Chem. 221 Notes (adapted from notes by Professor Alex .J. Levine)

Overview: These notes are a mixture of typing up of notes by Professor Alex Levine (~75-80%) with some extras or modifications.

The notes cover much of the minimum mathematics needed for physical chemists:

- Linear algebra, matrices and bra-kets, linear equations and least square fitting.
- Ordinary differential equations, special functions.
- Complex integration.
- Group theory – finite groups and Lie groups.
- Delta Functions, Continuous basis-sets, Fourier transforms and integrals

There are several missing parts, which may be inserted in later versions or covered in your other classes; the most important ones are:

- Partial differential equations (elliptical and hyperbolic), Sturm Liouville forms
- Probability
- Numerical algorithms: especially Monte Carlo and Metropolis forms.
- And for the advanced: Field theory (really part of physics)

The two recommended references are:

- Boas: Mathematical Methods in the Physical Sciences – very clear
- Mathews and Walker, Mathematical methods of Physics – my favorite, more condensed than Boas but more lecture-notes-like so "easily flows", and covers some material not in Boas.

Alternately, one could use:

- Arfken – Mathematical Methods for Physicists -- Very comprehensive.

Linear algebra:

Overview:

1. Vectors and Hilbert spaces, dot products

2. Coordinate transformations, matrices, rotation (orthogonal) matrices.

3. Bra-ket notation, completeness

4. Applying several operations or matrices

5. Determinants and Matrix Inverses, the Levi Civitta symbol

6. Hermitian conjugates. Hermitian and Unitary matrices.

7. Diagonalization of matrices.

8. Hermitian matrices – real eigenvalues, orthogonal eigenvectors.

9. Traces.

10. Functions of matrices $f(M) = Vf(\lambda)V^{-1}$

11. Non-orthogonal basis sets: orthogonalization (Grahm Schmidt, symmetric Diagonalization); or Generalized Diagonalization.

12. Linear homogenous and inhomogeneous equations; regularization terms; variational solution, and least square fitting.

Basis of unit vectors:

Vector: arrow with magnitude and direction; can be added. We are familiar with them in 2D and 3D, but they are general.

The starting point is a **basis**; in 3D it will be $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ (what we usually call a unit vectors in the x,y,z directions) and in general, for a d-dimensional space, it will be $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \ldots, \mathbf{e}_d$.

Further, we assume for most of the chapter that the basis is orthogonal and even orthonormal (later on we'll see how to take a general, non-orthogonal basis, and orthogonalize it). What orthonormal means, is that we have a **dot-product**, and the "dot product" of the unit vectors is 1 with themselves and zero with each other

$$\mathbf{e}_1 \bullet \mathbf{e}_1 = 1$$
$$\mathbf{e}_1 \bullet \mathbf{e}_2 = 0$$
$$\mathbf{e}_2 \bullet \mathbf{e}_2 = 1$$
$$etc.$$

Or generally

$$\mathbf{e}_i \bullet \mathbf{e}_j = \delta_{ij}$$

Where we defined the Kronecker delta

$$\delta_{ij} \equiv \begin{cases} 1 & \text{i=j} \\ 0 & \text{else} \end{cases}$$

This basis will give us what we call a coordinate system.

Representing vectors as list of numbers:

Given a coordinate system, associate a set of d numbers with the d-dimensional vector:

$$\boldsymbol{r} \text{ associated with: } (x_1, \ldots, x_d)$$

For 3D, for example, write:

$$\mathbf{r} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$$

(note that "1" refers to x-axis, 2: the y-axis, and 3: the z-axis.)

More generally we'll write

$$\mathbf{r} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 + \ldots = \sum_{j=1}^{d} x_j\mathbf{e}_j$$

Or even abbreviate using the:

Einstein summation convention **(repeated indicies are summed over)**

$$\mathbf{r} = x_j\mathbf{e}_j \equiv \sum_{j=1}^{d} x_j\mathbf{e}_j$$

Further, we allow **complex** coefficients and vectors.

Insert: Hilbert spaces. Everything we consider will be relevant to basis sets which are "infinite-dimensional" – "Hilbert spaces"; while this sounds scary, consider the orbitals of a 1D harmonic oscillator, or the orbitals of hydrogen atom; to describe a general function of space, $\psi(x)$, we can expand our function in terms of such orbitals, $\phi_j(x)$, i.e., $\psi(x) = \sum_j b_j\phi_j(x)$. Then, the coefficients, $(b_1, b_2, \ldots)$ can be viewed as coefficients of a vector.

The number of terms in such an expansion can be either infinite, formally, or if we expand numerically the number of terms will be finite. For example, if we expand a 1D function, or a 3D function, in terms of 1000 basis functions, then we are using a vector space of length d=1000. This vector space is much "bigger" of course then the space associated with 1D or 2D or 3D vectors, which have dimensions d=1, 2 and 3, respectively, but otherwise is similar.

Extracting the coefficients:

Given

$$\mathbf{r} = \sum_{j=1}^{d} x_j\mathbf{e}_j$$

we can extract the coefficient x_k by dot-product with $\mathbf{e}_k$:

$$\mathbf{e}_k \bullet \mathbf{r} = \mathbf{e}_k \bullet \sum_{j=1} x_j\mathbf{e}_j = \sum_{j=1} x_j\mathbf{e}_k \bullet \mathbf{e}_j$$

$$= \sum_{j=1} x_j\delta_{jk} = x_k \quad \text{(at the only j term that contributes is the k term)}$$

Note how this is easier with the Einstein summation convention,

$$\mathbf{e}_k \bullet \mathbf{r} = \mathbf{e}_k \bullet x_j\mathbf{e}_j = x_j\mathbf{e}_k \bullet \mathbf{e}_j = x_j\delta_{jk} = x_k$$

Either way, we find

$$x_k = \mathbf{e}_k \bullet \mathbf{r}$$

Dot product of 2 vectors

Given the definition, we can calculate a dot-product operation. For 2-D vectors, **and let's assume that the coefficients could be complex, we complex conjugate the left vector**:

$$\mathbf{A} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2$$
$$\mathbf{B} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2$$
$$\mathbf{A}\bullet\mathbf{B} = \left(a_1\mathbf{e}_1 + a_2\mathbf{e}_2\right)^* \bullet \left(b_1\mathbf{e}_1 + b_2\mathbf{e}_2\right)$$
$$= a_1^*b_1\mathbf{e}_1\bullet\mathbf{e}_1 + a_1^*b_2\mathbf{e}_1\bullet\mathbf{e}_2 + a_2^*b_1\mathbf{e}_2\bullet\mathbf{e}_1 + a_2^*b_2\mathbf{e}_2\bullet\mathbf{e}_2$$
$$= a_1^*b_1\cdot 1 + a_1^*b_2\cdot 0 + a_2^*b_1\cdot 0 + a_2^*b_2\cdot 1$$
$$= a_1^*b_1 + a_2^*b_2$$
$$= a_i^*b_i$$

And in general the proof is the same:

$$\mathbf{A}\bullet\mathbf{B} = \sum_i a_i^*\mathbf{e}_i \bullet \sum_j b_j\mathbf{e}_j = \sum_{i,j} a_i^*b_j\mathbf{e}_i\bullet\mathbf{e}_j = \sum_{i,j} a_i^*b_j\delta_{ij} = \sum_{i,j} a_i^*b_i$$

And with the repeated-index summation convention

$$\mathbf{A}\bullet\mathbf{B} = a_i^*b_i$$

Magnitude of vectors:

Note that we need the magnitude if vectors. In 3-D real vectors have the norm:

$$|\boldsymbol{r}| = \sqrt{x_1^2 + x_2^2 + x_3^2}.$$

For complex vectors, we need a conjugation. For example, even in 1D, consider the vector

$$\mathbf{r}_{1d} = (1+i)\mathbf{e}_1 \quad \text{i.e., } \mathbf{r}_{1d} = a_1\mathbf{e}_1 \quad \text{where } a_1 = 1+i$$

The magnitude of that vector is

$$|\,\mathbf{r}_{1d}\,| = \sqrt{a_1^*a_1} = \sqrt{(1-i)(1+i)} = \sqrt{1^2+1^2} = \sqrt{2}$$

This is why we took the complex conjugation – so the norm is real and positive.

The dot-product of a vector with itself is therefore squared norm. For 3D vectors

$$\mathbf{r}^*\bullet\mathbf{r} = \left(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3\right)^* \bullet \left(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3\right)$$
$$\mathbf{r}^*\bullet\mathbf{r} = x_i^*x_i = |\,\mathbf{r}\,|^2$$

And in general

$$|\mathbf{r}|==\left(\sum_i |x_i|^2\right)^{\frac{1}{2}}=\left(\sum_{i=1}^{d} x_i^* x_i\right)^{\frac{1}{2}}=\left(x_i^* x_i\right)^{\frac{1}{2}}.$$

where in the last equality we again used the Einstein summation convention.

A side note on generalized norms:

Formally, the norm we use this is one member of a general family of norms, L_k norms, defined as

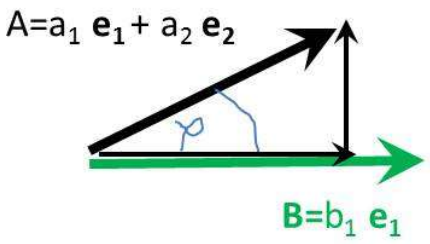
$$|\mathbf{r}|_{L_k} \equiv \left(\sum_{i=1}^{d} |x_i|^k\right)^{\frac{1}{k}}.$$

The norm as we know it is L_2; However, many modern applications benefit from using the L_1 norm, $\sum_{i=1}^{d} |x_i|$,and much of the advances in using the L_1 norm were made in the UCLA math. department since 2000.

End of side note (From now on norm will be for us only the L_2 norm).

Back to regular norms.

Geometrical dentition of dot product:



As you know, the dot-product of two vectors is the norm of the two vectors times the cosine of their angle (this can be proved as we'll do for 2-D -- and also 3-D -- space, but is actually the definition of a cosine of an angle in higher order space)

Let's see that in 2D.

Given A and B, take the x-axis along, say, B. (see figure). In this subsection, consider the vectors real, for simplicity.

Then:

$$\mathbf{A}\bullet\mathbf{B}=\left(a_1\mathbf{e}_1+a_2\mathbf{e}_2\right)\bullet b_1\mathbf{e}_1=a_1b_1$$

But

$b_1 = |\,\mathbf{B}\,|$

$a_1 = |\,\mathbf{A}\,|\cos\phi$ (see Figure). So:

$$\mathbf{A}\bullet\mathbf{B}=|\mathbf{A}||\mathbf{B}|\cos\phi$$

Q.E.D.

Better definition of vectors:

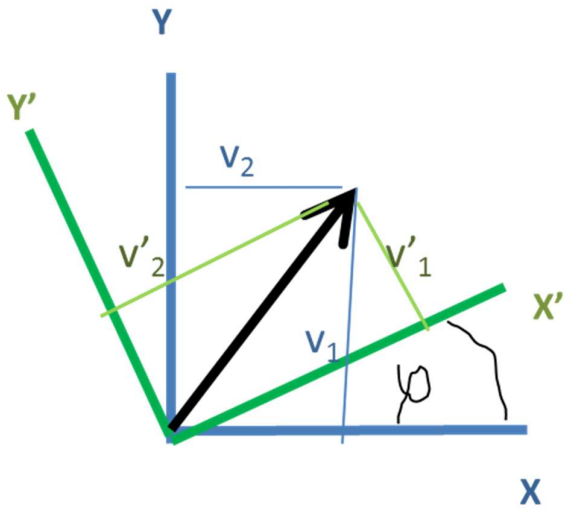
A vector is an n-tuple of #'s that transform in a specific way under rotation, as explained later.

Formally: vectors are "tensors of rank 1". (We will not discuss tensors – but they are important and learn about them later).

Example: coordinate transformation

Let's specialize first, for simplicity, to 2D **REAL** vectors

Consider same vector in 2 diff. coordinate systems: (See figure).



i.e., relate the representation of the vector, i.e., its coefficients, for two different basis sets:

$$\mathbf{v}=v_1\mathbf{e}_1+v_2\mathbf{e}_2$$
$$\mathbf{v}=v'_1\,\mathbf{e}'_1+v'_2\,\mathbf{e}'_2$$

Two diff. sets of #'s, v'_1, v'_2 and. v_1, v_2, both represent the same physical vector.

Note that (see figure):

$$\mathbf{e}'_1\bullet\mathbf{e}_1=\cos\phi$$
$$\mathbf{e}'_1\bullet\mathbf{e}_2=\sin\phi$$

The simplest way to obtain the coordinate transformation (i.e., to obtain v'_1, v'_2 from v_1, v_2) is algebraic:

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 = v'_1\,\mathbf{e}'_1 + v'_2\,\mathbf{e}'_2$$
$$v'_1 = \mathbf{e}'_1 \bullet \mathbf{v}$$
$$v'_1 = \mathbf{e}_1\,'\bullet\left(v_1\mathbf{e}_1 + v_2\mathbf{e}_2\right) = v_1\mathbf{e}_1\,'\bullet\mathbf{e}_1 + v_2\mathbf{e}_1\,'\bullet\mathbf{e}_2 = v_1\cos\phi + v_2\sin\phi$$

And similarly, $v'_2 = v_2 cos\phi - v_1 sin\phi$

Coordinate transformation through matrices

Using the algebraic form we then write, in general (following the 2nd, algebraic way):

$$\mathbf{v} = v_k\mathbf{e}_k$$
$$\mathbf{v} = v'_j\,\mathbf{e}'_j$$
$$v'_l = \mathbf{e}'_l \bullet \mathbf{v} = \mathbf{e}'_l \bullet v_k\mathbf{e}_k = \mathbf{e}'_l \bullet \mathbf{e}_k v_k$$

i.e., defining:

$$M_{lk} = \mathbf{e}'_l \bullet \mathbf{e}_k$$

we get:

$$v'_l = M_{lk}v_k = \sum_k M_{lk}v_k$$

I.e., in matrix form we write:

$$v' = Mv$$

where v' and v are now column vectors, and M is a matrix. In pictorial form,

$$\begin{pmatrix} \square \end{pmatrix} = \begin{pmatrix} \square & \square & \square & \square \\ & & & \end{pmatrix}\begin{pmatrix} \square \\ \square \\ \square \\ \square \end{pmatrix}$$

v' M v

I.e., the k'th element of the column vector v' is obtained by summing element by element the k'th row times the full column v.

Note that the **rotation we discuss not a rotation of the vector itself, v**, instead what we rotate is the coordinate system used. (Later we'll consider rotations of vectors).

In 2D this is quite interesting; we convert the equations we saw ($v'_1 = v_1 \cos\phi + v_2 \sin\phi, \quad v'_2 = -v_1 \sin\phi + v_2 \cos\phi$) to a matrix form

$$\begin{pmatrix} v'_1 \\ v'_2 \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

I.e., the rotation matrix is

$$M = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}$$

You can verify that indeed this rotation matrix is the same as the general expression, $M_{lk} = \mathbf{e}'_l \bullet \mathbf{e}_k$:,

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} = \begin{pmatrix} \mathbf{e}'_1 \bullet \mathbf{e}_1 & \mathbf{e}'_1 \bullet \mathbf{e}_2 \\ \mathbf{e}'_2 \bullet \mathbf{e}_1 & \mathbf{e}'_2 \bullet \mathbf{e}_2 \end{pmatrix}$$

Back to M: M is the transformation matrix between the old and new coordinate system.

Bra-ket notation:

We are going to introduce a notation that you may find cumbersome and unnecessary at this stage, but will be very important later: the bra-ket notation. We'll use it interchangeably with the regular notation. It is extremely important when we consider complex vectors and functions, and makes quantum mechanics much more elegant, as Dirac has shown (after finishing this course read his 1930 Quantum Mechanics book – the elegance of the presentation and ideas is sheer enjoyment).

We will introduce it for real and complex vectors.

In the bra-ket notation vectors are denoted by $|\mathbf{v}\rangle$ (ket), and a conjugate of a vector by $\langle\mathbf{v}|$ (a bra).

The ket $|\mathbf{v}\rangle$ is really a vector, i.e., is independent of the coordinate system used.

The power of the bra-ket notation comes when we look at the dot-product.

For general complex vectors, their dot product is

$$\langle\mathbf{a}|\mathbf{b}\rangle = \mathbf{a}^* \bullet \mathbf{b}$$

Note that the order determines the result:

$$\langle\mathbf{a}|\mathbf{b}\rangle = \langle\mathbf{b}|\mathbf{a}\rangle^*$$

Linearity

The ket is, for us, a vector, so it is linear, i.e.,

$$|a\mathbf{a} + b\mathbf{b}\rangle = a|\mathbf{a}\rangle + b|\mathbf{b}\rangle$$

Note that the bra is not exactly linear, i.e., the coefficients are complex conjugated, **i.e.,**

$\langle a\mathbf{a}| = a^* \langle \mathbf{a}|$

The reason is that the bra is essentially a complex conjugated vector, as we saw.

Let's say that we're in a 2D space. Then, we can use a coordinate system defined by two kets, $|\mathbf{e}_1\rangle, |\mathbf{e}_2\rangle$. These vectors are orthogonal, so

$$\langle \mathbf{e}_1 | \mathbf{e}_1 \rangle = \mathbf{e}_1 \bullet \mathbf{e}_1 = 1$$
$$\langle \mathbf{e}_1 | \mathbf{e}_2 \rangle = 0$$
etc.

Write a general vector in this 2D space, i.e., a general ket, as

$$|\mathbf{v}\rangle = v_1 |\mathbf{e}_1\rangle + v_2 |\mathbf{e}_2\rangle$$

This is of course completely equivalent to $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2$. We allow the vector to be complex, so the coefficients can be complex.

"Dotting" by the basis-set gives

$$\langle \mathbf{e}_1 | \mathbf{v} \rangle = \langle \mathbf{e}_1 | v_1\mathbf{e}_1 + v_2\mathbf{e}_2 \rangle = v_1 \langle \mathbf{e}_1 | \mathbf{e}_1 \rangle + v_2 \langle \mathbf{e}_1 | \mathbf{e}_2 \rangle = v_1$$

Similarly

$$\langle \mathbf{e}_2 | \mathbf{v} \rangle = v_2$$

Analogously, for a general basis fulfilling

$$\langle \mathbf{e}_i | \mathbf{e}_j \rangle = \delta_{ij},$$

the coefficients of a general vector

$$|\mathbf{v}\rangle = v_j |\mathbf{e}_j\rangle$$

are obtained as

$$v_j = \langle \mathbf{e}_j | \mathbf{v} \rangle.$$

Note the relation between a ket and its representation. Often, we confuse it. For example, rewriting what we did on dot products, we get

$$\langle \mathbf{u} | \mathbf{v} \rangle = \langle u_j \mathbf{e}_j | u_k \mathbf{e}_k \rangle = u_j^* v_k \langle \mathbf{e}_j | \mathbf{e}_k \rangle,$$

i.e.,

$$\langle \mathbf{u} | \mathbf{v} \rangle = u_j^* v_j$$

If we define a 2*1 matrix (which we colloquially call a vector):

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$
$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$,

then, define the Hermitian conjugate of the column vector as the complex conjugate of the transpose of the vector (the same definition will work below for matrices), i.e.,

$$u^\dagger = \begin{pmatrix} u_1^* & u_2^* \end{pmatrix}$$

Note that u was a 2*1 matrix, called column vector, and $u^\dagger$ is a 1*2 matrix, a row vector.

So we can write the dot product as

$$\langle \mathbf{u} | \mathbf{v} \rangle = u^\dagger v = \begin{pmatrix} u_1^* & u_2^* \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Note that v is not really a vector in the same sense that the ket |**v**> is, since it depends on coordinates; it is merely the **representation** of the ket in a basis-set. But we will often go back and forth and call v, the 2*1 tuple of numbers, a vector.

Finally, the dot product is the norm squared, as we saw:

$$\langle \mathbf{v} | \mathbf{v} \rangle = v^\dagger v = v_j^* v_j = \sum_j | v_j |^2$$

All the expressions we derived are valid for a general case, of course, of a d-dimensional vector.

Completeness

Finally, we can derive a very useful (and especially transparent expression in the bra-ket notation) of completeness:

$$\sum_j |\mathbf{e}_j\rangle\langle\mathbf{e}_j| = 1 \qquad (\text{or } |\mathbf{e}_j\rangle\langle\mathbf{e}_j| = 1 \quad \text{in the Einsten's summation convention})$$

Proof:

Apply this to an arbitrary vector

$$\mathbf{c} = \sum_k c_k |\mathbf{e}_k\rangle$$

And we'll get the same initial vector

$$\sum_j |\mathbf{e}_j\rangle\langle\mathbf{e}_j|\mathbf{c} = \sum_{kj} c_k |\mathbf{e}_j\rangle\langle\mathbf{e}_j|\mathbf{e}_k\rangle = \sum_{kj} c_k |\mathbf{e}_j\rangle\delta_{jk} = \sum_{kj} c_k |\mathbf{e}_k\rangle = \mathbf{c}\,,\ \textit{Q.E.D.}$$

Matrices and operators in the bra-ket notation:

A general operator on kets or vectors is defined as a **mapping** from one vector to another:

$$\mathbf{L}|\mathbf{u}\rangle = |\mathbf{v}\rangle$$

We'll usually limit ourselves to linear operators. Linear operators are those that fulfill:

$$\mathbf{L}|a\mathbf{a}+b\mathbf{b}\rangle = a\mathbf{L}|\mathbf{a}\rangle + b\mathbf{L}|\mathbf{b}\rangle$$

We can write linear operator **<u>through a linear matrix,</u>** as follows (we use in the derivation the Einstein summation convention):

Starting as:

$|\mathbf{u}\rangle \equiv \mathbf{L}|\mathbf{v}\rangle$

We then insert

$|\mathbf{v}\rangle = v_k |\mathbf{e}_k\rangle$

giving

$|\mathbf{u}\rangle = \mathbf{L}v_k |\mathbf{e}_k\rangle = v_k \mathbf{L}|\mathbf{e}_k\rangle$

But since $|\mathbf{u}\rangle$ is a ket, we can also expand it

$|\mathbf{u}\rangle = u_j |\mathbf{e}_j\rangle$

where the coefficients are obtained by "dotting" by a bra:

$u_j = \langle\mathbf{e}_j|\mathbf{u}\rangle = \langle\mathbf{e}_j|v_k\mathbf{L}|\mathbf{e}_k\rangle = \langle\mathbf{e}_j|\mathbf{L}|\mathbf{e}_k\rangle v_k$

i.e.,

$$u_j = L_{jk} v_k$$

where the matrix L_{jk} is related to the operator as

$$L_{jk} = \langle\mathbf{e}_j|\mathbf{L}|\mathbf{e}_k\rangle$$

In matrix form, we can write this as

$$u = Lv$$

where u is a column vector, L is a matrix, and v is a column vector:, as we saw.

$$\begin{pmatrix} \square \end{pmatrix} = \begin{pmatrix} \square & \square & \square & \square \\ & & & \end{pmatrix} \begin{pmatrix} \square \\ \square \\ \square \\ \square \end{pmatrix}$$

u L v

The ket and vector and matrix notations are so intertwined, we'll use them interchangeably

So to summarize: We have 3 ways to represent vectors

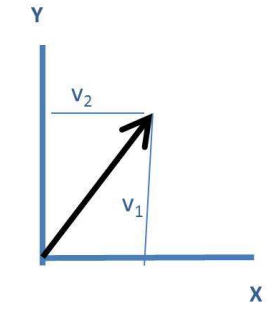
- As abstract vectors, **v**
- As abstract kets, |**v**>
- As column vectors, i.e., d*1 matrices.

And similarly operators can be viewed in 2 ways:

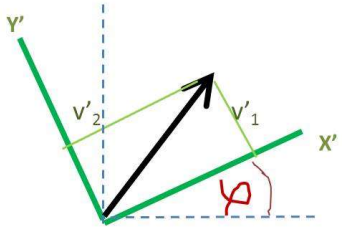
- As abstract operators
- As d*d matrices

Note on operations on vectors and coordinate transformation:

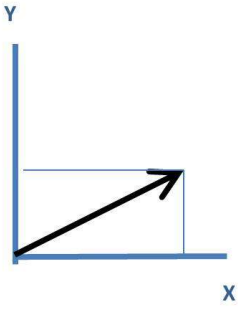
A fine point which can get lost is the difference between operating on a vector and coordinate transformation. When we operate on a vector, for example rotate it, the vector itself changes; when we do a coordinate transformation, the vector does not change but the basis set does. See figure for the difference.



I. Original vector



II. Same vector in rotated coordinate system



III. Now, instead, coordinate system fixed but vector rotated in opposite direction ($-\phi$) -- same coefficients as in II.

Example for operator: rotation operator

Let's consider the linear operator of rotating a vector by an angle ϕ, $\mathbf{R}_\phi$, which we abbreviate often just as **R**.

R is defined formally by rotating the x-axis and the y-axis, i.e. (see figure above):

$$\mathbf{R}|\mathbf{e}_1\rangle = \cos\phi|\mathbf{e}_1\rangle + \sin\phi|\mathbf{e}_2\rangle$$
$$\mathbf{R}|\mathbf{e}_2\rangle = -\sin\phi|\mathbf{e}_1\rangle + \cos\phi|\mathbf{e}_2\rangle$$

]And therefore, since

$$R_{ij} = \langle \mathbf{e}_i | \mathbf{R} | \mathbf{e}_j \rangle$$

We get

$$R_{11} = R_{22} = \cos\phi, \quad R_{12} = -\sin\phi, \quad R_{21} = \sin\phi$$

i.e.,

$$R = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

Note that R is not the same as M, the coordinate transformation matrix we had before,

$$M = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}.$$

In fact, $R_\phi = M_{-\phi}$.

This is understandable – see figure above – rotating a vector by an angle has the same effect on its coordinates as rotating the coordinate system in the opposite direction.

End of rotation operator(for now, more later).

Another example: stretching along an axis;

You can verify, that in 2D the operator that stretches a vector by a factor of 3 along its y axis is

$$\mathbf{L} = |\mathbf{e}_1\rangle\langle\mathbf{e}_1| + 3|\mathbf{e}_2\rangle\langle\mathbf{e}_2|$$

(Apply this on a vector $|\mathbf{v}\rangle = v_1|\mathbf{e}_1\rangle + v_2|\mathbf{e}_2\rangle$, and get $\mathbf{L}|\mathbf{v}\rangle = v_1|\mathbf{e}_1\rangle + 3v_2|\mathbf{e}_2\rangle$).

The associated matrix is

$$L = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

Applying two operations one after the other.

Let's apply two operations on a vector, **L** and **K** (first **K,** then **L).**

$$|\mathbf{b}\rangle = \mathbf{LK}|\mathbf{u}\rangle$$

$$b_j = \langle \mathbf{e}_j | \mathbf{LK} | \mathbf{u} \rangle = \langle \mathbf{e}_j | \mathbf{L} | \mathbf{e}_m \rangle \langle \mathbf{e}_m | \mathbf{K} | \mathbf{u} \rangle \quad \text{(implied summation)}$$

$$b_j = \langle \mathbf{e}_j | \mathbf{L} | \mathbf{e}_m \rangle \langle \mathbf{e}_m | \mathbf{K} | \mathbf{e}_n \rangle \langle \mathbf{e}_n | \mathbf{u} \rangle$$

$$b_j = L_{jm} K_{mn} u_n$$

$$b_j = (LK)_{jn} u_n$$

I.e., a product of operator **LK** is represented by a matrix, which is a matrix product of the matrices of the individual operators *LK*.

Example

If we want, for example, to rotate a vector clockwise by an angle ϕ and then stretch it by a factor of 3 along the y-axis, we can represent the matrix for the two operations by the product of the matrices

$$LR = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ 3\sin\phi & 3\cos\phi \end{pmatrix}$$

Note that since the operators do not commute (***LR*** ≠ ***RL*** **),** the matrices do not commute.

Determinants and Matrix Inverses

I presume that you saw before determinants, and know how to evaluate it formally, so I'll just present the formulae; for convenience, so as to avoid a plethora of indices, we will use a 3*3 matrix, and define the matrix as 3 row vectors,

$$M = \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \\ \boldsymbol{c} \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

And

$$\det(M) = a_1 \det\begin{pmatrix} b_2 & b_3 \\ c_2 & c_3 \end{pmatrix} - a_2 \det\begin{pmatrix} b_1 & b_3 \\ c_1 & c_3 \end{pmatrix} + a_3 \det\begin{pmatrix} b_1 & b_2 \\ c_1 & c_2 \end{pmatrix}$$

with

$$det\begin{pmatrix} b_2 & b_3 \\ c_2 & c_3 \end{pmatrix} = b_2c_3 - b_3c_2$$

etc.

A more revealing form is

$$\det(M) = \sum_{ijk} \epsilon_{ijk} a_i b_j c_k$$

where ϵ_{ijk} is the Levi-Civitta symbol (here for 3 indices, since we consider vectors of length 3 and therefore matrices of 3*3):

$\epsilon_{123} = 1$, and in general any permutation of a pair of indices changes sign,

For reference once and all let's write down the 3!=6 nonvanishing symbols for a 3-index Levi-Civitta tensor, but the rule above should have been sufficient for you to reconstruct them:

$$\epsilon_{123} = 1, \epsilon_{213} = -1, \epsilon_{312} = 1, \epsilon_{132} = -1, \epsilon_{231} = 1, \epsilon_{321} = -1.$$

Proof: we need to expand over ijk that are different; the expansion gives (with abbreviations, you can fill in the details):

$$\sum_{ijk} \epsilon_{ijk} a_i b_j c_k = \epsilon_{123} a_1 b_2 c_3 + \epsilon_{132} a_1 b_3 c_2 + \cdots = a_1 b_2 c_3 - a_1 b_3 c_2 + \cdots = \det(M)$$

This leads to several other interesting facts:

First, a cross product can be represented by the Levi Civitta tensor,

i.e.,

$$\boldsymbol{d} = \boldsymbol{b} \times \boldsymbol{c}$$

Is the same as

$$d_i = \sum_{jk} \epsilon_{ijk} b_j c_k$$

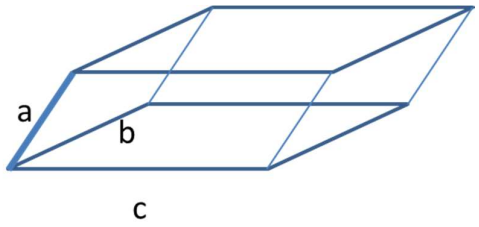
Proof: take, e.g., $i = 1$. Then, the second formula gives

$d_1 = \epsilon_{123} b_2 c_3 + \epsilon_{132} b_3 c_2 = b_2 c_3 - b_3 c_2$, Q.E.D.

Now from the definition of the determinant as $\det(M) = \sum_{ijk} \epsilon_{ijk} a_i b_j c_k$, it follows that

$$\det(M) = \sum_i a_i \sum_{jk} \epsilon_{ijk} b_j c_k = \sum_i a_i (\boldsymbol{b} \times \boldsymbol{c})_i = \boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})$$

while this expression is obviously valid only in 3D, it reveals something very useful; specifically, recall that the length of $\boldsymbol{b} \times \boldsymbol{c}$ is the same as the area of the paralleloid defined by **b, c** (see figure); then when we dot with **a** we get that $\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})$ (and therefore **det(M)**) is the **volume described by the 3 vectors, a,b,c**.



