2 Sturm–Liouville Theory

So far, we've examined the Fourier decomposition of functions defined on some interval (often scaled to be from $-\pi$ to π). We viewed this expansion as an infinite dimensional analogue of expanding a finite dimensional vector into its components in an orthonormal basis. But this is just the tip of the iceberg. Recalling other games we play in linear algebra, you might well be wondering whether we couldn't have found some other basis in which to expand our functions. You might also wonder whether there shouldn't be some role for *matrices* in this story. If so, read on!

2.1 Self-adjoint matrices

We'll begin by reviewing some facts about matrices. Let V and W be finite dimensional vector spaces (defined, say, over the complex numbers) with $\dim V = n$ and $\dim W = m$. Suppose we have a linear map $M : V \to W$. By linearity, we know what M does to any vector $\mathbf{v} \in V$ if we know what it does to a complete set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of basis vectors in V. Furthermore, given a basis $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ of W we can represent the map M in terms of an $m \times n$ *matrix* $\mathbf{M}$ whose components are

$$\mathbf{M}_{ai} = (\mathbf{w}_a, M\mathbf{v}_i) \qquad \text{for } a = 1, \ldots, m \;\text{ and }\; i = 1, \ldots, n\,, \tag{2.1}$$

where $(\ ,\ )$ is the inner product in W.

We'll be particularly interested in the case $m = n$, when the matrix $\mathbf{M}$ is square and the map M takes $M : V \to W \cong V$ is an isomorphism of vector spaces. For any $n \times n$ matrix M we define it's eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ to be the roots of the characteristic equation $P(\lambda) \equiv \det(\mathbf{M} - \lambda\mathbf{I}) = 0$, where $\mathbf{I}$ is the identity matrix. This characteristic equation has degree n and the fundamental theorem of algebra assures us that we'll always be able to find n roots (generically complex, and not necessarily distinct). The *eigenvector* $\mathbf{v_i}$ of $\mathbf{M}$ that corresponds to the eigenvalue λ_i is then defined by $\mathbf{M}\mathbf{v_i} = \lambda_i\mathbf{v_i}$ (at least for non-degenerate eigenvalues).

Given a complex $n \times n$ matrix $\mathbf{M}$, its *Hermitian conjugate* $\mathbf{M}^\dagger$ is defined to be the complex conjugate of the transpose matrix, $\mathbf{M}^\dagger \equiv (\mathbf{M}^{\mathrm{T}})^*$, where the complex conjugation acts on each entry of $\mathbf{M}^{\mathrm{T}}$. A matrix is said to be *Hermitian* or *self-adjoint* if $\mathbf{M}^\dagger = \mathbf{M}$. There's a neater way to define this: since for two vectors we have $(\mathbf{u}, \mathbf{v}) = \mathbf{u}^\dagger \cdot \mathbf{v}$, we see that a matrix $\mathbf{B}$ is the adjoint of a matrix $\mathbf{A}$ iff

$$(\mathbf{Bu}, \mathbf{v}) = (\mathbf{u}, \mathbf{Av}) \tag{2.2}$$

because the vector $(\mathbf{Bu})^\dagger = \mathbf{u}^\dagger\mathbf{B}^\dagger$. The advantages of this definition are that *i*) it does not require that we pick any particular components in which to write the matrix and *ii*) it applies whenever we have a definition of an inner product $(\ ,\ )$.

Self-adjoint matrices have a number of very important properties. Firstly, since

$$\lambda_i(\mathbf{v_i}, \mathbf{v_i}) = (\mathbf{v_i}, \mathbf{Mv_i}) = (\mathbf{Mv_i}, \mathbf{v_i}) = \lambda_i^*(\mathbf{v}_i, \mathbf{v}_i) \tag{2.3}$$

the eigenvalues of a self-adjoint matrix are always real. Secondly, we have

$$\lambda_i(\mathbf{v_j}, \mathbf{v_i}) = (\mathbf{v_j}, \mathbf{M v_i}) = (\mathbf{M v_j}, \mathbf{v_i}) = \lambda_j(\mathbf{v}_j, \mathbf{v}_i) \tag{2.4}$$

or in other words

$$(\lambda_i - \lambda_j)(\mathbf{v}_j, \mathbf{v}_i) = 0 \tag{2.5}$$

so that eigenvectors corresponding to distinct eigenvalues are orthogonal *wrt* the inner product (,). After rescaling the eigenvectors to have unit norm, we can express any $\mathbf{v} \in V$ as a linear combination of the orthonormal set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of eigenvectors of some self-adjoint $\mathbf{M}$. If $\mathbf{M}$ has degenerate eigenvalues (*i.e.* two or more distinct vectors have the same eigenvalue) then the set of vectors sharing an eigenvalue form a vector subspace of V and we simply choose an orthonormal basis for each of these subspaces. In any case, the important point here is that self-adjoint matrices provide a *natural* way to pick a basis on our vector space.

