Fredholm determinants

Henry P. McKean[1]

1 CIMS, 251 Mercer Street, New York, NY 10012 USA

Abstract: The article provides with a down to earth exposition of the Fredholm theory with applications to Brownian motion and KdV equation.

MSC: 46E20, 35Q53

Keywords: Fredholm detemninant • Brownian motion • KdV equation
© *Versita Sp. z o.o.*

Preface

The work of I. Fredholm [10] is one of the loveliest chapters in mathematics, coming just at the turn of the century, when serious attention was first given to analysis in infinite–dimensional (function) space. Nowadays, it is explained mostly in an abstract and, if I may say so, a superficial way. So I thought it would be nice to do it from scratch, concretely, following Fredholm in using only the simplest methods and adding a few illustrations of what it is good for. The treatment has much in common with Peter Lax [16] which you may wish to look at, too; see also Courant–Hilbert [6] and/or Whittaker–Watson [25], and, of course Fredholm's 1903 paper, itself.

Cramer's rule expresses the inverse of a square d matrix by ratios of sub–determinants, as I trust you know. There, the dimension is finite. What Fredholm did was to adapt this rule to infinite dimensions, i.e. to function space. Poincaré had posed the problem, giving as his opinion that it would not be solved in his lifetime. Fredholm did it just two years later.

Chapter 1 is to remind you about determinants and Cramer's rule in $d < \infty$ dimensions, first in the conventional form and then in Fredholm's form which is specially adapted to the passage from $d = 1, 2, 3, \ldots$ to $d = \infty$. Obviously, if this is to work, you want only matrices which are close enough to the identity, i.e. something like $I + K$ where I is the identity, K is "small", and you need to express $\det(I + K)$ as $1 + [\text{terms in } K, K^2, \ldots]$ under effective control. This is a matter of bookkeeping. Look at a $I + \lambda K$ with a variable λ, develop *its* determinant in powers of λ (it is a polynomial after all), and put λ back equal to 1. This produces the formula

$$\det(I + K) = 1 + \sum_{1 \le i \le d} K_{ij} + \sum_{1 \le i < j \le d} \det \begin{bmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{bmatrix} + \sum_{1 \le i < j < k \le d} \det \begin{bmatrix} K_{ii} & K_{ij} & K_{ik} \\ K_{ji} & K_{jj} & K_{jk} \\ K_{ki} & K_{kj} & K_{kk} \end{bmatrix} + \ldots + \det K,$$

which is the key to the passage to $d = \infty$ carried out in Chapter 2. Think now of K as an integral operator

$$K : f \mapsto \int_0^1 K(x, y) f(y) \, dy,$$

acting on $C[0, 1]$ say, and approximate its action by that of the matrix

$$\left[K\left(\frac{i}{n}, \frac{j}{n}\right) \frac{1}{n} \right]_{1 \le i, j \le d}$$

acting in dimension d. Then the preceding formula for the determinant of the *matrix* $I + K$ makes plain the good sense of Fredholm's recipe for the determinant of the *operator* $I + K$:

$$\det (I + K) = 1 + \int_{0 \le x \le 1} K(x, x) \, dz + \int_{0 \le x_1 < x_2 \le 1} \det \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) \\ K(x_2, x_1) & K(x_2, x_2) \end{bmatrix} d^2 x$$

$$+ \int_{0 \le x_1 < x_2 < x_3 \le 1} \det \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & K(x_1, x_3) \\ K(x_2, x_1) & K(x_2, x_2) & K(x_2, x_3) \\ K(x_3, x_1) & K(x_3, x_2) & K(x_3, x_3) \end{bmatrix} d^3 x + \text{ad infinitem.}$$

Chapter 2 develops this idea along with Cramer's rule, in detail. (Don't worry. I'll make a better notation when the time comes.) Particular attention is paid to the special but important case of symmetric kernels: $K(x, y) = K(y, x)$; of these I give only the simplest examples.