Parelloid volume: $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$

This is a general property of determinants; a determinant gives (up to an overall minus sign) the n-dimensional volume encased by the n row vectors consisting the determinants – or n column vectors, since, as we won't prove,

$$\det(M) = \det(M^T),$$

So we could use either column or row vectors.

(Also, we're not consistent on the symbol for the dimensionality sometimes using "d", sometimes "n", sometimes "N").

Other properties of determinants: **we can add or subtract to any row any constant times the other row (or to any column a product of the other column), without changing the determinant**;

This is easily proved in the Levi Civitta notation, since then, e.g., when we add the 3rd to the first row:

$$\det(\mathrm{M}') \equiv \det\begin{pmatrix} a_1 + f c_1 & a_2 + f c_2 & a_3 + f c_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} = \sum_{ijk} \epsilon_{ijk} \, (a_i + f c_i) b_j c_k$$

$$= \sum_{ijk} \epsilon_{ijk} \, a_i b_j c_k + f \sum_{ijk} \epsilon_{ijk} b_j c_i c_k = \det(M) + 0$$

Where the last term vanishes since ϵ_{ijk} is antisymmetric in i, k while it is multiplied by $c_i c_k$ which is symmetric in i,k (so we have terms such as $\epsilon_{123} c_1 b_2 c_3 + \epsilon_{321} c_3 b_2 c_1 = c_1 b_2 c_3 - c_3 b_2 c_1 = 0$)

Note: in practice determinants are not evaluated from these defining formulas for matrices with dimensions bigger than 6, since the effort grows factorially in this simplest way (like n!). Instead, one uses the Gauss-Jordan Elimination, not covered here, with a number of operations of about n^3.

Inversion of **M**: is only covered here for square ("n*n") matrices. As mentioned, it is practically done by Gauss Jordan elimination (won't be reviewed),which takes n^3 operations – see book. Alternately, for formal purposes, one writes the inverse as co-factors – this is something you saw before and won't be covered here.

The only important thing to remember is that M^{-1} is proportional to 1/det(M) – so if the determinant of M is zero, i.e., if the vectors of M are dependent, then M has no inverse.

This makes sense, and we'll prove it now (without reference to the explicit formulae).

Proof: If det(M) is zero, the column vectors are linearly dependent (as are the row vectors). That means that they are spanned by a basis with less than "n" vectors, where n is the dimension of the matrix (number of columns and rows). Therefore, there is a vector, **g,** that's not in the basis spanned by the column vectors of M.

Further: Now let's think on a matrix as a collection of column vectors (not row vectors as we had), i.e.,

$$\boldsymbol{M} = (\boldsymbol{m_1}, \boldsymbol{m_2}, \dots \boldsymbol{m_n})$$

where

$$\boldsymbol{m_1} \equiv \begin{pmatrix} M_{11} \\ M_{21} \\ \dots \end{pmatrix}$$

etc.,

Then, multiplying a vector $\boldsymbol{u}$ *by* $\boldsymbol{M}$*, amounts to a weighted summation of the column vectors, i.e.,*

$$\boldsymbol{Mu} = u_1\boldsymbol{m}_1 + u_2\boldsymbol{m}_2 + \cdots.$$

(Proof: $(\boldsymbol{Mu})_k = \sum_j M_{kj}u_j = M_{k1}u_1 + M_{k2}u_2 + \cdots = u_1\,(\boldsymbol{m_1})_k + u_2(\boldsymbol{m}_2)_k + \cdots$*)*

So it $\boldsymbol{M}$ *had an inverse, then we could define* $\boldsymbol{u} = \boldsymbol{M^{-1}g}$*, and then*

$$\boldsymbol{g} = \boldsymbol{MM^{-1}g} = \boldsymbol{Mu} = u_1\boldsymbol{m}_1 + u_2\boldsymbol{m}_2 + \cdots. =$$

So $\boldsymbol{g}$ *would have been spanned by the column vectors of M, which is a contradiction.*

To conclude: for the matrix to have an inverse, its column vectors must be a basis, therefore its determinant needs to be non-zero.

Properties of inverse:

$$(AB)^{-1} = B^{-1}A^{-1}$$

Proof:

$$(B^{-1}A^{-1})(AB) = B^{-1}A^{-1}AB = B^{-1}I\,B = I$$

Similar properties for transpose:

$$(AB)^T = B^T A^T$$

Proof: (using the summation convention)

$$(AB)^T{}_{ij} = (AB)_{ji} = A_{jk}B_{ki} = A^T_{kj}B^T_{ik} = B^T_{ik}A^T_{kj} = (B^T A^T)_{ij} \qquad Q.E.D.$$

*Note: An **orthogonal matrix** is defined as a matrix fulfilling*

$$R^{-1} = R^T$$

The natural extension of these to complex matrices is unitary matrices, discussed shortly.

Hermitian conjugates of operators

The Hermitian conjugate of a matrix is defined as

$$(M^\dagger)_{ij} = (M_{ji})^*$$

i.e., the complex conjugate of the transpose.

$$M^\dagger \equiv (M^\dagger)^* \quad \text{(definition)}$$

In bra-ket notation, the Hermitian-conjugate of an operator is defined to be an operator such that

$$\mathbf{L}^\dagger|\mathbf{b}\rangle \quad \text{means that for any ket } \langle\mathbf{c}|$$

$$\langle\mathbf{c}|\mathbf{L}^\dagger|\mathbf{b}\rangle \equiv \langle\mathbf{b}|\mathbf{L}|\mathbf{c}\rangle^*$$

By definition, the matrix representation of a Hermitian conjugate of an operator is the Hermitian conjugate of the matrix of the operator; here's the proof:

$$\left(\mathbf{L}^\dagger\right)_{ij} = \langle\mathbf{e}_i|\mathbf{L}^\dagger|\mathbf{e}_j\rangle = \langle\mathbf{e}_j|\mathbf{L}|\mathbf{e}_i\rangle^* = L_{ij}{}^*$$

As with kets and vectors, we'll move back and forth between Hermitian operators and matrices.

Note: $(AB)^\dagger = B^\dagger A^\dagger$ (prove this!)

An important point is that when we operate on a ket by an operator, the associated bra is related to the Hermitian conjugate of that operator. Let's prove this in matrix form first

If

$a = Lb$ $\quad \begin{pmatrix} \\ \end{pmatrix} = \begin{pmatrix} & \\ & \end{pmatrix} \begin{pmatrix} \\ \end{pmatrix}$,

where a and b are column vectors, i.e.,

$a_j = L_{jk} b_k$

then take the hermtian conjugate, i.e.,

complex conjugate and transpose (turning a into a row vector):

$a_j^* = b_k^* L_{jk}{}^* = b_k^* L^\dagger{}_{kj}$

or formally

$a^\dagger = b^\dagger L^\dagger$ $\quad (\quad) = (\quad)\left(L^\dagger\right)$

In bra-ket notation:

$|\mathbf{a}\rangle = \mathbf{L}|\mathbf{b}\rangle$ implies

$\langle \mathbf{a}| = \langle \mathbf{b}|\mathbf{L}^\dagger$

proof: take a general ket $|\mathbf{c}\rangle$; dot the 2nd line with this ket,

$\langle \mathbf{a}|\mathbf{c}\rangle = \langle \mathbf{c}|\mathbf{a}\rangle^* = \langle \mathbf{c}|\mathbf{L}|\mathbf{b}\rangle^* =$ (by defintion) $\langle \mathbf{b}|\mathbf{L}^\dagger|\mathbf{c}\rangle$

So $\langle \mathbf{a}|$ is equivalent completly to $\langle \mathbf{b}|\mathbf{L}^\dagger$, *i.e.*, we can identify both.

Hermitian and Unitary Matrices – extension of symmetric and orthogonal to complex matrices

There are two types of matrices associated with Hermitian conjugates: Hermitian matrices (or operators) and unitary matrices (or operators).

A Hermitian matrix, also called Self-adjoint, is defined as a matrix that's equal to its complex conjugate, i.e.,

$$M^\dagger = M \quad \text{(Hermitian matrices)}$$

Similarly, a Hermitian operator fulfills $\mathbf{M}^\dagger = \mathbf{M}$.

A unitary matrix (or operator) fulfills:

$M^\dagger = M^{-1}$ (Unitary matrices)

$\mathbf{M}^\dagger = \mathbf{M}^{-1}$ (Unitary operator) (so $\mathbf{M}^\dagger\mathbf{M} = \mathbf{M}\mathbf{M}^\dagger = 1$)

Example: A unitary operator preserves the norm of a vector.

Say

$$|\mathbf{a}\rangle = \mathbf{U}|\mathbf{b}\rangle$$

where **U** is unitary. Then

$$\langle \mathbf{a}| = \langle \mathbf{b}|\mathbf{U}^{\dagger}$$

$$\langle \mathbf{a}|\mathbf{a}\rangle = \langle \mathbf{b}|\mathbf{U}^{\dagger}\mathbf{U}|\mathbf{b}\rangle = \langle \mathbf{b}|1|\mathbf{b}\rangle = \langle \mathbf{b}|\mathbf{b}\rangle$$

i.e., the norm is preserved, as stipulated.

Another type of matrices you should remember is idempotent matrices, fulfilling

$$P^2 = P$$

Note that this implies that the eigenvalues of P (something we'll talk about later) are all 0 or 1. These matrices are important in electronic structure, since they are the density matrix of the electrons in the Hartree Fock or in the Kohn-Sham DFT approaches.

Hermitian and unitary matrices have importance in Diagonalization of matrices, covered next.

Diagonalization of matrices

For every linear operator or matrix **L** we can find an "easy' basis, where the action of the operator (or matrix) is "easy", i.e., is just a rescaling:

$$\mathbf{L}\left|\mathbf{v}_k\right\rangle = \lambda_k \left|\mathbf{v}_k\right\rangle$$

And furthermore, the eigenvectors can be normalized, i.e., $\left\langle \mathbf{v}_k \middle| \mathbf{v}_k \right\rangle = 1$, but there are cases when this eigenvector basis will also be automatically orthogonal – we'll learn later.

Example

For example, take a basis, $\mathbf{e}_1$ and $\mathbf{e}_2$ and define an operator **L** that exchanges these vectors. In a matrix form,

$$L = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

This matrix corresponds to the operator that exchanges the basis-elements, i.e.,

$$\mathbf{L}\left|\mathbf{e}_1\right\rangle = \left|\mathbf{e}_2\right\rangle$$

$$\mathbf{L}\left|\mathbf{e}_2\right\rangle = \left|\mathbf{e}_1\right\rangle$$

So for a general vector

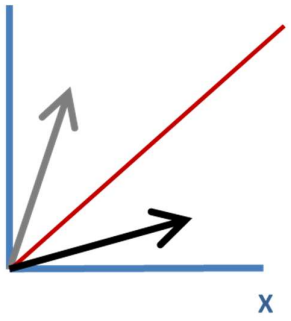
$$\left|\mathbf{a}\right\rangle = a_1\left|\mathbf{e}_1\right\rangle + a_2\left|\mathbf{e}_2\right\rangle$$

we operate with **L** as follows:

$$\mathbf{L}\left|\mathbf{a}\right\rangle = \mathbf{L}\left(a_1\left|\mathbf{e}_1\right\rangle + a_2\left|\mathbf{e}_2\right\rangle\right) = a_1\mathbf{L}\left|\mathbf{e}_1\right\rangle + a_2\mathbf{L}\left|\mathbf{e}_2\right\rangle = a_1\left|\mathbf{e}_2\right\rangle + a_2\left|\mathbf{e}_1\right\rangle$$

What will be the eigenvalue of this operator? We can guess it due to its simplicity

As you can see in the figure, this operator corresponds to a mirror exchange around the x-y (north-east) line.



Original vector: gray; after reflection through (red)x-y line, the x and y components of the vector exchange, resulting in the new, black, vector

Therefore, we can guess that one eigenvector will be a vector along the x-y line, which will be unchanged:

$$|\mathbf{v}_1\rangle = \frac{1}{\sqrt{2}}\left(|\mathbf{e}_1\rangle + |\mathbf{e}_2\rangle\right)$$

i.e.,

$$\mathbf{v}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

Where the 1/sqrt(2) is of course to normalize the eigenvector. The operator does not do anything to this vector, i.e., as you can verify

$$\mathbf{L}|\mathbf{v}_1\rangle = |\mathbf{v}_1\rangle,$$

i.e.,

$$\lambda_1 = 1$$

The other eigenvector will be perpendicular in this case to the x-y line, i.e., it will lie along the north-west (or south-east) line, i.e.,

$$|\mathbf{v}_2\rangle = \frac{1}{\sqrt{2}}\left(|\mathbf{e}_1\rangle - |\mathbf{e}_2\rangle\right)$$

i.e.,

$$\mathbf{v}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

And

$$\mathbf{L}|\mathbf{v}_2\rangle = -|\mathbf{v}_2\rangle,$$

i.e.,

$$\lambda_2 = -1$$

Diagonalizing in the general case:

In the general case, we can write

$$L\mathbf{v}_k = \lambda_k \mathbf{v}_k$$

i.e.,

$$(L - \lambda_k I)\mathbf{v}_k = 0$$

I.e., the matrix $(L - \lambda_k I)$ has a degenerate eigenvector. As we learned this means that

$$\det(L - \lambda_k I) = 0$$

We can use this equation to find λ_k, and then from it to find the eigenvector.

Example: let's use a more complicated example.

$$L = \begin{pmatrix} 3 & -1 \\ 2 & 1 \end{pmatrix}$$

So

$$L - \lambda I = \begin{pmatrix} 3-\lambda & -1 \\ 2 & 1-\lambda \end{pmatrix}$$

$$0 = \det(L - \lambda I) = (3-\lambda)(1-\lambda) - (-1)2 = \lambda^2 - 4\lambda + 5$$

i.e.,

$$\lambda = \frac{4 \pm \sqrt{16-20}}{2} = \frac{4 \pm \sqrt{-4}}{2} = \frac{4 \pm 2\sqrt{-1}}{2} = \frac{4 \pm 2i}{2} = 2 \pm i$$

The eigenvector can be found from this, I won't cover how this is done.

Note: the determinant of $L - \lambda$ will be a polynomial in λ of length d (where d is the size of the matrix), so , according to the theory of polynomials, there will be up to d different eigenvalues, i.e., a matrix of 10*10 will have at most 10 different eigenvalues, etc. This is in line with what we said that the eigenvectors forma basis, since a basis cannot include more than the dimension of the system (or in matrix form, a basis cannot include more than the number of rows or column of the matrix).

Diagonalization – general properties:

We wrote that diagonalizing linear operators amounts to finding the vectors, and associated eigenvalues, such that

$$\boldsymbol{L}|\boldsymbol{v}_k\rangle = \lambda_k|\boldsymbol{v}_k\rangle$$

Therefore, in any desired orthonormal basis, labeled $\{|\boldsymbol{e}_j\rangle\}$, we expand the operator, and by now you know what's the resulting equation, i.e., dot product by $\langle \boldsymbol{e_j}|$, and insert $\sum_i |\boldsymbol{e}_i\rangle\langle \boldsymbol{e}_i| = 1$, to get

$$\langle \boldsymbol{e_j}|\boldsymbol{L}|\boldsymbol{v}_k\rangle = \lambda_k|\boldsymbol{v}_k\rangle$$

$$\sum_i \langle \boldsymbol{e_j}|\boldsymbol{L}|\boldsymbol{e}_i\rangle\langle \boldsymbol{e}_i|\boldsymbol{v}_k\rangle = \lambda_k|\boldsymbol{v}_k\rangle$$

We write the eigenvectors in the original basis as

$|\boldsymbol{v}_k\rangle = \sum_m V_{mk}\,|\boldsymbol{e}_m\rangle$ (where $V_{mk} = \langle \boldsymbol{e_m}|\boldsymbol{v}_k\rangle$), i.e., V is the matrix where the eigenvectors are column vectors, i.e.,

$$V = (\boldsymbol{v_1}, \boldsymbol{v_2}, \dots)$$

So it follows that

$$\sum_i L_{ji} V_{ik} = \lambda_k V_{ji.}$$

And if we define the diagonal eigenvalue matrix (presuming the basis set has "n" vectors):

$$\lambda \equiv \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

We get:

$$LV = V\lambda$$

Further, multiplying by V^{-1} on the right we get

$$L = V\lambda V^{-1}$$

EIGENVALUES OF HERMITIAN MATRICES ARE REAL

Proof: Given an eigenvalue λ_k define the $\boldsymbol{a} = |\boldsymbol{v}_k\rangle, and\ \boldsymbol{b} = \boldsymbol{a}$.

$$\lambda_k = \langle \boldsymbol{v_k}|\boldsymbol{L}|\boldsymbol{v}_k\rangle \rightarrow$$

$$\lambda_k = \langle \boldsymbol{b}|\boldsymbol{L}|\boldsymbol{a}\rangle = \langle \boldsymbol{a}|\boldsymbol{L}^+|\boldsymbol{b}\rangle^* = \langle \boldsymbol{a}|\boldsymbol{L}|\boldsymbol{b}\rangle^* = \langle \boldsymbol{v_k}|\boldsymbol{L}|\boldsymbol{v_k}\rangle^* = \lambda_k^* \quad Q.E.D.$$

Note that this implies that eigenvalues of real symmetric matrices, which are a subset of Hermitian matrices, are real. (But eigenvalues of general matrices could be complex, as we saw).

Eigenvectors of Hermitian matrices are orthogonal (unless their eigenvalues are the same, in which case they still can be made to be orthogonal)

Proof: take two eigenvectors $|\boldsymbol{v}_j\rangle, |\boldsymbol{v}_k\rangle$. Then

$$0 = \langle \boldsymbol{v_k}|\boldsymbol{L} - \boldsymbol{L}^+|\boldsymbol{v_j}\rangle = \lambda_j\langle \boldsymbol{v_k}|\boldsymbol{v_j}\rangle - \langle \boldsymbol{v_k}|\boldsymbol{L}^+|\boldsymbol{v_j}\rangle = \lambda_j\langle \boldsymbol{v_k}|\boldsymbol{v_j}\rangle - \langle \boldsymbol{v_j}|\boldsymbol{L}|\boldsymbol{v_k}\rangle^*$$

$$= \lambda_j\langle \boldsymbol{v_k}|\boldsymbol{v_j}\rangle - \lambda_k^*\langle \boldsymbol{v_k}|\boldsymbol{v_j}\rangle = (\lambda_j - \lambda_k^*)\langle \boldsymbol{v_k}|\boldsymbol{v_j}\rangle$$

So either $(\lambda_j - \lambda_k^*) = 0, i.e., \lambda_j = \lambda_k$ (since the eigenvalues are real), or $\langle \boldsymbol{v_k}|\boldsymbol{v_j}\rangle = 0$, i.e., the eigenvectors are orthogonal.

Note that if there is more the one eigenvector with the same specific eigenvalue (e.g., as happens for the 2S and 2Px,2Py,2Pz orbitals of hydrogen), then we can still orthogonalize the eigenvectors, and make orthogonal (and orthonormal) eigenvectors.

We can of course also make each eigenvector normalized,

$$\langle \boldsymbol{v_k}|\boldsymbol{v_k}\rangle = 1\ (not\ summed\ over\ k)$$

Note that in that case (of orthonormal eigenvectors), then the matrix of eigenvectors is unitary, i.e.,

$$V^+V = 1$$

(Proof: $(V^+V)_{jk} = \sum_i V_{ji}^+ V_{ik} = V_{ij}^* V_{ik} = \langle \boldsymbol{v_k}|\boldsymbol{v_j}\rangle = \delta_{kj}$)

Mathews and Walker show that this is true not only for Hermitian matrices, but whenever a matrix L and its conjugate L^+ commute:

$$V^{+}V = 1 \quad when\ L\ , L^{+}\ commute$$

A little known example (beyond Hermitian matrices) is unitary matrices, i.e., we can find a unitary eigenvector matrix V for a matrix L if L is itself unitary!

Traces

Trace:

$$Tr(A) = A_{ii} \equiv \sum_i A_{ii}$$

(Note again that at times I use the Einstein summation convention, other times I don't; also, it won't be used outside this chapter.)

An important property of trace is that even for non-commuting matrices,

$$Tr(AB) = Tr(BA)$$

Proof:

$$Tr(AB) = (AB)_{ii} = A_{ij}B_{ji} = B_{ji}A_{ij} = (BA)_{jj} = Tr(BA)$$

Therefore (cyclic property)

$$Tr(ACD) = Tr(CDA)$$

(proof: just define B=CD, and apply the theorem above).

This leads to another property which will be very important later for matrix diagonalization:

Functions of Matrices:

Two useful facts:

1) **Any function $f(M)$ of any matrix which could be diagonalized (essentially all) as $M = V\lambda V^{-1}$, can be written as $f(M) = Vf(\lambda)V^{-1}$** ;
   Proof: assume the matrix can have a taylor expansion, i.e.,

   $f(M) = f_0 + f_1 M + f_2 M^2 + \cdots$ (where the f's are numbers)

   then we can write

$$f(M) = f(V\lambda V^{-1}) = f_0 + f_1 V\lambda V^{-1} + f_2 V\lambda V^{-1} V\lambda V^{-1} + \cdots$$
$$= V(f_0 + f_1\lambda + f_2\lambda^2 + \cdots)V^{-1} = Vf(\lambda)V^{-1} \quad Q.E.D.$$

2) We can relate the determinant of any matrix to the trace of its log:

$$\boldsymbol{M = exp(B) \rightarrow det(M) = exp(Tr(B))}$$

Proof: let's diagonalize

$$B = V\lambda V^{-1} \rightarrow M = exp(B) = V\, exp(\lambda)\, V^{-1} \rightarrow$$

$$det(M) = det(V)\, det(exp(\lambda))\, det(V^{-1})$$

But (we don't prove, let's accept it) the determinant of a product is the product of a determinant,

$$det(AB) = det(A)det(B), \quad \text{so } \det(V^{-1}) = \det(V)^{-1}$$

$$(\text{as } \det(VV^{-1}) = \det(I) = 1$$

and the determinant of a diagonal matrix is the product of its elements, i.e.,

$$det(M) = det\begin{bmatrix} e^{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\lambda_n} \end{bmatrix} = e^{\lambda_1} \ldots e^{\lambda_n} = e^{(\lambda_1 + \cdots \lambda_n)}$$

$$\det(M) = e^{Tr(\lambda)} = e^{Tr(\quad^{-1})} = e^{Tr(B)} \quad Q.E.D.$$

where we used

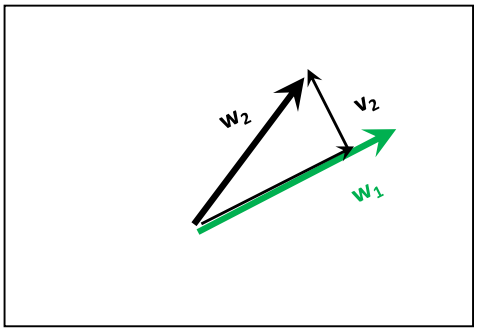
$$Tr(ABC) = Tr(CAB)$$

$$so \quad \boldsymbol{Tr(M)} = T\ (V\lambda V^{-1}) = Tr(V^{-1}V\lambda) = \boldsymbol{Tr(\lambda)}$$

Orthogonalizing basis sets:

Given a linearly independent set of "n" vectors, $|\boldsymbol{w_1}\rangle, |\boldsymbol{w_2}\rangle$, …we can make them into an orthonormal set, $|\boldsymbol{g_1}\rangle, |\boldsymbol{g_2}\rangle$, …by one of several different methods.

The simplest one is the Grahm Schmidt approach (see attached picture),



where we

(i) First, orthonormalize the first vector:

$$|g_1\rangle = \frac{1}{\sqrt{\langle w_1|w_1\rangle}}|w_1\rangle$$

(ii) Then, orthogonalize the 2nd vector to make it orthogonal to the first, i.e., define

$$|w'_2\rangle = |w_2\rangle - |g_1\rangle\langle g_1|w_2\rangle,$$

Followed by normalization of the resulting vector:

$$|g_2\rangle = \frac{1}{\sqrt{\langle w'_2|w'_2\rangle}}|w'_2\rangle$$

Continue like this with the next vectors, each time normalizing them to the previous one, and then normalizing i.e.,

$$|w'_k\rangle = |w_k\rangle - \sum_{j=1}^{k-1} |g_j\rangle\langle g_j|\, w_k\rangle$$

$$|g_k\rangle = \frac{1}{\sqrt{\langle w'_k|w'_k\rangle}}|w'_k\rangle$$

Formally, this produces an orthonormal set ($\langle g_j|g_k\rangle = \delta_{jk}$).

In practice, when you try this on the computer, you may encounter round-off errors, e.g., when you orthogonalize the 21st vector w.r.t. the 20th vector, your vector may not be orthogonal anymore on the computer to the 1st vector. – if this happens, you should repeat the Grahm Schmidt orthogonalization twice and in extreme cases even three times.

Also **the Grahm Schmidt approach is non-symmetric** in its treatment of the vectors – i.e., the first vector is not orthogonalized to any other vector, while the last it orthogonalized to all previous vectors.

Therefore, the Grahm Schmidt approach is perfect when you start from an arbitrary base and you want to produce the vectors before getting matrix elements. However, in many cases you have already the matrix given in a non-orthogonal basis, and you want to treat it symmetrically so it is better to start from a non-orthogonal basis, and treat it symmetrically then. For this, we need to study how to do diagonalization in non-orthogonal basis, and then only covert to orthogonal basis:

Non orthogonal basis sets: general, and diagonalization

As mentioned, often is it very convenient to work in nonorthogonal basis set,

$\{|\boldsymbol{w_1}\rangle, |\boldsymbol{w_2}\rangle, \ldots, |\boldsymbol{w_n}\rangle\}$ defined by the overlap matrix

$$S_{ij} = \langle \boldsymbol{w_i} | \boldsymbol{w_j} \rangle$$

(and of course for orthogonal basis set, S=I), This is very common in quantum chemistry, where the atomic basis sets that are sued are usually nonorthogonal. I.e., even if we take H_2 and use as our basis set two 1S orbitals, these orbitals are non-orthogonal. The Molecular orbitals are orthogonal, but there are a lot of advantages to using orthogonal orbitals.

The complicated thing about nonorthogonal basis sets is that we have to keep careful track of whether we are using a matrix or operator, as shown below.

Let's see what the nonorthogonality does to converting between operators and matrices. First, given a vector **b** which we expand by the basis-set,

$$|\boldsymbol{b}\rangle = \sum_j b_j \, |\boldsymbol{w_j}\rangle$$

The coefficient b_j are then found by dotting, i.e. define

$$f_i \equiv \langle \boldsymbol{w_i} | \boldsymbol{b} \rangle = \sum_j b_j \, \langle \boldsymbol{w_i} | \boldsymbol{w_j} \rangle = \sum_j S_{ij} b_j = (Sb)_i$$

i.e.,

$$b = S^{-1} f$$

Next, say we act on a vector, i.e.,

We can of course write the new vector in terms of the non-orthogonal basis set, i.e.,

$$|\boldsymbol{c}\rangle = \sum_j c_j \, |\boldsymbol{w_j}\rangle$$

And as before:

$$c_j = \sum_i (S^{-1})_{ji} \langle \boldsymbol{w_i} | \boldsymbol{L} | \boldsymbol{b} \rangle = \sum_{ik} (S^{-1})_{ji} \, \langle \boldsymbol{w_i} | \boldsymbol{L} | \boldsymbol{w_k} \rangle b_k$$

i.e.,

$$c = S^{-1} L b$$

Where

$$L_{ik} \equiv \langle \boldsymbol{w_i} | \boldsymbol{L} | \boldsymbol{w_k} \rangle$$

Symmetric Orthogonalization

Note that the discussion above leads to one simple way of orthogonalizing the basis: We first the eigenvector matrix of the Hermitian overlap matrix S,

$$S = UsU^{+}$$

(where s is a diangonal real matrix which can be shown to be positive deifnite), then use this to **define**

$$S^{-\frac{1}{2}} \equiv U\, s^{-\frac{1}{2}}\, U^{+}$$

And then define a new, orthogonal basis as

$$|\bar{\boldsymbol{w}}_k\rangle = \sum_j \left(S^{-\frac{1}{2}}\right)_{jk} |\boldsymbol{w}_j\rangle$$

(Prove that this basis set is orthogonal!)

(For your education – a completely different alternative is to use "bras" which are different than the kets and make the bra-ket product a unit matrix – you can read about it in the professional literature)

Now let's move to an alternative, where we don't make an orthogonal basis:

Diagonalization with non-orthogonal basis set: generalized eigenvalue problem

We learned that an eigenvector is defined as

$$\boldsymbol{L}|\boldsymbol{v}_k\rangle = \lambda_k|\boldsymbol{v}_k\rangle$$

Let's expand

$$|\boldsymbol{v}_k\rangle = \sum_j V_{jk}|\boldsymbol{w_j}\rangle$$

So

$$\boldsymbol{L}|\boldsymbol{v}_k\rangle = \sum_j V_{jk}\boldsymbol{L}|\boldsymbol{w_j}\rangle$$

"dot" the highlighted equation with $\langle \boldsymbol{w_i}|$ to get:

$$\langle \boldsymbol{w_i}|\boldsymbol{L}|\boldsymbol{v}_k\rangle = \sum_j \langle \boldsymbol{w_i}|\boldsymbol{w_j}\rangle V_{jk}\,\lambda_k$$

i.e.,

$$\sum_j \langle \boldsymbol{w_i}|\boldsymbol{L}|\boldsymbol{w}\rangle V_{jk} = (SV\lambda)_{ik}$$

Or

$$LV = SV\lambda$$

This is called the generalized-eigenvalue problem. There are several ways to solve it; the simplest one is to multiply the specially highlighted equation by $S^{-\frac{1}{2}}$, leading to

$$(S^{-\frac{1}{2}}\, L\, S^{-\frac{1}{2}})\, S^{\frac{1}{2}} V = S^{\frac{1}{2}} V\, \lambda$$

So if we define $\bar{L} = S^{-\frac{1}{2}} L\, S^{-\frac{1}{2}}$, we diagonalize $\bar{L}$ by finding $\bar{L} = \bar{V}\bar{\lambda}\bar{V}^{+}$, and then $V = S^{-\frac{1}{2}}\bar{V}$.

Of course, you have to be careful of the $S^{-\frac{1}{2}}$ which can diverge; read the book by Szabo and Ostlund, Chapters 2-3, on how to handle this properly.

Finally, we reach the last topic in this chapter:

Linear System of Equations

Let's consider a linear system of "n" equations for an unknown set of variables $\{a_j,\, j = 1, \ldots, n\}$. For a start, say we have as many equations as variables:

$$\begin{aligned} M_{11}a_1 + M_{12}a_2 + \cdots + M_{1n}a_n &= b_1 \\ M_{21}a_1 + M_{22}a_2 + \cdots + M_{2n}a_n &= b_2 \\ &\ldots \\ M_{n1}a_1 + M_{n2}a_2 + \cdots + M_{nn}a_n &= b_n \end{aligned}$$

we can write this of course as

$$\boldsymbol{Ma = b}$$

e.g., $\begin{pmatrix} 2 & -3 & 9 \\ -0.4 & 2.7 & 3 \\ 12 & 8 & -1.4 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 7 \\ -17.2 \\ 5 \end{pmatrix}$.

