Residues: Math and Physics

Gregg M. Gallatin gregg@appliedmathsolutions.com
Revision Date: March 12, 2025

November 19, 2025

Contents

1 Introduction

This is just a random collection of definitions, standard well known results, relationships, formulae, etc., from Math and Physics. There is nothing new except possibly the proof of the Bessel function closure relation. I have not seen it anywhere else, but no doubt it's out there somewhere, published long before I stumbled on this way of proving it myself.

I have tried to use pretty much standard notation throughout. A couple of standard math symbols that may need definition are: $\forall$ means "for all" or sometimes people say "for any" (either way means the same to me), $\in$ means "element of", $\exists$ means "there exists", "w.r.t." means "with respect to". I use the symbol "$\sim$" to mean "proportional to" or "scales as".

The word "Consider" will be used quite a bit, so get ready for it. I will often write out the step-by-step proof of a particular relation in excrutiating detail. Rather boring and very cluttered and hence awful to look at, but useful in the end, since it shows the individual steps in the derivation without having to write it out in words.

NOTE 1: This is a work in progress, I will continue to add to it. Keep track of the Revision Date above.

NOTE 2: This is not a hardcore math document, i.e., the style is not "theorem, proof, example". The level of rigor is that of pretty much any theoretical physics text.

NOTE 3: The sections are not in any particular order.

NOTE 4: The material included here ranges from some Quantum Field Theory stuff (energy and momentum conservation coming from time and space translation invariance), all the way "back" to why a capacitor always stores only half the energy supplied by the battery, generator or voltage source. Half is always lost to the resistor no matter how few Ohms it has.This fact is actually a big problem for computer chip design.

As an example of the "pure" math side, I show why Fourier transforms and Fourier series coefficients are least squares solutions.

Hope it's useful. Please, if you use anything from it, please reference it. Thanks!

Comments on notation:

It is common for physicists to write integrals in the form of an operation on a function and I will use that notation. That is instead of

$$\int f(x)\,dx$$

which is more or less standard in math texts, I will use the physics way of writing it

$$\int dx f(x)$$

Integration, $\int dx$, is then acting on the function $f(x)$.

There are two "delta functions" we will use quite often.

The Kronecker delta function will be written as $\delta_K(n,m)$ or as $\delta_{K,n,m}$ and the K will be dropped in expressions where it is obvious the Kronecker delta is being used. Here n and m, called "indices", are integers covering a specified range of values. The Kronecker delta function is just the elements of the identity matrix, $\delta_{K,n,m} = 1$ for $n = m$ and is zero otherwise. It appears in matrix manipulations and summations over given discrete ranges of n and m.

The Dirac delta function will be written as $\delta_D(x,y)$ or as $\delta_D(x-y)$ and the D will be dropped in expressions where it is obvious the Dirac delta is being used. Here x and y are continuous variables covering a specified range of values. The Dirac delta function is effectively the equivalent of the identity matrix but for integrals. It is defined by

$$f(x) = \int dy \delta_D(x-y) f(y)$$

for any (nonpathological) function $f(x)$.

2 Some Background Matrix Stuff

In many places we will use the matrix-vector notation and the "Einstein summation convention": (doubly) repeated indices are automatically assumed to be summed over the appropriate range. Example: ordinary matrix multiplication, indicated here by "$\cdot$", of matrix M on column vector x to yield column vector y can be written several different ways

$$y = M \cdot x$$
$$y_i = \sum_j M_{i,j} x_j = M_{i,j} x_j$$

Here y_i is the i^{th} component or element of column vector y, x_j is the j^{th} component or element of column vector x and $M_{i,j}$ is the row i and column j element of matrix M. The sum on j is over the range of values of j. A matrix inverse will be indicated with superscript "-1" and the transpose by superscript "T". Hence, for M a square invertible matrix, $M^{-1} \cdot M = I =$ the identity matrix of the same size as M and $x^T \cdot y =$ the "dot" product of x and y. This is the same as the dot product when x and y are treated as vectors, $\vec{x}$ and $\vec{y}$, that is $x^T \cdot y = \vec{x} \cdot \vec{y}$. Just different notation for the same thing.

Functions of matrices are defined by their power series expansion, e.g.,

$$\exp[M] = \sum_{n=1}^{\infty} \frac{1}{n!} M^n = \sum_{n=1}^{\infty} \frac{1}{n!} (M \cdot M \cdot \ldots \cdot M) = \lim_{p \to \infty} \left[\left(I + \frac{M}{p} \right)^p \right]$$

where there are n factors of M in the middle expression and I is the identity matrix. Proof of the "$\lim_{p\to\infty}$" expression above is in the next section.

NOTE: Since the function of a matrix is defined by its power series, it follows that if the matrix M is purely diagonal, i.e., all the off diagonal elements are zero, then any function of M,

$$f(M) = \begin{bmatrix} f(M_{1,1}) & 0 & \cdots & 0 \\ 0 & f(M_{2,2}) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & f(_{N,N}) \end{bmatrix}$$

That is, for diagonal matrices the function is distributed along the diagnal and all the off diagonal elements remain 0.

2.1 $\det[M] = \exp[\mathbf{tr}[\ln[M]]]$

We will now show that, for M an $N \times N$ matrix.

$$\det[M] \equiv \sum_{i_1, k_2, \cdots = 1}^{N} \epsilon_{i_1, i_2, \cdots, i_N} M_{1,i_1} M_{2,i_2} \cdots M_{N,i_N} = \exp[\mathrm{tr}[\ln[M]]]$$

Here "tr" indicates the trace which is the sum of the diagonal elements, e.g., $\mathrm{tr}[M] = \sum_{i=1}^{N} M_{i,i}$, and $\epsilon_{a,b,c,\cdots}$ is the totally antisymmetric "Levi-Civita tensor"

$$\epsilon_{a,b,c,\cdots} = \begin{cases} = +1 \text{ for } a,b,c,\cdots \text{ an even permutation of 1,2,3,}\cdots \\ = -1 \text{ for } a,b,c,\cdots \text{ an odd permutation of 1,2,3,}\cdots \\ = 0 \text{ if any of the } a,b,c,\cdots \text{ are equal to each other} \end{cases}$$

We will use the Einstein Summation Convention throughout this entire document. Note that in the sum above each summed over index $i_1,\ i_2,\ldots$ appears twice and only twice, once on ϵ and once on M. The Einstein Summation Convention simply drops the summation sign, simplifying the notation, and states that all doubly repeated indices are understood to be summed over their appropriate ranges.

Two examples: Via the Einstein Summation Convention we have

$$\begin{aligned} \epsilon_{i_1,i_2,\cdots,i_N} M_{1,i_1} M_{2,i_2} \cdots M_{N,i_N} &= \sum_{i_1,k_2,\cdots=1}^{N} \epsilon_{i_1,i_2,\cdots,i_N} M_{1,i_1} M_{2,i_2} \cdots M_{N,i_N} \\ \sum_{i=1}^{N} M_{i,i} &= M_{i,i} = \mathrm{tr}\,[M] \end{aligned}$$

Basically in all matrix expressions, indices appear once and only once when not summed over, and twice and only twice when summed over. Any expression where the same index appears three or more times generally, but certainly not always, indicates that a mistake has been made somewhere.

To proceed, begin by showing that

$$\exp[M] = \lim_{p\to\infty} \left[\left(I + \frac{M}{p} \right)^p \right]$$

Using the binomial theorem

$$\left(I + \frac{M}{p} \right)^p = \sum_{m=0}^{p} \frac{p!}{m!\,(p-m)!} \left(\frac{M}{p} \right)^m I^{p-m}$$

but obviously I^{p-m} can be ignored since I is the identity matrix, and we now have

$$\left(I + \frac{M}{p} \right)^p = \sum_{m=0}^{p} \frac{p!}{m!\,(p-m)!} \left(\frac{M}{p} \right)^m$$

By definition

$$\frac{p!}{(p-m)!} = p\,(p-1)\,(p-2)\cdots(p-m+1)$$

In the limit as $p \to \infty$ with m finite

$$\frac{p!}{(p-m)!} \to p^m$$

The reason we can consider m finite even though the sum over m runs from 0 up to p is because of the $m!$ in the denominator in the sum: the large m terms are essentially negligable. Putting this altogether gives

$$\begin{aligned}\lim_{p\to\infty}\left(I+\frac{M}{p}\right)^p &= \sum_{m=0}^{p}\frac{M^m}{m!}\\ &= \exp[M]\end{aligned}$$

So now consider

$$\begin{aligned}\det[\exp[M]] &= \lim_{p\to\infty}\det\left[\left(I+\frac{M}{p}\right)^p\right]\\ &= \lim_{p\to\infty}\left(\left(\det\left[\left(I+\frac{M}{p}\right)\right]\right)^p\right)\end{aligned}$$

where we have used identity $\det[A^n]=(\det[A])^n$. Note that this identity follows trivially from "det = exp tr ln".

Noting that the elements of I are $I_{m,n}=\delta_{m,n}=1$ for $m=n$ and are zero otherwise Here we are obviously dealing with the Kronecker delta and so we have

$$\begin{aligned}\det\left[\left(I+\frac{M}{p}\right)\right] &= \epsilon_{i_1,i_2,\cdots,i_N}\left(\delta_{1,i_1}+\frac{M_{1,i_1}}{p}\right)\left(\delta_{2,i_2}+\frac{M_{2,i_2}}{p}\right)\cdots\left(\delta_{N,i_N}+\frac{M_{N,i_N}}{p}\right)\\ &= \epsilon_{i_1,i_2,\cdots,i_N}\delta_{1,i_1}\delta_{2,i_2}\cdots\delta_{N,i_N}+\epsilon_{i_1,i_2,\cdots,i_N}\frac{M_{1,i_1}}{p}\delta_{2,i_2}\cdots\delta_{N,i_N}\\ &+\epsilon_{i_1,i_2,\cdots,i_N}\delta_{1,i_1}\frac{M_{2,i_2}}{p}\delta_{3,i_3}\cdots\delta_{N,i_N}+\cdots+\epsilon_{i_1,i_2,\cdots,i_N}\delta_{1,i_1}\delta_{2,i_2}\cdots\delta_{N-1,i_{N-1}}\frac{M_{N,i_N}}{p}\\ &+\text{terms of order }\frac{1}{p^2}\\ &= \epsilon_{1,2,3,\cdots,N}+\epsilon_{i_1,2,3,\cdots,N}\frac{M_{1,i_1}}{p}+\epsilon_{1,i_2,3\cdots,N}\frac{M_{2,i_2}}{p}\\ &+\cdots+\epsilon_{1,2,\cdots,N-1,i_N}\frac{M_{N,i_N}}{p}+\text{terms of order }\frac{1}{p^2}\\ &= 1+\frac{M_{1,1}}{p}+\frac{M_{2,2}}{p}+\cdots+\frac{M_{N,N}}{p}+\text{terms of order }\frac{1}{p^2}\\ &= 1+\frac{\mathrm{tr}[M]}{p}+\text{terms of order }\frac{1}{p^2}\end{aligned}$$

Hence we have

$$\begin{aligned}\det[\exp[M]] &= \lim_{p\to\infty}\left[\left(1+\frac{\mathrm{tr}[M]}{p}\right)^p\right]\\ &= \exp[\mathrm{tr}[M]]\end{aligned}$$

Replacing M with $\ln[M]$ we have

$$\det[\exp[\ln[M]]]=\det[M]=\exp[\mathrm{tr}[\ln[M]]]$$

NOTE: "det = exp tr ln" works as well when M is a differential operator instead of a matrix in which case $\delta_K(\cdots)$ is replaced with $\delta_D(\cdots)$. Hence it gets used quite a bit in functional integral approach to quantum field theory.

The above result immediately yields the derivative of $\det[M]$ w.r.t. any one element of M. Consider $M \to M + \Delta M$ and evaluate, step-by-step

$$\begin{aligned}
\det[M+\Delta M] - \det[M] &= \exp[\mathrm{tr}[\ln[M+\Delta M]] - \exp[\mathrm{tr}[\ln[M]]] \\
&= \exp\left[\mathrm{tr}\left[\ln\left[M\left(I + M^{-1}\cdot\Delta M\right)\right]\right]\right] - \exp[\mathrm{tr}[\ln[M]]] \\
&= \exp\left[\mathrm{tr}\left[\ln[M] + \ln\left[I + M^{-1}\cdot\Delta M\right]\right]\right] - \exp[\mathrm{tr}[\ln[M]]] \\
&= \exp[\mathrm{tr}[\ln[M]]]\left(\exp\left[\mathrm{tr}\left[\ln\left[I + M^{-1}\cdot\Delta M\right]\right]\right] - 1\right) \\
&= \det[M]\left(\exp\left[\mathrm{tr}\left[\ln\left[I + M^{-1}\cdot\Delta M\right]\right]\right] - 1\right) \\
&\text{For } \Delta M \text{ small expand } \ln \text{ then } \exp \\
&= \det[M]\left(\exp\left[\mathrm{tr}\left[M^{-1}\cdot\Delta M\right]\right] - 1\right) \\
&= \det[M]\,\mathrm{tr}\left[M^{-1}\cdot\Delta M\right]
\end{aligned}$$

For $\Delta M_{i,j} = \varepsilon\delta_{a,i}\delta_{b,j}$ (δ = Kronecker delta) so that $\Delta M = \varepsilon$ for $i = a$ and $j = b$ where a and b are particular values of i and j and $\Delta M = 0$ otherwise, this gives the derivative of $\det[M]$ w.r.t. the particular element $M_{a,b}$

$$\begin{aligned}
\frac{d(\det[M])}{dM_{a,b}} &= \lim_{\varepsilon\to 0}\frac{\det[M+\Delta M] - \det[M]}{\varepsilon} \\
&= \det[M]\lim_{\varepsilon\to 0}\frac{\mathrm{tr}\left[M^{-1}\cdot\Delta M\right]}{\varepsilon} \\
&= \det[M]\lim_{\varepsilon\to 0}\frac{\left(M^{-1}\right)_{i,k}\Delta Mk,i}{\varepsilon} \\
&= \det[M]\lim_{\varepsilon\to 0}\frac{\left(M^{-1}\right)_{i,k}\varepsilon\delta_{a,k}\delta_{b,i}}{\varepsilon} \\
&= \det[M]\left(M^{-1}\right)_{b,a}
\end{aligned}$$

2.2 Examples of det = exp tr ln

The fact that the determinant of a matrix equals the exponential of the trace of the natural logarithm of the matrix may seem very strange, so here are two examples showing how it works. Note that the trace is key since without it you would have exp(ln) which is an identity operation.

1. Consider the matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

where nominally a, b, c, d are real numbers but they can be complex.

First we have to evaluate $\ln(M)$. To do that we will express M in terms of

Pauli Matrices. The 3 standard Pauli Matrix (in standard notation) are

$$\begin{aligned}\sigma_1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ \sigma_2 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\ \sigma_3 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\end{aligned}$$

with, of course, $i = \sqrt{-1}$.

Define the commutator $[\cdots, \cdots]$ and anticommutator $\{\cdots, \cdots\}$ of 2 square matrices, A and B

$$\begin{aligned}[A, B] &= A \cdot B - B \cdot A \\ \{A, B\} &= A \cdot B + B \cdot A\end{aligned}$$

It is straightforward to show that

$$\begin{aligned}[\sigma_i, \sigma_j] &= 2i\epsilon_{i,j,k}\sigma_k \\ \{\sigma_i, \sigma_j\} &= 2\delta_{i,j}I\end{aligned}$$

with I the 2×2 identity matrix and $\delta_{i,j}$ the Kronecker delta. All indices i, j, k above can take the values $1, 2, 3$ and k is summed from 1 to 3 since it is repeated twice (Einstein Summation Convention). Don't confuse the factor $i = \sqrt{-1}$ above with the index i.

For ease of notation define

$$\sigma_0 = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and let Greek indices $\mu, \nu, \cdots$ range over $0, 1, 2, 3$. The four matrices σ_μ form a complete basis for expanding any 2×2 matrix. Indeed

$$M = C_\mu \sigma_\mu$$

with of course μ summed over $0, 1, 2, 3$ and

$$\begin{aligned}C_0 &= \frac{a+d}{2} \\ C_1 &= \frac{b+c}{2} \\ C_2 &= i\frac{b-c}{2} \\ C_3 &= \frac{a-d}{2}\end{aligned}$$

We now have

$$\begin{aligned}\ln[M] &= \ln\left[C_0\left(\sigma_0+\frac{C_i}{C_0}\sigma_i\right)\right]\\ &= \ln\left[(C_0 I)\cdot\left(I+\frac{C_i}{C_0}\sigma_i\right)\right]\\ &= \ln[C_o]\,I+\ln\left[I+\frac{C_i}{C_0}\sigma_i\right]\end{aligned}$$

where have substituted $\sigma_0 = I$, wrote C_0 as $C_0 I$ just to be clear it should be treated as a matrix and used the fact that $\ln[A\cdot B] = \ln[A]+\ln[B]$ when $[A,B]=0$ which applies since $\left[C_0 I, I+\frac{C_i}{C_0}\sigma_i\right]=0$.

Using the Taylor expansion

$$\ln\left[I+\frac{C_i}{C_0}\sigma_i\right] = \sum_{n=1}^{\infty}\frac{(-1)^{n+1}\left(\frac{C_i}{C_0}\sigma_i\right)^n}{n}$$

Evaluate successive terms in the series to see the pattern. To do this use

$$\begin{aligned}\sigma_i\cdot\sigma_j &= \frac{1}{2}\left([\sigma_i,\sigma_j]+\{\sigma_i,\sigma_j\}\right)\\ &= \epsilon_{i,j,k}\sigma_j+\delta_{i,j}I\end{aligned}$$

which maps higher powers of σ_i back to combinations of I and σ_i to the first power

$$\begin{aligned}\left(\frac{C_i}{C_0}\sigma_i\right)^2 &= \frac{C_i C_j}{C_0^2}\sigma_i\cdot\sigma_j\\ &= \frac{C_i C_j}{C_0^2}\left(\epsilon_{i,j,k}\sigma_j+\delta_{i,j}I\right)\\ &= \frac{C_i C_i}{C_0^2}I\\ &\equiv \frac{C^2}{C_0^2}I\end{aligned}$$

$$\begin{aligned}\left(\frac{C_i}{C_0}\sigma_i\right)^3 &= \left(\frac{C_i}{C_0}\sigma_i\right)^2\cdot\left(\frac{C_j}{C_0}\sigma_j\right)\\ &= \frac{C^2}{C_0^2}\frac{C_i}{C_0}\sigma_i\end{aligned}$$

$$\begin{aligned}\left(\frac{C_i}{C_0}\sigma_i\right)^4 &= \left(\frac{C_i}{C_0}\sigma_i\right)^2\cdot\left(\frac{C_i}{C_0}\sigma_i\right)^2\\ &= \left(\frac{C^2}{C_0^2}\right)^2 I\end{aligned}$$

$$\begin{aligned}\left(\frac{C_i}{C_0}\sigma_i\right)^5 &= \left(\frac{C_i}{C_0}\sigma_i\right)^4 \cdot \left(\frac{C_j}{C_0}\sigma_j\right) \\ &= \left(\frac{C^2}{C_0^2}\right)^2 \frac{C_i}{C_0}\sigma_i\end{aligned}$$

Considering *even* and *odd* powers

$$\begin{aligned}\left(\frac{C_i}{C_0}\sigma_i\right)^{even} &= \left(\frac{C^2}{C_0^2}\right)^{even/2} I \\ \left(\frac{C_i}{C_0}\sigma_i\right)^{odd} &= \left(\frac{C^2}{C_0^2}\right)^{(odd-1)/2} \left(\frac{C_i}{C_0}\sigma_i\right)\end{aligned}$$

Taking the trace gives

$$\begin{aligned}\mathrm{tr}\left[\left(\frac{C_i}{C_0}\sigma_i\right)^{even}\right] &= \left(\frac{C^2}{C_0^2}\right)^{even/2} \mathrm{tr}\left[I\right] \\ &= 2\left(\frac{C^2}{C_0^2}\right)^{even/2} \\ \mathrm{tr}\left[\left(\frac{C_i}{C_0}\sigma_i\right)^{odd}\right] &= \left(\frac{C^2}{C_0^2}\right)^{(odd-1)/2} \left(\frac{C_i}{C_0}\mathrm{tr}\sigma_i\right) \\ &= 0\end{aligned}$$

and we have, letting $even = 2n$,

$$\begin{aligned}\mathrm{tr}\left[\ln\left[I + \frac{C_i}{C_0}\sigma_i\right]\right] &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}\,\mathrm{tr}\left[\left(\frac{C_i}{C_0}\sigma_i\right)^n\right]}{n} \\ &= \sum_{n=1}^{\infty} \frac{(-1)^{2n+1}\,\mathrm{tr}\left[\left(\frac{C_i}{C_0}\sigma_i\right)^{2n}\right]}{2n} \\ &= -\sum_{n=1}^{\infty} \frac{\left(\frac{C^2}{C_0^2}\right)^n}{n} \\ &= \ln\left[1 - \frac{C^2}{C_0^2}\right]\end{aligned}$$

Also

$$\mathrm{tr}\left[\ln\left[C_0\right] I\right] = 2\ln\left[C_0\right] = \ln\left[C_0^2\right]$$

Putting things altogether

$$\begin{aligned}\det\left[M\right] &= \exp\left[\mathrm{tr}\left[\ln\left[M\right]\right]\right] \\ &= \exp\left[\ln\left[C_0^2\right] + \ln\left[1 - \frac{C^2}{C_0^2}\right]\right] \\ &= \exp\left[\ln\left[C_0^2 - C^2\right]\right] \\ &= ad - bc\end{aligned}$$

2. Consider a matrix M which is the outer product of an N element column vector u with its Hermitian conjugate, $u^\dagger$, i.e.,

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \text{ and } u^\dagger = [u_1^*, u_2^*, \cdots, u_N^*]$$

so that

$$\begin{aligned} M &= uu^\dagger = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} [u_1^*, u_2^*, \cdots, u_N^*] \\ &= \begin{bmatrix} u_1u_1^* & u_1u_2^* & \cdots & u_1u_N^* \\ u_2u_1^* & u_2u_2^* & \cdots & u_2u_N^* \\ \vdots & & \ddots & \\ u_Nu_1^* & u_Nu_2^* & \cdots & u_Nu_N^* \end{bmatrix} \end{aligned}$$

Take the inner product of $u^\dagger$ with u to be X, i.e.,

$$u^\dagger \cdot u = X$$

where "$\cdot$" is matrix multiplication.

Note: X is just a number.

We now have

$$\begin{aligned} M \cdot M &= uu^\dagger \cdot uu^\dagger \\ &= u\left(u^\dagger \cdot u\right)u^\dagger \\ &= Xuu^\dagger \\ &= XM \end{aligned}$$

From this is follows that

$$M^n = X^{n-1}M$$

Also

$$\text{tr}\left[M\right] = \text{tr}\left[uu^\dagger\right] = \sum_{n=1}^{N} u_n u_n^\dagger = X$$

Determine the eigenvalues of M.

To do that we need to find the values of λ which satisfy

$$0 = \det\left[\lambda I - M\right]$$

Here I is the $N \times N$ identity matrix. For an $N \times N$ matrix there are N values of λ, assuming at least one is nonzero we can rewrite the above equation as

$$0 = \det\left[\lambda\left(I - \frac{M}{\lambda}\right)\right]$$

Using the standard property of determinants $\det[cA] = c^N \det[A]$ where c is a number and A is any $N \times N$ matrix we have

$$0 = \lambda^N \det\left[I - \frac{M}{\lambda}\right]$$

Sidebar: Note that we can prove $\det[cA] = c^N \det[A]$ very simply using "det = exp tr ln":

$$\begin{aligned}
\det[cA] &= \exp[\mathrm{tr}[\ln[cA]]] \\
&= \exp[\mathrm{tr}[\ln[cI] + \ln[A]]] \\
&= \exp[\ln[c]\,\mathrm{tr}[I] + \mathrm{tr}[\ln[A]]] \\
&= \exp[\ln[c]\,N + \mathrm{tr}[\ln[A]]] \\
&= c^N \det[A]
\end{aligned}$$

End Sidebar:

The λ^N factor makes it appear that there are N "$\lambda = 0$" solutions, but before jumping to that conclusion, evaluate the remaining det using "exp tr ln".

$$\begin{aligned}
\det\left[I - \frac{M}{\lambda}\right] &= \exp\left[\mathrm{tr}\left[\ln\left[I - \frac{M}{\lambda}\right]\right]\right] \\
&= \exp\left[\mathrm{tr}\left[-\sum_{n=1}^{\infty} \frac{M^n}{n\lambda^n}\right]\right] \\
&= \exp\left[\mathrm{tr}\left[-\sum_{n=1}^{\infty} \frac{X^{n-1}}{n\lambda^n} M\right]\right] \\
&= \exp\left[-\sum_{n=1}^{\infty} \frac{X^{n-1}}{n\lambda^n} \mathrm{tr}[M]\right] \\
&= \exp\left[-\sum_{n=1}^{\infty} \frac{X^n}{n\lambda^n}\right] \\
&= \exp\left[\ln\left[1 - \frac{X}{\lambda}\right]\right] \\
&= 1 - \frac{X}{\lambda}
\end{aligned}$$

And so finally we have

$$\begin{aligned}
0 &= \det[\lambda I - M] \\
&= \lambda^N \left(1 - \frac{X}{\lambda}\right) \\
&= \lambda^{N-1} (\lambda - X)
\end{aligned}$$

Hence M has $N - 1$ "$\lambda = 0$" eigenvalues with the single nonzero eigenvalue being $\lambda = X = \mathrm{tr}[M]$.

3 Fourier Stuff

Consider the sum

$$f_N(x) = \sum_{n=-N}^{+N} \exp[in2\pi x/P]$$

where $i = \sqrt{-1}$ and x and P are real valued. By definition $f_N(x)$ is periodic

$$f_N(x) = f_N(x+P) \ \forall x$$

Integrate over one period and normalize by the period

$$\frac{1}{P}\int_{-P/2}^{+P/2} dx f_N(x) = \sum_{n=-N}^{+N} \frac{1}{P}\int_{-P/2}^{+P/2} dx \exp[in2\pi x/P] = 1$$

where we have used the fact that

$$\frac{1}{P}\int_{-P/2}^{+P/2} dx \exp[in2\pi x/P] = \delta_K(n,0) = \begin{cases} 1 \text{ for } n = 0 \\ 0 \text{ for } n \neq 0 \end{cases}$$

δ_K is the Kronecker delta function, $\delta_K(u,v) = 1$ for $u = v$ and is zero otherwise. The sum is finite so can switch it with the integral. The important point in the above result is that it is independent of N.

The sum can be evaluated exactly, which is shown just below

$$f_N(x) = \frac{\sin\left[\left(N+\frac{1}{2}\right)2\pi\frac{x}{P}\right]}{\sin\left[\pi\frac{x}{P}\right]} = \frac{\sin\left[(2N+1)\pi\frac{x}{P}\right]}{\sin\left[\pi\frac{x}{P}\right]}$$

Consider the sum $\sum_{n=0}^{\infty} x^n = 1/(1-x)$ which is valid for $|x| < 1$ but does converge for $x = \exp[i\theta]$. Consider the finite sum $\sum_{n=0}^{N} x^n = \sum_{n=0}^{\infty} x^n - \sum_{n=N+1}^{\infty} x^n = \sum_{n=0}^{\infty} x^n - x^{N+1}\sum_{n=0}^{\infty} x^n = \left(1-x^{N+1}\right)/(1-x)$. The sum $\sum_{n=-1}^{-N} x^n = -1 + \sum_{n=0}^{N}\left(x^{-1}\right)^n = -1 + \left(1-\left(x^{-1}\right)^{N+1}\right)/\left(1-x^{-1}\right)$. Putting the two terms together and doing some algebra gives

$$\sum_{n=-N}^{N} x^n = -1 + \frac{x^{-N}\left(1+x^N\right)\left(1-x^{N+1}\right)}{1-x}$$

Substituting $x \to \exp[i2\pi x/P]$ and using $\exp[i\theta] = \cos[\theta] + i\sin[\theta]$ gives the result for $f_N(x)$ shown above.

The peak value in the range $-P/2$ to $+P/2$ is

$$f_N(0) = 2N+1$$

as expected.

The position of the first zero is

$$\sin\left[(2N+1)\,\pi\frac{x}{P}\right] = \sin\left[\pm\pi\right] = 0$$
$$\Rightarrow x = \pm\frac{P}{2N+1}$$

With increasing N the peak value of $f_N(x)$ increases, its width decreases and the integral

$$\int_{-P/2}^{+P/2} dx f_N(x) = P$$

for all N. Hence in the range $-P/2$ to $+P/2$ we have

$$f_{N\to\infty}(x) = P\delta_D(x) = \text{Dirac delta function}$$

(The subscript D on δ is there to distinguish it from the Kronecker delta, δ_K. Properties of the Dirac delta function are discussed below.) Accounting for periodicity we have

$$f_{N\to\infty}(x) = P\sum_{n=-\infty}^{+\infty}\delta_D(x-nP)\ \ \forall x$$

As discussed below the Dirac delta function has units of 1/(its argument). $f_N(x)$ as it is defined via the sum of exponentials is dimensionless. The factor of P in the result for $f_{N\to\infty}(x)$ above occurs so that the righthand side is dimensionless as well.

3.1 Deriving Fourier Series

Consider a periodic function $g(x)$,

$$g(x) = g(x+P)\ \ \forall x$$

Assuming $g(x)$ is nonpathological (analytic) we can Taylor expand the right hand side

$$\begin{aligned} g(x+P) &= g(x) + \partial_x g(x)\,P + \frac{1}{2}\partial_x^2 g(x)\,P^2 + \cdots \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}P^n\partial_x^n g(x) \\ &= \exp\left[P\partial_x\right] g(x) \end{aligned}$$

In the last step we have used the standard power series representation of the exponential.

$$\sum_{n=0}^{\infty}\frac{1}{n!}\left(P\partial_x\right)^n = \exp\left[P\partial_x\right]$$

Using periodicity, $g(x+P) = g(x)$, gives

$$0 = \partial_x g(x) P + \frac{1}{2}\partial_x^2 g(x) P^2 + \cdots = \sum_{n=1}^{\infty} \frac{1}{n!}\partial_x^n g(x) P^n$$

This is a linear homogeneous differential equation with constant coefficients and so the general form of the solution (assuming no degeneracy in the values of λ) is

$$g(x) = \exp[\lambda x]$$

Substituting this in the differential equation gives

$$0 = \lambda P + \frac{1}{2}(\lambda P)^2 + \cdots = \exp[\lambda P] - 1$$

or

$$\exp[\lambda P] = 1$$

Hence

$$\lambda = i\frac{n2\pi}{P}$$

where $i = \sqrt{-1}$ and n = all integers (positive, negative and zero). The differential equation is infinite order and therefore requires an infinite number of coefficients. Hence the general form for $g(x)$ is

$$g(x) = \sum_{n=-\infty}^{+\infty} a_n \exp\left[in2\pi\frac{x}{P}\right]$$

which is the Fourier series representation of $g(x)$. Given $g(x)$, the a_n are given by

$$a_n = \frac{1}{P}\int_{-P/2}^{+P/2} dx g(x) \exp\left[-in2\pi\frac{x}{P}\right]$$

which follows directly upon substituting the Fourier series representation of $g(x)$ in the above integration, exchanging summation and integration (generally valid in real world problems) and doing the integral.

3.2 Fourier transforms from Fourier series

Consider

$$\begin{aligned}\frac{1}{P}\sum_{n=-\infty}^{\infty} \exp\left[in2\pi\frac{x}{P}\right] &= \sum_{n=-\infty}^{+\infty} \delta_D(x - nP) \\ &= \frac{2\pi}{P}\sum_{n=-\infty}^{+\infty} \delta_D\left(2\pi\frac{x}{P} - n2\pi\right)\end{aligned}$$

or, to write it in a different form, let

$$\theta = 2\pi \frac{x}{P}$$

and we get

$$\frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} \exp\left[in\theta\right] = \sum_{n=-\infty}^{+\infty} \delta_D\left(\theta - n2\pi\right)$$

Note

$$\begin{aligned}
\lim_{P\to\infty} \frac{1}{P} \sum_{n=-\infty}^{\infty} \exp\left[in2\pi \frac{x}{P}\right] &= \frac{1}{2\pi} \lim_{P\to\infty} \sum_{n=-\infty}^{+\infty} \left(\frac{2\pi}{P}\right) \exp\left[in\left(\frac{2\pi}{P}\right)x\right] \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\beta \exp\left[i\beta x\right] \\
&= \delta_D\left(\beta\right)
\end{aligned}$$

where we used $\Delta\beta = 2\pi/P \to d\beta$ and $n2\pi/P \to \beta$ for $P \to \infty$ and in the last step since $P \to \infty$ only the $n = 0$ term contributes to $\sum_n \delta_D\left(x - nP\right)$. Using this same approach we have

$$\begin{aligned}
g\left(x\right) &= \lim_{P\to\infty} \sum_{n=-\infty}^{+\infty} a_n \exp\left[in2\pi \frac{x}{P}\right] \\
&= \lim_{P\to\infty} \sum_{n=-\infty}^{+\infty} \frac{2\pi}{P} \frac{a_n}{2\pi/P} \exp\left[in2\pi \frac{x}{P}\right] \\
&= \lim_{P\to\infty} \sum_{n=-\infty}^{+\infty} \Delta\beta a\left(n\Delta\beta\right) \exp\left[in\Delta\beta x\right] \\
&= \int_{-\infty}^{+\infty} d\beta a\left(\beta\right) \exp\left[i\beta x\right]
\end{aligned}$$

which is the Fourier transform.

NOTE: We needed to define

$$a\left(n\Delta\beta\right) = \frac{a_n}{2\pi/P} = \frac{a_n}{\Delta\beta}$$

because $a\left(\beta\right)$ needs to be a "density", i.e., it has to have (units of $g\left(x\right)$)/(units of β) which is exactly what is needed to have (units of $g\left(x\right)$) on both sides of the equation.

3.3 Fourier Transform of a Periodic Function: Frequency Comb

We take take any function, $f(x)$, and make it a periodic function, $F(x)$, simply write

$$F(x) = \sum_{n=-\infty}^{+\infty} f(x - nP)$$

where P is the period of F, i.e.,

$$F(x + nP) = F(x) \ \forall x \text{ and } n = \text{Integer}$$

NOTE: The $f(x)$ can be any function in the sense that $f(x)$ does not have to vanish outside of 1 period P, e.g., $f(x)$ does not have to be zero for $x > P/2$ or $x < -P/2$. The $f(x)$ from different periods of can overlap each other.

Fourier transform $F(x)$

$$\begin{aligned}
\tilde{F}(\beta) &\equiv \int \frac{dx}{2\pi} e^{-i\beta x} F(x) \\
&= \sum_{n=-\infty}^{+\infty} \int \frac{dx}{2\pi} e^{-i\beta x} f(x - nP) \\
\text{Let } x &\rightarrow x + nP \text{ for each } n \\
&= \sum_{n=-\infty}^{+\infty} \exp[-in\beta P] \int \frac{dx}{2\pi} e^{-i\beta x} f(x) \\
&= \left(\sum_{n=-\infty}^{+\infty} \exp[-in\beta P] \right) \tilde{f}(\beta)
\end{aligned}$$

But as shown above $\frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} \exp[in\theta] = \sum_{n=-\infty}^{+\infty} \delta_D(\theta - n2\pi)$. Letting $\theta = \beta P$ we get

$$\begin{aligned}
\tilde{F}(\beta) &= 2\pi \sum_{n=-\infty}^{+\infty} \delta_D(\beta P - n2\pi) \tilde{f}(\beta) \\
&= \frac{2\pi}{P} \sum_{n=-\infty}^{+\infty} \delta_D\left(\beta - n\frac{2\pi}{P}\right) \tilde{f}\left(n\frac{2\pi}{P}\right)
\end{aligned}$$

where we have used the fact that $\delta_D(ax) = 1/|a| \, \delta_D(x)$ with P positive and $\delta_D(x - x_0) f(x) = \delta_D(x - x_0) f(x_0)$ for any function $f(x)$.

A periodic sum of Dirac delta functions, as above, is known as a "frequency comb" or a "Dirac comb". Here the frequencies are spatial. The same thing works with temporal frequencies and is often done using lasers to produce periodic pulses of light which are, to all intents and purposes, exact copies of each other.

4 Dirac Delta Function

The Dirac delta function $\delta_D(x)$ is "defined" as

$$\delta_D(x) = \begin{cases} \infty \text{ for } x = 0 \\ 0 \text{ for } x \neq 0 \end{cases}$$

with the property that

$$\int dx \delta_D(x) = 1$$

as long as the the range of integration includes $x = 0$.

This is actually not a valid way to define an ordinary function and so $\delta_D(x)$ must be thought of as the limit of an ordinary function.

For example a useful representation of the Dirac delta function is the limit of a Gaussian with the root-mean-square width σ, which integrates to 1 over the range $-\infty < x < +\infty$, as $\sigma \to 0$,

$$\delta_D(x) = \lim_{\sigma \to 0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2}{2\sigma^2}\right]$$

Integrating gives

$$\int_{-\infty}^{+\infty} dx \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2}{2\sigma^2}\right] = 1$$

independent of the value of σ. The reason we had to integrate from $-\infty$ to $+\infty$ is the Gaussian is nonzero over that range for any finite value of σ.

Another representation of $\delta_D(x)$ that is strictly 0 outside a finite range is

$$\delta_D(x) = \begin{cases} \frac{1}{2\varepsilon} \text{ for } -\varepsilon \leq x \leq \varepsilon \\ 0 \text{ otherwise} \end{cases}$$

which obviously integrates to 1.

There are many other useful representations. Perhaps one of the most useful is the Fourier representation

$$\delta_D(x) = \int_{-\infty}^{+\infty} \frac{d\beta}{2\pi} e^{i\beta x} = \int_{-\infty}^{+\infty} \frac{d\beta}{2\pi} e^{-i\beta x}$$

where the second equality follows from the change of integration variable $\beta \to -\beta$. This is derived below in the section on Fourier Stuff.

4.1 Properties of Dirac Delta Functions

For any nonpathological $f(x)$ we have

1.
$$\int_a^b dx f(x) \delta_D(x - x_0) = \begin{cases} f(x_0) \text{ for } a < x_0 < b \\ 0 \text{ otherwise} \end{cases}$$

This can be shown by first noting that $\delta_D(x-x_0)$ is nonzero only for $x=x_0$ and so we can simply evaluate f at x_0 in which case it comes out of the integral leaving $\int dx\delta_D(x-x_0)=1$ which follows trivially after making the change of integration variable $x\rightarrow x+x_0$. For x_0 equal to either a or b nominally you get $f(x_0)/2$ because only half the delta function is inside the range of integration but this has to be verified in each case.

2.

$$\int_a^b dxf(x)\,\delta_D(c\times(x-x_0))=\begin{cases}\frac{f(x_0)}{|c|} \text{ for } a<x_0<b\\ 0 \text{ otherwise}\end{cases}$$

where as usual $|\cdots|$ means absolute value and c is a constant. This can be shown in various ways. For example, let $x=x'+x_0$ then let $x''=cx'$ which gives

$$\begin{aligned}\int_a^b dxf(x)\,\delta_D(c\times(x-x_0)) &= \frac{1}{c}\int_{c(a-x_0)}^{c(b-x_o)} dxf\left(\frac{x''}{c}+x_0\right)\delta(x'')\\ &= \frac{f(x_0)}{c}\int_{c(a-x_0)}^{c(b-x_o)} dx\delta(x'')\end{aligned}$$

Now if $c>0$ then $\int_{c(a-x_0)}^{c(b-x_o)} dx\delta(x'')=1$. Remember we have assumed that $a<x_0<b$ and so $a-x_0<0<b-x_0$. But if $c<0$ then we have to swap the limits and we have $\int_{c(b-x_0)}^{c(a-x_o)} dx\delta(x'')=-1$ and therefore the $1/c$ factor becomes $1/|c|$. Thus

$$\delta_D(c\times(x-x_0))=\frac{1}{|c|}\delta_D(x-x_0)$$

3.

$$\begin{aligned}\int dxf(x)(\delta_D(g(x))) &= \sum_n \frac{f(x_n)}{\left|\partial_x g(x)_{x=x_n}\right|}\\ \text{where } g(x_n) &= 0\end{aligned}$$

i.e., the x_n are the roots of $g(x)$. This identity can be proven by noting first that the only contributions to the integral will come from the points where $x=x_n$. Separating the integral into a sum of integrals with integration ranges individually enclosing each of the x_n we have

$$\int dxf(x)(\delta_D(g(x)))=\sum_n\int_{a_n}^{b_n} dxf(x)\,\delta_D(g(x))$$

with $a_n<x_n<b_n$. Let $x=x'+x_n$ in each integral then Taylor expand $g(x+x_n)=g(x_n)+\partial_x g(x)_{x=x_n}x+\cdots=0+\partial_x g(x)_{x=x_n}x+\cdots$. The higher order terms "$+\cdots$" won't contribute because of the infinitely thin width of the

delta function and so we have

$$\begin{aligned}\int dx f(x)\left(\delta_D\left(g(x)\right)\right) &= \sum_n \int_{a_n-x_n}^{b_n-x_n} dx f(x+x_n)\,\delta_D\left(g(x+x_n)\right) \\ &= \sum_n \int_{a_n-x_n}^{b_n-x_n} dx f(x+x_n)\,\delta_D\left(\left(\partial_x g(x)_{x=x_n}\right)\times x\right) \\ &= \sum_n \frac{1}{\left|\partial_x g(x)_{x=x_n}\right|}\int_{a_n-x_n}^{b_n-x_n} dx f(x+x_n)\,\delta_D(x) \\ &= \sum_n \frac{f(x_n)}{\left|\partial_x g(x)_{x=x_n}\right|}\end{aligned}$$

where we have used the identity in 2. above in the next to last step.

5 Gaussian Integral

To integrate the Gaussian $\exp\left[-x^2\right]$ over the range $-\infty$ to $+\infty$ write the integral as the squareroot of the square of the integral, i.e.,

$$\int_{-\infty}^{+\infty} dx e^{-x^2} = \sqrt{\left(\int_{-\infty}^{+\infty} dx e^{-x^2}\right)^2}$$

which is valid because $\exp\left[-x^2\right] \geq 0\ \forall x$, and so the integral is positive. Rewriting the product of the integrals as a double integral and switching to polar coordinates gives

$$\begin{aligned}\int_{-\infty}^{+\infty} dx e^{-x^2} &= \sqrt{\int_{-\infty}^{+\infty} dx dy e^{-(x^2+y^2)}} \\ &= \sqrt{\int_0^{2\pi} d\theta \int_0^{\infty} dr r e^{-r^2}} \\ &= \sqrt{2\pi \int_0^{\infty} dr \left(-\frac{1}{2}\partial_r e^{-r^2}\right)} \\ &= \sqrt{-\pi\left(e^{-\infty^2}-1\right)} \\ &= \sqrt{\pi}\end{aligned}$$

To integrate $\exp\left[-ax^2+bx\right]$ complete the square in the exponent and make the appropriate change of variables and you get

$$\int_{-\infty}^{+\infty} dx \exp\left[-ax^2+bx\right] = \sqrt{\frac{\pi}{a}}\exp\left[\frac{b^2}{4a}\right]$$

This works for a and b having complex values, as long as $\mathrm{Re}\left[a\right] > 0$. The limit $\mathrm{Re}\left[a\right] \to 0$ can be taken after evaluating the integral. Hence we can write

(changing notation a bit)

$$\int_{-\infty}^{+\infty} dx \exp\left[-iax^2 + ibx\right] = \sqrt{\frac{\pi}{ia}} \exp\left[-\frac{b^2}{4ia}\right]$$

where $i = \sqrt{-1}$.

6 Inequalities

Let $\vec{x}, \vec{y}, \cdots$ be vectors in a real valued vector space V with inner or dot product "$\cdot$". $\vec{x}, \vec{y}, \cdots$ can be ordinary vectors in a D dimensional Euclidean space or functions in a Hilbert space. Because the vectors are members of a real valued vector space we have

$$|\vec{x}|^2 \equiv \vec{x} \cdot \vec{x} \geq 0 \text{ for any } \vec{x} \in V$$

Note with this definition the "length" of $\vec{x}$ is $|\vec{x}| = \sqrt{\vec{x} \cdot \vec{x}}$

6.0.1 Cauchy-Schwarz

Consider the combination

$$\vec{y} - \frac{\vec{y} \cdot \vec{x}}{|\vec{x}|} \frac{\vec{x}}{|\vec{x}|}$$

This is $\vec{y}$ with the part of $\vec{y}$ lying in the $\vec{x}$ direction subtracted off. Since the square of any vector in the space is greater than or equal to zero we have

$$\left| \vec{y} - \frac{\vec{y} \cdot \vec{x}}{|\vec{x}|} \frac{\vec{x}}{|\vec{x}|} \right|^2 = \left(\vec{y} - \frac{\vec{y} \cdot \vec{x}}{|\vec{x}|} \frac{\vec{x}}{|\vec{x}|} \right) \cdot \left(\vec{y} - \frac{\vec{y} \cdot \vec{x}}{|\vec{x}|} \frac{\vec{x}}{|\vec{x}|} \right) \geq 0$$

Expanding and rearranging gives the Cauchy-Schwarz inequality

$$|\vec{x}|\,|\vec{y}| \geq \vec{x} \cdot \vec{y} : \text{ Cauchy-Schwarz Inequality}$$

6.0.2 Triangle

Noting that

$$|\vec{x} + \vec{y}|^2 = |\vec{x}|^2 + |\vec{y}|^2 + 2\vec{x} \cdot \vec{y}$$

and

$$(|\vec{x}| + |\vec{y}|)^2 = |\vec{x}|^2 + |\vec{y}|^2 + 2\,|\vec{x}|\,|\vec{y}|$$

then using the Cauchy-Schwarz inequality and taking a square root gives the Triangle Inequality

$$|\vec{x}| + |\vec{y}| \geq |\vec{x} + \vec{y}| \ : \text{ Triangle Inequality}$$

7 Probability Stuff

7.1 Basic definition

Let

$$\vec{X} = (X_1, X_2, \cdots, X_N)$$

be a set or string of N random variables X_n and let

$$P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right)$$

be the probability density or distribution for $\vec{X}$ to take the value $\vec{x}$ given the parameter values $\vec{\theta}$. The vector notation is purely for convenience, $\vec{x}$ and $\vec{\theta}$ can simply be thought of as lists or strings of values, that is

$$(x_1, x_2, \cdots, x_N) \equiv \vec{x}$$

is a set or string of N values and

$$(\theta_1, \theta_2, \cdots, \theta_M) \equiv \vec{\theta}$$

is a set or string of M parameter values. There is no immediate requirement that $\vec{x}$ or $\vec{\theta}$ are members of a particular vector space. In general all the x_n and all the θ_m values are real.

NOTE: $P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right)$ is generally referred to in the literature as the "joint" probability distribution for the values of x_1,$x_2, x_3, \cdots, x_N$.

The probability that the random variable, or string of variables

$$X_1, X_2, \cdots, X_N = \vec{X}$$

lies within the infinitesimal volume $dv = dx_1 dx_2 \cdots dx_N$ centered at $\vec{x}$ given the string of values in $\vec{\theta}$ is

$$P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right) dx_1 dx_2 \cdots dx_N = P\left(\vec{x} \mid \vec{\theta}\right) dv$$

Note: Stating that $P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right)$ is a probability "density" makes the implicit assumption that all the x_n values are continuous. We will assume all the θ_m values are continuous as well. In order for $P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right)$ to be a probability density its integral over all relevant ranges of $\vec{x}$ must be finite and should always be normalized to 1, or to 100 % if desired, i.e.

$$\text{Normalization: } \int dv \; P_{\vec{X}}\left(\vec{x} \;\middle|\; \vec{\theta}\right) = 1 \text{ Always}$$

A discrete probability distribution is where, for example, the values of a single random variable X can take only particular discrete values. For example if the probability that X can take the particular value x_n is $p_n\left(\vec{\theta}\right)$ where n

labels the possible values of X then the probability distribution for X to take the value x_n is

$$P_X(n) = \sum_m p_m\left(\vec{\theta}\right)\delta_{m,n}$$

where m ranges over all the possible values of n and $\delta_{n,m}$ is the Kronecker delta, i.e., the elements of the identity matrix, $\delta_{m,n} = 1$ for $m = n$ and is zero otherwise. Normalization for a discrete probability distribution takes the form

$$\text{Normalization: } \sum_n P_X(n) = \sum_{n,m} p_m\left(\vec{\theta}\right)\delta_{m,n} = \sum_m p_m\left(\vec{\theta}\right) = 1 \text{ Always}$$

7.2 Expectation values

The expectation values of the X_n are defined by

$$\left\langle F\left(\vec{X}\right)\right\rangle = E\left[F\left(\vec{X}\right)\right] = \int d\nu F(\vec{x})\, P_X\left(\vec{x} \mid \vec{\theta}\right)$$

where the $\langle\cdots\rangle = E[\cdots]$ is just diffferent notation for the same thing. Both ways of writing the expectation value are commonly used in the literature.

As a specifict example consider a single random variable X, the mean or average value and the mean square variation, also called the variance are given by

$$\begin{aligned}
\text{mean or average} &= \langle X\rangle = E[X] = \int dx\; xP_X\left(x \mid \vec{\theta}\right)\\
\text{mean square or variance} &= \left\langle (X - \langle X\rangle)^2\right\rangle\\
&= \int dx\;\; (x - \langle X\rangle)^2 P_X\left(x \mid \vec{\theta}\right)\\
&= \int dx\;\; \left(x^2 - 2x\langle X\rangle + \langle X\rangle^2\right) P_X\left(x \mid \vec{\theta}\right)\\
&= \int dx\; x^2 P_X\left(x \mid \vec{\theta}\right) - 2\langle X\rangle \int dx\; xP_X\left(x \mid \vec{\theta}\right)\\
&+ \langle X\rangle^2 \int dx\; P\left(x \mid \vec{\theta}\right)\\
&= \left\langle x^2\right\rangle - \langle x\rangle^2
\end{aligned}$$

7.2.1 Examples

Gaussian Distribution The Gaussian distribution for a single continuous variable X is given by

$$P_X(x|\theta_1 = \mu, \theta_2 = \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

where

$\theta_1 = \mu =$ the mean value $= \langle X \rangle$

$\theta_2 = \sigma =$ the root mean square value (RMS) $= \left\langle \left(X - \langle X \rangle^2 \right) \right\rangle = \left\langle X^2 \right\rangle - \langle X \rangle^2$

So the Gaussian is a two parameter distribution, $\theta_1 = \mu$ and $\theta_2 = \sigma$.

The Gaussian distribution is properly normalized since

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int dx \exp\left[-\frac{(x-\mu)}{2\sigma^2} \right] = 1$$

Poisson Distribution The Posson distribution which is derived below is

$$P_X\left(n \mid \theta_1 = \langle X \rangle \equiv \langle n \rangle\right) = \frac{\langle n \rangle^n}{n!} e^{-\langle n \rangle}$$

Here $X = n$ is a random variable which can take all integer values $0, 1, 2, 3, \cdots$, and $P_X\left(n, \langle n \rangle\right)$ is the probability of getting the specific integer n when the mean value of n is given by $\langle n \rangle$. NOTE: The Poisson probability distribution has only a single parameter, the average value $\theta_1 = \langle n \rangle$

The Poisson distribution is properly normalized since

$$\sum_{n=0}^{\infty} \frac{\langle n \rangle^n}{n!} \exp\left[- \langle n \rangle \right] = \exp\left[- \langle n \rangle \right] \sum_{n=0}^{\infty} \frac{\langle n \rangle^n}{n!} = \exp\left[- \langle n \rangle \right] \exp\left[\langle n \rangle \right] = 1$$

Calculating

$$\sum_{n=0}^{\infty} n P_X\left(n \mid \langle n \rangle\right) = \langle n \rangle = \text{the mean or average value}$$

$$\sum_{n=0}^{\infty} \left(n - \langle n \rangle\right)^2 P_X\left(n \mid \langle n \rangle\right) = \langle n \rangle = \text{the mean square variation or variance} = \sigma^2$$

For a Poisson distribution $\sigma =$ square root of the mean number $\langle n \rangle$.

7.3 Independent Identically Distributed (i.i.d.) Random Variables

In the joint probability distribution $P_{\vec{X}}\left(\vec{X} \mid \vec{\theta}\right)$ if the probability for any one X_n taking a given value is completely independent or uncorrelated with the values of any of the other X_n then the joint probability distribution factorized, i.e., it takes the form

$$P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right) = P_{X_1}\left(x_1 \mid \vec{\theta}_1\right) P_{X_2}\left(x_2 \mid \vec{\theta}_2\right) \cdots P_{X_N}\left(x_N \mid \vec{\theta}_N\right)$$

Now if further $\vec{\theta}_1 = \vec{\theta}_2 = \cdots \vec{\theta}_N \equiv \vec{\theta}$ and all the functions $P_{X_1}, P_{X_2}, \cdots, P_{X_N} \equiv P_X$ are the same, e.g. all the same Gaussian or all the same Poisson then we have

$$\begin{aligned} P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right) &= P_X\left(x_1 \mid \vec{\theta}\right) P_X\left(x_2 \mid \vec{\theta}\right) \cdots P_X\left(x_N \mid \vec{\theta}\right) \\ &= \text{Idependent Identitically Distributed ("i.i.d.")} \\ &\text{Also referred to a "homoscedastic"} \end{aligned}$$

7.4 Data Analysis for i.i.d additive noise (homoscedastic)

Consider measuring the value of some single parameter or variable x which is corrupted by additive "i.i.d" noise ε, that is, each measured value of x, $x_{Measured}$ is actually the true value x_{True} plus random noise ε

$$x_{Measured} = x_{True} + \varepsilon$$

with the actual values of ε in each measurement being totally uncorrelated with one another and they have the same variance, σ^2. This type of noise is also termed "homoscedastic". If for example the σ values are different for different measurements then the noise is "heteroscedastic". In both cases the noise is still uncorrelated, i.e., $\langle \varepsilon_n \varepsilon_m \rangle = \sigma^2 \delta_{n,m}$ for the homoscedastic case and $\langle \varepsilon_n \varepsilon_m \rangle = \sigma_n^2 \delta_{n,m}$ (no sum on n) for the heteroscedastic case.

To simplify the notation we will write

$$\begin{aligned} x_{Measured} &\equiv M \\ x_{True} &\equiv T \\ &\text{or} \\ M &= T + \varepsilon \end{aligned}$$

To be specific take the values of ε to follow a zero mean Gaussian distribution, that is we have (again simplifying notation by writing $P(\varepsilon)$ instead of $P_X(\varepsilon \mid \theta = \sigma)$)

$$P(\varepsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\varepsilon^2}{\sigma^2}\right]$$

For $n = 1$ to N measurements since the value of ε in any measurement is completely uncorrelated with the value in any other measurement we have the joint probability distribution, i.e., the probability to get ε_1 for the noise in the first measurement, ε_2 for the noise in the second and so on, is given by

$$P_{joint}(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N) \equiv P_{joint}(\vec{\varepsilon}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\sum_{n=1}^{N} \frac{\varepsilon_n^2}{2\sigma^2}\right]$$

and the value of the n^{th} measurement is

$$M_n = T + \varepsilon_n$$

The expectation value of difference between the measured value and the true value for any single measurement is

$$\langle M_n - T \rangle = \langle \varepsilon_n \rangle = \int d^N \varepsilon \varepsilon_n P_{joint}\left(\vec{\varepsilon}\right) = 0$$

Although "on average" we might expect get zero, the expected variation from 0, i.e., the square root of the variance is

$$\sqrt{\left\langle \left(M_n - T\right)^2 \right\rangle} = \left(\int d^n \varepsilon \varepsilon_n^2 P_{joint}\left(\vec{\varepsilon}\right) \right)^{1/2} = \sigma$$

and the probability of having measurement M_n lie within a range $T - \Delta/2$ to $T + \Delta/2$ with $\Delta \ll \sigma$ is given, from the Gaussian distribution by

$$p \simeq \frac{1}{\sqrt{2\pi}} \frac{\Delta}{\sigma} \ll 1$$

As we all know the "best" thing to do is to take multiple measurements and average to "beat down the noise". Taking the average is the Maximum Likelihood or Least Squares solution as discussed below. Define the average as

$$\begin{aligned} A &= \frac{1}{N} \sum_{n=1}^{N} M_n \\ &= \frac{1}{N} \sum_{n=1}^{N} \left(T + \varepsilon_n\right) \\ &= T + \frac{1}{N} \sum_{n=1}^{N} \varepsilon_n \end{aligned}$$

Again computing the expectation value of $A - T$ we have

$$\langle A - T \rangle = \frac{1}{N} \sum_{n=1}^{N} \langle \varepsilon_n \rangle = 0$$

so still "on average" we expect to get $\langle A - T \rangle = 0$. But now compute the probability of the measured average value A to lie with the range $T - \Delta/2$ to $T + \Delta/2$ in the same was as before, evaluate the expected root mean square difference σ_A between A and T, $\sqrt{\left\langle \left(A - T\right)^2 \right\rangle}$

$$\sigma_A = \sqrt{\left\langle \left(A - T\right)^2 \right\rangle} = \left(\frac{1}{N^2} \sum_{n,m=1}^{N} \langle \varepsilon_n \varepsilon_m \rangle \right)^{1/2}$$

For zero mean i.i.d distributed ε_n we have

$$\langle \varepsilon_n \varepsilon_m \rangle = \left\{ \begin{array}{c} \sigma^2 \text{ for } n = m \\ 0 \text{ for } n \neq m \end{array} \right. = \sigma^2 \delta_{n,m}$$

and so

$$\sigma_A = \sqrt{\left\langle (A-T)^2 \right\rangle} = \frac{\sigma}{\sqrt{N}}$$

As shown in the probability mapping section below, if each measured value follows a Gaussian probability distribution then the average of the measured values also follows a Gaussian probability distribution but with σ for a single measurement replaced by $\sigma/\sqrt{N}$ for the root mean square variation of the average of the measurements. Also, for any i.i.d. case in the limit where $N \gg 1$ the central limit theorem proved below shows that the probability distribution for the average for N large asymptotes to a Gaussian. Hence either way, the probability of measuring A to be within $T - \Delta/2$ and $T + \Delta/2$ with $\Delta \ll \sigma$ is given by

$$p_A \simeq \sqrt{N} \frac{1}{\sqrt{2\pi}} \frac{\Delta}{\sigma}$$

which is $\sqrt{N}$ times greater than for a single measurement. Tthis is why you use many measurements to compute far fewer values, because it "beats down the noise".

