\partial \dot{\lambda}_l} - \frac{\partial \hat{\mathcal{L}}}{\partial \lambda_l} = 0 \quad &\Rightarrow \quad Q_l = 0. \end{aligned} \tag{108}$$