The solution (i.e., the value of **a**) depends on whether we have

(i) **Homogenous system**, i.e. $\boldsymbol{b} = 0$ -- then, the system of equations $\boldsymbol{Ma} = 0$ implies that the column vectors of **M** are linearly dependent, i.e., det(**M**)=0

(ii) **Inhomogeneous system,** $\boldsymbol{b} \neq 0$, then $\boldsymbol{a = M^{-1}b}$

Variational Solution of Inhomogeneous equations

However, **in practice** $\boldsymbol{a = M^{-1}b}$ is **very dangerous numerically,** since many matrices are almost degenerate; **Instead,** it is highly recommended, if you need to, to use a **variational method,** something you'll see a lot

Define a "functional", i.e., a number that depends on the unknown vector (here it depends on "**a**"), and tries to minimize the deviation from **Ma=b, (i.e., the squared norm of** $\boldsymbol{Ma - b}$) while adding a penalty term:

$$\begin{aligned} J &= (\boldsymbol{Ma - b})^{+}(\boldsymbol{Ma - b}) + \eta \boldsymbol{a}^{+}\boldsymbol{a} \\ &= \boldsymbol{a}^{+}\boldsymbol{M}^{+}\boldsymbol{Ma} + \eta \boldsymbol{a}^{+}\boldsymbol{a} - \boldsymbol{a}^{+}\boldsymbol{b} - \boldsymbol{b}^{+}\boldsymbol{a} + \boldsymbol{b}^{+}\boldsymbol{b} \end{aligned}$$

$$= \sum_{jk} a_j^* (M^+ M)_{jk} a_k + \sum_j a_j^* a_j - b_j^* a_j - a_j b_j^* + b_j^* b_j$$

where η is a small number, a "regularization parameter"; then, we find the $\{a_l\}$ that gives the minimum value of the functional, by differenting J w.r.t. $\{a_l^*\}$ or w.r.t. $\boldsymbol{a}^+$

$$\frac{\partial J}{\partial \boldsymbol{a}^+} = 0 \rightarrow (\boldsymbol{M}^+\boldsymbol{M} + \boldsymbol{I}\eta)\boldsymbol{a} - \boldsymbol{M}^+\boldsymbol{b} = 0 \rightarrow$$

$$\boldsymbol{a} = (\boldsymbol{M}^+\boldsymbol{M} + \boldsymbol{I}\eta)^{-1}\boldsymbol{M}^+\boldsymbol{b}$$

(Note that we differentiate w.r.t. the complex conjugate variables, a_j^* as if the original variable is unchanged; this is generally true in complex functions of f(z,z*) – we can differentiate w.r.t. z independently of z*, and vice versa; if you don't believe me you can differentiate w.r.t the real part and imaginary part of a_j separately, and you can get the same result)

Advantages of the variational solution:

The variational solution will equal the regular solution if $\eta = 0$. (Verify!) However, in general it will be much more stable, i.e., there usually the solution is stable, within a range of η between small and very small (but usually not zero).

What η does is introduce a "penalty" if the solution is large, thereby ensuring that we won't have solutions which may formally be excellent but in practice are very large and susceptible to round-off errors, etc.

Least Square Fitting

Quite beyond this case of avoiding singularities in square-matrix inversions, this variational approach (with the "penalty term", $\eta\boldsymbol{a}^+\boldsymbol{a}$) indispensable when we have **problems where the number of equations is not equal to the number of solutions, i.e.,** where we'll **like to fulfill** still $\boldsymbol{M}\boldsymbol{a} = \boldsymbol{b}$ **"to the best of our ability" ,** but M is not a square matrix.

If the number of equations is smaller than the number of variables there are infinitely many solutions possible; in that case we'll like the "best" solution, which we can define to be the solution with the least squared norm $\eta\boldsymbol{a}^+\boldsymbol{a}$, so we have exactly the highlighted solution.

If there are more equations than variables, we usually won't be able to get an exact answer, but again we can minimize the total objective J, (i.e., the squared norm of $\boldsymbol{M}\boldsymbol{a} - \boldsymbol{b}$, plus squared norm of the solution $\eta\boldsymbol{a}^+\boldsymbol{a}$), again resulting in the same equations.

Differential equations

Overview

1. First order equations, integrating factor,
2. Linear 1st order equations
3. Linear 2nd order equations, different types of solutions
4. homogenous and inhomogeneous solutions
5. Saddle points and classifying differential equations
6. Actually solving ordinary 2nd order differential equations
7. Qualitative considerations

First order equations

$$\frac{dy}{dx} = f(x, y)$$

In standard notation we write

$$f = -\frac{P}{Q}$$

and we'll specify different P, Q, later $i.e.$,

$$\frac{dy}{dx} = -\frac{P(x, y)}{Q(x, y)}$$

Only 1st derivative here → these are 1st order equations

Also: Only total derivative, not partial derivative → labeled as "ordinary" differential equations.

Examples

I. If we can write P,Q as functions of x,y separately: P(x), Q(y):

$$\frac{dy}{dx} = -\frac{P(x)}{Q(y)}$$

Lead to separable differential equations:

$$Q(y)dy + P(x)dx \quad 0$$

Integrate from x_0, y_0 to x, y

$$\int_{y_0}^{y} Q(y)dy \quad \int_{x_0}^{x} P(x)dx \quad 0$$

We can throw the initial conditions, and just add a constant of integration to be evaluated some other way.

For example, Boyle's gas law, as fixed T and n,

$$\frac{dV}{dP} = -\frac{V}{P}$$

So

$$\frac{dV}{V} = -\frac{dP}{P} \quad \rightarrow \quad logV = -logP + Cons \ .$$

I.e.,

$$\log(VP) = Con \quad .$$

i.e.

$$PV = Const.$$

For an ideal gas, this "Const." is just nRT .

Example 2:

Another option: $P(x,y), Q(x,y)$ that yield an "exact differential", i.e., there's a function $\phi(x,y)$ such that

$$\frac{\partial \phi}{\partial x} = P(x,y)$$

$$\frac{\partial \phi}{\partial y} = Q(x,y)$$

Then

$$d\phi = \frac{\partial \phi}{\partial x}dx + \frac{\partial \phi}{\partial y}dy = P(x,y)dx + Q(x,y)dy = 0$$

So the solutions are $\phi(x,y) = const.$

Note that this is just like the potential, alluded to earlier in the course.

Integrating factor:

You can always find an integrating factor $\alpha(x,y)$ so that:

$$\alpha(x,y)Q(x,y)dy + \alpha(x,y)P(x,y)dx \quad 0$$

and now the LHS is a total differential.

In general this is tough but there are simple and commonly found cases where you can find a general prescription for the integrating factor:

Linear 1st order differential equations

"linear" means linear in y, the dependent variable. We write it as

$$f(x) = yp(x) - q(x)$$

so

$$\frac{dy}{dx} + p(x)y = q(x)$$

Then we want an integrating factor $\alpha(x)$ that's only a function of x, such that

$$\alpha\frac{dy}{dx} + \alpha py = \alpha q$$

while having the left side as a total integral, i.e., we aim for having α such that:

$$\alpha\frac{dy}{dx} + \alpha py = \frac{d(\alpha y)}{dx}$$

But the RHS equals $\alpha\frac{dy}{dx} + y\frac{d\alpha}{dx}$, so that we want

$$\alpha p = \alpha' \ ,$$

i.e.,

$$\alpha = e^{\int_{x_0}^{x} p(x')dx'} = Const.\, e^{\int^{x} p(x')dx'}$$

This "const." will be irrelevant below, as we'll divide and multiply by α.

So we then have:

$$\frac{d}{dx}(\alpha y) = \alpha q$$

i.e.,

$$\alpha y = \int^{x} \alpha(x')q(x')dx' + C$$

$$y = \frac{C}{\alpha(x)} + \frac{1}{\alpha(x)}\int^{x} \alpha(x')q(x')dx', \qquad \alpha(x) = e^{\int^{x} p(x')dx'}$$

Example: A ball in viscous liquid.

Newton's law, together with a drag force proportional to the velocity (applicable for low Reynolds number, a concept from hydrodynamics), give:

$$m\frac{dv}{dt} = -mg - \bar{\xi}v$$

where $\bar{\xi}$ is the drag force. Defining the drag force per mass of the falling object, i.e.,

$$\xi = \frac{\bar{\xi}}{m}$$

we get

$$\frac{dv}{dt} + \xi v = -g.$$

This equation has the form we dealt with, $\frac{dy}{dx} + p(x)y = q(x)$, with p and q now constant. The solution is therefore as explained to define

$$\alpha = e^{\xi t},$$

and multiply the equation above, to give:

$$e^{\xi t}\frac{dv}{dt} + \xi e^{\xi t}v = -ge^{\xi t},$$

i.e.,

$$\frac{d\left(e^{\xi t}v\right)}{dt} = -ge^{\xi t},$$

i.e.,

$$e^{\xi t}v - v_0 = -g\int_0^t e^{\xi t'}dt' = -\frac{g}{\xi}\left(e^{\xi t} - 1\right),$$

i.e.,

$$v = v_0 e^{-\xi t} - \frac{g}{\xi}\left(1 - e^{-\xi t}\right).$$

At infinity the velocity is negative, i.e., $-\frac{g}{\xi}$, as the particles downward fall due to gravity is balanced by the friction.

Similar equations are relevant, e.g., for chemical reactions, for motions of dipoles in solution, etc.

Linear 2nd order differential equations

A little more complicated because there are two independent solutions.

Let's warm up with the simplest case: homogenous linear case with const. coefficients:

$$y'' + ay' + by = 0 \quad (a, b\ const: linear;$$
$$RHS = 0 \ \rightarrow homogenous, no\ const.term)$$

This simple problem has a nice solution; use an ansatz,

$$y = e^{\alpha x}$$

$$\alpha^2 + a\alpha + b \quad 0$$

i.e.,

$$\alpha = \frac{-a \pm \Delta}{2} \qquad \Delta \equiv \sqrt{a^2 - 4b}.$$

Three cases:

- α real when $a^2 > 4b$, two solutions, $y = Ae^{\left(-\frac{a}{2}+\frac{\Delta}{2}\right)x} + Be^{\left(-\frac{a}{2}-\frac{\Delta}{2}\right)x}$
- α real when $a^2 = 4b$ (so $\Delta = 0$) , one solution, $y = Ae^{-\frac{a}{2}x}$
- α complex when $a^2 < 4b$, $y = Ae^{\left(-\frac{a}{2}+i\frac{|\Delta|}{2}\right)x} + Be^{\left(-\frac{a}{2}-i\frac{|\Delta|}{2}\right)x}$

Example: damped Harmonic oscillator,

$$\frac{d^2y}{dt^2} = -\xi\frac{dy}{dt} - \omega_0^2 y$$

where again ξ is the scaled friction coefficient, and $m\omega_0^2$ is the restoring force.

We rewrite:

$$\frac{d^2y}{dt^2} + \xi\frac{dy}{dt} + \omega_0^2 y = 0$$

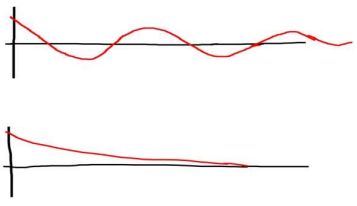
So we have three case (see figure):

- Overdamped case, $\xi > 2\omega_0$: two damped solutions,

$$y(t) = Ae^{-\frac{\left(\xi+\sqrt{\xi^2-4\omega_0^2}\right)t}{2}} + Be^{-\frac{\left(\xi-\sqrt{\xi^2-4\omega_0^2}\right)t}{2}}$$

- Critically damped, $\xi = 2\omega_0$:

Underdamped(top) and Overdamped (bottom) oscillations as a function of time.



$$y(t) = Ae^{-\frac{\xi}{2}t}$$

- Underdamped (oscillating) solution: ,$\xi < 2\omega_0$:

$$y(t) = Ae^{-\left(\xi+i\sqrt{4\omega_0^2-\xi^2}\right)\frac{t}{2}} + Be^{-\left(\xi-i\sqrt{4\omega_0^2-\xi^2}\right)\frac{t}{2}}.$$

Of course, when the friction is zero, we recover in this case the purely oscillating solution, $y(t) = Ae^{-i\omega t} + Be^{i\omega t}$.

Non-constant coefficients:

Frequently encountered. Examples:

QM harmonic oscillator with linearly increasing friction, yields Hermite's eqn.:

$$y'' - 2xy' + 2\alpha y = 0.$$

Particle in a circular tube; the resulting equation is Bessel's equation.

$$x^2y'' + xy' + (x^2 - m^2)y = 0,$$

and many other examples. We'll study these shortly; but first consider homogenous and non-homogenous solutions:

Homogenous and Non-Homogenous solutions:

Consider the general linear 2nd order equation (or more generally, any nth order solution):

$$y'' + P(x)y' + Q(x)y = R(x).$$

Then, we can classify the solution as a general mix of the following form:

$$y(x) = y_i(x) + c_1y_1(x) + c_2y_2(x).$$

Here:

1) y_i is a single solution to the inhhomogenous problem
2) $y_1$and y_2 are solutions of the homogenous equation,

$$y'' + P(x)y' + Q(x)y = 0$$

And c_1, c_2 are arbitrary coefficients.

(If the equation is higher than 2nd order, say 5th order, there will be up to 5 independent solutions to the homogenous equation.)

The reason for the statement is that first, we can by inspection add to the inhomogeneous solution any linear combination of the linear solution, and the equation will not be changed;

Further, if there are two inhomogeneous solutions, i.e., y_i and y_j, then

$$y_i'' + P(x)y_i' + Q(x)y_i = R(x)$$

$$y_j'' + P(x)y_j' + Q(x)y_j = R(x)$$

So the difference, $y_i - y_j$ fulfils the homogenous equation, i.e., it is enough to consider one solution of the inhomogeneous equation.

Finally, we won't prove why an n[th] order linear homogenous equation has n solution, that's something to accept (if you want to see a p[articular example, try $y'' - a^2 y = 0$, e.g., you see that it has two solutions,

$y_1 = \exp(+ax), y_2 = \exp(-a\ \ )$.)

Singular Points and Classifying Differential Equations

Given

$$y'' = f(y', y, x).$$

Then if y', y are finite at some point x_0 and y'' remains finite at that point, then x_0 is a regular point;

But if y', y are finite at some point x_0 and y'' is infinite x_0, then x_0 is a singular point.

There are further classifications: If we can write the equations in the linear homogenous form

$$y'' + P(x)y' + Q(x)y = 0.$$

Then, if P,Q are finite at x_0, it is an ordinary point.

If P, Q are diverge at x_0, but $(x - x_0)P$ and $(x - x_0)^2 Q$ remain finite at x_0 then x_0 is a regular (also known as non-essential) singularity

(You can also analyze the "limit" $x = \infty$ by doing a change of variables, $x = \frac{1}{z}$, and check the point $z = 0$.)

Example:

Bessel's equation,

$$x^2 y'' + xy' + (x^2 - m^2)y = 0 \quad \rightarrow$$

$$y'' + \frac{y'}{x} + \frac{(x^2 - m^2)}{x^2} y = 0$$

By inspection, a regular singularity at $x = 0$; a change of variables leads to nauseating derivation that eventually show that, for $z = \frac{1}{x}$,

$$\frac{d^2 y}{dz^2} + \frac{1}{z}\frac{dy}{dz} + \frac{1 - m^2 z^2}{z^4} y = 0$$

So z=0 (i.e., $x = \infty$) is an essential singularity (due to the $\frac{1}{z^4}$ part).

If the singularities are regular, the function is easier to deal with and still derive regular values.

Actually solving ordinary 2nd order differential equations

Used to be a big deal, now less important (for you) because essentially any equation that you'll encounter has been solved and programmed; however, the methodology is revealing.

Example (warm-up): Harmonic oscillator.

$$y'' + \omega^2 y = 0$$

Try

$$y(x) = \sum_{n=0}^{\infty} a_n x^{k+n}, \quad a_0 \neq 0 \qquad n \in Z(group\ of\ non-negative\ integers),$$

but k not neccessarily integer

Then

$$y''(x) = \sum_{n=0}^{\infty} (k+n)(k+n-1) a_n x^{k+n-2}$$

Let's break it down to the n=0, n=1 and n>1 parts, so we can reliable the last part, i.e.., arrive back at a polynomial of the form x^{n+k}:

$$y'' = k(k-1)a_0 x^{k-2} + (k+1)k a_1 x^{k-1} + \sum_{n=2}^{\infty} (k+n)(k+n-1) a_n x^{k+n-2}$$

And now we can reliable n→ n+2, so

$$y'' = k(k-1)a_0 x^{k-2} + (k+1)k a_1 x^{k-1} + \sum_{n=0}^{\infty} (k+n+2)(k+n+1) a_{n+2} x^{k+n}$$

So $y'' + \omega^2 y = 0$ gives:

$$k(k-1)a_0 x^{k-2} + (k+1)k a_1 x^{k-1} + \sum_{n=0}^{\infty} \left((k+n+2)(k+n+1)a_{n+2} + \omega^2 a_n\right) x^{n+k} = 0$$

For this to vanish, every coefficient needs to vanish (since each "x" term has a different power). So:

(i) $k(k-1)a_0 = 0 \rightarrow k = 0\ or\ k = 1\ \ (as\ a_0 \neq 0, by\ construction)$

(ii) $(k+1)k a_1 = 0 \ \rightarrow\ k = 0\ or\ a_1 = 0$ (k+1 cannot vanish since we know that k=0 or 1, from (i).

(iii) $(k+n+2)(k+n+1)a_{n+2} - a_n = 0 \rightarrow$

$$a_{n+2} = -\frac{\omega^2 a_n}{(k+n+2)(k+n+1)}$$

Let's first use the solution of k=0; then both conditions (i),(ii) are automatically fulfilled, and then we know

$$a_{n+2} = -\frac{\omega^2 a_n}{(n+2)(n+1)} \quad (k=0)$$

i.e.,

$$a_2 = -\frac{\omega^2 a_0}{2 \cdot 1}$$

$$a_4 = -\frac{\omega^2 a_2}{4 \cdot 3} = \frac{(-1)^2\, \omega^4}{4 \cdot 3 \cdot 2 \cdot 1}$$

And generally

$$a_{2n} = \frac{(-1)^n\, \omega^{2n}}{(2n)!}.$$

Setting $a_1 = 0$ (that's OK by the relations we had, plus the case where the coefficient of x^1 is non-zero is really the same as taking k=1 initially, discussed in a second), then we get:

$$k = 0 \quad \rightarrow \quad y(x) = a_0 \left(1 - \frac{\omega^2 x^2}{2} + \frac{\omega^4 x^4}{4!} - \frac{\omega^6 x^6}{6!} + \cdots\right) = a_0 \cos(\omega x).$$

What about the other case, i.e., k=1? Then we can do the math analogously, and get

$$k = 1 \rightarrow y(x) = a_0 \sin(\omega x).$$

So there are two independent solutions, both of which we know of already.

<u>Another example: Bessel's function,</u>

An example showing that things can be more tricky, is: series solution of the Bessel equation. The Bessel equation is, reminder:

$$x^2 y'' + xy' + (x^2 - m^2)y = 0$$

Plugging again a series solution:

$$y(x) = \sum_{n=0}^{\infty} a_n x^{k+n}, \quad a_0 \neq 0$$

We get:

$$\sum_{n=0}^{\infty} a_n (k+n)(k+n-1)x^{k+n} + a_n (k+n)x^{k+n} + a_n x^{k+n+2} - m^2 a_n x^{k+n} = 0.$$

All terms have the same power, starting at x^k except the third, which starts at x^{k+2}, so as before we first deal with the x^k, x^{k+1} terms, i.e., the $n = 0,1$ terms.

For the powers of x^k:

$$a_0(k(k-1)+k-m^2)=0 \;\rightarrow a_0(k^2-m^2)=0 \;\rightarrow \quad k=\pm m$$

For the powers of x^{k+1}:

$$a_1\big((k+1)k+k+1-m^2\big)=0 \;\rightarrow$$

$$a_1(k+1-m)(k+1+m)=0$$

But we know that $k=\pm m$. The only way that that can be together with the equations above are:

- Either $m=\pm\frac{1}{2}$, a special case of the spherical Bessel functions, not dealt with here;
- Otherwise, $a_1=0$.

Next, to the iterations; i.e., as before we relate the powers, i.e., rewrite the 3rd term in the main equation above as $a_{n-2}x^{k+n},\ n\geq 2$, so we get

$$a_n\big((k+n)(k+n-1)+(k+n)-m^2\big)+a_{n-2}=0 \;\rightarrow$$

$$a_n=-\frac{a_{n-2}}{(k+n)^2-m^2}$$

For $k=m$, get (using $a^2-b^2=(a+b)(a-b)$:)

$$\boldsymbol{a_n=-\frac{a_{n-2}}{(m+n)^2-m^2}=-\frac{a_{n-2}}{n(n+2m)}}.$$

For $k=-m$, get:

$$a_n=-\frac{a_{n-2}}{(n-m)^2-m^2}=-\frac{a_{n-2}}{n\,(n-2m)}$$

The first (bolded) solution will cover both cases if we allow m to be negative or positive.

Then, solving the recursion, gives readily (prove):

$$a_{2p}=\frac{(-1)^p a_0 m!}{2^{2p}p!\,(m+p)!}$$

So the solution is:

$$y(x)=a_0J_m(x),$$

Where we defined the m'th Bessel function:

$$J_m(x)=2^m m!\sum_{j=0}^{\infty}\frac{(-1)^j}{j!\,(m+j)!}\left(\frac{x}{2}\right)^{m+2j}$$

(note that $j=n/2$)

Note that when m is even, the Bessel function is even in x, and vice versa for odd m.

If m is positive or zero, everything here is fine. That's also the case when m is zero or negative. But when m is negative and integer, there is a problem. Then, the j=|m| term will blow up.

So $J_{-|m|}(x)$ cannot be independent of $J_{|m|}(x)$. It is usually defined as

$$J_{-m}(x) = (-1)^m J_m(x), \qquad m\ integer$$

So we see that for some values of the parameters (i.e., m integrer) there is only one regular solution; while for the harmonic oscliator there are always 2 solutions.

QUALITATIVE CONSIDERTATIONS

Often we get more from qualitative considerations then from actual solutions.

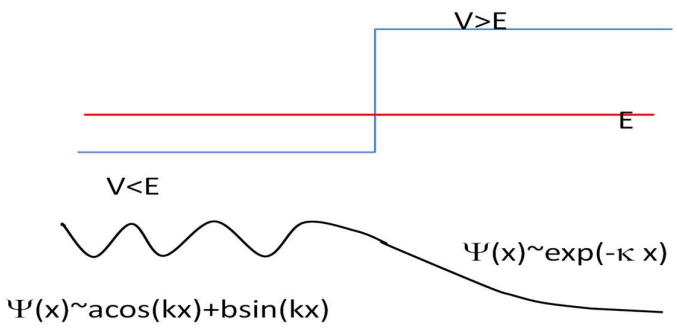
An example is a 1D Schrödinger-equation,

$$-\frac{1}{2m}\frac{d^2y}{dx^2} + V(x)y = Ey.$$

(we use units where $\hbar = 1$), which can be rewritten as one of two different forms, depending on whether the potential is lower or higher than the energy than the energy (also called the classically allowed and classically forbidden region, for obvious reasons):

$$\frac{d^2y}{dx^2} = \begin{cases} -k^2(x)y(x) & \text{where } k(x) = \sqrt{2m(E - V(x))} \text{ if } V(x) < E \\ \kappa^2(x)y(x) & \text{where } \kappa(x) = \sqrt{2m(V(x) - E)} \text{ if } E < V(x) \end{cases}$$

For a general V, of course, the solution of these equations has to be done explicitly (or on the computer, as usually done). But we can understand the qualitative feature by **approximating that the "local momentum"** (which refers either to $k(x)$ or $\kappa(x)$) **varies "slowly";** in that case we solve the Schrodinger equation **as if** the potential is constant so k or κ are fixed.



Difference between the solutions of the Schrödinger equation in classically forbidden regions (V>E) and classically allowed regions (V<E).

The solution (the WKB solution) is outlined in many advanced books (such as Mathew and Walker) and is reported below; for our purposes, however, we note that in the first case where the sign of the 2nd derivative is opposite that of y, then y oscillates like $y \sim exp(\pm ikx)$, while in the secnd case, where the second derivative has the same sign as the function, the solution is exponentially increasing or decreasing, going like $y \sim \exp(\pm \kappa x)$. (See figure for details.)

Note that if the energy is lower than the potential at large positive x, then at infinity only the exponentially decreasing solution is allowed otherwise y exploders at infinity; vice versa for large negative values of x. See the figure for details.

Finally, the actual WKB formula is:

$$y(x) = \frac{\alpha}{\sqrt{\kappa(x)}}\exp(-\int_{x_0}^{x}\kappa(x')dx') + \frac{\beta}{\sqrt{\kappa(x)}}\exp(\int_{x_0}^{x}\kappa(x')dx')$$

$$y(x) = \frac{a}{\sqrt{k(x)}}\exp(-i\int_{x_0}^{x}k(x')dx') + \frac{b}{\sqrt{k(x)}}\exp(i\int_{x_0}^{x}k(x')dx')$$

where x_0 is a nearby transition point (a point where $V(x_0) = E$), and a, b, α, β are constants.

The integrals (e.g., $\int_{x_0}^{x}\kappa(x')dx'$) are the extension of the linear terms (κx to the case that the local momentum ($\kappa(x), k(x)$) are not constant, i.e., when V(x) depends on x and is not locally constant.

The one new ingredient in the WLKB solution is the $\frac{1}{\sqrt{k(x)}}$ (or $\frac{1}{\sqrt{\kappa(x)}}$) term in front of the exponential. This term has a physical explanation – the higher the local memtum the faster the wavefunction moves (in the case of E>V), so the smaller its amplitude, due to a rule called "conservation of flux", that you'll see in QM.

By the way, the conservation of flux rule makes sense to anybody who travelled on freeways – in regions where the traffic moves fast the density of cars is low (big distance between each car), but in slower moving regions the cars are closer.

This is also true in earthquakes – that's why earthquakes make more damage in sandy or "liquefied-earth" regions, like Santa Monica, where they travel slower and have higher amplitude, and least damage in mountainous areas (think Mulholland drive) where they travel fast through rock, and therefore have smaller magnitude. So in earthquake descriptions (where the mathematics of the wave motion is similar), the waves also have the $\frac{1}{\sqrt{k(x)}}$ term.

Functions of a complex variable.

Important in all of physics: Quantum mechanics, hydrodynamics, general wave motion, etc. Important specifically since wave equations are much more natural in complex variables.

Overview:

1. Complex variables, functions of a complex variable, Re(f) and Im(f), complex conjugation
2. Cauchy Riemann conditions; analytical functions
3. Cauchy integral theorem, contour integrals
4. Integrals of $1/(z\text{-}z_0)$
5. Laurent Series
6. Residue theorem
7. Analytic functions can't decay in all directions
8. Cauchy's principle value
9. Saddle point integration for real variables steepest descent;
10. Gaussian integrals.
11. Saddle point integration of complex integrals

Complex variables:

Use $i = \sqrt{-1}$. A complex number is defined as $z = x + iy$, and the product of complex numbers, as you recall, is

$$z \cdot z' = (x + iy) \cdot (x' + iy') = (xx' - yy') + i(xy' + yx')$$

(The addition is trivial and will not be discussed here). If you are unhappy with something which is not formally defined, $\sqrt{-1}$, then think of complex numbers either as

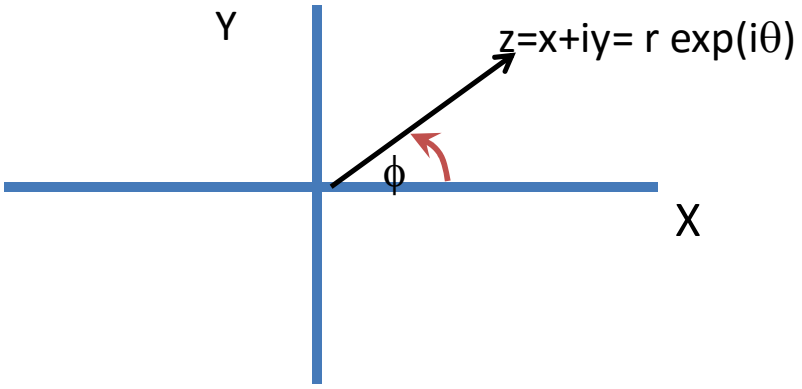
- Pairs of numbers z=(x,y), with specific rules for multiplying the pairs $(x, y) \cdot (x', y') = (xx' - yy', xy' + yx')$ -- this is how complex numbers are handled on the computer (with the trivial addition rules)
- Or, even better, as matrices of the form $\begin{pmatrix} x & y \\ -y & x \end{pmatrix}$, and as the following multiplication show such matrices are indeed isomorphic to the complex numbers, i.e.,

  when the matrix equivalent of z, matrix-multiplies the matrix equivalent of z', we get the matrix equivalent of zz':

$$z \cdot z' \to \begin{pmatrix} x & y \\ -y & x \end{pmatrix}\begin{pmatrix} x' & y' \\ -y' & x' \end{pmatrix} = \begin{pmatrix} xx' - yy' & xy' + yx' \\ -(yx + x'y') & -yy' + xx' \end{pmatrix}$$

This equivalence, not pursued further, is important in the discussion of Pauli matrices in quantum mechanics.

The complex plane (see figure)



$$z = x + iy = r\cos\theta + ir\sin\theta = re^{i\theta}$$

where

$$|z| \equiv r = \sqrt{x^2 + y^2}$$

$$\tan\theta = \frac{x}{y}$$

- r, θ are the polar coordinates:
- r is the modulus:
- θ is the argument $\theta = \arg(z)$

and $z = re^{i\theta}$ is the polar representation of *z.*

The modulus is the same as that of a 2-D vector, (x,y)

Note that (we won't prove, but follows from what you know on vectors):

$$|z_1| - |z_2| \le |z_1 - z_2| \le |z_1| + |z_2|$$

The polar representation makers it very easy to multiply two complex numbers, i.e.,

$$z_1 z_2 = r_1 r_2 e^{i(\theta_1 + \theta_2)}$$

So that:

- $|z_1 z_2| = |z_1||z_2| = r_1 r_2$
- $\arg(z_1 z_2) = \arg(z_1) + \arg(z_2) = \theta_1 + \theta_2$

Functions of complex variables

$$w(z) = u(x,y) + iv(x,y)\,, \quad where \;\; z = x + i$$

Example

$$f(z) = z^2 = x^2 - y^2 + 2ixy$$

Real and imaginary parts of a function

$$\text{Re}\, w(z) = u(x,y)$$

$$\text{Im}\, w(z) = v(x,y)$$

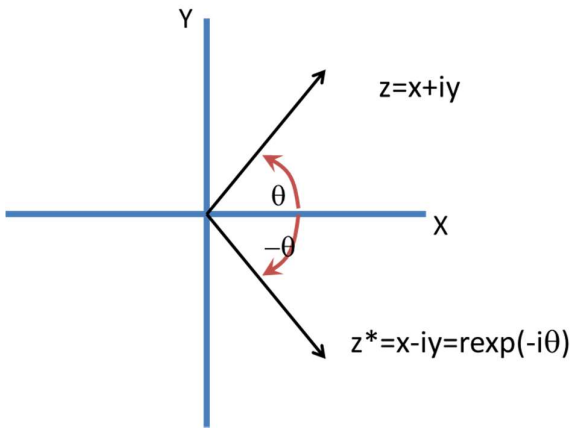
Note that both the real and imaginary parts of a function are real

Also, note that a complex function is a mapping from the complex plane to itself (see graph).

Complex conjugation : $z = x + iy,\;\; z^* = x - iy,$

Note that if we define $z = re^{i\theta}$, then $z^* = r\, e^{-i\theta}$

Also, $|z|^2 = zz^*$ (check this!, and see figure below)



We can analytically continue all elementary functions into the complex plane; however, watch out for multiply valued functions!