8 Biased versus Unbiased Estimators

If T is the true value of a random variable X, then anytime you have

$$f(T) = \langle f(X) \rangle = \int d\nu f(x) P(x) = f(T)$$

then the estimator $f(X)$ is said to be unbiased. In other words, if the average of the estimator yields the true answer then the estimator is said to be unbiased. For example the result in the previous section of $\langle A - T \rangle = 0$ or equivalently $\langle A \rangle = T$ is an example of an unbiased estimator. Note: This required the mean of the additive noise $\langle \varepsilon_n \rangle$ to be zero.

8.1 Biased Estimator Example: The $N-1$ Factor

The classic example of a biased estimator is the computation of σ or σ^2 using the average of the measured values instead of the true value, i.e., compute

$$\frac{1}{N} \sum_n \left(M_n - \frac{1}{N} \sum_m M_m \right)^2$$

Ideally this should be σ^2 since is the average of the square of the deviation of each measurement from the overall average of the measured values, but it isn't.

Assume zero mean additive i.i.d. noise. Noting that $\sum_n T = T\sum_n 1 = NT$, add and subtract T to get

$$\begin{aligned}\frac{1}{N}\sum_n\left(M_n - \frac{1}{N}\sum_m M_m\right)^2 &= \frac{1}{N}\sum_n\left((M_n - T) - \frac{1}{N}\sum_m (M_m - T)\right)^2 \\ &= \frac{1}{N}\sum_n\left(\varepsilon_n - \frac{1}{N}\sum_m \varepsilon_m\right)^2 \\ &= \frac{1}{N}\sum_n\left(\varepsilon_n^2 - \frac{2}{N}\varepsilon_n\sum_m \varepsilon_m + \frac{1}{N^2}\sum_{m,p}\varepsilon_m\varepsilon_p\right)\end{aligned}$$

Take the expectation value and use $\langle\varepsilon_n\varepsilon_m\rangle = \sigma^2\delta_{n,m}$

$$\begin{aligned}\left\langle\frac{1}{N}\sum_n\left(M_n - \frac{1}{N}\sum_m M_m\right)^2\right\rangle &= \frac{1}{N}\sum_n\left(\left\langle\varepsilon_n^2\right\rangle - \frac{2}{N}\sum_m\langle\varepsilon_n\varepsilon_m\rangle + \frac{1}{N^2}\sum_{m,p}\langle\varepsilon_m\varepsilon_p\rangle\right) \\ &= \frac{1}{N}\sum_n\left(\sigma^2 - \frac{2}{N}\sigma^2 + \frac{1}{N^2}\sum_m\sigma^2\right) \\ &= \frac{1}{N}\sum_n\left(\sigma^2 - \frac{2}{N}\sigma^2 + \frac{1}{N^2}N\sigma^2\right) \\ &= \frac{N-1}{N}\sigma^2\end{aligned}$$

But the true value of the variance of ε_n is σ^2, not $\left(\left(N-1\right)/N\right)\sigma^2$ and so this estimator for σ^2 is biased. Noting that the $1/N$ factor is just the overall $1/N$ we started with, we can adjust the estimator to

$$\frac{1}{N-1}\sum_n\left(M_n - \frac{1}{N}\sum_m M_m\right)^2$$

which is unbiased.

Finally note that for N "large" it really doesn't make much of a difference which way you do it.

8.2 Poisson → Gaussian for large $\langle X\rangle$

In the Poisson distribution let

$$n = \langle X\rangle + \Delta n$$

we have

$$P_X\left(n \mid \langle X\rangle\right) = \frac{\langle X\rangle^{\langle X\rangle+\Delta n}}{(\langle X\rangle + \Delta n)!}e^{-\langle X\rangle}$$

The Stirling approximation for the factorial of a large number, say L, is

$$L! \simeq \left(\frac{L}{e}\right)^L \sqrt{2\pi L}$$

where $e = e^1 = 2.71828\cdots$. Substituting the Stirling approximation with $L = \langle X\rangle + \Delta n$ into the Poisson distribution above and putting all the $\langle X\rangle + \Delta n$ terms into the exponent gives

$$\begin{aligned}
P_X\left(n \mid \langle X\rangle \gg 1\right) &\simeq \frac{1}{\sqrt{2\pi}} \exp\left[\begin{array}{c}(\langle X\rangle + \Delta n)\ln\left(\langle X\rangle\right) - (\langle X\rangle + \Delta n)\ln\left(\langle X\rangle + \Delta n\right) \\ +\langle X\rangle + \Delta n - \frac{1}{2}\ln\left(\langle X\rangle + \Delta n\right) - \langle X\rangle\end{array}\right] \\
&\simeq \frac{1}{\sqrt{2\pi}} \exp\left[-\left(\langle X\rangle + \Delta n + \frac{1}{2}\right)\ln\left(1 + \frac{\Delta n}{\langle X\rangle}\right) + \Delta n - \frac{1}{2}\ln\left(\langle X\rangle\right)\right] \\
&\simeq \frac{1}{\sqrt{2\pi\langle X\rangle}} \exp\left[-\left(\langle X\rangle + \Delta n + \frac{1}{2}\right)\ln\left(1 + \frac{\Delta n}{\langle X\rangle}\right)\right] \\
&\text{Taylor expand exponent to 2nd order in } \Delta n \\
&\simeq \frac{1}{\sqrt{2\pi\langle X\rangle}} \exp\left[-\frac{\Delta n}{2\langle X\rangle} + \left(1 - 2\langle X\rangle\right)\frac{\Delta n^2}{4\langle X\rangle^2}\right] \\
&\text{But for } \langle X\rangle \text{ large, } 1/\!\!/\langle X\rangle^2 \text{ is much smaller than } 1/\langle X\rangle \\
&\text{Drop} 1/\langle X\rangle^2 \text{ term} \\
&\simeq \frac{1}{\sqrt{2\pi\langle X\rangle}} \exp\left[-\frac{\Delta n^2 + \Delta n}{2\langle X\rangle}\right]
\end{aligned}$$

The peak of this approximation to $P_X\left(n \mid \langle X\rangle \gg 1\right)$ occurs at

$$\frac{\partial}{\partial \Delta n}\left((\Delta n)^2 + \Delta n\right) = 0 \;\rightarrow\; \Delta n = -\frac{1}{2}$$

We could shift Δn by $1/2$ to have the peak of the Gaussian and the Poisson distributions match, i.e. we could let

$$\Delta n' = \Delta n + \frac{1}{2} = n - \langle X\rangle + \frac{1}{2}$$

With this change of variables we would then have

$$P_X\left(n \mid \langle X\rangle \gg 1\right) \simeq \frac{1}{\sqrt{2\pi\langle X\rangle}} \exp\left[-\frac{\left(n - \langle X\rangle + 1/2\right)^2}{2\langle X\rangle}\right]$$

but now the average is $\langle X\rangle - 1/2$ instead of $\langle X\rangle$ and it gives a worse match, in a least squares sense, between the exact Poisson distribution and its Gaussian approximation. So no shift.

8.3 Derivation of Poisson Distribution

Consider the case where the probability to get one count (on say a photodetector for example) in some short time dt is given by

$$pdt$$

where p is a constant value. Also assume that the probability to get 2 or more counts in a time dt is negligible compare to pdt. The probability of getting n counts in the total time from $t = 0$ to $t = T$, $P(n, T)$ is the sum of two terms.

NOTE: Here T is a time, not be confused with the mean number of counts N above.

The first term is the probability to get n counts in the time $t = 0$ to $T - dt$ and zero counts from $T - dt$ to T which is $P(n, T - dt)(1 - pdt)$. The second term is the probability of getting $n - 1$ counts in the time $t = 0$ to $T - dt$ and 1 count in the time $T - dt$ to T which is given by $P(n - 1, T - dt)\, pdt$. Hence we have

$$P(n, T) = P(n, T - dt)(1 - pdt) + P(n - 1, T - dt)\, pdt$$

Expanding to first order in dt gives a coupled set of first order linear differential equations for $P(n, T)$

$$P(n, T) = P(n, T) - \frac{\partial}{\partial T} P(n, T)\, dt - P(n, T)\, pdt + P(n - 1, T)\, pdt$$

or

$$\frac{\partial}{\partial T} P(n, T) = p\left(P(n - 1, T) - P(n, T)\right) \text{ for } n \geq 1$$

$$\frac{\partial}{\partial T} P(0, T) = -pP(0, T) \text{ for } n = 0$$

Since for $T = 0$ there are zero counts we have the initial conditions

$$P(n, 0) = 0 \text{ for } n \geq 1$$
$$P(0, 0) = 1 \text{ for } n = 0$$

There are various ways to solve this set of equations. The most straightforward is to solve the $n = 0$ equation for $P(0, T)$, then substitute that solution into the $n = 1$ equation and solve for $P(1, T)$ and so on. There is also an elegantly obvious solution if the set of equations are written in matrix form (see below). In any case the solution is

$$P(n, T) = \frac{(pT)^n}{n!} \exp\left[-pT\right]$$

which, upon comparing with $P(n, T)$ given above shows that

$$\langle X \rangle = pT$$

and so we can write

$$P(n, T) \rightarrow P(n) = \frac{\langle X \rangle^n}{n!} \exp\left[-\langle X \rangle\right]$$

8.3.1 Matrix Solution

Define the (infinite length) column vector

$$P(T) = \begin{bmatrix} P(0,T) \\ P(1,T) \\ P(2,T) \\ \vdots \end{bmatrix}$$

and the (infinite $\times$ infinite) matrix

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

then the set of differential equations above for $P(n,T)$ can be combined into the matrix equation

$$\frac{\partial}{\partial T} P(T) = -pP(T) + pGP(T) = -p(I-G)P(T)$$

where I is the (infinite $\times$ infinite) identity matrix and matrix multiplication is understood. The initial conditions become

$$P(T=0) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

Since G is independent of T the solution to the matrix form of the differential equations for $P(n,T)$ is simply

$$\begin{aligned} P(T) &= \exp[-p(I-G)T]P(0) \\ &= \exp[-pT]\exp[GpT]P(0) \end{aligned}$$

Using $\exp[GpT] = \sum_{n=0}^{\infty} (GpT)^n / n!$ and comparing components on both sides yields the standard solution for $P(n,T)$ for the Poisson distribution as given above.

8.4 Characteristic Function

For a continuous variable x the "characteristic function" $Z(\beta)$ of a (properly normalized) probability distribution (density) $P_X(x)$ is the Fourier transform of $P_X(x)$ which is also the expectation value of of $\exp[i\beta x]$,

$$Z(\beta) = \int dx P_X(x) \exp[-i\beta x] = \langle \exp[-i\beta x] \rangle \ \text{:Characteristic function}$$

It follows from this that all the moments of $P_X(x)$ can be found simply by taking derivatives of $Z(\beta)$ with respect to β, including appropriate factors of i, and then setting $\beta = 0$

$$Z(\beta)|_{\beta=0} = \langle \exp[0] \rangle = \langle 1 \rangle = 1$$

$$\left(i\frac{\partial}{\partial \beta} Z(\beta) \right)|_{\beta=0} = \langle x \exp[-i\beta x] \rangle |_{\beta=0} = \langle x \rangle$$

and in general

$$\left(\left(i\frac{\partial}{\partial \beta} \right)^n Z(\beta) \right)|_{\beta=0} = \langle x^n \exp[-i\beta x] \rangle |_{\beta=0} = \langle x^n \rangle$$

Or to write this another way

$$\begin{aligned} Z(\beta) &= \int dx P_X(x) \exp[-i\beta x] \\ &= \sum_{n=0}^{\infty} \frac{(-i\beta)^n}{n!} \int dx P_X(x) x^n \\ &= \sum_{n=0}^{\infty} \frac{(-i\beta)^n}{n!} \langle x^n \rangle \\ &= 1 - i\beta \langle x \rangle - \frac{\beta^2}{2} \langle x^2 \rangle + \frac{i\beta^3}{3!} \langle x^3 \rangle + \cdots \end{aligned}$$

The same thing can be done in the discrete case using Fourier series.

The characteristic function always exists mathematically since probability densities by definition are always normalizable, i.e., their integral over all x can normalized to unity. .

8.5 Maximum Likelihood

What is a good, nominally the best, way to use measured data to determine the values of various parameters? The answer is Maximum likelihood is generally, but not always, the best.

Given the probabilty distribution for a x given the value of a parameter θ

$$P_X(x \mid \theta)$$

the likelihood function is defined as

$$L(\theta \mid k) = P_X(k \mid \theta)$$

where

$$k = \text{the measured or known value of } x$$

The idea is that for any given k the most probable or highest likelihood value for θ is the one that make $P_X(x \mid \theta)$ a maximum or equivalently maximizes

$L(\theta \mid k)$. Hence the "Maximum likelihood" value for θ, θ_{ML}, is the one for which

$$\frac{\partial}{\partial\theta} L(\theta \mid k)\Big|_{\theta=\theta_{ML}} = 0$$

In many cases it is more convenient to work with the logarithm of the likelihood function and so often one sees

$$\frac{\partial}{\partial\theta} \ln\left[L(\theta \mid k)\right]\Big|_{\theta=\theta_{ML}} = 0$$

Generally there must be at least as many values as there are parameters and so in general Maximum likelihood takes the form

$$\vec{\partial}_\theta \ln\left[L\left(\vec{\theta} \mid \vec{k}\right)\right]\Big|_{\vec{\theta}=\vec{\theta}_{ML}} = 0$$

where $\vec{\theta}$ just indicates a string or list of parameters $(\theta_1, \theta_2, \cdots)$ and $\vec{k}$ indicates a list or string of data values $(k_1, k_2, \cdots)$.

Note that nominally $L(\theta \mid k)$ is not normalized w.r t. θ and so $L(\theta \mid k)$ is not directly the probability distribution for θ given k.

In the next section we show that for uncorrelated Gaussian noise Maximum likelihood is equivalent to Least Squares.

8.6 Least Squares

8.6.1 Maximum likelihood and Least Squares

In this section we show how Maximum likelihood for measurments corrupted by additive Gaussian uncorrelated or "white" noise leads to least squares first for a single parameter and then for a set of parameters with correlated noise, i.e, "non-white" noise.

Single Parameter Gaussian White Noise Consider making N measurements of a variable or parameter θ where each measurement is corrupted by additive zero-mean noise ε with the same rms or σ width Gaussian distribution. That is, each measured value

$$x_n = \theta + \varepsilon_n$$

where the joint probability distribution for the ε_n values is

$$P_{Joint}(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\sum_{n-1}^{N} \frac{\varepsilon_n^2}{2\sigma^2}\right]$$

Since $\varepsilon_n = x_n - \theta$, the joint probability distribution for the measurements given θ is given by

$$P_{Joint}(x_1, x_2, \cdots, x_N \mid \theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\sum_{n-1}^{N} \frac{(x_n - \theta)^2}{2\sigma^2}\right]$$

Substituting actual measured values or known values $k_1, k_2, \cdots$ for $x_1, x_2, \cdots$ gives for the natural log of the likelihood function

$$\ln\left[L\left(\theta \mid k_1, k_2, \cdots, k_N\right)\right] = -\sum_{n-1}^{N} \frac{\left(k_n - \theta\right)^2}{2\sigma^2} - \frac{N}{2}\ln\left[2\pi\sigma^2\right]$$

Taking the derivative w.r.t. θ and setting it equal to zero is obviously equivalent to doing Least Squares (LS) on the set of values $(k_n - \theta)$ which gives, after some algebra, the Maximum likelihood solution

$$\theta_{ML} = \frac{1}{N}\sum_{n=1}^{N} k_n = \theta_{LS} \text{ in this case}$$

The point here is that Least Squares is the Maximum likelihood solution for the average value of a parameter or variable corrupted by additive uncorrelated, i.e., "white", Gaussian distributed noise.

Multiparameter Corrlated Gaussian Noise Consider a set of N variables or parameters $\theta = (\theta_1, \theta_2, \cdots, \theta_N)$ which are linearly related to a set of N measureable quantities $x = (x_1, x_2, \cdots, x_M)$ by a given know matrix A where $M \geq N$. Each x_m could be different type of measured quantity or some could be repeated measurements of the same quantity. Treating θ and x as column vectors and using "$\cdot$" to indicate matrix multiplication the relation between x and θ can be written as

$$x = A \cdot \theta$$

or

$$x_m = A_{m,n}\theta_n$$

where we are using the "Einstein summation convention".

Suppose the measurements of the elements of x are corrupted by additive zero-mean Gaussian distributed noise $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_M)$ with a specified covariance matrix C, then, treating ε as a column vector, the probability distribution for the ε_m is given by

$$P\left(\varepsilon\right) = P\left(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_M\right) = \frac{1}{\left(2\pi\right)^{M/2}\sqrt{\det\left[C\right]}}\exp\left[-\frac{1}{2}\varepsilon^T \cdot C^{-1} \cdot \varepsilon\right]$$

The covariance matrix is defined as the expectation value of the product of two of the ε_m,

$$\left\langle \varepsilon_p \varepsilon_q \right\rangle \equiv \int d^M \varepsilon \varepsilon_p \varepsilon_q P\left(\varepsilon\right)$$

Direct evaluation the integral gives

$$C_{p,q} = \left\langle \varepsilon_p \varepsilon_q \right\rangle$$

Note that since $\langle \varepsilon_p \varepsilon_q \rangle = \langle \varepsilon_q \varepsilon_p \rangle$, C and hence C^{-1} are both symmetric matrices: $C^T = C$ and $\left(C^{-1}\right)^T = C^{-1}$.

Accounting for the additive noise the measured data or known data k is given by

$$k = x + \varepsilon$$
$$\text{or}$$
$$k_m = x_m + \varepsilon_m$$

Letting $x = A \cdot \theta$, solving for ε and substituting it into $P(\varepsilon)$ we get the likelihood function

$$\begin{aligned} P(k - A \cdot \theta) &= L(\theta \mid k) \\ &= \frac{1}{(2\pi)^{M/2} \sqrt{\det[C]}} \exp\left[-\frac{1}{2}(k - A \cdot \theta)^T \cdot C^{-1} \cdot (k - A \cdot \theta)\right] \end{aligned}$$

and so

$$\ln[L(\theta \mid k)] = -\frac{1}{2}(k - A \cdot \theta)^T \cdot C^{-1} \cdot (k - A \cdot \theta) - \frac{1}{2}\det[C] - \frac{M}{2}\ln(2\pi)$$

Looking at the first it is clear that minimizing $\ln[L(\theta \mid k)]$ w.r.t. θ is equivalent to performing least squares on $(k - A \cdot \theta)^T \cdot C^{-1} \cdot (k - A \cdot \theta)$. Taking the derivative w.r.t. θ_n, using the fact C^{-1} is a symmetric matrix and and setting the result to zero for $\theta = \theta_{ML}$ gives

$$A^T \cdot C^{-1} \cdot (k - A \cdot \theta_{ML}) = 0$$

Rearranging gives

$$A^T \cdot C^{-1} \cdot A \cdot \theta_{ML} = A^T \cdot C^{-1} \cdot k$$

$A^T \cdot A$ is a square matrix and assuming it is invertible we have finally for the Maximum likelihood solution (often called the Maximum likelihood Estimate = MLE)

$$\begin{aligned} \theta_{ML} &= \left(A^T \cdot C^{-1} \cdot A\right)^{-1} \cdot A^T \cdot C^{-1} \cdot k \\ &\equiv S \cdot k \end{aligned}$$

8.6.2 Noise Propagation in Least Squares

From the form of $P(\varepsilon)$ we have

$$\langle \varepsilon_m \rangle = \int d^M \varepsilon \ \varepsilon_m P(\varepsilon) = 0$$
$$\langle \varepsilon_n \varepsilon_m \rangle = \int d^M \varepsilon \ \varepsilon_n \varepsilon_m P(\varepsilon) = C_{n,m}$$

which implies

$$\langle k \rangle = \langle x + \varepsilon \rangle = x$$

And so x which is the "true" value of k is also the mean value of k The "true" value of θ is

$$\theta_{True} = S \cdot \langle k \rangle = S \cdot x$$

and so

$$\langle \theta_{ML} - \theta_{True} \rangle = S \cdot \langle k - x \rangle = 0$$

or

$$\langle \theta_{ML} \rangle = \langle \theta_{True} \rangle$$

and so the estimate is unbiased.

The covariance of the deviation of θ_{ML} from the true value θ_{True} is given by

$$\begin{aligned} \langle (\theta_{ML,n} - \theta_{True,n}) (\theta_{ML,m} - \theta_{True,m}) \rangle &= S_{n,p} S_{m,q} \left\langle (k-x)_p (k-x)_q \right\rangle \\ &= S_{n,p} S_{m,q} \langle \varepsilon_p \varepsilon_q \rangle \\ &= S_{n,p} S_{m,q} C_{p,q} \\ &= \left(S \cdot C \cdot S^T \right)_{n,m} \end{aligned}$$

8.6.3 Fourier Series/Transform = Least Squares

Consider the Fourier series relation

$$f(x) = \sum_{n=-\infty}^{+\infty} a_n \exp\left[i n 2\pi \frac{x}{P} \right]$$

To find a_n using Least Squares we want to minimize

$$\int_{-P/2}^{+P/2} dx \left(f(x) - \sum_{m=-\infty}^{+\infty} a_m \exp\left[i m 2\pi \frac{x}{P} \right] \right)^2$$

w.r.t. a_n. Taking the derivative w.r.t. a_n for any particular n, using $da_m/da_n = \delta_K(m,n)$ and setting the result to zero gives

$$\int_{-P/2}^{+P/2} dx \left(f(x) - \sum_{m=-\infty}^{+\infty} a_m \exp\left[i m 2\pi \frac{x}{P} \right] \right) \exp\left[i n 2\pi x / P \right] = 0$$

Rearranging gives

$$\begin{aligned} \int_{-P/2}^{+P/2} dx \exp\left[i 2\pi n \frac{x}{P} \right] f(x) &= \sum_{m=-\infty}^{+\infty} a_m \underbrace{\int_{-P/2}^{P/2} dx \exp\left[i (n+m) 2\pi \frac{x}{P} \right]}_{= P \delta_K(-n,m)} \\ &= P a_{-n} \end{aligned}$$

Dividing by P and replacing n with $-n$ gives

$$a_n = \frac{1}{P}\int_{-P/2}^{+P/2} dx \exp\left[-in2\pi\frac{x}{P}\right] f(x)$$

The same thing works the same way for Fourier transforms but uses the Fourier representation of the Dircac delta function. Hence the coefficients a_n for Fourier series and $\tilde{f}(\beta)$ for Fourier transforms are Least Squares solutions. The formulae for $f(x)$ in terms of a_n for Fourier series and in terms of $\tilde{f}(\beta)$ for Fourier transforms are also Least Squares solutions.

8.7 Central Limit Theorem

The joint probability density or distribution for a set of N uncorrelated random variables $x_1, x_2, \cdots, x_N$ is given by the product of the probability distributions for each random variable

$$P_{Joint}(x_1, x_2, \cdots, x_N) = P_1(x_1) P_2(x_2) \cdots P_N(x_N)$$

To avoid notational clutter we have replaced $P_{X_i}(x_i)$ with $P_i(x_i)$.

A good example of this type of joint probability distribution is repeated measurements of a some parameter or variable, such as position, energy, momentum, voltage, etc, in which the measured value is affected by random noise. In this case all the individual probabilities are equal

$$P_1(x) = P_2(x) = \cdots = P_N(x) \equiv P(x)$$

Such a case is often labeled "i.i.d." standing for independent identically distributed. (NOTE: labeled is US spelling, labelled is UK spelling)

For multiple measurements of the same variable we expect the average of the measured values to better represent the true or noise-free value of the variable than any single measurement. Hence, consider the probability distribution for the the average value

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

of all the measured values and use the probability mapping technique from above to find the probability distribution for $\bar{x}$

$$P_{\bar{X}}(\bar{x}) = \int dx_1 \cdots dx_N \delta_D\left(\bar{x} - \frac{1}{N}\sum_{n=1}^{N} x_n\right) P(x_1) \cdots P(x_N)$$

Using the Fourier representation of the Dirac delta function $\delta_D(\cdots)$ we get

$$\begin{aligned}
P_{\bar{X}}(\bar{x}) &= \frac{1}{2\pi}\int d\omega \exp[i\omega\bar{x}] \int dx_1 \cdots dx_N \exp\left[-i\frac{\omega}{N}\sum_{n=1}^{N} x_n\right] P(x_1)\cdots P(x_N) \\
&= \frac{1}{2\pi}\int d\omega \exp[i\omega\bar{x}] \left(\int dx \exp[-i\omega x/N] P(x)\right)^N \\
&= \frac{1}{2\pi}\int d\omega \exp[i\omega\bar{x}] \left(Z\left(\frac{\omega}{N}\right)\right)^N
\end{aligned}$$

where

$$Z\left(\frac{\omega}{N}\right)$$

is the characteristic function of $P(x)$ defined above. Use

$$\left(Z\left(\frac{\omega}{N}\right)\right)^N = \exp\left[N\ln\left(Z\left(\frac{\omega}{N}\right)\right)\right]$$

In the limit as $N\to\infty$, for finite ω we have $\omega/N\to 0$. But from the properties of the characteristic function discussed above, its Taylor expansion around $\omega/N = 0$ is given, to second order, by

$$Z\left(\frac{\omega}{N}\right) \simeq 1 - i\langle x\rangle\frac{\omega}{N} - \frac{1}{2}\langle x^2\rangle\left(\frac{\omega}{N}\right)^2$$

Substituting this in the logarithm and using the Taylor series for $\ln[\cdots]$ to second order we get

$$\begin{aligned}
Z\left(\frac{\omega}{N}\right)^N &= \exp\left[N\ln\left[Z\left(\frac{\omega}{N}\right)\right]\right] \\
&\simeq \exp\left[N\ln\left[1 - i\langle x\rangle\frac{\omega}{N} - \frac{1}{2}\langle x^2\rangle\left(\frac{\omega}{N}\right)^2\right]\right] \\
&\simeq \exp\left[-i\langle x\rangle\omega - \frac{1}{2}\left(\langle x^2\rangle - \langle x\rangle^2\right)\frac{\omega}{N}^2\right]
\end{aligned}$$

Substituting this in the integral for $P_{\bar{X}}(\bar{x})$ yields

$$\begin{aligned}
P_{\bar{X}}(\bar{x}) &\simeq \frac{1}{2\pi}\int d\omega \exp\left[i\omega(\bar{x} - \langle x\rangle) - \frac{1}{2}\left(\langle x^2\rangle - \langle x\rangle^2\right)\frac{\omega^2}{N}\right] \\
&\simeq \frac{1}{\sqrt{2\pi\left(\frac{\langle x^2\rangle - \langle x\rangle^2}{N}\right)}} \exp\left[-\frac{(\bar{x} - \langle x\rangle)^2}{2\left(\frac{\langle x^2\rangle - \langle x\rangle^2}{N}\right)}\right] \quad \text{:Central Limit Theorem}
\end{aligned}$$

Hence we get a Gaussian distribution for $\bar{x}$ with mean $\langle x\rangle$ and rms variation

$$\bar{\sigma} = \sqrt{\frac{\langle x^2\rangle - \langle x\rangle^2}{N}} = \frac{\sigma}{\sqrt{N}}$$

where σ is the rms variation of x,

$$\sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

The fact that $P_{\bar{X}}(\bar{x})$ has the Gaussian form given above is the Central Limit Theorem.

8.8 Bayes Theorem

Here I give two simple "proofs" of Bayes theorem.

The first proof: Consider a probability which depends on two discrete random variables A and B,

$$P_{A,B}(a,b)$$

A and B can be different types of variable or the same type of variable partitioned into two distinct sets. $P_{A,B}(a,b)$ is the probability of getting the particular values a and b for A and B.

Normalization, we must have

$$\sum_{a,b} P_{A,B}(a,b) = 1$$

The probability for getting the value a for A independent of the value of b is

$$P_A(a) = \sum_b P_{A,B}(a,b)$$

and the probability for getting the value b for B independent of the value of a is

$$P_B(b) = \sum_a P_{A,B}(a,b)$$

Both $P_A(a)$ and $P_B(b)$ are properly normalized since

$$\sum_a P_A(a) = \sum_a \sum_b P_{A,B}(a,b) = \sum_{a,b} P_{A,B}(a,b) = 1$$

and

$$\sum_b P_B(b) = \sum_b \sum_a P_{A,B}(a,b) = \sum_{a,b} P_{A,B}(a,b) = 1$$

The conditional probability of getting a for A given a particular value b for B, $P(a|b)$ and the conditional probability for getting b for B given a particular value a for A, $P(b|a)$, are both nominally just $P_{A,B}(a,b)$. But neither $P(a|b) = P_{A,B}(a,b)$ nor $P(b|a) = P_{A,B}(a,b)$ is properly normalized. We need to have $\sum_a P(a|b) = 1$ with b fixed and $\sum_b P(b|a) = 1$ with a fixed. But, normalization

is easy to fix and we get

$$P(a|b) = \frac{P_{A,B}(a,b)}{\sum_a P_{A,B}(a,b)} = \frac{P_{A,B}(a,b)}{P_B(b)}$$

and

$$P(b|a) = \frac{P_{A,B}(a,b)}{\sum_b P_{A,B}(a,b)} = \frac{P_{A,B}(a,b)}{P_A(a)}$$

which are obviously properly normalized.

Combining these two results gives Bayes theorem

$$P(a|b) = \frac{P_A(a)}{P_B(b)} P(b|a) \quad \text{:Bayes Theorem}$$

The second proof: Consider two sets, call them A and B. Here "set" is meant in the mathematical sense as a collection of objects. Suppose that some of the objects are members of both set A and set B. These particular objects then are in the set which corresponds to the "intersection" of A and B. This intersection set is indicated as $A \cap B$. Let N_A be the total number of objects or members in A and N_B be the total number of objects or members of B. The total number of members of the intersection $A \cap B$ will be indicated by $N_{A \cap B}$.

If a member of set B is chosen randomly (meaning equal probability of picking any member of B) then the probability that it is also in A is given by

$$P_{A|B} = \frac{N_{A \cap B}}{N_B}$$

and if a member of set A is chosen randomly (meaning equal probability of picking any member of A) then the probability that it is also in B is given by

$$P_{B|A} = \frac{N_{A \cap B}}{N_A}$$

Multiply and divide $P_{A|B}$ by N_A/N where N is the total number of elements in both sets, counting any member of $A \cap B$ only once not twice and we have

$$\begin{aligned} P_{A|B} &= \frac{N_{A \cap B}}{N_B} \frac{N_A}{N} \frac{N}{N_A} \\ &= \frac{N_{A \cap B}}{N_A} \frac{N_A/N}{N_B/N} \end{aligned}$$

But N_A/N is the probability of picking a member of A randomly out of the union or combination of both sets, call it P_A and N_B/N is the probability of picking a member of B randomly out of the unrion or combination of both sets, call it P_B. Hence

$$P_{A|B} = \frac{P_{B|A} P_A}{P_B}$$

which again is Bayes theorem.

NOTE: Probability problems can be rather "gnarly" because they are usually stated in words and what is meant by the words in any given case can often be rather ambiguous. Hence I prefer the second derivation because it is very clear what is meant by the words.

8.9 Benfords Law

One might think that in "random" cases the probability that the first (nonzero) digit of a number is the same for all digits from 1 to 9. This is actually not the case for a wide range of "random" probability distributions. A result that is actually very interesting.

In 1881 American astronomer Simon Newcomb noticed that the pages of logarithms for numbers which started with the digit 1 were more used, worn and dirty than those that started with the digit 2, which were more used, worn and dirty than those that started with the digit 3, etc. In 1938 physicist Frank Benford analyzed a wide range of "random" data sets including: The surface areas of 335 rivers, The sizes of 3259 US populations, 104 physical constants, 1800 molecular weights, 5000 entries from a mathematical handbook, 308 numbers contained in an issue of Reader's Digest, The street addresses of the first 342 persons listed in American Men of Science and 418 death rates. All obeyed pretty much the same probability distribution with respect to the first digit of the numbers involved: The probability of the first digit being a 1 was around 30 % and decreased to around 5 % for the probability that the first digit was a 9. There are various serious mathematical proofs that this should be so. Here I give an argument (not a proof) based on the fact that probability distributions for continuous variables are densities and hence must have units of the reciprocal of the their argument. That is, $P_X(x)$ must have units of $1/x$.

The official statement of Benfords law is that the first digit d of a number occurs with probability

$$P(d) = \log[1+d] - \log[d] = \log\left[1+\frac{1}{d}\right] \text{ :Benfords Law}$$

where just for clarity "log" here is base 10.

The probability density for the value of a continuous "random" variable x, where x has units, must have units of $1/x$ and so at a minimum $P_X(x)$ must be proportional to $1/x$, that is,

$$P_X(x) \sim \frac{1}{x}$$

Note that since $P_X(x)\,dx$ is unitless a change of the units of x doesn't change the probability for x to be in a given range, $dx \to kdx$ and $P_X(x) \to (1/k)\,P_X(x)$ and so $P_X(x)\,dx \to P_X(x)\,dx$.

The probability for the first digit of x to be d where $d = 1, 2, \cdots, 9$ is given

by

$$\begin{aligned} P(d) &= \frac{\int_d^{d+1} dx P_X(x)}{\int_1^{10} dx P_X(x)} \\ &= \frac{\int_d^{d+1} dx/x}{\int_1^{10} dx/x} \\ &= \frac{\log[d+1] - \log[d]}{\log[10] - \log[1]} \\ &= \log[d+1] - \log[d] \\ &= \log\left[1 + \frac{1}{d}\right] \end{aligned}$$

The above argument is of course not a proof of Benfords law in any strict mathematical sense, it is simply an argument indicating the law should not be unexpected. Clearly the probabililty decreases with increasing values of d. The numerical values of the probability for $d = 1$ to 9 are given by 0.30103, 0.176091, 0.124939, 0.09691, 0.0791812, 0.0669468, 0.0579919, 0.0511525, 0.0457575.

8.10 Fisher Information

The Fisher information (matrix) is defined as

$$F_{ij} = \int dv \left(\frac{\partial}{\partial \theta_i} \ln\left(P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right)\right)\right)\left(\frac{\partial}{\partial \theta_j} \ln\left(P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right)\right)\right) P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right)$$

8.10.1 Fisher Information for a Gaussian

To get a sense of what the Fisher information (matrix) means consider a simple Gaussian distribution where $\mu = \theta$ corresponds to the mean value of x, i.e.,

$$P_X(x \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2}\right]$$

Defining the "expectation value" of any function of x, $f(x)$ by

$$\langle f(x) \rangle = \int_{-\infty}^{+\infty} dx f(x) P_X(x \mid \theta)$$

Direct evaluation of the integrals gives

$$\begin{aligned} \langle 1 \rangle &= \int_{-\infty}^{+\infty} dx P_X(x \mid \theta) = 1 \\ \langle x \rangle &= \int_{-\infty}^{+\infty} dx x P_X(x \mid \theta) = \theta \\ \left\langle (x - \langle x \rangle)^2 \right\rangle &= \left\langle x^2 \right\rangle - \langle x \rangle^2 = \int_{-\infty}^{+\infty} dx (x-\theta)^2 P_X(x \mid \theta) = \sigma^2 \end{aligned}$$

The first result shows that $P_X(x \mid \theta)$ is normalized to unity as required for probability densities. Since there is only one θ the Fisher information is a single number and direct evaluation of the integral

$$\begin{aligned} F &= \int_{-\infty}^{+\infty} dx \left(\frac{\partial}{\partial \theta} \ln \left(P_X(x \mid \theta) \right) \right)^2 P_X(x \mid \theta) \\ &= \frac{1}{\sigma^2} \end{aligned}$$

Next consider the case where there are N values in both strings or vectors $\vec{x}$ and $\vec{\theta}$ and the θ_n are the mean values of the x_n and that the distribution is again Gaussian with covariance matrix C, i.e.,

$$C_{i,j} = \langle (x_i - \theta_i)(x_j - \theta_j) \rangle$$

The probability density for this case is given by

$$\begin{aligned} P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right) &= \frac{\sqrt{\det[M]}}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2} \left(\vec{x} - \vec{\theta}\right) \cdot M \cdot \left(\vec{x} - \vec{\theta}\right) \right] \\ &= \frac{\sqrt{\det[M]}}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2} x_i M_{i,j} x_j \right] \end{aligned}$$

where all x_n are assumed to range from $-\infty$ to $+\infty$. It is understood that doubly repeated indices are summed over the relevant range, i.e.,

$$x_i M_{i,j} x_j \equiv \sum_{i,j=1}^{N} x_i M_{i,j} x_j = x^T \cdot M \cdot x$$

In the last equality think of x as the column vector form of $\vec{x}$, x^T as the row vector form of $\vec{x}$ and

$$M = C^{-1} = \text{Matrix inverse of } C$$

$$\det[M] = \text{Determinant of matrix } M$$

NOTE: Only the symmetric part of M contributes to $x^T \cdot M \cdot x$ and thus we might as well take M to be a symmetric matrix from the start. Also, for this Gaussian form of $P_{\vec{X}}\left(\vec{x} \mid \vec{\theta}\right)$ to be normalizable, M must be a "positive definite" matrix which means $x^T \cdot M \cdot x > 0 \ \forall x$, i.e, for any and all of the $x_n = -\infty$ to $+\infty$. Equivalently "postitive definite" here means all the eigenvalues of M are greater than zero.

Direct evaluation of the Fisher information matrix for the Gaussian distribution above gives

$$F_{i,j} = M_{i,j}$$

and so

$$F = \text{Inverse of the covariance matrix (in this case)}$$

Obviously for different probability distributions F will have a different form but the above result indicates the fundamental meaning of the Fisher Information.

8.10.2 Maximizing the Fisher Information

In the case where F equals the single value $1/\sigma^2$ it is obvious that maximizing F is equivalent to minimizing σ. But when F is a matrix what does "maximizing a matrix" mean? It could mean maximizing $\det[F]$ or it could be maximizing $\mathrm{tr}[F]$ or possibly $\mathrm{tr}\left[F^{-1}\right]$ where "tr" means the trace (= sum of the diagonal elements) or...? For a particular problem it may be obvious what to pick but the ambiguity has lead to a "measure" for F which can be smoothly interpolated between the determinant and the trace. Here "measure" means a single number indicating the "goodness" of F. That "measure" ϕ_p for F an $N \times N$ matrix is defined as

$$\phi_p[F] = \mathrm{tr}\left[\frac{1}{N}F^p\right]^{1/p} = \left(\mathrm{tr}\left[\frac{1}{N}F \cdot F \cdot \cdots \cdot F\right]\right)^{1/p}$$

where there are p factors of F in the last expression. For $p = 1$ this is the trace of F normalized by the number of diagonal elements and for $p = -1$ it's the trace of F^{-1} again normalized to the number of diagonal elements. We now show that

$$\phi_0[F] = \det[F]^{1/N}$$

and so $\phi_p[F]$ interpolates smoothly between a wide range of possible "measures" for F.

Let I be the $N \times N$ identity matrix so that $\mathrm{tr}[I] = N$, and consider

$$\begin{aligned}
\lim_{p\to 0}\left(\phi_p[F]\right) &= \lim_{p\to 0}\left(\mathrm{tr}\left[\frac{1}{N}F^p\right]^{1/p}\right) \\
&= \lim_{p\to 0}\left(\mathrm{tr}\left[\frac{1}{N}\exp\left[p\ln[F]\right]\right]^{1/p}\right) \\
&= \lim_{p\to 0}\left(\mathrm{tr}\left[\frac{1}{N}\left(I + p\ln[F]\right)\right]^{1/p}\right) \\
&= \lim_{p\to 0}\left(\frac{1}{N}\left(\mathrm{tr}[I] + p\mathrm{tr}\left[\ln[F]\right]\right)^{1/p}\right) \\
&= \lim_{p\to 0}\left(\left(1 + \frac{p}{N}\mathrm{tr}\left[\ln[F]\right]\right)^{1/p}\right) \\
&= \lim_{p\to 0}\left(\exp\left[\frac{p}{N}\mathrm{tr}\left[\ln[F]\right]\right]^{1/p}\right) \\
&= \exp\left[\frac{1}{N}\mathrm{tr}\left[\ln[F]\right]\right] \\
&= \left(\exp[\mathrm{tr}\left[\ln[F]\right]\right)^{1/N} \\
&= \det[F]^{1/N}
\end{aligned}$$

where we have repeatedly used $\exp[p \times \mathit{something}] \simeq 1 + (p \times \mathit{something})$ for

$p \to 0$. In the end we used the relation

$$\det[M] = \exp[\mathrm{tr}[\ln[M]]]$$

which is proved in the Introduction.

8.11 Probability Mapping

Let $P_A(a)$ be a discrete probability distribution, that is, $P_A(a)$ is the probability that the discrete random variable A takes the particular discrete value a. Here we are not concerned directly with the θ parameters and so to avoid clutter they are suppressed notationally. Of course for $P_A(a)$ to be a probability distribution it must be properly normalized, i.e.,

$$\sum_a P_A(a) = 1$$

The notation $\sum_a$ indicates summing over all possible values of a.

Suppose we want to determine the probability that some function of a, $f(a)$, takes the value $b = f(a)$. Obviously we need to sum up all the probabilities for a such that $f(a) = b$. This can be done included a constraint or filter in the sum over a. Letting

$$\delta_K(u,v) = \begin{cases} 1 \text{ for } u = v \\ 0 \text{ for } u \neq v \end{cases}$$

be the Kronecker delta function, we have, obviously

$$P_B(b) = \sum_a \delta_K(b, f(a)) P_A(a)$$

Given the definition of $\delta_K(u,v)$ the sum clearly includes only those probabilities for which $b = f(a)$.

Normalization: Given all possible values of A then $f(A)$ gives all possible values for B If we sum over all possible values of B then there will always be at least one value of A for which $b = f(a)$ and $P_B(b)$ is properly normalized

$$\begin{aligned} \sum_b P_B(b) &= \sum_b \sum_a \delta_K(b, f(a)) P_A(a) \\ &= \sum_a P_A(a) \\ &= 1 \end{aligned}$$

On the other hand if only a subset of all the possible values of B is considered then $P_B(b)$ is just the probability for the occurence of that subset and it is not properly normalized over that subset. Sometimes the mapping itself can be specific in that it picks out only a subset of the a values in which case again $P_B(b)$ is just the probability for that particular occurence and it is not properly normlized.

In the continuous case the sum should be replaced by an integral and the Kronecker delta function by a Dirac delta function,δ_D. Thus if we have $P_X(x)$ as the probability density for continuous random variable X to take the value x then

$$P_Y(y) = \int dx \delta_D(y - f(x)) P_X(x)$$

is the probability for random variable Y to take the value $y = f(x)$. Here

$$\delta_D(x) = \text{Dirac delta function}$$

Since in this case X and hence Y are both continuous variables, $P_X(x)$ must have units of $1/X$ and $P_Y(y)$ must have units of $1/Y$, but since by definition dx had units of X and $\delta_D(y - f(x))$ has units of $1/Y$, it follows that the units of $P_Y(y)$ are correct.

The same argument about normalization holds here. Integrating $P_Y(y)$ over all possible values of y gives

$$\int dy P_Y(y) = 1$$

$P_Y(y)$ is not properly normalized if only a subset of the possible values of y is used or if the mapping itself picks out a particular subset of x values.

The same technique can be used to map a continuous distribution over x to a discrete one over y by replacing the Dirac delta function in the integral with an appropriate sum of combinations of Heaviside step functions. A single Heaviside step function is defined by

$$\theta_H(u) = \begin{cases} 1 \text{ for } u > 0 \\ 0 \text{ for } u < 0 \end{cases}$$

There is some ambiguity in how to assign the value of $\theta_H(0)$. Often the value chosen makes no difference but in cases where it does the appropriate value to use is often defined by the particular problem at hand. Using combinations of θ_H to do a given desired mapping yields a sum of integrations over specific ranges of x and hence the resulting discrete probability distribution does not and should not have units of $1/Y$. This is because the Heaviside step function, unlike the Dirac delta function, is unitless.

Simple example: What is the probability for x to lie in the range $x_0 \leq x \leq x_1$. This is given by

$$\int dx \theta_H(x - x_0) \theta_H(x_1 - x) P_X(x) = \int_{x_0}^{x_1} dx P_X(x)$$

Note: This mapping is picking out a particular probability, it is not normalized unless $\int_{x_0}^{x_1} dx P_X(x) = 1$.

8.12 Strong Law of Large Numbers via Probability Mapping

Here we use the approach to probability mapping discussed above to prove the Strong Law of Large Numbers. There is also a weak law which I won't discuss.

Consider the average of N i.i.d. (independent identically distributed) random variables X_n

$$A = \frac{1}{N}\left(X_1 + X_2 + \cdots + X_N\right)$$

Since the X_n are i.i.d. the joint probability distribution for getting the value x_n for X_n for $n = 1, 2, \cdots, N$ is the product of the individual probabilities, i.e,.

$$P_X\left(x_1, x_2, \cdots, x_N\right) = P_X\left(x_1\right) P_X\left(x_2\right) \cdots P_X\left(x_n\right)$$

Using probability mapping the probability distribution for the random variable A to take the value a is given by

$$P_A\left(a\right) = \int dx_1 dx_2 \cdots dx_n \delta_D\left(a - \frac{1}{N}\sum_n x_n\right) P_X\left(x_1, x_2, \cdots, x_N\right)$$

Using the Fourier representation of the Dirac delta function

$$\delta_D\left(x\right) = \int \frac{d\beta}{2\pi} \exp\left[i\beta x\right]$$

we get

$$\begin{aligned} P_A\left(a\right) &= \int \frac{d\beta}{2\pi} e^{i\beta a}\left(\int dx e^{-i\beta x/N} P_X\left(x\right)\right)^N \\ &= \int \frac{d\beta}{2\pi} e^{i\beta a} Z\left(\beta\right)^N \end{aligned}$$

where $Z\left(\beta\right)$ is the characteristic function defined above. Using the series expansion of $Z\left(\beta\right)$ given above we have

$$\begin{aligned} P_A\left(a\right) &= \int \frac{d\beta}{2\pi} e^{i\beta a}\left(1 - i\frac{\beta}{N}\left\langle x\right\rangle - \frac{\beta^2}{2N^2}\left\langle x^2\right\rangle + \frac{i\beta^3}{3!N^3}\left\langle x^3\right\rangle + \cdots\right)^N \\ &= \int \frac{d\beta}{2\pi} e^{i\beta a}\left(1 - \frac{1}{N}\left(i\beta\left\langle x\right\rangle + \frac{\beta^2}{2N}\left\langle x^2\right\rangle - \frac{i\beta^3}{3!N^2}\left\langle x^3\right\rangle + \cdots\right)\right)^N \end{aligned}$$

But in the limit as $N \to \infty$

$$\left(1 + \frac{x}{N}\right)^N \to e^x$$

and so in the limit as $N \to \infty$

$$P_A\left(a\right) \to \int \frac{d\beta}{2\pi} e^{i\beta a - i\beta\left\langle x\right\rangle} = \delta\left(a - \left\langle x\right\rangle\right)$$

So in the limit as $N \to \infty$ the probability density for a is a Dirac delta function of $a - \left\langle x\right\rangle$ and hence the probability to get any value for a other than $\left\langle x\right\rangle$ is zero. This is the strong law of large numbers.

8.13 Random Walk via Probability Mapping

$$P_S(s) = \frac{1}{2}\left(\delta_K(s,+1) + \delta_K(s,-1)\right)$$

Calculate the probability $P_X(x)$ of being at position x after taking N steps where x of course is an integer. Use probability mapping

$$P_X(x) = \sum_{\{s\}} \delta_K\left(x, \sum_{n=1}^{N} s_n\right) P_S(s_1) P_S(s_2) \cdots P_S(s_N)$$

where $\sum_{\{s\}} = \sum_{s_1=-1}^{+1} \sum_{s_2=-1}^{+1} \cdots \sum_{s_N=-1}^{+1}$. Use the Fourier representation of the Kronecker delta function

$$\delta_K\left(x, \sum_{n=1}^{N} s_n\right) = \frac{1}{2\pi} \int_0^{2\pi} d\omega \exp\left[i\omega\left(x - \sum_{n-1}^{N} s_n\right)\right]$$

then

$$\begin{aligned}
P_X(x) &= \frac{1}{2\pi} \int_0^{2\pi} d\omega \exp[i\omega x] \sum_{\{s\}} \exp\left[-i\omega \sum_{n=1}^{N} s_n\right] P_S(s_1) P_S(s_2) \cdots P_S(s_N) \\
&= \frac{1}{2\pi} \int_0^{2\pi} d\omega \exp[i\omega x] \left(\sum_{s=-1}^{+1} \exp[-i\omega s] P_S(s)\right)^N \\
&= \frac{1}{2\pi} \int_0^{2\pi} d\omega \exp[i\omega x] \left(\frac{1}{2}\left(\exp[i\omega] + \exp[-i\omega]\right)\right)^N \\
&= \frac{1}{2\pi} \int_0^{2\pi} d\omega \exp[i\omega x] \cos[\omega]^N
\end{aligned}$$

If we expand $\left(\frac{1}{2}\left(\exp[i\omega] + \exp[-i\omega]\right)\right)^N$ using the binomial theorem then each term can be intergrated to yield a particular Kronecker delta, i.e.,

$$\begin{aligned}
P_X(x) &= \frac{1}{2^N} \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!} \frac{1}{2\pi} \int_0^{2\pi} d\omega \exp[i\omega x + in\omega - i(N-n)\omega] \\
&= \frac{1}{2^N} \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!} \frac{1}{2\pi} \int_0^{2\pi} d\omega \exp[i\omega(x + 2n - N)] \\
&= \frac{1}{2^N} \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!} \delta_K(x, N-2n)
\end{aligned}$$

8.14 Chi-Square Distribution via Probability Mapping

The Chi-Square distribution is the probability distribution for the sum, χ, of the squares of N random variables, x_n, each of which follow a Gaussian distribution

with zero mean and the same variance. That is

$$P(\chi) = \int_{-\infty}^{+\infty} dx_1 \cdots dx_N \delta_D\left(\chi - \left(x_1^2 + \cdots + x_N^2\right)\right) \frac{\exp\left[-\frac{1}{2\sigma^2}\left(x_1^2 + \cdots + x_N^2\right)\right]}{\left(2\pi\sigma^2\right)^{N/2}}$$

It seems customary to measure the x_n in units of σ which is equivalent to setting $\sigma = 1$ which we do now.

Using the delta function we have

$$P(\chi) = \frac{1}{(2\pi)^{N/2}} \exp\left[-\frac{\chi}{2}\right] \int_{-\infty}^{+\infty} dx_1 \cdots dx_N \delta_D\left(\chi - \left(x_1^2 + \cdots + x_N^2\right)\right)$$

Switching to spherical polar coordinates in N dimensions, with Ω_N the solid angle of a unit sphere in N dimensions, we have

$$\begin{aligned} P(\chi) &= \frac{\Omega_N}{(2\pi)^{N/2}} \exp\left[-\frac{\chi}{2}\right] \int_0^{\infty} dr r^{N-1} \delta_D\left(\chi - r^2\right) \\ &= \frac{\Omega_N}{(2\pi)^{N/2}} \exp\left[-\frac{\chi}{2}\right] \int_0^{\infty} dr r^{N-1} \delta_D\left(\left(\sqrt{\chi} - r\right)\left(\sqrt{\chi} + r\right)\right) \\ &= \frac{\Omega_N}{(2\pi)^{N/2}} \exp\left[-\frac{\chi}{2}\right] \frac{1}{2\sqrt{\chi}} \chi^{(N-1)/2} \end{aligned}$$

As shown in the section on solid angles in N dimensions

$$\Omega_N = \frac{2\pi^{N/2}}{\Gamma(N/2)}$$

and so

$$P(\chi) = \frac{\chi^{(N-2)/2}}{2^{N/2}\Gamma(N/2)} \exp\left[-\frac{\chi}{2}\right]$$

This is of course the probability distribution for $\chi \geq 0$. For $\chi < 0$, $P(\chi) = 0$. $P(\chi)$ is said to have "$N-1$ degrees of freedom" since χ is constrained to equal $x_1^2 + x_2^2 + \cdots + x_N^2$.

8.15 Sum of Poisson = Poisson via Probability Mapping

Consider two random integer valued variables, n and m, both of which satisfy Poisson statistics with $\langle n \rangle = N$ and $\langle m \rangle = M$. Then the probability distribution for $s = n + m$ is given by

$$\begin{aligned} P(s) &= \sum_{n,m=0}^{\infty} \delta_K(s, n+m) P(n|N) P(m|M) \\ &= \sum_{n,m=0}^{\infty} \delta_K(s, n+m) \frac{N^n}{n!} e^{-N} \frac{M^m}{m!} e^{-M} \end{aligned}$$

Using the identity

$$\delta_K(s, n+m) = \frac{1}{2\pi}\int_0^{2\pi} d\theta \exp\left[i(s-n-m)\theta\right]$$

and rearranging factors gives

$$\begin{aligned}
P(s) &= \frac{1}{2\pi}\int_0^{2\pi} d\theta \exp[is\theta] \sum_{n=0}^{\infty} \frac{\left(Ne^{i\theta}\right)^n}{n!} \sum_{m=0}^{\infty} \frac{\left(Me^{i\theta}\right)^m}{m!} \\
&= \frac{1}{2\pi}\int_0^{2\pi} d\theta \exp[is\theta] \exp\left[(N+M)\left(e^{i\theta}-1\right)\right] \\
&= \frac{1}{2\pi}\int_0^{2\pi} d\theta \exp[is\theta] \sum_{q=0}^{\infty} \frac{(N+M)^q e^{-iq\theta}}{q!} e^{-(N+M)} \\
&= \sum_{q=0}^{\infty} \frac{(N+M)^q}{q!} e^{-(N+M)} \frac{1}{2\pi}\int_0^{2\pi} d\theta \exp\left[i(s-q)\theta\right] \\
&= \sum_{q=0}^{\infty} \frac{(N+M)^q}{q!} e^{-(N+M)} \delta_K(s,q) \\
&= \frac{(N+M)^s}{s!} e^{-(N+M)}
\end{aligned}$$

Hence the sum of non-negative integer valued random variables which satisfy Poisson statistics also satisfies Poisson statistics.

8.16 Binomial Distribution via Probability Mapping

Consider a random system that has 2 and only two possible outcomes, call them a and b. Assume the occurrence of a has probability p and thus the occurrence of b has probability $1-p$. Define the variable s which can take the values either a or b.The probability distribution for the variable s then is given by

$$P(s) = \delta_{s,a}p + \delta_{s,b}(1-p)$$

where $\delta_{s,a} = 1$ for $s = a$ and is zero otherwise and $\delta_{s,b}$ is the same but for $s = b$, i.e., δ is a Kronecker delta function.

If sequential measurements of s, call them $s_1, s_2, \cdots$ are uncorrelated then the joint probability distribution for getting a particular sequence $s_1, s_2, \cdots, s_N$ for N measurements is given by

$$P_{joint}(s_1, s_2, \cdots, s_N) = P(s_1)P(s_2)\cdots P(s_N)$$

To find the probability, call it $P_a(m)$, that m of the measured s_q, $q = 1, 2, \cdots, N$ come out be a, use probability mapping. To do that consider the Kronecker delta function

$$\delta_K\left(m, \sum_{q=1}^{N} \delta_{s_q,a}\right) = \begin{cases} 1 \text{ for } \sum_{i=1}^{N} \delta_{s_q,a} = m \\ 0 \text{ for } \sum_{i=1}^{N} \delta_{s_q,a} \neq m \end{cases}$$

and calculate

$$P_a(m) = \sum_{\{s\}} P_{joint}(s_1, s_2, \cdots, s_N)\, \delta_K\left(m, \sum_{q=1}^{N} \delta_{s_q,a}\right)$$

Here $\sum_{\{s\}}$ indicates summing over all combinations of $s_1, s_2, \cdots, s_N$. The factor $\delta_K\left(m, \sum_{q=1}^{N} \delta_{s_q,a}\right)$ obviously amounts to summing up all the probabilities for which $\sum_{q=1}^{N} \delta_{s_q,a} = m$.

$\delta_K\left(m, \sum_{q=1}^{N} \delta_{s_q,a}\right)$ can be represented by a Fourier Transform

$$\delta_K\left(m, \sum_{q=1}^{N} \delta_{s_q,a}\right) = \frac{1}{2\pi} \int_0^{2\pi} d\omega \exp\left[i\omega\left(m - \sum_{q=1}^{N} \delta_{s_q,a}\right)\right]$$

Substituting gives

$$\begin{aligned}
P_a(m) &= \frac{1}{2\pi} \int_0^{2\pi} d\omega \exp[i\omega m] \sum_{\{s\}} P_{joint}(s_1, s_2, \cdots, s_N) \exp\left[-i\omega \sum_{q=1}^{N} \delta_{s_q,a}\right] \\
&= \frac{1}{2\pi} \int_0^{2\pi} d\omega \exp[i\omega m] \left\{ \begin{array}{c} \sum_{s_1} (\delta_{s_1,a} p + \delta_{s_1,b}(1-p)) \exp\left[-i\omega \delta_{s_1,a}\right] \\ \times \sum_{s_2} (\delta_{s_2,a} p + \delta_{s_2,b}(1-p)) \exp\left[-i\omega \delta_{s_2,a}\right] \\ \cdots \\ \times \sum_{s_N} (\delta_{s_N,a} p + \delta_{s_N,b}(1-p)) \exp\left[-i\omega \delta_{s_N,a}\right] \end{array} \right\} \\
&= \frac{1}{2\pi} \int_0^{2\pi} d\omega e^{i\omega m} \left(p e^{-i\omega} + (1-p)\right)^N \\
&= \sum_{n=0}^{N} \frac{N!}{n!(N-n)!} \frac{1}{2\pi} \int_0^{2\pi} d\omega e^{i\omega m} \left(p e^{-i\omega}\right)^n (1-p)^{N-n} \\
&= \sum_{n=0}^{N} \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \frac{1}{2\pi} \int_0^{2\pi} d\omega e^{i\omega(m-n)} \\
&= \sum_{n=0}^{N} \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \delta_{n,m} \\
&= \frac{N!}{m!(N-m)!} p^m (1-p)^{N-m}
\end{aligned}$$

which is the standard binomial probability distribution.

This result is rather obvious and can, with a little bit of thought, be written down directly without going through the above derivation. But it does provide another example of how "slick" the probability mapping approach is.