Chapter 3 presents three serious examples of what you can do with this machinery. The first application is to Gaussian probabilities with emphasis on Brownian motion, as initiated by Cameron and Martin [3]. The second is to the Kortweg–de Vries equation (KdV) describing long waves in shallow water. Here, I follow mostly Ch. Pöppe's [21] explanation of a formula of Dyson [8], representing the solution of KdV by means of a special type of Fredholm determinant. The third application is to the N-dimensional unitary ensemble, so-called, i.e. the group of $N \times N$ unitary matrices equipped with its invariant volume element. What is in question is the distribution of eigenvalues of a typical matrix both for $N < \infty$ and for $N = \infty$. Here, the principal names are Dyson [7], Jimbo *et al.* [12], Mehta [19], and Tracy–Widom [22].

I cannot close this preface without a little joke. I gave a lecture in Paris a while back: something about "Fred Holm's determinants" the advertisement said. Oh well! The name is Ivar. Don't forget it.

Prerequisites

Not much in fact. You will need: 1) standard linear algebra, both real and complex; 2) calculus in several variables, of course; 3) some acquaintance with Lebesgue measure (on the unit interval only) and the associated space $L^2[0, 1]$ of (real) measurable functions with $\int_0^1 f^2 < \infty$; 4) elementary complex function-theory. I will explain anything non-standard or more advanced. Good references are: for 1), Lax [15]; for 2), Courant [5]; for 3), Munroe [20]; for 4), Ahlfors [1].

1. An old story with a new twist

I collect here what you need to know about linear maps of the finite-dimensional space $\mathbb{R}^d$ to itself, first in the conventional format and then in Fredholm's way.

1.1. Simplest facts

$L : \mathbb{R}^d \to \mathbb{R}^d$ is the map, $R = \{y : y = Lx \text{ for some } x \in \mathbb{R}^d\}$ is its range, $N = \{x : Lx = 0\}$ its null space, and similarly for the transposed map $L^\dagger$: range $R^\dagger$ and null space $N^\dagger$. The figure displays these notions in a convenient way.
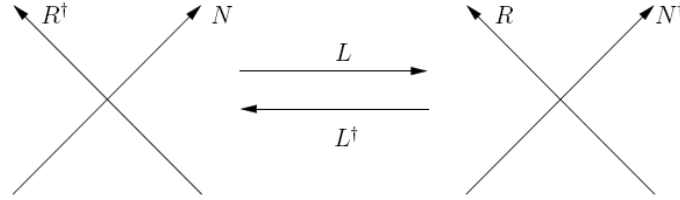
The position of $N^\dagger$ to the right indicates that it is the annihilator R^o of R: Indeed, $y \in R^o$ if and only if $(Lx, y) = (x, L^\dagger y)$ vanishes for every x, which is to say $L^\dagger y = 0$, and similarly to the left: $N = (R^\dagger)^o$. L kills N and so maps $R^\dagger$ one-to-one onto R; likewise, $L^\dagger$ kills $N^\dagger$ and maps R one-to-one onto $R^\dagger$. It follows that $\dim R^\dagger = \dim R \equiv r$ and so also $\dim N^\dagger = \dim N \equiv n = d - r$. The map $L : x \to y$ may be expressed by a matrix $M = [m_{ij}]_{1 \le i,j \le d}$ relative to the standard frame:

$$e_1 = (100 \ldots 0), \qquad e_2 = (010 \ldots 0), \qquad \ldots, \qquad e_d = (000 \ldots 1)$$

as in

$$y_i = \sum_{j=1}^d m_{ij} x_j, \qquad i = 1, \ldots, d.$$

R is the span of the columns of M, $R^\dagger$ the span of its rows, and $\dim R^\dagger = \dim R$ expresses the equality of "row rank" and "column rank".

1.2. Determinant and Cramer's rule

It is desirable to have a practical way to tell if $L : \mathbb{R}^d \to \mathbb{R}^d$ is one-to-one and onto (and so invertible), and also a recipe for its inverse L^{-1} in this case. Introduce (out of the air) the determinant:

$$\det M = \sum_{\pi \in S_d} \text{sign } \pi \prod_{i=1}^d m_{i\pi(i)},$$

the sum being taken over the "symmetric group" S_d of all permutations π of the "letters" $1, 2, \ldots, d$. Then L is one-to-one and onto if and only if $\det M \ne 0$, in which case L^{-1} is expressed by the inverse matrix $M^{-1} = [m_{ij}^{-1}]_{1 \le i,j \le d}$ formed according to Cramer's rule:

1) strike out the ith row and jth column of M and take the $(d-1) \times (d-1)$ determinant of what's left, this being the ij "cofactor" of M, so-called; **2)** supply the signature $(-1)^{i+j}$; **3)** transpose; **4)** divide by $\det M$. More briefly,

$$\bullet \qquad m_{ij}^{-1} = (\det M)^{-1}[(-1)^{i+j} \times \text{ the } ji \text{ cofactor of } M].$$

This is, I trust, familiar. If not, don't worry. I will derive it all in a geometrically appealing way in the next three sections.