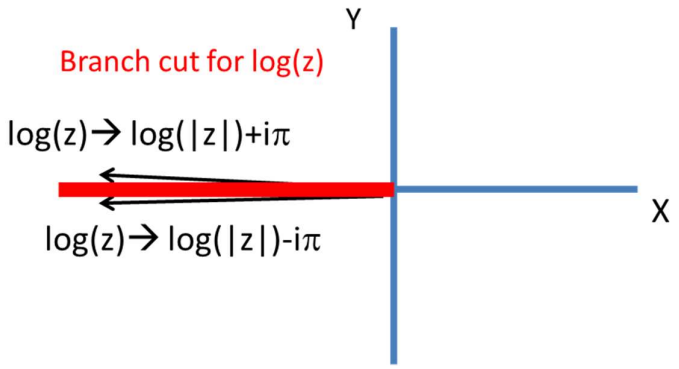
The best known example is:

$$\log(z) = \log\left(re^{i\theta}\right) = \log(r) + i$$

When we add $2n\pi$ to θ, z will not change ($e^{in\theta} = \cos(n\theta) + i \quad (\theta) = 1$) but log(z) will increase by $2n\pi$, i.e.,

$\log(z) = \log(r) + i\,(\theta + 2\pi n),\;\; n \in Z$ (Z means the group of integers, nothing to do with z, the typical suymbol for a complex number).

Thus log(z) is a multivalued function; to set it to singly valued we typically take n=0, and set $-\pi < \theta < \pi$, i.e., the negative x axis is a branch cut (see figure).

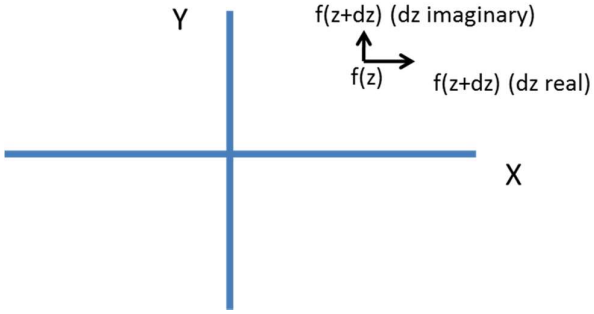


Cauchy Riemann Conditions for a function to be analytic and the derivatives of complex functions

Can we define df/dz for f(z)?

How about $\frac{df}{dz} = lim_{\delta z \to \infty} \frac{(f(z+dz)-f(z))}{dz}$

For the answer to be meaningful, it should be independent of the direction, i.ie., dz could be small and real, or small and imaginary, or any other direction (see figure).



Turns out it is enough to consider the two extreme choices (δz real or imaginary) to derive the necessary conditions for the derivative to be meaningful.

Let

$$\delta z = \delta x + i\,\delta y$$

$$f = u + i \;\rightarrow\; \delta f = \delta u + i\delta v$$

$$lim\frac{\delta f}{\delta z} = \frac{\delta u + i}{\delta x + i\,\delta y}$$

Say $\delta y = 0$, then:

$$lim\frac{\delta f}{\delta z} = \frac{\delta u + i\delta v}{\delta x} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}$$

But if $\delta x = 0$, then

$$lim\frac{\delta f}{\delta z} = \frac{\delta u + i\delta v}{i\delta y} = \frac{\partial u}{i\partial y} + \frac{\partial v}{\partial y}$$

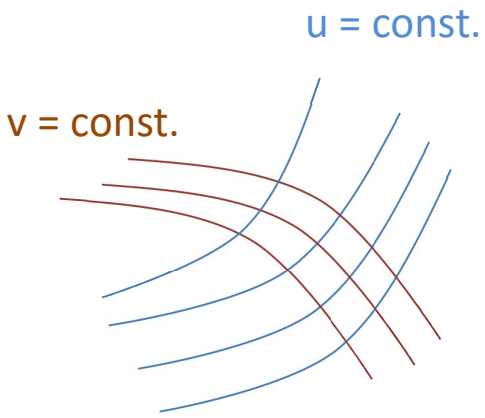
Comparing the 2 expressions, we need to have (since $\frac{1}{i} = -i$) :

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$$

For the derivative to be well defined, these Cauchy Riemann conditions need to be held.

The converse is also true, i.e., if the Cauchy Riemann conditions are true, then the derivative is well defined, i.e., it will be the same no matter what direction you come at. I won't prove that, but the proof is straightforward.

Corollary: From the Cauchy-Riemann conditions, we infer that const. u surfaces in the x-y plane are perpendicular to the const. v sources (see figure).



Proof: The const. u surfaces are perpendicular to ∇u, and similarly for v, so for the surfaces to be perpendicular we need ∇u to be perpendiucular to ∇v, i.e., we need to have $\nabla u \cdot \nabla v = 0$. But

$$\nabla u \cdot \nabla v = \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial \mathrm{v}}{\partial y} = \frac{\partial u}{\partial x}\frac{\partial u}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial \mathrm{u}}{\partial x} = 0, \qquad \text{Q. E. D.}$$

Analytical Functions

If f(z) has a well-defined derivative at a point z_0 and in some region about it, we'll call it differentiable.

Examples:

$f(z) = z^2$ analytic

$f(z) = z^*$ not analytic (prove!)

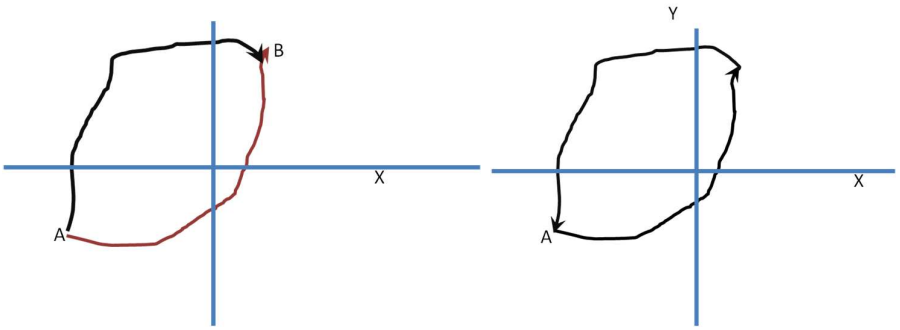
Cauchy integral theorem:

Define, along a path, $\int f(z)dz = \sum_j f(z_j)(z_{j+1} - z_j)$ (see figure below), just like we do for real functions, but along a path in the x-y plane.

Turns out that for analytic functions (in a given region) the path does not matter, since in an analytic region (i.e., a region where the function f(z) is analytic), then

$$\oint f(z)dz = 0 \quad (if\ f(z)\ analytic\ throughout\ the\ region\ "R"\ enclosed\ by\ the\ loop\ "C").$$

(Analogous to conservative quantities in thermodynamics).



Left: for a function f(z) analytic throughout the region enclosed by the two paths, the integral $\int f(z)dz$ from A to B equals in both paths (or any path in between). Right: equivalently, for a function analytic within a closed region, the closed-loop integral $\oint f(z)dz$ vanishes.

Proof: assume for simplicity that the derivative is continuous.

Part 1 of the proof is to note that:

$$\oint f(z)dz \quad \oint(u+iv)(dx+id\ ) = \oint(udx - vdy) + i\oint(udy + v\quad)$$

Part 2 is: separately, use Stokes theorem, for a vector $\mathbf{B} \equiv (B_x, B_y)$ and a path defined as $dl=(dx,dy)$ enclosing the flat area (with a volume element denoted by $d\mathbf{a} = (0,0,dx\,dy)$

$$\oint \boldsymbol{B} \cdot d\boldsymbol{l} = \int (\nabla \times \boldsymbol{B}) \cdot d\boldsymbol{a}$$

i.e.,

$$\oint (B_x \cdot dx + B_y dy) = \int \left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right) dxdy$$

Part 3 of the proof: combine parts 1 and 2 in two different ways.

3.i) First, define $B_x = u, B_y = -v$: then,

$$\oint (udx - vdy) = \oint (B_x \cdot dx + B_y dy) = \int \left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right) dxdy$$
$$= -\int \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) dxdy = 0$$

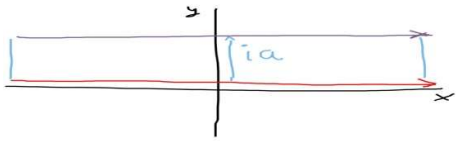
3.ii) Analogous, with $B_x = v, B_y = u$ → analogous proof leads to $\oint (udy + vdx) = 0$.

Corollary:

If a function vanishes close to the real axis and we have an integral along the real axis, we can shift that integrand up or down (i.e., add to it a constant or varying imaginary part, as long as in the region between the real axis and imaginary part the function has no singularities, i.e.,

$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} f(x + ia_0)dx$ (where a_0 can be dependent on x, as long as it is bounded—under the assumption that f has no poles between the real axis and a shift of it by ia_0 and if f(x+iy) vanishes when $|\boldsymbol{x}| \to \infty$ and y is fixed.

Proof: see figure;



Basically the difference between the integrals is the full contour integral minus the "sides" of the contour (light blue on both sides), and, if the function vanishes far away (at $|x| \to \infty$), we are left with the contour integral, which vanishes since there are no singularities.

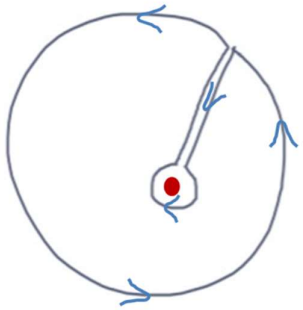
Integral over $1/(z-z_0)$

Next, we prove that

$$\oint \frac{f(z)}{z - z_0} dz = 2\pi i\, f(z_0)$$

With the same conditions the curve "C", **if it is counter clockwise (**otherwise you get a negative sign), **and f non-singular in region.**

Proof: let's assume that the integral is counterclockwise. No matter what the contour looks like is, we can change it (see figure) to an integral of a tiny loop labeled "T" around $z=z_0$, defined as $z = z_0 + re^{i\theta}$, where now r is tiny.



Equivalence of $\oint \frac{f(z)}{z-z_0} dz$ in a general contour around a singularity to an integral on a tiny loop (z_0 is labeled by a red point).

The reason is that, as indicated in the figure, we can make a single contour which encloses the outer region counterclockwise and the inner region clockwise (and a connecting line with two canceling contributions); that line integral will be therefore enclosing a region where the function is analytic (as it does not including the tiny inner region) and that line integral will therefore vanish.

I.e. (where "T" is the tiny loop)

$$0 = \oint_C \frac{f(z)}{z-z_0} dz - \oint_T \frac{f(z)}{z-z_0} dz$$

So

$$\oint_C \frac{f(z)}{z-z_0} dz = \oint_T \frac{f(z)}{z-z_0} dz$$

But in "T",

$$z = z_0 + re^{i\theta}$$

$$dz = d\left(r\ ^{i\theta}\right) = rde^{i\theta}\ (as\ r\ is\ fixed\ along\ \mathrm{T}) = re^{i\theta} id\theta$$

So

$$\oint_T \frac{f(z)}{z-z_0} dz = \oint_T \frac{f\left(z_0 + re^{i\theta}\right)}{re^{i\theta}} ire^{i\theta} d\theta = \oint_T f\left(z_0 + re^{i\theta}\right) id\theta \sim \oint_T f(z_0) id\theta$$

$$= f(z_0) i \oint d\theta = f(z_0) i 2\pi$$

Q.E.D.

Note that we were able to approximate f(z)~f(z_0) since the rest of the integrand had a well-defined limit.

Laurent series

We can extend this "Laurent Integral" to any exponent by considering the general integral (with n positive or negative but $n \neq 1$)

$$\oint_T \frac{dz}{(z-z_0)^n} = 0 \quad (n\ is\ integer, \neq 1).$$

Proof: we can consider a circular trajectory as before (the results will be the same for any trajectory in the same direction)

$$\oint_T \frac{dz}{(z-z_0)^n} = \oint_T \frac{ire^{i\theta}d\theta}{r^n(e^{i\theta})^n} =$$

$$= \frac{i}{r^{n-1}} \oint_T e^{-i(n-1)\theta}d\theta = \frac{i}{-i(n-1)ir^{n-1}} e^{i(n-1)\theta}|_0^{2\pi} = 0.$$

Note that the integral does not vanish for n=1, since then we cannot divide by n-1; for n=1 is it $2\pi i$, so when we combine:

$$\oint_T \frac{dz}{(z-z_0)^n} = \begin{cases} 0 & n\ integer, \neq 1 \\ 2\pi i & n = 1. \end{cases}$$

We can use this when we consider a general function, even one which is not analytic, but which has potentially a singularity or set of singularities around z_0; i.e., when we consider a Laruent series, which is an extension of Taylor series:

$$f(z) = \sum_{m=-\infty}^{\infty} a_m (z-z_0)^m.$$

(Note that if the function is analytic at and near $z_0$then $a_0 = f(z_0)$ and $a_n = 0$ for n<0.) ;

Examples

Example 1:

The function$\frac{1}{(1-z)}$ has a taylor series

$$\frac{1}{(1-z)} = 1 + z + z^2 + \cdots$$

So the function $\frac{1}{z(1-z)}$ has a Laurent series $\frac{1}{z(1-z)} = \frac{1}{z} + 1 + z + \cdots$
Of course this expansion only converges for $0 \neq |z| < 1$;

For this expansion, $a_{-1} = a_0 = a_1 = \cdots = 1$

Example 2

A "severely nonanalytic" function is

$$exp\left(-\frac{1}{z}\right) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n! z^n} = \sum_{n=-\infty}^{0} \frac{(-1)^n}{|n|!} z^n$$

Back from the examples to Laurent series

Inserting the Laurent series into the fundamental integral over a loop, we get that

$$\oint_T \frac{f(z)dz}{(z-z_0)^{n+1}} = \sum_{m=-\infty}^{\infty} \oint_T \frac{a_m (z-z_0)^m dz}{(z-z_0)^{n+1}} = \sum_{m=-\infty}^{\infty} a_m \oint_T \frac{dz}{(z-z_0)^{n+1-m}}$$

and since the integral vanishes unless n+1-m=1, i.e., unless n=m, we get

$$\oint_T \frac{f(z)dz}{(z-z_0)^{n+1}} = 2\pi i\, a_n$$

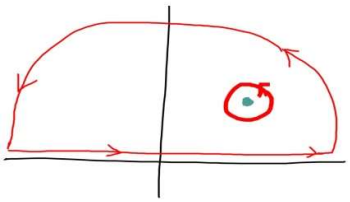
Residue theorem:

The integral expressions we got are fundamentally important. Let's take the case of *n=-1* ; then, we get

$$\oint_T f(z)dz = 2\pi i\, a_{-1}$$

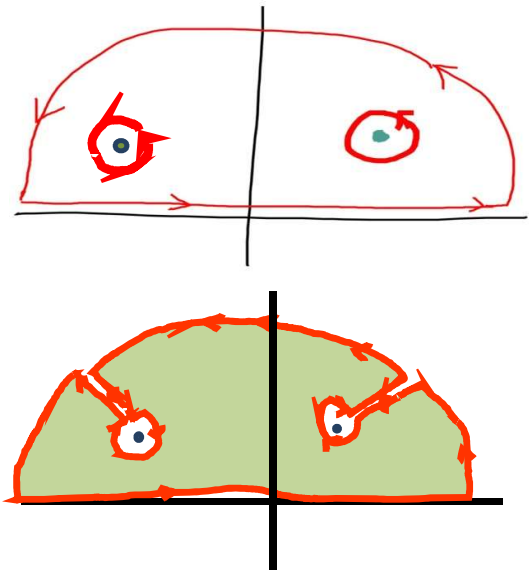
The a_{-1} term is called a "residue".

Note that we don't even need to specify in the integral where z_0 is! It is enough to know that there is a place within the contour where the function has a $\frac{1}{z-z_0}$behavior!



Similarly, as shown in the figure below, if the integral encloses a region with several singularities, and a function "sufficiently decays" we just need to add the contribution of each one, i.e.,

$$\oint_T f(z)dz = 2\pi i \sum residues\ within$$



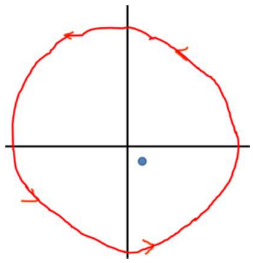
Proof that the closed contour integral over several singularities gives the same as the sum over all the residues within (left figure); as the bottom figure shows, the difference between the total integral and the integral over the residues is a closed-loop contour integral over a region (denoted by a green color) without singularities, so the total contour integral vanishes.

Corollary: analytic functions can't decay at all direction at infinity

A corollary of the residue theorem is that a singly-defined function without singularities cannot decay to 0 at all directions as |z|→ ∞; specifically, take a point z_0 where the function does not vanish, and consider the integral

$$\oint \frac{f(z)}{z - z_0} dz = 2\pi i\, f(z_0)$$

where now the integral is over a very large circular contour (see figure; the dot close to the center represents z_0). If the function would have decayed to 0, the left hand side would have vanish as the circle radius vanish, but the right hand side is non vanishing, so that would have been a contradiction.



There are stronger theorems which we won't prove which state that any well-defined function must either have a singularity or, alternately, diverge in some direction as $|z| \to \infty$;

Example – $f(x) = exp(-x^2)$; extending to the complex plane we get $f(z) = exp(-z^2)$; let's take z along, say, the positive imaginary axis, i.e., $z = i \cdot |z|$, then

$$f(z) = exp(-i^2 * |z|^2) = exp(|z|^2)!$$

Residue Example:

Define an important integral for quantum mechanics:

$$J = \int_{-\infty}^{\infty} \frac{e^{iwt}}{w - E_0 - i\eta} dw,$$

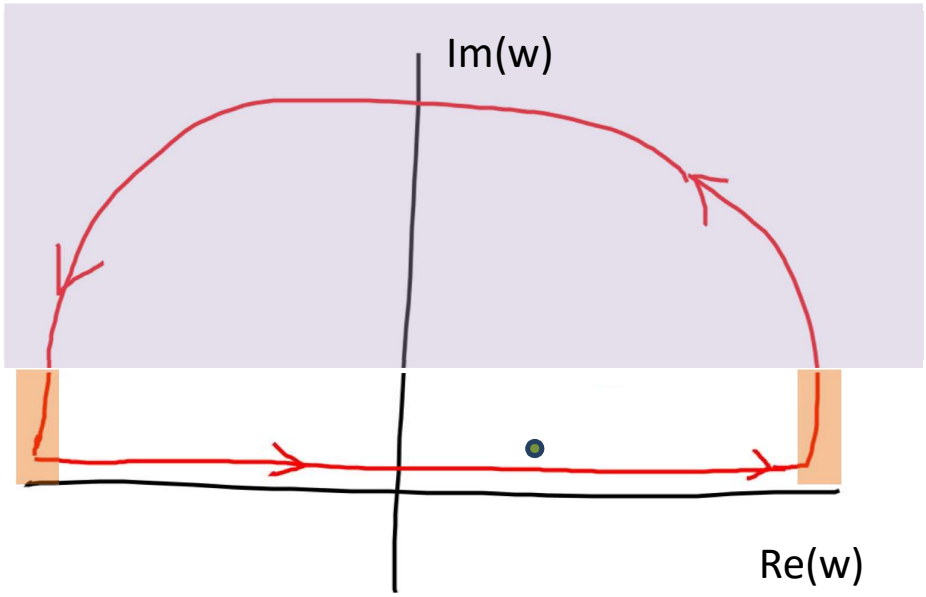
where η is positive but could be tiny (denoted sometimes therefore as $\eta = 0^+$), and t>0.

This is obviously an integral of the form

$$J = \int_{-\infty}^{\infty} \frac{f(w)}{w - w_0} dw.$$

We just need to transform this a closed contour one (counter-clock-wise, or change the sign if clock-wise).

We do that by adding to the integral a "half-circle contour" in the complex w plane, at large values of |w| in the positive half part (i.e., $Im\omega \geq 0$).



Proof that $J = \int_{-\infty}^{\infty} \frac{e^{iwt}}{w-E_0-i\eta} dw$ can be extended to the complex plane. In the purple region exp(iwt) is very small (since Im(w) is sufficiently large). In the orange region, the $1/(w - E_0 - i\eta)$ part is small and the length of the orange region is finite, so the contribution of the orange region vanishes when the circle is sufficiently large. The blue dot denotes $E_0 + i\eta$.

In the limit of a large radius of the half-circle, the integral will vanish, because, as shown in the figure, throughout most of the half-circle the contribution of exp(iwt) will be vanishingly small – since when Imω is large and positive, then

$$\exp(iwt) = \exp(i(w_r + iw_i)t) = \exp(iw_r t)\exp(-w_i t)$$

will vanish (due to the $\exp(-w_i t)$ term). Also, the $1/\ (\omega - E_0 + i\eta)$ will make the integral vanishingly small in the finite length parts of the semicircle which are close to the real-w axis (the "orange" regions)

So therefore:

$$J = \oint \frac{f(w)}{w - w_0} = 2\pi i\, f(w_0)$$

where

$$w_0 = E_0 + i\eta.$$

The residue is

$$residue = f(w_0) = \exp(iw_0 t) = \exp(iE_0 t)\exp(-\eta t).$$

So

$$J = 2\pi \mathrm{i}\ \exp(iE_0 t)\exp(-\eta\ ).$$

Note that if $\eta < 0$, the singularity would have lied outside the contour and the integral would have vanished!

Cauchy's Principle Value

So far we handled integrals where the pole is outside the contour;

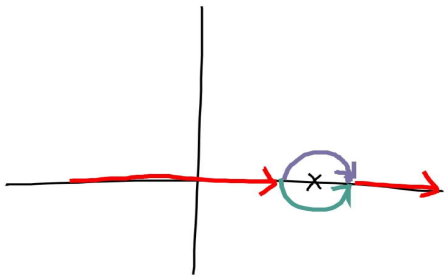
Now we'll consider a 1st order pole (i.e., a function which diverges like const./(z-z_0) for some value of z_0) if the position of the pole (z_0) lies along the contour. So we'll represent such functions as $\frac{g(x)}{x-x_0}$.

Assume the contour and the singularity are along the real-axis (i.e., replace z_0 by x_0) for simplicity.

Usually the integration limits are minus-infinity to infinity, and in that case we abbreviate and not put them at all:

$$P\int \frac{g(x)}{x-x_0}dx = \int_{-\infty}^{x_0-\epsilon} \frac{g(x)}{x-x_0}dx + \int_{x_0+\epsilon}^{\infty} \frac{g(x)}{x-x_0}dx$$

where $\epsilon \to 0$.



Principal value integral (red); the purple and green curves denote the extra loops needed to continue the integral above or below the real axis.

We can relate this to a contour integral which goes either above or below the singularity; specially, (see figure, and note that if the integral is above it is clockwise, i.e., minus the direction we usually consider):

$$\int_{-\infty,above}^{\infty} \frac{g(x)}{x-x_0}dx = P\int \frac{g(x)}{x-x_0}dx + \int_{x_0-\epsilon,above}^{x_0+\epsilon} \frac{g(x)}{x-x_0}dx$$

$$= P\int \frac{g(x)}{x-x_0}dx + \int_{\theta=\pi}^{\theta=0} \frac{g\left(x_0+\epsilon e^{i\theta}\right)}{\epsilon e^{i\theta}}\epsilon de^{i\theta}$$

$$\cong P\int \frac{g(x)}{x-x_0}dx + g(x_0)\int_{\theta=\pi}^{\theta=0} id\theta$$

So:

$$\int_{-\infty,above}^{\infty} \frac{g(x)}{x-x_0}dx = P\int \frac{g(x)}{x-x_0}dx - i\pi g(x_0)$$

Similarly

$$\int_{-\infty,below}^{\infty} \frac{g(x)}{x-x_0}dx = P\int \frac{g(x)}{x-x_0}dx + i\pi g(x_0)$$

Note that $\int$ below-$\int$ above = counter-clokcwise-circular-integral=$2\pi i g(x_0)$, as expected.(see figures above).

Example: diffraction integral, $2\int_0^\infty \frac{\sin(x)}{x} dx$

We can use the formulae we did, with $x_0=0$, and since sinx/x is an even function, the overall integral equals

$$2\int_0^\infty \frac{sin(x)}{x} dx = \int_{-\infty}^\infty \frac{sin(x)}{x} dx$$

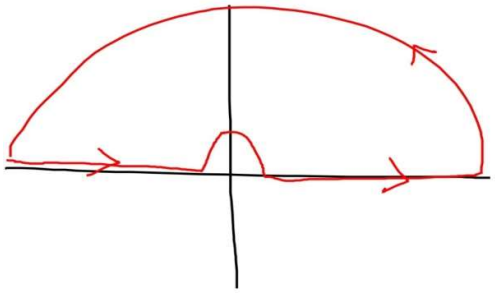
Also, since sinx/x is a well behaved function near zero, the integral equals the principle value:

$$\int_{-\infty}^\infty \frac{sin(x)}{x} dx = \mathrm{P}\int_{-\infty}^\infty \frac{\sin(x)}{x} dx$$

So since sin(x)=Im exp(ix)

$$2\int_0^\infty \frac{sin(x)}{x} dx = \mathrm{Im\ P}\int_{-\infty}^\infty \frac{e^{ix}}{x} dx$$

Consider however the contour integral from above (see figure)



$$0 = \oint_{above} \frac{e^{ix}}{x} dx = \mathrm{P}\int \frac{e^{ix}}{x} dx - i\pi$$

(where we used $\frac{sinx}{x} \to 1\ as\ x \to 0$)

So

$$\int_0^\infty \frac{sin(x)}{x} dx = \frac{1}{2} Im\ i\pi = \frac{\pi}{2}$$

Saddle point integration – steepest descent

(For further details, I recommend the treatment of Mathews and Walker,).

Warm-up – Steepest descent for real integrals.

Consider integrals if the general form

$$I = \int g(x)dx$$

Where g is peaked within the limits of the integral (or close to the contour of integration, and for now will be positive. We'll represent the integral as

$$I = \int e^{\alpha f(x)}dx$$

Where either f is naturally peaked, in which case we can take $\alpha = 1$, or f is a general function, not necessarily very peaked, and then we need to have $\alpha \gg 1$ in order for the following discussion to be valid.

The idea will be to find where the integral is peaked, i.e., $df/dx(x = x_0) = 0$, and then write

$$f(x) \sim f(x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$$

The idea is that this expansion generally works very well, since when it fails it is at values of x which are sufficiently far from the maximum that when we exponentiate the function the contributions will be insignificant compared with the maximum contribution, i.e.,

$e^{\alpha f(x)} \ll e^{\alpha f(x_0)}$ when the expansion above fails, so it makes no error in the integral.

Also, for above, notice that $f''(x_0)$ is negative, since f(x) is a maximum at x_0 .

Then we can integrate as following:

$$\int e^{\alpha f(x)} \sim e^{\alpha f(x_0)} \int e^{\frac{\alpha}{2}f''(x_0)(x-x_0)^2}.$$

To proceed, we need:

Gaussian integral interlude:

To remind you,

$$J \equiv \int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$$

Proof: square the integral, and change the variables in one of the integrals from x to y,

$$J^2 = \int_{-\infty}^{\infty} e^{-x^2}\, dx \int_{-\infty}^{\infty} e^{-y^2}\, dy = \iint e^{-r^2} dxdy,$$

where we defined $r^2 = x^2 + y^2$. Further, change coordinate system from x,y to r, ϕ, so $dxdy = rdrd\phi$, and define a new variable $q = r^2$, so $2rdr = dq$, so:

$$J^2 = \iint e^{-r^2} rdrd \qquad 2\pi \int_{-\infty}^{\infty} e^{-r^2}\, rdr = \pi \int_0^{\infty} e^{-q}\, dq = \pi, \quad \text{Q. E. D.}$$

Therefore, with a simple shift of variables:

$$\int_{-\infty}^{\infty} \mathrm{e}^{-ax^2}\, dx = \sqrt{\frac{\pi}{a}}$$

Corollary: Gaussian integrals:

$$\int_{-\infty}^{\infty} \mathrm{e}^{-ax^2+bx+c}\, dx = \sqrt{\frac{\pi}{a}} e^{\frac{b^2}{4a}+c}$$

Proof: define now $q = x - \frac{b}{2a}$, it is easy to prove that

$$-a\ ^2 + bx + c = -aq^2 + \frac{b^2}{4a} + c$$

By converting the integration variable from x to q, and dq=dx, we prove the expression.

END OF INTERLUDE

So back to steepest descent; from the integral above, we continue

$$\int e^{\alpha f(x)} dx \sim e^{\alpha f(x_0)} \int e^{\frac{\alpha}{2} f(x_0)(x-x_0)^2}$$

$$\int e^{\alpha f(x)} dx \cong \sqrt{\frac{2\pi}{-\alpha\ ''(x_0)}}\, e^{\alpha f(x_0)}$$

Where the minus sign is since the second derivative is negative (otherwise the integral is not well defined).

Example: The Gamma function.

$$\Gamma(\mathrm{x}+1) \equiv \int_0^{\infty} t^x e^{-t} dt$$

(when x is an integer, $\Gamma(x+1) = x!$) Write it as

$$\Gamma(\mathrm{x}+1) = \int_0^{\infty} e^f dt$$

$$f(t) = -t + x\ln(t)$$

The stationary point is at

$$\frac{df}{dt}(t = t_0) = 0 \to -1 + \frac{x}{t_0} = 0 \to t_0 = x$$

At which point, the function is

$$f(t_0) = -x + x\ln(x)$$

And the 2nd derivative is

$$\frac{d^2f}{dt^2} = -\frac{x}{{t_0}^2} = -\frac{1}{x}$$

Therefore, our prediction for the integrand is

$$f(t) \cong xlnx - \quad \frac{1}{2x}(t-x)^2$$

Therefore, we predict from the integral above (with $\alpha = 1$)that

$$\Gamma(\mathrm{x}+1) \cong \sqrt{\frac{2\pi}{-f''(\mathrm{t}_0)}}\, e^{f(t_0)} = \sqrt{2\pi x}\, e^{xlnx-} \quad = \sqrt{2\pi x}\, x^x e^{-x}$$

This is the first term in the asymptotic series for the Gamma function, i.e., Stirling's formula:

$$x! \equiv \Gamma(\mathrm{x}+1) = \sqrt{2\pi x}\, x^x e^{-x}\left(1 + \frac{1}{12x} + \frac{1}{288x^2} + \cdots\right)$$

END OF EXAMPLE.

Now we're ready for:

Saddle Point Integration of Complex Integrals

In practice, the equation we derived is valid also for complex integrals; however, proper derivation is very elucidating.

Preparation: For complex functions, no minima nor maxima, only saddle:

There's big difference between real functions and complex function:

First: Complex functions do not have a minimum of the real and imaginary parts, only extrema; and these extrema occur together, when df/dz=0

Proof;

$$f(z) \equiv u(x,y) + i \quad (x,y)$$

We know that (we proved)

$$\frac{df}{dz} = \frac{\partial u}{\partial x} + i\frac{\partial u}{\partial y},$$

i.e.,

$$., \quad \frac{df}{dz} = 0 \;\leftrightarrow \frac{\partial u}{\partial x} = 0, \frac{\partial u}{\partial y} = 0$$

The Cauchy Riemann rules are (recall):

$$\frac{\partial u}{\partial x} = -\frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = \frac{\partial v}{\partial x}$$

So when $\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = 0$, it follows that $\frac{\partial v}{\partial x} = 0, \frac{\partial v}{\partial y} = 0$.