8.16.1 Binomial to Poisson

For $p \ll 1$ the binomial probability distribution above is dominated by the small values of m. Assuming that the dominant values pf m satisfy $m \ll N$ we have

$$\begin{aligned} \frac{N!}{(N-m)!} &\simeq N^m \\ & \text{and} \\ (1-p)^{N-m} &\simeq (1-p)^N \simeq e^{-Np} \end{aligned}$$

In this case the binomial distribution above reduces to

$$P_a(m) \simeq \frac{(Np)^m}{m!} e^{-Np}$$

which is the Poisson distribution with the average value of m, $\langle m \rangle = Np$ as expected.

9 Grassmann Variables

9.1 Definition

Grassmann variables are defined to be anticommuting variables, that is, if ξ and η are Grassmann variables then

$$\xi\eta = -\eta\xi$$

Grassmann variables are also defined to be anticommuting with respect to themselves and so

$$\xi\xi = -\xi\xi = 0$$

Hence the most general form for any function of a single Grassmann variable, $f(\xi)$

$$f(\xi) = a + b\xi$$

where a and b can be ordinary numbers (real, imaginary or complex) or they can be a combination of Grassmann and ordinary numbers.

In particular

$$\exp[\lambda\xi] = 1 + \lambda\xi$$

is the complete series for the exponential of a Grassmann times another number λ which can be Grassmann or ordinary.

For 3 Grassmann numbers, ξ, η, χ we have

$$\xi\eta\chi = \chi\xi\eta$$

and hence the product of 2 Grassmann variables acts like an ordinary variable, i.e., it is commuting.

Derivatives are defined in the usual way

$$\frac{\partial}{\partial \xi}\xi = 1$$
$$\frac{\partial}{\partial \xi}\eta = 0$$

There are various arguments for how to define integration of Grassmann variables but the net result is

Complete Integration Table for Grassman Variables

$$\int d\xi = 0$$
$$\int d\xi \xi = -\int \xi d\xi = 1$$

That's it, the complete integration table for Grassmann variables.

For example

$$\int d\xi \exp\left[\lambda\xi\right] = \int d\xi \left(1 + \lambda\xi\right) = \lambda \int d\xi\xi = \lambda$$

NOTE: The same integral in the case where ξ is an ordinary variable is proportional to $1/\lambda$. This is why fermion fields (such as the fields representing electrons, positrons, neutrinos, quarks, etc.) are taken to be Grassman valued when using functional techniques in quantum field theory.

For a set of Grassmann variables ξ_n with $n = 1$ to N we have

$$\int d\xi_1 d\xi_2 \cdots d\xi_N \xi_{i_1}\xi_{i_2}\cdots\xi_{i_N} = \pm\epsilon_{i_1,i_2,\cdots,i_N}$$

where

$\epsilon_{a,b,c,\cdots}$ = Levi-Civita totally antisymmetric tensor (defined in the Introduction)

The $\pm$ ambiguity must be removed in any given case by defining a fixed convention for the sign or a fixed ordering of integration variables. In most case of practical interest the overall sign either doesn't matter or it follows from the problem at hand.

NOTE: The above result that integrals of Grassmans are essentially Levi-Civita antisymmetric tensors means Grassmann integration is just a convenient way of handling permutations.

9.2 Determinants and Grassmann Variables

Consider the integral

$$\int d\eta_1 d\eta_2 \cdots d\eta_N \int d\xi_1 d\xi_2 \cdots d\xi_N \exp\left[\xi_i M_{i,j} \eta_j\right]$$

where $M_{i,j}$ is a matrix of ordinary numbers or variables and doubly repeated indices are assumed to be summed over, i.e.,

$$\xi_i M_{i,j} \eta_j \equiv \sum_{i,j} \xi_i M_{i,j} \eta_j$$

the "Einstein summation convention".

Consider first the simple case

$$\begin{aligned}
\int d\eta \int d\xi \exp\left[\xi a \eta\right] &= \int d\eta \int d\xi \left(1 + \xi a \eta\right) \\
&= a \int d\eta \int d\xi \xi \eta \\
&= a \int d\eta \eta \\
&= a
\end{aligned}$$

where a is ordinary. To get this result we first used $\int d\eta \int d\xi = 0$ then $\int d\xi \xi = 1$ and finally $\int d\eta \eta = 1$.

Next consider a $2{\times}2$ matrix M. The only term to survive after expanding the exponential is the one with 2 $\xi's$ and 2 $\eta's$

$$\begin{aligned}
\int d\eta_1 d\eta_2 \int d\xi_1 d\xi_2 \exp\left[\xi_i M_{i,j} \eta_j\right] &= \int d\eta_1 d\eta_2 \int d\xi_1 d\xi_2 \frac{1}{2} \xi_{i_1} M_{i_1,j_1} \eta_{j_1} \xi_{i_2} M_{i_2,j_2} \eta_{j_2} \\
&= \frac{1}{2} M_{i_1,j_1} M_{i_2,j_2} \int d\eta_1 d\eta_2 \int d\xi_1 d\xi_2 \xi_{i_1} \eta_{j_1} \xi_{i_2} \eta_{j_2} \\
&= -\frac{1}{2} M_{i_1,j_1} M_{i_2,j_2} \int d\eta_1 d\eta_2 \left(\int d\xi_1 d\xi_2 \xi_{i_1} \xi_{i_2} \right) \eta_{j_1} \eta_{j_2} \\
&= -\frac{1}{2} M_{i_1,j_1} M_{i_2,j_2} \left(\pm \epsilon_{i_1,i_2}\right) \left(\pm \epsilon_{j_1,j_2}\right) \\
&= -\left(M_{1,1} M_{2,2} - M_{1,2} M_{2,1}\right) \\
&= -\det\left[M\right]
\end{aligned}$$

and the sign ambiguity $\pm\epsilon$ does not make a difference. On the other if we use a different order of the integration variables we get

$$\int d\eta_1 d\xi_1 d\eta_2 d\xi_2 \exp\left[\xi_i M_{i,j} \eta_j\right] = +\det\left[M\right]$$

which follows directly from the previous result since we have defined $d\xi_1 d\eta_2 = -d\eta_2 d\xi_1$.

For the general case expanding the exponent in a series the only term that does not integrate to zero is the term where the number of $\xi's$ and the number

of $\eta's$ are both exactly N. This term has the form

$$\frac{1}{N!}\int d\eta_1 d\eta_2 \cdots d\eta_N \int d\xi_1 d\xi_2 \cdots d\xi_N \xi_{i_1} M_{i_1,j_1} \eta_{j_1} \xi_{i_2} M_{i_2,j_2} \eta_{j_2} \cdots \xi_{i_N} M_{i_N,j_N} \eta_{j_N}$$
$$= \frac{1}{N!} M_{i_1,j_1} M_{i_2,j_2} \cdots M_{i_N,j_N} \int d\eta_1 d\eta_2 \cdots d\eta_N \int d\xi_1 d\xi_2 \cdots d\xi_N \xi_{i_1} \eta_{j_1} \xi_{i_2} \eta_{j_2} \cdots \xi_{i_N} \eta_{j_N}$$
$$= \pm \frac{1}{N!} M_{i_1,j_1} M_{i_2,j_2} \cdots M_{i_N,j_N} \epsilon_{i_1,i_2,\cdots,i_N} \epsilon_{j_1,j_2,\cdots,j_N}$$
$$= \pm \epsilon_{i_1,i_2,\cdots,i_N} M_{i_1,1} M_{i_2,2} \cdots M_{i_N,N}$$
$$= \pm \det[M]$$

The overall sign depends on N and for the given ordering the signs go as $+,-,-,+,+,-$ for $N = 1,2,3,4,5,6$, respectively.

In most cases where Grassmann variables or fields are used, such as in the functional integral approach to quantum field theory, the overall sign the integral is not important, essentially because normalization cancels it out. On the other hand we now show that $\det[M]$ is the Jacobeam of a change of integration variables and we know the same sign ambiguity occurs in the ordinary commuting case as well. This is why the Jacobean uses the absolute value of the determinant of the transformation. Hence, as in the ordinary case, the Grassmann result is generally written as

$$\int d\eta \int d\xi \exp\left[\xi_i M_{i,j} \eta_j\right] = |\det[M]|$$

NOTE: Often the $|\cdots|$ on the determinant are not shown explicity.

This result is the reciprocal of the Jacobean for ordinary commuting variables. Consider

$$\int dx \int dy \exp\left[x_i M_{i,j} y_j\right]$$

where the x_i, y_j and $M_{i,j}$ are all ordinary commmuting variables. Making the change of integration variable

$$z_i = M_{i,j} y_j$$

the Jacobean follows from

$$dz_1 dz_2 \cdots dz_N = \det[M]\, dy_1 dy_2 \cdots dy_N$$
or
$$dy_1 dy_2 \cdots dy_N = \frac{1}{\det[M]} dz_1 dz_2 \cdots dz_N$$
$$= \text{Jacobean} \times dz_1 dz_2 \cdots dz_N$$

and so

$$\int dx \int dy \exp\left[x_i M_{i,j} y_j\right] = \frac{1}{\det[M]} \int dx \int dz \exp\left[x_i z_i\right]$$
$$\sim \frac{1}{\det[M]}$$

But again $\det[M]$ can be negative and a change of integration variable cannot change the sign of the integral and so the Jacobean is defined as

$$\text{Jacobean for commuting variables} = \frac{1}{|\det[M]|}$$

The result for the Grassmann variables implies the Jacobean for Grassman variables is the reciprocal of the ordinary Jacobean, i.e.,

$$\lambda_i = M_{i,j}\eta_j \implies d\eta_1 d\eta_2 \cdots d\eta_N = \det[M]\, d\lambda_1 d\lambda_2 \cdots d\lambda_N$$

Taking the absolute value as in the ordinary case we have

$$\text{Jacobean for Grassmann variables} = |\det[M]|$$

10 Relativity, On Shell, Off Shell

Needs work

Relation between energy E and spatial momentum $\vec{p}$, rest mass m and speed of light c, all units MKS,

$$\begin{aligned} E^2 &= \vec{p}^2c^2 + m^2c^4 = \text{Relativistic energy (squared)} \\ \vec{p} &= m\gamma\vec{v} = \text{Relativistic momentum} \\ \gamma &= \frac{1}{\sqrt{1-\frac{\vec{v}^2}{c^2}}} \end{aligned}$$

Photons are massless so they obey

$$E = \sqrt{\vec{p}^2}c = |\vec{p}|\, c$$

$\vec{v}$ is velocity measured in a given inertial frame

$$\vec{v} = \frac{d\vec{x}}{dt}$$

where $\vec{x}$ is position and t is time in the frame.

4-momentum with index notation: subscript 0 is time, subscript 1,2,3 refer to x, y, z coordinates

$$p = (p_0, p_1, p_2, p_3) = \left(\frac{E}{c^2}, p_x, p_y, p_z\right)$$

$$p_\mu p^\mu = p_0^2 - \vec{p}^2 = \frac{E^2}{c^2} - \vec{p}^2 = m^2c^2$$

11 WKB (Wentzel Kramers Brillouin)

11.1 Derivation

Consider the partial differential equation (PDE)

$$\left(\vec{\partial}^2 + V\left(\vec{r}\right)\right)\phi\left(\vec{r}\right) = 0$$

where for $i = 1$ to N with repeated indices summed over that range,

$$\begin{aligned}
\hat{x}_i &= \text{Orthogonal unit vectors spanning the } N \text{ dimensional space of } \vec{r} \\
\vec{r} &= \hat{x}_i x_i = (x_1, x_2, \cdots, x_N) \\
\vec{\partial} &= \hat{x}_i \frac{\partial}{\partial x_i} = \hat{x}_i \partial_i = (\partial_1, \partial_2, \cdots, \partial_N) = \text{Gradient} \\
\vec{\partial}^2 &= \vec{\partial} \cdot \vec{\partial} = \partial_1^2 + \partial_2^2 + \cdots + \partial_N^2 = \text{Laplacian}
\end{aligned}$$

NOTE: The different ways of writing $\vec{r}$ and $\vec{\partial}$ above are just different notations for the same thing.

Let

$$\phi\left(\vec{r}\right) = \exp\left[S\left(\vec{r}\right)\right]$$

Substituting into the PDE and cancelling the overall factor of $\exp\left[S\left(\vec{r}\right)\right]$ gives (suppressing the $\vec{r}$ dependence for notational convenience)

$$\left(\vec{\partial}S\right)^2 + \left(V + \vec{\partial}^2 S\right) = 0$$

which can rearranged as

$$\left|\vec{\partial}S\right| = \sqrt{\left(\vec{\partial}S\right)\cdot\left(\vec{\partial}S\right)} = \pm i\sqrt{V + \vec{\partial}^2 S}$$

where $i = \sqrt{-1}$. The first equality in the last line is just a definition of the notation $|\cdots|$. Assume that $\vec{\partial}^2 S\left(\vec{r}\right)$ is smaller than $V\left(\vec{r}\right)$. Below we will give examples of how this can happen. But for now expand the square root to first order in $\vec{\partial}^2 S$

$$\begin{aligned}
\left|\vec{\partial}S\right| &\simeq \pm i\sqrt{V}\left(1 + \frac{1}{2}\frac{\vec{\partial}^2 S}{V}\right) \\
&\simeq \pm i\left(\sqrt{V} + \frac{1}{2}\frac{\vec{\partial}^2 S}{\sqrt{V}}\right)
\end{aligned}$$

So we have the modulus of the gradient of S but not its direction. Ideally we would like to have $\vec{\partial}S$ so that

$$S\left(\vec{r}\left(\ell\right)\right) - S\left(\vec{r}\left(0\right)\right) = \int_0^\ell d\ell'\ \hat{t}\left(\ell'\right)\cdot\vec{\partial}S\left(\vec{r}\left(\ell'\right)\right)$$

where

$$\begin{aligned}
\vec{r}(\ell) &= \text{a particular path or trajectory} \\
\ell &= \text{length along the path} \\
\hat{t}(\ell) &= \frac{d\vec{r}(\ell)}{d\ell} = \text{unit length vector tangent to the path} \\
\vec{\partial} S(\vec{r}(\ell')) &= \left[\vec{\partial} S(\vec{r})\right]_{\vec{r}=\vec{r}(\ell')}
\end{aligned}$$

The last line above means take the gradient of S first then replace $\vec{r}$ with $\vec{r}(\ell')$.

NOTE: $S(\vec{r})$ is a scalar function, i.e., at any $\vec{r}$ it has a given value, hence the value of $S(\vec{r}(\ell)) - S(\vec{r}(0))$ depends only on the points $\vec{r}(\ell)$ and $\vec{r}(0)$ and is independent of the path between those two points.

In 1 dimension (1D), the direction is given by default and so in 1D with position labeled by x we have

$$\begin{aligned}
\partial_x S(x) &\simeq \pm i\left(\sqrt{V(x)} + \frac{1}{2}\frac{\partial_x^2 S(x)}{\sqrt{V(x)}}\right) \\
&\simeq \pm i\left(\sqrt{V(x)} + \frac{1}{2}\frac{\partial_x(\partial_x S(x))}{\sqrt{V(x)}}\right) \\
&\simeq \pm i\left(\sqrt{V(x)} + \frac{1}{2}\frac{\partial_x\left(\pm i\sqrt{V(x)}\right)}{\sqrt{V(x)}}\right) \\
&\simeq \pm i\sqrt{V(x)} - \frac{1}{4}\partial_x \ln[V(x)]
\end{aligned}$$

where in the third line we have substituted $\pm i\sqrt{V(x)}$ for $\partial_x S(x)$ which is valid to the given order of the expansion. Integrating both sides we have

$$\begin{aligned}
S(x) &\simeq \pm i\int_{x_0}^{x} dx'\sqrt{V(x')} - \frac{1}{4}\ln[V(x)] + \frac{1}{4}\ln[V(x_0)] + S(x_0) \\
&\simeq \pm i\int_{x_0}^{x} dx'\sqrt{V(x')} - \frac{1}{4}\ln[V(x)] + \text{constant}
\end{aligned}$$

and

$$\exp[S(x)] \simeq \frac{C}{V(x)^{1/4}}\exp\left[\pm i\int_{x_0}^{x} dx'\sqrt{V(x')}\right] \; : \text{1D WKB Solution}$$

where $C = \exp[\text{constant}]$. This is the standard WKB result for 1D.

Note that the real part of S is an amplitude and the imaginary part is a phase. Below we will relate the phase to the eikonal, ray tracing and Fermats "principle of least time".

The are issues with the above approximate close to any turning points x_{TP} where $V(x_{TP}) = 0$. It looks like things "blow up" but by using Taylor expansions of $V(x)$ around x_{TP} a smooth finite solution can be found.

In 1D if $V(x) = k_0^2 = \text{constant}$, then $\phi(x) = \exp[S(x)] = \exp[\pm ik_0 x]$ which is the standard basis function solution. In multiple dimensions if $V(\vec{r}) = k_0^2 = \text{constant}$, then $\phi(\vec{r}) = \exp[S(\vec{r})] = \exp\left[\pm i\vec{k}\cdot\vec{r}\right]$ with the "length" of $\vec{k}$ fixed, $\left|\vec{k}\right| = \sqrt{\vec{k}\cdot\vec{k}} = k_0$, but the direction or the combination of directions of $\vec{k}$ are unspecified. Any combination is a solution. It is the boundary conditions, initial conditions, sources and/or sinks for the field $\phi(\vec{r})$ which determine the particular combination and weighting of the basis functions $\exp\left[i\vec{k}\cdot\vec{r}\right]$. The same is true for the WKB solution $\exp[S(\vec{r})]$, boundary conditions, initial conditions, etc., determine the relevant direction of $\vec{\partial}S$.

11.2 Eikonal, Rays, and Fermat

Consider a particular path $\vec{r}(\ell)$ with ℓ the length along the path and where for all ℓ the direction of the gradient, $\vec{\partial}S$, is tangent to the path

$$\left.\frac{\vec{\partial}S(\vec{r})}{\left|\vec{\partial}S(\vec{r})\right|}\right|_{\vec{r}=\vec{r}(\ell)} = \frac{d\vec{r}(\ell)}{d\ell} = \hat{t}(\ell) \;\; \forall\ell$$

Note that since

$$d\vec{r}\cdot d\vec{r} = \left(dx_1^2 + dx_2^2 + \cdots dx_N^2\right) = d\ell^2$$

we have

$$\frac{d\vec{r}(\ell)}{d\ell}\cdot\frac{d\vec{r}(\ell)}{d\ell} = \hat{t}(\ell)\cdot\hat{t}(\ell) = 1$$

To relate what we're doing here to optics and ray tracing let

$$V(\vec{r}) = (n(\vec{r})\,k)^2 = \text{Real and greater than 0 } \forall\ \vec{r}$$

where

$$k = \frac{2\pi}{\lambda} \text{ with } \lambda = \text{vacuum wavelength}$$
$$n(\vec{r}) = \text{position dependent index of refraction}$$

Consider just the $\sqrt{V}$ term in the exponent in S and choose the + sign. Since we have defined $d\vec{r}(\ell)/d\ell$ as being tangent to $\vec{\partial}S(\vec{r})$ at all points along the path we have

$$\vec{\partial}S(\vec{r}(\ell)) = \left|\vec{\partial}S(\vec{r}(\ell))\right|\frac{d\vec{r}(\ell)}{d\ell} = i\sqrt{V(\vec{r}(\ell))}\frac{d\vec{r}(\ell)}{d\ell} = ikn(\vec{r}(\ell))\frac{d\vec{r}(\ell)}{d\ell}$$

The right hand side is purely imaginary which means in this case S is pure phase and so define $\psi = S/i$ so that

$$\vec{\partial}\psi(\vec{r}(\ell)) = kn(\vec{r}(\ell))\frac{d\vec{r}(\ell)}{d\ell}$$

Take the derivative of both sides w.r.t. ℓ. Since the gradient $\vec{\partial}$ does not depend on ℓ, $d/d\ell\ \vec{\partial} = \vec{\partial}\ d/d\ell$,then using the notation $d/d\ell = \partial_\ell$, we have

$$\begin{aligned}
\partial_\ell \vec{\partial}\psi\left(\vec{r}\left(\ell\right)\right) &= \vec{\partial}\partial_\ell \psi\left(\vec{r}\left(\ell\right)\right) \\
&= \vec{\partial}\left(\frac{d\vec{r}\left(\ell\right)}{d\ell}\cdot\vec{\partial}\psi\left(\vec{r}\left(\ell\right)\right)\right) \\
&= \vec{\partial}\left(\frac{d\vec{r}\left(\ell\right)}{d\ell}\cdot\left(kn\left(\vec{r}\left(\ell\right)\right)\frac{d\vec{r}\left(\ell\right)}{d\ell}\right)\right) \\
&= \vec{\partial}\left(kn\left(\vec{r}\left(\ell\right)\right)\left(\underbrace{\frac{d\vec{r}\left(\ell\right)}{d\ell}\cdot\frac{d\vec{r}\left(\ell\right)}{d\ell}}_{=1}\right)\right) \\
&= k\vec{\partial}n\left(\vec{r}\left(\ell\right)\right)
\end{aligned}$$

Cancelling overall factors of k and equating the left and right hand sides above gives

$$\begin{aligned}
\vec{\partial}n\left(\vec{r}\left(\ell\right)\right) &= \partial_\ell\left(n\left(\vec{r}\left(\ell\right)\right)\frac{d\vec{r}\left(\ell\right)}{d\ell}\right) \\
&= \left(\frac{d\vec{r}\left(\ell\right)}{d\ell}\cdot\vec{\partial}n\left(\vec{r}\left(\ell\right)\right)\right)\frac{d\vec{r}\left(\ell\right)}{d\ell} + n\left(\vec{r}\left(\ell\right)\right)\left(\partial_\ell\frac{d\vec{r}\left(\ell\right)}{d\ell}\right)
\end{aligned}$$

or, rearranging and remembering $\partial_\ell = d/d\ell$

$$\frac{d^2\vec{r}\left(\ell\right)}{d\ell^2} = \frac{1}{n\left(\vec{r}\left(\ell\right)\right)}\left(\vec{\partial}n\left(\vec{r}\left(\ell\right)\right) - \frac{d\vec{r}\left(\ell\right)}{d\ell}\left(\frac{d\vec{r}\left(\ell\right)}{d\ell}\cdot\vec{\partial}n\left(\vec{r}\left(\ell\right)\right)\right)\right) = \text{Ray trajectory equation}$$

This is the standard differential equation for the trajectory of a ray, $\vec{r}\left(\ell\right)$ given the index of refraction as a function of position, $n\left(\vec{r}\right)$.

Substituting $S = i\psi$ in the integral above for $S\left(\vec{r}\left(\ell\right)\right)$ and using $\vec{\partial}\psi\left(\vec{r}\left(\ell\right)\right) = kn\left(\vec{r}\left(\ell\right)\right)d\vec{r}\left(\ell\right)/d\ell$ we have

$$\psi\left(\vec{r}\left(\ell\right)\right) = k\int_0^\ell d\ell'\ n\left(\vec{r}\left(\ell'\right)\right) + \psi\left(\vec{r}\left(0\right)\right)$$

Since only phase differences count we can set $\psi\left(\vec{r}\left(0\right)\right) = 0$.

11.2.1 Fermats Principle of Least Time

We now show that setting the change in $\psi\left(\vec{r}\left(\ell\right)\right)$ equal to zero for a small variation in the path $\vec{r}\left(\ell\right)$ yields the ray trajectory equation derived above. But since the index of refraction is defined as

$$n\left(\vec{r}\right) = \frac{c}{v\left(\vec{r}\right)} = \frac{\text{speed of light in vacuum}}{\text{speed of light at } \vec{r} \text{ in a material}}$$

it follows that

$$\frac{1}{kc}\psi\left(\vec{r}\left(\ell\right)\right)=\int_{0}^{\ell}d\ell'\frac{1}{v\left(\vec{r}\left(\ell'\right)\right)}=\text{time for ray to propgate from }\vec{r}\left(0\right)\text{ to }\vec{r}\left(\ell\right)$$

Hence the variation of $\psi\left(\vec{r}\left(\ell\right)\right)$ being 0 for small change in $\vec{r}\left(\ell\right)$ indicates the ray trajectory is an extremum w.r.t. time which is Fermats principle of "least time". We only show the trajectory is an extremum. We do not show it always corresponds to a minimum because, in some cases, it doesn't. In some cases it corresponds to a maximum, as considered in the section on Circular Ray Trajectories below.

We work with $n\left(\vec{r}\left(\ell\right)\right)$ instead of $v\left(\vec{r}\left(\ell\right)\right)$ since it is less messy. Before evaluating the variation note that because by definition

$$\frac{d\vec{r}\left(\ell\right)}{d\ell}\cdot\frac{d\vec{r}\left(\ell\right)}{d\ell}=1$$

we have to do a constrained variation. This can be done using a Lagrange multiplier. But it's much easier and cleaner to use an arbitrary parameter, say σ, to indicate or label position along the ray trajectory with σ not necessarily the length along the ray trajectory.This works as long as ℓ increases monotonically with σ,

$$\frac{d\ell}{d\sigma}>0\ \forall\sigma$$

so that each value of σ indicates one and only one position on the trajectory.

Noting that

$$d\ell=\sqrt{d\vec{r}\cdot d\vec{r}}$$

it follows that

$$d\ell=\frac{d\ell}{d\sigma}d\sigma=\sqrt{\frac{d\vec{r}\left(\sigma\right)}{d\sigma}\cdot\frac{d\vec{r}\left(\sigma\right)}{d\sigma}}d\sigma=\sqrt{\left|\frac{d\vec{r}\left(\sigma\right)}{d\sigma}\right|^{2}}d\sigma$$

Hence consider

$$\psi\left(\vec{r}\left(\sigma\right)\right)=k\int_{0}^{\sigma}d\sigma'\sqrt{\left|\frac{d\vec{r}\left(\sigma'\right)}{d\sigma'}\right|^{2}}n\left(\vec{r}\left(\sigma'\right)\right)$$

We will implement the variation by using functional derivatives. They act like ordinary derivatives up until the functional derivative is being taken w.r.t to the same function as itself then they yield a Dirac delta function. Functional derivative notation usually, but not always is written using δ or sometimes D, instead of ∂ or d. The functional derivative of a function $f\left(x\right)$ w.r.t. itself $f\left(x'\right)$ is given by

$$\frac{\delta f\left(x\right)}{\delta f\left(x'\right)}=\delta_{D}\left(x-x'\right)$$

Use the ordinary chain rule until the functional derivative is w.r.t. the identical function, then use the above result.

Take the functional derivative of $\psi\left(\vec{r}\left(\sigma\right)\right)$ with respect to $\vec{r}\left(\sigma''\right)$. We can do this using $\vec{r}\left(\sigma\right) = x_i\left(\sigma\right)\hat{x}_i$ and $\delta/\delta\vec{r}\left(\sigma''\right) = \delta/\delta x_j\left(\sigma''\right)\hat{x}_j$. (Rememeber: Einstein summation convention and the $\hat{x}_i$ are fixed unit vectors along, say, the x, y, z directions).

Setting the result to 0 gives

$$0 = \hat{x}_j \cdot \int_0^\sigma d\sigma' \left(\begin{array}{c} \frac{1}{2}\frac{1}{\sqrt{\left|\frac{d\vec{r}\left(\sigma'\right)}{d\sigma'}\right|^2}} 2\frac{dx_i\left(\sigma'\right)}{d\sigma'}\frac{\delta}{\delta x_j\left(\sigma''\right)}\frac{dx_i\left(\sigma'\right)}{d\sigma'} n\left(\vec{r}\left(\sigma'\right)\right) \\ +\sqrt{\left|\frac{d\vec{r}\left(\sigma'\right)}{d\sigma'}\right|^2}\frac{\delta x_i\left(\sigma'\right)}{\delta x_j\left(\sigma''\right)}\partial_i n\left(\vec{r}\left(\sigma'\right)\right) \end{array} \right)$$

or, since each vector component must individually be zero

$$\begin{aligned} 0 &= \int_0^\sigma d\sigma' \left(\begin{array}{c} \frac{1}{\sqrt{\left|\frac{d\vec{r}\left(\sigma'\right)}{d\sigma'}\right|^2}}\frac{dx_i\left(\sigma'\right)}{d\sigma'}\frac{\delta}{\delta x_j\left(\sigma''\right)}\frac{dx_i\left(\sigma'\right)}{d\sigma'} n\left(\vec{r}\left(\sigma'\right)\right) \\ +\sqrt{\left|\frac{d\vec{r}\left(\sigma'\right)}{d\sigma'}\right|^2}\frac{\delta x_i\left(\sigma'\right)}{\delta x_j\left(\sigma''\right)}\partial_i n\left(\vec{r}\left(\sigma'\right)\right) \end{array} \right) \\ &= \int_0^\sigma d\sigma' \left(\begin{array}{c} \frac{1}{\sqrt{\left|\frac{d\vec{r}\left(\sigma'\right)}{d\sigma'}\right|^2}}\frac{dx_i\left(\sigma'\right)}{d\sigma'}\frac{d}{d\sigma'}\frac{\delta x_i\left(\sigma'\right)}{\delta x_j\left(\sigma''\right)} n\left(\vec{r}\left(\sigma'\right)\right) \\ +\sqrt{\left|\frac{d\vec{r}\left(\sigma'\right)}{d\sigma'}\right|^2}\frac{\delta x_i\left(\sigma'\right)}{\delta x_j\left(\sigma''\right)}\partial_i n\left(\vec{r}\left(\sigma'\right)\right) \end{array} \right) \end{aligned}$$

where we have used $\frac{\delta}{\delta x_j\left(\sigma''\right)}\frac{d}{d\sigma'} = \frac{d}{d\sigma'}\frac{\delta}{\delta x_j\left(\sigma''\right)}$. But by definition of functional derivatives $\frac{\delta x_i\left(\sigma'\right)}{\delta x_j\left(\sigma''\right)} = \delta_K\left(i,j\right)\delta_D\left(\sigma' - \sigma''\right)$ and so

$$0 = \int_0^\sigma d\sigma' \left(\begin{array}{c} \frac{1}{\sqrt{\left|\frac{d\vec{r}\left(\sigma'\right)}{d\sigma'}\right|^2}}\frac{dx_j\left(\sigma'\right)}{d\sigma'}\frac{d}{d\sigma'}\delta_D\left(\sigma' - \sigma''\right) n\left(\vec{r}\left(\sigma'\right)\right) \\ +\sqrt{\left|\frac{d\vec{r}\left(\sigma'\right)}{d\sigma'}\right|^2}\partial_j n\left(\vec{r}\left(\sigma'\right)\right)\delta_D\left(\sigma' - \sigma''\right) \end{array} \right)$$

Integrate the first term by parts and use the property of the Dirac delta function to evaluate the integral with the result

$$0 = -\frac{d}{d\sigma''}\left(\frac{1}{\sqrt{\left|\frac{d\vec{r}\left(\sigma''\right)}{d\sigma''}\right|^2}}\frac{dx_j\left(\sigma''\right)}{d\sigma''} n\left(\vec{r}\left(\sigma''\right)\right)\right) + \sqrt{\left|\frac{d\vec{r}\left(\sigma''\right)}{d\sigma''}\right|^2}\partial_j n\left(\vec{r}\left(\sigma''\right)\right)$$

We can now set $\sigma = \ell$ in which case $\left|d\vec{r}\left(\sigma\right)/d\sigma\right|^2 = \left|d\vec{r}\left(\ell\right)/d\ell\right|^2 = 1$ and dropping the double prime notation on σ we have

$$0 = \partial_j n\left(\vec{r}\left(\ell\right)\right) - \frac{d}{d\ell}\left(\frac{dx_j\left(\ell\right)}{d\ell} n\left(\vec{r}\left(\ell\right)\right)\right)$$

Rearranging gives

$$\frac{d^2x_j(\ell)}{d\ell^2} = \frac{1}{n(\vec{r}(\ell))}\left(\partial_j n(\vec{r}(\ell)) - \frac{dx_j(\ell)}{d\ell}\frac{dx_i(\ell)}{d\ell}\partial_i n(\vec{r}(\sigma))\right)$$

or in terms of vectors

$$\frac{d^2\vec{r}(\ell)}{d\ell^2} = \frac{1}{n(\vec{r}(\ell))}\left(\vec{\partial}n(\vec{r}(\ell)) - \frac{d\vec{r}(\ell)}{d\ell}\left(\frac{d\vec{r}(\ell)}{d\ell}\cdot\vec{\partial}n(\vec{r}(\ell))\right)\right) = \text{Ray trajectory equation}$$

which is exactly the same result from the previous subsection.

11.2.2 Circular Ray Trajectories

Consider the index distribution $n(r)$ in 2D with $r = \sqrt{x^2+y^2}$.

Lets see if we can find a particular index distribution, $n(r)$, such that a circle of radius R, i.e., $\vec{r}(\ell) = R\hat{r}(\ell/R) = R\hat{r}(\theta)$ satisfies the ray trajectory equation given above. With ℓ being distance along the circumference of the circle, $\theta = \ell/R$ is, by definition, the usual angle in polar (or cylindrical) coordinates. Also, $\hat{r}(\theta)$ is the unit vector in the r direction, and $\hat{\theta}(\theta)$ is the unit vector in the θ direction, both depend on angle θ.

First we have

$$\vec{\partial}n(r) = \vec{\partial}r\ \partial_r n(r) = \hat{r}(\theta)\,\partial_r n(r)$$

and

$$\partial_\ell\vec{r}(\ell/R) = \partial_\ell\vec{r}(\theta) = R\frac{1}{R}\partial_\theta\hat{r}(\theta) = \hat{\theta}(\theta) = \hat{\theta}(\ell/R)$$

Hence, for the circular trajectory we have

$$\frac{d\vec{r}(\ell)}{d\ell}\cdot\vec{\partial}n(\vec{r}(\ell)) = 0$$

The ray trajectory equation is now reduced to

$$\frac{d^2\vec{r}(\ell)}{d\ell^2} = \frac{\vec{\partial}n(\vec{r}(\ell))}{n(\vec{r}(\ell))}$$

But

$$\frac{d^2\vec{r}(\ell)}{d\ell^2} = \frac{d^2R\hat{r}(\ell/R)}{d\ell^2} = \frac{d}{d\ell}\hat{\theta}(\ell/R) = -\frac{1}{R}\hat{r}(\ell/R)$$

Putting things together gives

$$-\frac{1}{R}\hat{r}(\ell/R) = \frac{\hat{r}(\ell/R)\,\partial_r n(r)}{n(r)}$$

or

$$-\frac{1}{R} = \frac{\partial_r n(r)}{n(r)}$$

which has the solution $n(r) \sim \exp[-r/R]$. Fixing the integration constant, $n(r=R) = n_0$, we have finally

$$n(r) = n_0\exp[-r/R]$$

So an index distribution in 2D which decreases exponentially with r has circles as one solution to the ray trajectory equation. The circular solution only works at one radius, R.

NOTE 1: $n(r)$ is unitless (just a number) but r has units of length. So something with units of length has to enter into the function $n(r)$ to cancel the units of r which is why in the end $n(r)$ is actually $n(r/R)$. Units matter!

NOTE 2: The ray can go either way around the circle. The shortest direction around the circle between any two points is the "least time" extremum, and the longest direction around the circle between the same two points is the "most time" extremum.

11.3 Level Set Method

The level set method was created to model surface evolution, i.e., to evolve an N dimensional manifold living in $N+1$ dimensional space. Physically the goal is to model etch and deposition processes. The starting equation is very similar to the ray trace equation above and the resulting level set approach to its solution is very similar to WKB above.

To be specific consider a 2D surface living in 3D. The equivalent of the ray trace equation is written

$$\partial_t \vec{r}(t) = R\,\hat{n}\left(\vec{r}(t)\right)$$

where $\vec{r}(t) = (x(t), y(t), z(t))$ is the position of a point on the surface at time t, $\hat{n}\left(\vec{r}(t)\right)$ is the local normal to the surface at that point and R is the "speed" at which that point moves forward in time. Tracing out $\vec{r}(t)$ is equivalent to tracing a ray in the optical case.

To talk about the whole surface and not just a single point we need to let $\vec{r}(t) \rightarrow \vec{r}(t, \vec{s})$ where $\vec{s}$ is a 2D coordinate system living on (or in, if you prefer) the surface which labels position on the surface itself.

The rate R can and does depend on lots of things. If R is a constant then the surface evolution is isotropic and homogeneous. If it depends only on position, $\vec{r}(t, \vec{s})$, then the evolution is isotropic but not homogeneous. It can also depend directly on time t It can also depend on derivatives of $\vec{r}(t, \vec{s})$ both with respect to t and $\vec{s}$. For example, if it depends on $\partial_t \vec{r}(t)$ then the evolution rate or speed of the surface depends on the direction of propagation of the surface at each point which is nonisotropic but still homogeneous evolution as long as R does not depend directly on $\vec{r}(t, \vec{s})$. Both the isotropic and nonisotropic, but still homogeneous, cases are very relevant to many etch and deposition processes.

The dependence of R on $\vec{s}$ accounts for the dependence of the surfaces evolution on it's shape. This dependence can be local as in the case where R depends on the local curvature of the surface. This would involve a dependence on second derivatives of $\vec{r}(t, \vec{s})$ with respect to the components of $\vec{s}$. For example, a bump on the surface (positive curvature) could evolve differently from dip on the surface (negative curvature). The dependence can also be nonlocal in that R at a given position depends on the overall shape of the surface itself. An example would be a surface which evolves to have a bubble which is almost but

not completely closed off. The part of the surface inside the bubble could be "starved" of whatever is causing surface evolution (deposition materials or etch factors) and so evolve much more slowly than other parts of the surface.

The bubble completely closing off brings up the real possibility of the surface changing its topology during the evolution. If the bubble completely closes off then the original manifold has become two manifolds, one with, nominally, the topology of a sphere. Changes in topology are very hard to track using the $\vec{r}(t,\vec{s})$ representation of the surface, points and/or regions of the surface need to be added or removed from consideration during the evolution. The level set method was developed in part to provide an "easy way" of handling topology changes.

To develop the level set method consider a 2D surface in 3D described in x,y,z coordinates by

$$z = h(x,y,t)$$

with h the "height" of the surface at x,y and time t.

Writing things this way assumes the surface does not fold back over itself, i.e., each (x,y) position corresponds to one and only one point on the surface. If the surface does fold over on itself, like, say, a sphere or slightly distorted sphere, then either a coordinate system different than (x,y,z), such as sperical polar, needs to be used or the surface needs to be treated as the union of a set of patches or sub-regions appropriately glued together. Here we consider the simple case where $z = h(x,y,t)$ works, i.e., $\vec{s} = (x,y)$ works to uniquely label positions on the surface.

Define the field

$$\phi(z,y,z,t) = z - h(x,y,t)$$

It is straightforward to show that

$$\hat{n}(x,y,t) = \frac{\vec{\partial}\phi(x,y,z,t)}{\left|\vec{\partial}\phi(x,y,z,t)\right|} = \frac{(-\partial_x h, -\partial_y h, 1)}{\sqrt{1+(\partial_x h)^2+(\partial_y h)^2}}$$

is the local normal to the surface at position $\vec{s} = (x,y)$ at time t. Here $\vec{\partial} = (\partial_x, \partial_y, \partial_z) =$ the 3D gradient.

The important point is that

$$\phi(x,y,z,t) = 0 \Longrightarrow z = h(x,y,t)$$

And so, at any time t, solving for the contour of ϕ with value 0, gives the surface at time t.

Determine how ϕ evolves in time.

$$\Delta\phi = \vec{\partial}\phi \cdot \Delta\vec{r} + \partial_t \phi \Delta t$$

If we started with ϕ satisfying $\phi = 0$ at time t and we want $\phi = 0$ to be the surface at time $t + \Delta t$ then we must have

$$\begin{aligned} \Delta\phi &= \vec{\partial}\phi \cdot \Delta\vec{r} + \partial_t \phi \Delta t = 0 \\ &\text{or} \\ \partial_t \phi \Delta t &= -\vec{\partial}\phi \cdot \Delta\vec{r} \end{aligned}$$

Using $\Delta\vec{r} = R\hat{n}\Delta t$ gives

$$\partial_t \phi = -R\vec{\partial}\phi \cdot \hat{n}$$

after cancelling the Δt factors. But $\hat{n} = \vec{\partial}\phi / \left|\vec{\partial}\phi\right| = \vec{\partial}\phi / \sqrt{\vec{\partial}\phi \cdot \vec{\partial}\phi}$ and so

$$\partial_t \phi = -R\frac{\vec{\partial}\phi \cdot \vec{\partial}\phi}{\sqrt{\vec{\partial}\phi \cdot \vec{\partial}\phi}} = -R\left|\vec{\partial}\phi\right|$$

This is very reminiscent of the WKB result above where the evolution equation depended on $\left|\vec{\partial}S\right|$. Indeed writing things in terms of ray propagation or ray trajectories gave $\left.\frac{\vec{\partial}S(\vec{r})}{\left|\vec{\partial}S(\vec{r})\right|}\right|_{\vec{r}=\vec{r}(\ell)} = \frac{d\vec{r}(\ell)}{d\ell}$. Substituting $d\ell = Rdt$ and rearranging gives $\partial_t \vec{r} = R\vec{\partial}S / \left|\vec{\partial}S\right| = R\hat{n}$ where $\hat{n}$ is the local unit normal to the S =constant surface, which is optics is called the "wavefront".

12 Energy and Momentum Conservation

12.1 Microscopic Newton and Maxwell Momentum Conservation

Here we show conservation of momentum for the "microscopic" (Jackson's nomenclature) Maxwells equations. First do the simple case, i.e., take the matter (point particles with charge and mass) to be classical (non-quantum) and non-relativistic. Conservation of momentum still works out if we do the full deal, i.e., let the matter be relativistic and/or quantum mechanical but then, for the "and" case, the full machinery if quantum field theory is required. I will put this derivation at the end.

12.2 Maxwells Equations (MKS units)

$$\begin{aligned}
\vec{\partial} \times \vec{E}\left(\vec{x}, t\right) &= -\partial_t \vec{B}\left(\vec{x}, t\right) \\
\vec{\partial} \times \vec{H}\left(\vec{x}, t\right) &= \vec{J} + \partial_t \vec{D}\left(\vec{x}, t\right) \\
\vec{\partial} \cdot \vec{D}\left(\vec{x}, t\right) &= \rho\left(\vec{\rho}, t\right) \\
\vec{\partial} \cdot \vec{B}\left(\vec{x}, t\right) &= 0
\end{aligned}$$

where

$$\begin{aligned}
\vec{D}\left(\vec{x}, t\right) &= \epsilon_0 \vec{E}\left(\vec{x}, t\right) \\
\vec{B}\left(\vec{x}, t\right) &= \mu_0 \vec{H}\left(\vec{x}, t\right)
\end{aligned}$$

and for a set $1, \cdots, N$ of point particles with trajectories $\vec{x}_n(t)$ and charges q_n

$$\rho(\vec{x}, t) = \sum_n q_n \delta(\vec{x} - \vec{x}_n(t))$$
$$\vec{J}(\vec{x}, t) = \sum_n q_n (\partial_t \vec{x}_n(t)) \delta(\vec{x} - \vec{x}_n(t))$$

Clearly this satisfies charge conservation.

NOTE: (1) To simplify notation we will drop writing the explicit $\vec{x}$ and t , i.e., $\vec{E}(\vec{x}, t) = \vec{E}$, $\vec{B}(\vec{x}, t) = \vec{B}$, $\cdots, \vec{x}_n(t) = \vec{x}_n$. (2) q_n must account for all the charges that create and interact with fields, I do mean all. (3) The total charge is assumed to be zero although this may not actually be totally necessary.

The standard (free space) form of the momentum density of the electromagnetic field is $\frac{1}{c^2}\vec{E} \times \vec{H} = \epsilon_0 \vec{E} \times \vec{B}$ (units are Newton $\times$ sec/m^3), hence the time rate of change of the total field momentum $\vec{P}$ is

$$\partial_t P_i = \epsilon_0 \int d^3x \left(\partial_t \vec{E} \times \vec{B} + \vec{E} \times \partial_t \vec{B}\right)_i$$

The integral is over all space, the fields vanish at infinity and the subscript i indicates the vector component in the x, y, z direction. Using Maxwell gives

$$\partial_t P_i = \epsilon_0 \int d^3x \left(\frac{1}{\epsilon_0}\left(\vec{\partial} \times \frac{1}{\mu_0}\vec{B} - \vec{J}\right) \times \vec{B} - \vec{E} \times \left(\vec{\partial} \times \vec{E}\right)\right)_i$$

Substituting the identity

$$\left(\vec{V} \times \left(\vec{\partial} \times \vec{V}\right)\right)_i = V_j \partial_i V_j - V_j \partial_j V_i$$

gives

$$\partial_t P_i = \epsilon_0 \int d^3x \left(\frac{1}{\epsilon_0 \mu_0}\left(B_j \partial_j B_i - B_j \partial_i B_j\right) - \left(\vec{J} \times \vec{B}\right)_i - E_j \partial_i E_j + E_j \partial_j E_i\right)$$

Integrating the first and last terms by parts (dropping the surface terms), then using $\vec{\partial} \cdot \vec{B} = 0$, $\vec{\partial} \cdot \vec{E} = \rho/\epsilon_0$ and $B_j \partial_i B_j + E_j \partial_i E_j = \frac{1}{2}\partial_i\left(\vec{B}^2 + \vec{E}^2\right)$ which is a pure divergence and hence becomes a surface integral which vanishes at infinity gives

$$\begin{aligned}
\partial_t \vec{P} &= -\int d^3x \left(\left(\vec{J} \times \vec{B}\right) + \vec{E}\rho\right) \\
&= -\int d^3x \left(\left(\sum_n q_n (\partial_t \vec{x}_n) \delta(\vec{x} - \vec{x}_n(t)) \times \vec{B}\right) + \vec{E} \sum_n q_n \delta(\vec{x} - \vec{x}_n(t))\right) \\
&= -\sum_n q_n (\partial_t \vec{x}_n) \times B(\vec{x}_n(t), t) - \sum_n q_n \vec{E}(\vec{x}_n(t), t) \\
&= -\sum_n \partial_t \vec{p}_n
\end{aligned}$$

where the last line is just Newton since $q_n\left(\partial_t \vec{x}_n\right) \times \vec{B}\left(\vec{x}_n\left(t\right), t\right) + q_n \vec{E}\left(\vec{x}_n\left(t\right), t\right)$ is the force on particle n. And so

$$\partial_t \left(\vec{P} + \sum_n \vec{p}_n \right) = 0$$

The total of the field plus matter momentum is conserved. Note that solving simultaneously the combined Newton and Maxwell system of equations gives all (classical nonrelativistic) absorption and radiation effects. Everything is there. If we replace the nonrelativistic Newtons laws with the relativistic ones, everything still works out, and then all classical relativistic absorption and radiation effects have been accounted for.

12.3 Space Translation Invariance ⇒ Momentum Conservation

For a classical nonrelativistic collection of N particles labeled with $n = 1, \cdots, N$, if their behavior doesn't depend on the position of the system then its invariant under spatial translations where all particle positions $\vec{x}_n \to \vec{x}_n + \vec{a}$ where $\vec{a}$ is the same constant position shift of all the particles. In order to include all the forces on all the particles the potential energy between the particles can only depend on the combinations $\vec{x}_n - \vec{x}_m$. The kinetic energy is automatically invariant since $\partial_t\left(\vec{x}_n + \vec{a}\right) = \partial_t \vec{x}_n$. (Of course if there is also no preferred direction in space then V can't depend on how the entire system is rotated and so V depends only on $|\vec{x}_n - \vec{x}_m|$ which leads to angular momentum conservation) To make things simple work in 1D and so V depends only on $|x_n - x_m|$. (Everything goes through just fine in 2D or 3D or... it's just more notationally cumbersome to write out.)

The Lagrangian for a system of particles interacting via a pair potential V is

$$L = \sum_n \frac{m_n \left(\partial_t x_n\right)^2}{2} - \frac{1}{2} \sum_{n,m} V\left(|x_n - x_m|\right)$$

The factor of $1/2$ in front of V is included to cancel the factor of 2 picked up since we are double counting the potential energy (the sums on n and m are independent and so allow for each pair of n and m values to occur twice). Also we assume $V\left(|x_n - x_n|\right) = V\left(0\right) =$ a constant, nominally 0.

The momentum is defined by

$$p_k \equiv \frac{\partial L}{\partial\left(\partial_t x_k\right)} = m_k \partial_t x_k$$

The equations of motion are

$$\partial_t \frac{\partial L}{\partial\left(\partial_t x_k\right)} = \frac{\partial L}{\partial\left(x_k\right)} \quad \Rightarrow \quad \partial_t p_k = -\frac{1}{2} \sum_{n,m} \frac{\partial}{\partial x_k} V\left(|x_n - x_m|\right)$$

Using $sgn(x) =$ the sign of x which gives $\partial_x |x| = \partial_x (sgn(x)\, x) = 2\delta_D(x)\, x + sgn(x) = sgn(x)$ and $\partial x_n / \partial x_k = \delta_{n,k}$ we get

$$\begin{aligned}
\partial_t p_k &= -\frac{1}{2} \sum_{n,m} \frac{\partial}{\partial x_k} V(|x_n - x_m|) \\
&= -\frac{1}{2} \sum_{n,m} \frac{\partial |x_n - x_m|}{\partial x_k} \frac{\partial V(|x_n - x_m|)}{\partial |x_n - x_m|} \\
&= -\frac{1}{2} \sum_{n,m} (\delta_{n,k} - \delta_{m,k})\, sgn(x_n - x_m) \frac{\partial V(|x_n - x_m|)}{\partial |x_n - x_m|} \\
&= -\frac{1}{2} \left(\sum_m sgn(x_k - x_m) \frac{\partial V(|x_k - x_m|)}{\partial |x_k - x_m|} - \sum_n sgn(x_n - x_k) \frac{\partial V(|x_n - x_k|)}{\partial |x_n - x_k|} \right) \\
&= -\sum_m sgn(x_k - x_m) \frac{\partial V(|x_k - x_m|)}{\partial |x_k - x_m|}
\end{aligned}$$

And so the time derivative of the total momentum $\sum_k p_k$ satisfies

$$\begin{aligned}
\partial_t \sum_k p_k &= -\sum_{k,m} sgn(x_k - x_m) \frac{\partial V(|x_k - x_m|)}{\partial |x_k - x_m|} \\
&= 0
\end{aligned}$$

since $\partial V(|x_k - x_m|) / \partial |x_k - x_m|$ is symmetric under $k \leftrightarrow m$ but $sgn(x_k - x_m)$ is antisymmetric under $k \leftrightarrow m$.

12.4 Microscopic Newton and Maxwell Energy Conservation

The standard free space form of the electromagnetic energy density is $\frac{\epsilon_0}{2}\vec{E}^2 + \frac{1}{2\mu_0}\vec{B}^2$ (units are Joule/m^3 = Newton/m^2). The time rate of change of the total field energy E (do not confuse the field energy E with the field itself $\vec{E}$) is

$$\begin{aligned}
\partial_t E &= \int d^3x \left(\epsilon_0 \vec{E} \cdot \partial_t \vec{E} + \frac{1}{\mu_0} \vec{B} \cdot \partial_t \vec{B} \right) \\
&= \frac{1}{\mu_0} \int d^3x \left(\vec{E} \cdot \left(\vec{\partial} \times \vec{B} \right) - \vec{B} \cdot \left(\vec{\partial} \times \vec{E} \right) \right) - \int d^3x \vec{E} \cdot \vec{J}
\end{aligned}$$

after substituting Maxwells equations. Using the identity

$$\vec{a} \cdot \left(\vec{\partial} \times \vec{b} \right) - \vec{b} \cdot \left(\vec{\partial} \times \vec{a} \right) = -\vec{\partial} \cdot \left(\vec{a} \times \vec{b} \right)$$

gives

$$\partial_t E = -\frac{1}{\mu_0} \int d^3x \vec{\partial} \cdot \left(\vec{E} \times \vec{B} \right) - \int d^3x \vec{E} \cdot \vec{J}$$

But

$$\int d^3x\vec{\partial}\cdot\left(\vec{E}\times\vec{B}\right)=\int_{|\vec{x}|=\infty}dS\hat{r}\cdot\left(\vec{E}\times\vec{B}\right)=0$$

where dS is the area element on sphere at $|\vec{x}|=\infty$ and $\hat{r}$ is the outward normal to the sphere surface. The integral vanishes since the fields vanish faster than $1/r^2$ as $\vec{r}$ approaches infinity.

Substituting for $\vec{J}$ from above gives

$$\begin{aligned}\int d^3x\vec{E}\cdot\vec{J} &= \int d^3x\vec{E}\cdot\left(\sum_n q_n\left(\partial_t\vec{x}_n\left(t\right)\right)\delta\left(\vec{x}-\vec{x}_n\left(t\right)\right)\right)\\ &= \sum_n q_n\left(\partial_t\vec{x}_n\left(t\right)\right)\int d^3x\vec{E}\left(\vec{x}\right)\cdot\delta\left(\vec{x}-\vec{x}_n\left(t\right)\right)\\ &= \sum_n \partial_t\vec{x}_n\left(t\right)\left(q_n\vec{E}\left(\vec{x}_n\left(t\right),t\right)\right)\end{aligned}$$

after using the electric force $\vec{F}_E$ on a particle with charge q is $q\vec{E}$. The magnetic force on a particle with charge q is $\vec{F}_B=q\partial_t\vec{x}\times\vec{B}$ which is perpendicular to the motion of the particle, hence it follows that the work done on the particle by the magnetic field $\vec{F}_B\cdot d\vec{x}=\vec{F}_B\cdot\partial_t\vec{x}dt=0$ and therefore the total (electric and magnetic) work done on the particle is $\vec{F}_E\cdot d\vec{x}=q\vec{E}\cdot\partial_t\vec{x}\left(t\right)dt$. Putting this altogether we have

$$0=\partial_t E+\int d^3x\vec{E}\cdot\vec{J}=\partial_t E+\sum_n\partial_t\vec{x}_n\left(t\right)\left(q_n\vec{E}\left(\vec{x}_n\left(t\right),t\right)\right)$$

Again this accounts for the total of all (classical nonrelativistic) absorption and radiation effects. If we replace the nonrelativistic Newtons laws with the relativistic ones, everything still works out, and then all classical relativistic absorption and radiation effects have been accounted for.

NOTE: Working the above analysis backwards, i.e., starting $\int d^3x\vec{J}\cdot\vec{E}$ is the time rate of change of the work done on the particles due the fields acts as a derivation, via energy conservation, of the expression for E for the electromagnetic field and of the Poynting Vector $\tilde{}\vec{E}\times\vec{B}$. This is how it is done in Jackson's book "Classical Electrodynamics" for example.

12.5 Time Translation Invariance ⇒ Energy Conservation

For a classical nonrelativistic collection of N particles labeled with $n=1,\cdots,N$, if their behavior doesn't depend on when the system is intialized and started then the system is invariant under time translations where all particle positions $\vec{x}_n\left(t\right)\rightarrow\vec{x}_n\left(t+T\right)$ where T is a constant time shift. Initial conditions are then $\vec{x}_n\left(0\right)\rightarrow\vec{x}_n\left(T\right)$, i.e., instead of starting the system at $t=0$, it is started at $t=T$. As in the case of space rotation and translation invariance, as above, the potential energy between the particles can only depend on

the combinations $|\vec{x}_n(t) - \vec{x}_m(t)|$ hence the potential energy is invariant under time translation since $V(|\vec{x}_n(t) - \vec{x}_m(t)|) \rightarrow V(|\vec{x}_n(t+T) - \vec{x}_m(t+T)|) = V(|\vec{x}_n(t) - \vec{x}_m(t)|)$. That is, V can only depend explicitly on the distances between particles and not on when they are at a given distance from each other. If V depended independently on the distances and on time as well, i.e., $V(|\vec{x}_n(t) - \vec{x}_m(t)|, t)$ then energy is not conserved because the energy of the source causing the explicit time dependence of the potential energy has not be included. The kinetic energy is automatically invariant since $\partial_t \vec{x}_n(t+T) = \partial_{t+T} \vec{x}_n(t+T)$, i.e., it is the velocity at time $t+T$. Given that both the kinetic and potential energies are time translation invariant the total energy is time translation invariant.

NOTE: The entire argument effectively reduces to V is time translation invariant when it is a function only of the distances between the particles. We now show that in this case energy is conserved.

To make things simple work in 1D and so V depends only on $x_n - x_m$. (Everything goes through just fine in 2D or 3D or... it's just more notationally cumbersome to write out.)

The Lagrangian for a system of particles interacting via a pair potential V is, as above,

$$L = \sum_n \frac{m_n (\partial_t x_n)^2}{2} - \frac{1}{2} \sum_{n,m} V(|x_n - x_m|)$$

(Again, to avoid cluttered notation we are not writing out the explicit t dependence of x_n)

The total energy which in this case is also called the Hamiltonian, H, is defined by

$$H = \sum_n p_n (\partial_t x_n) - L$$

where, as above,

$$p_n \equiv \frac{\partial L}{\partial (\partial_t x_n)} = m_n \partial_t x_n$$

and so

$$\begin{aligned} H &= \sum_n m_n (\partial_t x_n)^2 - \sum_n \frac{1}{2} m_n (\partial_t x_n)^2 + \frac{1}{2} \sum_{n,m} V(|x_n - x_m|) \\ &= \sum_n \frac{1}{2} m_n (\partial_t x_n)^2 + \frac{1}{2} \sum_{n,m} V(|x_n - x_m|) \end{aligned}$$

which is the standard result: Total Energy = Kinetic Energy + Potential Energy.