A self-adjoint matrix $\mathbf{M}$ is non-singular ($\det \mathbf{M} \neq 0$ so that $\mathbf{M}^{-1}$ exists) if and only if all its eigenvalues are non-zero. In this case, we can solve the linear equation $\mathbf{Mu} = \mathbf{f}$ for a fixed vector $\mathbf{f}$ and unknown $\mathbf{u}$. Formally, the solution is $\mathbf{u} = \mathbf{M}^{-1}\mathbf{f}$, but a practical way to determine $\mathbf{u}$ proceeds as follows. Suppose $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthonormal basis of eigenvectors of $\mathbf{M}$. Then we can write

$$\mathbf{f} = \sum_{i=1}^{n} f_i \mathbf{v_i} \qquad \text{and} \qquad \mathbf{u} = \sum_{i=1}^{n} u_i \mathbf{v_i}$$

where $f_i = (\mathbf{v}_i, \mathbf{f})$ *etc.* as before. We will know the vector $\mathbf{u}$ if we can find all its coefficients u_i in the $\{\mathbf{v}_i\}$ basis. But by linearity

$$\mathbf{Mu} = \sum_{i=1}^{n} u_i \, \mathbf{M v}_i = \sum_{i=1}^{n} u_i \lambda_i \mathbf{v}_i \;=\; \mathbf{f} = \sum_{i=1}^{n} f_i \mathbf{v}_i \;, \tag{2.6}$$

and taking the inner product of this equation with $\mathbf{v}_j$ gives

$$\sum_{i=1}^{n} u_i \lambda_i \, (\mathbf{v}_j, \mathbf{v}_i) = u_j \lambda_j \;=\; \sum_{i=1}^{n} f_i (\mathbf{v}_j, \mathbf{v}_i) = f_j \tag{2.7}$$

using the orthonormality of the basis. Provided $\lambda_j \neq 0$ we deduce $u_j = f_j/\lambda_j$ so that

$$\mathbf{u} = \sum_{i=1}^{n} \frac{f_i}{\lambda_i} \mathbf{v}_i \;. \tag{2.8}$$

If $\mathbf{M}$ is singular then either $\mathbf{Mu} = \mathbf{f}$ has no solution or else has a non-unique solution (which it is depends on the choice of $\mathbf{f}$).

2.2 Differential operators

In the previous chapter we learned to think of functions as infinite dimensional vectors. We'd now like to think of the analogue of matrices. Sturm and Liouville realised that these

could be thought of as *linear differential operators* $\mathcal{L}$. This is just a linear combination of derivatives with coefficients that can also be functions of x, *i.e.* $\mathcal{L}$ is a linear differential operator of order p if

$$\mathcal{L} = A_p(x)\frac{d^p}{dx^p} + A_{p-1}(x)\frac{d^{p-1}}{dx^{p-1}} + \cdots + A_1(x)\frac{d}{dx} + A_0(x) \ .$$

When it acts on a (sufficiently smooth) function $y(x)$ it gives us back some other function $\mathcal{L}y(x)$ obtained by differentiating $y(x)$ in the obvious way. This is a linear map between spaces of functions because for two (p-times differentiable) functions $y_{1,2}(x)$ and constants $c_{1,2}$ we have $\mathcal{L}(c_1y_1 + c_2y_2) = c_1\mathcal{L}y_1 + c_2\mathcal{L}y_2$. The analogue of the matrix equation $\mathbf{Mu} = \mathbf{f}$ is then the differential equation $\mathcal{L}y(x) = f(x)$ where we assume that both the coefficient functions $A_p(x), \ldots, A_0(x)$ in $\mathcal{L}$ and the function $f(x)$ are known, and that we wish to find the unknown function $y(x)$.