This proves that when the real part is stationary, so is the imaginary part.

Second: At any point $\nabla^2 u = \nabla^2 v = 0$

In general (at any point, also away from the minimum) we can prove that the 2-D Laplacian of both u and v vanishes;

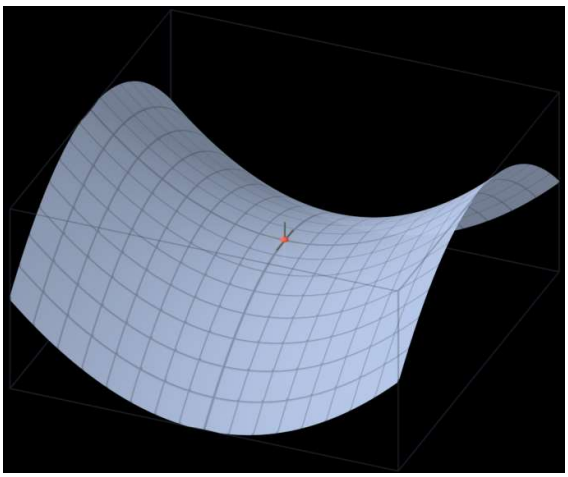
Proof: using the Cauchy Riemann theorem we get that:

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = -\frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 v}{\partial y \partial x} = 0$$

and similarly $\nabla^2 v = 0$.

The Laplacian condition means that at a maximum along x, i.e., a point where $\frac{\partial^2 u}{\partial x^2} < 0$ (the condition for a minimum, along with $\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = 0$), at that point u will have a mimimum along y, $\frac{\partial^2 u}{\partial y^2} > 0$, and vice versa (we could have isolated cases where both 2nd derivatives are zero, but the conclusion of an extremum will still be valid then, although harder to prove).

So the landscape of u looks like a saddle (see Mathews and Walker for more pictures)



Now to Saddle point integration:

Now, back to our discussion. When we have an integral in the complex plane

$$I \equiv \int e^{\alpha f} dz = \int e^{\alpha u} e^{i\alpha v} dz,$$

we have to make sure that we take a path where the real-part goes through a maximum. To get where that path is, we can again write near the stationary point

$$f(z) = f(z_0) + \frac{1}{2} f''(z_0)(z - z_0)^2.$$

Define, just for here, $-f''(z_0)$ as a complex number in polar coordinates, i.e.,

$$be^{i\chi} \equiv -f''(z_0)$$

where b, χ are real parameters (and b is positive). Further, define the deviation from z_0 also as a complex number in polar coordinates, i.e.,

$$z - z_0 = se^{i\phi}$$

so by the way

$$(z - z_0)^2 = s^2 e^{2i}$$

So putting together we find that near the stationary point (with $f_0 \equiv f(z_0)$)

$$f(z) = f(z_0) - \frac{1}{2} s^2 be^{i(2\phi+\chi)},$$

And taken together, we see that near the stationary point, the real and imaginary part of the function are:

$$u = u_0 - \frac{1}{2} s^2 b \cos(2\phi + \chi)$$

$$v = v_0 - \frac{1}{2} s^2 b \sin(2\phi + \chi).$$

The integration contour should be one where u decreases the most when we get away from the stationary point, i.e., where u looks like an inverted parabola for this, we need to have taken a direction so that "cos" is the largest, i.e.,

$$cos(2\phi + \chi) = 1 \text{ i.e., } 2\phi + \chi = 0 \; or \; 2\pi$$

$$i.e., \phi = -\frac{\chi}{2} \quad or = -\frac{\chi}{2} + \pi,$$

$$(i.e., \; e^{i\phi} = \pm e^{-\frac{i\chi}{2}}).$$

In that direction, v will be stationary to 2nd order, since then

$$sin(2\phi + \chi) = 0.$$

(A more general way to see it, albeit a little non-rigorously if you think about tit deeply, is to recall that ∇u and ∇v are orthogonal to each other, so that in the direction where u varies the most, v is stationary).

Therefore, in that direction,

$$f(z) = f_0 - \frac{1}{2} s^2 b$$

Also, when $z - z_0 = \pm se^{-i\frac{\chi}{2}}$, then: $dz = e^{-i\frac{\chi}{2}} ds$.

So:

$$I \equiv \int e^{\alpha f} dz \cong e^{\alpha f_0} e^{-i\frac{\chi}{2}} \int e^{-\alpha \frac{1}{2} s^2 b} ds = \sqrt{\frac{2\pi}{\alpha b}} e^{-i\frac{\chi}{2}} e^{\alpha f(z_0)}$$

But we defined

$$be^{i\chi} \equiv -f''(z_0)$$

So

$$\sqrt{\frac{1}{b}} e^{-i\frac{\chi}{2}} = \sqrt{\frac{1}{-f''(z_0)}}$$

And we get

$$\int e^{\alpha f(z)} dz \cong \sqrt{\frac{2\pi}{-\alpha f''(z_0)}} e^{\alpha f(z_0)}$$

i.e., exactly the same formula we had before! This is general – you often can continue formulae to the complex plane.

This is the end of the complex integration part, but you should read Mathews and Walker for more contour integration.

Appendix to complex variables: Differentiating a general (non-analytic) function w.r.t. ψ *and* ψ^*

Let's say we have a complex number or a vector or a function, ψ and we have a function G that depends on ψ and ψ^*.

$$G(\psi, \psi^*)$$

How to take derivatives of this function?

The answer is simple; let's state it first and then prove: when differentiating w.r.t. ψ keep ψ^* fixed, and vice versa, i.e., we can consider

$$\left.\frac{\partial G}{\partial \psi^*}\right|_{\psi}, \text{and} \quad \left.\frac{\partial G}{\partial \psi}\right|_{\psi^*}$$

This is very strange, i.e., how can you think of varying a function keeping its complex conjugate fixed? Fortunately, it works and is efficient and time savings, and we'll give the relations below.

Before exemplifying, let's see what this means.

Denote the real and imaginary parts of ψ as

$$\psi = u + iv$$

So

$$\psi^* = u - iv$$

An important thing to note is that on the one hand u, v are real; however, when we differentiate w.r.t. to them we don't restrict them, so formally they could be (during the differentiation!) both complex and real. This is difficult to absorb; but think of the Eqs. above then as a definition of ψ, ψ^* in general.

So we can write

$$\left.\frac{\partial G}{\partial u}\right|_{v}, \text{and} \left.\frac{\partial G}{\partial v}\right|_{u}$$

Then, formally,

$$\left.\frac{\partial G}{\partial \psi^*}\right|_{\psi} = \left.\frac{\partial G}{\partial u}\right|_{v} \left.\frac{\partial u}{\partial \psi^*}\right|_{\psi} + \left.\frac{\partial G}{\partial v}\right|_{u} \left.\frac{\partial v}{\partial \psi^*}\right|_{\psi}$$

So we need to determine $\left.\frac{\partial x}{\partial \psi^*}\right|_{\psi}$, $\left.\frac{\partial y}{\partial \psi^*}\right|_{\psi}$. This is done by writing from above

$$u = \frac{\psi + \psi^*}{2}$$

And

$$v = \frac{\psi - \psi^*}{2i}$$

So

$$\left.\frac{\partial u}{\partial \psi^*}\right|_{\psi} = \frac{1}{2} \ , \left.\frac{\partial v}{\partial \psi^*}\right|_{\psi} = -\frac{1}{2i} = \frac{i}{2}$$

So

$$\left.\frac{\partial G}{\partial \psi^*}\right|_{\psi} = \frac{1}{2}\left.\frac{\partial G}{\partial u}\right|_{v} + \frac{i}{2}\left.\frac{\partial G}{\partial v}\right|_{u}$$

And

$$\left.\frac{\partial G}{\partial \psi}\right|_{\psi^*} = \frac{1}{2}\left.\frac{\partial G}{\partial u}\right|_{v} - \frac{i}{2}\left.\frac{\partial G}{\partial v}\right|_{u}$$

The important thing is that these two quantities are linearly independent; and if we know them we can find the derivatives of the function w.r.t. the real and imaginary part. If you prefer, then think of these as the definitions of the gradients w.r.t. the function and its derivatives.

Example (related to the solution of linear problems that we talked about before; there it was for real function). Say we have a function

$$E = (ag - b)^*(ag - b) + cg^*g$$

i.e.,

$$E = (a^*g^* - b^*)(ag - b) + cg^*g$$

where c is real, g is complex (I change from ψ) and we want to set E at a minimum. Then we could have written this in terms of u, v where $g = u + iv$, i.e., we could've written:

$$E = (a^*u - ia^*v - b^*)(au + iav - b) + cu^2 + cv^2$$

And then differentiate w.r.t. u,v, and set the derivatives to zero ; but instead we differentiate the green expression above w.r.t. g^* and get

$$0 = \frac{\partial E}{\partial g^*} = a^*(ag - b) + cg$$

i.e.,

$$g = \frac{a^* b}{|a|^2 + c}$$

(Exercise: get the same expression from the gray equation above, in terms of u,v ; see how more cumbersome it is!)

Note that in this case, when we want to get an equation for g, we differentiate w.r.t. g^*. If we were to differentiate w.r.t. g, we would get an equation for g^* that will be the complex conjugate of the equation for g, so everything will work fine.

The same applies of course when we diff. w.r.t. a Hermitian conjugate (complex and transpose), but I'll leave it up to you.

GROUP THEORY (Combination of A.J. Levine's notes with Mathews and Walker, Chap.16)

Overview:

1. Definition of a group: closed, associative, identity, inverse.
2. Subgroups: invariant
3. Abelian groups
4. Examples: Continuous: Orthogonal, unitary, special unitary; finite groups – S_n, cyclic
5. S_3 example.
6. Cycles
7. The rearrangement theorem
8. Cayle's theorem: every finite group of order n isomorphic to subset of S_n
9. Classes and Invariant subgroups
10. Conjugate subgroups $(H'=gHg^{-1})$ --- identical or isomorphic ;
11. Invariant subgroups (any subgroup identical to all its conjugate subgroups).
12. Left and right cossets , gH , Hp
13. Lagrange's theorem – order of subgroups divides the group order.
14. Factor groups – set of cosets.
15. Homorphism and isomorphism
16. Symmetries
17. Representations:
    - Reducible and irreducible, multiplets
    - Properties of irreducible representations: Schur's lemma, orthogonality theorem
    - Characters
18. Continuous groups and generators;
    - Generators; Generators are Hermitian.
    - Finite rotations
    - Lee algebras
    - Examples for rotations in QM

Definition of a group:

A set of objects, G = {a, b, … } and a mupliplicication rule, such that:

- The group is closed under multiplication:

$$\text{If, a, b } \in G \text{ then } a \cdot b = G$$

- Multiplication is associative

$$(a \cdot b) \cdot c = a \cdot (b \cdot c)$$

- There exists an identity element, labeled "I", such that

$$a \cdot I = I \cdot a = a \text{ for all } a \in G$$

- There exists an inverse a^{-1} for each member of the group a, such that

$$a^{-1} \cdot a = a \cdot a^{-1} = I$$

In the following we'll often hide the $\cdot$, i.e., abbreviate ab for $a \cdot b$

A few other definitions:

- If $G' \in G$ and G′ ia closed under multiplication (see part a of the group definition), then G' is a **subgroup** of G.

A more subtle definition:

- If for any $g \in G$ and $g' \in G'$ we have that $gg'g^{-1} \in G$ then we say that G' is an invariant subgroup of G—see later.

Note the connection to a similarly transform of matrices or operators

- If $ab = ba$ for all $a, b \in G$, then G is Abelian or commutative.

Example: $\{-1, +1\}$ under multiplication: write the group multiplication table:

	$-\mathbf{1}$	$+1$
$-\mathbf{1}$	$+1$	-1
$+\mathbf{1}$	-1	$+1$

Further Examples for groups:

The first few groups exemplified are groups of matrices, where the operation is matrix multiplication:

- n*n orthogonal matrices: **O(n)**

- n*n orthogonal matrices with det. +1: **SO(n)**
- n*n unitary matrices: **U(n)**
- n*n unitary matrices with det. 1: **SU(n)**

The 4 examples above were for continuous groups.

Other examples: finite groups.

- S_n : Group of permutations of n objects.
- Cyclic groups: The cyclic group, C_n, is defined as the group of cyclical rotations of n objects. We'll exemplify these concepts on S_3 and C_3 now.

Comprehensive example: The permutation group of 3 objects "S_3":

We will label a permutation of 3 groups as, e.g., (2,3,1) which means:
If we were to start with 3 spheres, (A,B,C), then:

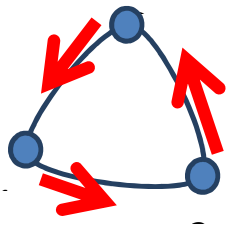
(2,3,1) permutation of (A,B,C) would have given (B,C,A).

(i.e., into the first place we will move the 2nd ball, into the 2nd place we'll move the 3rd ball, etc.).

Similarly, a permutation of 1 and 3 would be denoted as (3,1) (we could have more properly denoted it as (3,2,1), but since the 2nd ball is left in place, we often denote it for clarity as (3,1).)

The 3!=6 members of the group are:

- I – the identity member (nothing is changed)
- (3,2) – 2 and 3 are permuted
- (2,1) – 1 and 2 are permuted
- (3,1) – 1 and 3 are permuted
- (2,3,1) – see above, all 3 permuted, cyclically (see figure)
- (3,1,2) – all 3 permuted in the opposite direction.



Graphic rendering of the cyclic permutation (2,3,1)

This is a group, as we can see, i.e., the multiplication of two members is simply acting successively with each (starting from the right). For example:

$$(3,2)*(3,1,2) = (3,1)$$

Proof: Label the balls A,B,C. We start with balls ordered as (A,B,C)

We first act with the right member. Acting with (3,1,2) results in (C,A,B)

Acting with (3,2) on this (i.e., on (C,A,B)) permutes the balls in the 2nd and 3rd positions, so it gives (C,B,A), which is equivalent to what would (3,1) would have given if acted on the original (A,B,C) order.

Similarly, we can write the whole 6*6 group multiplication table, tedious but straightforward

Subgroups example:

Note that there are 4 natural subgroups for this group: permutations of 2 objects. And cyclical permutations

First subgroup: (3,2) and the identity I=(1,2,3) form a subgroup (associated with permuting just 2,3). This subgroup has the same multiplication table as that of the numbers (-1,1), i.e.,

	$(\mathbf{3},\mathbf{2})$	I
(3,2)	I	(3,2)
I	(3,2)	I

Groups which have the same multiplication table are related ("isomorphic"), as we'll show later.

Similarly, (3,1) (permuting 3 and 1) and I form a subgroup, and

{(2,1), I} also is a subgroup.

Cyclical permutations subgroup: Another subgroup is associated with switching all objects cyclically, as if they're arranged in a loop; it is called C_3. In addition to I, the cyclical subgroup contains:

- (2,3,1) (shifting all to the left, the leftmost one goes to the right)—see figure previous page.
- (3,1,2) shifting all to the right (and the right most goes to the left); it is the same as shifting twice to the left, i.e.,
  $(2,3,1)^2 = (2,3,1)*(2,3,1) = (3,1,2)$
  (Proof: do it yourself, acting on our three original spheres, A,B,C).

So the cyclical permutation subgroup is

C_3={ I, (2,31) and (3,1,2)}.

You can prove that the multiplication table for the cyclical permutations is equivalent to that of the three complex numbers, $1, \exp\left(\frac{2\pi}{3}\right), \exp\left(-\frac{2\pi}{3}\right)$. Later we'll see one matrix equivalent to this group.

Also note: in our specific case, the subgroups are all Abelian (a*b=b*a for every two members within each particular subgroup) but the overall group is not Abelian, e.g.,

$$(3,2)*(3,1,2) = (3,1) \quad \text{but}$$

$$(3,1,2) * (1,3,2) = (2,1)$$

Bigger groups have of course also non-Abelian subgroups (e.g., the permutation group of 4 objects has the permutations of the 1st three as a non-Abelian subgroup).

Now we will get to a very useful concept:

Cycles.

For each element "a", let's act by "a" once, then again, then again, etc.; i.e., we consider

$$a, a^2, a^3, a^4, \ldots a^n$$

At one point the element will return to 1, that will be the length of the cycle. For example, for our S_3 group, the order of (3,1) is 2 as (3,1)(3,1)=I .

In general:

All the elements of the cycle are different; that's easy to prove by contradiction. Say we had two elements in a cycle that were equal,

$$a^k = a^m$$

then, multiplying by $(a^{-1})^k$ we get

$$1 = a^{m-k}$$

So the length of the cycle is smaller than m, a contradiction.

From this, it follows that the length of the cycle is smaller or equal to the number of elements in the group (the "order" of the group).

One can prove an even stronger statement,

The length of the cycle has to be a divisor of the order of the group. For S_3, for example, with 6 elements, the length of the cycle can be 1, 2,3 or 6. We'll prove that "Lagrange's theorem" later.

The rearrangement theorem

In a finite group the action of an element on the set of **all** elements, **rearranges** the set.

Easy to prove:

Denote by n the size of the finite group, G, with elements $\{g_1, g_2, \dots, g_n\}$

For any $p \in G$ and $a, b \in G$, if $pa = pb \rightarrow a = b$ (multiply by p^{-1} on the left), so if $a \neq b$ then $pa \neq pb$.

So consider $pG = \{pg_1, pg_2, \dots, pg_n\}$,.

Each pg_j is distinct , as we have shown above and we have n of them, so the set pG contains exactly the same number as G and each of its elements is in G (by definition of group multiplication), so $pG = G$, Q.E.D.

Note also $pg_i \neq g_i$ unless p=I, the identity element.

Try this with our favorite group, permutations of three objects, officially denoted as S_3:

$$S_3 = \{ I, \quad (2,1), \quad (3,1), \quad (3,2), \quad (2,3,1), \quad (3,1,2) \}$$

E.g. (try this):

$$(2,3,1)S_3 = \{(2,3,1), \ (3,1), \quad (3,2) \quad (2,1), \quad (3,1,2), \quad I \}$$

so we're back to S_3 (order of elements of group is unimportant, just what the elements are!)

(Practice, if you're not sure, here's example, (2 31)(3,1) on (ABC)

= (2 3 1) on (CBA)

= (BAC)

= (3,1) on (ABC))

Cayley's Theorem: Every group G of order n is isomorphic to a subgroup of S_n, the group of permutations on n objects.

I.e., given ∈ G , we can define a permutation associated with a ; this permutation, labeled P_a, will be defined as

$$P_a = \begin{pmatrix} g_1 & g_2 & g_3 & \cdots & g_n \\ g_{a_1} & g_{a_2} & g_{a_3} & \cdots & g_{a_n} \end{pmatrix}$$

Where $g_{a_n} = ag_i$

Example: our favorite group S_3, permutations of 3 groups, has 6 elements into it. So action by a group member is by itself a permutation of these 6 objects. We can see it if we write what we got above, i.e., when we wrote S_3 in a particular order and acted on it by the element (2 3 1):

$$P_{(2\ 3\ 1)} = \begin{pmatrix} \text{I}, & (2,1), & (3,1), & (3,2), & (2,3,1), & (3,1,2) \\ (2,3,1), & (3,1), & (3,2) & (2,1), & (3,1,2), & \text{I} \end{pmatrix}$$

This is a permutation, since if we just label the 6 elements of S_3 as we wrote them as (1,2,3,4,5,6), i.e., element I is I, element 2 is (2,1), element 3 is (3,1), etc., then $P_{2,3\ 1}$ is equivalent to this permutation:

$$\begin{pmatrix} 1,2,3,4,5,6 \\ 5,3,4,2,6,1 \end{pmatrix}$$

Or just, if we abbreviate, (5 3 4 2 6 1).

Proof of Cayle's theorem:

We need to prove that this set of elements, P_a, is a group, i.e., that there is a multiplication such that multiplying two members of the set gives another member; but that's fairly obvious; in words, the multiplication of P_a and P_b will be the permutation obtained by action of a*b, i.e.,

$$P_a \cdot P_b = P_{a \cdot b}$$

For example, in our case

$$P_{(2\ 3\ 1)} = \begin{pmatrix} \text{I}, & (2,1), & (3,1), & (3,2), & (2,3,1), & (3,1,2) \\ (2,3,1), & (3,1), & (3,2) & (2,1), & (3,1,2), & \text{I} \end{pmatrix}$$

So

$$P_{(2\ 3\ 1)}P_{(2\ 3\ 1)} = \begin{pmatrix} \text{I}, & (2,1), & (3,1), & (3,2), & (2,3,1), & (3,1,2) \\ (3,1,2), & (3,2), & (2,1) & (2,1), & I, & (2,3,1) \end{pmatrix}$$

(proof: take e.g., the third element in other original group list , (3,1); $P_{2,3,1}$ maps it to (3,2), which is in turn mapped by the 2nd $P_{(2,3,1)}$ to(2,1); I used color to highlight this above.

Similarly,

$$P_{(2\,3\,1)}P_{(2\,3\,1)}P_{(2\,3\,1)} =$$

$$= \begin{pmatrix} \text{I}, & (2,1), & (3,1), & (3,2), & (2,3,1), & (3,1,2) \\ \text{I}, & (2,1), & (3,1), & (3,2), & (2,3,1), & (3,1,2) \end{pmatrix}$$

i.e.,

$$P_{(2\,3\,1)}P_{(2\,3\,1)}P_{(2\,3\,1)} = \text{I} = \text{identity (null)permutaion!}$$

Thus, in our case the cycle obtained by acting with $P_{(2\,3\,1)}$ had a length 3, i.e., after 3 iterations it went back to unity. Similarly, there are cycles with length 2. We can prove (see later) that **the length of each cycle is a divisor of the length of the group,** so for a group of length 6 (like S_3) there cannot be cycles of length 4 or 5, as 4 or 5 don't divide 6.

Classes and Invariant Subgroups

Definition:

Conjugate elements: Two elements a, b in a group G are called **conjugate**

$$a \sim b$$

if there exists an element p in the group, such that

$$\mathbf{b = pap^{-1}} \quad \textbf{conjugate elements}$$

Classes: sets of elements that are conjugate to each other.

Note:

- Each element is conjugate to itself, $a = IaI^{-1}$
- Elements are symmetrically conjugate, i.e.,

  $$a \sim b \rightarrow b \sim a;$$

  Proof: $b = pap^{-1} \rightarrow a = p^{-1}b(p^{-1})^{-1}$
- Conjugation is a transitive property:

  $$a \sim b, \quad b \sim c \rightarrow a \sim c.$$

  Proof: there exists p, g such that :

  $$b = pap^{-1}, \quad c = gbg^{-1} = gp\, ap^{-1}g^{-1} = (gp)a(gp)^{-1}$$

  (note that we used $(gp)^{-1} = p^{-1}g^{-1}$, as proved for matrices; proof:
  $= p^{-1}g^{-1}(gp) = p^{-1}g^{-1}gp = p^{-1}p = I$

The transitive property implies that **each element belongs only to one class**, since if $a \sim b$ anf $a \sim c$ then it implies that $b \sim c$, so b and c have to in the same class; i.e., all the elements that a is conjugate to are in the same class, so a cannot belong to two different classes.

Example:

Conjugate classes of S_3: three such classes:

(2,3,1), (3,1,2)
(1,2), (2,3), (1,3)
I

Definition: **conjugate subgroups.**

If $H \subseteq G$ is a subgroup, then we define

$\mathbf{H' \equiv \{aha^{-1}\ (\text{for all h} \in \text{H and a particular } a \in \text{G})\}}$

We say that H, H′ are conjugate subgroups.

Note: **conjugate subgroups are either identical or isomorphic**; the latter means that we can find one to one match of the two subgroup, and the product of two elements in one group is associated with a product of the associated elements in the other

Put differently: if $f = gh,$ and f, g, h are associated with f′, g′, h′ then $f' = g'h.$ Proof:

$$f' = afa^{-1} = agha^{-1} = aga^{-1}aha^{-1} = g'h'$$

Definition: **Invariant subgroup**

If H ⊆ G is identical to all its conjugate subgroups, then H is an invariant subgroup.

Makes sense and could be proved: H is an invariant subgroup **if** it contains elements of G in complete classes.

Example:

{I, (2,3,1), (3,1,2) } ⊆ S_3 is an invariant subgrouo

{I, (2,1) } ⊆ S_3 is NOT an invariant subgroup;

E.g., define $a = (3,1)$ (and therefore $a^{-1} = a = (3,1)$). Then,

(3,1)(2,1)(3,1) =(3,1,2),

So multiplying the (3,1) permutation with a on the left and a^{-1} on the right takes us to a different element (a conjugate element), i.e., to a conjugate subgroup which is different than the initial subgroup.

Cosets and Factor groups:

Def. of coset: Let

$$H = \{h_1, h_2, \dots\} \subseteq G$$

be a subgroup of *G*. Pick an element , p, of *G*,

$$p \in G, p \notin H$$

Then $\boldsymbol{pH} = \{\boldsymbol{ph_1}, \boldsymbol{ph_2}, \dots\}$ **is a left coset** of G, and $Hp = \{h_1 p, \dots\}$ is a right coset of G.

Claim: for each two elements of G, p and q, **either $\boldsymbol{pH} = \boldsymbol{qH}$ or they have no element in common!** This means we have a new way to break up groups.

Proof:

If $ph_i = qh_j \rightarrow q^{-1}p = h_j h_i^{-1} \in H, \quad i.e., \quad q^{-1}p = h_k$ for some $k \quad \rightarrow$

$$p = qh_k \quad \rightarrow \quad pH = qh_k H = qH$$

Example: left cosets in S_3:

$$H_1 = \{I, (231)\ (312)\}$$

First coset of H_1: H_1 itself : {(231) (312) I }
Second coset of H_1: (31) (21) (32)

Another example: left coset of H_2:

$$H_2 = \{I, (2,1)\}$$

First coset of H_2: H_2 itself : {(2,1) I }
Second coset of H_2: (3 1) (3 1 2) (obtained e.g., by multiplying p=(3,1) by H_2)
Second coset of H_2: (3,2) (2 1 3)

Lagrange's Theorem:

The order of a finite group must be an integer multiple of the order of any of its subgroups (as mentioned previously).

Proof: take a subgroup H of G

The left costs of H partition the group with each element appearing only once.

So the number n_G of group elements in G must by kn_H , where k is the number of cosets, i.e., $n_G = kn_H$ (and k integer), Q.E.D.

Factor groups:

If H is an invariant subgroup of G, the set of cosets can be taken as a group, called a "factor" or "quotient" group with the multiplication rule pH · qH = (pq)H.

This quotient group is labeled as **G/H,** the reason being that its length is n_G/n_H.

Proof/notes:

For an invariant subgroup, the left and right cosets are the same,

$$pHp^{-1} = H \quad \rightarrow \quad pH = Hp$$

Let's prove that the quotient set is really a group.

- First, closure under multiplication.

$$pHqH = \text{all products of the form: } ph_iqh_j.$$

But

$$ph_iqh_j = pqq^{-1}h_iqh_j = pqh_k,$$

where

$$h_k = qh_iq^{-1}h_j$$

Since H is invariant, qh_iq^{-1} belongs ot the group, and therefore, as H is a group, $(qh_iq^{-1})h_j$ then also belongs to the group, so h_k is a member of the subgroup H.

So:

$$pHqH = pqH \qquad \text{(for H invariant)}$$

- The identity is clear:

$$p = I \quad \rightarrow pH = IH = H$$

- Associativity clearly works, $(pH \cdot qH) \cdot rH = pH \cdot (qH \cdot rH) = pqrH$
- There is an inverse, i.e.

$$(pH)^{-1} = p^{-1}H$$

Proof: using $pHqH = pqH$ for invarient groups, we get (taking $q = p^{-1}$):

$$(p^{-1}H)(pH = p^{-1}pH = IH = H$$

Since we grouped the n_G elements of G into sets (the cosets) of size n_H), which do not overlap, we have indeed n_G/n_H elements in the factor group, G/H.

Example: H = {I, (3,12), (2,1,3)} is an invariant subgroup of S_3.

The cosets of H are H itself and another coset, labeled M, and defined as M=(1,2)H = {(31) (21) (32)}.

The multiplication table of S_3/H is

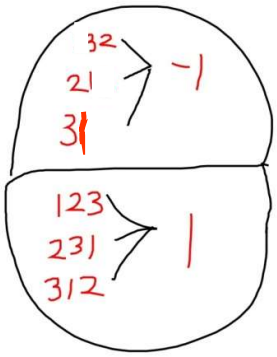
	M	H
H	H	M
M	M	H

Note that this table is the same as that of the cyclic group of two things, C_2 = {I, (2,1)}, or identically to S_2, and therefore the two groups are isomorphic (meaning essentially the same, we'll see later a precise definition)

So the factor group S_3/H is isomorphic to C_2 and H and M correspond to even and odd permutations.

In essence, the factor group corresponds to the original group where you "ignore" the differences between elements in each coset.

Here we turned S_3 into C_2 by ignoring which permutations it was and just keeping track of the evenness or oddness of the permutation.



Homomorphism and Isomorphism

Homomorphism: A mapping between one group and another that preservers multiplication.

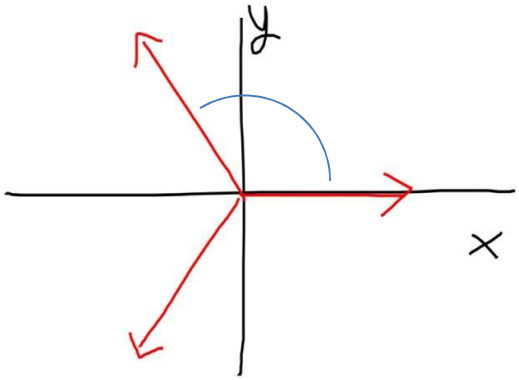
$g_i \in G \rightarrow g_i' \in G'$ and if $g_1 g_2 = g_3$ then $g_1' g_2' = g_3'$

Example: mapping $S_3 \rightarrow C_2$ (see figure)

Isomorphism: we alluded to that earlier. It is a special (one-to-one mapping) homorphism. If G and G' are isomorphic we say that:
G' is a **faithful representation** of G (and vice versa), and use the symbol $G \cong G'$.

An example is $S_3/H=C_2$ which is isomorphic to {1,-1}, i.e., {1,-1} are a faithful representation of C_2.

Another example is C_3; we already saw one faithful representation of it, 3 complex numbers , $\{1, e^{\frac{2\pi i}{3}}, e^{-\frac{2\pi i}{3}}\}$. (see figure):



Obviously, another faithful representation of C_3 is $\{1, R_2\left(\frac{2\pi}{3}\right), R_2\left(-\frac{2\pi}{3}\right)\}$; , i.e., rotations in 2D by 120 and 240 degrees (see above).

These 2D rotations can be represented therefore by the following rotation matrices

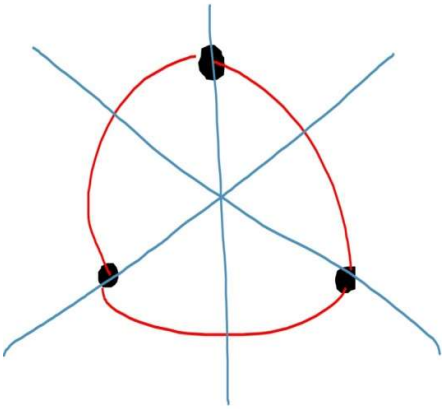
$$\left\{\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \begin{pmatrix}\frac{1}{2} & \frac{\sqrt{3}}{2}\\ -\frac{\sqrt{3}}{2} & \frac{1}{2}\end{pmatrix}, \begin{pmatrix}\frac{1}{2} & -\frac{\sqrt{3}}{2}\\ \frac{\sqrt{3}}{2} & \frac{1}{2}\end{pmatrix}\right\} = \{I, a, a^2\}$$

You can verify explicitly that indeed the 3rd matrix is the square of the 2nd (so we can call it a^2) and also $a^3 = 1$, as should be for the cyclic rotation group.