Taking the time derivative gives

$$\begin{aligned}
\partial_t H &= \sum_n m_n \left(\partial_t x_n\right)\left(\partial_t^2 x_n\right) + \frac{1}{2}\sum_{n,m} \partial_t V\left(|x_n - x_m|\right) \\
&= \sum_n m_n \left(\partial_t x_n\right)\left(\partial_t^2 x_n\right) + \frac{1}{2}\sum_{n,m}\sum_k \left(\partial_t x_k\right)\frac{\partial}{\partial x_k} V\left(|x_n - x_m|\right) \\
&= \sum_n m_n \left(\partial_t x_n\right)\left(\partial_t^2 x_n\right) + \frac{1}{2}\sum_k \left(\partial_t x_k\right)\sum_{n,m}\left(\delta_{k,n} - \delta_{k,m}\right) sgn\left(x_n - x_m\right)\frac{\partial V\left(|x_n - x_m|\right)}{\partial |x_n - x_m|} \\
&= \sum_k m_k \left(\partial_t x_k\right)\left(\partial_t^2 x_k\right) + \sum_k \left(\partial_t x_k\right)\sum_m sgn\left(x_k - x_m\right)\frac{\partial V\left(|x_k - x_m|\right)}{\partial |x_k - x_m|} \\
&= \sum_k \left(\partial_t x_k\right)\left(m_k\left(\partial_t^2 x_k\right) + \sum_m sgn\left(x_k - x_m\right)\frac{\partial V\left(|x_k - x_m|\right)}{\partial |x_k - x_m|}\right) \\
&= \sum_k \left(\partial_t x_k\right)\left(\partial_t p_k + \sum_m sgn\left(x_k - x_m\right)\frac{\partial V\left(|x_k - x_m|\right)}{\partial |x_k - x_m|}\right) \\
&= 0 \text{ via equations of motion from above}
\end{aligned}$$

12.6 Full Relativistic Field Theory Approach

This is the fun one. The action for a fully relativistic field theory is

$$\begin{aligned}
S &= \int d^4x L_{total} \\
&= \int d^4x \left(L_{fields} + L_{matter} + L_{interaction}\right)
\end{aligned}$$

where L is the Lagrangian (density). For example for QED

$$L_{total}^{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}\left(i\gamma^\mu\left(\partial_\mu - ieA_\mu\right) - m\right)\psi$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

with A_μ the 4-vector potential and $\psi$the Dirac (Grassman) field. Note that we are working in "field theorist units" where all numerical constants are unity, i.e., $c = \hbar = \epsilon_0 = \mu_0 = ... = 1$.

But to be general we can let $\phi\left(\vec{x}, t\right) \equiv \phi\left(x\right)$ be a generic field and L_{total} will depend on $\phi\left(x\right)$ and $\partial\phi\left(x\right)$ where $\partial\phi\left(x\right)$ is shorthand all four derivatives $\partial_\mu\phi$, i.e., let

$$L_{total} = L\left(\partial\phi\left(x\right), \phi\left(x\right)\right)$$

Symmetries and/or invariances imply conservation laws via Noethers theorem. Here the invariance follows from the fact that the action S is invariant under a

change of the dummy integration variable d^4x

$$\begin{aligned} S &= \int d^4x L\left(\partial\phi\left(x\right),\phi\left(x\right)\right) \\ &= \int d^4x' L\left(\partial'\phi\left(x'\right),\phi\left(x'\right)\right) \\ &= S' \end{aligned}$$

Now let $x'_\mu = x_\mu + \varepsilon_\mu\left(x\right)$ with $\varepsilon_\mu\left(x\right)$ infinitesimal.

$$\begin{aligned} d^4x' &= \left(1+\partial^\mu\varepsilon_\mu\right)d^4x \\ L\left(\partial'\phi\left(x'\right),\phi\left(x'\right)\right) &= L\left(\partial\phi,\phi\right) + \frac{\partial L}{\partial\phi}\left(\partial_\mu\phi\right)\varepsilon^\mu + \frac{\partial L}{\partial\left(\partial_\nu\phi\right)}\left(\partial_\mu\partial_\nu\phi\right)\varepsilon^\mu \end{aligned}$$

where all the functions on the right hand side depend on x. Since this will be under the integral $\int d^4x$ we can integrate the last term by parts and keeping only terms to first order in ε gives

$$\begin{aligned} \left(1+\partial^\mu\varepsilon_\mu\right)L\left(\partial'\phi\left(x'\right),\phi\left(x'\right)\right) &= L\left(\partial\phi,\phi\right) + \partial^\mu\varepsilon_\mu L\left(\partial\phi,\phi\right) - \frac{\partial L}{\partial\left(\partial_\nu\phi\right)}\left(\partial_\mu\phi\right)\partial_\nu\varepsilon^\mu \\ &\quad + \underbrace{\left(\frac{\partial L}{\partial\phi} - \partial_\nu\frac{\partial L}{\partial\left(\partial_\nu\phi\right)}\right)}_{=0 \text{ Equation of motion}}\left(\partial_\mu\phi\right)\varepsilon^\mu \end{aligned}$$

The last term vanishes due to the equation of motion and so

$$S' - S = 0 = \int d^4x\left(\delta^\nu_\mu L\left(\partial\phi,\phi\right) - \frac{\partial L}{\partial\left(\partial_\nu\phi\right)}\left(\partial_\mu\phi\right)\right)\partial_\nu\varepsilon^\mu$$

Integrating by parts and noting that the result holds for arbitrary ε_μ yields the result

$$\partial_\nu\left(\frac{\partial L}{\partial\left(\partial_\nu\phi\right)}\left(\partial_\mu\phi\right) - \delta^\nu_\mu L\left(\partial\phi,\phi\right)\right) = 0$$

And so the total "energy-momentum" tensor

$$T_{\mu\nu} = \frac{\partial L}{\partial\left(\partial_\mu\phi\right)}\left(\partial_\nu\phi\right) - \delta^\mu_\nu L\left(\partial\phi,\phi\right)$$

is conserved since it's 4-divergence vanishes

$$\partial^\mu T_{\mu\nu} = 0$$

If ϕ is a collection of fields indexed by i then

$$\begin{aligned} T^\mu_\nu &= \sum_i \frac{\partial L}{\partial\left(\partial_\mu\phi_i\right)}\left(\partial_\nu\phi_i\right) - \delta^\mu_\nu L\left(\partial\phi,\phi\right) \\ T_{\mu\nu} &= \sum_i \frac{\partial L}{\partial\left(\partial^\mu\phi_i\right)}\left(\partial_\nu\phi_i\right) - g_{\mu\nu}L\left(\partial\phi,\phi\right) \end{aligned}$$

where $g_{\mu\nu}$ is the Minkowski metric, $g_{00} = 1$, $g_{11} = g_{22} = g_{33} = 1$ with all other entries 0.

For Maxwell with $L = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ and $\phi_1 = A_0, \phi_2 = A_1$ etc, then (with i, j summed from 1 to 3) and symmetrizing the results we get

$$\begin{aligned} L &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} = -\frac{1}{4}\left(-2F_{0i}^2 + F_{ij}^2\right) = \frac{1}{2}\left(\vec{E}^2 - \vec{B}^2\right) \\ T_{00} &= -F_{0\mu}F_0{}^{\mu} - \frac{1}{2}\left(\vec{E}^2 - \vec{B}^2\right) = F_{0i}^2 - \frac{1}{2}\left(\vec{E}^2 - \vec{B}^2\right) = \frac{1}{2}\left(\vec{E}^2 + \vec{B}^2\right) \\ T_{0i} &= -F_{0\mu}F_i^{\mu} = F_{0j}F_{ij} = \varepsilon_{ijk}E_jB_k = \left(\vec{E} \times \vec{B}\right)_i \end{aligned}$$

which is the standard result (in units where ϵ_0 and μ_0 are 1).

13 Capacitor Energy Loss

Electric power, P, (Watts) is given by the product of current, I; (Amps) and voltage, V (Volts)

$$P = IV$$

This power can be dissipated, as in a resistor, or stored as in a capacitor or inductor. The interesting thing is that in any resistor capacitor (RC) circuit, during charging, the capacitor always stores only half the energy supplied by the voltage source, e.g., a battery. The proof is simple.

Energy, E (Joules) is the time integral of the power. For simplicity we assume the energy is 0 at time t = 0 and

then integrate to t = ∞

$$E = \int_0^\infty dtP = \int_0^\infty dtIV$$

Consider an uncharged capacitor wired in series with a resistor and a battery of voltage V_{batt}. The current is given by the time rate of change of the electric charge $Q(t)$ that flows past any point in the circuit, i.e.,

$$I(t) = \frac{dQ\,(t)}{dt}$$

For the battery the total power supplied to the circuit from $t = 0$ to $t = \infty$ is,assuming V_{batt} = constant,

$$\begin{aligned} E_{batt} &= \int_0^\infty dtI\,(t)\,V_{batt} \\ &= V_{batt}\int_0^\infty dt\frac{dQ\,(t)}{dt} \\ &= V_{batt}Q_{batt} \end{aligned}$$

where Q_{batt} is the total charge supplied by the battery.

The energy stored in the capacitor, E_C, is given by

$$E_C = \int_0^\infty dt I(t) V_C(t) = \int_0^\infty dt \frac{dQ(t)}{dt} V_C(t)$$

but the voltage across the capacitor

$$V_C(t) = \frac{Q_C(t)}{C}$$

where $Q_C(t)$ is the charge on the capacitor at time t which equals the total charge $Q(t)$ supplied by the battery up to time t and C is the capacitance, which is constant. Hence we have

$$\begin{aligned} E_C &= \int_0^\infty dt \frac{dQ(t)}{dt} V_C(t) \\ &= \frac{1}{C} \int_0^\infty dt \frac{dQ(t)}{dt} Q(t) \\ &= \frac{Q_{batt}^2}{2C} \\ &= \frac{1}{2} V_{batt} Q_{batt} \end{aligned}$$

That last step follows since at $t = \infty$ the capacitor is fully charged, i.e., $Q_C(t = \infty) = Q_{batt}$, the current ,$I(t = \infty) = 0$ and the voltage across the capacitor is equal (and opposite) to the voltage across the battery, i.e,. $V_C = |V_{batt}|$. Hence the amount of energy stored by the capacitor is always 1/2 the amount supplied to the circuit by the battery (or other voltage source) independent of the size (Ohms) of the resistor. The other half the supplied energy is dissipated in the resistor.

This can be shown in a more explicit way by solving the differential equation (DE) for the circuit which is

$$V_{batt} = I(t) R + \frac{Q_C(t)}{C}$$

where R is the resistance (Ohms).

Let

$$Q_C(t) = C V_{batt} + q(t)$$

Then the DE becomes

$$\frac{dq(t)}{dt} R + \frac{1}{C} q(t) = 0$$

which has the general solution

$$q(t) = q(0) e^{-t/RC}$$

The initial condition is $Q_C(0) = 0$, hence $q(0) = -C V_{batt} = Q_{batt}$. This gives

$$Q_C(t) = Q_{batt} \left(1 - e^{-t/RC}\right)$$

Now for the energy dissipated by the resistor we have, using $V_R(t) = I(t) R$

$$\begin{aligned} E_R &= \int_0^\infty dt I(t) V_R(t) \\ &= R \int_0^\infty dt I(t)^2 \\ &= R \int_0^\infty dt \left(\frac{dQ(t)}{dt} \right)^2 \\ &= R \int_0^\infty dt \left(\frac{dq(t)}{dt} \right)^2 \\ &= \frac{RQ_{batt}^2}{(RC)^2} \int_0^\infty dt e^{-2t/RC} \\ &= \frac{Q_{batt}^2}{RC^2} \frac{RC}{2} \\ &= \frac{Q_{batt}^2}{2C} \\ &= \frac{1}{2} V_{batt} Q_{batt} \end{aligned}$$

Note that the R cancels out. Hence independent of the value of R, half the energy supplied by the battery (or voltage source) is always dissipated in the resistor and half is stored in the capacitor. The important fact is that the energy $V_{batt}Q_{batt}/2$ dissipated by the resistor becomes heat. This has important implications for thermal management of semiconductor computer chips, i.e., how to keep chips cool and not waste energy on heat. Every transistor on a chip, not to mention other parts of the circuitry, has a resistance and a capacitance. And, as we have just shown, half the energy supplied by the voltage source always ends up as heat as each capacitor is charged. So half the energy supplied to do computation automatically ends up heating the chip.

14 Lagrangian, Action, Euler-Lagrange, Derivatives

First, some background. The Hamiltonian is defined as the total energy, kinetic plus potential, written in terms of generalized position coordinates and canonical momenta. Often, but not always, these coordinates and momenta are the standard x, y, z coordinates and mass times velocity, respectively, but not always. In quantum mechanics the Hamiltonian is represented as a operator and it propagates wave functions forward (or backward) in time. The key point here being that the Hamiltonian specifically picks out time. But relativity treats time and space on an equal footing. This means that relativistic quantum mechanics and quantum field theory done using the Hamiltonian approach had the added burden of having to show explicitly that the results produced were relativistically invariant. This fact led Feynman, following Schwinger, to wonder how to

do quantum mechanics using the Lagrangian, which is the kinetic energy minus the potential energy, expressed in terms of (generalized) coordinates and velocities (instead of momenta) and which treats space and time on an equal footing hence yielding results that are obviously relativistically invariant from the get go. This lead to the path integral approach to quantum mechanics (discussed in the next section) which directly generalizes to the functional integral approach to quantum field theory which is the standard approach used today.

14.1 Lagrangian

The Lagrangian $\mathcal{L}$ for classical mechanics is defined as the kinetic energy (KE) minus the potential energy (PE)

$$\mathcal{L} = KE - PE$$

in the simplest case in 3D in Cartesian coordinates for a single particle of mass m subject to the potential energy $V(\vec{r}, t)$, where we have included an explicit time dependent along with the dependence on $\vec{r}$, we have

$$\mathcal{L}(\vec{v}(t), \vec{r}(t)) = \frac{1}{2} m \vec{v}(t)^2 - V(\vec{r}(t))$$

where $\vec{r}(t)$ is the particle trajectory, i.e., position as a function of time and $\vec{v}(t) = \partial_t \vec{r}(t)$ is its associated velocity. The same $KE - PE$ form holds for multiple particle systems in arbitrary dimensions and using curvilinear coordinates (cylindrical, spherical polar, etc.).

14.2 Action

In classical mechanics the "Action", often denoted with the letter S, is defined as the integral of the Lagrangian over time.

$$S = \int_{t_s}^{t_f} dt \mathcal{L}(\partial_t \vec{r}(t), \vec{r}(t))$$

where t_s is the starting time and t_f is the end or finish time. Note that S is a "functional" of the trajectory $\vec{r}(t)$, that is, because of the integral over time S depends on the entire shape or functional form of $\vec{r}(t)$ from $t = t_s$ to t_f while S is just a function of t_s and $t_{f.}$. Given specific starting and ending positions, $\vec{r}_s = r(t = t_s)$ and $\vec{r}_f = \vec{r}(t - t_f)$, S is also a function of $\vec{r}_s$ and $\vec{r}_f$.

NOTE: The functional dependence of S on $\vec{r}(t)$ is often indicated as $S(\vec{r}(\cdot))$ to show that S depends on the functional form of $\vec{r}(t)$, it does not depend on t.

14.3 Euler-Lagrange

Question: For a given $\mathcal{L}$ what trajectory, i.e., what functional form of $\vec{r}(t)$ extremizes the Action. To find the equation that $\vec{r}(t)$ must satisfy we take the functional derivative of S w.r.t $\vec{r}(t')$, which we denote $\delta/\delta r_n(t')$ where n labels

the x, y, z,etc. component of $\vec{r}$. A functional derivative w.r.t. a given function acts like an ordinary derivative (using chain rule) until it acts on itself, then it acts like a Dirac delta function, e.g., $\delta \mathrm{f}(t) / \delta \mathrm{f}(t') = \delta_D (t - t')$.

Taking the functional derivative of S w.r.t $r_n(t')$ then gives

$$\begin{aligned}
\frac{\delta S}{\delta r_n(t')} &= \frac{\delta}{\delta r_n(t')} \int_{t_s}^{t_f} dt \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right) \\
&= \int_{t_s}^{t_f} dt \frac{\delta}{\delta r_n(t')} \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right) \\
&= \int_{t_s}^{t_f} dt \left(\frac{\partial \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right)}{\partial\left(\partial_t r_n(t)\right)} \frac{\delta\left(\partial_t r_n(t)\right)}{\delta r_n(t')} + \frac{\partial \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right)}{\partial\left(r_n(t)\right)} \frac{\delta r_n(t)}{\delta r_n(t')} \right)
\end{aligned}$$

but $\frac{\delta(\partial_t r_n(t))}{\delta r_n(t')} = \partial_t \frac{\delta r_n(t)}{\delta r_n(t')}$.

$$\begin{aligned}
\frac{\delta S}{\delta r_n(t')} &= \int_{t_s}^{t_f} dt \left(\frac{\partial \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right)}{\partial\left(\partial_t r_n(t)\right)} \partial_t \frac{\delta r_n(t)}{\delta r_n(t')} + \frac{\partial \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right)}{\partial\left(r_n(t)\right)} \frac{\delta r_n(t)}{\delta r_n(t')} \right) \\
&= \int_{t_s}^{t_f} dt \left(\frac{\partial \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right)}{\partial\left(\partial_t r_n(t)\right)} \partial_t \delta_D (t - t') + \frac{\partial \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right)}{\partial\left(r_n(t)\right)} \delta_D (t - t') \right)
\end{aligned}$$

Integrating the first term by parts and dropping the endpoint terms (we are varying $\vec{r}$ at t' and taking $t_s < t' < t_f$ so there is no variation at the endpoints $\vec{r}_s$ and $\vec{r}_f$ we get

$$\begin{aligned}
\frac{\delta S}{\delta r_n(t')} &= \int_{t_s}^{t_f} dt \delta_D (t - t') \left(-\partial_t \frac{\partial \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right)}{\partial\left(\partial_t r_n(t)\right)} + \frac{\partial \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right)}{\partial\left(r_n(t)\right)} \right) \\
&= -\partial_{t'} \frac{\partial \mathcal{L}\left(\partial_{t'} \vec{r}(t'), \vec{r}(t')\right)}{\partial\left(\partial_{t'} r_n(t')\right)} + \frac{\partial \mathcal{L}\left(\partial_{t'} \vec{r}(t'), \vec{r}(t')\right)}{\partial\left(r_n(t')\right)}
\end{aligned}$$

Setting $\frac{\delta S}{\delta r_n(t')} = 0$ to get the condition for $\vec{r}(t)$ to extremize S gives, after dropping the prime on t

$$\partial_t \frac{\partial \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right)}{\partial\left(\partial_t r_n(t)\right)} - \frac{\partial \mathcal{L}\left(\partial_t \vec{r}(t), \vec{r}(t)\right)}{\partial\left(r_n(t)\right)} = 0$$

This is known as the Euler-Lagrange equation or equations, since there is one for each n. Strictly speaking ∂_t should be written as d/dt but usually it is written as just ∂_t.

14.3.1 Example

The Euler-Lagrange equation for

$$\mathcal{L} = \frac{1}{2} m \left(\partial_t \vec{r}(t)\right)^2 - V\left(\vec{r}(t), t\right)$$

is

$$m \partial_t^2 \vec{r}(t) = -\vec{\partial} V\left(\vec{r}(t), t\right)$$

which is just Newtons Law: $\vec{F} = m\vec{a}$.

To be explicit about notation,

$$\vec{\partial} V\left(\vec{r}\left(t\right), t\right) \equiv \vec{\partial} V\left(\vec{r}, t\right)\Big|_{\vec{r}=\vec{r}(t)}$$

that is, after taking the gradient with respect to $\vec{r}$, replace $\vec{r}$ with $\vec{r}(t)$ in V. Some mathematics texts write this as $\left(\vec{\partial} V\right)\left(\vec{r}\left(t\right), t\right)$ to indicated taking the gradient first.

As discussed in the section below on stationary phase and classical mechanics, the Euler-Lagrange equation being the extremum of the Action explains, at least in part, how quantum mechanics reduces to classical mechanics.

Comment: In quantum field theory the Lagrangian is a function of the fields and of both the space derivatives and time derivatives of the fields and the Action is the integral of the Lagrangian over space and time. Hence if the Lagrangian is relativistically invariant then so is the Action and nominally so then are all the results derived from the Action.

14.4 Derivatives of the Action

The solution of Euler-Lagrange is the classical (Newtons Law) trajectory, $\vec{r}_{cl}\left(t\right)$. The Euler-Lagrange equation follows from setting the functional derivative of the Action to zero. For the classical Action, i.e., the Action evaluated for $\vec{r}(t) = \vec{r}_{cl}\left(t\right)$,

$$S_{cl} = S\left(\vec{r}_{cl}\left(\cdot\right)\right)$$

what do derivatives of S_{cl} w.r.t. the starting and ending positions and times yield?

We will work in 1D so that $\vec{r}_{cl}\left(t\right)$ is replaced with $x_{cl}\left(t\right)$ and just do the derivatives w.r.t. x_f and t_f. The Euler-Lagrange equation is generally a second order differential equation and so generally required two boundary or initial and final conditions to completely specify a given solution. This means that $x_{cl}\left(t\right)$ along with being a function of t also depends on the starting position and time x_s and t_s and the final position and time x_f and t_f, hence we should really write $x_{cl}\left(t\right)$ as $x_{cl}\left(t, x_f, t_f, x_s, t_s\right)$. Here we just consider the derivatives of $S\left(x_{cl}\left(t\right)\right)$ w.r.t. x_f and t_f. The results for the derivative of $S\left(x_{cl}\left(t\right)\right)$ w.r.t. x_s and t_s are

evaluated exactly the same way.

$$\begin{aligned}
\frac{\partial}{\partial x_f} S\left(x_{cl}\left(\cdot\right)\right) &= \frac{\partial}{\partial x_f}\int_{t_s}^{t_f} dt \mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right) \\
&= \int_{t_s}^{t_f} dt \frac{\partial}{\partial x_f}\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right) \\
&= \int_{t_s}^{t_f} dt \left(\frac{\partial\left(\partial_t x_{cl}\left(t\right)\right)}{\partial x_f}\frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)} + \frac{\partial x_{cl}\left(t\right)}{\partial x_f}\frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(x_{cl}\left(t\right)\right)}\right) \\
\text{use } \frac{\partial\left(\partial_t x_{cl}\left(t\right)\right)}{\partial x_f} &= \partial_t \frac{\partial x_{cl}\left(t\right)}{\partial x_f} \text{ since } t \text{ and } x_f \text{ are independent variables} \\
&= \int_{t_s}^{t_f} dt \left(\left(\partial_t \frac{\partial x_{cl}\left(t\right)}{\partial x_f}\right)\frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)} + \frac{\partial x_{cl}\left(t\right)}{\partial x_f}\frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(x_{cl}\left(t\right)\right)}\right) \\
&\quad \text{Integrate first term by parts} \\
&= \left(\frac{\partial x_{cl}\left(t\right)}{\partial x_f}\frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)}\right)\Bigg|_{t_s}^{t_f} \\
&\quad - \int_{t_s}^{t_f} dt \frac{\partial x_{cl}\left(t\right)}{\partial x_f}\left(\partial_t \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)} - \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)}\right)
\end{aligned}$$

The integral term vanishes because $x_{cl}\left(t\right)$ satisfies the Euler-Lagrange equation. The first term when evaluated at t_s also vanishes since $\partial x_{cl}\left(t_s\right)/\partial x_f = \partial x_s/\partial x_f = 0$ since x_s and x_f are independent of each other. That leaves

$$\begin{aligned}
\frac{\partial}{\partial x_f} S\left(x_{cl}\left(\cdot\right)\right) &= \frac{\partial x_{cl}\left(t_f\right)}{\partial x_f}\frac{\mathcal{L}\left(\partial_t x_{cl}\left(t_f\right), x_{cl}\left(t_f\right)\right)}{\partial\left(\partial_t x_{cl}\left(t_f\right)\right)} \\
&= \frac{\partial x_f}{\partial x_f}\frac{\mathcal{L}\left(\partial_t x_{cl}\left(t_f\right), x_{cl}\left(t_f\right)\right)}{\partial\left(\partial_t x_{cl}\left(t_f\right)\right)} \\
&= \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t_f\right), x_{cl}\left(t_f\right)\right)}{\partial\left(\partial_t x_{cl}\left(t_f\right)\right)} \\
&= \text{canonical momentum at } t_f
\end{aligned}$$

NOTE: Instead of integrating by parts above we could simply have used $\frac{\mathcal{L}(\partial_t x_{cl}(t), x_{cl}(t))}{\partial(\partial_t x_{cl}(t))} = \partial_t \frac{\mathcal{L}(\partial_t x_{cl}(t), x_{cl}(t))}{\partial(\partial_t x_{cl}(t))}$ which turns the integrand into an exact time derivative..

For $\mathcal{L} = \frac{1}{2} m \left(\partial_t x\right)^2 - V\left(x\left(t\right), t\right)$ we have explicitly

$$\frac{\partial}{\partial x_f} S\left(x_{cl}\left(\cdot\right)\right) = m \partial_t x_{cl}\left(t_f\right) = m v_f = p_f$$

Now take the derivative w.r.t. t_f

$$\begin{aligned}
\frac{\partial}{\partial t_f} S\left(x_{cl}\left(\cdot\right)\right) &= \frac{\partial}{\partial t_f} \int_{t_s}^{t_f} dt \mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right) \\
&= \mathcal{L}\left(\partial_t x_{cl}\left(t_f\right), x_{cl}\left(t_f\right)\right) + \int_{t_s}^{t_f} dt \frac{\partial}{\partial t_f} \mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right) \\
&= \mathcal{L}\left(\partial_t x_{cl}\left(t_f\right), x_{cl}\left(t_f\right)\right) \\
&\quad + \int_{t_s}^{t_f} dt \left(\frac{\partial\left(\partial_t x_{cl}\left(t\right)\right)}{\partial t_f} \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)} + \frac{\partial x_{cl}\left(t\right)}{\partial t_f} \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(x_{cl}\left(t\right)\right)} \right) \\
&= \mathcal{L}\left(\partial_t x_{cl}\left(t_f\right), x_{cl}\left(t_f\right)\right) \\
&\quad + \int_{t_s}^{t_f} dt \left(\left(\partial_t \frac{\partial x_{cl}\left(t\right)}{\partial t_f} \right) \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)} + \frac{\partial x_{cl}\left(t\right)}{\partial t_f} \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(x_{cl}\left(t\right)\right)} \right) \\
&\quad \text{Integrate first term by parts} \\
&= \mathcal{L}\left(\partial_t x_{cl}\left(t_f\right), x_{cl}\left(t_f\right)\right) + \left(\frac{\partial x_{cl}\left(t\right)}{\partial t_f} \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)} \right)\Bigg|_{t_s}^{t_f} \\
&\quad - \int_{t_s}^{t_f} dt \frac{\partial x_{cl}\left(t\right)}{\partial t_f} \left(\partial_t \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)} - \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)} \right)
\end{aligned}$$

Again, the integral term vanishes since $x_{cl}\left(t\right)$ satisfies the Euler-Lagrange equation. The first term when evaluated at t_s also vanishes since $x_{cl}\left(t_f\right)$ is fixed at value x_f independent of t_s. This leaves

$$\frac{\partial}{\partial t_f} S\left(x_{cl}\left(\cdot\right)\right) = \mathcal{L}\left(\partial_t x_{cl}\left(t_f\right), x_{cl}\left(t_f\right)\right) + \left(\frac{\partial x_{cl}\left(t_f\right)}{\partial t_f} \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t_f\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)} \right)$$

But for $t \to t_f$ we have

$$\begin{aligned}
x_{cl}\left(t\right) &= x_f + v_f\left(t - t_f\right) + \cdots \\
&\quad \text{so} \\
\frac{\partial x_{cl}\left(t \to t_f\right)}{\partial t_f} &= -v_f
\end{aligned}$$

hence

$$\begin{aligned}
\frac{\partial}{\partial t_f} S\left(x_{cl}\left(\cdot\right)\right) &= \mathcal{L}\left(\partial_t x_{cl}\left(t_f\right), x_{cl}\left(t_f\right)\right) - v_f \frac{\mathcal{L}\left(\partial_t x_{cl}\left(t_f\right), x_{cl}\left(t\right)\right)}{\partial\left(\partial_t x_{cl}\left(t\right)\right)} \\
&= -E_f \\
&= -\text{ Energy at } t_f
\end{aligned}$$

NOTE: Again, instead of integrating by parts above we could simply have used $\frac{\mathcal{L}\left(\partial_t x_{cl}(t), x_{cl}(t)\right)}{\partial\left(\partial_t x_{cl}(t)\right)} = \partial_t \frac{\mathcal{L}\left(\partial_t x_{cl}(t), x_{cl}(t)\right)}{\partial\left(\partial_t x_{cl}(t)\right)}$ which automatically cancels the relevant terms.

For $\mathcal{L} = \frac{1}{2}m\left(\partial_t x\right)^2 - V\left(x\left(t\right), t\right)$ we have explicitly

$$\begin{aligned}\frac{\partial}{\partial t_f} S\left(x_{cl}\left(\cdot\right)\right) &= \frac{1}{2}mv_f^2 - V\left(x_f, t_f\right) - mv_f^2 \\ &= -\left(\frac{1}{2}mv_f^2 + V\left(x_f, t_f\right)\right)\end{aligned}$$

15 Path Integrals

15.1 Derivation

Path integrals are usually considered to be related to Quantum Mechanics but that is not necessary at all. Although they do get used for quantum mechanics and especially quantum field theory in the form of functional integrals, path integrals follow for any partial differential equation which is first order in one (or more) of the variables.

Consider a Partial Differential Equation (PDE) of the form

$$\partial_t \phi\left(\vec{r}, t\right) = F\left(t, \vec{r}, \vec{\partial}\right)\phi\left(\vec{r}, t\right)$$

Here $\vec{r}$ could represent position in any number of dimensions and t could represent time, but neither of these representations is necessary. For example, in the paraxial wave equation "t " is position along the optic axis which is generally indicated with the variable z. But, even though the variable t does not have to have anything to do with time, for convenience, I will refer to it as time. Also, $F\left(t, \vec{r}, \vec{\partial}\right)$ can be any combination of functions of $t, \vec{r}$ and derivatives w.r.t. $\vec{r}$, i.e., $\vec{\partial}$. A specific example is discussed below.

We would like simply to write the solution as

$$\phi\left(\vec{r}, t\right) = \exp\left[\int_0^t dt F\left(t, \vec{r}, \vec{\partial}\right)\right]\phi\left(\vec{r}, 0\right)$$

but this is WRONG because F depends on both $\vec{r}$ and $\vec{\partial}$ and hence the Taylor expansion of $\exp\left[\int_0^t dt F\left(t, \vec{r}, \vec{\partial}\right)\right]$ will generate terms where $\vec{\partial}$ is acting on the $\vec{r}$ dependence of F and such derivative terms are not present in the original PDE.

The path integral approach gets around this, but at the expense of turning ordinary integrals into path integrals.

It follows from the PDE that over a short time (remember t can be any variable it does not have to be "time") step Δt we have

$$\phi\left(\vec{r}, t + \Delta t\right) \simeq \phi\left(\vec{r}, t\right) + \Delta t F\left(t, \vec{r}, \vec{\partial}\right)\phi\left(\vec{r}, t\right)$$

To first order in Δt this is equivalent to

$$\phi\left(\vec{r}, t + \Delta t\right) \simeq \exp\left[\Delta t F\left(t, \vec{r}, \vec{\partial}\right)\right]\phi\left(\vec{r}, t\right)$$

Iterating this N times we have

$$\phi\left(\vec{r},t+\left(N+1\right)\Delta t\right)\simeq\exp\left[\Delta tF\left(t+N\Delta t,\vec{r},\vec{\partial}\right)\right]\exp\left[\Delta tF\left(t+\left(N-1\right)\Delta t,\vec{r},\vec{\partial}\right)\right]\cdots$$
$$\cdots\exp\left[\Delta tF\left(t+2\Delta t,\vec{r},\vec{\partial}\right)\right]\exp\left[\Delta tF\left(t+\Delta t,\vec{r},\vec{\partial}\right)\right]\phi\left(\vec{r},t\right)$$

Setting $t'=\left(N+1\right)\Delta t$ and taking the limit as $\Delta t\rightarrow 0$ and $N\rightarrow\infty$ in such a way that t' stays fixed, the above product of integrals becomes a path integral.

The best way to see this is to consider a specific form for F. So, consider an F with the specific form

$$F\left(t,\vec{r},\vec{\partial}\right)=c\vec{\partial}^{2}+V\left(\vec{r},t\right)$$

where c is an arbitrary complex valued constant and $V\left(\vec{r}\right)$ can be complex valued as well. We now have

$$\phi\left(\vec{r},t+\Delta t\right)\simeq\phi\left(\vec{r},t\right)+c\Delta t\vec{\partial}^{2}\phi\left(\vec{r},t\right)+\Delta tV\left(\vec{r},t\right)\phi\left(\vec{r},t\right)$$

To first order in Δt this is the same as

$$\phi\left(\vec{r},t+\Delta t\right)\simeq\exp\left[c\Delta t\vec{\partial}^{2}\right]\exp\left[\Delta tV\left(\vec{r},t\right)\right]\phi\left(\vec{r},t\right)$$

We have written the above as a product of two exponentials for convenience in executing the next steps in the derivation. But note two things. First, if we had $V=0$ then $\phi\left(\vec{r},t+\Delta t\right)=\exp\left[c\Delta t\vec{\partial}^{2}\right]\phi\left(\vec{r},t\right)$ is exact for any value of Δt, large or small, since c is a constant independent of t. Second, $\exp\left[A+B\right]=\exp\left[A\right]\exp\left[B\right]$ if and only if, A and B commute, i.e., if and only if $AB=BA$. This can be proved by Taylor expanding $\exp\left[A+B\right]$ and noting that you need to be able to replace BA with AB everywhere in order to produce a series which sums to $\exp\left[A\right]\exp\left[B\right]$. Here we have $A=c\vec{\partial}^{2}$ and $B=V\left(\vec{r},t\right)$ but $\vec{\partial}^{2}\left(V\left(\vec{r},t\right)\phi\left(\vec{r},t\right)\right)=\left(\vec{\partial}^{2}V\left(\vec{r},t\right)\right)\phi\left(\vec{r},t\right)+2\vec{\partial}V\left(\vec{r},t\right)\cdot\vec{\partial}\phi\left(\vec{r},t\right)+V\left(\vec{r},t\right)\vec{\partial}^{2}\phi\left(\vec{r},t\right)$, hence $\vec{\partial}^{2}V\left(\vec{r},t\right)\phi\left(\vec{r},t\right)-V\left(\vec{r},t\right)\vec{\partial}^{2}\phi\left(\vec{r},t\right)\neq 0$ and so $c\vec{\partial}^{2}$ and $V\left(\vec{r},t\right)$ do not commute and $\exp\left[c\Delta t\vec{\partial}^{2}+\Delta tV\left(\vec{r},t\right)\right]\neq\exp\left[c\Delta t\vec{\partial}^{2}\right]\exp\left[\Delta tV\left(\vec{r},t\right)\right]$ to all orders in Δt. It is only true in the limit as $\Delta t\rightarrow 0$.

Noting that

$$\exp\left[V\left(\vec{r},t\right)\right]\phi\left(\vec{r},t\right)=\int d^{D}r'\delta_{D}\left(\vec{r}-\vec{r}'\right)\exp\left[V\left(\vec{r}',t\right)\right]\phi\left(\vec{r}',t\right)$$

where D is the number of dimensions of the space $\vec{r}$ lives in, we have

$$\phi\left(\vec{r},t+\Delta t\right)\simeq\int d^{D}r'\exp\left[c\Delta t\vec{\partial}^{2}\right]\delta_{D}\left(\vec{r}-\vec{r}'\right)\exp\left[\Delta tV\left(\vec{r}',t\right)\right]\phi\left(\vec{r}',t\right)$$

Using

$$\begin{aligned}\exp\left[c\Delta t\vec{\partial}^2\right]\delta_D\left(\vec{r}-\vec{r}'\right) &= \int\frac{d^Dp}{\left(2\pi\right)^D}\exp\left[c\Delta t\vec{\partial}^2\right]\exp\left[i\vec{p}\cdot\left(\vec{r}-\vec{r}'\right)\right]\\ &= \int\frac{d^Dp}{\left(2\pi\right)^D}\exp\left[-c\Delta t\vec{p}^2+i\vec{p}\cdot\left(\vec{r}-\vec{r}'\right)\right]\\ &= \left(\sqrt{\frac{1}{4\pi c\Delta t}}\right)^D\exp\left[-\frac{\left(\vec{r}-\vec{r}'\right)^2}{4c\Delta t}\right]\end{aligned}$$

we have

$$\begin{aligned}\phi\left(\vec{r},t+\Delta t\right) &\simeq \int d^Dr'\left(\sqrt{\frac{1}{4\pi c\Delta t}}\right)^D\exp\left[-\frac{\left(\vec{r}-\vec{r}'\right)^2}{4c\Delta t}\right]\exp\left[\Delta tV\left(\vec{r}',t\right)\right]\phi\left(\vec{r}',t\right)\\ &\simeq \int d^Dr'\left(\sqrt{\frac{1}{4\pi c\Delta t}}\right)^D\exp\left[-\frac{\left(\vec{r}-\vec{r}'\right)^2}{4c\Delta t}+\Delta tV\left(\vec{r}',t\right)\right]\phi\left(\vec{r}',t\right)\end{aligned}$$

where in the last step we have used the fact that $\left(\vec{r}-\vec{r}'\right)^2$ commutes with $V\left(\vec{r},t\right)$ and so we can put the two exponents together into one without error.

Letting $t\rightarrow t+\Delta t$ and replacing the dummy variable $\vec{r}'$ with $\vec{r}_1$ we have

$$\phi\left(\vec{r},t+2\Delta t\right)\simeq\int d^Dr_1\left(\sqrt{\frac{1}{4\pi c\Delta t}}\right)^D\exp\left[-\frac{\left(\vec{r}-\vec{r}_1\right)^2}{4c\Delta t}+\Delta tV\left(\vec{r}_1,t+\Delta t\right)\right]\phi\left(\vec{r}_1,t+\Delta t\right)$$

then using the above result for $\phi\left(\vec{r},t+\Delta t\right)$ with $\vec{r}=\vec{r}_1$ gives

$$\phi\left(\vec{r},t+2\Delta t\right)\simeq\int d^Dr_1d^Dr'\left(\sqrt{\frac{1}{4\pi c\Delta t}}\right)^{2D}\exp\left[\begin{array}{c}-\frac{\left(\vec{r}-\vec{r}_1\right)^2}{4c\Delta t}+\Delta tV\left(\vec{r}_1,t+\Delta t\right)\\ -\frac{\left(\vec{r}_1-\vec{r}'\right)^2}{4c\Delta t}+\Delta tV\left(\vec{r}',t\right)\end{array}\right]\phi\left(\vec{r}',t\right)$$

Repeating this process N times gives

$$\begin{aligned}\phi\left(\vec{r},t+\left(N+1\right)\Delta t\right) \simeq & \int d^Dr_N\cdots d^Dr_1d^Dr'\left(\sqrt{\frac{1}{4\pi c\Delta t}}\right)^{(N+1)D}\\ & \times\exp\left[\sum_{n=0}^{N}\left(-\frac{\left(\vec{r}_{n+1}-\vec{r}_n\right)^2}{4c\Delta t}+\Delta tV\left(\vec{r}_n,t+n\Delta t\right)\right)\right]\phi\left(\vec{r}',t\right)\end{aligned}$$

where $\vec{r}_{N+1}=\vec{r}$ and $\vec{r}_0=\vec{r}'$.

We can write $\vec{r}_n=\vec{r}\left(t+n\Delta t\right)$ where $\vec{r}\left(t\right)$ is any particular path or trajectory, as a function of t, through the D dimensional space, then in the limit as $\Delta t\rightarrow 0$ while $N\rightarrow\infty$ so that $\left(N+1\right)\Delta t=t'$ stays fixed we can write

$$\begin{aligned}\sum_{n=0}^{N}\left(-\frac{\left(\vec{r}_{n+1}-\vec{r}_n\right)^2}{4c\Delta t}+\Delta tV\left(\vec{r}_2,t+n\Delta t\right)\right) &= \sum_{n=1}^{N}\Delta t\left(-\frac{1}{4c}\left(\frac{\vec{r}_{n+1}-\vec{r}_n}{\Delta t}\right)^2+V\left(\vec{r}_n,t+n\Delta t\right)\right)\\ &= \int_0^{t'}dt\left(-\frac{1}{4c}\vec{v}\left(t\right)^2+V\left(\vec{r}\left(t\right),t\right)\right)\end{aligned}$$

with

$$\vec{v}(t) = \lim_{\Delta t \to 0} \frac{(\vec{r}(t+\Delta t) - \vec{r}(t))}{\Delta t}$$

where here t is any "time" value between t and $t + t'$.

Again using $\vec{r}_n = \vec{r}(t + n\Delta t)$ we see that the integral $\int d^D r_n$ is integrating over all possible positions of the path $\vec{r}(t)$ at "time" $t + n\Delta t$ and hence the string of integrals

$$\int d^D r_N \cdots d^D r_2 d^D r_1$$

is, in a piecewise linear sense, integrating over all possible paths starting at $\vec{r}'$ and ending at $\vec{r}$. Thus in the limit as $\Delta t \to 0$, $N \to \infty$ with $t' = (N+1)\Delta t$ fixed, this is equivalent to integrating over all possible paths from starting at $\vec{r}(t) = \vec{r}'$ and ending at $\vec{r}(t + t') = \vec{r}$. The factors

$$\left(\sqrt{\frac{1}{4\pi c \Delta t}}\right)^{(N+1)D}$$

are normalization factors.

Thus we can consider

$$\lim_{\substack{\Delta t \to 0 \\ N \to \infty \\ (N+1)\Delta t = t'}} \int d^D r_N \cdots d^D r_2 d^D r_1 \left(\sqrt{\frac{1}{4\pi c \Delta t}}\right)^{(N+1)D} \to \int_{\vec{r}(t)=\vec{r}'}^{\vec{r}(t+t')=\vec{r}} \delta\vec{r}(t)$$

as the (normalized) integral over all possible paths from $\vec{r}(t) = \vec{r}'$ to $\vec{r}(t + t') = \vec{r}$ and define the "propagator" as the path integral

$$K(\vec{r}, \vec{r}', t, t + t') = \int_{\vec{r}(t)=\vec{r}'}^{\vec{r}(t+t')=\vec{r}} \delta\vec{r}(t) \exp\left[\int_0^{t'} dt \left(-\frac{1}{4c}\vec{v}(t)^2 + V(\vec{r}(t), t)\right)\right]$$

The above results then reduces to

$$\phi(\vec{r}, t + t') = \int d^D r' K(\vec{r}, \vec{r}', t, t + t') \phi(r', t)$$

Without any loss in generality we can set $t = 0$ to get

$$\phi(\vec{r}, t') = \int d^D r' K(\vec{r}, \vec{r}', t') \phi(r', 0)$$

COMMENTS: There are only a few cases where the path integral can be evaluated exactly. Hence many approximate but powerful techniques have been developed for getting useful results in cases where the path integral cannot be evaluated exactly. The two most common are a perturbation approach in V and the second is a stationary phase evaluation of the path integral.

15.2 Stationary Phase and Classical Mechanics

For nonrelativistic quantum mechanics we have the Schrodinger equation

$$i\hbar\partial_t\psi\left(\vec{r},t\right)=\left(-\frac{\hbar^2}{2m}\vec{\partial}^2+V\left(\vec{r},t\right)\right)\psi\left(\vec{r},t\right)$$

where the propagator $K\left(\vec{r},\vec{r}',t,t'\right)$ satisfies

$$\begin{aligned} i\hbar\partial_t K\left(\vec{r},\vec{r}',t'\right) &= \left(-\frac{\hbar^2}{2m}\vec{\partial}^2+V\left(\vec{r},t\right)\right)K\left(\vec{r},\vec{r}',t'\right) \\ K\left(\vec{r},\vec{r}',t\rightarrow t',t'\right) &= \delta_D\left(\vec{r}-\vec{r}'\right) \end{aligned}$$

Comparing this to the path integral solution above we have $c\rightarrow i\hbar/2m$ and $V\rightarrow-\left(i/\hbar\right)V$ and so $K\left(\vec{r},\vec{r}',t\right)$ is given by

$$K\left(\vec{r},\vec{r}',t\right)=\int_{\vec{r}(0)=\vec{r}'}^{\vec{r},(t)=\vec{r}}\delta\vec{r}(t)\exp\left[\frac{i}{\hbar}\int_0^{t'}dt\left(\frac{m}{2}\vec{v}\left(t\right)^2-V\left(\vec{r}\left(t\right),t\right)\right)\right]$$

with $\vec{v}\left(t\right)=\partial_t\vec{r}\left(t\right)$.

Note that in classical mechanics the Langragian $\mathcal{L}$ is defined as Kinetic Energy minus Potential Energy (see above section) and so we have

$$\mathcal{L}\left(\vec{r}\left(t\right),\partial_t\vec{r}\left(t\right)\right)=\frac{m}{2}\vec{v}\left(t\right)^2-V\left(\vec{r}\left(t\right),t\right)$$

The "action" S is defined as the integral of the Lagrangian over time or

$$S=\int dt\mathcal{L}\left(\vec{r}\left(t\right),\partial_t\vec{r}\left(t\right)\right)$$

Note that S is a functional of $\vec{r}\left(t\right)$, that is, for any given trajectory or path, S evaluates to a real number, that is S is a map from trajectories to the reals. In 1D it is a map from the function $x\left(t\right)$ to the reals, in 3D it is a map from 3 functions $\vec{r}\left(t\right)=\left(x\left(t\right),y\left(t\right),z\left(t\right)\right)$ to the reals. Anything that maps functions to numbers is called a functional.

Taking the functional derivative (defined in another section) of S w.r.t. $\vec{r}\left(t\right)$ and setting it to zero yields the Euler-Lagrange equations, i.e.,

$$\left.\frac{\delta S}{\delta r_n\left(t\right)}\right|_{\vec{r}(t)=\vec{r}_{cl}(t)}=0\;\rightarrow\;\partial_t\frac{\partial\mathcal{L}}{\partial\left(\partial_t r_n\right)}-\left.\frac{\partial\mathcal{L}}{\partial r_n}\right|_{\vec{r}(t)=\vec{r}_{cl}(t)}=0$$

where $\vec{r}_{cl}\left(t\right)$ is the classical path or trajectory. Here n labels the x,y,z components of $\vec{r}$ and the assumption is the boundary conditions $\vec{r}\left(0\right)=\vec{r}'$ and $\vec{r}\left(t\right)=\vec{r}$ are held fixed when varying $\vec{r}\left(t\right)$. The Euler-Lagrange equation is the equivalent of Newtons $\vec{F}=m\vec{a}$. In other words, classical mechanics corresponds to finding the paths or trajectories which extremize the action, i.e., for which $\delta S/\delta\vec{r}\left(t\right)=0$. Such paths or trajectories are the "stationary points" of the action.

Using the definition of S we have

$$K\left(\vec{r},\vec{r}',t\right)=\int_{\vec{r}(0)=\vec{r}'}^{\vec{r},(t)=\vec{r}}\delta\vec{r}\left(t\right)\exp\left[\frac{i}{\hbar}S\left(\vec{r}\left(t\right),\partial_t\vec{r}\left(t\right)\right)\right]$$

To put this is words, the propagator for the Schroding equation, or quantum mechanics, is the sum over all paths from $\vec{r}\left(0\right)=\vec{r}'$ to $\vec{r}\left(t\right)=\vec{r}$ of the phase factors $\exp\left[\frac{i}{\hbar}S\left(\vec{r}\left(t\right),\partial_t\vec{r}\left(t\right)\right)\right]$. As just discussed above, classical mechanics corresponds to those paths for which S is "stationary" $\delta S/\delta\vec{r}\left(t\right)=0$, hence quantum mechanics corresponds to adding the phase fluctuations from nonclassical paths to the one (or multple ones) from the classical path(s).

The stationary phase approximation as discussed below amounts to expanding the phase to quadratic order about the stationary points of the phase. Here that amounts to letting $\vec{r}\left(t\right)=\vec{r}_{cl}\left(t\right)+\Delta r\left(t\right)$ with $\vec{r}\left(0\right)=\vec{r}'$ and $\vec{r}\left(t\right)=\vec{r}$ so that $\Delta\vec{r}\left(0\right)=\Delta\vec{r}\left(t\right)=0$ and expanding $S\left(\vec{r}\left(t\right),\partial_t\vec{r}\left(t\right)\right)$ to second order in $\delta\vec{r}\left(t\right)$. Noting that $\hbar\simeq10^{-34}$Js (Joule seconds) which is extremely small in everyday terms and so we expect that quantum fluctuations about the classical trajectory will be small, at least in certain relevant cases. Expanding S to second order in $\Delta\vec{r}\left(t\right)$ gives

$$\begin{aligned}S\left(\vec{r}\left(t\right),\partial_t\vec{r}\left(t\right)\right) &= S\left(\vec{r}_{cl}\left(t\right),\partial_t\vec{r}_{cl}\left(t\right)\right)+\int dt\left.\frac{\delta S}{\delta r_n\left(t\right)}\right|_{cl}\Delta r_n\left(t\right)\\ &\quad+\frac{1}{2}\int dtdt'\left.\frac{\delta^2S}{\delta r_n\left(t\right)\delta r_m\left(t'\right)}\right|_{cl}\Delta r_n\left(t\right)\Delta r_m\left(t'\right)+\cdots\end{aligned}$$

The subscript cl indicates setting $\vec{r}\left(t\right)=\vec{r}_{cl}\left(t\right)$ after taking the functional derivatives.

Given the Euler-Lagrange equations this reduces to

$$S\left(\vec{r}\left(t\right),\partial_t\vec{r}\left(t\right)\right)=S_{cl}+\frac{1}{2}\int dtdt'\left.\frac{\delta^2S}{\delta r_n\left(t\right)\delta r_m\left(t'\right)}\right|_{\vec{r}(t)=\vec{r}_{cl}(t)}\Delta r_n\left(t\right)\Delta r_m\left(t'\right)+\cdots$$

where $S_{cl}\equiv S\left(\vec{r}_{cl}\left(t\right),\partial_t\vec{r}_{cl}\left(t\right)\right)$. The path integral for the propagator to second order in $\Delta\vec{r}\left(t\right)$ then becomes

$$K\left(\vec{r},\vec{r}',t\right)=e^{iS_{cl}/\hbar}\int_{\Delta\vec{r}(0)=0}^{\Delta\vec{r}(t)=0}\delta\left(\Delta r\left(t\right)\right)\exp\left[\frac{i}{2\hbar}\int dtdt'\left.\frac{\delta^2S}{\delta r_n\left(t\right)\delta r_m\left(t'\right)}\right|_{cl}\Delta r_n\left(t\right)\Delta r_m\left(t'\right)\right]$$

Based on the integral of a Gaussian we expect that

$$\begin{aligned}&\int_{\Delta\vec{r}(0)=0}^{\Delta\vec{r}(t)=0}\delta\left(\Delta r\left(t\right)\right)\exp\left[\frac{i}{2\hbar}\int dtdt'\left.\frac{\delta^2S}{\delta r_n\left(t\right)\delta r_m\left(t'\right)}\right|_{cl}\Delta r_n\left(t\right)\Delta r_m\left(t'\right)\right]\\ \sim\;&\frac{1}{\sqrt{\det\left[\left.\frac{\delta^2S}{\delta r_n(t)\delta r_m(t')}\right|_{cl}\right]}}\end{aligned}$$

But $\left.\frac{\delta^2S}{\delta r_n(t)\delta r_m(t')}\right|_{cl}$ is a differential operator, not a matrix, so the question is how to evaluate the determinant of an operator. This is discussed in the next section.

15.3 Path Integral Determinant

As shown in the previous section a path integral is essentially the infinite dimensional limit of a finite dimensional "volume" integral.

Consider that in D dimensions with $\vec{y} = (y_1, y_2, \cdots, y_N)$ we have (Einstein Summation Convention)

$$\begin{aligned}\int_{-\infty}^{+\infty} d^D y \exp\left[-\sum_{n,m=1}^{D} y_n M_{n,m} y_m\right] &= \int_{-\infty}^{+\infty} d^D y \exp\left[-y_n M_{n,m} y_m\right] \\ &= \frac{\pi^{D/2}}{\sqrt{\det[M]}}\end{aligned}$$

Because we have $y_i M_{i,j} y_j$ only the symmetric part of the matrix M contributes and so, w.l.o.g., we can take M to be symmetric. For M symmetric and real, all its eigenvalues are real. Diagonalizing M gives

$$\begin{aligned}\int_{-\infty}^{+\infty} d^D y \exp\left[-y_n M_{n,m} y_m\right] &= \int_{-\infty}^{+\infty} d^D y' \exp\left[-\sum_{n=1}^{D} \lambda_n y_n'^2\right] \\ &= \int_{-\infty}^{+\infty} dy_1' e^{-\lambda_1 y_1'^2} \int_{-\infty}^{+\infty} dy_2' e^{-\lambda_2 y_2'^2} \cdots \int_{-\infty}^{+\infty} dy_D' e^{-\lambda_D y_D'^2} \\ &= \frac{\pi^{D/2}}{\sqrt{\lambda_1 \lambda_2 \cdots \lambda_D}} \\ &= \frac{\pi^{D/2}}{\sqrt{\det[M]}}\end{aligned}$$

where λ_n are the eigenvalues of M. Obviously this only works if all the $\lambda_n > 0$, hence it only works if M is a positive definite matrix.

The other way to see this is, if the square root of M can be found, i.e., a matrix Q such that $M = Q \cdot Q = Q^2$, then changing variables to $y_n' = Q_{n,m} y_m$ (Einstein Summation Convention) has the Jacobian $\left|\det\left[Q^{-1}\right]\right| = 1/\left|\det[Q]\right| = 1/\det[Q]$ where the last equality follows from the fact that M is positive definite. Hence we have

$$\begin{aligned}\int_{-\infty}^{+\infty} d^D y \exp\left[-y_n M_{n,m} y_m\right] &= \int_{-\infty}^{+\infty} \frac{d^D y'}{\det[Q]} \exp\left[-\sum_{n=1}^{D} y_n'^2\right] \\ &= \frac{\pi^{D/2}}{\det[Q]} \\ &= \frac{\pi^{D/2}}{\det\left[M^{1/2}\right]} \\ &= \frac{\pi^{D/2}}{\sqrt{\det[M]}}\end{aligned}$$

Taking the initial time as 0 and the final time as T and using notation $q(t)$ for position in 1D we have for $\mathcal{L} = m(\partial_t q(t))^2/2 - V(q(t))$ and $0 < t < T$

$$S = \int_0^T dt' \left(\frac{m}{2} (\partial_{t'} q(t'))^2 - V(q(t')) \right)$$

$$\begin{aligned}
\frac{\delta S}{\delta q(t)} &= \int_0^T dt' \left(m(\partial_{t'} q(t')) \left(\partial_{t'} \frac{\delta q(t')}{\delta q(t)} \right) - \frac{\partial V(q(t'))}{\partial (q(t'))} \frac{\delta q(t')}{\delta q(t)} \right) \\
&= \int_0^T dt' \left(m(\partial_{t'} q(t')) (\partial_{t'} \delta_D(t - t')) - \frac{\partial V(q(t'))}{\partial (q(t'))} \delta_D(t - t') \right) \\
&= \int_0^T dt' \left(-m \left(\partial_{t'}^2 q(t') \right) - \frac{\partial V(q(t'))}{\partial (q(t'))} \right) \delta_D(t - t') \\
&= -m \left(\partial_t^2 q(t) \right) - \frac{\partial V(q(t))}{\partial (q(t))} \\
&= -m \left(\partial_t^2 q(t) \right) - V'(q(t))
\end{aligned}$$

$$\begin{aligned}
\frac{\delta^2 S}{\delta q(t) \delta q(t')} &= -\left(m \partial_t^2 \frac{\delta q(t)}{\delta q(t')} + \frac{\partial^2 V(q(t))}{\partial (q(t)) \partial (q(t'))} \frac{\delta q(t)}{\delta q(t')} \right) \\
&= -\left(m \partial_t^2 + \frac{\partial^2 V(q(t))}{\partial (q(t)) \partial (q(t'))} \right) \delta_D(t - t') \\
&= -\left(m \partial_t^2 + V''(q(t)) \right) \delta_D(t - t')
\end{aligned}$$

with $V'(q) = \partial_q V(q)$ and $V''(q) = \partial_q^2 V(q)$.

Note that

$$\frac{\delta S}{\delta q(t)} = 0 \ \rightarrow \ m \partial_t q(t) = -V'(q(t)) = \text{Newton}$$

To simplify notation from the previous section set

$$\Delta r(t) = y(t)$$

Setting $q(t) = q_{cl}(t)$ to determine the lowest order quantum fluctuations about

the classical solution we want to evaluate

$$\begin{aligned}
\Delta &\equiv \int_{y(0)=0}^{y(T)=0} \delta y(t) \exp\left[\frac{i}{\hbar}\int_0^T dt \frac{1}{2}\left.\frac{\delta^2 S}{\delta q(t)\,\delta q(t')}\right|_{cl} y(t)\, y(t')\right] \\
&= \int_{y(0)=0}^{y(T)=0} \delta y(t) \exp\left[-\frac{i}{\hbar}\int_0^T dt \frac{1}{2}\left[\left(\partial_t^2 + V''(t)\right)\delta_D(t-t')\right] y(t)\, y(t')\right] \\
&= \int_{y(0)=0}^{y(T)=0} \delta y(t) \exp\left[-\frac{i}{2\hbar}\int_0^T dt y(t)\left(\partial_t^2 + V''(t)\right) y(t)\right] \\
&\equiv \frac{1}{\sqrt{\det\left(\frac{i}{\pi\hbar}\left(\partial_t^2 + V''(t)\right)\delta_D(t-t')\right)}}
\end{aligned}$$

For nonperiodic classical trajectories: $q_{cl}(0) \neq q_{cl}(T)$ there is generally a single solution for $q_{cl}(t)$. For periodic classical trajectories $q_{cl}(0) = q_{cl}(T) = q_0$, a specified value. In this case there will be multiple solutions (with different energy) with the basic period of the path given by T/n for integer n. This case requires accounting for the multiple trajectories. We will only consider the single trajectory case.

We can solve the integral $\int_0^T dt y(t)\left(\partial_t^2 + V''(t)\right) y(t)$ by diagonalizing the operator

$$\partial_t{}^2 + V''(t)$$

subject to the boundary conditions $y(0) = y(T) = 0$.

To diagonalize it make the following change of variables (I have no idea how you see ahead of time that this approach will work, but it does), let

$$\partial_t y(t) = \partial_t z(t) + K(t)\, y(t)$$

and choose K so that

$$\dot{y}^2 - V''(t)\, y^2 = \dot{z}^2 \pm \partial_t\left(\text{something} \times y^2\right)$$

(either sign works)

Here we are using the notation:

$$\dot{y} \equiv \partial_t y(t),$$

That is, a "dot" over any function indicates the derivative of that function w.r.t. time.