1.3. Signed volume

The determinant, which seems at first so complicated and unmotivated, has a simple geometrical meaning: It is the volume V of the parallelepiped spanned by the vectors $f_1 = Le_1, \ldots, f_d = Le_d$, with a suitable signature attached. Plainly, L is one-to-one and onto if and only if $\dim R = d$, which is to say that the family $F = [f_1, \ldots, f_d]$ is a "frame", *i.e.* that it spans $\mathbb{R}^d$, and this in turn is the same as to say that the associated (signed) volume $V = V(F)$ does not vanish. I discuss this signed volume one dimension at a time.

d = 1. $e_1 = +1$ is the standard frame; any other frame F is just a number $f_1 \neq 0$, and $V(F) = f_1$ is the length $|f_1|$ with the signature of f_1 attached. The frame f_1 is "proper" or "improper" according as f_1 points in the same direction as $e_1 = +1$ or not. Obviously, you cannot deform a proper into an improper frame without passing through $f_1 = 0$ which is not a frame at all, *i.e.* the proper and improper frames are disconnected. This is all simple stuff made complicated. The next step conveys the idea more plainly.

d = 2. Here is the standard frame.
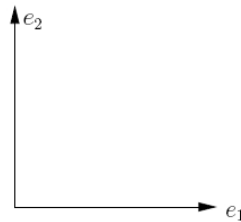
A second frame $F = [f_1, f_2]$ can be lined up with it by a (counter–clockwise) rotation aligning f_1 with e_1 as in the next figure.
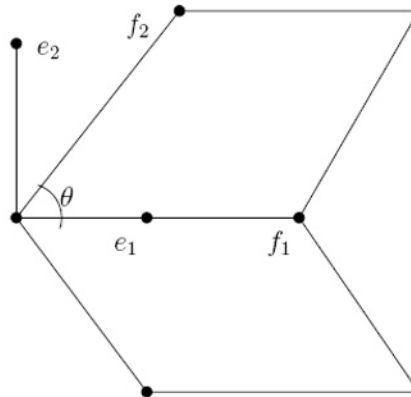
The resulting frame is "proper" or "improper" according as the rotated f_2 lies in the upper or the lower half plane. It cannot lie on the line directed by e_1 as that would make it depend upon f_1; in particular, the proper and improper frames are disconnected, one from the other, as in $d = 1$. The signed volume V is now declared to be the area of the parallelogram spanned by F, taken with the signature plus or minus according as the frame is proper or not, *i.e.*

$$V = |f_1| \, |f_2| \times \sin \theta,$$

so that the plus sign prevails for $0 < \theta < \pi$ and the minus sign for $\pi < \theta < 2\pi$. Let's spell it out: The absolute volume is

$$|V| = |f_1| \times \left|\text{the coprojection of } f_2 \text{ on } f_1\right| = |f_1| \times |f_2 - e_1(f_2, e_1)|$$
$$= |f_1| \times \sqrt{|f_2|^2 - |f_1|^{-2}(f_2, f_1)^2} \qquad e_1 \text{ being } f_1 \text{ divided by its length } |f_1|$$
$$= \sqrt{|f_1|^2 |f_2|^2 - (f_2, f_1)^2} = \sqrt{(f_{11}^2 + f_{12}^2)(f_{21}^2 + f_{22}^2) - (f_{21}f_{11} + f_{22}f_{12})^2}$$
$$= \sqrt{(f_{11}f_{22} - f_{12}f_{21})^2} = \pm(f_{11}f_{22} - f_{12}f_{21}).$$