Symmetries:

As chemists we're usually interested in groups for describing symmetries;

For example, an isosceles molecule of identical atoms, X_3, will have an S_3 symmetry – in addition to the cyclic rotations C_3 it will also have **three mirror planes** (light blue in the figure) where the molecule can be reflected, see figure below.



Once one atom is modified, only one inversion symmetry remains, belonging to $S_2 \subseteq S_3$

That's a general feature – starting from a symmetric molecule or object, by changing the ingredient(s) we will usually reduce the symmetry.

REPRESANATIONS: Reducible and irreducible

Multiplets

Very important for QM and the use of symmetry and matrices in it.

Example: take a Hamiltonian H (H now denotes the Hamiltonian, not a subgroup). Say that H is invariant under a group G of symmetry transformations denoted by the operator R, i.e.,

$$H = RHR^{-1}, \qquad where\ R \in G$$

i.e., the transformed H equals the original H. (Example: translations; or rotations for a Hamiltonian with spherical symmetry, etc.)

Note that the condition above means that

$$HR = RH \ \rightarrow \ [R, H] = 0$$

This implies that there is a set of degenerate energy eigenstates related by the symmetry operation; i.e., given

$$H|\psi\rangle = E|\psi\rangle,$$

then every new vector obtained by acting by R is also an eigenstate with the same energy:

$$|\psi_R\rangle \equiv R|\psi\rangle$$

$$H|\psi_R\rangle = HR|\psi\rangle = RH|\psi\rangle = RE|\psi\rangle = E|\psi_R\rangle.$$

The full set of states $\{R|\psi\rangle,$ for all $R \in G\}$ is called a **multiplet.**

Representations

We will find matrix representation(s), $\{M(g_1), M(g_2), \dots, M(g_n)\}$ in the space of $|\psi\rangle$ so that for each element of the group, g_a, the matrix $M(g_a)$ represents the action of g on that vector

$$g_a|\psi_j\rangle = \sum_j M(g_a)_{ij} \ |\psi_j\rangle \qquad (i.e., M(g_a)_{ij} = \langle\psi_i|g^a|\psi_j\rangle)$$

A representation fulfills

$$g_a g_b = g_c \rightarrow M(g_a)M(g_b) = M(g_c).$$

If all the matrices are different (or at least linearly independent) the representation is isomorphic to G, otherwise it is homomorphic.

Note that I used "a" as a superscript, not to confuse with the basis set. Also, note that the choice of matrices and how they operate on a vector is ours – we can make it more or less complicated. This is discussed now:

Irreducible and reducible representations

There are special representations that are **irreducible :** if R can take any element of the set of vectors $V_\psi = \{|\psi_j\rangle\}$ to all other elements then it is an irreducible representation;

On the other hand, there are examples of **reducible representations.**

INSERT: Matrix Rotation Upon Basis Change:

First, let's recall what happens to the representation when we transform a basis. We did that before, but it is good exercise to recall again. When we rotate a basis

$$\chi_i = \sum_l U_{il}^* \psi_l$$

where U is unitary, then

$$M'_{ij}(g) = \langle \chi_i | g | \chi_j \rangle = \sum_{lj} U_{il}\, U_{jk}^* \langle \psi_l | g | \psi_k \rangle$$

i.e.,

$$M'(g) = UM(g)U^+$$

END OF INSERT.

Next, let's consider a particular reducible representation, labeled $D(g)$. Being reducible means that there is some subset $|\chi_j\rangle$ where acting with the group on an element of the vector does not take the result outside of a. In matrix form this means that after rotation by the corresponding U (moving from the initial basis set to the specific basis set), then

$$UD(g)U^+ = D'(g) = \begin{pmatrix} D'^1(g) & 0 \\ 0 & D'^2(g) \end{pmatrix} \qquad (reducible\ matrices)$$

Say for example there are 10 vectors in the basis, and $D^1(g)$ is a set of 3*3 matrices while $D^2(g)$ is a set of 7*7 matrices; then this means that every group member rotates the first 3 vectors among themselves, and the last seven among themselves.

In that case, $D^1(g)$ and $D^2(g)$ are two indepdendent representations and we write (dropping the primes, and remembering that this is true possibly only after matrix rotation by a unitary matrix U):

$$D = D^1 \oplus D^2$$

It is clear that we can write any representation then as a sum of irreducible representations, i.e., those that cannot be represented themselves as a further sum of other representations. The irreducible representations are thus the building blocks of any representation;

For another example, a matrix representation (i.e., a set of "h" matrices, where h is the group length—not the number of vectors!) that can be brought to the form

$$D(g) = \begin{pmatrix} D^1(g) & 0 & 0 \\ 0 & D^1(g) & 0 \\ 0 & 0 & D^2(g) \end{pmatrix}$$

will be also written as

$$D = 2D^1 \oplus D^2$$

Example: Irreducible representations of S_3 (from Mathews and Walker).

Table 16–1 The Irreducible Representations of the Group S_3

g	$D^{(1)}$ $n_i = 1$	$D^{(2)}$ 1	$D^{(3)}$ 2
[1 2 3]	1	1	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
[2 3 1]	1	1	$\frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix}$
[3 1 2]	1	1	$\frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}$
[3 2 1]	1	−1	$\frac{1}{2}\begin{pmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}$
[2 1 3]	1	−1	$\frac{1}{2}\begin{pmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix}$
[1 3 2]	1	−1	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$

Properties of Irreducible representations (won't be proven)

(1) Every representation is equivalent to a unitary representation, i.e., one in which each matrix is represented by a unitary matrix.

(2) A matrix which commutes with every matrix of an irreducible representation is a multiple of the unit matrix, i.e.,

If for all g, $AD(g) = D(g)A$, and D irreducible → A is a multiple of the unit matrix.

(3) Schur's Lemma: If $D^{(1)}(g), D^{(2)}(g)$ are two irreducible representations and there's a matrix A such that for all group members g

$$D^{(1)}(g)A = AD^{(2)}(g)$$

then either $A = 0$ (trivial case) or $n_1 = n_2$, det(A)≠ 0, and the two representations are equivalent to each other.

(4) **Practically very important: Orthogonality theorem.**

Let a group G contain h elements, and consider two irreducible representation, $D^{(i)}(g), D^{(j)}(g)$.

$D^{(i)}(g)$ is a matrix so we denote its elements as: $D^{(i)}_{\alpha\beta}(g)$. Then

$$\sum_g \left(D^{(i)}_{\alpha\beta}(g)\right)^* D^{(j)}_{\gamma\delta}(g) = \frac{h}{n_i}\delta_{ij}\delta_{\alpha\gamma}\delta_{\beta\delta}.$$

Let's think what this theorem means. Let's consider the $\alpha\beta$ elment of the i'th irreducible representation; "h" is the number of group elements, to we can think of $D^{(i)}_{\alpha\beta}(g)$ (for a particular $\alpha\beta$ *and* i) as a vector with h elements.

The orthogonality theorem says that there this vector is orthogonal to any other such vector from other irreducible representations, and is further orthogonal to other vectors from a different location (γ, δ) in the same representation.

Finally, when dotted with itself, the vector gives $\frac{h}{n_i}$

(5) Each representation "i" is made of "h" matrices which are each $n_i * n_i$ in dimension. I.e., each representation has $(n_i)^2$ independent vectors of length h associatred with it (due to the orthgooanlity therem). As there can be no more than "h" independent vectors of length h, we get

$$\sum_j n_j^2 \le h$$

It turns out that we can prove that there is an equality sign, i.e.,

$$\sum_j n_j^2 = h$$

Example for the orthogonality theorem: S_3.

We can exemplify these properties using the table for the group S_3. Let's consider $\alpha\beta = 2{,}1$ and i=3 (the 2*2 matrix representation). Then:

$$D^{(3)}_{21}(g) = \left(0, -\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{2}, 0\right) = \frac{\sqrt{3}}{2}(0,-1,1,1,-1,0)$$

- When dotted with $D^{(1)}_{11}(g) = (1,1,1,1,1,1)$ or with $D^{(2)}_{11}(g) = (1,1,1,-1,-1,-1)$the answer is obviously zero.
- When dotted with $D^{(3)}_{22}(g) = (1, -\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, 1)$ the answer is zero; similarly if we dot with $D^{(3)}_{12}(g) = \frac{\sqrt{3}}{2}(0,1,-1,1,-1,0)$ or $D^{(3)}_{11}(g) = (1, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -1)$.
- When dotted with itself, we get $\left(\frac{\sqrt{3}}{2}\right)^2 * (0+1+1+1+0) = 3 = \frac{6}{2} = \frac{h}{n_3}$

Also, $n_1 = n_2 = 1, n_3 = 2$ so $\sum_j n_j^2 = 1+1+4 = 6 = h$ $\quad Q.E.D.$

CHARACTERS

Matrix representations are not unique, since we can always do a similarity transform by any unitary matrix U:

$$D(g) \rightarrow D'(g) = S\ \ (g)S^{+}, \qquad SS^{+} = 1$$

We therefore characterize the irreducible representations by **characters,** defined as traces of the representation

$$\chi^{(i)}(g) = \sum_{\alpha} D^{(i)}_{\alpha\alpha}(g)$$

Example: S_3. Take the matrices in the previous table, and calculate the trace of each:

Table 16–2 Character Table for S_3

g	$\chi^{(1)}$	$\chi^{(2)}$	$\chi^{(3)}$
[1 2 3]	1	1	2
[2 3 1]	1	1	−1
[3 1 2]	1	1	−1
[3 2 1]	1	−1	0
[2 1 3]	1	−1	0
[1 3 2]	1	−1	0

Note

(i) The character of the identity element – here [1,2,3] – is the dimensionality of the representation, since the trace of a unit matrix is its dimensionality.

(ii) You notice immediately that within the three classes of the groups, the characters are equal.

Here are the three classes

- $C_1 = I$ (=(1,23)
- C_2 = Cyclical elements { (2,3,1), (3,1,2) }
- C_3 = Single permutations { (1,3,2), (3,2,1), (2,1,3) }

The general proof that the within each class the characters are equal is straightforward:

If g_1 and g_2 are in the same subclass, then there's an "h" in the group such that

$$g_1 = hg_2h^{-1}$$

So, in any representation (using Tr(ABC)=Tr(CAB)

$$\chi^{(i)}(g_1) = Tr\left(D^{(i)}(g_1)\right) = Tr\left(\ \ ^{(i)}(hg_2h^{-1})\right) = Tr\left(\ \ ^{(i)}(h)D^{(i)}(g_2)D^{(i)}(h^{-1}))\right)$$

$$= Tr\left(D^{(i)}(h^{-1})D^{(i)}(h)D^{(i)}(g_2)\right)$$

$$= Tr\left(D^{(i)}(h^{-1}h)\left(D^{(i)}(g_2)\right)\right) = Tr\left(\left(D^{(i)}(g_2)\right)\right) = \chi^{(i)}(g_2) \qquad Q.E.D.$$

It is a matter of a few lines proof to get from the highlighted orthogonality theorem to an analogous and more compact theorem for the characters:

$$\sum_k p_k \chi^{(i)}(C_k)\chi^{(j)}(C_k) = h\delta_{ij}$$

where the sum is over the characters, and p_k is the number of elements in each character, and we write

$$\chi^{(i)}(C_k) \equiv \chi^{(i)}(g), \qquad \text{where g is any member of } C_k$$

(Hints for the proof: start from the orthogonality theorem, pick $\beta = \alpha, \delta = \gamma$, and then sum over α, γ; the LHS has characters of matrices, and they will be identical within the p_k elements. The RHS gets an n_i factor once we sum over both α, γ)

The advantage of the present orthogonality theorem for characters is its compactness; let's label the number of classes by "s" (for our group there are s=3 classes); then the $\chi^{(i)}(C_k)$ are a sort of orthogonal vector (with weights p_k).

Table 16–3 Character Table for S_3

Class	$\chi^{(1)}$	$\chi^{(2)}$	$\chi^{(3)}$
C_1	1	1	2
C_2	1	1	−1
C_3	1	−1	0

Therefore, when we have a **reducible** representation, we can know immediately how it is composed out of the irreducible representations;

For example, say we write a general reducible representation as

$$D(g) = c_1 D^{(1)}(g) + c_2 D^{(2)}(g) + \cdots$$

We can then find the c_j,... coefficients by noting that the equailtiy holds for the traces, as for each g

$$\chi(g) = c_1 \chi^{(1)}(g) + c_2 \chi^{(2)}(g) + \cdots$$

i.e., for each class

$$\chi(C_k) = c_1 \chi^{(1)}(C_k) + c_2 \chi^{(2)}(C_k) + \cdots$$

So multiplying by the vector of characters for a particular irreducible representation, $\chi^{(j)}(C_k)$, and summing (incluing p_k), gives the coefficient of that representation:

$$\sum_k p_k \chi^{(j)}(C_k)\chi(C_k) = hc_j \ !$$

So given any representation, we just need to compute the trace of each element of the representation (or more precisely, just one trace for each class), and we can then find out, from the highlighted equation above, how this representation is computed of the irreducible representations!

Mathews and Walker give examples – read them in 436-440, and also read the physical applications part (pages 440-448)– very revealing!

CONTINOUS GROUPS AND GENERATORS

E.g., rotations by an arbitrary angle

$$R(\phi) = \begin{pmatrix} cos\phi & sin\phi \\ -sin\phi & cos\phi \end{pmatrix} = I\cos\phi + i\sigma_2 \sin\phi$$

where a reminder :

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

Also, since $(\sigma_2)^2 = I$ (prove this!), then

$$R(\phi) = \mathrm{e}^{i\sigma_2\phi} = I\cos\phi + i\sigma_2 \sin\phi$$

Proof:

$$\mathrm{e}^{i\sigma_2\phi} = 1 + i\sigma_2\phi + \frac{1}{2}(i\sigma_2\phi)^2 + \frac{1}{6}(i\sigma_2\phi)^3 + \cdots$$

And using $(\sigma_2)^2 = 1,\ (\sigma_2)^3 = \sigma_2, etc.$

$$\mathrm{e}^{i\sigma_2\phi} = 1 - \frac{1}{2}\phi^2 + \cdots + i\sigma_2\left(\phi + \frac{1}{6}(i\sigma_2\phi)^3 + \cdots\right) = \cos\phi + i\sigma_2\sin\phi, \quad \text{Q.E.D.}$$

From $R(\phi) = \mathrm{e}^{i\sigma_2\phi}$ we get that

$$R(\phi_1)R(\phi_2) = R(\phi_1 + \phi_2)$$

Consider now **infinitesimal rotations** ($\delta\phi$ *small*) , i.e., rotations which are very close to the identity:

$$R(\delta\phi) = e^{i\,\delta\phi\,\sigma_2} \cong 1 + i\,\delta\phi\,\sigma_2 = \begin{pmatrix} 1 & -\delta\phi \\ \delta\phi & 1 \end{pmatrix}$$

General definition: generators

In general a definition: for any group G, where the elements are denoted as $R(\phi)$ (whether they correspond to rotations or are something else), then a **generator** of the group G will be defined as

$$S = -i\frac{dR}{d\phi}\Big|_{\phi=0} \qquad for\ R \in G$$

Example: for the group of rotations in 2D, from the expression above we see that the associated generator is

$$S = \sigma_2$$

Theorem: every generator is traceless (i.e., its matrix representation is traceless).

Proof:

$$R \sim (1 - i\delta\phi\, S) \sim exp(-i\delta\phi\, S) = exp(A), \quad A \equiv -i\delta\phi\, S$$

Now recall what we learned at the end of the Linear Algebra part:

we can relate the determinant of any matrix to the trace of its log:

$$\boldsymbol{M = exp(B) \;\rightarrow det(M) = exp(Tr(B))}$$

Using this and noting that R is an orthogonal matrix, we get, writing R=exp(A),

$$1 = de\;(R) = exp(T\;(-i\delta\phi\, S) = exp(-i\delta\phi Tr(S)$$

So Tr(A)=0, i.e., Tr(S)=0, Q.E.D.

Another (simpler) theorem: every generator is Hermitian.

Proof:

$$R \sim (1 - i\delta\phi\, S)$$

R is orthogonal, and is therefore since R is real, it is unitary (norm conserving):

$$R^{+}R = 1$$

$$1 = (1 - i\delta\phi\, S)^{+}(1 - i\delta\phi\, S) = (1 + i\delta\phi\, S^{+})(1 - i\delta\phi\, S)$$
$$= 1 + i\delta\phi(S^{+} - S) + O((\delta\phi)^{2})$$

So $S^{+} - S = 0, \quad \rightarrow \quad S^{+} = S, \quad Q.E.D.$

So to summarize: **generators are traceless and Hermitian.**

Finite rotations:

So far we alluded to exponentiation, by writing, small rotation angle $\delta\phi$,

$$R(\delta\phi) \sim (1 - i\delta\phi\, S) \sim exp(-i\delta\phi\, S),$$

while this is true to the order we looked at (linear at $\delta\phi$ the advantage of the exponential form is that if we form a finite rotation is will still be true, i.e., writing a finite rotation as a sum of many small rotations,

$$\phi = \delta\phi + \cdots + \delta\phi$$

we get

$$R(\phi) = R(\delta\phi)R(\delta\phi) \ldots R(\delta\phi) = \exp(-i\delta\phi\, S)\exp(-i\delta\phi\, S) \ldots \exp(-i\delta\phi\, S)$$
$$= \exp(-i(\delta\phi + \delta\phi + \cdots \delta\phi\;\; S)$$

i.e.,

$$R(\phi) = e^{-i\phi S}$$

Q.E.D.

Combining several rotations: Lee algebras.

Next, consider what happens when we apply several rotations successively. **Rotations are generally non-commutative (when we go past a 2D rotation in a single plane);**

Example (try it!): take a pencil pointing up along the z-axis; then rotate a little around the z-axis (i.e., around itself), not changing its position; and rotate then a little around the y-axis, so it will end along the x-z plane.

If you were to do these two rotations in the opposite order, you'll first to rotate along the y-axis, bring the pencil to the x-z plane; and then you'll rotate around z, thereby making it point a little along y., i.e., in a different place then in the first series of rotations.

While rotations are non-commutative in a general group, we can still consider groups of rotations (in a general space) that form a continuous group (which could possibly be a subgroup of the original group). For this however to be true, we need to be able to **express every series of rotations as a new single rotation.**

Let's see what this implies on the generators. Consider action by two small rotations, one denoted by j and the other by k, each by a different small angle; we'll denote the angles, for brevity, as ϕ and ϵ, without the "deltas", but remember that they are both small. (Remember that we are not just talking about rotations in 3D, but in general dimensions).

Now let's expand the quantities to 2nd order in the angles;

$$R_j(\phi)R_k(\epsilon) = \exp\left(-i\phi S_j\right)\exp(-i\epsilon S_k)$$
$$= \left(1 - i\phi S_j - \frac{\phi^2}{2}\left(S_j\right)^2\right)\left(1 - i\epsilon S_k - \frac{\epsilon^2}{2}(S_k)^2\right)$$
$$= 1 - i\phi S_j - \frac{\phi^2}{2}\left(S_j\right)^2 - i\epsilon S_k - \frac{\epsilon^2}{2}(S_k)^2 - \phi\epsilon S_j S_k + \text{terms of 3rd order}$$

Now we want to be able to expand this as a new rotation. The new rotation should obviously be related to the sum of the two individual rotations, although with an extra amount, i.e., we write

$$R_j(\phi)R_k(\epsilon) = \exp(-i\phi S_j - i\ S_k - \mathrm{i}\Delta)$$

where Δ is an operator we need to find. We will see that Δ is of second order, so we only need to expand in terms linear in it (i.e., no need to consider terms like Δ^2, of 4th order, or terms like $\epsilon\Delta$ or $\epsilon\phi$, of 3rd order.) Let's then expand this definition:

$$R_j(\phi)R_k(\epsilon) = \exp\left(-i\phi S_j - i\epsilon\ _k - \mathrm{i}\Delta\right)$$

$$= 1 - i\phi S_j - i\epsilon\ _k - \mathrm{i}\Delta - \frac{\left(\phi S_j + \epsilon S_k + \Delta\right)^2}{2}$$

$$= 1 - i\phi S_j - i\epsilon S_k - \mathrm{i}\Delta - \frac{1}{2}\left(\phi^2\left(S_j\right)^2 + \epsilon^2(S_k)^2 + \phi\epsilon S_j S_k + \phi\epsilon S_k S_j\right)$$

$$+ \text{3rd order terms}$$

Equating the blue-highlighted equations term by term, we see that

$$-\phi\epsilon S_j S_k = -i\Delta - \frac{1}{2}\left(\phi\epsilon\ _j S_k + \phi\epsilon\ _k S_j\right)$$

i.e.,

$$-i\Delta = -\phi\epsilon S_j S_k + \frac{1}{2}\left(\phi\epsilon S_j S_k + \phi\epsilon S_k S_j\right) = \frac{\phi\epsilon}{2}\left(S_k S_j - S_j S_k\right)$$

i.e.,

$$\Delta = \frac{\phi\epsilon}{2} i[S_j, S_k]$$

It is clear therefore that in order for the overall exponent in the yellow-highlighted equation above to be a rotation, i.e., to be expressed in terms of generators, we need to have **the commutators of the generators be expressed as a sum of generators, i.e.,** we need to have:

$$\left[S_j, S_k\right] = \sum_m C_{jk}^m S_m.$$

A group of operators fulfilling this requirement is called a Lee Algebra.

Note the properties that the C's must have. First, the commutator is antisymmetric in its indices, and therefore the coefficients need to be antisymmetric:

$$\left[S_j, S_k\right] = -\left[S_k, S_j\right] \quad \rightarrow \quad C_{jk}^m = -C_{kj}^m$$

Also, in general the commutator of commutator fulfills the Jacobi identity (note the cyclical order of the operators ABC, BCA, CAB)

$$\left[A, [B, C]\right] + \left[B, [C, A]\right] + \left[C, [A, B]\right] = 0$$

(Proof: expand it, you'll get terms that cancel each other

$$ABC - ACB + BCA - CBA + BCA - BAC - CAB + ACB + CAB - CB + BAC = 0 \qquad Q.E.D.$$

Plugging in the Lie Algebra equations, we get

$$0 = \left[S_i, [S_j, S_k]\right] + \left[S_j, [S_k, S_i]\right] + \left[S_k, [S_i, S_j]\right]$$

The first term is

$$\left[S_i, [S_j, S_k]\right] = \sum_m C_{jk}^m [S_i, S_m] = \sum_{m,l} C_{im}^l C_{jk}^m S_l$$

So the Jacobi equation then (2 eqs. above us) becomes, when we rotate the (i,j,k) indices:

$$0 = \sum_{m,l} \left(C_{im}^l C_{jk}^m + C_{jm}^l C_{ki}^m + C_{km}^l C_{ij}^m\right) S_l$$

We'll assume that we took a minimal set of generators, i.e., a set that's linearly independent; so therefore for the combination to vanish each coefficient of "S_l" must vanish, i.e.,

$$\sum_m \left(C_{im}^l C_{jk}^m + C_{jm}^l C_{ki}^m + C_{km}^l C_{ij}^m\right) = 0$$

These equations are often sufficient to construct the algebra, i.e., given a few C's we can construct the rest from the Jacobi identities.

The reason we like to use generators: there are only a finite number of them in each problem. E.g., for rotations in 2D, there is an infinite number of rotation matrices R(ϕ) but just one generator.

We can carry over from the rotation matrices to the generators properties like commutation with a Hamiltonian, H, i.e., if for all ϕ

$$[R(\phi), H] = 0 \quad \rightarrow$$

then take a small angle ϕ

$$R = exp(-i\phi S) \sim \mathrm{I} - i\phi \qquad \rightarrow$$

$$0 = [I - i\phi S, H] = -i\phi [S, H],$$

i.e.,

$$[S, H] = 0.$$

I.e., the Hamiltonian commutes with the generator, so they share eigenvectors.

Example:

Look more specifically at the rotation group in 3D. What are the generators?

$$R_z(\phi) = \begin{pmatrix} cos\,\phi & sin\,\phi & 0 \\ -sin\,\phi & cos\,\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$R_x(\psi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\,\psi & sin\,\psi \\ 0 & -sin\,\psi & cos\,\psi \end{pmatrix}$$

$$R_y(\theta) = \begin{pmatrix} cos\,\theta & 0 & sin\,\theta \\ 0 & 1 & 0 \\ -sin\,\theta & 0 & cos\,\theta \end{pmatrix}$$

$$S_z = -i\frac{dR(\phi)}{d\phi}\bigg|_{\phi=0} = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$S_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}$$

$$S_y = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}$$

You can verify, that these form a Lie-Algebra, e.g.,

$$[S_x, S_z] = S_xS_z - S_zS_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}$$

$$= i^2\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - i^2\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$= -1\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} - (-1)\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} = iS_y$$

So:

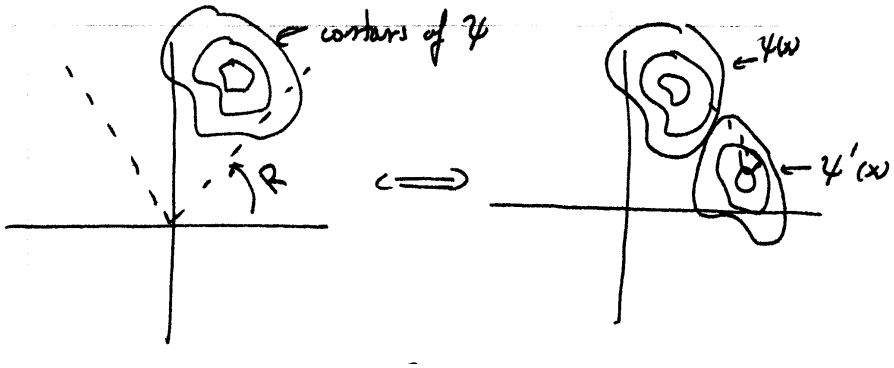
$$C_{xz}^{y} = i$$

Example: rotation operators and generators in QM:

Rotating a wavefunction. If:

$$\boldsymbol{r}' = R(\phi)\boldsymbol{r}$$

where R is a general rotation then we define (see figure):



$$\psi'(\boldsymbol{r}) \equiv R\psi(\boldsymbol{r}) \equiv \psi(\boldsymbol{r}')$$

Note that in a sense the rotation of the wavefunction is in the opposite sense to the rotation of the coordinates, just like what we learned in linear algebra!

Consider, for example, a rotation around the z-axis by a small angle $\delta\phi$.

Then,

$$\boldsymbol{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

and

$$R_z(\delta\phi) = \begin{pmatrix} cos\,\delta\phi & sin\,\delta\phi & 0 \\ -sin\,\delta\phi & cos\,\delta\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \cong \begin{pmatrix} 1 & \delta\phi & 0 \\ -\delta\phi & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ (up to order } (\delta\phi)^2)$$

So:

$$\boldsymbol{r}' = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R_z(\delta\phi)\,\boldsymbol{r} = \begin{pmatrix} 1 & \delta\phi & 0 \\ -\delta\phi & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + \delta\phi\, y \\ y - \delta\phi\, x \\ z \end{pmatrix}$$

So:

$$\psi'(x,y,z) = R_z(\delta\phi)\psi(x,y,z) = \psi(x + \delta\phi\, y, y - \delta\phi\, x, z) =$$

$$= \psi(x,y,z) - \delta\phi\left\{x\frac{\partial\psi}{\partial y} - y\frac{\partial\psi}{\partial x}\right\} + O((\delta\phi)^2)$$

i.e.,

$$R_z(\delta\phi)\psi = (1 - i\delta\phi L_z)\psi = \exp(-i\delta\phi L_z)\,\psi$$

Where

$$L_z = -i\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)$$

So the generator of the rotation operator, defined analogously to the definition on matrices we had before, is defined to be

$$Generator = i\frac{R_z(\delta\phi) - 1}{\delta\phi} = L_z$$

And, from the highlighted equation above, we get upon a finite rotation $\phi = \delta\phi + \cdots + \delta\phi$:

$$R_z(\phi)\psi = \exp(-i\delta\phi L_z)\exp(-i\delta\phi L_z)\ \ldots\ \exp(-i\delta\phi L_z)\ \psi$$

i.e.,

$$R_z(\phi)\psi = \exp(-i\phi L_z)\,\psi$$

i.e., when operating on functions,

$$R_z(\phi) = \exp(-i\phi L_z)$$

END OF GROUP THEORY!

V. Delta Functions, Continuous basis-sets, Fourier transforms.

Delta functions review:

A delta function associated with a point x_0 is a function that essentially is very small at all points except x_0 , is infinite at x_0, and its integral is 1

$$\int_{x=-\infty}^{x=\infty} \delta(x - x_0)dx = 1$$

– just like what well expect of the charge density of a classical point particle (when you integrate the charge density over all space you get a constant number, the charge; and away from the particle the charge density will be zero).

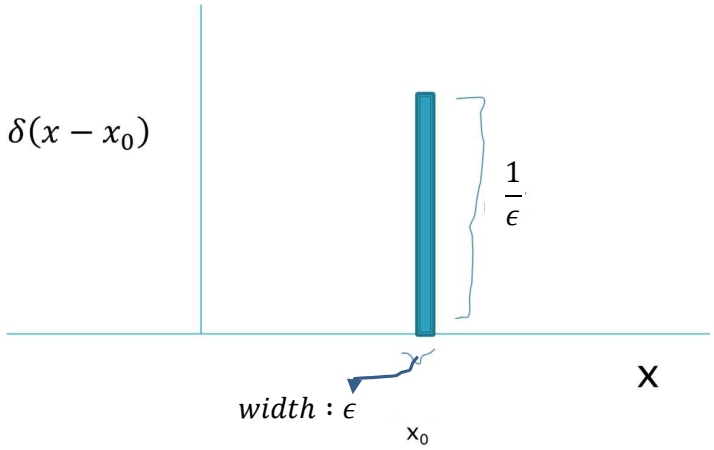
We write it as:

$$g_{x_0}(x) = \delta(x - x_0)$$

The delta function is really a limiting behavior of series of function; for example, the simplest thing will be to imagine that space is divided (we'll consider 1D mainly) to segments of length "ϵ"; then

$$\delta(x - x_0) = \frac{1}{\epsilon} \quad if\left(\ |x - x_0| < \frac{\epsilon}{2}\right); \quad 0 \ otherwise.$$

Pictorially:



There are of course smoother options, e.g.

$$\delta(x - x_0) = \frac{1}{\pi}\frac{\epsilon}{(x - x_0)^2 + \epsilon^2}$$

Where 'ϵ" is a small number, like dx above.

In general any function whose area is 1, u(x), we can turn into a delta function, by rescaling

$$\delta(x - x_0) = \frac{1}{\epsilon}u\left(\frac{x - x_0}{\epsilon}\right) \quad \text{as } \epsilon \text{ gets smaller.}$$

Exercise: do this for a normalized Gaussian, i.e., $\frac{1}{\sqrt{\pi}\sigma}\exp\left(-\frac{x^2}{\sigma^2}\right)$.