Since the second term on the right hand side of $\dot{y}^2 - V''(t)\, y^2$ above is an exact derivative which vanishes at the endoints (assuming the "something" is finite there) this reduces the path integral to that of a free particle, i.e.,

This change of variables gives

$$\begin{aligned}\dot{y}^2 - V''y^2 &= \dot{z}^2 + 2Ky\dot{z} + K^2y^2 - V''y^2 \\ &= \dot{z}^2 + 2Ky\left(\dot{y} - Ky\right) + K^2y^2 - V''y^2 \\ &= \dot{z}^2 + 2\dot{y}yK - \left(K^2 + V''\right)y^2 \\ &= \dot{z}^2 + \left(\partial_t\left(y^2\right)\right)K - \left(K^2 + V''\right)y^2\end{aligned}$$

Thus, if we choose K so that

$$\partial_t K = -\left(K^2 + V''\right)$$

then we get the form we want.

Let

$$K = \frac{\partial_t f}{f}$$

then

$$\partial_t K = \frac{{\partial_t}^2 f}{f} - \left(\frac{\partial_t f}{f}\right)^2 = \frac{{\partial_t}^2 f}{f} - K^2$$

Thus, letting $K = \partial_t f/f$ automatically generates the $-K^2$ term and so we will get the $-V''$ if we pick f so that

$$\frac{{\partial_t}^2 f}{f} = -V''$$

or equivalently

$$\left({\partial_t}^2 + V''\right)f = 0$$

Hence f is the 0 eigenvalue solution to the operator we want to diagonalize. Note that we want f to be nonvanishing everywhere in the interval 0 to T otherwise K will nominally blow up.

To find $y(t)$ as a function of $z(t)$ integrate $\dot{y} = \dot{z} + Ky = \dot{z} + \left(\dot{f}/f\right)y$ to get

$$y(t) = z(t) + \int_0^t dt' \frac{\dot{f}(t')}{f(t')} y(t')$$

More properly this should be written as

$$z(t) = y(t) - \int_0^t dt' \frac{\dot{f}(t')}{f(t')} y(t')$$

which expresses z (on the left) in terms of y (on the right). Note that $y(0) = 0 \Rightarrow z(0) = 0$ (as long as $\dot{f}/f$ does not contain a singularity at $t = 0$). The value of $z(T)$ is not obvious. How this is handled is shown below.

The inverse of this relation can be found by solving

$$\begin{aligned}\dot{z} &= \dot{y} - Ky \\ &= \dot{y} - \frac{\dot{f}}{f}\, y\end{aligned}$$

Use an integrating factor. Let $y(t) = \exp\left[\int_0^t dt' \dot{f}(t')/f(t')\right]\gamma(t) = f(t)/f(0)\,\gamma(t)$ and we have

$$\begin{aligned}
\dot{z} &= \dot{y} - \frac{\dot{f}}{f}\, y \\
&= \frac{\dot{f}}{f} y + \exp\left[\int_0^t dt' \dot{f}(t')/f(t')\right]\dot{\gamma} - \frac{\dot{f}}{f}\, y \\
&= \exp\left[\int_0^t dt' \dot{f}(t')/f(t')\right]\dot{\gamma} \\
&= \frac{f(t)}{f(0)}\dot{\gamma}
\end{aligned}$$

Solving for $\gamma(t)$ and multiplying it by $f(t)/f(0)$ to get $y(t)$ gives

$$y(t) = \frac{f(t)}{f(0)}\gamma(t) = \frac{f(t)}{f(0)} f(0) \int_0^t dt' \frac{\dot{z}(t')}{f(t')} = f(t) \int_0^t dt' \frac{\dot{z}(t')}{f(t')}$$

If we integrate by parts and use $z(0) = 0$ this can be rewritten as

$$\begin{aligned}
y(t) &= z(t) + f(t) \int_0^t dt' \frac{\dot{f}(t')}{f(t')^2} z(t') \\
&= z(t) + f(t) \int_0^T dt' \theta(t - t') \frac{\dot{f}(t')}{f(t')^2} z(t')
\end{aligned}$$

which is how it appears in the 1974 PRD paper by Dashen, Hasslacher and Neveu. The second form above for $y(t)$ has all the t dependence inside the integral. This is useful for evaluating the Jacobian.of the change of integration variables in the path integral as shown below.

As metioned above, with the change of variables the action becomes

$$\int_0^T dt \left(\frac{1}{2}\dot{y}^2 - V''(t)\, y^2\right) = \int_0^T dt \frac{1}{2}\dot{z}^2$$

The condition that $y(T) = 0$ will be implemented in the path integral by including a delta function. The same could be done with the condition $y(0) = 0$ but in the discrete time formalism we will find it more straightforward to set the value of $y(0) = 0$ and not integrate over it. Note that setting $y(0) = 0$ also sets $z(0) = 0$ as discussed above.

The delta function on $y(T)$ can implemented by multplying the path integral with a judiciously chosen form for "1"

$$\begin{aligned}
1 &= \int d(y(T))\, \delta(y(T)) \\
&= \int \frac{d\alpha}{2\pi} \int d(y(T)) \exp\left[i\alpha y(T)\right] \\
&= \int \frac{d\alpha}{2\pi} \int d(y(T)) \exp\left[i\alpha f(T) \int_0^T dt \frac{\dot{z}(t)}{f(t)}\right]
\end{aligned}$$

In the last line we have substituted the solution for $y(t)$ in terms of $z(t)$. This is useful since it now looks like a source term in the path integral, only coupled to $\dot{z}$ rather than to z.

Putting all the change of integration variable parts together we get

$$\begin{aligned}\Delta &= \int_{y(0)=0}^{y(T)=0} \delta y(t) \exp\left[\frac{i}{\hbar}\int_0^T dt \frac{1}{2}\left(\dot{y}^2 - V''(t)\, y^2\right)\right] \\ &= \int \frac{d\alpha}{2\pi}\int d(y(T)) \exp\left[i\alpha y(T)\right] \int_{z(0)=0}^{y(T)} \delta y(t) \exp\left[\frac{i}{\hbar}\int_0^T dt \frac{1}{2}\left(\dot{y}^2 - V''(t)\, y^2\right)\right] \\ &= \int \frac{d\alpha}{2\pi}\int_{z(0)=0} \delta z(t) \det\left[\frac{\delta y(t)}{\delta z(t')}\right] \exp\left[\frac{i}{\hbar}\int_0^T dt \left(\dot{z}(t)^2 + \hbar\alpha \frac{f(T)}{f(t)}\dot{z}(t)\right)\right]\end{aligned}$$

In the last line we have made the replacement

$$\int d(z(T)) \int_{z(0)=0}^{z(T)} \delta z(t) = \int_{z(0)=0} \delta z(t)$$

That is integrate over all paths $z(t)$ with $z(0) = 0$ and include $z(T)$ in the integration as well.when doing the path integral. The factor $\det\left[\delta y(t)/\delta z(t')\right]$ is the Jacobian for the change of integration variables. Note that it should be evaluated to include the $d(y(T))$ to $d(z(T))$ transformation as well as the regular $\delta y(t)$ to $\delta z(t)$ transformation. The Jacobian will be evaluated below. For now just note that since the change of variables from $y(t)$ to $z(t)$ is linear this means the Jacobian is independent of $z(t)$ and so can be pulled outside the path integral.

To evaluate the path integral discretize it by letting $n = 0, 1, ..., N$ with $t_n = n\varepsilon$ with $\varepsilon = T/\!\!/N$ and $z_n = z(t_n)$. Note that with this definition $t_0 = 0$ and $t_N = T$. The boundary condition $z(0) = 0$ is implemented by setting $z_0 = 0$. The boundary condition $y(T) = 0$ is being handled by the delta function. In discrete form we have

$$\Delta = \lim_{\substack{N->\infty \\ N\varepsilon = T}} \det\left[\frac{\delta y(t)}{\delta z(t')}\right] \int \frac{d\alpha}{2\pi}\int \frac{dz_1}{\mu}\cdots\frac{dz_N}{\mu} \exp\left[\frac{i}{\hbar}\sum_{n=1}^{N}\left(\begin{array}{c}\frac{(z_n - z_{n-1})^2}{2\varepsilon} \\ +\hbar\alpha\frac{f(T)}{f_n}(z_n - z_{n-1})\end{array}\right)\right]$$

where $\mu = \sqrt{2\pi i\varepsilon\hbar}$ is the proper measure (normalization so that each gaussian integral when done on its own, uncoupled to the others, is unity). With the delta function handling the $y(T) = 0$ condition we have to integrate over dz_N as well but this makes it easy to change the discrete integration variables to

$$b_n = z_n - z_{n-1}$$

which gives

$$\begin{aligned}\Delta &= \lim_{\substack{N->\infty \\ N\varepsilon=T}} \det\left[\frac{\delta y(t)}{\delta z(t')}\right] \int \frac{d\alpha}{2\pi} \prod_{n=1}^{N} \left(\int \frac{db}{\mu} \exp\left[\frac{i}{\hbar}\frac{b^2}{2\varepsilon} + i\alpha\frac{f(T)}{f_n}b\right]\right) \\ &= \lim_{\substack{N->\infty \\ N\varepsilon=T}} \det\left[\frac{\delta y(t)}{\delta z(t')}\right] \int \frac{d\alpha}{2\pi} \left(\exp\left[-\frac{i\hbar\alpha^2 f(T)^2}{2} \sum_{n=1}^{N} \frac{\varepsilon}{{f_n}^2}\right]\right) \\ &= \det\left[\frac{\delta y(t)}{\delta z(t')}\right] \int \frac{d\alpha}{2\pi} \left(\exp\left[-\frac{i\hbar\alpha^2 f(T)^2}{2} \int_0^T \frac{dt}{f(t)^2}\right]\right) \\ &= \det\left[\frac{\delta y(t)}{\delta z(t')}\right] \frac{1}{\sqrt{2\pi i\hbar f(T)^2 \int_0^T \frac{dt}{f(t)^2}}}\end{aligned}$$

The Jacobian will be evaluated using the relation $\det(X) = \exp(tr(\ln(X)))$. Taking the functional derivative of $y(t)$ in terms of $z(t)$ gives

$$\begin{aligned}\frac{\delta y(t)}{\delta z(t')} &= \frac{\delta}{\delta z(t')}\left(z(t) + f(t)\int_0^t ds \frac{\dot{f}(s)}{f(s)^2} z(s)\right) \\ &= \frac{\delta}{\delta z(t')}\left(z(t) + f(t)\int_0^T ds\ \theta(t-s)\frac{\dot{f}(s)}{f(s)^2} z(s)\right) \\ &= \delta(t-t') + \frac{f(t)\dot{f}(t')}{f(t')^2}\theta(t-t')\end{aligned}$$

and so

$$\begin{aligned}\det\left[\frac{\delta y(t)}{\delta z(t')}\right] &= \det\left[\delta(t-t') + \frac{f(t)\dot{f}(t')}{f(t')^2}\theta(t-t')\right] \\ &= \exp\left[tr\left[\ln\left[\delta(t-t') + \frac{f(t)\dot{f}(t')}{f(t')^2}\theta(t-t')\right]\right]\right]\end{aligned}$$

Expanding the ln we find that only the first term survives the trace (tr) due to the θ functions. The trace operation, tr, is defined as

$$tr(F(t,t') \equiv \int dtdt'\delta(t-t')F(t,t') = \int dtF(t,t)$$

The second order term in the expansion of ln contains

$$tr\left(\int dt''\theta(t-t')\theta(t''-t')\right) = \int dtdt''\theta(t-t')\theta(t''-t) = 0$$

This vanishes because both θ functions cannot be satisfied at the same time.

Hence

$$\begin{aligned}\det\left[\frac{\delta y(t)}{\delta z(t')}\right] &= \exp\left[tr\left[\frac{f(t)\dot{f}(t')}{f(t')^2}\theta(t-t')\right]\right] \\ &= \exp\left[\int_0^T dt\frac{\dot{f}(t)}{f(t)}\theta(0)\right] \\ &= \exp\left[\frac{1}{2}\int_0^T dt\frac{\dot{f}(t)}{f(t)}\right] \\ &= \left(\frac{f(T)}{f(0)}\right)^{1/2}\end{aligned}$$

Note that we have used $\theta(0) = 1/2$ which follows from combining

$$\theta(t) = \lim_{\varepsilon\to 0}\int\frac{dw}{2\pi i}\frac{e^{iwt}}{w-i\varepsilon}$$

with $1/(w-i\varepsilon) = PV(1/w) + i\pi\delta(w)$ where PV stands for the Cauchy Principal Value.

Next we have to solve for $f(t)$.Note first that by definition the classical solution $q_{cl}(t)$ satisfies

$$\left(\partial_t{}^2 + V'(q_{cl}(t))\right)q_{cl}(t) = 0$$

where, of course, $V'(q_{cl}(t)) \equiv (\partial_y V(y))_{y=q_{cl}(t)}$. Taking the time derivative of this equation yields

$$\left(\partial_t{}^2 + V''(q_{cl}(t))\right)\dot{q}_{cl}(t) = 0$$

where $V''(q_{cl}(t)) = \left(\partial_y{}^2 V(y)\right)_{y=q_{cl}(t)}$ and so one solution for $f(t)$ is $\dot{q}_{cl}(t)$.

We are solving a second order differential equation and so there must be two independent solutions. The second solution can be found using the Wronskian. Let two linearly indepedent solutions of a differential equation be $h_1(t)$ and $h_2(t)$, then the Wronskian is defined as

$$W(t) = h_1\partial_t h_2 - h_2\partial_t h_1$$

For a differential equation of the form we have here, i.e, $\left(\partial_t{}^2 + Q(t)\right)h(t) = 0$ we find that taking ∂_t of $W(t)$ yields zero, i.e., $\partial_t W(t) = 0$ and so $W(t) =$ constant (NOTE: $W =$ constant holds, in general terms, only for the particular form of differential equation considered here. In general W is not constant.). If we have found one nontrivial solution h_1, by whatever technique, then we can find another linearly independent solution h_2 by solving $W =$ constant $= h_1\partial_t h_2 - h_2\partial_t h_1$ for h_2 in term of h_1. Doing this yields

$$h_2(t) = h_1(t)\int^t\frac{dt'}{h_1(t')^2}$$

The lower limit on the integral is not specified since it will be determined by the boundary conditions. For the same reason the overall constant on the right hand side has been set to one. This is OK since the general solution is a linear combination of h_1 and h_2 with coefficients that are determined by the boundary conditions, i.e., the general solution is $h(t) = \alpha h_1(t) + \beta h_2(t)$. Apply this to the case here by setting the first solution we found $\dot{q}_{cl}(t)$ equal to $h_1(t)$, then compute $h_2(t)$ and finally take a linear combination we get

$$f(t) = \alpha \dot{q}_{cl}(t) + \beta \dot{q}_{cl}(t) \int^t \frac{dt'}{\dot{q}_{cl}(t)^2}$$

Here α and β are the constants of integration which are specified by the boundary conditions.

16 Propagators versus Greens Functions

Throughout this section we use the notation $\delta(\cdots)$ for the Dirac delta function.

16.1 Diffusion Type Equations

Consider the "free" diffusion equation

$$\left(\partial_t - D\ \vec{\partial}^2\right) \phi(\vec{r}, t) = 0$$

The general solution can be written as

$$\phi(\vec{r}, t) = \int d^3k\ a\left(\vec{k}\right) \exp\left[i\vec{k}\cdot\vec{r} - D\vec{k}^2 t\right]$$

$\phi(\vec{r}, t)$ can be related to $\phi(\vec{r}, 0)$, the initial field distribution, in the following way. Fourier Transform

$$\phi(\vec{r}, 0) = \int d^3k\ a\left(\vec{k}\right) \exp\left[i\vec{k}\cdot\vec{r}\right]$$

to get

$$a\left(\vec{k}\right) = \int \frac{d^3r}{(2\pi)^3}\ \phi(\vec{r}, 0) \exp\left[-i\vec{k}\cdot\vec{r}\right]$$

Substitute this into $\phi(\vec{r}, t)$ to get

$$\begin{aligned}
\phi(\vec{r}, t) &= \int d^3k \int \frac{d^3r'}{(2\pi)^3}\ \phi(\vec{r}', 0) \exp\left[i\vec{k}\cdot(\vec{r} - \vec{r}') - D\vec{k}^2 t\right] \\
&= \int d^3r' \left(\int \frac{d^3k}{(2\pi)^3}\ \exp\left[i\vec{k}\cdot(\vec{r} - \vec{r}') - D\vec{k}^2 t\right]\right) \phi(\vec{r}', 0) \\
&\equiv \int d^3r' K(\vec{r} - \vec{r}', t)\, \phi(\vec{r}', 0)
\end{aligned}$$

with

$$K\left(\vec{r}-\vec{r}',t\right)=\int\frac{d^3k}{\left(2\pi\right)^3}\ \exp\left[i\vec{k}\cdot\left(\vec{r}-\vec{r}'\right)-D\vec{k}^2t\right]$$
$$=\left(\frac{1}{4\pi Dt}\right)^{3/2}\exp\left[-\frac{\left(\vec{r}-\vec{r}'\right)^2}{4Dt}\right]$$

$K\left(\vec{r}-\vec{r}',t\right)$ is called "the propagator". Note that $K\left(\vec{r}-\vec{r}',t\right)$ has the initial condition

$$K\left(\vec{r}-\vec{r}',0\right)=\delta\left(\vec{r}-\vec{r}'\right)=\delta\left(x-x'\right)\delta\left(y-y'\right)\delta\left(z-z'\right)$$

which means that

$$\phi\left(\vec{r},0\right)=\int d^3r'K\left(\vec{r}-\vec{r}',0\right)\phi\left(\vec{r}',0\right)$$
$$=\int d^3r'\delta\left(\vec{r}-\vec{r}'\right)\phi\left(\vec{r}',0\right)$$
$$=\phi\left(\vec{r},0\right)$$

as one would expect. Thus the propagator actually does propagate the initial field $\phi\left(\vec{r},0\right)$ forward in time.

Now consider the Greens function for the same differential operator defined (up to boundary and initial conditions which are sorted out below) by

$$\left(\partial_t-D\ \vec{\partial}^2\right)G\left(\vec{r}-\vec{r}',t-t'\right)=\delta\left(\vec{r}-\vec{r}'\right)\delta\left(t-t'\right)$$

The fact that the differential operator $\left(\partial_t-D\ \vec{\partial}^2\right)$ is independent of time and space coordinates, i.e., independent of t and $\vec{r}$ implies that the Greens function is time and space translation invariant, i.e., in general one should write $G\left(\vec{r},\vec{r}',t,t'\right)$ but because of the time and space translation invariance it can be written as $G\left(\vec{r}-\vec{r}',t-t'\right)$. How this works is shown in the Appendix. Because of the translation invariance we can, without loss of generality, set $\vec{r}'$ and t' both equal to zero while deriving the general form for G. Nonzero values of $\vec{r}'$ and t' can be accounted for by simply replacing $\vec{r}$ with $\vec{r}-\vec{r}'$ and t with $t-t'$ in $G\left(\vec{r},t\right)$.

Using Fourier Transforms we can write

$$G\left(\vec{r},t\right)=\int\frac{d^3kd\omega}{\left(2\pi\right)^4}g\left(\vec{k},\omega\right)\exp\left[i\vec{k}\cdot\vec{r}-i\omega t\right]$$

Substituting into the left hand side of the differential equation for G gives

$$\left(\partial_t-D\ \vec{\partial}^2\right)G\left(\vec{r}',t\right)=\int\frac{d^3kd\omega}{\left(2\pi\right)^4}g\left(\vec{k},\omega\right)\left(\partial_t-D\vec{\partial}^2\right)\exp\left[i\vec{k}\cdot\vec{r}-i\omega t\right]$$
$$=\int\frac{d^3kd\omega}{\left(2\pi\right)^4}g\left(\vec{k},\omega\right)\left(-i\omega+D\vec{k}^2\right)\exp\left[i\vec{k}\cdot\vec{r}-i\omega t\right]$$

This must be equal to the right hand side

$$\int \frac{d^3kd\omega}{(2\pi)^4} g\left(\vec{k},\omega\right)\left(-i\omega + D\vec{k}^2\right)\exp\left[i\vec{k}\cdot\vec{r} - i\omega t\right] = \delta\left(\vec{r}-\vec{r}'\right)\delta\left(t-t'\right)$$

Noting that by definition

$$\int \frac{d^3kd\omega}{(2\pi)^4} \exp\left[i\vec{k}\cdot\vec{r} - i\omega t\right] = \delta\left(\vec{r}-\vec{r}'\right)\delta\left(t-t'\right)$$

we find that

$$g\left(\vec{k},\omega\right) = \frac{1}{-i\omega + D\vec{k}^2} = \frac{i}{\omega + iD\vec{k}^2}$$

Hence there is a pole at $\omega = -iD\vec{k}^2$, i.e., in the lower half ω-plane and we can use this pole to evaluate the ω integral in the Fourier Transform solution for $G\left(\vec{r},t\right)$. For $t > 0$ we can close the ω contour in the lower half plane and pick up the pole (and a factor of $-2\pi i$). For $t < 0$ we can close the ω contour in the upper half plane but since there is no pole there we get zero. Doing these steps yields

$$\begin{aligned}
G\left(\vec{r},t\right) &= i\int \frac{d^3k}{(2\pi)^4}\exp\left[i\vec{k}\cdot\vec{r}\right]\int d\omega \frac{1}{\omega + iD\vec{k}^2}\exp\left[-i\omega t\right] \\
&= i\theta\left(t\right)\left(-2\pi i\right)\int \frac{d^3k}{(2\pi)^4}\exp\left[i\vec{k}\cdot\vec{r} - i\left(-iD\vec{k}^2\right)t\right] \\
&= \theta\left(t\right)\int \frac{d^3k}{(2\pi)^3}\exp\left[i\vec{k}\cdot\vec{r} - D\vec{k}^2 t\right] \\
&= \theta\left(t\right)K\left(\vec{r},t\right)
\end{aligned}$$

The step function factor

$$\theta\left(t\right) = \begin{cases} 1 \text{ for } t > 0 \\ 0 \text{ for } t < 0 \end{cases}$$

is present because we get 0 for $t < 0$ and we get $\int \frac{d^3k}{(2\pi)^3}\exp\left[i\vec{k}\cdot\vec{r} - D\vec{k}^2 t\right] = K\left(\vec{r},t\right)$ for $t > 0$. The fact that the pole contributes only for $t > 0$ and hence the overall factor $\theta\left(t\right)$ means that G propagates "disturbances" forward in time. In other words you specify initial conditions and the propagator and Greens function both then account for the development of those conditions forward in time. Note that G satisfies the initial condition $G\left(\vec{r},t\to 0_+\right) = K\left(\vec{r},0\right) = \delta\left(\vec{r}\right)$.

Check the relation $G = \theta K$ directly.

$$\begin{aligned}\left(\partial_t - D\ \vec{\partial}^2\right) G\left(\vec{r}-\vec{r}',t-t'\right) &= \left(\partial_t - D\ \vec{\partial}^2\right)\theta\left(t-t'\right) K\left(\vec{r}-\vec{r}',t-t'\right) \\ &= \delta\left(t-t'\right) K\left(\vec{r}-\vec{r}',t-t'\right) \\ &+ \theta\left(t-t'\right) \underbrace{\left(\partial_t - D\ \vec{\partial}^2\right) K\left(\vec{r}-\vec{r}',t-t'\right)}_{=0} \\ &= \delta\left(t-t'\right) K\left(\vec{r}-\vec{r}',0\right) \\ &= \delta\left(t-t'\right)\delta\left(\vec{r}-\vec{r}'\right)\end{aligned}$$

Finally because G and K are constructed as Fourier Transforms they vanish at spatial infinity. Note that we can always add a homogeneous solution(s) ϕ to G to get whatever boundary conditions we want.

16.2 Schrodinger Type Equations

The same analysis goes through almost unchanged for the Schrodinger type equation

$$\left(i\partial_t + \frac{\hbar}{2m}\vec{\partial}^2\right)\psi\left(\vec{r},t\right) = 0$$

where m is the mass of the particle. The difference is now the pole in ω sits on the real ω axis and we must choose to circle it in the correct way in order to obtain propagation forward in time, i.e, the retarded propagator. The relation between G and K now contains a factor of $-i$,i.e., $G\left(\vec{r},t\right) = -i\theta\left(t\right) K\left(\vec{r},t\right)$. Explicitly, for the Schrodinger case we have

$$\psi\left(\vec{r},t\right) = \int d^3k\ a\left(\vec{k}\right)\exp\left[i\vec{k}\cdot\vec{r} - i\frac{\hbar\vec{k}^2}{2m}t\right]$$

and so

$$\psi\left(\vec{r},0\right) = \int d^3k\ a\left(\vec{k}\right)\exp\left[i\vec{k}\cdot\vec{r}\right]$$

Inverse Fourier Transforming gives

$$a\left(\vec{k}\right) = \int \frac{d^3r}{\left(2\pi\right)^3}\exp\left[-i\vec{k}\cdot\vec{r}\right]\psi\left(\vec{r},0\right)$$

and substituting back into $\psi\left(\vec{r},t\right)$ and rearranging we get

$$\begin{aligned}\psi\left(\vec{r},t\right) &= \int d^3r'\left(\int \frac{d^3k}{\left(2\pi\right)^3}\exp\left[i\vec{k}\cdot\left(\vec{r}-\vec{r}'\right) - i\frac{\hbar\vec{k}^2}{2m}t\right]\right)\psi\left(\vec{r}',0\right) \\ &\equiv \int d^3r' K\left(\vec{r}-\vec{r}',t\right)\psi\left(\vec{r}',0\right)\end{aligned}$$

with

$$\begin{aligned} K\left(\vec{r}-\vec{r}',t\right) &= \int \frac{d^3k}{\left(2\pi\right)^3}\exp\left[i\vec{k}\cdot\left(\vec{r}-\vec{r}'\right)-i\frac{\hbar\vec{k}^2}{2m}t\right] \\ &= \frac{1}{\left(2\pi\right)^3}\left(\frac{\pi}{\hbar it/2m}\right)^{3/2}\exp\left[-\frac{\left(\vec{r}-\vec{r}'\right)^2}{i4\hbar t/2m}\right] \\ &= \left(\frac{m}{2\pi i\hbar t}\right)^{3/2}\exp\left[i\frac{m}{2\hbar t}\left(\vec{r}-\vec{r}'\right)^2\right] \end{aligned}$$

16.3 Helmholtz Equation

The Helmoholtz equation is the wave equation for single frequency, i.e., monochromati^{2}c, waves. The wave equation is given by

$$\left(\frac{1}{c^2}\partial_t^2-\vec{\partial}\right)\Phi\left(\vec{r},t\right)=0$$

where c is the speed of propagation of the waves. We are working in 3D and so $\vec{\partial}^2=\partial_x^2+\partial_y^2+\partial_z^2$.

Monochromatic For the monochromatic case substitute

$$\Phi\left(\vec{r},t\right)=\phi\left(\vec{r}\right)e^{-i\omega t}$$

where $\omega=2\pi f=ck=2\pi/\lambda$ with $\lambda=$ wavelength and f being the temporal frequency (in say Hertz). With this substitution the wave equation reduces to the Helmholtz equation

$$\left(\vec{\partial}^2+k^2\right)\phi\left(\vec{r}\right)=0$$

This is a homogeneous linear differential equation with constant coefficients and so the basic solution is an exponential

$$\exp\left[i\vec{p}\cdot\vec{r}\right]=\exp\left[i\left(p_xx+p_yy+p_zz\right)\right]$$

Substituting the basic solution into Helmholtz yields

$$\vec{p}^2=k^2$$

Letting $p_z=\gamma$ and changing notation to $\beta_x=p_x$ and $\beta_y=p_y$ the solution is simply

$$\gamma\left(\vec{\beta}\right)=\pm\sqrt{k^2-\vec{\beta}^2}$$

Propagating versus Evanescent For $k^2 > \vec{\beta}^2$, $\gamma\left(\vec{\beta}\right)$ is real but if $k^2 < \vec{\beta}^2$ then $\gamma\left(\vec{\beta}\right) = \pm i\sqrt{\vec{\beta}^2 - k^2} = \pm i\left|\gamma\left(\vec{\beta}\right)\right|$. In the real case $\exp\left[\pm i\gamma\left(\vec{\beta}\right)z\right]$ oscillates along z but for the imaginary case $\exp\left[\pm i\gamma\left(\vec{\beta}\right)z\right] = \exp\left[\mp\left|\gamma\left(\vec{\beta}\right)\right|z\right]$ which exponentially decreases or increases with increasing z. The real case is termed "propgating" and the imaginary case is termed "evanescent". Both are important and are needed to get a full solution in many "real world" problems but there are also lots of "real world" problems where $\tilde{\phi}_{\pm}\left(\vec{\beta}\right)$ are nonzero only for $\vec{\beta}^2 < k^2$ in which case only the propagating terms are required.

Using superposition (the differential equation is linear in ϕ and so superposition holds) we can write any solution in the form

$$\phi\left(\vec{r}\right) = \int d^2\beta \left(\tilde{\phi}_+\left(\vec{\beta}\right)e^{i\vec{\beta}\cdot\vec{\rho}+i\gamma\left(\vec{\beta}\right)z} + \tilde{\phi}_-\left(\vec{\beta}\right)e^{i\vec{\beta}\cdot\vec{\rho}-i\gamma\left(\vec{\beta}\right)z}\right)$$

where $\vec{\rho} = (x, y)$. There are two coefficients $\tilde{\phi}_+$ and $\tilde{\phi}_-$ of course because the Helmholtz equation is second order.

With the choice $e^{-i\omega t}$ the full time dependent basic solution is $e^{i\vec{p}\cdot\vec{r}-i\omega t}$. The phase of this function is given by $\vec{p}\cdot\vec{r} - \omega t$ and surfaces of constant phase, termed wavefronts, are solutions to $\vec{p}\cdot\vec{r} - \omega t =$ constant. Separating $\vec{r}$ into the sum of components parallel and perpendicular to $\vec{p}$, i.e., $\vec{r} = \vec{r}_{\parallel} + \vec{r}_{\perp}$ gives $\vec{r}\cdot\vec{p} = \vec{r}_{\parallel}\cdot\vec{p}$ and $\vec{r}_{\perp}\cdot\vec{p} = 0$. Hence the phase increases for shifts in position in the direction of $\vec{r}_{\parallel}$ and is constant for changes in position in the plane perpendicular to $\vec{r}_{\parallel}$. In other words the surfaces of constant phase or the wavefronts are planes with their normal in direction of $\vec{p}$. ($\exp\left[i\vec{p}\cdot\vec{r} - \omega t\right]$ is called a "plane wave" for this reason.) To maintain a constant phase as t increases, the position of every wavefront must move in the $+\vec{r}_{\parallel}$ direction. Letting the unit vector in the direction of $\vec{p}$ be $\hat{p} = \vec{p}/\left|\vec{p}\right| = \vec{p}/k$ so that $\vec{r}_{\parallel}(t) = r_{\parallel}(t)\,\hat{p}$ with $r_{\parallel}(t)$ the position in the $\hat{p}$ direction of a given wavefront gives $\vec{r}_{\parallel}(t)\cdot\vec{p} - \omega t = r_{\parallel}(t)\,k - \omega t =$ constant. Taking the derivative of both sides w.r.t. t yields $\partial_t r_{\parallel}(t) = \omega/k = \lambda f =$ the speed of propagation of the wavefronts and hence of the plane wave itself.

16.3.1 Helmholtz propagator

Consider a case where the solution for $\phi\left(\vec{r}\right)$ is a superposition of plane waves all propagating generically in the $+z$ direction, i.e., $\tilde{\phi}_-\left(\vec{\beta}\right) = 0$, in which case

$$\phi\left(\vec{r}\right) = \phi\left(\vec{\rho}, z\right) = \int d^2\beta\tilde{\phi}_+\left(\vec{\beta}\right)e^{i\vec{\beta}\cdot\vec{\rho}+i\gamma\left(\vec{\beta}\right)z}$$

For $z = 0$ we have

$$\phi\left(\vec{\rho}, 0\right) = \int d^2\beta\tilde{\phi}_+\left(\vec{\beta}\right)e^{i\vec{\beta}\cdot\vec{\rho}}$$

Inverse Fourier transforming gives

$$\tilde{\phi}_+\left(\vec{\beta}\right) = \int \frac{d^2\rho}{(2\pi)^2} e^{-i\vec{\beta}\cdot\vec{\rho}} \phi\left(\vec{\rho}, 0\right)$$

Substituting above and rearranging the order of the integrals gives

$$\phi\left(\vec{\rho}, z\right) = \int d^2\rho' \int \frac{d^2\beta}{(2\pi)^2} e^{i\vec{\beta}\cdot\left(\vec{\rho}-\vec{\rho}'\right)+i\gamma\left(\vec{\beta}\right)z} \phi\left(\vec{\rho}', 0\right)$$

Hence the propagator is given by

$$K\left(\vec{\rho}-\vec{\rho}', z\right) = \int \frac{d^2\beta}{(2\pi)^2} e^{i\vec{\beta}\cdot\left(\vec{\rho}-\vec{\rho}'\right)+i\gamma\left(\vec{\beta}\right)z}$$

or

$$\phi\left(\vec{\rho}, z\right) = \int d^2\rho' K\left(\vec{\rho}-\vec{\rho}', z\right) \phi\left(\vec{\rho}', 0\right)$$

Note that

$$K\left(\vec{\rho}-\vec{\rho}', 0\right) = \delta\left(\vec{\rho}-\vec{\rho}'\right)$$

Setting $z = 0$ in K therefore gives back exactly the input field. This is an exact solution for all z, in empty space, given input field $\phi\left(\vec{\rho}, 0\right)$.

Fresnel Propagator Expanding

$$\gamma\left(\vec{\beta}\right) = k\sqrt{1-\frac{\vec{\beta}^2}{k^2}} \simeq k - \frac{\vec{\beta}^2}{2k}$$

which, after integration over $\vec{\beta}$, yields the Fresnel propagator.

$$K_{Fresnel}\left(\vec{\rho}-\vec{\rho}', z\right) = \frac{1}{i\lambda z} e^{ikz+ik\left(\vec{\rho}-\vec{\rho}'\right)^2/2z}$$

Obviously the Fresnel propagator is an approximation to the exact propagator.

16.3.2 Helmholtz Greens Functions

Plane Wave Representation The Greens function G must statisfy the inhomogeneous equation

$$\left(\vec{\partial}^2 + k^2\right) G\left(\vec{r}-\vec{r}'\right) = \delta\left(\vec{r}-\vec{r}'\right) \equiv \delta\left(x-x'\right)\delta\left(y-y'\right)\delta\left(z-z'\right)$$

Using the superposition principle we can write

$$G\left(\vec{r}-\vec{r}'\right) = \int \frac{d^3p}{(2\pi)^3} \tilde{G}\left(\vec{p}\right) e^{i\vec{p}\cdot\left(\vec{r}-\vec{r}'\right)}$$

and so

$$\left(\vec{\partial}^2 + k^2\right) G\left(\vec{r} - \vec{r}'\right) = \int \frac{d^3p}{\left(2\pi\right)^3} \tilde{G}\left(\vec{p}\right) \left(-\vec{p}^2 + k^2\right) e^{i\vec{p}\cdot\left(\vec{r}-\vec{r}'\right)}$$

Using the plane wave or Fourier representation of the delta function

$$\delta\left(\vec{r} - \vec{r}'\right) = \int \frac{d^3p}{\left(2\pi\right)^2} e^{i\vec{p}\cdot\left(\vec{r}-\vec{r}'\right)}$$

the solution for $\tilde{G}\left(\vec{p}\right)$ is given by

$$\tilde{G}\left(\vec{p}\right) = -\frac{1}{\vec{p}^2 - k^2}$$

Hence

$$G\left(\vec{r} - \vec{r}'\right) = -\int \frac{d^3p}{\left(2\pi\right)^3} \frac{e^{i\vec{p}\cdot\left(\vec{r}-\vec{r}'\right)}}{\vec{p}^2 - k^2}$$

Do the p_z integral first and use the $\vec{\beta},\ \gamma$ and $\vec{\rho}$ notation from above

$$\begin{aligned} G\left(\vec{r} - \vec{r}'\right) &= -\int \frac{d^2\beta}{\left(2\pi\right)^3} e^{i\vec{\beta}\cdot\left(\vec{\rho}-\vec{\rho}'\right)} \int dp_z \frac{e^{ip_z\cdot\left(z-z'\right)}}{p_z^2 - \left(k^2 - \vec{\beta}^2\right)} \\ &= -\int \frac{d^2\beta}{\left(2\pi\right)^3} e^{i\vec{\beta}\cdot\left(\vec{\rho}-\vec{\rho}'\right)} \int dp_z \frac{e^{ip_z\cdot\left(z-z'\right)}}{\left(p_z + \gamma\left(\vec{\beta}\right)\right)\left(p_z - \gamma\left(\vec{\beta}\right)\right)} \end{aligned}$$

For $z > z'$ we close the contour in the upper half plane and include the $p_z = \gamma\left(\vec{\beta}\right)$ pole. This gives

$$\begin{aligned} G\left(\vec{r} - \vec{r}'\right) &= G\left(\vec{\rho} - \vec{\rho}', z - z'\right) \\ &= -i \int \frac{d^2\beta}{\left(2\pi\right)^2} \frac{e^{i\vec{\beta}\cdot\left(\vec{\rho}-\vec{\rho}'\right) + i\gamma\left(\vec{\beta}\right)\left(z-z'\right)}}{2\gamma\left(\vec{\beta}\right)} \end{aligned}$$

Setting $z' = 0$ it is obvious that

$$2\partial_z G\left(\vec{\rho} - \vec{\rho}', z\right) = K\left(\vec{\rho} - \vec{\rho}', z\right)$$

Spherical Wave Representation Start with

$$G\left(\vec{r} - \vec{r}'\right) = -\int \frac{d^3p}{\left(2\pi\right)^3} \frac{e^{i\vec{p}\cdot\left(\vec{r}-\vec{r}'\right)}}{\vec{p}^2 - k^2}$$

but work in spherical polar coordinates (r, θ, φ). First set $\vec{r}' = 0$ since, to get it back, we merely have to replace it with $\vec{r} - \vec{r}'$ and second, w.l.o.g., we can take $\vec{r}$ to be along the z direction for the evaluation of the integral. Then we have

$$G\left(\vec{r}\right) = -\frac{1}{\left(2\pi\right)^3} \int_0^\infty p^2 dp \int_0^\pi \sin\left(\theta\right) d\theta \int_0^{2\pi} d\varphi \frac{e^{\iota pr\cos(\theta)}}{p^2 - k^2}$$

where $p = |\vec{p}|$ and $r = |\vec{r}|$. The $d\varphi$ integral yields 2π. Using $\sin(\theta)\, d\theta = -d(\cos(\theta))$ and replacing the dummy integration variable $\cos(\theta)$ with X we have

$$\begin{aligned} G(\vec{r}) &= -\frac{1}{(2\pi)^2}\int_0^\infty dp \frac{p^2}{p^2-k^2}\int_{-1}^1 dX e^{iprX} \\ &= -\frac{1}{(2\pi)^2}\int_0^\infty dp \frac{p^2}{p^2-k^2}\frac{1}{ipr}\left(e^{ipr}-e^{-ipr}\right) \\ &= -\frac{1}{(2\pi)^2 ir}\int_0^\infty dp \frac{p}{p^2-k^2}\left(e^{ipr}-e^{-ipr}\right) \end{aligned}$$

Let $p \to -p$ in the e^{-ipr} integration

$$\begin{aligned} \int_0^\infty dp \frac{p}{p^2-k^2}\left(-e^{-ipr}\right) &= \int_0^{-\infty} d(-p)\frac{-p}{p^2-k^2}\left(-e^{ipr}\right) \\ &= \int_{-\infty}^0 dp \frac{p}{p^2-k^2}e^{ipr} \end{aligned}$$

and so

$$\begin{aligned} G(\vec{r}) &= -\frac{1}{(2\pi)^2 ir}\int_{-\infty}^\infty dp \frac{p}{p^2-k^2}e^{ipr} \\ &= -\frac{1}{(2\pi)^2 ir}\int_{-\infty}^\infty dp \frac{p}{(p-k)(p+k)}e^{ipr} \end{aligned}$$

But $r \geq 0$ so we need to close the contour in upper half plane. Including the $p = k$ pole inside the contour yields

$$\begin{aligned} G(\vec{r}) &= -\frac{1}{(2\pi)^2 ir}2\pi i\frac{k}{2k}e^{ikr} \\ &= -\frac{1}{4\pi r}e^{ikr} \end{aligned}$$

which, accounting for the $e^{-i\omega t}$ time dependence is a outgoing spherical wave. The phase here is kr and so surfaces of constant phase correspond to spheres. Had we included the $p = -k$ pole inside the contour we would have an incoming spherical wave.

NOTE: Setting $k = 0$ gives the Greens function for the Laplacian, i.e,

$$G_{Laplacian}(\vec{r}) = -\frac{1}{4\pi r}$$

with

$$\vec{\partial}^2 G_{Laplacian}(\vec{r}) = \delta(\vec{r})$$

Finally since $r = |\vec{r}|$ replacing $\vec{r}$ with $\vec{r} - \vec{r}'$ gives

$$G(\vec{r}-\vec{r}') = -\frac{1}{4\pi|\vec{r}-\vec{r}'|}e^{ik|\vec{r}-\vec{r}'|}$$

Paraxial Wave Equation The paraxial wave equation, which has the same form as the Schrodinger equation is obtained from the Helmholtz equation by making the substitution

$$\phi\left(\vec{r}\right) = e^{ikz}\psi\left(\vec{\rho}, z\right)$$

where $\psi\left(\rho, z\right)$ is taken to vary very slowly with z compared to e^{ikz}, i.e.,

$$\frac{\left|\partial_z^2\psi\left(\vec{\rho}, z\right)\right|}{k\left|\partial_z\psi\left(\vec{\rho}, z\right)\right|} \ll 1$$

which is the "paraxial approximation".

Substituting $\phi\left(\vec{r}\right) = e^{ikz}\psi\left(\vec{\rho}, z\right)$ into the Helmholtz equation gives, using $\vec{\partial}_\perp^2 = \partial_x^2 + \partial_y^2$,

$$\left(\vec{\partial}_\perp^2 - k^2 + 2ik\partial_z + \partial_z^2\right)\psi\left(\vec{\rho}, z\right) + k^2\psi\left(\vec{\rho}, z\right) = 0$$

Cancelling the "k^2" terms and dropping the $\partial_z^2\psi$ term gives

$$\left(2ik\partial_z + \vec{\partial}_\perp^2\right)\psi\left(\vec{\rho}, z\right) = 0$$

which has the same mathematical form as the free (no potential energy) Schrodinger equation in 2D with z playing the role of time.

Had we included a position dependent index of refraction, $n\left(\vec{r}\right) = n\left(\vec{\rho}, z\right)$ which, simplifying things a bit, modifies the Helmholtz equation to be

$$\left(\vec{\partial}^2 + n\left(\vec{r}\right)^2 k^2\right)\phi\left(\vec{r}\right) = 0$$

and then made the paraxial approximation we would get

$$\left(2ik\partial_z + \vec{\partial}_\perp^2 + \left(n\left(\vec{\rho}, z\right)^2 - 1\right)k^2\right)\psi\left(\vec{\rho}, z\right) = 0$$

with $\left(n\left(\vec{\rho}, z\right)^2 - 1\right)k^2$ being the equivalent of the negative of the potential energy in the Schrodinger equation.

16.4 Wave Equation Greens Function

The wave equation is

$$\left(\frac{1}{c^2}\partial_t^2 - \vec{\partial}^2\right)\phi\left(\vec{r}, t\right) = 0$$

where as usual $\vec{r} = (x_1 = x,\ x_2 = y,\ x_3 = z)$ and c is the speed of propagation of the waves.

The wave equation is a homogeneous linear partial differential equation with constant coefficients and so, barring degeneracy, the basic form of the solution is an exponential which we will write as $\exp\left[i\vec{p}\cdot\vec{r} - i\omega t\right]$.

The Greens function must satisfy

$$\left(\frac{1}{c^2}\partial_t^2 - \vec{\partial}^2\right) G\left(\vec{r}-\vec{r}',t-t'\right) = \delta\left(t-t'\right)\delta\left(\vec{r}-\vec{r}'\right)$$

Combining the fact that

$$\left(\frac{1}{c^2}\partial_t^2 - \vec{\partial}^2\right) e^{i\vec{p}\cdot\vec{r}-i\omega t} = \left(-\frac{\omega^2}{c^2}+\vec{p}^2\right) e^{i\vec{p}\cdot\vec{r}-i\omega t}$$

with the fact that, from Fourier transforms,

$$\begin{aligned}
\delta\left(t-t'\right)\delta\left(\vec{r}-\vec{r}'\right) &= \delta\left(t-t'\right)\delta\left(x-x'\right)\delta\left(y-y'\right)\delta\left(z-z'\right) \\
&= \int_{-\infty}^{+\infty} \frac{d\omega d^3p}{\left(2\pi\right)^4} e^{i\vec{p}\cdot\left(\vec{r}-\vec{r}'\right)-i\omega\left(t-t'\right)}
\end{aligned}$$

we get the general form for G

$$\begin{aligned}
G\left(\vec{r}-\vec{r}',t-t'\right) &= \int_{-\infty}^{+\infty} \frac{d\omega d^3p}{\left(2\pi\right)^4} \frac{e^{i\vec{p}\cdot\left(\vec{r}-\vec{r}'\right)-i\omega\left(t-t'\right)}}{\left(\vec{p}^2-\frac{\omega^2}{c^2}\right)} \\
&= \int_{-\infty}^{+\infty} \frac{d\omega}{\left(2\pi\right)} e^{-i\omega\left(t-t'\right)} \int_{-\infty}^{+\infty} \frac{d^3p}{\left(2\pi\right)^3} \frac{e^{i\vec{p}\cdot\left(\vec{r}-\vec{r}'\right)}}{\left(\vec{p}^2-\frac{\omega^2}{c^2}\right)}
\end{aligned}$$

Looking at the spherical wave representation of the Greens function for the Helmhotz equation from above we have

$$G\left(\vec{r}-\vec{r}',t-t'\right) = \int_{-\infty}^{+\infty} \frac{d\omega}{\left(2\pi\right)} e^{-i\omega\left(t-t'\right)} \frac{4\pi}{\left|\vec{r}-\vec{r}'\right|} e^{ik\left|\vec{r}-\vec{r}'\right|}$$

with $k=\omega/c$. Changing the integration variable to k using $\omega = ck$ gives

$$\begin{aligned}
G\left(\vec{r}-\vec{r}',t-t'\right) &= \frac{4\pi c}{r}\int_{-\infty}^{+\infty} \frac{dk}{\left(2\pi\right)} e^{ik\left|\vec{r}-\vec{r}'\right|-ick\left(t-t'\right)} \\
&= \frac{4\pi c}{\left|\vec{r}-\vec{r}'\right|}\delta\left(\left|\vec{r}-\vec{r}'\right|-c\left(t-t'\right)\right)
\end{aligned}$$

where the last line follows from the Fourier transform representation of Dirac delta functions.

It's very interesting that the Greens function for the wave equation reduces to a Dirac delta function in precisely 3 space plus 1 time dimension. The units of a Dirac delta function are the reciprocal of the units of its argument, hence $\delta\left(t-t'\right)$ has units of 1/time and $\delta\left(\vec{r}-\vec{r}'\right) = \delta\left(x-x'\right)\delta\left(y-y'\right)\delta\left(z-z'\right)$ has units of 1/length^3. The units of $\left(\frac{1}{c^2}\partial_t^2 - \vec{\partial}^2\right)$ are 1/length^2 hence we have that the units of G must be 1/(length $\times$ time) and this is precisely the units of G above.

17 Scalar and Vector Diffraction

17.1 Maxwells Equations→Wave Equation

The macroscopic Maxwells equations (Jackson's nomenclature) are

$$\begin{aligned}\vec{\partial}\times\vec{E} &= -\partial_t\vec{B}\\ \partial\times\vec{H} &= \vec{J}+\partial_t\vec{D}\\ \partial\cdot\vec{D} &= \rho\\ \vec{\partial}\cdot\vec{B} &= 0\end{aligned}$$

where $\vec{D}=\varepsilon\vec{E}$ and $\vec{B}=\mu\vec{H}$ with ε and μ being material dependent and ρ and $\vec{J}$ being the "free" electric charge and electric current densities respectively.

NOTE: The above equations are valid in the "linear approximation" to the macroscopic Maxwells equations. Ideally, to account for properties of charges in a material, at least at the classical nonrelatvistic level, one would need to solve, simultaneously, Newtons equations for the motion of all the charged particles in the material, e.g., the motion of all the electrons and protons (or nuclei), and, the microscopic Maxwells equations with ρ and $\vec{J}$ being the total charge and current densities of all the particles. Obviously this is not possible at least in any practical sense. The "linear approximation" treats all the bound charges in the material, i.e., the electrons and nuclei, to be bound into atoms and molecules, and the solution to Newtons laws is approximated as a linear response of the material to the local electric and magnectic fields. That is the electric and magnitic polarizabilities of the material are taken to depend linearly on $\vec{E}$ and $\vec{B}$. In this way the free space values ε_0 and μ_0 are converted to ε and μ. Going beyond this linear regime is generally known as "nonlinear optics". It should be remembered that the higher order nonlinear terms are always there (along with relativistic and quantum effects) but they are generally extremely small and hence negligable. It was only with the invention of lasers that electric and magnetic fields could be produced with large enough magnitudes to the make the effect of the nonlinear terms non-negligable. The so called "free" charge and current densities, ρ and $\vec{J}$ above, then pertain to any left over unbound charge that is present.

Take the curl of the curl of $\vec{E}$ equation and treat ε and μ as constants, set ρ and $\vec{J}$ both to zero,

$$\begin{aligned}\vec{\partial}\times\left(\vec{\partial}\times\vec{E}\right) &= -\partial_t\left(\vec{\partial}\times\vec{B}\right)\\ &= -\mu\partial_t\left(\vec{\partial}\times\vec{H}\right)\\ &= -\mu\varepsilon\partial_t^2\vec{E}\end{aligned}$$

But, in component form,

$$\vec{\partial}\times\left(\vec{\partial}\times\vec{E}\right)=\hat{x}_i\epsilon_{ijk}\partial_j\epsilon_{klm}\partial_l E_m$$

where $\hat{x}_i$ with $1 = 1, 2, 3$ are the x, y, z unit vectors, respectively, and ϵ_{ijk} is the totally antisymmetric Levi-Civita tensor, i.e., it equals one when i, j, k is an even permulation of 1,2,3, minus one for an odd permutation and is zero otherwise. Hence we have $\epsilon_{ijk}\epsilon_{klm} = \epsilon_{ijk}\epsilon_{lmk}$. k is doubly repeated and hence summed over 1,2,3, that is

$$\epsilon_{ijk}\epsilon_{lmk} = \epsilon_{ij1}\epsilon_{lm1} + \epsilon_{ij2}\epsilon_{lm2} + \epsilon_{ij3}\epsilon_{lm3}$$

Consider the first term which is non-zero only if i, j, l, m take the values 2 and 3. If $i = l$ then $j = m$ and $\epsilon_{ij1}\epsilon_{lm1} = \epsilon_{ij1}\epsilon_{ij1} = (\epsilon_{ij1})^2 = +1$ (no sum on i and j) whereas if $i = m$ then $j = l$ and $\epsilon_{ij1}\epsilon_{lm1} = \epsilon_{ij1}\epsilon_{ji1} = -(\epsilon_{ij1})^2 = -1$. Doing the same for the other terms and putting the results together gives the identity

$$\epsilon_{ijk}\epsilon_{lmk} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}$$

Using this above gives

$$\begin{aligned}
\hat{x}_i\epsilon_{ijk}\epsilon_{klm}\partial_j\partial_l E_m &= \hat{x}_i\left(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right)\partial_j\partial_l E_m \\
&= \hat{x}_i\left(\partial_i\partial_m E_m - \partial_j\partial_j E_i\right) \\
&= \vec{\partial}\left(\vec{\partial}\cdot\vec{E}\right) - \vec{\partial}^2\vec{E}
\end{aligned}$$

But $\vec{\partial}\cdot\vec{E} = 0$ since $\rho = 0$ and ϵ is constant. Combining these results gives

$$\left(\vec{\partial}^2 - \mu\epsilon\partial_t^2\right)\vec{E} = 0$$

or in component form

$$\left(\vec{\partial}^2 - \mu\epsilon\partial_t^2\right)E_i = 0$$

and each component of $\vec{E}$ satisfies the wave equation. A similar analysis yields

$$\left(\vec{\partial}^2 - \mu\epsilon\partial_t^2\right)B_i = 0$$

But $\left(\vec{\partial}^2 - \mu\epsilon\partial_t^2\right)E_i = 0$ is a linear homogeneous differential equation with constant coefficients and so the basic solution (neglecting degeneracies) is an exponential which, in monochromatic plane wave form, is $\exp\left[i\vec{k}\cdot\vec{r} - i\omega t\right]$. Defining $c = 1/\sqrt{\mu\epsilon}$ which is the speed of light in the material and subsituting this into the wave equation yields the constraint

$$\vec{k}^2 = \frac{\omega^2}{c^2} \Longrightarrow \vec{k} = \pm\frac{\omega}{c}\hat{k}$$

with $\hat{k}$ the unit vector in the direction of propagation. $\vec{\partial}\cdot\vec{E} = 0 \Longrightarrow \vec{k}\cdot\vec{E} = 0$ and so $\vec{E}$ is perpendicular to the direction of propagation of the plane wave. Similarly $\vec{\partial}\cdot\vec{B} = 0 \Longrightarrow \vec{k}\cdot\vec{B} = 0$ and $\vec{B}$ is perpendicular to the direction of propagation of the plane wave.

Writing $\vec{E}(\vec{r},t) = E(\vec{r})e^{-i\omega t}$ and using the fact that because the wave equation for $\vec{E}$ (and for $\vec{B}$) is linear, superposition holds and we can write any solution for $\vec{E}(\vec{r})$ with $\vec{r} = (\vec{\rho},z) = (x,y,z)$ as

$$\vec{E}(\vec{r}) = \vec{E}(\vec{\rho},z) = \int d^2\beta \sum_p \hat{e}_p\left(\hat{\beta}\right)\left(\tilde{E}_p^+\left(\vec{\beta}\right)e^{i\vec{k}\cdot\vec{\rho}+i\gamma(\vec{\beta})z} + \tilde{E}_p^-\left(\vec{\beta}\right)e^{i\vec{k}\cdot\vec{\rho}-i\gamma(\vec{\beta})z}\right)$$

Here $\gamma\left(\vec{\beta}\right) = \sqrt{k^2-\vec{\beta}^2} = \sqrt{\omega^2/c^2-\vec{\beta}^2}$ and so $\vec{k} = \vec{\beta}+\gamma\left(\vec{\beta}\right)z$ with $\vec{k}\cdot\vec{k} = k^2 = \omega^2/c^2$, $\hat{e}_p\left(\hat{\beta}\right)$, with $p = 1,2$, are two unit vectors perpendicular to $\vec{k}$ and to each other defined so that $\hat{e}_1\left(\hat{\beta}\right)\times\hat{e}_2\left(\hat{\beta}\right) = \hat{k}\left(\hat{\beta}\right)$. $\tilde{E}_p^{\pm}\left(\vec{\beta}\right)$ are the amplitudes of the $\pm z$ propagating plane waves as function of polarization, p, and $\vec{\beta}$.

See the discussion in the section on "Propagators versus Greens Functions" for a description of the difference between evanescent and propagating plane waves. The solution for $\vec{E}(\vec{\rho},z)$ has the same general form as that for $\phi(\vec{\rho},z)$ the section on "Propagators versus Greens Functions" and following the same steps to derive the vector propagator for $\vec{E}(\vec{\rho},z)$ gives, again considering only $+z$ propagation

$$\sum_p \hat{e}_p\left(\hat{\beta}\right)\tilde{E}_p^+\left(\vec{\beta}\right) = \int \frac{d^2\rho}{(2\pi)^2}e^{-i\vec{\beta}\cdot\vec{\rho}}\vec{E}(\vec{\rho},0)$$

Take the dot product with $\hat{e}_{p'}\left(\hat{\beta}\right)$ and us the fact that given how $\hat{e}_p\left(\hat{\beta}\right)$ have been defined we have

$$\hat{e}_{p'}\left(\hat{\beta}\right)\cdot\hat{e}_p\left(\hat{\beta}\right) = \delta_{p',p}$$

with $\delta_{p',p}$ being the Kronecker delta. Hence we get

$$\tilde{E}_p\left(\vec{\beta}\right) = \int \frac{d^2\rho}{(2\pi)^2}e^{-i\vec{\beta}\cdot\vec{\rho}}\hat{e}_p\left(\hat{\beta}\right)\cdot\vec{E}(\vec{\rho},0)$$

Substituting back into the general solution above for $\vec{E}(\vec{\rho},z)$ (with $\tilde{E}_p^-\left(\vec{\beta}\right) = 0$) gives (being careful about dummy integrations values)

$$\vec{E}(\vec{\rho},z) = \int d^2\rho' \int \frac{d^2\beta}{(2\pi)^2}e^{i\vec{\beta}\cdot(\vec{\rho}-\vec{\rho}')+i\gamma(\vec{\beta})z}\sum_{p=1}^{2}\hat{e}_p\left(\hat{\beta}\right)\left(\hat{e}_p\left(\hat{\beta}\right)\cdot\vec{E}(\vec{\rho},0)\right)$$

In component form this is

$$E_i(\vec{\rho},z) = \int d^2\rho' \int \frac{d^2\beta}{(2\pi)^2}e^{i\vec{\beta}\cdot(\vec{\rho}-\vec{\rho}')+i\gamma(\vec{\beta})z}\sum_{p=1}^{2}\hat{e}_{p,i}\left(\hat{\beta}\right)\hat{e}_{p,j}\left(\hat{\beta}\right)E_j(\vec{\rho},0)$$

where $\hat{e}_{p,i}\left(\hat{\beta}\right)$ is the i^{th} or x,y,z component of the polarization unit vector $\hat{e}_p\left(\hat{\beta}\right)$.