For most of our applications in mathematical physics, we'll be interested in *second order*[9] linear differential operators[10]

$$\mathcal{L} = P(x)\frac{d^2}{dx^2} + R(x)\frac{d}{dx} - Q(x) \ . \tag{2.9}$$

Recall that for any such operator, the *homogeneous equation* $\mathcal{L}y(x) = 0$ has precisely two non-trivial linearly independent solutions, say $y = y_1(x)$ and $y = y_2(x)$ and the general solution $y(x) = c_1y_1(x) + c_2y_2(x)$ with $c_i \in \mathbb{C}$ is known as the *complementary function*. When dealing with the *inhomogeneous equation* $\mathcal{L}y = f$, we seek any single solution $y(x) = y_p(x)$, and the general solution is then a linear combination

$$y(x) = c_py_p(x) + c_1y_1(x) + c_2y_2(x)$$

of the particular and complementary solutions. In many physical applications, the function f represents a driving force for a system that obeys $\mathcal{L}y(x) = 0$ if left undisturbed.

In the cases (I assume) you've seen up to now, actually *finding* the particular solution required a good deal of either luck or inspired guesswork – you 'noticed' that if you differentiated such-and-such a function you'd get something that looked pretty close to the solution you're after, and perhaps you could then refine this guess to find an exact solution. Sturm–Liouville theory provides a more systematic approach, analogous to solving the matrix equation $\mathbf{Mu} = \mathbf{f}$ above.

2.3 Self-adjoint differential operators

The 2$^{\text{nd}}$-order differential operators considered by Sturm & Liouville take the form

$$\mathcal{L}y \equiv \frac{d}{dx}\left(p(x)\frac{dy}{dx}\right) - q(x)y \, , \tag{2.10}$$

where $p(x)$ is *real* (and once differentiable) and $q(x)$ is real and continuous. This may look to be a tremendous specialization of the general form (2.9), with $R(x)$ restricted to be

[9] It's a beautiful question to ask 'why only second order'? Particularly in quantum theory.

[10] The sign in front of $Q(x)$ is just a convention.

$P'(x)$, but actually this isn't the case. Provided $P(x) \neq 0$, starting from (2.9) we divide through by $P(x)$ to obtain

$$\frac{d^2}{dx^2} + \frac{R(x)}{P(x)}\frac{d}{dx} - \frac{Q(x)}{P(x)} = \mathrm{e}^{-\int_0^x R(t)/P(t)\,\mathrm{d}t}\,\frac{d}{dx}\left(\mathrm{e}^{\int_0^x R(t)/P(t)\,\mathrm{d}t}\,\frac{d}{dx}\right) - \frac{Q(x)}{P(x)} \tag{2.11}$$

Thus setting $p(x)$ to be the integrating factor $p(x) = \exp\left(\int_0^x R(t)/P(t)\,\mathrm{d}t\right)$ and likewise setting $q(x) = Q(x)p(x)/P(x)$, we see that the forms (2.10) and (2.9) are equivalent. However, for most purposes (2.10) will be more convenient.

The beautiful feature of these Sturm–Liouville operators is that they are self-adjoint with respect the inner product

$$(f, g) = \int_a^b f(x)^* g(x)\,\mathrm{d}x\,, \tag{2.12}$$

provided the functions on which they act obey appropriate boundary conditions. To see this, we simply integrate by parts twice:

$$\begin{aligned}
(\mathcal{L}f, g) &= \int_a^b \left[\frac{d}{dx}\left(p(x)\frac{df^*}{dx}\right) - q(x)f^*(x)\right] g(x)\,\mathrm{d}x \\
&= \left[p\,\frac{df^*}{dx}\,g\right]_a^b - \int_a^b p(x)\frac{df^*}{dx}\frac{dg}{dx} - q(x)f(x)^*g(x)\,\mathrm{d}x \\
&= \left[p\,\frac{df^*}{dx}\,g - p\,f^*\frac{dg}{dx}\right]_a^b + \int_a^b f(x)^*\left[\frac{d}{dx}\left(p(x)\frac{dg}{dx}\right) - q(x)\,g(x)\right] \\
&= \left[p(x)\left(\frac{df^*}{dx}\,g - f^*\frac{dg}{dx}\right)\right]_a^b + (f, \mathcal{L}g)
\end{aligned} \tag{2.13}$$

where in the first line we have used fact that p and q are real for a Sturm–Liouville operator. So we see that $(\mathcal{L}f, g) = (f, \mathcal{L}g)$ provided we restrict ourselves to functions which obey the boundary conditions

$$\left[p(x)\left(\frac{df^*}{dx}\,g - f^*\frac{dg}{dx}\right)\right]_a^b = 0\,. \tag{2.14}$$