d = 3. Here is the standard frame again, obeying the right–hand screw rule.
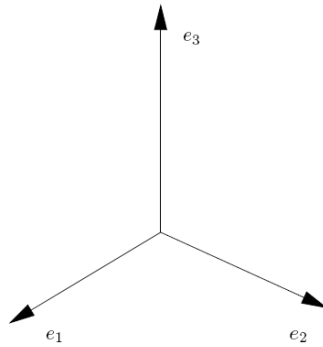
The absolute volume for the frame $F = [f_1, f_2, f_3]$ is

$$|V| = |f_1| \times \left|\text{coprojection of } f_2 \text{ on } f_1\right| \times \left|\text{coprojection of } f_3 \text{ on the plane of } f_1 \text{ \& } f_2\right|,$$

but what signature should be attributed to V itself? To decide, consider the following four steps: 1) rotate F so as to align f_1 with e_1 and adjust its length so they match; 2) rotate about e_1 to bring f_2 to the "upper" half of the plane of e_1 and e_2 and deform it to e_2 plain; 3) f_3 now lies either above or below the plane of e_1 & e_2 and can be deformed to $\pm e_3$: in short, F can be deformed (via honest frames) to one of the two frames $[e_1, e_2, \pm e_3]$, and it is *this* signature, attached to e_3, that is ascribed to V. In a fancier language, the frames of $\mathbb{R}^3$ fall into two connected deformation classes, "proper" and "improper", distinguished by the above signature, much as for $d \leq 2$, which is to say that $\mathbb{R}^3$ is capable of two different "orientations" exemplified by the right– and left–hand screw rules.

d ≥ 4. It is much the same in higher dimensions. A frame determines an increasing family of subspaces $\mathbb{R}^n = \text{span}\{f_k : 1 \leq k \leq n\}$, ending with $\mathbb{R}^d$ itself; the absolute volume is the product, from $n = 1$ up to d, of the length of the coprojection of f_n on $\mathbb{R}^{n-1}$; and the signature of V is determined by the deformation class, of which there are two, as before: Any frame may be deformed to $[e_1, e_2, \ldots, \pm e_d]$, and it is *this* signature, attached to e_d, that is ascribed to V. The definition of the signed volume is now complete, but what has it to do with the determinant? I postpone the punchline in favor of two simple observations.

Observation 1.

V is "skew" in a sense I will now explain. Let $\pi \in S_d$ be a permutation of the letters $1, \ldots, d$. Then $V(\pi F) = V\left[f_{\pi(1)}, f_{\pi(2)}, \ldots\right]$ represents the same absolute volume as $V(F)$: it is only the signature that changes, as in $V(\pi F) = \chi(\pi)V(F)$ with $\chi(\pi) = \pm 1$. Plainly, this χ is a character of S_d meaning that $\chi(\pi_1 \pi_2) = \chi(\pi_1)\chi(\pi_2)$, and it is a fact that S_d has only one such character besides the trivial character $\chi \equiv 1$: Obviously, every permutation is a product of "transpositions", exchanging two letters, fixing the rest. Now you will check that if χ ascribes the value $+1$ to *any* transposition, then it is the same for *every* transposition, in which case χ is trivial, i.e. $\chi \equiv 1$. Otherwise, it ascribes the

value -1 to every transposition and $\chi(\pi) = (-1)^{\#}$, $\#$ being the number of transpositions in π. This is nothing but the signature of π appearing in the formula for the determinant. The meaning of "V skew" is now this: $V(\pi F) = \chi(\pi) \times V(F)$.

Aside. Here you have a somewhat indirect proof that $\chi(\pi) = (-1)^{\#}$, alias the *signature* of π, is well-defined, i.e. that the parity of $\#$ is always the same, however π may be decomposed into transpositions. Another way is to examine how π changes the signature of the "discriminant" $\prod_{i>j}(x_i - x_j)$ of d distinct real numbers $x_1 < x_2 < \cdots < x_d$. Try it.

Observation 2.

V is now extended to any family of vectors $F = [f_1, \ldots, f_d]$, independent or not, by putting $V = 0$ in the second case, when the parallelepiped collapses. Then (most important) V is linear in each of its arguments $f_1, \ldots, f_d$. Because V is skew, you have to check this only for f_d. That is obvious for $d = 2$ from the explicit formula $V = f_{11}f_{22} - f_{12}f_{21}$ and extends easily to $d \geq 3$: in computing V, the part of f_d in the span of f_n, $1 \leq n < d$, changes nothing; it is only the projection of f_d upon the normal g to that span which counts, as in $V(F) = V[f_1, \ldots, f_{d-1}, g] \times (f_d \bullet g)$. This makes the linearity obvious.