Now, noting that $\hat{e}_1\left(\hat{\beta}\right), \hat{e}_2\left(\hat{\beta}\right)$ and $\hat{k}\left(\hat{\beta}\right)$ form a complete set of unit basis vectors in 3D we have

$$\vec{E} = \hat{e}_1\left(\hat{\beta}\right)\left(\hat{e}_1\left(\hat{\beta}\right)\cdot\vec{E}\right) + \hat{e}_2\left(\hat{\beta}\right)\left(\hat{e}_2\left(\hat{\beta}\right)\cdot\vec{E}\right) + \hat{k}\left(\hat{\beta}\right)\left(\hat{k}\left(\hat{\beta}\right)\cdot\vec{E}\right)$$

Rearranging gives

$$\begin{aligned}\hat{e}_1\left(\hat{\beta}\right)\left(\hat{e}_1\left(\hat{\beta}\right)\cdot\vec{E}\right) + \hat{e}_2\left(\hat{\beta}\right)\left(\hat{e}_2\left(\hat{\beta}\right)\cdot\vec{E}\right) &= \vec{E} - \hat{k}\left(\hat{\beta}\right)\left(\hat{k}\left(\hat{\beta}\right)\cdot\vec{E}\right) \\ &= \vec{E} - \frac{\vec{k}\left(\hat{\beta}\right)}{k}\left(\frac{\vec{k}\left(\hat{\beta}\right)}{k}\cdot\vec{E}\right)\end{aligned}$$

which in component form is

$$\begin{aligned}\left[\hat{e}_1\left(\hat{\beta}\right)\left(\hat{e}_1\left(\hat{\beta}\right)\cdot\vec{E}\right) + \hat{e}_2\left(\hat{\beta}\right)\left(\hat{e}_2\left(\hat{\beta}\right)\cdot\vec{E}\right)\right]_i &= E_i - \frac{k_i k_j}{k^2}E_j \\ &= \left(\delta_{i,j} - \frac{k_i k_j}{k^2}\right)E_j\end{aligned}$$

Hence we have

$$E_i\left(\vec{\rho}, z\right) = \int d^2\rho' \int \frac{d^2\beta}{\left(2\pi\right)^2} e^{i\vec{\beta}\cdot\left(\vec{\rho}-\vec{\rho}'\right)+i\gamma\left(\vec{\beta}\right)z}\left(\delta_{i,j} - \frac{k_i k_j}{k^2}\right)E_j\left(\vec{\rho}, 0\right)$$

and the vector propagator is given by

$$\begin{aligned}K_{i,j}\left(\vec{\rho}-\vec{\rho}', z\right) &= \int \frac{d^2\beta}{\left(2\pi\right)^2} e^{i\vec{\beta}\cdot\left(\vec{\rho}-\vec{\rho}'\right)+i\gamma\left(\vec{\beta}\right)z}\left(\delta_{i,j} - \frac{k_i k_j}{k^2}\right) \\ &= \left(\delta_{i,j} + \frac{\partial_i\partial_j}{k^2}\right)K\left(\vec{\rho}-\vec{\rho}', z\right)\end{aligned}$$

where $K\left(\vec{\rho}-\vec{\rho}', z\right)$ is the scalar propagator derived above.

17.2 Diffraction?

First, note that both the scalar propagator $K\left(\vec{\rho}-\vec{\rho}', z\right)$ and the vector propagator $K_{i,j}\left(\vec{\rho}-\vec{\rho}', z\right)$ are exact propagators for monochromatic waves. There is no approximation.

Second, note that both propagators reduce to a delta function for $z = 0$ and therefore they give back the $z = 0$ starting field distribution exactly. (For the vector propagator we have to assume input field $\vec{E}\left(\vec{\rho}, 0\right)$ satisfies Gauss law, i.e., $\left(\vec{\partial}\cdot\vec{E}\left(\vec{\rho}, z\right)\right)_{z=0} = 0$, in which case the output field satisfies Gauss law.)

So, if both propagators are exact why do people talk about "Kirchoff Diffraction", for example, as a approximation. The approximation is in the input fields $\phi\left(\vec{\rho}, 0\right)$ and $\vec{E}\left(\vec{\rho}, 0\right)$. In most or almost all cases the input fields themselves are approximations to an exact solution. For example in the case of diffraction

through an aperture (a hole in a thin opaque plate or screen) the exact field $\vec{E}(\vec{\rho},0)$ must satisfy Maxwells equations and the associated boundary conditions on the edges of the aperture. But the effect of the boundary conditions on the field, which is large at the edge of the aperture, decays away, nominally exponentially with distance, in a few wavelengths from the edge. Hence, if the aperture is many wavelengths in size (in all directions), then, more than a few wavelengths from the edge, the transmitted field through the aperture is equal, to a good approximation, to the incident field on the aperture. Hence using incident field distribution and ignoring the effect of the edge caused by the boundary conditions is where the approximation lies.

NOTE: The word "thin" in the statement "a hole in a thin opaque plate or screen" refers to the ratio of the plate thickness to the size of the hole being much less than 1. As the thickness of the plate increases, with the hole size fixed, the "hole" is trending toward becoming a tube. When the thickness is larger than the aperture size then it is definitely a tube. Think fiber optics. In this case the system must be treated as a waveguide, i.e., the eigenmodes must be accounted for using either Maxwell and the boundary conditions to get an exact result, or using some valid approximation to Maxwell and the boundary conditions. But, it is still true that the diffraction downstream of the aperture is given exactly, using the above propagators with the field distribution (calculated using the eigenmodes of the tube for example) exiting the aperture.

18 Stationary Phase, Steepest Descent, Stirling

18.1 Stationary Phase

Consider the integral where $f(x)$ is a real valued function of x

$$I = \int_a^b dx \exp\left[if(x)\right]$$

where $b > a$.

The dominant contribution to the integral comes from the region where the phase, $f(x)$, is constant and/or varies slowly. For a smoothly varying function $f(x)$ the regions where $f(x)$ is slowly varying are centered around the "stationary phase points" x_n of $f(x)$, i.e., positions where

$$\partial_x f(x)|_{x=x_n} = 0$$

NOTE: An equivalent notation is sometimes used

$$\partial_x f(x)|_{x=x_n} = (\partial_x f)(x_n) \equiv f'(x_n)$$

The meaning is first compute the function which is the derivative of $f(x)$ w.r.t. to x then evaluate it at x_n hence $(\partial f)(x_n)$. If $f(x)$ does not have any explicit dependence on x_n then we have

$$\partial_x f(x)|_{x=x_n} = (\partial_x f)(x_n) \equiv f'(x_n) = \partial_{x_n} f(x_n)$$

Consider just a single stationary phase point x_0 which satisfies

$$a \ll x_0 \ll b$$

by which we mean

$$x_0 - a \gg \sqrt{\frac{1}{(\partial_x^2 f)(x_0)}}$$

and

$$b - x_0 \gg \sqrt{\frac{1}{(\partial_x^2 f)(x_0)}}$$

so that x_0 is far from the endpoints of the integration. Expanding $f(x)$ to second order around x_0

$$\begin{aligned} f(x_0 + \eta) &= f(x_0) + (\partial_x f)(x_0)\eta + \frac{1}{2}\left(d_x^2 f\right)(x_0)\eta^2 + \cdots \\ &= f(x_0) + \frac{1}{2}\left(\partial_x^2 f\right)(x_0)\eta^2 + \cdots \end{aligned}$$

yields

$$\begin{aligned} I &\simeq \exp[if(x_0)] \int_{-\infty}^{+\infty} d\eta \exp\left[i\frac{1}{2}\left(\partial_x^2 f\right)(x_0)\eta^2\right] \\ &\simeq \exp[if(x_0)] \sqrt{\frac{2\pi}{(\partial_x^2 f)(x_0)}} \end{aligned}$$

where we have let η range from $-\infty$ to $+\infty$ under the assumption that contributions to the integrals from positions far from x_0 are negligable. This is the method of stationary phase.

18.2 Stirlings Approximation

The same idea as stationary phase can work for integrals of the form

$$I = \int_a^b dx \exp[f(x)]$$

if $f(x)$ has a peak somewhere between a and b and the rms width of $f(x)$ is much smaller than $b - a$. Again $f(x)$ is real valued.

Consider the specific example of Stirlings aproximation for the factorial function $n!$, defined via the Gamma function $\Gamma(1+n) = n!$ with $n \geq 0$,

$$\begin{aligned} \Gamma(1+n) &= \int_0^\infty dx x^n \exp[-x] \\ &= \int_0^\infty dx \exp[n\ln(x) - x] \end{aligned}$$

Note that $x^n \exp[-x] \to 0$ for both $x \to 0$ and $x \to \infty$ and is alway non-negative hence it has a peak value somewhere between $x = 0$ and $x = \infty$. The position of the peak, x_0, is given by

$$\partial_x \left(n \ln(x) - x \right)\big|_{x=x_0} = 0 \Longrightarrow x_0 = n$$

For $n \gg 0$, but finite, the peak position $x_0 = n$ is far from both 0 and ∞ (any finite number is always infinitely far from ∞). Let $x = x_0 + \varepsilon = n + \varepsilon$ and expand about $\varepsilon = 0$ to second order

$$\begin{aligned}
& n \ln(x) - x \\
& = n \ln(n + \varepsilon) - (n + \varepsilon) \\
& = n \left(\ln(n) + \ln\left(1 + \frac{\varepsilon}{n}\right) \right) - n - \varepsilon \\
& \simeq n \ln(n) + n \left(\frac{\varepsilon}{n} - \frac{1}{2} \left(\frac{\varepsilon}{n} \right)^2 \right) - n - \varepsilon + \cdots \\
& \simeq n \ln(n) - n - \frac{1}{2} \frac{\varepsilon^2}{n} + \cdots
\end{aligned}$$

Hence, substituting into the integral above

$$\begin{aligned}
n! = \Gamma(1 + n) & \simeq \exp[n \ln(n) - n] \times \int_{-\infty}^{+\infty} d\varepsilon \exp\left[-\frac{1}{2} \frac{\varepsilon^2}{n} \right] \\
& \simeq \exp[n \ln(n) - n] \sqrt{2\pi n} \\
& \simeq n^n e^{-n} \sqrt{2\pi n} \\
& \simeq \left(\frac{n}{e} \right)^n \sqrt{2\pi n}
\end{aligned}$$

which is Stirlings approximation for $n!$.

19 Harmonic Oscillator Path Integral

Lagrangian

$$L(x, \partial_t x, t) = \frac{m}{2} (\partial_t x)^2 - \frac{m\omega^2}{2} x^2 + f(t)\, x$$

Equation of motion

$$\partial_t \frac{\partial L}{\partial (\partial_t x)} - \frac{\partial L}{\partial x} = 0 \quad \Rightarrow \quad {\partial_t}^2 x + \omega^2 x = f/m$$

Boundary Conditions

$$\begin{aligned}
x(t_a) &= x_a \\
x(t_b) &= x_b
\end{aligned}$$

The general solution is the sum of a homogeneous solution $x_{HS}(t)$ and a particular integral $x_{PI}(t)$. The funtamental form of the homogeneous solution is a linear combination of $\exp[-i\omega t]$ since

$$\left(\partial_t{}^2 + \omega^2\right) \exp[-i\omega t] = 0$$

The particular integral has the form

$$x_{PI}(t) = \int dt' G(t-t') f(t')/m$$

with the Greens function G satisfying

$$\left(\partial_t{}^2 + \omega^2\right) G(t-t') = \delta(t-t')$$

We need to incorporate the boundary conditions. This will be done by requiring the homogeneous solution to satisfy the boundary conditions, i.e, $x_{HS}(t_a) = x_a$ and $x_{HS}(t_b) = x_b$ and fixing $x_{PI}(t_b) = x_{PI}(t_a) = 0$.

Given this the homogeneous solution can be wrtten as

$$x_{HS}(t) = \frac{x_b \sin(\omega(t-t_a)) + x_a \sin(\omega(t_b-t))}{\sin(\omega(t_b-t_a))}$$

which clearly satisfies the boundary conditions. Also we want $x(t)$ to be a real function.

To get a G that satifies this we will need to add homogeneous solutions to the "raw" G defined by

$$G(t) = -\int \frac{dq}{2\pi} \frac{e^{-iqt}}{q^2-\omega^2}$$

We do not need G to be causal since we have chosen the end point position but let us start with the causal or retarded G, i.e., $G(t<0) = 0$ and and then add a homogeneous solution to get the necessary boundary conditions on G. First, pick how to circulate the pole in the denominator so that we get the retarded propagator. With the numerator of the form $\exp[-iqt]$ we need to displace both poles $\pm\omega$ into the lower half plane. Then for $t > 0$ which requires closing the contour in the lower half plane we encircle the poles and get a nonzero result but for $t < 0$ which requires closing the contour in the upper half plane there are no poles and we get 0. To do this let

$$\begin{aligned} q^2 - \omega^2 &= (q-\omega)(q+\omega) \\ &\to (q-(\omega-i\varepsilon))(q+(\omega+i\varepsilon)) \end{aligned}$$

where $\varepsilon = 0_+$. Using this we have

$$\begin{aligned} G(t) &= -\int \frac{dq}{2\pi} \frac{e^{-iqt}}{(q-(\omega-i\varepsilon))(q+(\omega+i\varepsilon))} \\ &= \frac{2\pi i}{2\pi}\left(\frac{\exp[-i(\omega-i\varepsilon)t]}{2\omega} + \frac{\exp[-i(-(\omega+i\varepsilon))t]}{-2\omega}\right)\theta(t) \\ &= \frac{i}{2\omega}(-2i\sin(\omega t))\,\theta(t) \\ &= \frac{\sin(\omega t)}{\omega}\theta(t) \end{aligned}$$

where $\theta(t)$ is the Heaviside step function. Note that $G(t)$ is real and so given a real $f(t)$ the solution $x(t)$ will be real.

Test this result

$$\begin{aligned} \partial_t \frac{\sin(\omega t)}{\omega}\theta(t) &= \cos(\omega t)\,\theta(t) + \frac{\sin(\omega t)}{\omega}\delta(t) = \cos(\omega t)\,\theta(t) \\ {\partial_t}^2 \frac{\sin(\omega t)}{\omega}\theta(t) &= -\omega\sin(\omega t)\,\theta(t) + \cos(\omega t)\,\delta(t) = -\omega^2 \underbrace{\frac{\sin(\omega t)}{\omega}\theta(t)}_{G(t)} + \delta(t) \end{aligned}$$

and so

$$\left({\partial_t}^2 + \omega^2\right) G(t) = \delta(t)$$

as required and the $\theta(t)$ factor shows it is the retarded Greens function.

Using this we have

$$\left({\partial_t}^2 + \omega^2\right)\int_{t_a}^{t_b} dt' G(t-t')\,f(t')/m = \int_{t_a}^{t_b} dt'\,\delta(t-t')\,f(t')/m = f(t)/m$$

which accounts for the inhomogeneous term on the right hand side of the differential equation.

Now add a homogenous solution to $\int Gf/m$ so that the integral vanishes at t_a and t_b First note that

$$\begin{aligned} \int_{t_a}^{t_b} dt' G(t-t')\,f(t')/m &= \int_{t_a}^{t_b} dt' \frac{\sin(\omega(t-t'))}{\omega}\theta(t-t')\,f(t')/m \\ &= \int_{t_a}^{t} dt' \frac{\sin(\omega(t-t'))}{\omega} f(t')/m \end{aligned}$$

which vanishes automatically at $t = 0$. At $t = t_b$

$$\begin{aligned} \int_{t_a}^{t_b} dt' G(t_b-t')\,f(t')/m &= \int_{t_a}^{t_b} dt' \frac{\sin(\omega(t_b-t'))}{\omega}\theta(t_b-t')\,f(t')/m \\ &= \int_{t_a}^{t_b} dt' \frac{\sin(\omega(t_b-t'))}{\omega} f(t')/m \end{aligned}$$

This is a constant independent of t.If we multiply it by a homogeneous solution which vanishes at $t=t_a$ and is 1 at $t=t_b$ and subtract this from $\int Gf/m$ then the result will vanish at both t_a and t_b and will solve the inhomogenous equation with the desired boundary conditions. The homogeneous solution which does this is $\sin(\omega(t-t_a))/\sin(\omega(t_b-t_a))$ and we get

$$\begin{aligned}x_{PI}(t) &= \int_{t_a}^{t_b} dt' G(t-t') f(t')/m - \left(\int_{t_a}^{t_b} dt' \frac{\sin(\omega(t_b-t'))}{\omega} f(t')/m\right) \frac{\sin(\omega(t-t_a))}{\sin(\omega(t_b-t_a))} \\ &= \int_{t_a}^{t_b} dt' \left[\frac{\sin(\omega(t-t'))}{\omega}\theta(t-t') - \frac{\sin(\omega(t_b-t'))}{\omega}\frac{\sin(\omega(t-t_a))}{\sin(\omega(t_b-t_a))}\right] f(t')/m\end{aligned}$$

Using $1=\theta(t-t')+\theta(t'-t)$ we can combine the terms in square brackets into a new Greens function which satisfies the desired boundary conditions

$$\begin{aligned}g(t-t') &= \frac{1}{\omega}\left(\sin(\omega(t-t'))\theta(t-t') - \sin(\omega(t_b-t'))\frac{\sin(\omega(t-t_a))}{\sin(\omega(t_b-t_a))}\right) \\ &= \frac{1}{\omega}\left(\left(\sin(\omega(t-t'))\sin(\omega(t_b-t_a)) - \sin(\omega(t_b-t'))\sin(\omega(t-t_a))\right)\frac{\theta(t-t')}{\sin(\omega(t_b-t_a))}\right. \\ &\quad \left. - \sin(\omega(t_b-t'))\sin(\omega(t-t_a))\frac{\theta(t'-t)}{\sin(\omega(t_b-t_a))}\right) \\ &= \frac{1}{\omega}\left(\left(\sin(\omega(t_a-t'))\sin(\omega(t_b-t))\right)\frac{\theta(t-t')}{\sin(\omega(t_b-t_a))}\right. \\ &\quad \left. + \sin(\omega(t_b-t'))\sin(\omega(t_a-t))\frac{\theta(t'-t)}{\sin(\omega(t_b-t_a))}\right)\end{aligned}$$

This is the standard result for a harmonic oscillator Greens function which vanishes when either t or t' equals t_a or t_b. It can be seen to be just a linear combination of homogeneous solutions chosen in such a way that it is continuous but with a discontinuous first derivative so that the second derivative contains a term proportional to a delta function.

We now have the full solution

$$x(t) = \frac{x_b\sin(\omega(t-t_a)) + x_a\sin(\omega(t_b-t))}{\sin(\omega(t_b-t_a))} + \int_{t_a}^{t_b} dt' g(t-t') f(t')/m$$

where $t_b-t_a=T$.

Now evaluate the action for the classical solution trajectory $x(t)$

$$\begin{aligned}S &= \int_{t_a}^{t_b} dt L(x, \partial_t x) \\ &= \int_{t_a}^{t_b} dt \left(\frac{m}{2}(\partial_t x)^2 - \frac{m\omega^2}{2}x^2 + f(t)x\right)\end{aligned}$$

Integrate the first term by parts to get

$$\begin{aligned}
S_{classical} &= \frac{m}{2}\, x\partial_t x|_{t_a}^{t_b} - \int_{t_a}^{t_b} dt\; x\left(\left(\frac{m}{2}{\partial_t}^2 + \frac{m\omega^2}{2}\right)x - f(t)\right) \\
&= \frac{m}{2}\left(x_b\left(\partial_t x\right)_{t_b} - x_a\left(\partial_t x\right)_{t_a}\right) - \int_{t_a}^{t_b} dt\; x\left(\left(\frac{m}{2}{\partial_t}^2 + \frac{m\omega^2}{2}\right)x - f(t)\right) \\
&= \frac{m}{2}\left(x_b\left(\partial_t x\right)_{t_b} - x_a\left(\partial_t x\right)_{t_a}\right) + \frac{1}{2}\int_{t_a}^{t_b} dt\; x f(t)
\end{aligned}$$

where in the third line we have used the fact that $\left(\frac{m}{2}{\partial_t}^2 + \frac{m\omega^2}{2}\right)x = f/2$. Substituting the solution for x yields

$$\begin{aligned}
S_{classical} &= \frac{m\omega}{2\sin\left(\omega\left(t_b - t_a\right)\right)}\left[\begin{array}{c} x_b\left(x_b\cos\left(\omega\left(t_b - t_a\right)\right) - x_a\right) \\ -x_a\left(x_b - x_a\cos\left(\omega\left(t_b - t_a\right)\right)\right)\end{array}\right] \\
&+ \frac{m}{2}\frac{x_b}{\sin\left(\omega\left(t_b - t_a\right)\right)}\int_{t_a}^{t_b} dt\sin\left(\omega\left(t - t_a\right)\right)f(t)/m \\
&+ \frac{m}{2}\frac{x_a}{\sin\left(\omega\left(t_b - t_a\right)\right)}\int_{t_a}^{t_b} dt\sin\left(\omega\left(t_b - t\right)\right)f(t)/m \\
&+ \frac{1}{2}\int_{t_a}^{t_b} dt'\frac{x_b\sin\left(\omega\left(t - t_a\right)\right) + x_a\sin\left(\omega\left(t_b - t\right)\right)}{\sin\left(\omega\left(t_b - t_a\right)\right)}f(t') \\
&+ \frac{1}{2}\int_{t_a}^{t_b} dtdt' f(t)\, g\left(t - t'\right) f(t')/m \\
&= \frac{m\omega}{2\sin\left(\omega\left(t_b - t_a\right)\right)}\left[\begin{array}{c} x_b\left(x_b\cos\left(\omega\left(t_b - t_a\right)\right) - x_a\right) \\ -x_a\left(x_b - x_a\cos\left(\omega\left(t_b - t_a\right)\right)\right) \\ +\frac{2}{m\omega}x_b\int_{t_a}^{t_b} dt\sin\left(\omega\left(t - t_a\right)\right)f(t) \\ +\frac{2}{m\omega}x_a\int_{t_a}^{t_b} dt\sin\left(\omega\left(t_b - t\right)\right)f(t)\end{array}\right] \\
&+ \frac{1}{2m}\int_{t_a}^{t_b} dtdt' f(t)\, g\left(t - t'\right) f(t')
\end{aligned}$$

Need to complete

20 Fluctuation Dissipation Theorem

20.1 Derivation

Start with the partition function for a generic field (or order parameter) $\phi\left(\vec{r}\right)$

$$Z = \int D\phi e^{-\beta H[\phi]}$$

where $D\phi$ indicates functional integration, $\beta = 1/k_B T$ and $H\left[\phi\right]$ = Hamiltonian = functional of $\phi\left(\vec{r}\right)$.

The expectation values of ϕ are given by

$$\left\langle \phi\left(\vec{r}_1\right)\phi\left(\vec{r}_2\right)\cdots\phi\left(\vec{r}_n\right)\right\rangle = \frac{1}{Z}\int D\phi\ \phi\left(\vec{r}_1\right)\phi\left(\vec{r}_2\right)\cdots\phi\left(\vec{r}_n\right)\ e^{-\beta H[\phi]}$$

where the $1/Z$ is for normalization.

In order to compute expectation values of ϕ add the term

$$\int d\vec{r} J\left(\vec{r}\right)\phi\left(\vec{r}\right)$$

to the Hamiltonian to get

$$Z\left[J\right] = \int D\phi e^{-\beta H[\phi]-\beta\int d\vec{r}J\phi}$$

$Z\left[J\right]$ is the generating functional of moments or correlation functions of ϕ, i.e.,

$$\begin{aligned}\frac{1}{Z}\ \frac{1}{(-\beta)^n}\frac{D^n Z\left[J\right]}{DJ\left(\vec{r}_1\right)DJ\left(\vec{r}_2\right)\cdots DJ\left(\vec{r}_n\right)}\bigg|_{J=0} &= \frac{1}{Z}\int D\phi\ \phi\left(\vec{r}_1\right)\phi\left(\vec{r}_2\right)\cdots\phi\left(\vec{r}_n\right)\ e^{-\beta H[\phi]} \\ &= \left\langle \phi\left(\vec{r}_1\right)\phi\left(\vec{r}_2\right)\cdots\phi\left(\vec{r}_n\right)\right\rangle\end{aligned}$$

$D/DJ\left(\vec{r}\right)$ indicates functional differentiation.

Now note that J can be thought of as a linear external "force" or driving term on ϕ and so the susceptibility of ϕ with respect to J is given by

$$\chi\left(\vec{r},\vec{r}'\right) = \left.\frac{D\left\langle\phi\left(\vec{r}\right)\right\rangle_J}{DJ\left(\vec{r}'\right)}\right|_{J=0}$$

where the notation $\left\langle\cdots\right\rangle_J$ indicates J has not yet been set to zero, i.e., $\left\langle\phi\left(\vec{r}\right)\right\rangle_J$ is the expectation value of ϕ in the presence of the "force" or driving term J.

But from above

$$\left\langle\phi\left(\vec{r}\right)\right\rangle_J = -\frac{1}{\beta Z}\frac{DZ\left[J\right]}{DJ\left(\vec{r}\right)} = -\frac{k_B T}{Z}\frac{DZ\left[J\right]}{DJ\left(\vec{r}\right)}$$

and so

$$\begin{aligned}\left.\frac{D\left\langle\phi\left(\vec{r}\right)\right\rangle_J}{DJ\left(\vec{r}'\right)}\right|_{J=0} &= -\frac{k_B T}{Z}\left.\frac{D^2 Z\left[J\right]}{DJ\left(\vec{r}\right)DJ\left(\vec{r}'\right)}\right|_{J=0} \\ &= -k_B T\beta^2\left\langle\phi\left(\vec{r}\right)\phi\left(\vec{r}'\right)\right\rangle \\ &= -\frac{1}{k_B T}\left\langle\phi\left(\vec{r}\right)\phi\left(\vec{r}'\right)\right\rangle\end{aligned}$$

which yields

$$\chi\left(\vec{r},\vec{r}'\right) = -\frac{1}{k_B T}\left\langle\phi\left(\vec{r}\right)\phi\left(\vec{r}'\right)\right\rangle$$

This is how I understand the fluctuation-dissipation theorem.

20.1.1 Gillespies Paper (Am. J. Phys. 61, p. 1077 (1993))

Heres are two quick ways to get from Gillespies Eq(1) to his Eq(3). There's no new physics here just some fun math.

The differential equation is

$$(M\partial_t + \gamma) V(t) = f\Gamma(t)$$

M and f are constants, $\Gamma(t)$ is a "random" function with the given statistics

$$\langle \Gamma(t) \rangle = 0$$
$$\langle \Gamma(t)\Gamma(t') \rangle = \delta(t-t')$$

The goal is to solve for the statistics of the function $V(t)$.

Method 1. The solution to the differential equation can be written as

$$V(t) = f\int_0^\infty dt' G(t-t')\Gamma(t')$$

The Greens function is given by

$$\begin{aligned} G(t) &= \int \frac{d\omega}{2\pi} \frac{e^{-i\omega t}}{(-iM\omega+\gamma)} \\ &= \frac{1}{-i2\pi M}\int d\omega \frac{e^{-i\omega t}}{(\omega + i\gamma/M)} \\ &= \frac{1}{M} e^{-\gamma t/M}\theta(t) \end{aligned}$$

and so

$$V(t) = \frac{f}{M}\int_0^t dt' e^{-\gamma(t-t')/M}$$

which vanishes for $t = 0$.

From the solution for $V(t)$ it follows that

$$\begin{aligned} \langle V(t)V(t') \rangle &= f^{\,2}\int_0^\infty dt_1 dt_2 G(t-t_1)G(t'-t_2)\langle \Gamma(t)\Gamma(t') \rangle \\ &= \frac{f^{\,2}}{M^2} e^{-\gamma(t+t')/M}\int_0^\infty dt_1 e^{2\gamma t_1/M}\ \theta(t-t_1)\theta(t'-t_1) \end{aligned}$$

For $t > t'$ we get

$$\langle V(t)V(t') \rangle = \frac{f^{\,2}}{2\gamma M} e^{-\gamma(t+t')/M}\left(e^{2\gamma t'/M} - 1\right)$$

Now let $t' \to t$

$$\left\langle V(t)^2 \right\rangle = \frac{f^{\,2}}{2\gamma M}\left(1 - e^{-2\gamma t/M}\right)$$

and so

$$\left\langle V\left(t\to\infty\right)^{2}\right\rangle =\frac{f^{\ 2}}{2\gamma M}$$

Demanding

$$\frac{M}{2}\left\langle V\left(t\to\infty\right)^{2}\right\rangle =\frac{1}{2}k_{B}T$$

gives

$$f=\sqrt{2\gamma k_{B}T}$$

Method 2 The probability distribution functional for $V\left(t\right)$ given $\Gamma\left(t\right)$ is

$$P\left[V;\Gamma\right]\sim\delta_{D}\left[\frac{1}{f}\left(M\partial_{t}+\gamma\right)V-\Gamma\right]$$

For the slightly more general case of $\left\langle \Gamma\left(t\right)\right\rangle =0$ and $\left\langle \Gamma\left(t\right)\Gamma\left(t'\right)\right\rangle =H\left(t,t'\right)$, a Gaussian probability distribution functional for $\Gamma\left(t\right)$ is given by

$$P\left[\Gamma\right]\sim\exp\left[-\int dtdt'\Gamma\left(t\right)\frac{H^{-1}\left(t,t'\right)}{2}\Gamma\left(t'\right)\right]$$

The normalizations will be sorted out below. You can keep track of it by including the appropriate determinant factors but it just makes things ugly.

The probability for $V\left(t\right)$ is then given by

$$\begin{aligned}P\left[V\right] & =\int D\Gamma P\left[V;\Gamma\right]P\left[\Gamma\right]\\ & \sim\int D\Gamma\exp\left[-\int dtdt'\Gamma\left(t\right)\frac{H^{-1}\left(t,t'\right)}{2}\Gamma\left(t'\right)\right]\delta\left[\frac{\left(M\partial_{t}+\gamma\right)V}{f}-\Gamma\right]\\ & \sim\exp\left[-\int dtdt'\left(\left(M\partial_{t}+\gamma\right)V\right)\left(t\right)\frac{H^{-1}\left(t,t'\right)}{2f^{\ 2}}\left(\left(M\partial_{t}+\gamma\right)V\right)\left(t'\right)\right]\end{aligned}$$

For the specific case of $\left\langle \Gamma\left(t\right)\Gamma\left(t'\right)\right\rangle =\delta\left(t-t'\right)$ we have $H\left(t,t'\right)=\delta\left(t-t'\right)$ and $H^{-1}\left(t,t'\right)=\delta\left(t-t'\right)$ and so

$$\begin{aligned}P\left[V\right] & \sim\exp\left[-\int dt\frac{\left(\left(M\partial_{t}+\gamma\right)V\left(t\right)\right)^{2}}{2f^{\ 2}}\right]\\ & \sim\exp\left[-\int dt\frac{V\left(t\right)\left(-M^{2}{\partial_{t}}^{2}+\gamma^{2}\right)V\left(t\right)}{2f^{\ 2}}\right]\end{aligned}$$

after integration by parts in the exponent.

The generating functional for moments of $V\left(t\right)$ is given by

$$\begin{aligned}Z\left[J\right] & =\int DV\ P\left[V\right]e^{i\int dtV\left(t\right)J\left(t\right)}\\ & \sim\exp\left[-\int dtdt'J\left(t\right)\frac{f^{\ 2}}{2}g\left(t-t'\right)J\left(t'\right)\right]\end{aligned}$$

where

$$\left(-M^2{\partial_t}^2+\gamma^2\right)g\left(t-t'\right)=\delta\left(t-t'\right)$$

Note that

$$Z\left[J=0\right]=1=\int DV\ P\left[V\right]$$

and therefore the result is properly normalized (all the det's have cancelled).

Now

$$\left\langle V\left(t\right)V\left(t'\right)\right\rangle = f^{\ 2}g\left(t-t'\right)$$

But

$$\begin{aligned} g\left(t\right) &= \frac{1}{M^2}\int\frac{d\omega}{2\pi}\frac{e^{-i\omega t}}{\left(\omega^2+\gamma^2/M^2\right)} \\ &= \frac{1}{M^2}\int\frac{d\omega}{2\pi}\frac{e^{-i\omega t}}{\left(\omega+i\gamma/M\right)\left(\omega-i\gamma/M\right)} \\ &= \frac{e^{-\gamma|t|/M}}{2\gamma M} \end{aligned}$$

And so

$$\frac{M}{2}\left\langle V\left(t\right)^2\right\rangle=\frac{f^{\ 2}}{4\gamma}=\frac{1}{2}k_BT\quad\Rightarrow f=\sqrt{2\gamma k_BT}$$

Same as before.

21 Proof: Bessel Function Closure Relation

Using the integral definition of $J_0\left(x\right)$ in terms of $\exp\left[ix\cos\left[\theta\right]\right]$ gives

$$\begin{aligned} \int_0^\infty d\rho\ \rho J_0\left(a\rho\right)J_0\left(b\rho\right) &= \int_0^\infty d\rho\ \rho J_0\left(a\rho\right)J_0\left(-b\rho\right) \\ &= \frac{1}{\left(2\pi\right)^2}\int_0^{2\pi}d\theta' d\theta\int_0^\infty d\rho\ \rho\exp\left[ia\rho\cos\left(\theta\right)-ib\rho\cos\left(\theta'\right)\right] \end{aligned}$$

The key point is the integrals over θ and θ' can be shifted by an arbitrary angle without changing the result. Shift both by the same angle ϕ and average over ϕ.

$$\int_0^\infty d\rho\,\rho J_0\left(a\rho\right)J_0\left(-b\rho\right)=\frac{1}{\left(2\pi\right)^2}\int_0^{2\pi}d\theta' d\theta\int_0^{2\pi}\frac{d\phi}{2\pi}\int_0^\infty d\rho\,\rho\exp\left[ia\rho\cos\left(\theta-\phi\right)-ib\rho\cos\left(\theta'-\phi\right)\right]$$

Letting $\vec{a}=a\cos\left(\theta\right)\hat{x}+a\sin\left(\theta\right)\hat{y}$, $\vec{b}=b\cos\left(\theta'\right)\hat{x}+b\sin\left(\theta'\right)\hat{y}$ and $\vec{\rho}=\rho\cos\left(\phi\right)\hat{x}+\rho\sin\left(\phi\right)\hat{y}=x\hat{x}+y\hat{y}$ where $\hat{x}$ and $\hat{y}$ are the x,y unit vectors respectively, we

have

$$\begin{aligned}
&\frac{1}{(2\pi)^2}\int_0^{2\pi} d\theta' d\theta \int_0^{2\pi}\frac{d\phi}{2\pi}\int_0^{\infty} d\rho\ \rho \exp\left[ia\rho\cos(\theta-\phi) - ib\rho\cos\left(\theta'-\phi\right)\right] \\
&= \frac{1}{2\pi}\int_0^{2\pi} d\theta' d\theta \int \frac{dxdy}{(2\pi)^2}\exp\left[i\vec{a}\cdot\vec{\rho} - i\vec{b}\cdot\vec{\rho}\right] \\
&= \frac{1}{2\pi}\int_0^{2\pi} d\theta' d\theta\delta\left(\vec{a}-\vec{b}\right) \\
&= \frac{1}{2\pi}\int_0^{2\pi} d\theta' d\theta\delta\left(a\cos(\theta) - b\cos\left(\theta'\right)\right)\delta\left(a\sin(\theta) - b\sin\left(\theta'\right)\right)
\end{aligned}$$

Let $a\cos(\theta) = a_x$ and $a\sin(\theta) = a_y$ so that $a = a_x^2 + a_y^2$, and consider the integral

$$\begin{aligned}
&\frac{1}{2\pi}\int_0^{2\pi} d\theta'\delta\left(a_x - b\cos\left(\theta'\right)\right)\delta\left(a_y - b\sin\left(\theta'\right)\right) \\
&= \frac{1}{2\pi}\int_0^{2\pi} d\theta'\frac{a_x + b\cos\left(\theta'\right)}{a_x + b\cos\left(\theta'\right)}\delta\left(a_x - b\cos\left(\theta'\right)\right)\delta\left(a_y - b\sin\left(\theta'\right)\right) \\
&= \frac{1}{2\pi}\int_0^{2\pi} d\theta' 2b\cos\left(\theta'\right)\delta\left(a_x^2 - b^2\cos\left(\theta'\right)^2\right)\delta\left(a_y - b\sin\left(\theta'\right)\right) \\
&= \frac{1}{2\pi}\int_0^{2\pi} d\theta' 2b\cos\left(\theta'\right)\delta\left(a_x^2 - b^2\left(1-\sin\left(\theta'\right)^2\right)\right)\delta\left(a_y - b\sin\left(\theta'\right)\right) \\
&= \frac{1}{2\pi}\int_0^{2\pi} d\theta' 2\cos\left(\theta'\right)\delta\left(a_x^2 - b^2\left(1-\frac{a_y^2}{b^2}\right)\right)\delta\left(\frac{a_y}{b} - \sin\left(\theta'\right)\right) \\
&= \frac{1}{2\pi}\int_0^{2\pi} d\theta' 2\cos\left(\theta'\right)\delta\left(a^2 - b^2\right)\delta\left(\frac{a_y}{b} - \sin\left(\theta'\right)\right) \\
&= \frac{1}{2\pi}\int_0^{2\pi} d\theta'\frac{1}{b}\cos\left(\theta'\right)\delta(a-b)\,\delta\left(\frac{a_y}{b} - \sin\left(\theta'\right)\right) \\
&= \frac{1}{2\pi b}\delta(a-b)\underbrace{\int d\left(\sin\left(\theta'\right)\right)\delta\left(\frac{a_y}{b} - \sin\left(\theta'\right)\right)}_{=1}
\end{aligned}$$

In the last step, because $\delta(a-b)$ requires $a = b$ it follows that a_y/b must lie between -1 and $+1$ and so the integral over $d\left(\sin\left(\theta'\right)\right)$ must equal 1. Substituting this result above we get

$$\int_0^{\infty} d\rho\ \rho J_0(a\rho) J_0(b\rho) = \frac{1}{b}\delta(a-b)$$

22 Fokker-Planck Equation

22.1 Derivation

Consider the following differential equation

$$\frac{\partial}{\partial t}X(t) = V(X(t)) + N(t)$$

where $V(X(t))$ is a given function of $X(t)$ and $N(t)$ is a noise term, that is, $N(t)$ is to be treated as a random function.

Let the probability distribution for $N(t)$ to be the particular function $\nu(t)$ be

$$P_{N(t)}(\nu(t))$$

It follows from the probability mapping approach discussed above that the probability for X at time t, $X(t)$, to have the particular value x is given by

$$P_{X(t)}(x) = \int \delta\nu(\cdot)\,\delta_D(x - X(t, x_0, \nu(\cdot)))\,P_{N(t)}(\nu(\cdot))$$

where $\delta\nu(\cdot)$ indicates integration over all functions $\nu(t)$. (The "$\cdot$" notation is explained below.) Here $X(t, x_0, \nu(\cdot))$ is the solution to the differential equation with initial condition $X(0) = x_0$ and with the particular function noise function $\nu(t)$ and $\delta_D(\cdots)$ is the Dirac delta function.

To simplify notation we will write $X(t, x_0, \nu(\cdot)) = X(t)$. Don't forget that $X(t)$ depends on $\nu(\cdot)$. The "$\cdot$" notation is used to make it clear that $\delta\nu(\cdot)$ and $\nu(\cdot)$ do not depend on time. Think of $\delta\nu(\cdot)$ as being the product

$$d\nu(0)\,d\nu(\Delta t)\,d\nu(2\Delta t)\,d\nu(3\Delta t)\cdots$$

and $\nu(\cdot)$ as the set of values

$$\{\nu(0), \nu(\Delta t), \nu(2\Delta t), \nu(3\Delta t), \cdots\}$$

as $\Delta t \to 0$. Listing out the values this way makes it clear that $\delta\nu(\cdot)$ and $\nu(\cdot)$ don't depend on t. Another way to think about it is, since $\nu(t)$ is a random (assume real valued) function, then at any particular time t it can have any value between $-\infty$ and $+\infty$.

Take the derivative w.r.t. t on both sides. (The only t dependence on the

right is the t dependence of X)

$$\begin{aligned}
\frac{\partial}{\partial t}P_{X(t)}(x) &= \int \delta\nu(\cdot)\frac{\partial}{\partial t}\delta_D(x-X(t))P_{N(t)}(\nu(\cdot)) \\
&= \int \delta\nu(\cdot)\frac{\partial X(t)}{\partial t}\frac{\partial}{\partial X(t)}\delta_D(x-X(t))P_{N(t)}(\nu(\cdot)) \\
&= -\int \delta\nu(\cdot)\frac{\partial X(t)}{\partial t}\frac{\partial}{\partial x}\delta_D(x-X(t))P_{N(t)}(\nu(\cdot)) \\
&= -\frac{\partial}{\partial x}\int \delta\nu(\cdot)\frac{\partial X(t)}{\partial t}\delta_D(x-X(t))P_{N(t)}(\nu(\cdot)) \\
&= -\frac{\partial}{\partial x}\int \delta\nu(\cdot)(V(X(t))+\nu(t))\delta_D(x-X(t))P_{N(t)}(\nu(\cdot)) \\
&= -\frac{\partial}{\partial x}\int \delta\nu(\cdot)(V(x)+\nu(t))\delta_D(x-X(t))P_{N(t)}(\nu(\cdot)) \\
&= -\frac{\partial}{\partial x}V(x)P_{X(t)}(x) \\
&\quad -\frac{\partial}{\partial x}\int \delta\nu(\cdot)\nu(t)\delta_D(x-X(t))P_{N(t)}(\nu(\cdot))
\end{aligned}$$

The reason for writing $\nu(t)$ and not $\nu(\cdot)$ is that the factor $\nu(t)$ is being evaluated at the specific time t.

The statistics for the noise function is often chosen to be zero mean and have a specific autocorrelation function, i.e.,

$$\begin{aligned}
\langle\nu(t)\rangle &= 0 \\
\langle\nu(t)\nu(t')\rangle &= C(t-t')
\end{aligned}$$

where the $t-t'$ is chosen so that the noise statistics are time translation invariant. One can choose $C(t,t')$ instead which makes the noise, nominally, not time translation invariant. The most convenient specific form for $P_{N(t)}(\nu(t))$ that yields these statistics is a zero mean Gaussian in function space, i.e.,

$$P_{N(t)}(\nu(t)) = \frac{1}{\text{normalization}}\exp\left[-\frac{1}{2}\int dtdt'\nu(t)C(t,t')^{-1}\nu(t')\right]$$

where $C(t,t')^{-1}$ is the inverse of $C(t,t')$,

$$\int dt''C(t,t'')^{-1}C(t'',t') = \delta_D(t-t')$$

Since $P_{N(t)}(\nu(t))$ occurs to the first power in every term in the differential equation above, the "normalization" factor cancels out.

Note that $C(t,t')^{-1}$ is generally a differential operator. For example, for

$$C(t,t') = \exp[-\alpha|t-t'|]$$

we have

$$C(t,t')^{-1} = \frac{1}{2\alpha}\left(-\frac{\partial^2}{\partial t^2}+\alpha^2\right)\delta_D(t-t')$$

since

$$\begin{aligned}\int dt''C\left(t,t''\right)^{-1}C\left(t'',t'\right) &= \int dt''\frac{1}{2\alpha}\left(-\frac{\partial^2}{\partial t^2}+\alpha^2\right)\delta_D\left(t-t''\right)\exp\left[-\alpha\left|t''-t'\right|\right]\\ &= \frac{1}{2\alpha}\left(-\frac{\partial^2}{\partial t^2}+\alpha^2\right)\exp\left[-\alpha\left|t-t'\right|\right]\\ &= \delta_D\left(t-t'\right)\end{aligned}$$

This is obtained using $\left|t-t'\right| = sgn\left(t-t'\right)\left(t-t'\right)$ where $sgn\left(\cdots\right)$ is the sign of "$\cdots$"

$$sgn\left(x\right) = \begin{cases} +1 \text{ for } x>0 \\ -1 \text{ for } x<0 \end{cases}$$

It follows from this that $\partial/\partial t\ sgn\left(t-t'\right) = 2\delta_D\left(t-t'\right)$.(Draw a graph of $sgn\left(t\right)$ from negtive to positive values of $t-t'$ and think of taking a derivative.)

$$\begin{aligned}\frac{\partial^2}{\partial t^2}\exp\left[-\alpha\left|t-t'\right|\right] &= -\alpha\frac{\partial}{\partial t}\left(\frac{\partial}{\partial t}\left|t-t'\right|\right)\exp\left[-\alpha\left|t-t'\right|\right]\\ &= -\alpha\frac{\partial}{\partial t}\left(\frac{\partial}{\partial t}\left(sgn\left[\left(t-t'\right)\right]\right)\left(t-t'\right)\right)\exp\left[-\alpha\left|t-t'\right|\right]\\ &= -\alpha\frac{\partial}{\partial t}\left(2\delta_D\left(t-t'\right)\left(t-t'\right)+sgn\left(t-t'\right)\right)\exp\left[-\alpha\left|t-t'\right|\right]\\ \text{use } \delta_D\left(t-t'\right)f\left(t-t'\right) &= \delta_D\left(t-t'\right)f\left(0\right)\\ &\quad \text{for any } f\left(t\right) \text{ to get}\\ &= -\alpha\frac{\partial}{\partial t}\left(sgn\left(t-t'\right)\exp\left[-\alpha\left|t-t'\right|\right]\right)\\ &= -2\alpha\delta_D\left(t-t'\right)+\alpha^2 sgn\left(t-t'\right)^2\exp\left[-\alpha\left|t-t'\right|\right]\\ \text{but } sgn\left(t-t'\right)^2 &= 1 \text{ and so}\\ \frac{\partial^2}{\partial t^2}\exp\left[-\alpha\left|t-t'\right|\right] &= -2\alpha\delta_D\left(t-t'\right)+\alpha^2\exp\left[-\alpha\left|t-t'\right|\right]\end{aligned}$$

which gives, as above

$$\frac{1}{2\alpha}\left(-\frac{\partial^2}{\partial t^2}+\alpha^2\right)\exp\left[-\alpha\left|t-t'\right|\right] = \delta_D\left(t-t'\right)$$

Often the following choice is made for $C\left(t-t'\right)$

$$C\left(t-t'\right) = \sigma^2\delta_D\left(t-t'\right)$$

in which case we have

$$C\left(t-t'\right)^{-1} = \frac{1}{\sigma^2}\delta_D\left(t-t'\right)$$

This follows from the simple fact that "The inverse of the identity matrix is the identity matrix." With this choice for the autocorrelation function we have

$$P_{N(t)}\left(\nu\left(t\right)\right) = \exp\left[-\frac{1}{2\sigma^2}\int dt'\nu\left(t'\right)^2\right]$$

where again we have dropped the normalization factor since it cancels out of the differential equation and we are being careful to use the dummy integration varible t' in the integral.

We can now evaluate $\int \delta\nu(\cdot)\,\nu(t)\,\delta_D(x - X(t))\,P_{N(t)}(\nu(\cdot))$. We have

$$\int \delta\nu(\cdot)\,\nu(t)\,\delta_D(x - X(t))\,P_{N(t)}(\nu(\cdot)) = -\sigma^2 \int \delta\nu(\cdot)\,\delta_D(x - X(t))\,\frac{\delta}{\delta\nu(t)} \exp\left[-\frac{1}{2\sigma^2}\int dt'\nu(t')^2\right]$$

This is because $\delta/\delta\nu(t)$ is a (functional) derivative. It acts exactly the same as an ordinary derivative until you get to $\delta\nu(t')/\delta\nu(t)$ in which case you get a Dirac delta function, i.e.,

$$\frac{\delta\nu(t')}{\delta\nu(t)} = \delta_D(t - t')$$

For example

$$\begin{aligned}
\frac{\delta}{\delta\nu(t)}\int dt' w(t')\,\nu(t')^n &= \int dt' w(t')\,n\nu(t')^{n-1}\,\frac{\delta\nu(t')}{\delta\nu(t)} \\
&= \int dt' w(t')\,n\nu(t')^{n-1}\,\delta_D(t - t') \\
&= nw(t)\,\nu(t)^{n-1}
\end{aligned}$$

Integrating by parts (the Gaussian kills off the "surface" terms) gives

$$\begin{aligned}
&\int \delta\nu(\cdot)\,\nu(t)\,\delta_D(x - X(t))\,P_{N(t)}(\nu(\cdot)) \\
= \;& \sigma^2 \int \delta\nu(\cdot)\left(\frac{\delta}{\delta\nu(t)}\delta_D(x - X(t))\right)\exp\left[-\frac{1}{2\sigma^2}\int dt'\nu(t')^2\right] \\
= \;& \sigma^2 \int \delta\nu(\cdot)\left(\frac{\delta X(t)}{\delta\nu(t)}\frac{\delta}{\delta X(t)}\delta_D(x - X(t))\right)\exp\left[-\frac{1}{2\sigma^2}\int dt'\nu(t')^2\right] \\
= \;& -\sigma^2 \int \delta\nu(\cdot)\left(\frac{\delta X(t)}{\delta\nu(t)}\frac{\partial}{\partial x}\delta_D(x - X(t))\right)\exp\left[-\frac{1}{2\sigma^2}\int dt'\nu(t')^2\right] \\
= \;& -\sigma^2\frac{\partial}{\partial x}\int \delta\nu(\cdot)\left(\frac{\delta X(t)}{\delta\nu(t)}\delta_D(x - X(t))\right)\exp\left[-\frac{1}{2\sigma^2}\int dt'\nu(t')^2\right]
\end{aligned}$$

Now we have to evaluate $\delta X(t)/\delta\nu(t)$. Integrating both sides of the differential equation for $X(t)$ gives

$$X(t) = \int_0^t dt' V(X(t')) + \int_0^t dt'\nu(t')$$

Taking the functional derivative $\delta/\delta\nu(t)$ of both sides gives

$$\frac{\delta X(t)}{\delta\nu(t)} = \int_0^t dt' \frac{\delta V(X(t'))}{\delta\nu(t)} + \int_0^t dt' \frac{\delta\nu(t')}{\delta\nu(t)}$$

t is the endpoint of the integration and we are varying ν at time t, hence, due to causality, any change in $X(t')$ can occur only when $t' = t$. But the delta

function factor $\delta_D(x - X(t))$ is holding $X(t')$ fixed at the value x at $t' = t$, hence we have $\delta_D(x - X(t))\,\delta V(X(t'))/\delta\nu(t) = 0$. The second term evaluates to

$$\int_0^t dt'\delta_D(t' - t) = \frac{1}{2}$$

The 1/2 is because the delta function is centered at the endpoint t and so half the area under the delta function is outside the range of integration. (Consider representing the delta function as a Gaussian $\sqrt{2\pi\sigma^2}\exp\left[-(t'-t)^2/2\sigma^2\right]$ in the limit $\sigma \to 0$. Integrating the Gaussian from 0 to t yields 1/2 in the limit as $\sigma \to 0$.) Hence we have finally, putting all the pieces together

$$\frac{\partial}{\partial t}P_{X(t)}(x) = -\frac{\partial}{\partial x}V(x)\,P_{X(t)}(x) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}P_{X(t)}(x)$$

which is the Fokker-Planck Equation.

23 Euler Sine Formula Simple Proof

The Euler Sine Formula is the identity

$$\frac{\sin(x)}{x} = \prod_{n=1}^{\infty}\left(1 - \left(\frac{x}{n\pi}\right)^2\right)$$

This relation may look strange but note that the right hand side is zero for all $x = n\pi$ for $n =$ any nonzero positive or negative integer, it is 1 at $x = 0$ and is symmetric about $x = 0$ with only even powers of x occuring in the product which certainly matches the behavior of $\sin(x)/x$

There are no doubt multiple techniques for proving this. Here's the simplest one I know. I found I had written it in the margin of my copy of Feynman and Hibbs "Quantum Mechanics and Path Integrals" in the section on using Fourier series to solve path integrals. No doubt it is a standard proof which may even be assigned as a problem in some texts, but I haven't seen it anywhere.

Let

$$f(x) = \prod_{n=1}^{\infty}\left(1 - \left(\frac{x}{n\pi}\right)^2\right)$$

Develop a differential equation for $f(x)$ by taking a derivative w.r.t. x on both sides. It is straightforward to show

$$\partial_x f(x) = -\frac{2x}{\pi^2}\left(\sum_{n=1}^{\infty}\frac{1}{n^2 - \left(\frac{x}{\pi}\right)^2}\right)f(x)$$

Rewrite this as

$$\partial_x \ln(f(x)) = -\frac{2x}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2 - \left(\frac{x}{\pi}\right)^2}$$

Then using the "well known" result for the series (which we prove below)

$$\sum_{n=1}^{\infty}\frac{1}{n^2-\left(\frac{x}{\pi}\right)^2}=\frac{1}{2\left(\frac{x}{\pi}\right)^2}-\frac{\pi\cot\left(\pi\frac{x}{\pi}\right)}{2\frac{x}{\pi}}$$

gives directly

$$\begin{aligned}\partial_x\ln\left(f\left(x\right)\right) &= -\frac{2x}{\pi^2}\left(\frac{1}{2\left(\frac{x}{\pi}\right)^2}-\frac{\pi\cot\left(\pi\frac{x}{\pi}\right)}{2\frac{x}{\pi}}\right)\\ &= -\frac{1}{x}+\cot\left(x\right)\\ &= -\partial_x\ln\left(x\right)+\partial_x\left(\ln\left(\sin\left(x\right)\right)\right)\\ &= \partial_x\ln\left(\frac{\sin\left(x\right)}{x}\right)\end{aligned}$$

and so finally

$$f\left(x\right)=\frac{\sin\left(x\right)}{x}$$

From the original infinite product defintion, $f\left(0\right)=1$ as a boundary conditions and so we have that the integration constant is zero.

The above series can be evaluated using the technique of converting sums to integrals. The details are presented in Bernard Friedmans superb book "Lectures on Applications-Oriented Mathematics" in the chapter on complex integration. (page 153 in the 1969 edition).

Consider series of the form

$$S=\sum_{n=1}^{\infty}\frac{1}{n^2-a^2}$$

If we can find a function $g\left(z\right)$ with poles at $z=n$ (a positive integer) all with residue 1 then

$$\frac{1}{2\pi i}\int_C dz\frac{g\left(z\right)}{z^2-a^2}=S+\text{pole contribution from } z=\pm a$$

where C is a suitably chosen closed contour in the $z=x+iy$ complex plane.

Of course $\sin\left(\pi z\right)$ has zeros at $z=n$ with residue $\cos\left(\pi n\right)=\left(-1\right)^n$. Hence we can take

$$g\left(z\right)=\frac{\cos\left(\pi z\right)}{\sin\left(\pi z\right)}=\cot\left(\pi z\right)$$

Consider evaluating this on the closed square contour

$$z=\begin{cases}\pm\left(N+\frac{1}{2}\right)+iy \text{ vertical sides}\\ \pm\left(N+\frac{1}{2}\right)i+x \text{ horizontal sides}\end{cases}$$

with N a positive integer. Rewriting $g(z)$ in terms of exponentials gives

$$\begin{aligned} g(z) &= i\frac{e^{i\pi z}+e^{-i\pi z}}{e^{i\pi z}-e^{-i\pi z}} \\ &= i\frac{1+e^{-i2\pi z}}{1-e^{-i2\pi z}} \end{aligned}$$

On the left and right sides of the square contour we have

$$\begin{aligned} g\left(z=\pm\left(N+\frac{1}{2}\right)+iy\right) &= i\frac{1+e^{\mp i2\pi(N+1/2+iy)}}{1-e^{\mp i2\pi(N+1/2+iy)}} \\ &= i\frac{1-e^{\pm 2\pi y}}{1+e^{\pm 2\pi y}} \end{aligned}$$

and so

$$0 \leq \left|g\left(z=\pm\left(N+\frac{1}{2}\right)+iy\right)\right| \leq 1$$

for all N and y.

On the top and bottom sides of the square contour

$$\begin{aligned} g\left(z=\pm\left(N+\frac{1}{2}\right)i+x\right) &= i\frac{1+e^{\mp i2\pi((N+1/2)i+x)}}{1-e^{\mp i2\pi((N+1/2)i+x)}} \\ &= i\frac{1+e^{\pm 2\pi((N+1/2)\mp i2\pi x)}}{1-e^{\pm 2\pi((N+1/2)\mp i2\pi x)}} \end{aligned}$$

which gives

$$\begin{aligned} \left|g\left(z=\pm\left(N+\frac{1}{2}\right)i+x\right)\right| &= \left|\frac{1+e^{\pm 2\pi((N+1/2)\mp 2i\pi x)}}{1-e^{\pm 2\pi((N+1/2)\mp 2i\pi x)}}\right| \\ &= \left(\frac{1+2e^{\pm 2\pi(N+1/2)}\cos(2\pi x)+e^{\pm 4\pi(N+1/2)}}{1-2e^{\pm 2\pi(N+1/2)}\cos(2\pi x)+e^{\pm 4\pi(N+1/2)}}\right)^{1/2} \end{aligned}$$

For $N \to \infty$ we have

$$\left|g\left(z=\pm\left(N+\frac{1}{2}\right)i+x\right)\right| \to 1$$

for all x.

Hence the integral along the contour is, for N large, less than

$$\begin{aligned} &\int_{-(N+1/2)}^{+(N+1/2)} dx \left|\frac{1}{\left(\left(N+\frac{1}{2}\right)i+x\right)^2-a^2}\right| + \int_{-(N+1/2)}^{+(N+1/2)} dx \left|\frac{1}{\left(-\left(N+\frac{1}{2}\right)i+x\right)^2-a^2}\right| \\ &+ \int_{-(N+1/2)}^{+(N+1/2)} dy \left|\frac{1}{\left(\left(N+\frac{1}{2}\right)+iy\right)^2-a^2}\right| + \int_{-(N+1/2)}^{+(N+1/2)} dy \left|\frac{1}{\left(-\left(N+\frac{1}{2}\right)+iy\right)^2-a^2}\right| \\ \to\ & 0 \text{ as } N \to \infty \end{aligned}$$

Hence the integral yields zero for $N \to \infty$.

Noting that we have both the $z = n$ poles from $g(z)$ and the $z = \pm a$ poles from the $1/\left(z^2 - a^2\right)$ factor inside the contour for $N + 1/2 > a$, we get, for $N \to \infty$,

$$\begin{aligned} 0 &= \frac{1}{2\pi i}\int_C dz \frac{g(z)}{z^2 - a^2} \\ &= 2\sum_{n=1}^{\infty} \frac{1}{n^2 - a^2} - \frac{1}{a^2} + \frac{\pi}{a}\cot(\pi a) \end{aligned}$$

And so we have

$$\begin{aligned} S &= \sum_{n=1}^{\infty} \frac{1}{n^2 - a^2} \\ &= \frac{1}{2a^2} - \frac{\pi}{2a}\cot(\pi a) \end{aligned}$$

23.1 The Basel Problem

The Basel Problem was the name given to finding the closed form solution to the sum $\sum_{n=1}^{\infty} 1/n^2$ which is the Riemann Zeta function $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$ for $s = 2$. The problem was proposed by Pietro Mengoli in 1650 and solved by Leonhard Euler in 1734. Eulers first proof was not rigorous. Euler did develop a rigorous proof in 1741.