Examples of such boundary conditions are to require that *all* our functions satisfy

$$\begin{aligned}
b_1 f'(a) + b_2 f(a) &= 0 \\
c_1 f'(b) + c_2 f(b) &= 0\,,
\end{aligned} \tag{2.15}$$

where $b_{1,2}$ and $c_{1,2}$ are constants, not both zero. I emphasize that we must choose the *same* constants for all our functions. These boundary conditions ensure that (2.14) vanishes at each boundary separately. If the function $p(x)$ obeys $p(a) = p(b)$ then we can likewise ask that all our functions are periodic, so that $f(a) = f(b)$ and $f'(a) = f'(b)$; this ensures that the contributions at each boundary cancel in (2.14). Finally, it may sometimes be that $p(a) = p(b) = 0$, though in this case the endpoints of the interval $[a, b]$ are singular points of the differential equation.

2.4 Eigenfunctions and weight functions

Whatever boundary conditions we choose, provided they satisfy (2.14) we have $(\mathcal{L}f, g) = (f, \mathcal{L}g)$ and $\mathcal{L}$ is self-adjoint. Just as in the finite dimensional case of self-adjoint matrices, these self-adjoint differential operators automatically inherit many useful properties.

We start by defining the notion of an *eigenfunction* of our differential operator. It's convenient to do this in a slightly surprising way. A *weight function* $w(x)$ is a real-valued, non-negative function that has at most finitely many zeros on the domain $[a, b]$. A function $y(x)$ is said to be an *eigenfunction of* $\mathcal{L}$ *with eigenvalue* λ *and weight* $w(x)$ if

$$\mathcal{L}y(x) = \lambda w(x) y(x) \tag{2.16}$$

where we note the presence of the weight function on the right hand side. In fact, given such an eigenfunction we can always find a corresponding eigenfunction $\tilde{y}(x)$ with weight function 1 by setting $\tilde{y}(x) = \sqrt{w(x)}\, y(x)$ and replacing $\mathcal{L}y$ by $\frac{1}{\sqrt{w}}\mathcal{L}\left(\frac{\tilde{y}}{\sqrt{w}}\right)$, so the weight function does not really represent anything new, but it's conventional (and will turn out to be convenient) to keep it explicit.

The weight function plays a role in the inner product. We define the *inner product with weight* w to be

$$(f, g)_w \equiv \int_a^b f(x)^* g(x)\, w(x)\, \mathrm{d}x \tag{2.17}$$

so that the measure includes a factor of w. Notice that since w is real

$$(f, g)_w = (f, wg) = (wf, g) \tag{2.18}$$

where the inner products on the *rhs* are the standard ones with measure $\mathrm{d}x$ only. This inner product is again non-degenerate in the sense that $(f, f)_w = 0$ implies $f = 0$ if f is continuous (at least in a neighbourhood of any zeros of $w(x)$). This is because

$$(f, f)_w = \int_a^b |f(x)|^2\, w(x)\, \mathrm{d}x$$

is the integral of a continuous positive function $w(x)|f(x)|^2$. By assumption w has only finitely many zeros on $[a, b]$ while f is continuous, so the integral gives zero if and only if $f(x) = 0$ identically on $[a, b]$.

The first property of Sturm–Liouville operators is that their eigenvalues are always real. The proof is exactly the same as in the finite dimensional case: if $\mathcal{L}f = \lambda w f$ then

$$\lambda\,(f, f)_w = (f, \lambda w f) = (f, \mathcal{L}f) = (\mathcal{L}f, f) = \lambda^*(f, f)_w \tag{2.19}$$

using the self-adjointness of $\mathcal{L}$ and the fact that the inner product $(\ ,\ )$ is *anti*-linear in its first entry. Note that if f has eigenvalue λ, then because the eigenvalues, weight w and coefficients $p(x)$ and $q(x)$ are real

$$\mathcal{L}(f^*) = (\mathcal{L}f)^* = (\lambda w f)^* = \lambda w(x)\, f(x)^* \tag{2.20}$$

so that f^* is also an eigenfunction of $\mathcal{L}$ with the same eigenvalue. Thus, taking $\mathrm{Re}\, f$ and $\mathrm{Im}\, f$ if necessary, we can always choose our eigenfunctions to be real-valued.