The fundamental equation for a delta-function is, that for any function f(x):

$$\int_{x=-\infty}^{x=\infty} f(x)\delta(x-x_0)dx = f(x_0)$$

We see that simply – the delta function should be zero at any point except x_0, so the only value of x where we need f(x) is at $x=x_0$, so we can approximate

$$f(x)\delta(x-x_0) = f(x_0)\delta(x-x_0)$$

And then pull $f(x_0)$ out of the integreal.

Interestingly, this highlighted expression allows us to extend the definition of the delta function even to functions which do not vanish at x different from x_0, but are very quickly oscillating; primarily

$$\frac{1}{\pi}\frac{\sin\left(\frac{x-x_0}{\epsilon}\right)}{x}$$

Note that this function has an integral 1, is going to a very large number $1/(\pi\epsilon)$ at its max peak at x=x0, but does not vanish when x is not x_0; instead it is highly oscillatory at any point except x=x0. (Plot it to be convinced). Because it is highly oscillatory except at x=x0, the yellow highlighted relation will still be valid.

Some properties of delta functions:

- **The delta function is symmetric**

$$\delta(x-x_0) = \delta(x_0-x)$$

- **Convolution of delta functions**

$$\int_{x=-\infty}^{x=\infty} \delta(x-x_1)\delta(x-x_0)dx = \delta(x_0-x_1)$$

Follows from the previously highlighted expression, by setting $f(x) = \delta(x-x_0)$.

- **Change of variables:**

$$\delta(u(x)-u_0) = \frac{1}{\left|\frac{d(u)}{d(x)}\right|}\delta(x-x_0)$$

Where x_0 is defined as the value of x where $u(x_0) = u_0$. If more than one value of x_0 can be found that fulfills $u(x_0) = u_0$, then we need to sum over all these values.

(Also, in multiple dimensions, we should replace |du/dx| by the Jacobian of the transformation, where you learned about Jacobians in calculus.)

Proof: let's assume we deal with "usual" delta functions, (i.e., not of the form sin(ax)/x), so we just need to prove the delta function vanishes everywhere except at x_0 and to prove the integral of the LHS and RHS is the same;

But we know that the LHS will vanish unless $u(x)=u_0$, i.e., unless $x=x_0$.

Further, the integral of the LHS is

$$\int \delta(u(x)-u_0)\,dx = \int \delta(u(x)-u_0)\left|\frac{dx}{du}\right| du = \int \left|\frac{dx}{du}\right| \delta(u(x)-u_0)du$$

$$= \left|\frac{dx}{du}\right|_{x=x_0} \int \delta(u-u_0)du = \left|\frac{dx}{du}\right|_{x=x_0}$$

Which obviously equals the integral of the RHS of the last highlighted equation.

- **Delta function as a derivative:**

$$\delta(x-x_0) = \frac{d}{dx}\theta(x-x_0) = -\frac{d}{dx_0}\theta(x_0-x)$$

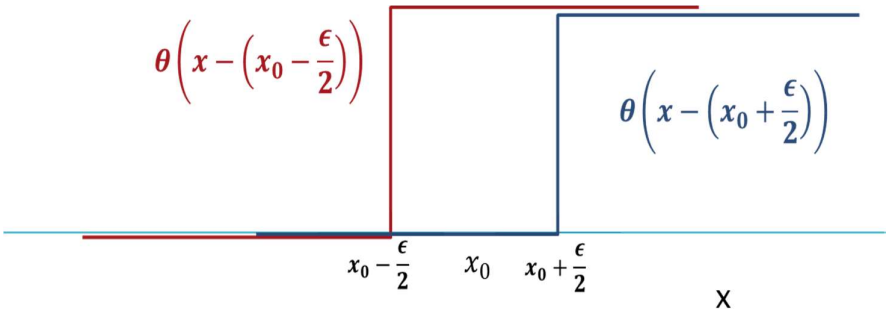
Where we introduced the **step function**:

$$\theta(z) = \begin{cases} 1 & z > 0 \\ 0 & z < 0 \end{cases}$$

This is clearly seen from the following figure, where we write the derivative as:

$$\delta(x-x_0) = -\frac{d}{dx_0}\theta(x_0-x) \cong \frac{\left(\theta\left(x-\left(x_0+\frac{\epsilon}{2}\right)\right)-\theta\left(x-\left(x_0-\frac{\epsilon}{2}\right)\right)\right)}{\epsilon}$$

Only the region where the blue and red lines do not overlap does not vanish – leading exactly to the "isolated-bin" picture of the delta function we had before.

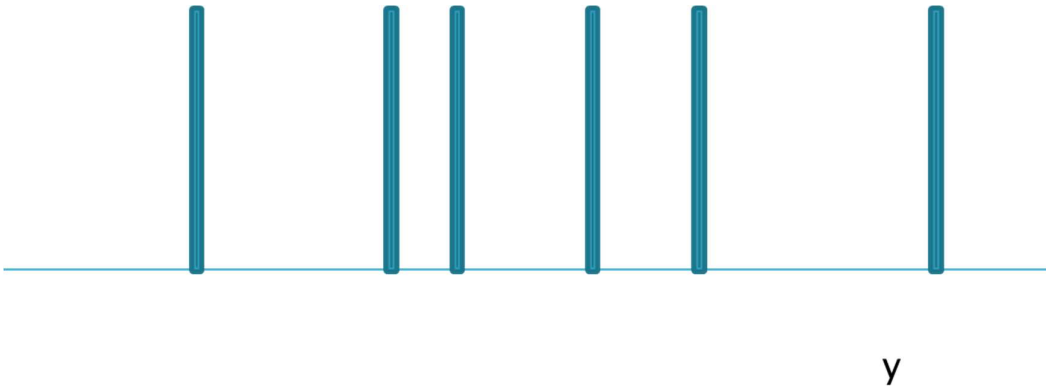


- <u>Numerical usefulness:</u>

  Often we need to represent a **"density of states", i.e.**

$$\rho(y) = \sum_j \delta(y-y_j)$$

  i.e., a bunch of "bins". For example, if we have a single-photon detector, y may be a time variable, and y_j is then the arrival time of each photon (see picture):



Or if we have high-resolution spectra, the "y" may be specific frequencies.

Or if we have grades in a large class, we may ask what's the density of grades, and their distribution.

The density of state is important since it allows us to represent summations of a function at specific points as integrals, which are often much easier:

$$\sum_j f(y_j) = \int_{-\infty}^{\infty} \rho(y) f(y) dy$$

Often, the density of states is obtained numerically best by first fitting another function:

$$G(y) = \sum_j \theta(y - y_j)$$

G will be, for example, the number of students getting higher than a score "y";

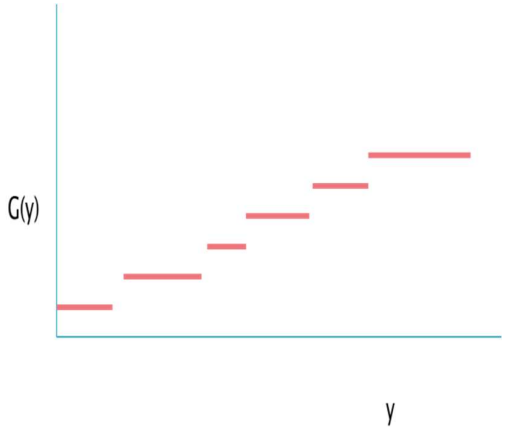
Then we will write, from the expressions before:

$$\rho(y) = \frac{dG(y)}{dy}$$

(proof: G is a sum of theta's; the derivative of each gives a delta function, i.e., gives ρ).

G is a function with "steps".

Then often the best way to represent ρ smoothly is to first fit G to a smooth function, and then differentiate it w.r.t. y.



Interestingly, this gives a compact way to calculate the density of states in QM, if we know the Hamiltonian:

$$\rho(E) = \mathrm{Tr}(\delta(H - E))$$

Proof: calculate the LHS in a basis set of the eigenvectors, so

$$\mathrm{Tr}\big(\delta(H-E)\big) = \sum_n \langle\phi_m|\delta(H-E)|\phi_m\rangle = \sum_n \langle\phi_m|\delta(\epsilon_m-E)|\phi_m\rangle$$

$$= \sum_n \delta(\epsilon_m-E)\langle\phi_m|\phi_m\rangle = \sum_n \delta(\epsilon_m-E)\,, \quad \text{Q. E. D.}$$

The highlighted expression may not sound very useful, but in practice there are modern ways to calculate it for very large matrices where direct diagonalization will be very expensive.

Interlude: delta function as integral over exp(ipx)

We need a result:

$$\int \exp(ipx)\, dx = 2\pi\delta(p)$$

To prove this, recall the Gaussian integral:

$$G(p) = \int \exp(ipx)\exp(-\eta x^2)\, dx = \sqrt{\frac{\pi}{\eta}}\exp\left(-\frac{p^2}{4\eta}\right)$$

(Reminder on the proof of the Gaussian integral: we write

$$s \equiv x - i\frac{p}{2\eta}$$

So:

$$ipx - \eta\ ^2 = -\eta s^2 - \frac{p^2}{4\eta^2}$$

$$ds = dx$$

So

$$\int \exp(ipx)\exp(-\eta x^2)\,\mathrm{dx} = \exp\left(-\frac{p^2}{2\eta^2}\right)\int \exp(-\eta s^2)\, ds$$

$$= \sqrt{\frac{\pi}{\eta}}\exp\left(-\frac{p^2}{4\eta^2}\right)$$

Note that the result of the Gaussian integral, is itself a Gaussian;

End of proof for Gaussian integral)

$\sqrt{\frac{\pi}{\eta}}\exp\left(-\frac{p^2}{4\eta^2}\right)$ is a pointed function of p, i.e., it goes to zero if $p \neq 0$ as $\eta \to 0$, so in the limit of $\eta \to 0$, it will be proportional to a delta-function; we just need to determine its own integral, i.e.,

$$\int G(p)dp \quad \int \sqrt{\frac{\pi}{\eta}}\exp\left(-\frac{p^2}{4\eta^2}\right)dp = \sqrt{\pi}\int \exp\left(-\frac{y^2}{4}\right)dy = \sqrt{\pi}\left(2\sqrt{\pi}\right) = 2\pi$$

where

$$y = \frac{p}{\sqrt{\eta}}$$

Therefore, we can write, in the limit that η is tiny, that

$$G(p) = 2\pi\delta(p) \quad (\eta \to 0)$$

Q.E.D.

Delta functions and linear algebra

Delta functions allow us to extend linear algebra to a continuous space and to a continuous basis. Those are not exactly the same thing, i.e., a continuous space can be described by a discrete or continuous or a mixture basis, but on a fundamental levels they originate from the same place.

Plan:

We will **first need to move from continuous space to discrete space**. There we'll show that "functions" in continuous space are associated with vectors in discrete space, and show that Dirac deltas map to (Kronecker deltas)/spacing.

Then we'll move **back from discrete space to continuous space,** so we can handle continuous basis.

This supposedly round-about way will teach us about continuum functions much deeper than if we were to arbitrarily introduce them.

Continuous → discrete space:

To understand continuous basis, let's think of space (1D for now, nothing is fundamentally different in 2D or 3D) as really being discrete but made of tiny separation, dx, i.e., think of space as restricted to be between a minimum value "a" to a maximum value "b", and made of discrete steps,

$$x_1 = a$$
$$x_j = a + (j-1)dx$$
$$\dots$$
$$x_N = a + (N-1)dx = b$$

where "N" will be the total number of points,

$$N = \frac{b-a}{dx} + 1$$

N could be 1000, or a million, or as much as you want, depending on how discretized the space is (how small dx is).

A continuous function f(x) will be approximated by its values at the discrete grid space, i.e., it will be represented by a vector

$$\boldsymbol{f} = (f_1, f_2, \dots f_N)^T$$

(we put a transpose, since it is easier to write a row vector, while we want our final vector to be a column vector). Here,

$$f_1 = f(x_1), \dots, f_j = f\big(x_j\big), \dots, f_N = f(x_N)$$

For example, every time you are plotting a function on the computer, you are really do this – you are "pixelating" or discretizing your function and get a vector of values.

Now, let's consider the eigenvalues and eigenfunction of the "x" operator, i.e., the eigenvalues (and eigenfunctions) of the operator defined as:

$$\boldsymbol{x}f(x_j) = x_j f(x_j)$$

(I use a bold face to designate that **x** is an operator, i.e., when it acts on a function it maps it to a different function. Later we'll remove the bold-face)

These eigenvalues ("η_n") and associated eigenfunctions, $X_n(x_j)$, fulfill:

$$x_j X_n(x_j) = \eta_n X_n(x_j)$$

Obviously, the eigenstates of this operator are "Kronecker deltas"; i.e.,

$$X_{jn} \equiv X_n(x_j) = \delta_{jn}$$

and the eigenstates are just the values of x on the grid

$$\eta_n = x_n$$

For example, the 5th eigenvector will be

$$X_{j5} = (0,0,0,0,1,0,0,0,0,0,0,0, , \dots ,0)^T$$

and

$$\eta_5 = x_5.$$

Since this is a basis, we can use now "bras" and "kets", e.g., we can define a "x" ket, which will be an eigenstate of the x operator; for example, the ket associated with the 5th point will fulfill:

$$\boldsymbol{x}|x_5\rangle = x_5|x_5\rangle$$

Or in general

$$\boldsymbol{x}|x\rangle = x|x\rangle$$

SIDE NOTE:

Don't get confused! The bold **x** is an operator, $|x\rangle$ is a vector, and the "x" on the RHS outside the vector is a number, that lables the vector.

For example, say that dx=0.01, and that a and b range are, -1000 and 1000 to pick some arbitrary numbers.

Then, the ket

$$|6.82\rangle$$

Will be part of the basis (it will be the j=100684'th ket of this basis, if you calculate). This ket IS NOT THE SAME AS

$$2|3.41\rangle$$

The former is essentially a "delta function" associated with the point x=6.82, and is a vector of length 1; the latter is a vector at length 2 associated with the point x=3.41.

END OF SIDE NOTE

The orthogonality relation is

$$\langle x|x'\rangle = \delta_{xx'}$$

Where we use the Kronecker deltas.

Further, we have as usual for a basis:

$$1 = \sum_x |x\rangle\langle x|$$

(it is highlighted in blue, **since we are going to modify it)**

So that a general function (i.e., a general vector) can be written as

$$|f\rangle = \sum_x |x\rangle\langle x|\, f\rangle = \sum_x f(x)|x\rangle$$

where we used

$$\langle x|f\rangle = f(x)$$

This relation is obvious is we think of specific value of x, for example the 5th value, x_5; then,

$$\langle x_5|f\rangle = (0,0,0,0,1,0,0,\ldots.0)(f_1, f_2, f_3, f_4, f_5, f_6, \ldots f_N)^T = f_5$$

Operators and delta functions in the discretized x basis

We learned before that operators become matrices when we considered discretized basis. For example, take the operator x. It is really a diagonal matrix in our presentation (i.e. ,in the x_j basis), so it will be

$$\begin{bmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_N \end{bmatrix}$$

Other operators are similarly defined.

Relation between Kronecker deltas and Dirac Deltas

Finally we note that when we move from Continuum to discrete representations, the Dirac delta function is transformed to a scaled Kronecker delta, i.e.,

$$\delta(x - x') \to \frac{\delta_{xx'}}{dx}$$

To see this, note that the fundamental property of Dirac delta functions is

$$\int f(x')\delta(x - x')dx' = f(x)$$

And indeed, when we use a discretized version of the integral, then

$$\sum_{x'} f(x')\frac{\delta_{xx'}}{dx}dx = f(x)$$

showing that these are the same things in a discretized space (since an integral is a sum multiplied by dx).

NOW BACK, DISCRETE → CONTINUUM

NOW WE WILL REDEFINE THE KETS in preparation to moving to a continuum basis:

$$|x\rangle \rightarrow \frac{1}{\sqrt{dx}}|x\rangle$$

Then,

$$\langle x|x'\rangle = \frac{\delta_{xx'}}{dx} = \delta(x - x')$$

The last relation is something we saw earlier, in the plot of the delta function; i.e., a Dirac delta function can be thought of as a Kronecker delta, divided by dx (there it as ϵ, here we call it dx.)

Then the "blue" orthogonality relation becomes in our language

$$\langle x|x'\rangle = \delta(x - x')$$

and

$$1 = \sum_x |x\rangle\langle x|\ dx = \int |x\rangle\langle x|\ dx$$

i.e., a sum over many values of x times dx is exactly the integral!

And therefore

$$|f\rangle = \int f(x)|x\rangle dx$$

Where we still have

$$\langle x|f\rangle = f(x)$$

Simply put, the eigenfunctions of the "x" operator are now Dirac delta functions, and they are orthogonal with a delta function normalization.

A side note: since we replaced the sum by an integral, we can also get rid of "a" and "b", i.e., assume that space (and the integrals) run from minus infinity to infinity.

Everything we learned in linear algebra still maps here; for example, the overlap of two functions can be written as (in the first equality we insert $1=\int |x\rangle\langle x|\ dx$):

$$\langle g|f\rangle = \int \langle g|x\rangle\langle x|f\rangle dx \qquad \int g^*(x)f(x)dx$$

(Note the complex conjugate, $\langle g|x\rangle = \langle x|g\rangle^* = g^*(x)$)

Now we start seeing the power of the bra-ket notation. Look at the equation for an overlap of two function from above. It tells us the we can think of an overlap as nothing more than the sum (or integral) of overlaps of our functions with the eigenstate of the "x" operator – i.e., there is nothing special about x, and we see

that we could have obviously applied the eigenstates of any other linear operator, as the book by Dirac explains.

Discrete and continuous basis sets

Discrete Basis Sets

The interesting thing is that even though we now have continuous space and continuous basis, we could still have **discrete basis sets.**

As an example, we can use fit any function by eigenfunctions of the harmonic oscillators, i.e., write

$$f(x) = \sum_n f_n g_n(x)$$

Where we use the eigenstates of the Harmonic oscillator

$$hg_n(x) = \epsilon_n g_n(x)$$

and

$$h = K + \frac{1}{2}m\omega^2 x^2$$

$$\epsilon_n = \hbar\omega\left(n + \frac{1}{2}\right)$$

K is the kinetic energy operator :

$$K = -\frac{1}{2m}\frac{d^2}{dx^2}$$

And m is the mass (don't confuse with m used as an index below).

Of course, many other Hamiltonians have discrete-only eigenfunctions (any Hamiltonian where the potential is not bounded from above, as you'll learn in Quantum Mechanics). Note that these eigenstates are orthogonal, i.e.,

$$\langle g_n | g_m \rangle = \int g_n^*(x) g_m(x) dx = \delta_{nm}$$

Also note that we could have written the equation above in terms of formal bras and kets:

$$|f\rangle = \sum_n f_n |g_n\rangle$$

Where as we learned

$$f_n = \langle g_n | f \rangle$$

Finally, all these properties can be summarized by the equation we know:

$$1 = \sum_n |g_n\rangle\langle g_n|$$

Note that the basis is discrete but infinite, i.e., the summation in the bold equation above extends from n=0 to infinity. The basis function are orthogonal as we learned, since h **is Hermitian;, i.e.,**

$$\langle\psi|h|\chi\rangle = \int \psi^*(x)\left(-\frac{1}{2m}\frac{d^2}{dx^2}+\frac{1}{2}m\omega^2x^2\right)\chi(x)dx =$$

$$\left(\int \chi^*(x)\left(-\frac{1}{2m}\frac{d^2}{dx^2}+\frac{1}{2}m\omega^2x^2\right)\psi(x)dx\right)^* = \langle\chi|h|\psi\rangle^*$$

(Obviously, the same proof also applies when we use other potentials.)

Here, we used integration by parts to get that when you integrate over all space, and your two functions decay at infinity, then

$$\int \psi^*(x)\frac{d^2}{dx^2}\chi(x)dx = -\int \frac{d}{dx}\psi^*(x)\frac{d}{dx}\chi(x)dx \qquad \int \chi(x)\frac{d^2}{dx^2}\psi^*(x)dx$$
$$= \left(\int \psi(x)\frac{d^2}{dx^2}\chi^*(x)dx\right)^*$$

The momentum basis

Before going on, let's talk about another continuum basis set, in addition to $|x\rangle$; the momentum basis, $|p\rangle$.

This basis fulfills, by definition,

$$p_{operator}|p\rangle = p|p\rangle$$

where (using the convention $\hbar = 1$):

$$p_{operator} = -i\frac{d}{dx}$$

It is easy to see that the following will fulfill this equation:

$$|p\rangle = c\int_{-\infty}^{+\infty}\exp(ipx)\,|x\rangle dx$$

(well, to do it properly mathematically we have to be quite careful, but let's not worry too much).

Think of this as summing many kets. "c" is a constant to be determined shortly

Since $p_{operator}$ is a Hermitian operator, we know that the eigenstates are orthogonal; we should look at the normalization only. We know that

$$\langle x|p\rangle = c\int_{-\infty}^{+\infty}\exp(ipx')\,\langle x|x'\rangle dx' = \int_{-\infty}^{+\infty}\exp(ipx')\,\delta(x-x')dx' = c\exp(ipx)$$

So

$$\langle p'|p\rangle = \int \langle p'|x\rangle\langle x|p\rangle dx = c^*c\int \exp\big((ipx - \quad 'x)\big)\,d$$
$$= |c|^2\int \exp(i(p-p')x)\,dx$$

And from the integral we did earlier, $\int \exp(ipx)\,dx = 2\pi\delta(p)$, we get

$$\langle p'|p\rangle = 2\pi|c|^2\delta(p-p')$$

So the choice

$$c = \frac{1}{\sqrt{2\pi}}$$

Leads to orthonormal basis

$$\langle p'|p\rangle = \delta(p-p')$$

So to conclude

$$|p\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp(ipx)\,|x\rangle dx$$

Side note: analogy between(x,p) and (p,-x).

To remind you: In general, whenever we expand an orthonormal basis I in terms of another orthonormal basis II, and then do the opposite (expand II in terms of I) the expansion coefficients are—they are the complex conjugate of each other.

The same is true for continuum orthonormal basis sets, so we can write;

$$|x\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp(-ipx)\,|p\rangle dp$$

(Formal proof: as we proved the expansion coefficient of |p> in terms of |x> is simply

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi}}\exp(ipx)$$

We reversed the role of p and x, i.e., expand |x> in terms of a basis |p>, and the expansion coefficients will be

$$\langle p|x\rangle = \langle x|p\rangle^* = \frac{1}{\sqrt{2\pi}}\exp(-ipx$$

So the relation between x and p is exactly the same as between p and –x!

This analogy can be carried even further. We can write

$$\boldsymbol{x_{operator} = i\frac{d}{dp}} \quad !$$

Proof: in a sense we don't need the proof, since we trust that p=-id/dx, and we just reverse the role of x and p and reverse one of the signs.

We'll get back to the x,p relation next chapter, when we consider Fourier transforms.

Now at last:

Mixed discrete-Continuum basis sets

It turns out the a general Hermitian operator, labeled "L", could have either a

- discrete spectrum (the eigenvalues are all separate – the harmonic oscillator),
- a continuum spectrum (all eigenvalues are continuous) – the "x'" operator is such, its eigenvalues range from $-\infty$ to $+\infty$ continuously; same for the p operator.
- **Mixed-spectrum**, i.e., regions with continuous eigenvalues and regions with isolated eigenvalues, so we can write

$$1 = \sum_n |\lambda_n\rangle\langle\lambda_n| + \int d\lambda\, |\lambda\rangle\langle\lambda|$$

So, e.g., any ket is expanded as

$$|f\rangle = \sum_n |\lambda_n\rangle\langle\lambda_n|f\rangle + \int d\lambda\, |\lambda\rangle\langle\lambda|f\rangle$$

And the discrete and continuum basis set members are orthonormal, as follow:

$$\langle\lambda_n|\lambda\rangle = 0,$$
$$\langle\lambda_n|\lambda_m\rangle = \delta_{nm}$$
$$\langle\lambda|\lambda'\rangle = \delta(\lambda - \lambda')$$

Two examples for mixed spectra:

- A particle in a box, but that box is has finite height e.g.,:

$$V(x) = \begin{cases} 0 & |x| > x_0 \\ -W & |x| < x_0 \end{cases}$$

  In that case there could be a finite number of discrete states (or zero if the well depth is too shallow) and the rest are continuum states, with energies between 0 and infinity.

- Electron in hydrogen – there are infinitely many discrete Columbic states (1s, 2s,2px,2py,2pz,etc…), but in addition there are continuum states with E>0.

Finally,

you should know that there is something which is called "resonances", which are discrete eigenstates of the Hamiltonian which are not bound, and are "buried in the continuum"; they are very interesting for predisocciation and other p process where the particle is not bound formally but in practice spends a long time looking like it is bound.

Also:

(1) **One complication on continuum states- normalization** – if we label them by a different variable, we need to change the normalization; e.g., the eigenstates of p are also eigenstates of the kinetic energy operator, with eigenvalue E= $\frac{p^2}{2m}$; so we could write

$$|E\rangle = b|p\rangle$$

(we actually have another state with the same energy with –p, but let's forget about this for a second, and assume that p and p' below are both positive). "c" can be obtained from

$$|b|^2\langle p'|p\rangle = \langle E'|E\rangle = \delta(E-E') = \delta\left(\frac{p^2}{2m}-\frac{p'^2}{2m}\right)$$

Since we assumed for now that p and p' are both positive, the argument of the delta function can only vanish if p=p', so we use the rules on delta function to get:

$$\delta(E-E') = \delta\left(\frac{p^2}{2m}-\frac{p'^2}{2m}\right) = \frac{1}{\left|\frac{dE}{dp}\right|}\delta(p-p') = \frac{1}{\frac{p}{m}}\langle p|p'\rangle$$

So

$$b^2 = \frac{1}{\frac{p}{m}} = \left|\frac{dp}{dE}\right|$$

Next: Fourier transforms.

Fourier transforms:

Plan:

- Introduce Fourier transforms from different perspectives
- Properties of Fourier transforms: gradients
- Convolutions and Filter
- Green's Functions
- Correlation functions
- Parceval theorem
- Fast Fourier transforms
- Numerical derivatives and spectral methods

We won't have time to deal with the cousins of Fourier transforms: **Laplace transforms,** but you should be aware of them,

Fourier transforms – from different perspectives.

Let's arrive at Fourier transform in somewhat different ways.

(i)Fourier Transforms from bras and kets.

Take a general "ket" $|f\rangle$. Expand it in terms of |x> and also in terms of momentum eigenvectors |p>. Let's recall that (from the previous chapter)

$$|f\rangle = \int f(x)|x\rangle dx = \int f(p)|p\rangle dp$$

$$|p\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(ipx)\,|x\rangle dx$$

So: do a Hermitian conjugate, get:

$$\langle p| = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-ipx)\,\langle x|dx$$

Also

$$f(x) = \langle x|f\rangle$$

So using the equations above, get

$$\tilde{f}(p) \equiv \langle p|f\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-ipx)\,\langle x|f\rangle dx$$

So

$$\tilde{f}(p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-ipx)\, f(x) dx$$

And similarly, if we were to express $\langle x|$ in terms of $\langle p|$ we'll get:

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(ipx)\tilde{f}(p)\, dp$$

[**Note** that we put a "tilde" above f. The reason is that we want to make sure we understand that f(x) refers to a particular representation (the "x" representation), and $\tilde{f}(p)$ refers to another representation. Put differently, make sure to distinguish f(x=1.34) and $\tilde{f}(p = 1.34)$; if we would have used the same "f" you could have got confused.

END OF NOTE]

(ii) Fourier transforms directly from delta-functions

The 2nd approach is not to worry about bras and kets to just define $\tilde{f}(p)$ from the equation above (the one before the last), and to claim that f(x) fulfils the last equation; this is easily proved as:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(ipx)\tilde{f}(p)\, dp$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(ipx) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-ipx')\, f(x')dx'\, dp$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} \exp(ip(x - \ \ ')dp \right) f(x')dx'$$

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} 2\pi\delta(x - \ \ ')\, f(x')dx' = f(x) \qquad Q.E.D.$$

(iii) 3rd approach: starting with Fourier summation

The last way is the traditional way to teach Fourier transforms, as a limit of Fourier series. Thus, given a function f(x) which is assumed periodic, i.e.,

$$f(x + L) = f(x)$$

We'll defined the function as a Fourier sum, and then see what happens to the formulae as $L \to \infty$.

Since now L is periodic, we write it in terms of sin's and cos's:

$$f(x) = \sum_{n=0}^{\infty} a_n \cos\left(\frac{2\pi n}{L}x\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{2\pi n}{L}x\right)$$

Since $\sin(\frac{2\pi}{L}x)$ and $\cos(\frac{2\pi n}{L}x)$ are linear combination of $\exp\left(i\frac{2\pi n}{L}x\right)$ and $\exp\left(-i\frac{2\pi n}{L}x\right)$we can write

$$f(x) = \frac{1}{\sqrt{\mathrm{L}}} \sum_{n=-\infty}^{\infty} g_n \exp\left(i\frac{2\pi n}{L}x\right) = \sum_{n=-\infty}^{\infty} g_n \phi_n(x)$$

where

$$\phi_n(x) \equiv \frac{1}{\sqrt{\mathrm{L}}} \exp\left(i\frac{2\pi n}{L}x\right)$$

Note that in the last highlighted equation n takes on positive and negative value (e.g.., the cos and sin with, say, n=5, appear as linear combination of the exponential terms with n=5 and n=-5. The $\frac{1}{\sqrt{L}}$ is for normalization, see below.

(Exercise: work out g_n in terms of a_n, b_n. Be careful with n=0.)

Let's prove that these function are an orthonormal basis set, where the dot-product is defined here as an integral over the period:

$$\langle \phi_{n'} | \phi_n \rangle = \int_{-\frac{L}{2}}^{\frac{L}{2}} \phi_{n'}^*(x)\phi_n(x)dx = \frac{1}{L}\int_{-\frac{L}{2}}^{\frac{L}{2}} \exp\left(i\frac{2\pi(n-n')}{L}x\right)dx$$

$$= (\text{for n} \neq \text{n}') \frac{\left(\exp\left(\frac{i2\pi(n-n')}{L}\frac{L}{2}\right) - \exp\left(-\frac{i2\pi(n-n')}{L}\frac{L}{2}\right)\right)}{\frac{i2\pi(n-n')}{L}L}$$

$$= \frac{(\exp(i\pi(n- \ \ ')) - \exp(-i\pi(n- \ \ ')))}{\frac{i2\pi(n-n')}{L}L} = 0$$

And the n=n' term is of course 1.

So we can write therefore:

$$g_n = \langle \phi_n | f \rangle = \int_{-\frac{L}{2}}^{\frac{L}{2}} \phi_n^*(x) f(x) dx$$

i.e.,

$$g_n = \frac{1}{\sqrt{L}}\int_{-\frac{L}{2}}^{\frac{L}{2}} \exp\left(-i\frac{2\pi n}{L}x\right) f(x)dx$$

Now let's define the momentum as linearly proportional to the n index:

$$p = \left(\frac{2\pi}{L}\right)n$$

So

$$dp \quad \frac{2\pi}{L}dn = \frac{2\pi}{L} \quad (\text{as dn} = 1)$$

And

$$\tilde{f}(p) \equiv \sqrt{\frac{L}{2\pi}}g_n = \frac{1}{\sqrt{2\pi}}\int_{-\frac{L}{2}}^{\frac{L}{2}} \exp(-ipx)\, f(x)dx$$

So from the highlighted equation in the previous page, multiplying by $1 = \frac{dp}{dp}$

$$f(x) = \frac{dp}{dp}\frac{1}{\sqrt{\text{L}}}\sum_{n=-\infty}^{\infty} g_n \exp\left(i\frac{2\pi n}{L}x\right) = \frac{\sqrt{2\pi}}{L\, dp}\sum_{p=-\infty}^{\infty} \tilde{f}(p)\exp(ipx)dp$$

i.e.., converting the sum*dp to an integral,

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{f}(p) \exp(ipx) dp$$

These blue-highlighted equations are exactly the Fourier transformation; the only difference is that to get to a Fourier transform, we need to take the limit L→ ∞ so the integral over x extends to +-infintiy and the spacing dp becomes infinitesimaly small, i.e., p becomes a contnous variable.