From above we have that

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{1}{n^2} &= \lim_{a \to 0} \sum_{n=1}^{\infty} \frac{1}{n^2 - a^2} \\ &= \lim_{a \to 0} \left(\frac{1}{2a^2} - \frac{\pi}{2a}\cot(\pi a) \right) \\ &= \frac{1}{2a^2} - \frac{\pi}{2a}\left(\frac{1}{\pi a} - \frac{\pi a}{3} + \cdots \right) \\ &= \frac{\pi^2}{6} \end{aligned}$$

24 Feynman Denominator Formula

The Feynman Denominator Formula is

$$\frac{1}{x_1 x_2 \cdots x_N} = \int_0^1 dt_1 dt_2 \cdots dt_N \delta_D \left(1 - \sum_{n=1}^{N} t_n \right) \frac{(N-1)!}{\sum_{n=1}^{N} x_n t_n}$$

Here we derive this result.

It is straightforward to write

$$\frac{1}{x_1 x_2 \cdots x_N} = \int_0^\infty ds_1 ds_2 \cdots ds_N \exp\left[-\sum_{n=1}^N x_n s_n\right]$$

assuming all the x_n have a positive real part. Since over their ranges of integration all the s_n are nonnegative we have $\sum_{n=1}^N s_n \geq 0$ from which it follows that for any combination of s_n values in the range 0 to ∞,

$$\int_0^\infty dT \delta_D\left(T - \sum_{n=1}^N s_n\right) = 1$$

One could be concerned about the case where all the $s_n = 0$ but this is nominally a "set of measure zero" and so we will ignore it. Inserting this way of writing 1 into the integral above gives

$$\frac{1}{x_1 x_2 \cdots x_N} = \int_0^\infty dT \delta_D\left(T - \sum_{n=1}^N s_n\right) \int_0^\infty ds_1 ds_2 \cdots ds_N \exp\left[-\sum_{n=1}^N x_n s_n\right]$$

Let

$$s_n = T t_n$$

and use

$$\begin{aligned}
\delta_D\left(T - \sum_{n=1}^N s_n\right) &= \delta_D\left(T\left(1 - \sum_{n=1}^N t_n\right)\right) \\
&= \frac{1}{T}\delta_D\left(1 - \sum_{n=1}^N t_n\right)
\end{aligned}$$

to get

$$\begin{aligned}
\frac{1}{x_1 x_2 \cdots x_N} &= \int_0^1 dt_1 dt_2 \cdots dt_N \delta_D\left(1 - \sum_{n=1}^N t_n\right) \int_0^\infty dT T^{N-1} \exp\left[-T\sum_{n=1}^N x_n t_n\right] \\
&= \int_0^1 dt_1 dt_2 \cdots dt_N \delta_D\left(1 - \sum_{n=1}^N t_n\right) \frac{(N-1)!}{\left(\sum_{n=1}^N x_n t_n\right)^N}
\end{aligned}$$

where the upper limit of 1 on the dt_n integrals follows from the fact that $\delta_D\left(1 - \sum_{n=1}^N t_n\right)$ indicates the sum of all the t_n values can at most be 1 otherwise the integral vanishes.

Finally, to evaluate the T integral we have used the trick

$$\begin{aligned}
\int_0^\infty dT T^p \exp\left[-\alpha T\right] &= (-\partial_\alpha)^p \int_0^\infty dT \exp\left[-\alpha T\right] \\
&= (-\partial_\alpha)^p \frac{1}{\alpha} \\
&= \frac{p!}{\alpha^{p+1}}
\end{aligned}$$

25 Density of States

25.1 Definition

The density of states is the number of quantum energy eigenstates of a given system per unit energy interval.

For discrete states $E_0, E_1, E_2, \cdots$ we have

$$\rho(E) = \sum_{n=0}^{\infty} \delta_D(E - E_n)$$

Note that δ_D always has units of 1/(units of its argument) and so $\delta_D(E - E_n)$ has units of 1/energy as required. Also for any E that is not equal to any E_n then $\rho(E) = 0$, there are no states at that energy and for $\rho(E = E_n) = \infty$ since there is one state at E_n but it is discrete and so it is spread over an energy width of 0.

For continuous states $E\left(\vec{k}\right)$ where $\vec{k}$ is continuous and represents, generically, momentum we have

$$\rho(E) = \int d^D k \delta_D\left(E - E\left(\vec{k}\right)\right)$$

It is useful to consider the case where $E\left(\vec{k}\right) \sim \vec{k}^p$ and consider how $\rho(E)$ scales with the number of space dimensions D and the power of the momentum p. Note that for the Schrodinger equation we have $p = 2$ whereas for the Dirac equation we have $p = 1$.

NOTE: In both the discrete and continuous cases complete knowledge of $\rho(E)$ implies complete knowledge of the energy eigenvalues, E_n and $E\left(\vec{k}\right)$.

Writing the integral in polar coordinates in D dimensions where Ω_D is the D-dimensional angular part gives

$$\begin{aligned}
\rho(E) &= \int d^D k \delta_D\left(E - \vec{k}^p\right) \\
&= \Omega_D \int_0^{\infty} dk k^{D-1} \delta_D\left(E - k^p\right)
\end{aligned}$$

Let $k = E^{1/p} + \eta$

$$\begin{aligned}
&= \Omega_D \int d\eta \left(E^{1/p} + \eta\right)^{D-1} \delta_D\left(E - \left(E^{1/p} + \eta\right)^p\right) \\
&= \Omega_D \int d\eta \left(E^{1/p} + \eta\right)^{D-1} \delta_D\left(E - \left(E + pE^{(p-1/p)}\eta + \cdots\right)\right) \\
&= \Omega_D \int d\eta \left(E^{1/p} + \eta\right)^{D-1} \frac{1}{pE^{(p-1)/p}} \delta_D(\eta) \\
&= \frac{\Omega_D}{p} E^{D/p-1}
\end{aligned}$$

For the Schrodinger equation $E \sim k^2$ whereas for the Dirac Equation $E \sim k$, hence we have

$$\begin{aligned}
\rho_{\text{Schrodinger}}(E) &\sim E^{D/2-1} \\
&= E^{-1/2} \text{ in 1D} \\
&= E^0 = \text{constant in 2D} \\
&= E^{1/2} \text{ in 3D}
\end{aligned}$$

and

$$\begin{aligned}
\rho_{\text{Dirac}}(E) &\sim E^{D-1} \\
&= E^0 = \text{ constant in 1D} \\
&= E^1 \text{ in 2D} \\
&= E^2 \text{ in 3D}
\end{aligned}$$

Very different scaling of the density of states in different dimensions for Schrodinger $\left(E \sim k^2\right)$ and Dirac $(E \sim k)$ type equations.

25.2 Greens Function Local Density of States

The Greens function, actually the propagator (see section on difference between the two) is given by

$$\begin{aligned}
G(\vec{r}, \vec{r}', t) &= \sum_n u_n(\vec{r})\, u_n(\vec{r}')^*\, e^{-iE_n t} \text{ Discrete Case} \\
&= \int d^D k u\left(\vec{k}, \vec{r}\right) u\left(\vec{k}, \vec{r}'\right)^* e^{-iE(\vec{k})t} \text{ Continuous Case}
\end{aligned}$$

where $u_n(\vec{r})$ and $u\left(\vec{k}, \vec{r}\right)$ are orthonormal eigenfunctions of some operator, e.g., a Hamiltonian $\hat{H}$, i.e.

$$\begin{aligned}
\hat{H} u_n(\vec{r}) &= E_n u_n(\vec{r}) \text{ Discrete Case} \\
\hat{H} u\left(\vec{k}, \vec{r}\right) &= E\left(\vec{k}\right) u\left(\vec{k}, \vec{r}\right) \text{ Continuous Case}
\end{aligned}$$

Define local density of states $\rho(\vec{r}, E)$ for the discrete case.

$$\begin{aligned}
\rho(\vec{r}, E) &= \int_{-\infty}^{+\infty} \frac{dt}{2\pi} e^{iEt} G(\vec{r}, \vec{r}, t) \\
&= \int_{-\infty}^{+\infty} \frac{dt}{2\pi} e^{iEt} \sum_n u_n(\vec{r})\, u_n(\vec{r})^*\, e^{-iE_n t} \\
&= \sum_n u_n(\vec{r})\, u_n(\vec{r})^* \int_{-\infty}^{+\infty} \frac{dt}{2\pi} e^{i(E-E_n)t} \\
&= \sum_n u_n(\vec{r})\, u_n(\vec{r})^*\, \delta_D(E - E_n)
\end{aligned}$$

then since

$$\int d^D r u_n(\vec{r}) u_n(\vec{r})^* = 1$$

we have

$$\rho(E) = \int d^D r \rho(\vec{r}, E)$$

This requires a little more thought for the continuous case since

$$\int d^D r u\left(\vec{k}, \vec{r}\right) u\left(\vec{k}, \vec{r}\right)^* = \delta_D\left(\vec{k} - \vec{k}\right) = \delta_D(0) = \infty.$$

26 Partition Function, Equipartition Theorem

26.1 Partition Function

If a system (think particles in a box) is in thermal equilibrium, then the probability $P(E)$ of it having total energy E (think sum of all the particle energies, kinetic and potential) is proportional to $\exp[-E/k_B T]$ where k_B is Boltzmanns constant and T is the temperature in Kelvin, i.e.,

$$P(E) \sim \exp\left[-\frac{E}{k_B T}\right]$$

$\exp[-E/k_B T]$ is often referred to as the Boltzmann "probability" or factor.

Why $\exp[-E/k_B T]$? There is a lovely derivation, actually only a justification, of this in a footnote in the first few pages of Feynmans book on statistical mechanics. It goes more or less like this.

Assume the only relevant macroscopic quantity of the system on which the probability can depend is its total energy E and that the total momentum, angular momentum, charge (of any type) are all on average irrrelevant or all on average zero. Here "on average" means the average over a time scale which is long compared to the time scale of (thermal) fluctuations of the system. Since energy is only defined up to an additive constant it follows that the ratio of probabilities for the system to have two different energies E_1 and E_2 satisfies

$$\frac{P(E_2)}{P(E_1)} = \frac{P(E_2 + E_{arb})}{P(E_1 + E_{arb})}$$

where E_{arb} here is an arbitrary value of energy. That is this relation must hold for all values of E_{arb}. The only function that satisfies this requirement is $P(E) \sim a^{bE}$, where a and b are constants. Without loss of generality (w.l.o.g.) this may be rewritten as $P(E) \sim \exp[-cE]$ where c is a constant. The minus sign and the fact that $c = 1/k_B T$ all follow from physics, that is, comparing this exponential form to the rules of thermodynamics. This is the argument used by Maxwell to derive the Maxwell-Boltzmann distribution.

NOTE: If the system is perfectly isolated then, via energy conservation, its total energy must be fixed at some value, say, E_0 in which case $P(E) =$

$\delta_D(E - E_0)$. This distribution is known as the "microcanonical ensemble". The result $P(E) \sim \exp[-E/k_BT]$ is known as the "canonical ensemble" and in this case the system must be considered to be in contact with a nominally much larger system known as the "heat bath" with which it can exchange energy but not "particles". If the system can exchange both "particles" and energy with the heat bath then $P(E)$ is generalized to $P(E,N) \sim \exp[-(E-\mu N)/k_BT]$ where μ is the so called "chemical potential" and N is the total number of particles in the system itself. This is known as the "grand canonical ensemble".

To be a true probability $P(E)$ must be properly normalized.

If the possible values of E are discrete, i.e., E can take any of the values E_n with $n = 1, 2, \cdots$, then the normalization factor is

$$\begin{aligned} Z &= \sum_n e^{-E_n/k_BT} : \text{canonical ensemble} \\ Z &= \sum_{n,N} e^{-(E_n-\mu N)/k_BT} : \text{grand canonical ensemble} \end{aligned}$$

The sum on N means: sum over all possible numbers of particles in the system from 0 to ∞.

If E is continuous and depends on a combination of "degrees of freedom", X_n, e.g., momenta, position, angular momenta, angular position, etc., as $E(X_1, X_2, \cdots)$ then

$$\begin{aligned} Z &= \int dX_1 dX_2 \cdots e^{-E(X_1,X_2,\cdots)/k_BT} : \text{canonical ensemble} \\ Z &= \sum_N \int dX_1 dX_2 \cdots e^{-(E(X_1,X_2,\cdots)-\mu N)/k_BT} : \text{grand canonical ensemble} \end{aligned}$$

where the number of $X's$ generally varies in a deterministic way with the number of "particles" N.

In both cases Z is called the "partition function" but it is just the normalization factor and so

$$P(E_n) \equiv P_n = \frac{\exp[-E_n/k_BT]}{Z}$$

The partition function contains all of thermodynamics since it acts as a generating function, that is, derivatives of the partition function with respect to various parameters yield different thermodynamic quantities. To see this define the Helmholtz Free energy F as

$$\begin{aligned} F &= -k_BT \ln(Z) \\ &= -k_BT \ln\left(\sum_n \exp[-E_n/k_BT]\right) \end{aligned}$$

Entropy is defined as

$$S = -k_B \sum_n P_n \ln(P_n)$$

Taking the derivative of F with respect to the explicit T dependence of F yields

$$\frac{dF}{dT} = -k_B \ln(Z) - k_B T \frac{1}{Z} \frac{dZ}{dT}$$

but

$$\begin{aligned}
\frac{dZ}{dT} &= \frac{d}{dT} \sum_n \exp\left[-\frac{E_n}{k_B T}\right] \\
&= \sum_n \left(\frac{E_n}{k_B T^2} \exp\left[-\frac{E_n}{k_B T}\right] \right) \\
&= \frac{Z}{k_B T^2} \sum_n E_n P_n \\
&= \frac{Z}{k_B T^2} U
\end{aligned}$$

where

$$\langle E \rangle = \sum_n E_n P_n \equiv U$$

is, by definition, the average energy of the system.

$$\begin{aligned}
\frac{dF}{dT} &= -k_B \ln(Z) - k_B T \frac{1}{Z} \frac{Z}{k_B T^2} \sum_n E_n P_n \\
&= -k_B \left(\ln(Z) + \sum_n \frac{E_n}{k_B T} P_n \right)
\end{aligned}$$

Multiplying the $\ln(Z)$ term by 1 in the form $1 = \sum_n P_n = \sum_n \exp\left[-E_n/k_B T\right]/Z$ gives

$$\begin{aligned}
\frac{dF}{dT} &= -k_B \sum_n P_n \left(\ln(Z) + \frac{E_n}{k_B T} \right) \\
&= -k_B \sum_n P_n \left(-\ln\left(\frac{\exp\left[-E_n/k_B T\right]}{Z} \right) \right) \\
&= k_B \sum_n P_n \ln(P_n) \\
&= -S
\end{aligned}$$

and so we have

$$S = -\frac{dF}{dT}$$

where the derivative d/dT does not include any T dependence of the E_n. Such dependence might come from varying the system volume with the temperature. Varying E_n with T is excluded in the above result and (d/dT) is then often written using the notation$(d/dT)_V$ which means the volume is held constant while varying T.

Note it also follows from the above that

$$\begin{aligned} S &= k_B \ln(Z) + k_B T \frac{1}{Z}\frac{dZ}{dT} \\ &= -\frac{F}{T} + \frac{U}{T} \\ &\text{which gives} \\ F &= U - TS \end{aligned}$$

The heat capacity is the change in average energy with temperature

$$C_V = \frac{dU}{dT}$$

where the subscript V on C indicates holding the volume constant or equivalently taking derivatives only with respect to (w.r.t.) the explicit T dependence and so

$$C_V = -\frac{d^2F}{dT^2}$$

Energy is force times distance and so in general force is the change in energy with some distance or length. Pressure p is force per area on the system and so pressure is the negative of the change in energy of the system w.r.t. a change in volume while holding the P_n fixed, i.e.,

$$\begin{aligned} p &= -\sum_n \left(\frac{dE_n}{dV}\right) P_n \\ &= k_B T \sum_n \left(-\frac{1}{k_B T}\frac{dE_n}{dV}\right) \frac{\exp\left[-E_n/k_B T\right]}{\sum_m \exp\left[-E_m/k_B T\right]} \\ &= k_B T \sum_n \left(-\frac{1}{k_B T}\frac{dE_n}{dV}\right) \frac{\exp\left[-E_n/k_B T\right]}{Z} \\ &= -k_B T \left(\frac{d\ln(Z)}{dV}\right)_T \\ &= \left(\frac{dF}{dV}\right)_T \end{aligned}$$

where the subscript T means the temperature is held constant.

26.2 Equipartition Theorem

Consider the case where the energy depends continuously on some set of degrees of freedom, $X_1, X_2, \cdots$, i.e., $E_n \to E(X_1, X_2, \cdots)$. For example the X_n might be the phase space coordinates of N particles in 3 dimensions in which case there are $6N$ different X variables,

$$X_1, X_2, \cdots \to p_{n,x}, p_{n,y}, p_{n,z}, q_{n,x}, q_{n,y}, q_{n,z} \text{ for } n = 1, 2, \cdots, N$$

Here the $p_{n,x}, p_{n,y}, p_{n,z}$ are the x, y, z components of the "momentum" of particle n and $q_{n,x}, q_{n,y}, q_{n,z}$ are its "position" coordinates.

In terms of the X_n with $n = 1, 2, \cdots, N$ we have

$$P\left(X_1, X_2, \cdots\right) = \frac{\exp\left[-E\left(X_1, X_2, \cdots\right)/k_B T\right]}{\int d^N X \exp\left[-E\left(X_1, X_2, \cdots\right)/k_B T\right]}$$

with

$$Z = \int d^N X \exp\left[-E\left(X_1, X_2, \cdots\right)/k_B T\right]$$

being the partition function.

Suppose $E\left(X_1, X_2, \cdots\right)$ has the form

$$E\left(X_1, X_2, \cdots\right) = c_1 X_1^{p_1} + c_2 X_2^{p_2} + \cdots$$

that is $E\left(X_1, X_2, \cdots\right)$ is a sum of the X_n raised to powers $p_1, p_2, ...$ multiplied by constants $c_1, c_2, ...$ Then the expectation value of any one particular term in the sum,

$$E_n = c_n X_n^{p_n}$$

for a given n is given by

$$\begin{aligned}
\langle E_n \rangle &= \langle c_n X_n^{p_n} \rangle \\
&= \frac{1}{p_n} \left\langle X_n \frac{dE\left(X_1, X_2, \cdots\right)}{dX_n} \right\rangle \\
&= \frac{1}{p_n} \int d^N X X_n \frac{dE\left(X_1, X_2, \cdots\right)}{dX_n} P\left(X_1, X_2, \cdots\right) \\
&= \frac{1}{p_n} \frac{\int d^N X X_n \frac{dE(X_1, X_2, \cdots)}{dX_n} \exp\left[-E\left(X_1, X_2, \cdots\right)/k_B T\right]}{\int d^N X \exp\left[-E\left(X_1, X_2, \cdots\right)/k_B T\right]} \\
&= -\frac{1}{p_n} k_B T \frac{\int d^N X X_n \left(\frac{d}{dX_n} \exp\left[-E\left(X_1, X_2, \cdots\right)/k_B T\right]\right)}{\int d^N X \exp\left[-E\left(X_1, X_2, \cdots\right)/k_B T\right]} \\
&\quad \text{Integrate by parts, assume surface terms vanish} \\
&= \frac{1}{p_n} k_B T \frac{\int d^N X \exp\left[-E\left(X_1, X_2, \cdots\right)/k_B T\right]}{\int d^N X \exp\left[-E\left(X_1, X_2, \cdots\right)/k_B T\right]} \\
&= \frac{1}{p_n} k_B T
\end{aligned}$$

If $p_n = 2$ then we have

$$\langle E_n \rangle = \frac{1}{2} k_B T$$

Hence each degree of freedom which contributes quadratically in an additive fashion to the total energy carries, on average, $\frac{1}{2} k_B T$ of thermal energy. This is the standard statement of the Equipartition Theorem. Note that the higher the power dependence, i.e., the larger the value of p_n, the less thermal energy is stored, on average, in that degree of freedom. Interesting.

27 Shot and Thermal (Johnson) Noise

Shot noise corresponds to any process with discrete random events: raindrops hitting a roof, photons being absorbed one at a time, electrons moving through a circuit, the hissing sound of bacon frying, etc. Shot noise is almost always associated with Poisson statistics. In fact, if the probability of receiving one "count" of something in an infinitesimal time interval dt is given by pdt then this leads directly to Poisson statistics as shown in a previous section.

Thermal (or Johnson) Noise on the other hand comes from fluctuations which occur when a system is at a finite (nonzero) temperature.

For example consider a capacitor, unwired to anything, at a finite temperature T (in Kelvin). The energy stored in a capacitor of capacitance C with voltage V_{C} between the plates is, as shown in a previous section, given by

$$E = \frac{1}{2}CV_{\mathrm{C}}^2 = \frac{Q^2}{2C}$$

where Q is the (instantaneous) charge on the capacitor. At thermal equilibrium there is, via Boltzmann equipartition of energy, $\frac{1}{2}k_BT$ of thermal energy in each degree of freedom of a system where k_B is Boltzmann's constant. Assuming V counts as 1 degree of freedom, setting $E = \frac{1}{2}k_BT$ yields

$$CV_{\mathrm{C}}^2 = k_BT \rightarrow V_{\mathrm{C}} = \sqrt{\frac{k_BT}{C}}$$

Note that V is the rms or root mean square voltage fluctuation across the capacitor. The average of voltage across the capacitor, measured over a long period of time is zero.

For a resistor, consider the resistor (with resistance R) being wired in a single loop circuit in series with the capacitor (C) but not wired to anything else. In this configuration the voltage across the resistor is instantaneously equal to the voltage across the capacitor (neglecting speed of light effects) and so we have

$$V_{\mathrm{R}} = V_{\mathrm{C}} = \sqrt{\frac{k_BT}{C}}$$

But the time constant of an resistor-capacitor or RC circuit is given by

$$\tau = RC$$

and so, the rms voltage across the resistor is

$$V_{\mathrm{R}} = \sqrt{\frac{k_BRT}{\tau}}$$

or

$$V_{\mathrm{R}}^2 = k_BRT\left(\frac{1}{\tau}\right)$$

The $1/\tau$ factor can be thought of as a frequency interval Δf in Hertz.

The fact that k_B, T, and R are all (nominally) frequency independent means that up to some very high frequency where other physical effects come into play, such as the speed of light, the mean square voltage is the same at all frequencies, i.e., k_BRT is frequency independent. But, an RC circuit is a low pass filter and so this uniform frequency distribution of squared voltage must be filtered to get the actual thermally induced mean voltage across the resistor if it were unconnected to anything. This filtering leads to the "famous" factor of 4, i.e. $V_{\mathrm{R}} = \sqrt{4k_BRT}$ as we now show.

To see this consider the thermally generated voltage fluctuations to be coming from a voltage source or electromotive force (commonly referred to as an e.m.f.) and treat the resistor as ideal, solve for the voltage across the capacitor, which we know from above, in terms of this e.m.f, then solve for the (mean square) e.m.f.

In an RC circuit, with RC in series, and with a voltage souce $V_{\mathrm{S}}(\omega)$ where $\omega = 2\pi f$ is the radian frequency with f the frequency in Hertz, the current $I(\omega)$ is given by

$$I(\omega) = \frac{V_{\mathrm{S}}(\omega)}{R + \frac{1}{i\omega C}}$$

The voltage across the capacitor $V_{\mathrm{C}}(\omega)$ is

$$V_{\mathrm{C}}(\omega) = \frac{I(\omega)}{i\omega C} = \frac{V_{\mathrm{S}}(\omega)}{1 + i\omega RC}$$

which gives the mean (absolute) square voltage

$$\left|V_{\mathrm{C}}(\omega)\right|^2 = \frac{\left|V_{\mathrm{S}}(\omega)\right|^2}{1 + (\omega RC)^2}$$

NOTE: We have to use the absolute value squared since we are using i to account for the 90 degree phase difference between the voltage across the capacitor and the voltage across the resistor. This would not be necessary if the phase was handled explicitly in terms of sines and cosines. Note also that the low pass aspect is obvious from the $\omega^2 = (2\pi f)^2$ dependence in the denominator, as the frequency f or equivalently the radian frequency ω increases the amount of voltage transferred from the source to the capacitor decreases and so, more-or-less, only the low frequency components of the source voltage are transferred to the capacitor.

To get the total voltage across the capacitor we need to integrate over f from 0 to ∞. This gives

$$V_{\mathrm{C}}^2 = \frac{k_BT}{C} = \int_0^\infty df \frac{\left|V_{\mathrm{S}}(\omega)\right|^2}{1 + (\omega RC)^2} = \frac{1}{2\pi}\int_0^\infty d\omega \frac{\left|V_{\mathrm{S}}(\omega)\right|^2}{1 + (\omega RC)^2}$$

But we know from above that the thermally induced square voltage,

$$\left|V_{\mathrm{source}}(\omega)\right|^2 \tilde{} \; k_BRT$$

is frequency independent, hence we have, using $V_C^2 = k_B T/C$

$$\begin{aligned}
\frac{k_B T}{C} &= \frac{|V_S|^2}{2\pi}\int_0^\infty d\omega \frac{1}{1+(\omega RC)^2} \\
&= \frac{|V_S|^2}{2\pi RC}\int_0^\infty d(\omega RC)\frac{1}{1+(\omega RC)^2} \\
&= \frac{|V_S|^2}{2\pi RC}\int_0^\infty dz\frac{1}{1+z^2} \\
&= \frac{|V_S|^2}{2\pi RC}\frac{1}{2}\int_{-\infty}^{+\infty} dz\frac{1}{1+z^2} \\
&= \frac{|V_S|^2}{2\pi RC}\frac{1}{2}\frac{2\pi i}{2i} \\
&= \frac{|V_S|^2}{4RC}
\end{aligned}$$

or

$$|V_S|^2 = 4k_B RT$$

The actual squared source voltage needs to be 4 times larger account for the low pass filtering of the RC circuit. Hence the thermally induced mean square voltage fluctuation across a resistor at temperature T is given by $4k_B RT$.

28 Euler-McLauren Formula

The Euler-McLauren formula is a relation between sums and integrals. It is used to improve the accuracy of the numerical result when computing integrals as sums. Define $F(N)$ as the sum

$$F(N) \equiv f(0) + \cdots + f(N) = \sum_{n=0}^{N} f(n)$$

Here $f(x)$ is assumed to be smooth, well behaved, etc.

From the definition of $F(N)$ we have

$$F(N+1) - F(N) = f(N+1)$$

But we have from Taylor series

$$\begin{aligned}
f(x+a) &= f(x) + a\partial_x f(x) + \frac{a^2}{2}\partial_x^2 f(x) + \cdots \\
&= \sum_{n=0}^{\infty}\frac{(a\partial_x)^n}{n!} f(x) \\
&= \exp[a\partial_x] f(x)
\end{aligned}$$

hence

$$F(x+1) = \exp[\partial_x] F(x)$$
$$f(x+1) = \exp[\partial_x] f(x)$$

Substituting this above with $a = 1$ we get

$$\begin{aligned} 0 &= F(N+1) - F(N) - f(N+1) \\ &= \left[(\exp[\partial_x] - 1) F(x) - \exp[\partial_x] f(x) \right]_{x=N} \\ &= F(N) - \left[\frac{\exp[\partial_x]}{\exp[\partial_x] - 1} f(x) \right]_{x=N} \end{aligned}$$

Rearranging gives

$$\begin{aligned} F(N) &= \left[\frac{\exp[\partial_x]}{\exp[\partial_x] - 1} f(x) \right]_{x=N} \\ &= \left[\left(1 + \frac{1}{\exp[\partial_x] - 1} \right) f(x) \right]_{x=N} \\ &= f(N) + \left[\frac{1}{\exp[\partial_x] - 1} f(x) \right]_{x=N} \end{aligned}$$

Letting ∂_x^{-1} indicate integration from 0 to x with an additive constant, c, we can write

$$\begin{aligned} F(N) &= f(N) + \left[\frac{\partial_x}{\exp[\partial_x] - 1} \partial_x^{-1} f(x) \right]_{x=N} \\ &= f(N) + \left[\frac{\partial_x}{\exp[\partial_x] - 1} \left(\int_0^x dx f(x) + c \right) \right]_{x=N} \end{aligned}$$

Now noting that

$$\frac{t}{\exp[t] - 1} = \sum_{n=0}^{\infty} \frac{b_n}{n!} t^n$$

where b_n are the Bernoulli numbers, we have

$$\begin{aligned} F(N) &= f(N) + \sum_{n=0}^{\infty} \frac{b_n}{n!} \left[\partial_x^n \left(\int_0^x dx f(x) + c \right) \right]_{x=N} \\ &= f(N) + b_0 \left(\int_0^N dx f(x) + c \right) + \sum_{n=1}^{\infty} \frac{b_n}{n!} \left[\partial_x^n \left(\int_0^x dx f(x) + c \right) \right]_{x=N} \\ &= f(N) + \int_0^N dx f(x) + c + b_1 f(N) + \frac{b_2}{2} (\partial_x f)(N) \\ &+ \frac{b_3}{3!} \left(\partial_x^2 f \right)(N) + \frac{b_4}{4!} \left(\partial_x^3 f \right)(N) + \cdots \end{aligned}$$

Here we have used the fact that $b_0 = 1$ and c must be chosen so that $F(0) = f(0)$ which gives

$$c = -b_1 f(0) - \frac{b_2}{2}(\partial_x f)(0) - \frac{b_3}{3!}\left(\partial_x^2 f\right)(0) - \frac{b_4}{4!}\left(\partial_x^3 f\right)(0) - \cdots$$

Using the Benoulli number values $b_1 = -1/2, b_2 = 1/6, b_3 = 0, b_4 = -1/30, b_5 = 0, \cdots$ gives

$$\begin{aligned} F(N) &= f(N) + \int_0^N dx f(x) - \frac{1}{2}(f(N) - f(0)) + \frac{1}{12}((\partial_x f)(N) - (\partial_x f)(0)) \\ &\quad - \frac{1}{30}\frac{1}{4!}\left(\left(\partial_x^3 f\right)(N) - \left(\partial_x^3 f\right)(0)\right) + \cdots \\ &= \frac{f(N) + f(0)}{2} + \int_0^N dx f(x) + \frac{1}{12}((\partial_x f)(N) - (\partial_x f)(0)) \\ &\quad - \frac{1}{30}\frac{1}{4!}\left(\left(\partial_x^3 f\right)(N) - \left(\partial_x^3 f\right)(0)\right) + \cdots \end{aligned}$$

This result is the the Euler-Maclaurin formula or series. This is, with a small difference, the derivation given on page 132 of Bernard Friedman's wonderful little book, "Lectures on Applications Oriented Mathematics", once Friedmans different definition of the Bernoulli numbers is accounted for.

Rearranging gives

$$\int_0^N dx f(x) = \sum_{n=0}^N f(n) - \frac{f(N) + f(0)}{2} - \frac{1}{12}((\partial_x f)(N) - (\partial_x f)(0)) + \cdots$$

This shows how to easily "correct" discrete sums to be closer to the true integral value: The dominant correction amounts to simply subtracting the mean or average of the starting and ending values of $f(x)$.

29 Embedding, Induced Metric, Curvature

29.1 Whitney Embedding Theorem

We begin by stating, but not proving the (weak) Whitney Embedding Theorem:

> Any smooth Riemannian manifold of dimension d can be embedded, i.e. mapped in a one-to-one smooth non-selfintersecting way, into a Euclidean space of dimension $2d + 1$.

Examples: A one dimensional manifold, i.e., a path or trajectory, can be embedded in 3D with no self intersections. If embedded in 2D it will generally have self intersections. Think of projecting a knot in a piece of string onto a sheet of paper. The surface of a sphere is a 2D manifold as is the surface of a torus. Both can be embedded in 3D with no self intersections. The surface of finite length cylinder with any (non-self-intersecting) cross sectional area can be

embedded in 2D. Bending it in a circle and connecting the ends makes a torus. But if the ends are connected "from the inside", similar to the "twist" that makes a Mobius strip, then you get a Klein bottle and the Klein bottle cannot be embedded in 3D without self intersections. It requires 4D. So, depending on the manifold, sometimes fewer than $2d+1$ dimensions are required. The (weak) Whitney Embedding theorem guarantees you can always make it work in $2d+1$, it's finding a map for any given Riemannian manifold which has no self-intersections and preserves any given metric on the manifold that is the hard part.

29.2 Induced Metric

Consider a 2D surface embedded in standard "x, y, z" 3D Euclidean space. Given any small enough subarea of the surface it can be represented as

$$z = H(x, y)$$

Note that in general $H(x, y)$ will be definable only over a finite area in the x, y plane.

With this definition position on the surface in 3D is given by

$$\vec{r}(x, y) = (x, y, H(x, y))$$

Note that we are using x and y here to label position along or "in" the surface in 2D. Also note that $\vec{r}(x, y)$ "lives in 3D" but depends only on x and y.

Infinitesimal displacements along or tangent to the surface in 3D are given by

$$\begin{aligned} d\vec{r} &= (dx, dy, \partial_x H dx + \partial_y H dy) \\ &\equiv \left(dx, dy, \left(d\vec{\rho} \cdot \vec{\partial}_\perp\right) H\right) \\ &= d\rho \left(\hat{\rho}_x, \hat{\rho}_y, \left(\hat{\rho} \cdot \vec{\partial}_\perp\right) H\right) \end{aligned}$$

with $d\vec{\rho} = (dx, dy)$, $\hat{\rho} = d\vec{\rho}/d\rho \equiv \left(\hat{\rho}_x, \hat{\rho}_y\right)$, $d\rho = \sqrt{dx^2 + dy^2}$,and $\vec{\partial}_\perp = (\partial_x, \partial_y)$. With these definitions we have $\hat{\rho}_x = \cos(\phi)$ and $\hat{\rho}_y = \sin(\phi)$ where ϕ is the angle between $\hat{\rho}$ and the x axis.Thus the tangent vector in the $d\vec{\rho}$ direction is given

by

$$\begin{aligned}
\hat{t}\left(\hat{\rho}\right) &= \frac{d\vec{r}}{\left|d\vec{r}\right|} \\
&= \frac{\left(dx, dy, \left(d\vec{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)}{\left|\left(dx, dy, \left(d\vec{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)\right|} \\
&= \frac{\left(\hat{\rho}_x, \hat{\rho}_y, \left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)}{\left|\left(\hat{\rho}_x, \hat{\rho}_y, \left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)\right|} \\
&= \frac{\left(\hat{\rho}_x, \hat{\rho}_y, \left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)}{\sqrt{1+\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)^2}}
\end{aligned}$$

NOTE: Obviously $\hat{t}$ depends on the choice of $\hat{\rho}$ (as well as on x and y). The $\hat{\rho}$ dependence is indicated explicitly by writing $\hat{t}\left(\hat{\rho}\right)$.

An infinitesimal distance ds along or "in" the surface in 3D is the square root of

$$\begin{aligned}
ds^2 &= d\vec{r}\cdot d\vec{r} \\
&= \left(dx, dy, d\vec{\rho}\cdot\vec{\partial}_{\perp}H\right)\cdot\left(dx, dy, d\vec{\rho}\cdot\vec{\partial}_{\perp}H\right) \\
&= dx^2 + dy^2 + \left(dx\partial_x H + dy\partial_y H\right)^2 \\
&= \left(1+\left(\partial_x H\right)^2\right)dx^2 + \left(1+\left(\partial_y H\right)^2\right)dy^2 + 2\left(\partial_x H\right)\left(\partial_y H\right)dxdy
\end{aligned}$$

If ξ_i with $i = 1, \cdots, N$ label the positions "in" a manifold of dimension N then the standard form for ds^2 in terms of the metric $g_{i,j}$ is

$$\begin{aligned}
ds^2 &= \sum_{i,j=1}^{N} g_{i,j}d\xi_i d\xi_j \\
&\equiv g_{i,j}d\xi_i d\xi_j
\end{aligned}$$

where the second line follows from invoking (yet again) the Einstein summation convention. Note that, in N dimensions, the $g_{i,j}$ are the i,j elements of an $N\times N$ matrix or tensor g.

Comparing the two forms of ds^2 above gives, for our 2D surface, the induced metric

$$g\left(x,y\right) = \left[\begin{array}{cc} 1+\left(\partial_x H\right)^2 & \left(\partial_x H\right)\left(\partial_y H\right) \\ \left(\partial_x H\right)\left(\partial_y H\right) & 1+\left(\partial_y H\right)^2 \end{array}\right]$$

We say induced because the metric above depends on how we embed the 2D surface in 3D, i.e., on our choice of $H\left(x,y\right)$. If the 2D manifold had a pre-specified or given metric $g_0\left(\xi_1,\xi_2\right)$ then as long we pick $H\left(x,y\right)$ such that $\left(g_0\left(\xi_1 = x, \xi_2 = y\right)\right) = g\left(x,y\right)$ for all x and y, then the induced metric is the

same as the given metric. Note that this type of relation between the induced and given metric holds in any number of dimensions.

With the matrix g as written above and with "$\cdot$" indicating matrix multiplication, we get

$$\begin{aligned} ds^2 &= (dx, dy) \cdot g \cdot \begin{pmatrix} dx \\ dy \end{pmatrix} \\ &= (dx, dy) \cdot \begin{bmatrix} 1 + (\partial_x H)^2 & (\partial_x H)(\partial_y H) \\ (\partial_x H)(\partial_y H) & 1 + (\partial_y H)^2 \end{bmatrix} \cdot \begin{pmatrix} dx \\ dy \end{pmatrix} \\ &= \left(1 + (\partial_x H)^2\right) dx^2 + \left(1 + (\partial_y H)^2\right) dy^2 + 2(\partial_x H)(\partial_y H)\, dxdy \end{aligned}$$

as we should.

Because g is symmetric it can be diagonalized by changing the x, y coordinate system. One choice is to define coordinates u and v by $z = H(x, y) = u =$ constant, in which case constant u curves in the x, y plane correspond to constant z values on the surface $H(x, y)$. Given this definition $d\vec{u}$ at everypoint in the x, y plane lies in the direction of the gradient of $H(x, y)$, i.e.,

$$\begin{aligned} d\vec{u} &= du \frac{\left(\vec{\partial}_\perp H\right)}{\left|\left(\vec{\partial}_\perp H\right)\right|} \\ &= du \frac{(\partial_x H, \partial_y H)}{\sqrt{(\partial_x H)^2 + (\partial_y H)^2}} \end{aligned}$$

Then $d\vec{v}$ can be taken to be everywhere perpendicular to $d\vec{u}$

$$d\vec{v} = dv \frac{(-\partial_y H, \partial_x H)}{\sqrt{(\partial_x H)^2 + (\partial_y H)^2}}$$

so that we have

$$d\vec{u} \cdot d\vec{v} = 0$$

with v corresponding to position along each constant u curve.

In the u, v coordinate system we have, from Pythagoras,

$$ds^2 = \left(1 + \left(\vec{\partial}_\perp H\right)^2\right) du^2 + dv^2$$

and so

$$g(u, v) = \begin{bmatrix} 1 + \left(\vec{\partial}_\perp H\right)^2 & 0 \\ 0 & 1 \end{bmatrix}$$

Of course finding the explicit forms of $x(u, v)$ and $y(u, v)$ requires solving (simultaneously)

$$\begin{aligned} u &= H(x(u, v), y(u, v)) \\ 0 &= \frac{\partial H(x(u, v), y(u, v))}{\partial v} \end{aligned}$$

29.3 Curvature (Extrinsic)

We begin by constructing the unit normal = unit vector everywhere perpendicular to the surface.

Unit Normal: Consider the function

$$F(x,y,z) = z - H(x,y)$$

The 3D gradient of F points, in 3D, in the direction of the maximum rate of change of F. This is easy to see. The change in F, dF, in the direction (in 3D) $d\vec{r} = (dx, dy, dz)$ is given by

$$dF = d\vec{r} \cdot \vec{\partial} F$$

where $\vec{\partial} = (\partial_x, \partial_y, \partial_z)$. But

$$d\vec{r} \cdot \vec{\partial} F = dr \left| \vec{\partial} F \right| \cos(\theta)$$

where $dr = |d\vec{r}| = \sqrt{dx^2 + dy^2 + dz^2}$, $\left|\vec{\partial} F\right| = \sqrt{(\partial_x F)^2 + (\partial_y F)^2 + (\partial_z F)^2}$ and θ is the angle between $d\vec{r}$ and $\vec{\partial} F$. Hence for a given dr, dF is maximum for $\theta = 0$, that is $d\vec{r}$ parallel to $\vec{\partial} F$. Any direction perpendicular to $\vec{\partial} F$, i.e., $\theta = \pi/2$ has, by definition $dF = 0$ and so is tangent to the surface $z = H(x,y)$. Thus we have that the normal (= unit vector in 3D perpendicular to the surface at every point) is given by

$$\hat{n}(x,y) = \frac{\vec{\partial} F}{\left|\vec{\partial} F\right|} = \frac{(-\partial_x H, -\partial_y H, 1)}{\sqrt{1 + \left(\vec{\partial}_\perp H\right)^2}}$$

Note that $\hat{n}(x,y)$ "lives in 3D"„ i.e., it has x, y and z components, but depends only on the 2D coordinates x and y.

Tangent Vectors: Since

$$\hat{n}(x,y) \cdot \hat{n}(x,y) = 1 \text{ for all } x, y$$

we have

$$\left(d\vec{\rho} \cdot \vec{\partial}_\perp\right) (\hat{n}(x,y) \cdot \hat{n}(x,y)) = 0$$

or using $\left(d\vec{\rho} \cdot \vec{\partial}_\perp\right) = d\rho \left(\hat{\rho} \cdot \vec{\partial}_\perp\right)$ and dividing by $d\rho$

$$\left(\left(\hat{\rho} \cdot \vec{\partial}_\perp\right) \hat{n}(x,y)\right) \cdot \hat{n}(x,y) = 0$$

which means either $\left(\left(\hat{\rho} \cdot \vec{\partial}_\perp\right) \hat{n}(x,y)\right) = 0$ or $\left(\hat{\rho} \cdot \vec{\partial}_\perp\right) \hat{n}(x,y)$ is tangent to the surface in 3D. Hence, if it doesn't vanish, $\left(\left(\hat{\rho} \cdot \vec{\partial}_\perp\right) \hat{n}(x,y)\right)$ will be parallel or tangent to $d\vec{r}$ for some value of $\hat{\rho} = (dx/d\rho, dy/d\rho)$, i.e., it must be proportional to $\hat{t}(\hat{\rho})$ defined above for some value of $\hat{\rho}$.

Explicitly, taking the inner or "dot" product of $\hat{n}(x,y)$ with $d\vec{r} = (dx, dy, (\partial_x H dx + \partial_y H dy)) = d\rho\hat{t}(\hat{\rho})$ and $\hat{n}$ gives

$$\begin{aligned} d\vec{r}\cdot\hat{n} &= (dx, dy, (\partial_x H dx + \partial_y H dy)) \cdot \frac{(-\partial_x H, -\partial_y H, 1)}{\sqrt{1 + \left(\vec{\partial}_\perp H\right)^2}} \\ &= \frac{-\partial_x H dx - \partial_y H dy + \partial_x H dx + \partial_y H dy}{\sqrt{1 + \left(\vec{\partial}_\perp H\right)^2}} \\ &= 0 \end{aligned}$$

which shows explicity that $\hat{n}$ is normal to the surface.

A similar result follows for $\hat{t}(\hat{\rho})$ since it has unit length everywhwere, i.e.,

$$\hat{t}(\hat{\rho}) \cdot \hat{t}(\hat{\rho}) = 1 \text{ for all } x, y$$

and so

$$\left(\left(\hat{\rho}' \cdot \vec{\partial}_\perp\right)\hat{t}(\hat{\rho})\right) \cdot \hat{t}(\hat{\rho}) = 0$$

for any $\hat{\rho}'$. Hence, if $\left(d\vec{\rho}' \cdot \vec{\partial}_\perp\right)\hat{t}(\hat{\rho})$ doesn't vanish then it points normal or perpendicular to the surface, i.e., parallel or antiparallel to $\hat{n}(x,y)$.

Let's evaluate $\left(\hat{\rho}\cdot\vec{\partial}_\perp\right)\hat{n}$ explicitly. To simplify notation define

$$\begin{aligned} &\text{Notation:} \\ &\text{First Derivatives} \\ H_x &= \partial_x H \qquad H_y = \partial_y H \\ &\text{Second Derivatives} \\ H_{xx} &= \partial_x^2 H \qquad H_{xy} = \partial_x\partial_y H \qquad H_{yy} = \partial_y^2 H \\ &\text{and use} \\ \vec{\partial} z &= (0,0,1) \end{aligned}$$

$$
\begin{aligned}
\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n} &= \left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\frac{\vec{\partial}\left(z-H\right)}{\sqrt{1+\left(\vec{\partial}_{\perp}H\right)^{2}}} \\
&= \frac{\left(-\hat{\rho}_{x}H_{xx}-\hat{\rho}_{y}H_{xy},-\hat{\rho}_{x}H_{xy}-\hat{\rho}_{y}H_{yy},0\right)}{\sqrt{1+H_{x}^{2}+H_{y}^{2}}} \\
&\quad +\left(H_{x},H_{y},-1\right)\left(\frac{\left(\hat{\rho}_{x}\left(H_{x}H_{xx}+H_{y}H_{xy}\right)+\hat{\rho}_{y}\left(H_{x}H_{xy}+H_{y}H_{yy}\right)\right)}{\left(1+H_{x}^{2}+H_{y}^{2}\right)^{3/2}}\right) \\
&= \frac{\begin{array}{c}\left(-\hat{\rho}_{x}H_{xx}-\hat{\rho}_{y}H_{xy},-\hat{\rho}_{x}H_{xy}-\hat{\rho}_{y}H_{yy},0\right)\left(1+H_{x}^{2}+H_{y}^{2}\right)\\ +\left(H_{x},H_{y},-1\right)\left(\hat{\rho}_{x}\left(H_{x}H_{xx}+H_{y}H_{xy}\right)+\hat{\rho}_{y}\left(H_{y}H_{xy}+H_{y}H_{yy}\right)\right)\end{array}}{\left(1+H_{x}^{2}+H_{y}^{2}\right)^{3/2}} \\
&= \frac{\left(\begin{array}{c}\hat{\rho}_{x}\left(H_{x}H_{y}H_{xy}-\left(1+H_{y}^{2}\right)H_{xx}\right)+\hat{\rho}_{y}\left(H_{x}H_{y}H_{yy}-\left(1+H_{y}^{2}\right)H_{xy}\right),\\ \hat{\rho}_{x}\left(H_{x}H_{y}H_{xx}-\left(1+H_{x}^{2}\right)H_{xy}\right)+\hat{\rho}_{y}\left(H_{x}H_{y}H_{xy}-\left(1+H_{x}^{2}\right)H_{yy}\right),\\ -\left(\hat{\rho}_{x}\left(H_{x}H_{xx}+H_{y}H_{xy}\right)+\hat{\rho}_{y}\left(H_{y}H_{xy}+H_{y}H_{yy}\right)\right)\end{array}\right)}{\left(1+H_{x}^{2}+H_{y}^{2}\right)^{3/2}}
\end{aligned}
$$

Consider some special cases. First, note that the x,y componentf of $\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}$ does not, in general, point in the same direction as $\hat{\rho}$. Second, note that if $H_{xx}=H_{yy}=H_{xy}=0$ at a given point then we get $\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}=(0,0,0)$ at that point. This makes sense, in order to get a nonzero result for $\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}$ at any point the surface must have some curvature, i.e. at least one nonvanishing second derivative at that point. If the surface is locally "flat", $H_{xx}=H_{yy}=H_{xy}=0$, then $\hat{n}$ doesn't change for infinitesimal displacements from that point. Third, note that if $H_{x}=H_{y}=0$ at a given point then

$$
\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}=-\left(\hat{\rho}_{x}H_{xx}+\hat{\rho}_{y}H_{xy},\hat{\rho}_{x}H_{xy}+\hat{\rho}_{y}H_{yy},0\right)
$$

which makes sense. The vanishing of the first derivatives means that, at that point, the surface is perfectly horizontal and so any tangent vector at that point will have only x and y components, no z component and, further, the curvature will be proportional to the second derivatives.

Curvature is defined as the reciprocal or multiplicative inverse of the radius of curvature. For a circle of radius R, the arc length swept out on the circle ds for a change in angle $d\theta$ is

$$
ds=Rd\theta=\sqrt{1+\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)^{2}}
$$

NOTE: This is extrinsic curvature as it depends on how the surface is embedded in 3D. A curve embedded in 3D will have extrinsic curvature, e.g., the curvature of a circle, but it does not have intrinsic curvature which is defined by the Riemann curvature tensor.

Hence the curvature κ of a circle of radius R is given by

$$\kappa = \frac{1}{R} = \frac{d\theta}{ds}$$

Since $\hat{n}(x,y)$, by definition, has unit length everywhere, the difference between $\hat{n}(x,y)$ and $\hat{n}(x+dx,y+dy)$ can only be a rotation. Note that $\hat{n}$ can rotate in any direction and so $\left(d\vec{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}$ does not necessarilly point on the $d\vec{\rho}$ direction.

As discussed above, if it doesn't vanish,

$$\left(\hat{n}(x+dx,y+dy)-\hat{n}(x,y)\right)=\left(d\vec{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}$$

is tangent to the surface and therefore

$$\begin{aligned} d\theta &= \left(\left(d\rho\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}\right)\cdot\hat{t}(\hat{\rho}) \\ &= -\hat{n}\cdot\left(\left(d\rho\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{t}(\hat{\rho})\right) \end{aligned}$$

where the second line follows from $\hat{n}\cdot\hat{t}(\hat{\rho})=0$ for all x and y. Note that $d\theta$ depends on the direction $\hat{\rho}$.

The curvature $\kappa(\hat{\rho})$ of the 1D path or trajectory of the intersection of $H(x,y)$ with the plane containing $\hat{\rho}$ and $\hat{n}$ is thus given by

$$\begin{aligned} \kappa(\hat{\rho}) &= \frac{d\theta}{ds} \\ &= \frac{\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}\right)\cdot\hat{t}(\hat{\rho})}{\sqrt{1+\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)^{2}}} \\ &= -\frac{\hat{n}\cdot\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{t}(\hat{\rho})\right)}{\sqrt{1+\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)^{2}}} \end{aligned}$$

The last two lines follow from $\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\left(\hat{n}\cdot\hat{t}\right)=0$ and it is a matter of choice which one to use.

Note: There is a sign convention associated with the curvature which is (I believe) opposite to the one used above.

Begin comment

We can write this another way using $\hat{n}=\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{t}(\hat{\rho})\right)/\left|\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{t}(\hat{\rho})\right|$. Note that $\hat{n}$ expressed this way is independent of $\hat{\rho}$, hence, assuming, $\left|\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{t}\right|\neq$

0, we have

$$\begin{aligned}\kappa\left(\hat{\rho},\hat{\rho}'\right) &\equiv -\frac{\left(\left(\hat{\rho}'\cdot\vec{\partial}_{\perp}\right)\hat{t}\left(\hat{\rho}'\right)\right)\cdot\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{t}\left(\hat{\rho}\right)\right)}{\left|\left(\hat{\rho}'\cdot\vec{\partial}_{\perp}\right)\hat{t}\left(\hat{\rho}'\right)\right|\sqrt{1+\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)^{2}}} \\ &= -\frac{\left(\left(\hat{\rho}'\cdot\vec{\partial}_{\perp}\right)\hat{t}\left(\hat{\rho}'\right)\right)\cdot\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{t}\left(\hat{\rho}\right)\right)}{\sqrt{\left(\left(\hat{\rho}'\cdot\vec{\partial}_{\perp}\right)\hat{t}\left(\hat{\rho}'\right)\right)\cdot\left(\left(\hat{\rho}'\cdot\vec{\partial}_{\perp}\right)\hat{t}\left(\hat{\rho}'\right)\right)}\sqrt{1+\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)^{2}}}\end{aligned}$$

for any $\hat{\rho}$ and $\hat{\rho}'$. This expresses the curvature entirely in terms of $\hat{t}$.

For $\hat{\rho}=\hat{\rho}'$ this reduces to

$$\kappa\left(\hat{\rho}\right)=\frac{\sqrt{\left(\left(\hat{\rho}'\cdot\vec{\partial}_{\perp}\right)\hat{t}\left(\hat{\rho}'\right)\right)\cdot\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{t}\left(\hat{\rho}\right)\right)}}{\sqrt{1+\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)^{2}}}$$

The same game can be played using $\hat{t}\left(\hat{\rho}\right)=\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}\right)/\left|\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}\right|$, again assuming $\left|\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}\right|\neq 0$. In this way $\kappa\left(\hat{\rho}\right)$ can be expressed entirely in terms of $\hat{n}$.

End comment

Evaluating the $\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)\hat{n}\right)\cdot\hat{t}$ form for $\kappa\left(\hat{\rho}\right)$ above gives, with "$\cdot$" meaning matrix multplication,

$$\kappa\left(\hat{\rho}\right)=\frac{\left(\hat{\rho}_x,\hat{\rho}_y\right)\cdot\begin{pmatrix}\partial_x^2 H & \partial_x\partial_y H\\ \partial_x\partial_y H & \partial_y^2 H\end{pmatrix}\cdot\begin{pmatrix}\hat{\rho}_x\\ \hat{\rho}_y\end{pmatrix}}{\sqrt{1+\left(\vec{\partial}_{\perp}H\right)^{2}}\left(1+\left(\left(\hat{\rho}\cdot\vec{\partial}_{\perp}\right)H\right)^{2}\right)}$$

Examples:

1.

$$H=H\left(x\right)$$

$$\kappa\left(\hat{\rho}\right)=\frac{\hat{\rho}_x^2\partial_x^2}{\sqrt{1+\left(\partial_x H\right)^2}\left(1+\hat{\rho}_x^2\left(\partial_x H\right)^2\right)}$$

2.

$$H=xy$$

$$\kappa\left(\hat{\rho}\right)=\frac{2\hat{\rho}_x\hat{\rho}_y}{\left(1+\left(\hat{\rho}_x y+\hat{\rho}_y x\right)^2\right)\sqrt{1+x^2+y^2}}$$

3.

$$H=\frac{x^2}{2R_x}+\frac{y^2}{2R_y}$$

$$\kappa\left(\hat{\rho}\right)=\frac{\hat{\rho}_x^2/R_x+\hat{\rho}_y^2/R_y}{\left(1+\left(\hat{\rho}_x x/R_x+\hat{\rho}_y y/R_y\right)^2\right)\sqrt{1+\left(x/R_x\right)^2+\left(y/R_y\right)^2}}$$

4.

$$H=-\sqrt{R-x^2-y^2}$$

$$\kappa\left(\hat{\rho}\right)=\frac{1}{R}$$

30 Riemann Curvature

This section has all the math but not much explanation. Need to work on it.

NOTE: Riemann curvature is "intrinsic curvature", it is not the same thing as the extrinsic curvature $\hat{\kappa}$ discussed above. For example, a circle is the embedding of a 1D space into the 2D plane. The circle has a constant extrinsic curvature given by 1/(circle radius). If the circle is distorted then the extrinsic curvature becomes position dependent. But for all (nonpathological) embeddings its Riemann curvature is 0. More generally think of a 1D space embedded in 3D, i.e., a curve in 3D. Let the coordinate along the curve be the length ℓ along the curve from some given starting point. Then $ds^2=d\ell^2$ and the metric, which is a 1×1 matrix in this case is $g=1$. All its derivatives vanish and so the Riemann curvature (derived below) which depends on the second derivatives of g is explicitly zero no matter how twisted or contorted the 1D curve is in 3D.