Just as in the finite dimensional case, eigenfunctions f_1 and f_2 with distinct eigenvalues, but the *same* weight function, are orthogonal wrt the inner product with weight w, since:

$$\lambda_i (f_j, f_i)_w = (f_j, \mathcal{L} f_i) = (\mathcal{L} f_j, f_i) = \lambda_j (f_j, f_i)_w \tag{2.21}$$

so that if $\lambda_i \neq \lambda_j$ then

$$(f_j, f_i)_w = \int_a^b f_j(x)^* f_i(x)\, w(x)\, \mathrm{d}x = 0\,. \tag{2.22}$$

Thus, exactly as in the finite dimensional case, given a self-adjoint operator $\mathcal{L}$ we can form an orthonormal set $\{Y_1(x), Y_2(x), \ldots\}$ of its eigenfunctions by setting

$$Y_n(x) = y_n(x) \Big/ \sqrt{\int_a^b |y_n|^2\, w\, \mathrm{d}x} \tag{2.23}$$

where $y_n(x)$ is the unnormalised eigenfunction. I emphasize again the presence of the weight function in these orthogonality and normalization conditions.

Finally, after making a particular choice of boundary conditions, one can also show[11] that the eigenvalues form a countably infinite sequence $\lambda_1, \lambda_2, \ldots$, with $|\lambda_n| \to \infty$ as $n \to \infty$, and that the corresponding set of orthonormal eigenfunctions $Y_1(x), Y_2(x), \ldots$ form a *complete* basis for functions on $[a, b]$ satisfying these boundary conditions. That is, any function $f(x)$ on $[a, b]$ that obeys the chosen boundary conditions may be expanded as[12]

$$f(x) = \sum_{n=1}^{\infty} f_n\, Y_n(x)\,, \qquad \text{where} \qquad f_n = \int_a^b Y_n^*(x) f(x)\, w(x)\, \mathrm{d}x = (Y_n, f)_w\,. \tag{2.24}$$

The significant feature here is that the function $f(x)$ is expanded as a *discrete* sum, just as we saw for Fourier series. This is really remarkable, because the definition of the Y_ns – that they be normalised eigenfunctions of $\mathcal{L}$ – involves no hint of discreteness. In fact, we'll see later in the course that the discreteness arises because the domain $[a, b]$ is *compact*, and because of our boundary conditions (2.14).

2.5 Some examples

Let's take a look at some simple examples of the general theory above. The simplest non-trivial case is just to take the domain $[a, b]$ to be $[-L, L]$ and impose the homogeneous boundary conditions that all our functions are periodic *i.e.* $f(-L) = f(L)$ and $f'(-L) = f'(L)$. If we also choose $p(x) = 1$ and $q(x) = 0$ then the Sturm–Liouville operator reduces to

$$\mathcal{L} = \frac{d^2}{dx^2}\,, \tag{2.25}$$

[11] But, sadly, not in this course.

[12] As warned, in this course we will no longer worry about convergence of these infinite sums, although see section 2.8.

which is easily seen to be self-adjoint when acting on functions obeying these boundary conditions. Finally, we choose the weight function to be $w(x) = 1$ identically.

The eigenfunction equation becomes

$$\mathcal{L}y(x) = -\lambda y(x) \tag{2.26}$$

where we've introduced a minus sign for convenience (just by relabelling the eigenvalues). If $\lambda < 0$ then the only solution *that obeys the periodic boundary conditions* is the trivial case $y(x) = 0$. However, if $\lambda \geq 0$ then a basis of solutions is given by

$$y_n(x) = \exp\left(\mathrm{i}\frac{n\pi x}{L}\right) \qquad \text{for} \qquad \lambda_n = \left(\frac{n\pi}{L}\right)^2 \quad \text{with} \quad n \in \mathbb{Z}\,. \tag{2.27}$$

Thus we have recovered the Fourier series of section 1.3 as a special case! Note that these eigenvalues are degenerate, with $\lambda_n = \lambda_{-n}$; as we saw before, whenever $y(x)$ is a complex–valued eigenfunction of a SL operator, then $y^*(x)$ is also an eigenfunction with the same eigenvalue. If instead of asking for the functions to be periodic, we'd asked specifically that $f(-L) = f(L) = 0$, then we would find just the sinusoidal Fourier series which has non-degenerate eigenvalues.