END OF FOURIER TRANSFORM DERIVATION

Fourier transforms properties: derivatives and solving linear equations involving derivatives.

The Fourier transform of a derivative is the same as acting on the original Fourier transform by ip .

Proof: (assuming that f vanishes at infinity and doing intergration by parts

$$\widetilde{f'}(p) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \exp(-ipx)\frac{df}{dx}dx = -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \frac{d}{dx}[\exp(-ipx)]\, f dx$$

$$= \frac{ip}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \exp(-ipx)\, f dx$$

i.e.,

$$\widetilde{f'}(p) = ip\tilde{f}(p)$$

Notice that we can rewrite this symbolically as:

$$p = -i\nabla$$

Where we introduced the gradient operator, which in 1-D is just the derivative, while in higher dimensions it will be $\nabla = (\frac{d}{dx}, \frac{d}{dy}, \dots)$. That's the quantum mechanical expression we all know – the reason for this was the realization that particles are wavefunctions, so when you apply the "derivative operator" you're just multiplying by momentum (well, it is slightly more complicated, but let's not get too much to QM).

This is very useful when we come to calculate numerically complicated actions by an operator with derivatives.

For example, often we need the operator

$$\frac{1}{\nabla^2 + \mu} f(x)$$

i.e., we need to solve for an unknown function $u(x)$ that fulfills

$$(\nabla^2 + \mu)u(x) = f$$

Then to solve, we just need to convert this to Fourier space, and then we'll get:

$$(-p^2 + \mu)\,\tilde{u}(p) = \tilde{f}(p), \quad \text{i.e.,}$$

$$\tilde{u}(p) = \frac{1}{-p^2 + \mu}\tilde{f}(p)$$

So we need to Fourier-transform from f(x)→ $\tilde{f}(p)$, determine $\tilde{u}(p)$ and then Fourier transform back to u(x). Later we'll describe how this is done numerically, but suffice to say that Fourier transforms are very efficient numerically, so this makes this approach important numerically.

Convolutions:

Another important property is **that convolutions in x space become products in p space (and vice versa).** I.e., a convolution of two function, w(x) and f(x), is defined as

$$v(x) = \int w(x - x')f(x')dx'$$

We claim:

$$\tilde{v}(p) = \sqrt{2\pi}\ \tilde{w}(p)\ \tilde{f}(p)$$

Proof:

$$\tilde{v}(p) = \frac{1}{\sqrt{2\pi}} \iint \exp(-ipx)\, w(x - x')f(x')dx'\, dx =$$

$$= \frac{1}{\sqrt{2\pi}} \iint \exp(-ip(x' + y)\, w(y)f(x')dx'\, dy \quad (\text{where } y \equiv x - x')$$

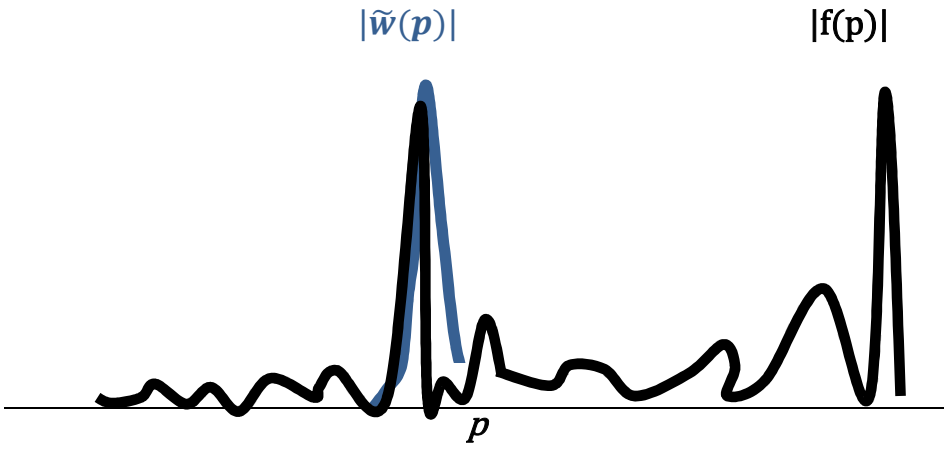
$$= \sqrt{2\pi}\frac{1}{\sqrt{2\pi}}\left(\int \exp(-ipy)\, w(y)dy\right)\frac{1}{\sqrt{2\pi}}\left(\int \exp(-ipx')\, f(x')dx'\right)$$

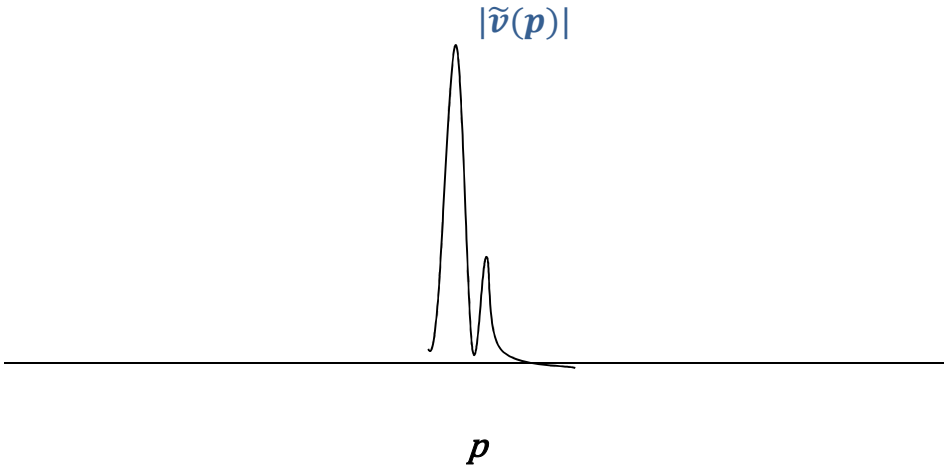
$$= \sqrt{2\pi}\ \tilde{w}(p)\ \tilde{f}(p) \quad Q.E.D.$$

This gives a practical way to evaluate convolutions, since as mentioned and we'll see later **it is very efficient numerically to do Fourier transforms (see the discussion on FFT in a few pages)– so in practice, on the computer, often convolutions are done by Fourier transforming each one of the convoluted functions and then Fourier transforming the product back from x to p (or back from w to t).**

In the next section we'll actually use a non-FFT approach to calculate convolutions for a special class of functions, i.e., Lorentzians filters; later we'll get to general filters.

See the next two figures that show action of a filter to "filter and clean" spectra.




$|\tilde{v}(p)|$
p

Filters and convolutions:

Let's say we have a physical device, e.g., a radio, that is tuned to extract specific frequencies. An example is that we get a signal f(t) and build, through electrical connections, a circuit made of capacitors, inductors, resistors, etc., that yields, an output signal, g(t), that fulfills

$$\frac{d^2g}{dt^2} = -\alpha\frac{dg}{dt} - {\omega_0}^2 g + f(t)$$

(Note that we'll use a frequency variable which is ω, not p, since our fundamental variable is time and the notation for thr Fourier transform variable assorted with t is ω, not p; but everything applies also to using x instead of t, and then labeling p instead of frequency).

For example, g could be obtained from electrical equivalent to a spring, and we can change ω_0 at will (use a "tuner"). Then we know that

$$-\omega^2\tilde{g}(\omega) = -i\alpha\omega\tilde{g}(\omega) - {\omega_0}^2\tilde{g}(\omega) + \tilde{f}(\omega)$$

i.e.,

$$\tilde{g}(\omega) = \frac{1}{\omega_0^2 - \omega^2 + i\alpha}\tilde{f}(\omega).$$

It is clear that g will be peaked near $\omega{\sim}\omega_0$.

Further, we can even find out the explicit relation between g and f, i.e.,

$$g(t) = \int u(t-t')f(t')dt'$$

where from the yellow relations from above the Filter function, u, has as its Fourier transform,

$$\tilde{u}(\omega) = -\frac{1}{\sqrt{2\pi}}\frac{1}{\omega^2 - i\alpha\omega - \omega_0^2}$$

This is called a **Lorentzians filter;** it falls off quite slowly with frequency.

Let's calculate what will be u(t-t'), i.e., the Fourier transform of $\tilde{u}(\omega)$.

(Side note: One thing we know is that u(t-t') should vanish if t<t' (the output should only depend on past and present signals, no future ones); let's see if that's borne out.)

We know that the denominator can be written as

$$\tilde{u}(\omega) = -\frac{1}{\sqrt{2\pi}}\frac{1}{(\omega - \nu_+)(\omega - \nu_-)}$$

Where a simple calculation of where the denominator vanishes shows that the poles (essentially values of ω where the function explodes, i.e., in this case the denominator vanishes) are at:

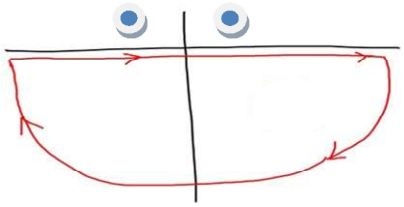
$$\nu_\pm = \frac{i\alpha}{2} \pm \bar{\omega}_0$$

Where

$$\bar{\omega}_0 = \sqrt{\omega_0^2 - \frac{\alpha^2}{4}}$$

Therefore (presuming the square root is real, i.e., α is smaller than $2\omega_0$, i.e., damping is not huge) the Fourier transform will be easy to calculate, using a half-circle contour.

If t<0 the contour should be taken in the lower half (see picture with the poles denoted by blue), since at the lower half exp($i\omega t$) will decay rapidly if t<0 and im(ω) < 0.

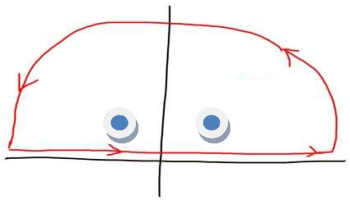


So:

$$u(t) = 0 \text{ for } t < 0$$

as is physically reasonable!

For t>0, take the upper half, as exp(iwt) decays then in the upper half of the complex plane:



Then:

$$u(t) \text{ (for } t > 0) = \frac{1}{\sqrt{2\pi}} \int \tilde{u}(\omega) exp(i\omega t) = -\frac{1}{2\pi} \oint \frac{exp(i\omega t) d\omega}{(\omega - \nu_+)(\omega - \nu_-)}$$

$$= -\frac{2\pi i}{2\pi} \sum \text{residues of } \left(\frac{exp(i\omega t)}{(\omega - \nu_+)(\omega - \nu_-)} \right) \text{ at poles}$$

There are two residues, associated with the two poles at ν_+ and ν_-. For the pole at ν_+, the residues is ther "other part", i.e., the part which is not $1/(\omega - \nu_+)$:

$$\text{residue at } (\omega = \nu_+) = \frac{exp(i\omega t)}{(\omega - \nu_-)} = \frac{exp(i\nu_+ t)}{(\nu_+ - \nu_-)} = \frac{\exp\left(-\frac{\alpha t}{2}\right)\exp(i\bar{\omega}_0 t)}{2\bar{\omega}_0}$$

For the other residue, we'll get the same result but with a minus sign (the denominator a is above now $\frac{1}{\nu_- - \nu_+} = -\frac{1}{2\bar{\omega}_0}$) and the sign of i is reversed, so when adding we get

$$u(t) = -i\left(\frac{\exp\left(-\frac{\alpha t}{2}\right)\exp(i\bar{\omega}_0 t)}{2\bar{\omega}_0} - \frac{\exp\left(-\frac{\alpha t}{2}\right)\exp(-i\bar{\omega}_0 t)}{2\bar{\omega}_0}\right)$$

i.e.,

$$u(t) = \exp\left(-\frac{\alpha t}{2}\right)\frac{\sin(\bar{\omega}_0 t)}{\bar{\omega}_0}$$

So to conclude

$$g(t) = \int u(t-t')f(t')dt' = \int \exp\left(-\frac{\alpha(t-t')}{2}\right)\frac{\sin(\bar{\omega}_0(t-t'))}{\bar{\omega}_0}f(t')dt'$$

Notice that the "convolution" or "filter" function oscillates at a frequency close to the desired frequency, and is also damped as a function of time.

Filters and Green's Functions:

The discussion on filters is important in two contexts:

1) Actual filters used in things like radios. Note that in practice a combination of several such filters will be used, leading to filters with "sharper" frequency response, i.e., u(w) will be more sharply peaked around w_0, and there will be little amplitude for u(w) when w is far from w_0; this can be done quite easily by adding two or more filters, but we won't get into that.

2) Without realizing it we started talking about **Green's functions.** For a general operator L, these are defined as:

$$LG(x,x') = \delta(x-x')$$

where L operates on the x coordinate. This sounds very difficult, but let's see how it worked out in our example. Here, we replace x by t and

$$L = \frac{d^2}{dt^2} + \alpha\frac{dg}{dt} + {\omega_0}^2$$

So

$$LG(t,t) = \delta(t-t')$$

becomes

$$\left(\frac{d^2}{dt^2} + \alpha\frac{dg}{dt} + {\omega_0}^2\right)G(t,t') = \delta(t-t')$$

We notice that in this case G depends only on the difference, i.e., we can write

$$G(t,t') = u(t-t')$$

$$\left(\frac{d^2}{dt^2} + \alpha\frac{dg}{dt} + {\omega_0}^2\right)u(t-t') = \delta(t-t')$$

This is just the u that we had before!

The importance of G is that once we know it we can solve any equation of the form

$$L\psi = f$$

as (switching back to using x as our independent variable):

$$\psi(x) = \int G(x,x')f(x')dx'$$

Proof: consider $L\psi = L\int G(x,x')f(x')dx'$. Since L operates on x we can insert it into the integral over x', resulting in:

$$L\psi = \int LG(x,x')f(x')dx' = \int \delta(x-x')f(x')dx' = f(x) \quad \text{Q. E. D.}$$

Green's functions are really important in Q.M., and you will learn about them in Q.M. class and beyond.

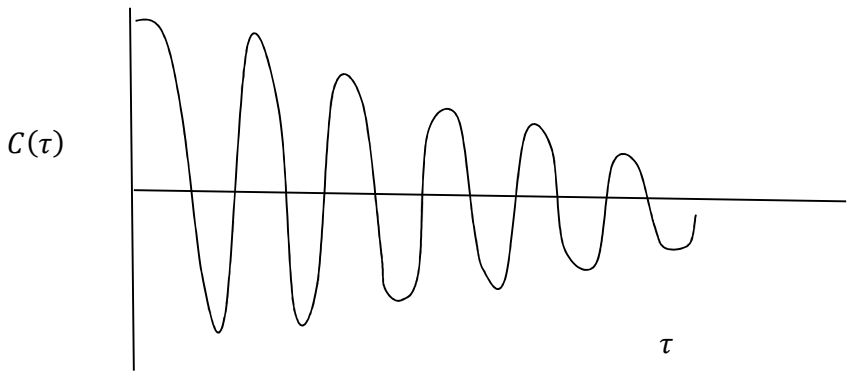
Correlation functions

Consider the correlation function of a signal, defined as:

$$C(\tau) \equiv \int \psi^*(t)\psi(t+\tau)dt$$

For example, ψ could mean the current going through a device, or the electric field coming from a soruce (light, star, laser, or anything), or any other signal in time.

If $\tau = 0$, then $C(0)$ is positive quantiy. If $\tau \gg 0$ (what do we mean by >>0 depends on the problem) then $\psi(t)$ will be independent of $\psi(t+\tau)$ so the value of the correlation function will go to zero (at least if the average of ψ vanishes over time, let's not get into that). See figure for example:



In this picture, the oscillations frequency, which we'll label T, depends on the frequency of the signal (if it has a well-defined frequency), and here the damping time will be $\tau \sim 4T$; for coherent laser signals the damping time could be millions of oscillation periods or more, while for broad-wavelength sources, such as the sun, C will damp very fast.

Since the correlation function looks like a convolution, we will get a relation to Fourier transform, using a similar methodology to earlier:

$$\tilde{C}(\omega) \equiv \frac{1}{\sqrt{2\pi}} \int exp(i\omega\tau)\, \psi^*(t)\psi(t+\tau)dtd\tau$$

$$\frac{1}{\sqrt{2\pi}} \int exp(i\omega(y-t))\, \psi^*(t)\psi(y)dtdy \qquad (where\ y \equiv t+\tau)$$

$$\frac{1}{\sqrt{2\pi}} \int exp(-i\omega t)\, \psi^*(t)dt \int exp(i\omega y)\, \psi(y)dy$$

$$= \sqrt{2\pi}\psi^*(\omega)\psi(\omega)$$

i.e.,

$$\tilde{C}(\omega) = \sqrt{2\pi}|\psi(\omega)|^2$$

And using the formula we all memorized by now:

$$C(\tau) = \frac{1}{\sqrt{2\pi}} \int C(\omega)\, exp(i\omega\tau)\, d\omega$$

We can write the correlation function in time as:

$$C(\tau) = \int |\psi(\omega)|^2\, ex \;\; (i\omega\tau)\, d\omega$$

The correlation function is also extremely important in all branches of chemistry and physics, as it also relates to influence, i.e., if I **make a change of something** (e.g., excite a dipole, apply electric field, inject an electron, rotate) **at time t, how long (how much time, τ) till the effects of this change dissipate**.

Perceval's theorem:

Set $\tau = 0$ in the last highlighted eq. and the one in the prev. page, to get Parceval's theorem:

$$\int |\psi(\omega)|^2 d\omega \; = \int |\psi(t)|^2 dt$$

(notice that the same theorem is valid if we talk about p-x rather than w-t)

This theorem is obvious to us if we think of $\psi(t)$ as a representation of an abstract vector, i.e.,

$$\psi(t) = \langle t|\psi\rangle$$

$$\psi(\omega) = \langle \omega|\psi\rangle$$

I.e., the LHS and the RHS of the highlighted eq. above are simply $\langle\psi|\psi\rangle$!

Perceval's theorem tells you that the power of the signal can be measured either in frequency or in time. It also tells you that the components of signal at any frequency contribute each a positive contribution to the power; for example, if our signal has a lot of "noise" at frequencies other than the desired one, they will all contribute to the signal, and in the ideal case should be filtered out.

Fast Fourier transforms

Finally we come to the reason Fourier transform are so popular and important these days – the fast Fourier transform algorithm; by its name, it is a method to carry Fourier transforms very efficiently for large signals.

But first let's discuss in detail:

(i) Numerical Fourier transforms

Since we deal with numerical Fourier transforms, they have a finite number of points; i.e., we really deal with Fourier sums; further, even in Fourier sums we have integrals over time (and sums over frequency), while here we'll have sums over both times and frequencies. So we have a signal made of "N" points, which for convenience will be labeled now from 0 to N-1 (and not from 1 to N)

$$f_0, \dots, f_{N-1}$$

where, similarly to earlier

$$f_0 = f(t_{min})$$

$$f_1 = f(t_{min} + \Delta t)$$

…

$$f_j = f(t_{min} + j\Delta t)$$

…

$$f_{N-1} = f(t_{min} + (N-1)\Delta t)$$

Here, Δt is not necessarily "small".

We want to numerically obtain the Fourier transform at N frequencies, which we defined, analogous to the usual Fourier summation, as:

$$\tilde{f}_k = \sum_{j=0,..,N-1} exp\left(i\omega_k t_j\right) f_j \, \Delta t$$

where we'll obtain them at

$$\omega_k = k \;\; d\omega$$

Where

$$d\omega = \frac{2\pi}{L}$$

Where

$$L = N\Delta t$$

Is the length of the signal (to be precise, it is the**(length of the signal)+Δt**, since it is associated with the periodicity; e.g.. if our signal has 32 points (from 0 to 31), we are assuming implicitly that there is a 33rd, 34th , etc. points, and the 33rd point is equivalent to the 1st, so L is the distance in time between them, i.e., $32\Delta t$)

Also, there are N independent values of k, since we can form N independent combinations from summing over the N values of f_j; we'll label these values

$$k = 0, \dots, N-1$$

(this choice of k's is only for the derivations' convenience, the true value of k's really are shifted and are

$$k = -\frac{N}{2}, \dots, \frac{N-1}{2}$$

i.e., k should really be viewed as having positive and negative values; but that's a manageable complication, and for the derivation's' sake we'll remain with k=0,…,N-1.

It is straightforward to see that we can assume that t_{min}=0 (the necessary corrections to this assumption are numerically cheap); this will result then in

$$\omega_k t_j = \frac{2\pi kj}{N} \quad (assuming\ t_{min} = 0)$$

Further, we'll remove the multiplication of the sum by Δt in the summation above, resulting in:

$$\tilde{f}_k = \sum_{j=0,..,N-1} exp\left(-i\frac{2\pi kj}{N}\right) f_j \quad (we\ omit\ the\ multiplication\ by\ \Delta t)$$

A piece of good news is that even though we have a discrete summation, the formula for the inverse FFT remains exactly the same (with a modified pre-factor):

$$f_j = \frac{1}{N} \sum_{k=0,\dots,N-1} exp\left(i\frac{2\pi kj}{N}\right) \tilde{f}_k$$

Sampling theorem and number of time points: one may ask how many points are needed to get a "good signal".

It is easy to see that if our discrete signal is due to an underlying continuous signal in time, with two closely spaced frequencies, e.g.,

$$f(t) = a_1\, exp(i\omega_1 t) + a_2\, exp(i\omega_2 t) + ot \qquad frequencies$$

Then in order to "differentiate these frequencies, i.e., get a picture of a Fourier transform with two separate picks, one at w_1 and the other at w_2, we need to have

$$|\boldsymbol{\omega_2 - \omega_1}| > \boldsymbol{d\omega}\left(= \frac{2\pi}{N\Delta t}\right)$$

i.e. we need a minimum number of time points:

$$N > \frac{2\pi}{\Delta t\,|\omega_2 - \omega_1|}$$

Further, we need to ensure that the time step is not too large so we don't confuse an oscillating signal with a non-oscillating one (see figure):

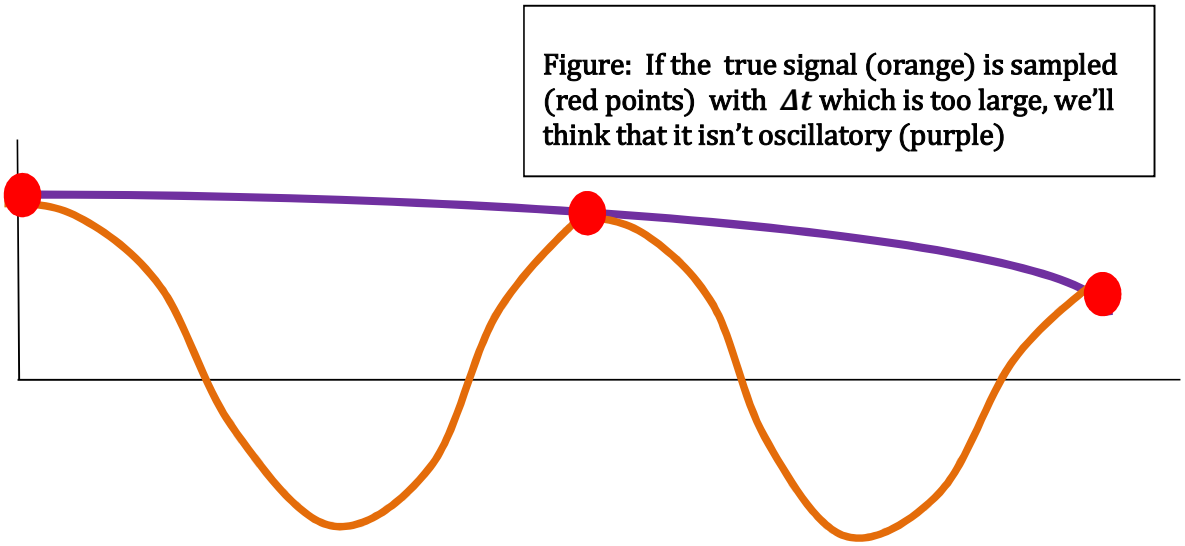


Figure: If the true signal (orange) is sampled (red points) with Δt which is too large, we'll think that it isn't oscillatory (purple)

This is equivalent to requiring that:

$$\Delta t > \frac{2\pi}{range\ of\ contributing\ frequencies\ in\ signal}$$

These two requirements (known as sampling theorem Nyquist criteria, and uncertainty relation, depending on your field**), can however be circumvented with modern algorithms** (one is Filter-Diagonalization, developed in my group).

Now to the ice on the pudding:

Fast Fourier transforms:

Ostensibly a numerical Fourier transform can be an expensive evaluation, i.e., looks like it costs N^2 operations

(i.e., for each j do a summation over N values of "k" -- – totaling N^2 points.)

However, these is a saving feature: $exp\left(i\frac{2\pi}{N}kj\right) = W^{kj}$ $(where\ W \equiv exp\left(i\frac{2\pi}{N}\right))$ and i**t takes on only N different values** (rather than N^2 values):

$$W^0 = 1, W = exp\left(-i\frac{2\pi}{N}\right),\ W^2 = exp\left(-i\frac{2\pi}{N}\right)^2, \dots, W^{N-1},$$
$$since\ W^N = exp\left(-i\frac{2\pi}{N}\right)^N = 1 = W^0.$$

This enables the FFT algorithm, discovered independently many times, **which costs only N $\log_2 N$ operations, rather than N^2.**

The trick is to realize that we can write the FFT of 2N points from two FFTs of N points:

I.e., say that we have 2N points, and that we have:

- separately done the Fourier transform of the N even members $f^e = (f_0, f_2, f_4, \dots f_{2N-2})$
- and the N odd members, $f^o = (f_1, f_3, f_5, \dots, f_{2N-1})$,

i.e.., that we have

$$(\tilde{f}^e)_k = \sum_{j=0,\dots,N-1} W^{k\cdot j} f_{2j} \quad k = 0, \dots, N-1$$

$$(\tilde{f}^o)_k = \sum_{j=0,\dots,N-1} W^{k\cdot j} f_{(2j+1)} \quad k = 0, \dots, N-1$$

Then, to get the FFT of the full 2N vector $f = (f_1, f_2, \dots f_{2N})$ we first define a factor similar to W but for 2N:

$$U = exp\left(-i\frac{2\pi}{2N}\right)$$

(so $U^2 = W$).

Then, write the Full FFT of the 2N points as:

$$\tilde{f}_k = \sum_{j=0,\dots,2N-1} U^{k\cdot j} f_j \qquad k = 0, \dots, 2N-1$$

And now divide the sum to even terms (2j) and odd (2j-1):

$$= \sum_{j=0,\dots,N-1} U^{k\cdot 2j} f_{2j} + \sum_{j=0,\dots,N-1} U^{k\cdot(2j+1)} f_{2j+1} \qquad k = 0,\dots,2N-1$$

But in the first sum, $U^{k\cdot 2j} = (U^2)^{kj} = W^{kj}$ and similarly in the 2nd sum $U^{k\cdot(2j+1)} = UW^{kj}$ so we get

$$\tilde{f}_k = \sum_{j=0,\dots,N-1} W^{kj} f_{2j} + U^k \sum_{j=0,\dots,N-1} W^{kj} f_{2j+1} \qquad k = 0,\dots,2N-1$$

Now the first term looks just like $(\tilde{f}^e)_k$ and the last is $(\tilde{f}^o)_k$ so we get:

$$\tilde{f}_k = (\tilde{f}^e)_k + U^k (\tilde{f}^o)_k$$

The only thing is that k is defined now from 0 to 2N-1, while we said that we did a Fourier transform of the N points in the even and odd parts separately, so we had it from 0,...,N-1 only;
But that's OK since you can prove that both $(\tilde{f}^e)_k$ and$(\tilde{f}^o)_k$ are peioridc with period N, i.e., if we know N values we get automatically their next N values.

This algorithm can be applied recursively:

E.g.,

To get, say, the Fourier transform of N=64 points we need to

(i) Get TWO Fourier transform of N=32; and
Do the highlighted sum above – cost of N=64 complex multiplications (and N additions, which are cheaper).

Now to do TWO Fourier transform with N=32 above we need

(ii) FOUR Fourier transforms with N=16, and then do TWO different summations as highlighted above, each for 32 points, i.e., 64 complex multiplications.

So to do this we need:

(iii) Eight Fourier transforms (F.T.) with N=8, and 64 complex multiplications, i.e.,
(iv) 16 F.T. with N=4, and another 64 complex operations, i.e.,
(v) 32 F.T. with N=2, and another 64 complex operations,;
(vi) and the 32 F.T. with N=2 cost themselves 64 operations. (it is easy to do a FT with 2 points!)

So overall we had $64+64+64+64+64+64 = 6*64$ operations; obviously for any N which is a power of 2 ($N=2^M$) the cost is $N \cdot M$, i.e., $\boldsymbol{N\, log_2\, N}$ operations as promised.

Example: for a 3D imaging, where the same cost applies (with N=total number of points$=N_x \cdot N_y \cdot N_z$), it is quite feasible to have $N=(1000)^3 = 10^9$ points. Then, using an FFT algorithm we need only

$$N\, log_2\, N \sim 10^9 \cdot 30 = 3 \cdot 10^{10}$$

i.e., 30 billion complex multiplications and additions, which will take about a quarter-minute to half a minute on modern 3GHZ workstations; if we were to try to do this without the FFT algorithm, it would have taken $N^2=10^9 \cdot 10^9 = 10^{18}$ operations, which will take, with the same computer, 50 years…

For reference on this chapter so far (as well as on almost all things numerical), use the book : **Numerical Recipes.**

NUMERICAL DERIVATIVES WITH SPECTRAL METHODS:

One final comment – the FFT enables, in addition to studying spectra and doing convolution, to have a very accurate numerical derivative. I.e., given N values of a function

$$f(x_j) \ , where\ x_j = x_{min} + j\ dx,$$

the traditional way to get a derivative would have been to calculate an approximation to the derivative from, say,

$$\frac{df}{dx}(x_j) \cong \frac{f(x_{j+1}) - f(x_j)}{2\ dx}$$

or another similar formula (we use blue since this formula will be replaced). However, with FFT we can imagine that the function is rigorously a sum of N exponentials, i.e., imagine that

$$f(x) = \sum_k exp(ip_k x)\ \tilde{f}_k$$

(we won't worry about overall normalization constant), then we will

(i) Get the $\tilde{f}_k$ by a numerical FFT;

(ii) **Differentiate analytically the exponentials** $exp(\mathrm{ip_k x})$ **to get:**

$$\frac{df}{dx}(x = x_j) = \sum_k ip_k\ exp(ip_k x_j)\ \tilde{f}_k$$

(iii) DO this highlighted summation by inverse Fourier transform.

For reference, see D. Kosloff and R. Kosloff, A Fourier Method Solution for the Time Dependent Schrödinger equation as a Tool in Molecular Dynamics, J. Comp. Phys., 52, 35-53 (1983) and Ronnie Kosloff, Time Dependent Methods in Molecular Dynamics, J. Phys. Chem., 92, 2087-2100 (1988).

END OF Fourier Transforms

END OF COURSE!