N dimensional space: Cartesian $=R^N$

$$\begin{aligned}\vec{x} &= x_i\hat{u}_i\\ &\quad \text{Summation convention}\\ \hat{u}_i\cdot\hat{u}_j &= \delta_{i,j}\\ \hat{u}_i &= \text{Cartesian}\\ \vec{x}\left(\vec{\xi}\right) &= x_i\left(\vec{\xi}\right)\hat{u}_i\end{aligned}$$

Notation : Drop arrow on ξ in function arguments: $\vec{x}\left(\vec{\xi}\right)\longrightarrow\vec{x}\left(\xi\right)$, etc.
Same for functions of $\vec{x}$, drop arrow

$$\begin{aligned}\vec{\xi} &= N \text{ dimensional curvilinear coords in } R^N\\ &= \left(\xi^1,\xi^2,\cdots,\xi^N\right)\\ &\quad \text{Use notation } \xi^{Greek} \text{ and } x_{Latin}\end{aligned}$$

Why distinguish upper lower indices?
Because $d\xi^\mu$ and $\partial/\partial\xi^\mu$ transform oppositely

ξ^μ = manifold coordinates

$$\begin{array}{rcl} & & \text{Transformation} \\ d\eta^\mu & = & \left(\partial\eta^\mu/\partial\xi^\nu\right) d\xi^\nu \\ \partial/\partial\eta^\mu & = & \left(\partial\xi^\nu/\partial\eta^\mu\right) \partial/\partial\xi^\nu \end{array}$$

$$\begin{array}{rcl} d\eta^\mu \partial/\partial\eta^\mu & = & \left(\partial\eta^\mu/\partial\xi^\nu\right) d\xi^\nu \left(\partial\xi^\alpha/\partial\eta^\mu\right) \partial/\partial\xi^\alpha \\ & = & \left(\partial\xi^\alpha/\partial\eta^\mu\right) \left(\partial\eta^\mu/\partial\xi^\nu\right) d\xi^\nu \partial/\partial\xi^\alpha \\ & = & \delta^\alpha_\nu d\xi^\nu \partial/\partial\xi^\alpha \\ & = & d\xi^\mu \partial/\partial\xi^\mu \\ & & \text{Invariant (formwise)} \end{array}$$

$$\begin{array}{rcl} & & \text{Notation} \\ \partial_{Latin} & = & \dfrac{\partial}{\partial x_{Latin}} \\ \partial_{Greek} & = & \dfrac{\partial}{\partial \xi^{Greek}} \end{array}$$

Need to remember x_i and ξ^μ appear as both functions and variables, i.e.,

$$\begin{array}{rcll} \partial_i \xi^\mu & = & \dfrac{\partial \xi^\mu\left(x\right)}{\partial x_i} & \text{Here } x_i \text{ is a variable, } \xi^\mu \text{ is a function} \\ \partial_\mu x_i & = & \dfrac{\partial x_i\left(\xi\right)}{\partial \xi^\mu} & \text{Here } \xi^\mu \text{ is a variable, } x_i \text{ is a function} \end{array}$$

$$\begin{array}{rcl} d\vec{r} & = & d\xi^\mu \partial_\mu \vec{r} \\ & = & d\xi^\mu \vec{e}_\mu \\ \vec{e}_\mu & = & \text{local basis vector along } \xi^\mu \text{ coords in } R^N \end{array}$$

Derive the metric $g_{\mu\nu}$

$$\begin{array}{rcl} ds^2 & = & d\vec{r} \cdot d\vec{r} \\ & = & d\xi^\mu \vec{e}_\mu \cdot d\xi^\nu \vec{e}_\nu \\ & = & \left(\vec{e}_\mu \cdot \vec{e}_\nu\right) d\xi^\mu d\xi^\nu \\ & \equiv & g_{\mu\nu} d\xi^\mu d\xi^\nu \end{array}$$

$$\begin{array}{rcl} g_{\mu\nu} & = & \vec{e}_\mu \cdot \vec{e}_\nu \\ & = & \left(\partial_\mu x_i\right) \left(\partial_\nu x_j\right) \hat{u}_i \cdot \hat{u}_j \\ & = & \left(\partial_\mu x_i\right) \left(\partial_\nu x_j\right) \delta_{ij} \\ & = & \left(\partial_\mu x_i\right) \left(\partial_\nu x_i\right) \end{array}$$

$$g_{\mu\nu} = g_{\nu\mu}$$

Metric Inverse

$$\begin{aligned} g_{\mu\alpha}\left(\partial_j\xi^\alpha\right)\left(\partial_j\xi^\nu\right) &= \left(\partial_\mu x_i\right)\left(\partial_\alpha x_i\right)\left(\partial_j\xi^\alpha\right)\left(\partial_j\xi^\nu\right) \\ &= \left(\partial_\mu x_i\right)\left(\partial_j\xi^\alpha\right)\left(\partial_\alpha x_i\right)\left(\partial_j\xi^\nu\right) \\ &= \left(\partial_\mu x_i\right)\left(\partial_j x_i\right)\left(\partial_j\xi^\nu\right) \\ &= \left(\partial_\mu x_i\right)\delta_{ij}\left(\partial_j\xi^\nu\right) \\ &= \left(\partial_\mu x_i\right)\left(\partial_i\xi^\nu\right) \\ &= \partial_\mu\xi^\nu \\ &= \delta_\mu^\nu \end{aligned}$$

$$g^{\mu\nu} = \left(\partial_j\xi^\mu\right)\left(\partial_j\xi^\nu\right) = \left(g^{-1}\right)_{\mu\nu}$$

Vector field

$$\vec{v}\left(\xi\right) = v^\mu\left(\xi\right)\vec{e}_\mu\left(\xi\right)$$

Covariant Derivatives

$$\partial_\nu\vec{v}\left(\xi\right) = \left(\partial_\nu v^\mu\left(\xi\right)\right)\vec{e}_\mu\left(\xi\right) + v^\mu\left(\xi\right)\left(\partial_\nu\vec{e}_\mu\left(\xi\right)\right)$$

$\partial_\nu\vec{e}_\mu\left(\xi\right)$ is a vector so it can be expanded in terms of $\vec{e}_\gamma$

This only works if ξ_μ and x_i are the same dimension.

i.e., the ξ_μ are curvilinear coordinates in x_i space.

$$\partial_\nu\vec{e}_\mu\left(\xi\right) = \Gamma_{\nu\mu}^\gamma\left(\xi\right)\vec{e}_\gamma\left(\xi\right)$$

$$\begin{aligned} \partial_\nu\vec{e}_\mu &= \partial_\nu\partial_\mu x_i\hat{u}_i \\ &= \partial_\nu\partial_\mu x_i\delta_{ij}\hat{u}_j \\ &= \left(\partial_\nu\partial_\mu x_i\right)\left(\frac{\partial x_j}{\partial x_i}\right)\hat{u}_j \\ &= \left(\partial_\nu\partial_\mu x_i\right)\left(\frac{\partial\xi^\gamma}{\partial x_i}\frac{\partial x_j}{\partial\xi^\gamma}\right)\hat{u}_j \\ &= \left(\partial_\nu\partial_\mu x_i\right)\left(\partial_i\xi^\gamma\right)\left(\left(\partial_\gamma x_j\right)\hat{u}_j\right) \\ &= \left(\partial_\nu\partial_\mu x_i\right)\left(\partial_i\xi^\gamma\right)\vec{e}_\gamma \end{aligned}$$

$$\begin{aligned} \Gamma_{\nu\mu}^\gamma &= \left(\partial_\nu\partial_\mu x_i\right)\left(\partial_i\xi^\gamma\right) \\ &= \left(\frac{\partial^2 x_i\left(\xi\right)}{\partial\xi_\nu\partial\xi_\mu}\right)\left(\frac{\partial\xi^\gamma\left(x\right)}{\partial x_i}\right) \end{aligned}$$

$$\Gamma_{\nu\mu}^\gamma = \Gamma_{\mu\nu}^\gamma$$

$$\begin{aligned} \partial_\nu\vec{v} &= \left(\partial_\nu v^\mu\right)\vec{e}_\mu + v^\mu\Gamma_{\nu\mu}^\gamma\vec{e}_\gamma \\ &= \left(\partial_\nu v^\mu\right)\delta_\mu^\gamma\vec{e}_\gamma + v^\mu\Gamma_{\nu\mu}^\gamma\vec{e}_\gamma \\ &= \left(\partial_\nu v^\mu\right)\delta_\mu^\gamma\vec{e}_\gamma + v^\mu\Gamma_{\nu\mu}^\gamma\vec{e}_\gamma \\ &= \left(\left(\delta_\mu^\gamma\partial_\nu + \Gamma_{\nu\mu}^\gamma\right)v^\mu\right)\vec{e}_\gamma \\ &= \left(\left(\delta_\mu^\gamma\partial_\nu + \Gamma_{\mu\nu}^\gamma\right)v^\mu\right)\vec{e}_\gamma \end{aligned}$$

$$D^{\gamma}_{\mu\nu} = \left(\delta^{\gamma}_{\mu}\partial_{\nu} + \Gamma^{\gamma}_{\mu\nu}\right)$$

$$\begin{aligned}
\text{Trajectory:} \quad & \xi^{\mu}(t) \\
t &= \text{monotonically increasing parameter along trajectory} \\
t &= 0 \rightarrow 1 \\
& \text{Trajectory minimizes distance}
\end{aligned}$$

$$\begin{aligned}
& \text{In } R^N \text{ straight line is shortest distance between two points} \\
\text{hence } x_i(\xi(t)) &= (x_i(\xi(1)) - x_i(\xi(0)))\, t + x_i(\xi(0))
\end{aligned}$$

$$\begin{aligned}
\partial_t x_i(\xi(t)) &= (\partial_t \xi^{\mu})(\partial_{\mu} x_i) = (x_i(\xi(1)) - x_i(\xi(0))) = \text{constant} \\
\partial_t^2 x_i(\xi(t)) &= \left(\partial_t^2 \xi^{\mu}\right)(\partial_{\mu} x_i) + (\partial_t \xi^{\mu})(\partial_t \xi^{\nu})(\partial_{\mu}\partial_{\nu} x_i) = 0 \\
& \text{multiply by } \partial_i \xi^{\gamma}
\end{aligned}$$

$$\begin{aligned}
0 &= (\partial_i \xi^{\gamma})\left(\partial_t^2 \xi^{\mu}\right)(\partial_{\mu} x_i) + (\partial_i \xi^{\gamma})(\partial_t \xi^{\mu})(\partial_t \xi^{\nu})(\partial_{\mu}\partial_{\nu} x_i) \\
&= \left(\partial_t^2 \xi^{\mu}\right)(\partial_i \xi^{\gamma})(\partial_{\mu} x_i) + (\partial_t \xi^{\mu})(\partial_t \xi^{\nu})(\partial_{\mu}\partial_{\nu} x_i)(\partial_i \xi^{\gamma}) \\
&= \left(\partial_t^2 \xi^{\mu}\right)\delta^{\gamma}_{\mu} + (\partial_t \xi^{\mu})(\partial_t \xi^{\nu})\,\Gamma^{\gamma}_{\mu\nu} \\
&= \left(\partial_t^2 \xi^{\gamma}\right) + \Gamma^{\gamma}_{\mu\nu}(\partial_t \xi^{\mu})(\partial_t \xi^{\nu}) \\
&= \text{geodesic equation}
\end{aligned}$$

$$\begin{aligned}
& \text{Derivative of metric} \\
\partial_{\gamma} g_{\mu\nu} &= \partial_{\gamma}\left((\partial_{\mu} x_i)(\partial_{\nu} x_i)\right) \\
&= (\partial_{\gamma}\partial_{\mu} x_i)(\partial_{\nu} x_i) + (\partial_{\mu} x_i)(\partial_{\gamma}\partial_{\nu} x_i) \\
&= (\partial_{\gamma}\partial_{\mu} x_i)\,\delta_{ij}\,(\partial_{\nu} x_j) + (\partial_{\mu} x_i)\,\delta_{ij}\,(\partial_{\gamma}\partial_{\nu} x_j) \\
&= (\partial_{\gamma}\partial_{\mu} x_i)(\partial_i \xi^{\alpha})(\partial_{\alpha} x_j)(\partial_{\nu} x_j) + (\partial_{\mu} x_i)(\partial_j \xi^{\alpha})(\partial_{\alpha} x_i)(\partial_{\gamma}\partial_{\nu} x_j) \\
&= \Gamma^{\alpha}_{\mu\gamma}(\partial_{\alpha} x_j)(\partial_{\nu} x_j) + \Gamma^{\alpha}_{\nu\gamma}(\partial_{\alpha} x_i)(\partial_{\mu} x_i) \\
&= \Gamma^{\alpha}_{\mu\gamma} g_{\alpha\nu} + \Gamma^{\alpha}_{\nu\gamma} g_{\alpha\mu} \\
&= \text{Linear combination of the metric} \\
& \text{"Metric Preserving"}
\end{aligned}$$

$$\partial_{\gamma} g_{\mu\nu} - \Gamma^{\alpha}_{\mu\gamma} g_{\alpha\nu} - \Gamma^{\alpha}_{\nu\gamma} g_{\alpha\mu} = 0$$

$$\Gamma^{\gamma}_{\mu\nu} \text{ in terms of } g_{\mu\nu}$$

$$\begin{aligned}
\partial_{\gamma} g_{\mu\nu} &= \Gamma^{\alpha}_{\mu\gamma} g_{\alpha\nu} + \Gamma^{\alpha}_{\nu\gamma} g_{\alpha\mu} \\
\partial_{\mu} g_{\gamma\nu} &= \Gamma^{\alpha}_{\mu\gamma} g_{\alpha\nu} + \Gamma^{\alpha}_{\nu\mu} g_{\alpha\gamma} \\
\partial_{\nu} g_{\mu\gamma} &= \Gamma^{\alpha}_{\mu\nu} g_{\alpha\gamma} + \Gamma^{\alpha}_{\nu\gamma} g_{\alpha\mu}
\end{aligned}$$

$$\begin{aligned}\partial_\mu g_{\gamma\nu}+\partial_\nu g_{\mu\gamma}-\partial_\gamma g_{\mu\nu} &= \Gamma^\alpha_{\mu\gamma}g_{\alpha\nu}+\Gamma^\alpha_{\nu\mu}g_{\alpha\gamma}\\ &\quad +\Gamma^\alpha_{\mu\nu}g_{\alpha\gamma}+\Gamma^\alpha_{\nu\gamma}g_{\alpha\mu}\\ &\quad -\Gamma^\alpha_{\mu\gamma}g_{\alpha\nu}-\Gamma^\alpha_{\nu\gamma}g_{\alpha\mu}\\ &= 2\Gamma^\alpha_{\mu\nu}g_{\alpha\gamma}\end{aligned}$$

$$\begin{aligned}\Gamma^\alpha_{\mu\nu} &= \frac{1}{2}\left(g^{-1}\right)_{\alpha\gamma}\left(\partial_\mu g_{\gamma\nu}+\partial_\nu g_{\mu\gamma}-\partial_\gamma g_{\mu\nu}\right)\\ &= \frac{1}{2}g^{\alpha\gamma}\left(\partial_\mu g_{\gamma\nu}+\partial_\nu g_{\mu\gamma}-\partial_\gamma g_{\mu\nu}\right)\end{aligned}$$

$$\begin{aligned}\partial_t\vec{v}\left(\xi\left(t\right)\right) &= \left(\partial_t\xi^\mu\right)\left(\partial_\mu\vec{v}\right)\\ &= \left(\partial_t\xi^\mu\right)\left(D^\gamma_{\mu\nu}v^\nu\right)\vec{e}_\gamma\end{aligned}$$

$$\begin{aligned}\partial_t\vec{v}\left(\xi\left(t\right)\right) &= 0\\ &\Longrightarrow \left(D^\gamma_{\mu\nu}v^\nu\right)=0\\ &\quad \text{Parallel Transport}\end{aligned}$$

Curvature

$$\begin{aligned}\partial_\beta\vec{e}_\mu &= \Gamma^\nu_{\beta\mu}\vec{e}_\nu\\ \partial_\alpha\left(\partial_\beta\vec{e}_\mu\right) &= \left(\partial_\alpha\Gamma^\nu_{\beta\mu}\right)\vec{e}_\nu+\Gamma^\nu_{\beta\mu}\left(\partial_\alpha\vec{e}_\nu\right)\\ &= \left(\partial_\alpha\Gamma^\nu_{\beta\mu}\right)\vec{e}_\nu+\Gamma^\nu_{\beta\mu}\Gamma^\sigma_{\alpha\nu}\vec{e}_\sigma\end{aligned}$$

$$\begin{aligned}\text{second term:}\quad \text{swap}\quad & \nu\longleftrightarrow\sigma\\ = \quad & \left(\partial_\alpha\Gamma^\nu_{\beta\mu}\right)\vec{e}_\nu+\Gamma^\sigma_{\beta\mu}\Gamma^\nu_{\alpha\sigma}\vec{e}_\nu\end{aligned}$$

$$\text{swap: } \alpha\longleftrightarrow\beta$$

$$\partial_\beta\left(\partial_\alpha\vec{e}_\mu\right)=\left(\partial_\beta\Gamma^\nu_{\alpha\mu}\right)\vec{e}_\nu+\Gamma^\nu_{\alpha\mu}\Gamma^\sigma_{\beta\nu}\vec{e}_\sigma$$

$$\text{subtract}$$

$$\partial_\alpha\left(\partial_\beta\vec{e}_\mu\right)-\partial_\beta\left(\partial_\alpha\vec{e}_\mu\right)=\left(\partial_\alpha\Gamma^\nu_{\beta\mu}-\partial_\beta\Gamma^\nu_{\alpha\mu}+\Gamma^\sigma_{\beta\mu}\Gamma^\nu_{\alpha\sigma}-\Gamma^\sigma_{\alpha\mu}\Gamma^\nu_{\beta\sigma}\right)\vec{e}_\nu$$

$$\begin{aligned}&\text{Riemann Curvature Tensor Definition}\\ R^\nu_{\alpha\beta\mu} &= \partial_\alpha\Gamma^\nu_{\beta\mu}-\partial_\beta\Gamma^\nu_{\alpha\mu}+\Gamma^\sigma_{\beta\mu}\Gamma^\nu_{\alpha\sigma}-\Gamma^\sigma_{\alpha\mu}\Gamma^\nu_{\beta\sigma}\\ &\text{NOTE: By definition this is the matrix coefficient of } \vec{e}_\nu\end{aligned}$$

NOTE: The way the indices are ordered on R varies. The ordering above corresponds to that in the book "Gauge Fields, Knots and Gravity" by Baez and

Munian. A different ordering occurs in Steven Weinberg's "Gravitation and Cosmology". Weinberg's ordering is the same that occurs in Tony Zee's "Quantum Field Theory in a Nutshell". For example if we use the notation $\bar{R}$ for Weinberg's ordering and R for the ordering used here then we have $R^{\nu}_{\alpha\beta\mu} = \bar{R}^{\nu}_{\mu\beta\alpha}$

$$
\begin{aligned}
& \text{But } \partial_\alpha\left(\partial_\beta \vec{e}_\mu\right) - \partial_\beta\left(\partial_a \vec{e}_\mu\right) \text{ should be zero since} \\
0 \;=\; & \partial_\alpha\partial_\beta - \partial_\beta\partial_\alpha \\
& \text{Check this explicitly} \\
& \text{Be careful to distinguish } \xi^\mu \text{ as a coordinate variable} \\
& \text{from } \xi^\mu \text{ as a function of } x_i \\
& \text{e.g., } \left(\partial_i \xi^\mu\right) \text{ is a function of } x \text{ so } \partial_\nu\left(\partial_i \xi^\mu\right) \text{ is zero}
\end{aligned}
$$

$$
\begin{aligned}
& \partial_\alpha \Gamma^{\nu}_{\beta\mu} - \partial_\beta \Gamma^{\nu}_{\alpha\mu} + \Gamma^{\sigma}_{\beta\mu}\Gamma^{\nu}_{\alpha\sigma} - \Gamma^{\sigma}_{\alpha\mu}\Gamma^{\nu}_{\beta\sigma} \\
= \;& \partial_\alpha\left(\partial_\beta\partial_\mu x_i\right)\left(\partial_i \xi^\nu\right) - \partial_\beta\left(\partial_\alpha\partial_\mu x_i\right)\left(\partial_i \xi^\nu\right) \\
& + \left(\partial_\beta\partial_\mu x_i\right)\left(\partial_i \xi^\sigma\right)\left(\partial_\alpha\partial_\sigma x_j\right)\left(\partial_j \xi^\nu\right) \\
& - \left(\partial_a\partial_\mu x_i\right)\left(\partial_i \xi^\sigma\right)\left(\partial_\beta\partial_\sigma x_j\right)\left(\partial_j \xi^\nu\right) \\
= \;& \left(\left(\partial_\alpha\partial_\beta\partial_\mu x_i\right) - \left(\partial_\beta\partial_\alpha\partial_\mu x_i\right)\right)\left(\partial_i \xi^\nu\right) \\
& + \left(\partial_\beta\partial_\mu x_i\right)\left(\partial_\alpha\left(\left(\partial_i \xi^\sigma\right)\partial_\sigma x_j\right)\right)\left(\partial_j \xi^\nu\right) \\
& - \left(\partial_a\partial_\mu x_i\right)\left(\partial_\beta\left(\left(\partial_i \xi^\sigma\right)\partial_\sigma x_j\right)\right)\left(\partial_j \xi^\nu\right) \\
= \;& \left(\partial_\beta\partial_\mu x_i\right)\left(\partial_\alpha \delta_{ij}\right)\left(\partial_j \xi^\nu\right) \\
& - \left(\partial_a\partial_\mu x_i\right)\left(\partial_\beta \delta_{ij}\right)\left(\partial_j \xi^\nu\right) \\
= \;& 0
\end{aligned}
$$

Important point to note about all of the above: $\Gamma^{\mu}_{\alpha\beta}$ and hence $R^{\mu}_{\alpha\beta\nu}$ can be written entirely in terms of $g_{\mu\nu}$. The fact that everything can be written in terms of $g_{\mu\nu}$ is, effectively, the Gauss "Theorema Egregium", which literally translates as remarkable or outstandingly good theorem. Of course, today, the meaning of egregious is exactly the opposite.

Now, given the above, the question is

$$\text{How can } R^{\nu}_{\alpha\beta\mu} \neq 0 \text{ ?}$$

Consider a sub-manifold M, e.g., a "surface", of dimension D embedded in Euclidean space, R^N, of dimension $N > D$. "Embedded" means, basically, that M as a submanifold of R^N has no self intersections. We will assume the embedding is smooth, i.e., all the functions $x_i\left(\vec{\xi}\right)$ have derivatives at least up to second order.

NOTE: The Whitney Embedding theorem guarantees that a manifold of dimension D can always be embedded in Euclidean space of dimension $2D+1$.

Now

$$
\begin{aligned}
\vec{x} &= (x_1, x_2, \cdots, x_N) \\
\vec{\xi} &= (\xi_1, \xi_2, \cdots, \xi_D)
\end{aligned}
$$

The fact that $D < N$ changes things. Let $\vec{x}(\xi)$ map points in M as a function of the ξ^μ coordinates or parameters into R^N. Consider an infinitesimal displacement vector in R^N, $d\vec{x}$, that is tangent to M at some point, i.e.,

$$d\vec{x} = d\xi^\mu \partial_\mu \vec{x}(\xi) = d\xi^\mu \partial_\mu x_i(\xi)\,\hat{u}_i$$

where again the $\hat{u}_i$ are the Cartesian unit vectors of R^N. The length squared of $d\vec{x}$ is

$$\begin{aligned} ds^2 &= d\vec{x}\cdot d\vec{x} \\ &= d\xi^\mu \partial_\mu x_i(\xi)\,\hat{u}_i d\xi^\nu \partial_\nu x_j(\xi)\,\hat{u}_j \\ &= \vec{e}_\mu(\xi)\cdot\vec{e}_\mu(\xi)\,d\xi^\mu d\xi^\nu \\ &\equiv g_{\mu\nu}(\xi)\,d\xi^\mu d\xi^\nu \end{aligned}$$

where

$$\vec{e}_\mu(\xi) = \partial_\mu x_i(\xi)\,\hat{u}_i$$

are vectors tangent to M. So we still have a metric as a function of the ξ_μ.

But now

$$\partial_\nu \vec{e}_\mu \neq \Gamma^\sigma_{\nu,\mu}\vec{e}_\sigma$$

since $\partial_\nu \vec{e}_\mu(\xi)$ will not, in general, be tangent to M.

But $\Gamma^\sigma_{\nu,\mu}$ is a function of the metric $g_{\mu,\nu}(\xi)$, and its derivatives, so the Riemann curvature tensor is a function of the metric and its derivatives and we can simply calculate it for a given metric. Theorema Egregium.

31 Density Matrix

Usually in quantum mechanics problems one is computing results for a specific wave function or a specific combination of wave functions. But what if the combination itself had a probability distribution. A simple example would be that a colleague is supposed to prepare a specific wave function for you to do measurements on but this colleague is a bit of a practical joker and so at random sends you different wave functions to measure. Or, more realistically, the system you are dealing with is interacting with another system which randomly changes the wave function of the system you are measuring. In either case, when analyzing your results you must account for this extra randomness, which is "on top" of the inherent quantum randomness of the system. From the point of view of superposition, this can be thought of as the numerical coefficients in the superposition having their own probability distribution instead of being fixed. You can also think of it as the different possible wave functions with each having their own probability of occurring.

Let $|\psi_i\rangle$ be a complete set of quantum basis states. The corresponding wave functions are given by $\psi_i(x) = \langle x|\psi_i\rangle$ where $|x\rangle$ are position eigenstates. By definition an arbitrary state is given by a superposition

$$|\psi\rangle = \sum_i c_i\,|\psi_i\rangle$$

where the c_i are complex valued numbers. Assume the $|\psi_i\rangle$ are orthonormal, $\langle\psi_i|\psi_j\rangle = \delta_{ij}$ where δ_{ij} is the Kronecker delta, i.e., the elements of the identity matrix with i and j labeling rows and columns, respectively.

We always want the wave function to be normalized

$$\begin{aligned}\langle\psi|\psi\rangle &= \sum_{i,j} c_i^* c_j \langle\psi_i|\psi_j\rangle \\ &= \sum_i |c_i|^2 \\ &= 1\end{aligned}$$

The corresponding wave function is given by

$$\langle x|\psi\rangle = \sum_i c_i \psi_i(x)$$

A so called pure state has a given set of the c_i.

Stating the obvious, quantum mechanics itself is probabalistic. If the c_i themselves are also probabalistic then the state is said to be "not pure". Let $P[c]$ be the joint probability distribution for the c_i. For the pure state

$$\left\langle\psi|\hat{A}|\psi\right\rangle = \sum_{i,j} c_i^* c_j \left\langle\psi_i|\hat{A}|\psi_j\right\rangle \equiv \sum_{i,j} c_i^* c_j A_{i,j}$$

is the expectation value of the operator $\hat{A}$ for the state $|\psi\rangle$. For the non-pure state we have to average over the probability distribution $P[c]$ as well. Indicate this average by the notation $\langle\cdots\rangle_P$.

Note that for all possible values of c_i we want to have the state normalized, i.e., $\langle\psi|\psi\rangle = \sum_i |c_i|^2 = 1$

Using $P[c]$ we have

$$\begin{aligned}\left\langle\left\langle\psi|\hat{A}|\psi\right\rangle\right\rangle_P &= \int d^N c P[c] \left\langle\psi|\hat{A}|\psi\right\rangle \\ &= \sum_{i,j} \int d^N c P[c] c_i^* c_j A_{i,j} \\ &\equiv \sum_{i,j} \rho_{j,i} A_{i,j} \\ &= \mathrm{tr}[\rho \cdot A]\end{aligned}$$

where "$\cdot$" indicates matrix multiplication with the matrix $\rho_{j,i}$ defined by

$$\int d^N c P[c] c_i^* c_j = \int d^N c P[c] c_j c_i^*$$

where the equality follows from the fact that c_i are just complex numbers.

We can also write ρ as an operator $\hat{\rho}$

$$\begin{aligned}\left\langle |\psi\rangle\langle\psi| \right\rangle_P &= \left(\sum_{i,j}\int d^N cP\left[c\right] c_i c_j^*\right)|\psi_i\rangle\left\langle\psi_j\right| \\ &= \sum_{i,j}\rho_{i,j}|\psi_i\rangle\left\langle\psi_j\right| \\ &= \hat{\rho}\end{aligned}$$

Note that

$$\begin{aligned}\mathrm{tr}\left[\hat{\rho}\right] &= \sum_k\sum_{i,j}\rho_{i,j}\langle\psi_k|\psi_i\rangle\left\langle\psi_j|\psi_k\right\rangle \\ &= \sum_k \rho_{k,k} \\ &= \int d^N cP\left[c\right]\sum_k |c_k|^2 \\ &= \int d^N cP\left[c\right] \\ &= 1\end{aligned}$$

In thermal equilibrium, if $|n\rangle$ are eigenstates of the Hamiltonian, $\hat{H}$, i.e., $\hat{H}|n\rangle = E_n|n\rangle$ with E_n the energy eigenvalues then $\hat{\rho}$ is purely diagonal in the $|n\rangle$ basis and has the from

$$\hat{\rho} = \sum_n \frac{e^{-\beta E_n}}{Z}|n\rangle\langle n|$$

where

$$Z = \sum_n e^{-\beta E_n}$$

the partition function and $\beta = 1/k_B T$ with k_B the Boltzmann constant and T the temperature in Kelvin.

32 Solid Angles

32.1 Solid Angle in N dimensions.

Consider the Gaussian integral in N dimensions with $x = (x_1, x_2, \cdots, x_N)$. In the Cartesian x_n coordinates the integrals factorize and we have

$$\begin{aligned}\int_{-\infty}^{+\infty} d^N x \exp\left[-\sum_{n=1}^N x_n^2\right] &= \left(\int_{-\infty}^{+\infty} dx e^{-x^2}\right)^N \\ &= \pi^{N/2}\end{aligned}$$

after using the standard result

$$\int_{-\infty}^{+\infty} dx e^{-x^2} = \sqrt{\pi}$$

Rewrite the integral in polar coordinates with $r^2 = \sum_{n=1}^{N} x_n^2$

$$\begin{aligned}\int_{-\infty}^{+\infty} d^N x \exp\left[-\sum_{n=1}^{N} x_n^2\right] &= \int_0^{\infty} dr r^{N-1} e^{-r^2} \int d\Omega_{N-1} \\ &= \Omega_{N-1} \int_0^{\infty} dr r^{N-1} e^{-r^2}\end{aligned}$$

where Ω_{N-1} is the surface area of the $N-1$ dimensional surface of a unit radius sphere in N dimensions, i.e., it is the solid angle in N dimensions.

Letting $r^2 = t$ in the dr integral we have

$$\begin{aligned}\int_0^{\infty} dr r^{N-1} e^{-r^2} &= \frac{1}{2} \int_0^{\infty} \frac{dt}{\sqrt{t}} \left(\sqrt{t}\right)^{N-1} e^{-t} \\ &= \frac{1}{2} \int_0^{\infty} dt t^{N/2-1} e^{-t} \\ &= \frac{1}{2} \Gamma\left(N/2\right)\end{aligned}$$

where $\Gamma(\cdots)$ is the Euler Gamma function. For positive integer values of z, $\Gamma(z) = (z-1)!$.

Hence we have

$$\Omega_{N-1} \frac{1}{2} \Gamma\left(N/2\right) = \pi^{N/2}$$

or

$$\Omega_{N-1} = \frac{2\pi^{N/2}}{\Gamma\left(N/2\right)}$$

Now consider another way of approaching the same problem. Use a Dirac delta function to constrain $\sqrt{\sum_{n=1}^{N} x_n^2}$ to have unit length, i.e., to live on the $N-1$ dimensional surface of a unit sphere in N dimensions

$$\Omega_{N-1} = \int d^N x \delta_D \left(1 - \sqrt{\sum_{n=1}^{N} x_n^2}\right)$$

Use the identity

$$\delta_D \left(1 - \sqrt{\sum_{n=1}^{N} x_n^2}\right) = 2\delta_D \left(1 - \sum_{n=1}^{N} x_n^2\right)$$

and substitute the Fourier representation of the delta function and assume we can exchange orders of integration

$$\begin{aligned}
\Omega_{N-1} &= 2\int d^N x \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \exp\left[i\omega\left(1-\sum_{n=1}^{N} x_n^2\right)\right] \\
&= 2\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{i\omega}\left(\int_{-\infty}^{+\infty} dx e^{-i\omega x^2}\right)^N \\
&= 2\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi}\left(\frac{\pi}{i\omega}\right)^{N/2} e^{i\omega} \\
&= \pi^{N/2-1}\int_{-\infty}^{+\infty} d\omega \left(\frac{1}{i\omega}\right)^{N/2} e^{i\omega}
\end{aligned}$$

For $N = 2Q = Even$ we have

$$\int d^N x \delta_D\left(1-\sqrt{\sum_{n=1}^{N} x_n^2}\right) = \pi^{Q-1}\int_{-\infty}^{+\infty} d\omega \left(\frac{1}{i\omega}\right)^{Q} e^{i\omega}$$

but

$$\left(\frac{1}{i\omega}\right)^{Q} = (-1)^{Q-1}\frac{1}{(Q-1)!}\left(\partial_z^{Q-1}\frac{1}{z+i\omega}\right)_{z=0}$$

$$\begin{aligned}
\int_{-\infty}^{+\infty} d\omega \left(\frac{1}{i\omega}\right)^{Q} e^{i\omega} &= (-1)^{Q-1}\frac{1}{(Q-1)!}\left(\partial_z^{Q-1}\int_{-\infty}^{+\infty} d\omega \frac{1}{z+i\omega} e^{i\omega}\right)_{z=0} \\
&= (-1)^{Q-1}\frac{1}{(Q-1)!}\left(\frac{1}{i}\partial_z^{Q-1}\int_{-\infty}^{+\infty} d\omega \frac{1}{\omega - iz} e^{i\omega}\right)_{z=0} \\
&\quad \text{Close contour upper half complex plane, } \mathrm{Re}\,[z] \text{ positive} \\
&= (-1)^{Q-1}\frac{1}{(Q-1)!}\left(\frac{1}{i}\partial_z^{Q-1} 2\pi i e^{-z}\right)_{z=0} \\
&= 2\pi\frac{1}{(Q-1)!}
\end{aligned}$$

and so for $N = 2Q = Even$

$$\begin{aligned}
\Omega_{N-1} &= \frac{\pi^{N/2-1}2\pi}{(N/2-1)!} \\
&= \frac{2\pi^{N/2}}{\Gamma(N/2)}
\end{aligned}$$

Replacing $(N/2-1)!$ with $\Gamma(N/2)$ analytically continues the result to $N = odd$.

NOTE: This is the same result as obtained using the Gaussian integral approach.

32.2 Circular cone solid angle

To get the solid angle of a circular cone restrict the range of the x_N integral to be from $\cos(\theta_0)$ to 1, where θ_0 is the angle between $\vec{r} = (x_1, x_2, \cdots, x_N)$ and the x_N axis. Indicating the circular cone solid angle by $\Omega_{N-1}(\theta_0)$ we have

$$\begin{aligned}
\Omega_{N-1}(\theta_0) &= \int_{-\infty}^{+\infty} d^{N-1}x \int_{\cos(\theta_0)}^{1} dx_N \delta_D\left(1 - \sqrt{\vec{r}^2}\right) \\
&= 2\int_{-\infty}^{+\infty} d^{N-1}x \int_{\cos(\theta_0)}^{1} dx_N \delta_D\left(1 - \vec{r}^2\right) \\
&= 2\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{i\omega} \left(\int_{-\infty}^{+\infty} dx e^{-i\omega x^2}\right)^{N-1} \int_{\cos(\theta_0)}^{1} dx_N e^{-i\omega x_N^2} \\
&= 2\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{i\omega} \left(\frac{\pi}{i\omega}\right)^{(N-1)/2} \int_{\cos(\theta_0)}^{1} dx_N e^{-i\omega x_N^2}
\end{aligned}$$

Consider the case $N = odd$ and let $(N-1)/2 = Q$ and we have

$$\begin{aligned}
\Omega_{N-1}(\theta_0) &= \pi^{Q-1} \int_{\cos(\theta_0)}^{1} dx_N \int_{-\infty}^{+\infty} d\omega \left(\frac{1}{i\omega}\right)^{Q} e^{i\omega\left(1-x_N^2\right)} \\
&= \pi^{Q-1}(-1)^{Q-1} \frac{1}{(Q-1)!} \int_{\cos(\theta_0)}^{1} dx_N \left(\partial_z^{Q-1} \int_{-\infty}^{+\infty} d\omega \frac{1}{z + i\omega} e^{i\omega\left(1-x_N^2\right)}\right)_{z=0} \\
&= \pi^{Q-1}(-1)^{Q-1} \frac{1}{(Q-1)!} \int_{\cos(\theta_0)}^{1} dx_N \left(\frac{1}{i}\partial_z^{Q-1} \int_{-\infty}^{+\infty} d\omega \frac{1}{\omega - iz} e^{i\omega\left(1-x_N^2\right)}\right)_{z=0} \\
&= \pi^{Q-1}(-1)^{Q-1} \frac{1}{(Q-1)!} \int_{\cos(\theta_0)}^{1} dx_N \left(\frac{1}{i}\partial_z^{Q-1} \int_{-\infty}^{+\infty} d\omega \frac{1}{\omega - iz} e^{i\omega\left(1-x_N^2\right)}\right)_{z=0} \\
&= \pi^{Q-1}(-1)^{Q-1} \frac{1}{(Q-1)!} \int_{\cos(\theta_0)}^{1} dx_N \left(\frac{1}{i}\partial_z^{Q-1} 2\pi i e^{-z\left(1-x_N^2\right)}\right)_{z=0} \\
&= 2\pi^{Q} \frac{1}{(Q-1)!} \int_{\cos(\theta_0)}^{1} dx_N \left(1 - x_N^2\right)^{Q-1}
\end{aligned}$$

Dropping the subscript N from x_N

$$\begin{aligned}
\int_{\cos(\theta_0)}^{1} dx \left(1 - x^2\right)^{Q-1} &= \sum_{q=0}^{Q-1} \frac{(Q-1)!}{q!\,(Q-1-q)!} \int_{\cos(\theta_0)}^{1} dx x^{2q} \\
&= \sum_{q=0}^{Q-1} \frac{(Q-1)!}{q!\,(Q-1-q)!} (-1)^q \frac{1 - \cos(\theta_0)^{2q+1}}{2q+1}
\end{aligned}$$

Note that

$$\int_0^x dx \left(1 - x^2\right)^{Q-1} = x \times Hypergeometric2F1[\frac{1}{2}, 1-Q, \frac{3}{2}, x]$$

using the notation from Mathematica, but we will stick with the finite series representation.

Example: $N = 3, Q = 1$

$$\begin{aligned}\int_{\cos(\theta_0)}^{1} dx \left(1-x^2\right)^0 &= \int_{\cos(\theta_0)}^{1} dx \\ &= 1-\cos\left(\theta_0\right)\end{aligned}$$

which is correct.

Putting it all together and using $(Q-1)! = \Gamma(Q) = \Gamma\left((N-1)/2\right)$ and remembering we must have $N = odd$ we get

$$\Omega_{N-1}\left(\theta_0\right) = 2\pi^{(N-1)/2}\frac{1}{\Gamma\left((N-1)/2\right)}\sum_{q=0}^{\frac{N-1}{2}-1}\frac{\left(\frac{N-1}{2}-1\right)!}{q!\left(\frac{N-1}{2}-1-q\right)!}\left(-1\right)^q\frac{1-\cos\left(\theta_0\right)^{2q+1}}{2q+1}$$

For a single cone in $N = odd$ dimesions with a (half) opening angle of θ_0, it's surface area covers the following fraction of the full surface area of the sphere

$$\begin{aligned}&\frac{2\pi^{(N-1)/2}\frac{1}{\Gamma((N-1)/2)}\sum_{q=0}^{\frac{N-1}{2}-1}\frac{\left(\frac{N-1}{2}-1\right)!}{q!\left(\frac{N-1}{2}-1-q\right)!}\left(-1\right)^q\frac{1-\cos(\theta_0)^{2q+1}}{2q+1}}{\frac{2\pi^{N/2}}{\Gamma(N/2)}} \\ = \quad &\frac{1}{\sqrt{\pi}}\frac{\Gamma\left(N/2\right)}{\Gamma\left((N-1)/2\right)}\sum_{q=0}^{\frac{N-1}{2}-1}\frac{\left(\frac{N-1}{2}-1\right)!}{q!\left(\frac{N-1}{2}-1-q\right)!}\left(-1\right)^q\frac{1-\cos\left(\theta_0\right)^{2q+1}}{2q+1}\end{aligned}$$

33 Acoustic Pulses

Acoustic Wave Equation (compressible gas)

$$\left(\vec{\partial}^2 - \frac{1}{c^2}\partial_t^2\right)\phi\left(\vec{r},t\right) = 0$$

Here ϕ is the velocity potential, i.e., $\vec{\partial}\phi\left(\vec{r},t\right) = \vec{v}\left(\vec{r},t\right)$ where $\vec{v}\left(\vec{r},t\right)$ is the velocity of the gas atoms/molecules at position $\vec{r}$ and time t. c is the speed of sound.

Consider the 1+1 dimesional problem where an infinite rigid but movable "wall" is placed parallel to the yz plane with it's position along the x axis given by $X(t)$. Let the gas fill the semi-infiinite half space to the right of the wall, i.e., $x > X(t)$. When the wall speed $\partial_t X(t)$ is much less than the speed of sound the atoms/molecules just to the right of the wall must move with the speed of the wall, i.e. $v\left(X(t),t\right) = \partial_t X(t)$. This is only one boundary condition, nominally we need two since the wave equation is second order. But since the wave equation in this case can be factorized as $\left(\partial_x - \partial_t/c\right)\left(\partial_x + \partial_t/c\right)\phi = 0$, we can solve either $\left(\partial_x - \partial_t/c\right)\phi = 0$ or $\left(\partial_x + \partial_t/c\right)\phi = 0$ to get waves/pulses moving in the $-x$ or the $+x$ direction, respectively, hence we only need one boundary condition which is $\left(\partial_x\phi\right)\left(X(t),t\right) \equiv \left(\partial_x\phi\left(x,t\right)\right)_{x=X(t)} = v\left(X(t),t\right) = \partial_t X(t)$.

For $\partial_t X(t) << c$ we have the following approximate solution

$$\phi(x,t) = -cX\left(t - \frac{x - X(t)}{c}\right)$$

NOTE: This solution assumes implicitly that x is always to the right of the wall, i.e,. $x \geq X(t)$.

First show that this satisfies the wave equation up to terms of order $\partial_t X/c$. Let $\xi(x,t) = t - (x - X(t))/c$, then

$$\begin{aligned}\left(\partial_x + \frac{1}{c}\partial_t\right)(-cX(\xi(x,t))) &= -c(\partial_\xi X)(\partial_x \xi) + \frac{1}{c}(-c)(\partial_\xi X)(\partial_t \xi) \\ &= -c(\partial_\xi X)\left(-\frac{1}{c}\right) - (\partial_\xi X)\left(1 + \frac{\partial_t X(t)}{c}\right)\end{aligned}$$

But given $\partial_t X(t)/c << 1$ we have

$$\left(\partial_x + \frac{1}{c}\partial_t\right)(-cX(\xi(x,t))) = \left(\partial_x + \frac{1}{c}\partial_t\right)\left(-cX\left(t - \frac{x - X(t)}{c}\right)\right) \cong 0$$

Show that it satsifies the boundary conditions.

$$\partial_x \phi(x,t)|_{x=X(t)} = \partial_\xi X(\xi(x,t))|_{x=X(t)} = (\partial_\xi X)(t) = \partial_t X(t)$$

The density variation $\delta\rho(x,t)$ about a constant background density, ρ_0, is related to the speed of the molecules via

$$\frac{v(x,t)}{c} = \frac{\delta\rho(x,t)}{\rho_0}$$

and so

$$\frac{\delta\rho(x,t)}{\rho_0} = \frac{\partial_x \phi(x,t)}{c} = -\partial_x X\left(t - \frac{x - X(t)}{c}\right)$$

Example 1: For the wall moving at a constant speed in steady state (all transient effects have propagated away), $X(t) = Vt$, we have

$$\frac{\delta\rho(x,t)}{\rho_0} = -\partial_x\left(V \times \left(t - \frac{x - Vt}{c}\right)\right) = \frac{V}{c}$$

and so all molecules everywhere are moving at the same speed V and so $\partial_x v(x,t) = 0$.

In order to generate acoustic waves or pulses the wall must change speed, i.e., it must accelerate.

Example 2: Consider the wall to have been oscillating forever with radian frequency ω and amplitude X_0 and to be at $x = 0$ at $t = 0$, then

$$X(t) = X_0 \sin[\omega t]$$

The velocity and acceleration are given by

$$\begin{aligned} V(t) &= \partial_t X(t) = X_0 \omega \cos[\omega t] \equiv V_0 \cos[\omega t] \\ &\quad \text{and} \\ A(t) &= \partial_t V(t) = -X_0 \omega^2 \sin[\omega t] \equiv A_0 \sin[\omega t] \end{aligned}$$

respectively. Using the above solution we have that, for $V_0 \ll c$, as required by the above solution, the relative density variation (effectively a pressure wave) is given by

$$\begin{aligned} \frac{\delta\rho(x,t)}{\rho_0} &= -\partial_x \left(X_0 \sin\left[\omega \left(t - \frac{x - X_0 \sin[\omega t]}{c} \right) \right] \right) \\ &= -X_0 \cos\left[\omega \left(t - \frac{x - X_0 \sin[\omega t]}{c} \right) \right] \\ &\quad \times \partial_x \left(\omega \left(t - \frac{x - X_0 \sin[\omega t]}{c} \right) \right) \\ &= \frac{V_0}{c} \cos\left[\omega \left(t - \frac{x - X_0 \sin[\omega t]}{c} \right) \right] \\ &\simeq \begin{cases} \frac{V_0}{c} \cos[\omega t] = \frac{V(t)}{c} & \text{for } x = X(t) \\ \frac{V_0}{c} \cos\left[\frac{\omega}{c} (x - ct) \right] & \text{for } x \gg X_0 \end{cases} \end{aligned}$$

Thus for $x = X(t)$, i.e., "at the wall", the relative density is proportional to $V(t)$. For $x \gg X_0$, i.e., "far away from the wall", we have a wave of relative density with amplitude $V_0/c \ll 1$ moving in the $+x$ direction with speed c and oscillating with radian frequency ω, exactly as expected.

Example 3: Suppose the wall is motionless up to time $t = 0$ and then starts oscillating sinusoidally at $t = 0$, i.e.,

$$X(t) = X_0 \sin[\omega t]\, \theta(t)$$

Here $\theta(t)$ is the Heaviside step function: $\theta(t) = 0$ for $t < 0$ and $\theta(t) = 1$ for $t > 0$.

Comment: In many problems specifying the value of $\theta(0)$ is not necessary. In those problems where it is necessary to know $\theta(0)$, the value to choose is usually dictated by the problem itself. Common choices are $\theta(0) = 0$ or $1/2$ or 1.

Substituting into the solution above we have

$$\begin{aligned} \frac{\delta\rho(x,t)}{\rho_0} &= -\partial_x \left(X_0 \sin\left[\omega \left(t - \frac{x - X_0 \sin[\omega t]\,\theta(t)}{c} \right) \right] \theta\left(t - \frac{x - X_0 \sin[\omega t]\,\theta(t)}{c} \right) \right) \\ &= \frac{V_0}{c} \cos\left[\omega \left(t - \frac{x - X_0 \sin[\omega t]}{c} \right) \right] \theta\left(t - \frac{x - X_0 \sin[\omega t]}{c} \right) \\ &\quad + \frac{X_0}{c} \sin\left[\omega \left(t - \frac{x - X_0 \sin[\omega t]}{c} \right) \right] \delta_D\left(t - \frac{x - X_0 \sin[\omega t]}{c} \right) \end{aligned}$$

At the end we have assumed $t > 0$ so that $\theta(t) = 1$. Also we have used the general relation $\partial_\alpha \theta(\alpha) = \delta_D(\alpha)$ which holds for any variable α.

The Dirac delta function δ_D in the second term sets the sine factor multiplying it to zero and so finally we have

$$\frac{\delta\rho(x,t)}{\rho_0} = \frac{V_0}{c}\cos\left[\omega\left(t - \frac{x - X_0\sin[\omega t]}{c}\right)\right]\theta\left(t - \frac{x - X_0\sin[\omega t]}{c}\right)$$

For $t > 0$ we have for x "at the wall", $x = X(t) = X_0\sin[\omega t]$ and therefore "at the wall" we have

$$\frac{\delta\rho(x,t)}{\rho_0} = \frac{V_0}{c}\cos[\omega t]\,\theta(t) = \frac{V_0}{c}\cos[\omega t]$$

just as in the previous example.

For $x \gg X(t) = X_0\sin[\omega t]$ we have

$$\begin{aligned}\frac{\delta\rho(x,t)}{\rho_0} &\simeq \frac{V_0}{c}\cos\left[\omega\left(t - \frac{x}{c}\right)\right]\theta\left(t - \frac{x}{c}\right)\\ &\simeq \begin{cases}\frac{V_0}{c}\cos\left[\omega\left(t - \frac{x}{c}\right)\right] & \text{for } x < ct\\ 0 \text{ for } x > ct\end{cases}\end{aligned}$$

This again is exactly as expected. The wall started oscillating at $t = 0$ and the density (or if you prefer, pressure) waves travel at speed c. For $x > ct$, the waves haven't arrived there yet and so $\delta\rho/\rho_0$ should be zero. To put this result differently, the front end of the wave train is at $x = ct$.

34 Navier-Stokes, Sound, Bernoulli

$$\begin{aligned}&\text{Navier-Stokes} \quad :\\ \partial_t\vec{v} + \left(\vec{v}\cdot\vec{\partial}\right)\vec{v} &= -\vec{\partial}p + \mu\vec{\partial}^2\vec{v}\end{aligned}$$

$$\begin{aligned}\vec{v} &= \vec{v}(\vec{r},t) = \text{local fluid flow velocity } (\mathrm{m/s})\\ \vec{r} &= \text{position}\\ t &= \text{time}\\ \rho_0 &= \text{mean or average density } (\mathrm{kg/m^3}) = \text{constant}\\ p &= p(\vec{r},t) = \frac{P(\vec{r},t)}{\rho_0} = \text{"reduced pressure" } \left(\mathrm{m^2/s^2}\right)\\ P(\vec{r},t) &= \text{pressure } (\mathrm{N/m^2} = \text{Pascal})\\ \mu &= \text{kinematic viscosity} = \frac{\mu_D}{\rho_0}\left(\mathrm{m^2/s}\right)\\ \mu_D &= \text{dynamic viscosity } (\mathrm{kg/(m\times s)})\end{aligned}$$

Given these definitions every term in the Navier-Stokes equation has units of $\mathrm{m/s^2}$.

$$\begin{aligned} \text{Continuity Equation} \quad &: \\ \partial_t \rho(\vec{r},t) \quad &= \quad -\vec{\partial}\cdot(\rho(\vec{r},t)\,\vec{v}(\vec{r},t)) \end{aligned}$$

where $\rho(\vec{r},t)$ = local density $(\mathrm{kg/m^3})$

34.1 Sound Waves

Let $\rho = \rho_0 + \delta\rho$. The continuity equation then becomes

$$\partial_t\left(\frac{\delta\rho}{\rho_0}\right) = -\vec{\partial}\cdot\vec{v} + \text{ H.O.T.}$$

where H.O.T. stands for Higher Order Terms.

Taking the position and time dependence of $P(\vec{r},t)$ to depend only on the position and time dependence of $\rho(\vec{r},t)$, i.e., assuming that $P(\vec{r},t) \to P(\rho(\vec{r},t))$, we have

$$P(\vec{r},t) = P_0 + \left(\frac{\partial P(\vec{r},t)}{\partial\rho}\right)_{\rho=\rho_0}\delta\rho(\vec{r},t) + \text{H.O.T.}$$

where $P_0 = P(\rho_0) =$ constant. Note that $P(\vec{r},t) \to P(\rho(\vec{r},t))$ yields

$$\left(\frac{\partial P(\vec{r},t)}{\partial\rho}\right)_{\rho=\rho_0} \equiv c^2 = \text{constant (units} = \mathrm{m^2/s^2})$$

NOTE: For many/most gases including an ideal gas (at constant temperature), $P \sim \rho$ and so $\partial P(\vec{r},t)/\partial\rho = P(\vec{r},t)/\rho$. This relation holds if temperature of the gas is unform and constant, i.e., isothermal fluctuations. But if the fluctuations in pressure and density are so fast there can be no flow of heat, i.e., they are adiabatic, then the gas temperature must fluctuate as well in which case we have $p \sim \rho^\gamma$ where γ is the ratio of the two specific heats of the gas. This is the more common way of expressing the speed of sound.

As we now show c is the speed of sound.

As defined above

$$p(\vec{r},t) = \frac{P(\vec{r},t)}{\rho_0} = \frac{P_0}{\rho_0} + c^2\frac{\delta\rho(\vec{r},t)}{\rho_0} + \text{H.O.T.}$$

and so

$$\vec{\partial}p(\vec{r},t) = c^2\vec{\partial}\frac{\delta\rho(\vec{r},t)}{\rho_0}$$

In the linear regime, i.e., when $\left(\vec{v}\cdot\vec{\partial}\right)\vec{v} \simeq 0$, the Navier-Stokes equation reduces to

$$\partial_t\vec{v} = -c^2\vec{\partial}\frac{\delta\rho}{\rho_0} + \mu\vec{\partial}^2\vec{v}$$

Taking the divergence, $\left(\vec{\partial}\cdot\right)$, yields.

$$\partial_t \vec{\partial}\cdot\vec{v} = -c^2\vec{\partial}^2\frac{\delta\rho}{\rho_0} + \mu\vec{\partial}^2\left(\vec{\partial}\cdot\vec{v}\right)$$

But from the continuity equation as written above, ignoring the H.O.T., we have $\vec{\partial}\cdot\vec{v} = -\partial_t\left(\delta\rho/\rho_0\right)$. Substituting this relation yields

$$-\partial_t^2\left(\frac{\delta\rho}{\rho_0}\right) = -c^2\vec{\partial}^2\left(\frac{\delta\rho}{\rho_0}\right) - \mu\vec{\partial}^2\partial_t\left(\frac{\delta\rho}{\rho_0}\right)$$

Rearraging gives

$$\left(\partial_t^2 - c^2\vec{\partial}^2 + \mu\partial_t\vec{\partial}^2\right)\left(\frac{\delta\rho}{\rho_0}\right) = 0$$

which is the wave equation for sound (= pressure or density waves) which travel with speed c and are damped by the term proportional to the kinematic viscosity μ.

34.2 Bernoulli

First, for simplicity consider 1D, i.e., x axis only, assume steady state, i.e., $\partial_t v = 0$, use $v\partial_x v = \partial_x\left(\frac{1}{2}v^2\right)$ and $p = P/\rho_0$ then the Navier-Stokes equation reduces to

$$\partial_x\left(P + \frac{1}{2}\rho_0 v^2 - \mu_D\partial_x v\right) = 0$$

which implies that

$$P + \frac{1}{2}\rho_0 v^2 - \mu_D\partial_x v = \text{constant}$$

This is Bernoullis equation for 1D fluid flow.

Now consider Navier-Stokes assuming steady state and incompressible flow in which case it reduces to

$$\left(\vec{v}\cdot\vec{\partial}\right)\vec{v} = -\frac{1}{\rho}\vec{\partial}p + \frac{\mu}{\rho}\ \vec{\partial}^2\vec{v}$$

with $\vec{\partial}\cdot\vec{v} = 0$ because of incompressiblility.

With the assumption that $\vec{v}$ is "curl free", i.e.,

$$\vec{\partial}\times\vec{v} = 0$$

it then follows that $\vec{v}$ can be written as the gradient of a scalar field, i.e.,

$$\vec{v} = \vec{\partial}\phi$$

In component notation, where the i, j subscripts label the x, y, z components of vectors we have

$$\begin{aligned}\left(\left(\vec{v}\cdot\vec{\partial}\right)\vec{v}\right)_i &= v_j\partial_j v_i \\ &= \left(\partial_j\phi\right)\partial_j\partial_i\phi \\ &= \partial_j\left(\frac{1}{2}\left(\partial_i\phi\right)^2\right) \\ &\rightarrow \vec{\partial}\left(\frac{1}{2}\left(\vec{\partial}\phi\right)^2\right)\end{aligned}$$

Hence under these assumptions Navier-Stokes reduces to

$$\vec{\partial}\left(p+\frac{\rho}{2}\left(\vec{\partial}\phi\right)^2-\mu\ \vec{\partial}^2\phi\right)=0$$

and so we have

$$p+\frac{\rho}{2}\left(\vec{\partial}\phi\right)^2-\mu\ \vec{\partial}^2\phi=\text{constant}$$

This is Bernoulli in 3D written in terms of the "potential", ϕ. If viscosity can be ignored then we have

$$p+\frac{\rho}{2}\vec{v}^2=\text{constant}$$

which is the more standard version.

35 Nonlinear Harmonic Oscillator

In 1D, the standard freely oscillating harmonic oscillator is described by

$$\left(\partial_t^2+\omega^2\right)x\left(t\right)=0$$

where $x\left(t\right)$ is the position of the oscillator at time t and ω is its natural frequency of oscillation in radians/second, i.e, $2\pi f$ where f is the oscillation frequency in Hertz.

A damped driven harmonic oscillator is described by

$$\left(\partial_t^2+\gamma\partial_t+\omega^2\right)x\left(t\right)=F\left(t\right)$$

where γ represents the damping, i.e., friction, and $F\left(t\right)$ is the driving force.

Note that the above differential equation is second order, ∂_t^2, and so requires two conditions (two initial or two final or one each) to specify a specific solution.

Consider adding a nonlinear term $\varepsilon x\left(t\right)^3$ to the above equation. We now have a nonlinear damped driven harmomic oscillator,

$$\left(\partial_t^2+\gamma\partial_t+\omega^2\right)x\left(t\right)+\varepsilon x\left(t\right)^3=F\left(t\right)$$

Let $F\left(t\right)=A\sin\left(t\right)$ and $\omega^2=1+\nu$. For $\nu=0$ the natural frequency and the drive frequency are both 1 and hence ν relates to the detuning of the natural frequency from the drive frequency.

Ignore transients and just consider steady state oscillation. In steady state both sides of the above equation oscillate at the same frequency in which case we can set $x(t) = X\sin(t+\phi)$. The goal is to find how X, the amplitude of oscillation depends on A, the amplitude of the driving force. Substituting we get

$$\nu X\sin(t+\phi) + \gamma X\cos(t+\phi) + \varepsilon X^3\sin(t+\phi)^3 = A\sin(t)$$

NOTE: This equation has an issue. The left hand side has terms oscillating at both 1 and 3 radians/(unit time). The right hand side oscillates only at 1 radian/(unit time). So there is an inconsistency, but since we are just trying to get a general picture of how X depends on A, we will proceed.

Use the identities

$$\begin{aligned}
\sin(t+\phi) &= \cos(t)\sin(\phi) + \sin(t)\cos(\phi) \\
\cos(t+\phi) &= \cos(t)\cos(\phi) - \sin(t)\sin(\phi) \\
\sin(t+\phi)^3 &= \frac{3}{4}\sin(t+\phi) - \frac{1}{4}\sin(3(t+\phi))
\end{aligned}$$

Then, to eliminate the dependence on t and $3t$, multiply the above equation by $\sin(t)$ and integrate over one period which gives

$$\nu X\cos(\phi) - \gamma X\sin(\phi) + \frac{3}{4}\varepsilon X^3\cos(\phi) = A$$

Repeat, but this time multiply by $\cos(t)$ and integrate over one period, which gives

$$\nu X\sin(\phi) + \gamma X\cos(\phi) + \frac{3}{4}\varepsilon X^3\sin(\phi) = 0$$

To eliminate the ϕ dependence, add the squares of these last two equations. This gives

$$\left(\frac{3}{4}\varepsilon X^3 + \nu X\right)^2 + (\gamma X)^2 = A^2$$

For $\varepsilon = 0$ this is the standard harmonic oscillator result

$$X(\nu) = \frac{A}{\sqrt{\nu^2+\gamma^2}}$$

And so $X(\nu)$ for $\varepsilon = 0$ is the standard Lorentizian-like curve or resonance with a peak at $\nu = 0$ and width on the order of γ. For $\varepsilon \neq 0$, as A increases the peak curls over like a wave breaking on a beach and at some critical value of A, for a given ε, you have 3 real valued solutions for X. This leads to hysteresis wherein X jumps from one solution to another as you sweep through the curled over peak.