For a more interesting example, suppose we're interested in solving the differential equation[13]

$$\frac{1}{2}H'' - xH' = -\lambda H(x) \qquad \text{for} \qquad x \in \mathbb{R}\,, \tag{2.28}$$

subject to the condition that $H(x)$ behaves like a polynomial as $|x| \to \infty$ (so that in particular $\mathrm{e}^{-x^2/2}H(x) \to 0$ as $|x| \to \infty$). The reason for this strange looking condition will be revealed below.

Equation (2.28) is not yet in Sturm–Liouville form, so we first compute the integrating factor

$$p(x) = -\int_0^x 2t\,\mathrm{d}t = -x^2 \tag{2.29}$$

and rewrite (2.28) as

$$\frac{d}{dx}\left(\mathrm{e}^{-x^2}\frac{dH}{dx}\right) = -2\lambda\,\mathrm{e}^{-x^2}\,H(x) \tag{2.30}$$

(multiply through by $2\mathrm{e}^{-x^2}$ to recover the form (2.28)). This equation is known as *Hermite's equation* and it plays an important role in combinatorics, probability and in the quantum mechanics of a harmonic oscillator. We can now understand the condition that $H(x)$ grows at most polynomially at large $|x|$: in checking self–adjointness of the Sturm–Liouville operator on the unbounded domain $\mathbb{R}$, we do not need to consider boundary terms, but we do need to ensure the integrals

$$\int_{-\infty}^{\infty} f^*\frac{d}{dx}\left(\mathrm{e}^{-x^2}\frac{dg}{dx}\right)\,\mathrm{d}x = \int_{-\infty}^{\infty}\frac{df^*}{dx}\,\mathrm{e}^{-x^2}\,\frac{dg}{dx}\,\mathrm{d}x$$

[13] I'm cheating here by working on an unbounded domain $x \in \mathbb{R}$ rather than $x \in [a, b]$. Much of the theory holds, but notice the rather strange boundary condition we impose. This is just to ensure that the integrals over the entire real axis that arise when we check self-duality of $\mathcal{L}$ are bounded.

actually remain finite! For f and g regular along $\mathbb{R}$ this will be so as long as the integral is suppressed at large $|x|$. This is what our decay condition is designed to ensure.

I'll state without proof that equation (2.30) has non-trivial solutions that are regular for all $x \in \mathbb{R}$ iff the eigenvalue λ is a non-negative integer n, and you can check that these solutions are given by

$$H_n(x) = (-1)^n \, \mathrm{e}^{x^2} \, \frac{d^n}{dx^n} \, \mathrm{e}^{-x^2} \, . \tag{2.31}$$

Carrying out the differentiation in (2.31) we find for example, $H_0(x) = 1$, $H_1(x) = 2x$, $H_2(x) = 4x^2 - 2$ and $H_3(x) = 8x^3 - 12x$. In general $H_n(x)$ is a real polynomial of degree n, known as a *Hermite polynomial*. The Hermite polynomials are orthogonal with respect to the weight function $w(x) = \mathrm{e}^{-x^2}$ and obey the normalization condition

$$(H_m, H_n)_{\mathrm{e}^{-x^2}} = \int_{-\infty}^{\infty} H_m(x) \, H_n(x) \, \mathrm{e}^{-x^2} \, \mathrm{d}x = \delta_{m,n} \, 2^n \sqrt{\pi} \, n! \, , \tag{2.32}$$

where we note the decay condition is again crucial to ensure that this inner product remains finite.

2.6 Inhomogeneous equations and Green's functions

Finally, we return to the infinite dimensional analogue of the inhomogeneous matrix equation $\mathbf{Mu} = \mathbf{f}$ for a self-adjoint matrix $\mathbf{M}$. In the context of Sturm–Liouville differential operators, we seek to solve the inhomogeneous differential equation

$$\mathcal{L}\phi(x) = w(x)F(x) \tag{2.33}$$

where again we choose to include the weight function in the definition of the forcing term on the right hand side. By the remarks above, the functions $\phi(x)$ and $F(x)$ can be expanded in a complete set of eigenfunctions of $\mathcal{L}$. So we suppose that the set $\{Y_1(x), Y_2(x), \ldots\}$ form a complete set of such eigenfunctions with

$$\mathcal{L}Y_n(x) = \lambda_n \, w(x) \, Y_n(x) \qquad \text{and} \qquad (Y_m, Y_n)_w = \delta_{m,n} \tag{2.34}$$

and expand

$$\phi(x) = \sum_{n=1}^{\infty} \phi_n \, Y_n(x) \, , \qquad\qquad F(x) = \sum_{n=1}^{\infty} F_n \, Y_n(x) \, . \tag{2.35}$$

As in the matrix case, it is assumed that the function $F(x)$, and hence the coefficients $F_n = (Y_n, F)_w$ are known, while the coefficients ϕ_n must be found. But again, this can be done exactly in analogy with the finite dimensional case. Since $\mathcal{L}$ is a linear operator we have

$$\begin{aligned} \mathcal{L}\phi &= \sum_{n=1}^{\infty} \phi_n \, \mathcal{L}Y_n = w \sum_{n=1}^{\infty} \phi_n \lambda_n \, Y_n \\ &= wF = w \sum_{n=1}^{\infty} F_n Y_n \end{aligned} \tag{2.36}$$

and taking the inner product with Y_m gives $\phi_m \lambda_m = F_m$. Thus, provided none of the eigenvalues of $\mathcal{L}$ are zero we have found the particular solution

$$\phi_p(x) = \sum_{n=1}^{\infty} \frac{F_n}{\lambda_n} Y_n(x) \,. \tag{2.37}$$

As above, the general solution is now a sum of this particular solution and the complementary function $\phi_c(x)$ satisfying $\mathcal{L}\phi_c = 0$.

It's worth looking a little more closely at the structure of the solution (2.37). Substituting in the definition of the forcing term coefficients F_n and exchanging the sum and integral we have

$$\begin{aligned} \phi_p(x) &= \sum_{n=1}^{\infty} \frac{(Y_n, F)_w}{\lambda_n} Y_n(x) = \sum_{n=1}^{\infty} \left[\frac{1}{\lambda_n} Y_n(x) \int_a^b Y_n^*(t)\, F(t)\, w(t)\, \mathrm{d}t \right] \\ &= \int_a^b G(x,t) f(t)\, \mathrm{d}t \end{aligned} \tag{2.38}$$

where $f(t) = w(t)F(t)$ is the right hand side of the original inhomogeneous equation (2.33) and we've defined the *Green's function*

$$G(x;t) \equiv \sum_{n=1}^{\infty} \frac{Y_n(x) Y_n^*(t)}{\lambda_n} \,. \tag{2.39}$$

The Green's function is a function of two variables $(x,t) \in [a,b] \times [a,b]$. The important point about the Green's function is that it depends on the differential operator $\mathcal{L}$ both through its eigenfunctions and (more subtly) through the boundary conditions we chose to ensure $\mathcal{L}$ is self-adjoint, but it does *not* depend on the forcing function f. Thus if we know the Green's function we can use (2.38) to construct a particular solution of $\mathcal{L}y = f$ for an *arbitrary* forcing term. In this way, the Green's function provides a formal inverse to the differential operator $\mathcal{L}$ in the sense that

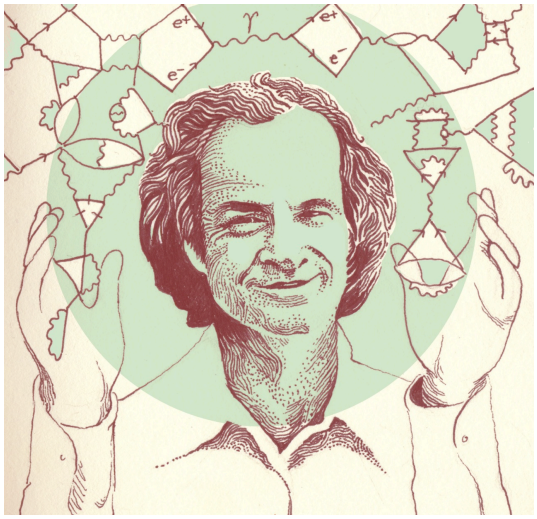
$$\text{if} \qquad \mathcal{L}y(x) = f(x) \qquad \text{then} \qquad y(x) = \int_a^b G(x,t)\, f(t)\, \mathrm{d}t$$

again in analogy with the finite dimensional case where $\mathbf{Mu} = \mathbf{f}$ implies $\mathbf{u} = \mathbf{M}^{-1}\mathbf{f}$ for a non-singular matrix. The notion of a Green's function and the associated integral operator as an inverse of $\mathcal{L}$ is very important. We'll meet it again later in many more general contexts. One of them is depicted here[14]:

2.7 Parseval's identity II

Recall that Pythagoras' theorem says that the length squared of a vector is the sum of the (mod-)squared of its components in *any* orthonormal basis. In the case of the Fourier basis, we obtained an infinite dimensional version of this in equation (1.50). We now establish

[14] I stole this picture from Hannah Wilson Illustration, as you'll quickly discover if, like me, you google Feynman diagrams.



a version of Parseval's identity in this more general context of the weighted inner product $(\ ,\ )_w$.

Let $\{Y_1(x), Y_2(x), \ldots\}$ be a complete set of functions that are orthonormal with respect to some weight function $w(x)$, so that $(Y_m, Y_n)_w = \delta_{m,n}$. Then expanding

$$f(x) = \sum_{n=1}^{\infty} f_n\, Y_n(x) \qquad \text{with} \qquad f_n = (Y_n, f)_w$$

as in (2.24) we have

$$\begin{aligned}(f,f)_w &= \int_a^b \left[\sum_{m=1}^{\infty} f_m^*\, Y_m^*(x)\right]\left[\sum_{n=1}^{\infty} f_n\, Y_n(x)\right] w(x)\,\mathrm{d}x \\ &= \sum_{n,m} f_m^*\, f_n\, (Y_m, Y_n)_w = \sum_{n=1}^{\infty} |f_n|^2\,.\end{aligned} \tag{2.40}$$

This is Parseval's identity for the case of the inner product with weight w.

2.8 Least squares approximation

In the real world, our computers have finite power and memory, and we typically don't have the resources to handle a very large number of eigenfunctions. So in practical applications, it's important to know how accurately we can represent a function by expanding it in just a limited, incomplete set of eigenfunctions. Suppose we consider the *finite* sum

$$g(x) \equiv \sum_{i=1}^{n} c_i\, Y_i(x) \tag{2.41}$$

that just includes some finite number n of the eigenfunctions, for *some* constants c_i. We ask how we should choose these constants if we want g to represent a given function $f(x)$

as closely as possible. One notion of what we mean by 'closely' is to ask that the distance between g and f should be minimized in the $(\ ,\ )_w$ norm, or in other words that

$$(f-g, f-g)_w = \int_a^b |f-g|^2 \; w(x)\,\mathrm{d}x$$

should be made as small as possible by varying the c_is. Using the definition (2.41) of $g(x)$ and the expansion (2.24) of f we have

$$\begin{aligned} \frac{\partial}{\partial c_k}(f-g, f-g)_w &= \frac{\partial}{\partial c_k}\left[(f,f) - (g,f) - (f,g) + (g,g)\right] \\ &= -\frac{\partial}{\partial c_k}\left[\sum_{i=1}^{n} c_i^* f_i + f_i^* c_i - |c_i|^2\right] \\ &= -f_k^* + c_k^* \end{aligned} \tag{2.42}$$

and likewise

$$\frac{\partial}{\partial c_k^*}(f-g, f-g)_w = -f_k + c_k\,, \tag{2.43}$$

where the evaluation of (g,g) uses Parseval's identity. These derivatives vanish iff $c_k = f_k$, and since

$$\frac{\partial^2}{\partial c_j\,\partial c_k}(f-g, f-g)_w = 0 = \frac{\partial^2}{\partial c_j^*\,\partial c_k^*}\,, \qquad \text{while} \qquad \frac{\partial^2}{\partial c_j\,\partial c_k^*}(f-g, f-g)_w = \delta_{j,k} \geq 0$$

the extremum is indeed a minimum. Therefore, if we wish to approximate a function $f(x)$ by representing it as a linear combination of just a few eigenfunctions of some Sturm–Liouville operator, the best we can do is to choose the coefficients $c_k = (Y_k, f)_w$ exactly as in its true expansion. This is also an important first step in checking that the expansion (2.24) of $f(x)$ does indeed converge on the original function as the number of included terms tends to infinity.