1.4.   Why V is the determinant

It remains to compute V from the two properties, 1) skewness and 2) the linearity elicited in Section 1.3, plus the normalization 3) $V[e_1, \ldots, e_d] = 1$. Write $f_i = \sum_{j=1}^{d} f_{ij}e_j$ for $1 \leq i \leq d$. Then

$$V = \sum_{j_1=1}^{d} \cdots \sum_{j_d=1}^{d} f_{1j_1} \ldots f_{dj_d} V[e_{1j_1}, \ldots, e_{1j_d}] \qquad \text{by 2).}$$

Here, it is only summands with distinct $j_1, \ldots, j_d$ that survive, V being skew by 1), so you may write

$$V = \sum_{\pi \in S_d} V[e_{\pi(1)}, \ldots, e_{\pi(d)}] \prod_{i=1}^{d} f_{i\pi(i)} = \sum_{\pi \in S_d} \text{sign}\,\pi \prod_{i=1}^{d} f_{i\pi(i)} \qquad \text{by 1) and 3)}$$
$$= \det\left[f_{ij}\right]_{1 \leq i,j \leq d},$$

and if now $f_1 = Le_1$, etc., so that the preceding matrix is the old M transposed, then it is seen that $\det M^{\dagger}$ is nothing but the signed volume of the image of the unit cube C under L. Actually, $\det M^{\dagger} = \det M$, as in rule 1 of the next section, so with an imprecise notation, $\det M = V(LC)$.

Aside. Besides its interpretation as a signed volume, $\det M$ has also a spectral meaning: it is the value of the characteristic polynomial $\det(M - \lambda I)$ at $\lambda = 0$; as such, it is the product of the (generally complex) eigenvalues of M. These are either real or else they appear in conjugate pairs, M being real, in accord with the reality of the determinant.

1.5.   Simplest rules

Here, A, B, C, etc. are $d \times d$ matrices.

Rule 1.

$\det C^{\dagger} = \det C$: indeed,
$$\sum \chi(\pi) \prod_{i=1}^{d} c_{i\pi(i)} = \sum \chi(\pi) \prod_{j=1}^{d} c_{j\pi^{-1}(j)} = \sum \chi\left(\pi^{-1}\right) \prod_{j=1}^{d} c_{j\pi(j)}$$

since π^{-1} runs once through the symmetric group as π does so, and $\chi(\pi) = \chi(\pi^{-1})$.

Rule 2.

$\det(AB) = \det A \times \det B = \det(BA)$. This is also pretty clear. The absolute value $|\det C|$ represents the factor by which the volume of *any* geometrical figure in $\mathbb{R}^d$ is changed by application of the associated linear map, as you will see by dividing space up into little cubes, so the formula is correct, up to a signature to the right in case none of the determinants vanish. But then each of the associated frames can be deformed without change of signature to $[e_1, \ldots, e_d]$ or to $[e_1, \ldots, -e_d]$, and you can check the rest by hand: for example, if A is proper (orientation-preserving) and B is improper (orientation-reversing), then AB is improper and the signs work out as follows $(-1) = (+1) \times (-1)$.

Rule 3.

$\det(C^{-1}) = (\det C)^{-1}$ if $\det C \neq 0$, as is immediate from Rule 2.

Rule 4.

$\det C$ does not depend on the basis of $\mathbb{R}^d$. This is important: It means that the determinant is an attribute of the associated *geometrical* transformation $\mathbb{R}^d \to \mathbb{R}^d$ and not an artifact of its *presentation* by the matrix C. The point is that under change of basis C is replaced by $D^{-1}CD$ with $\det D \neq 0$ and $\det(D^{-1}CD) = \det C$ by Rules 2 and 3.

Rule 5.

$\ln \det C = \mathrm{sp} \ln C$ if C is near the identity. Here sp is the trace. Now it is better to write $C = I + K$ with small K so that $\ln C = \ln(I + K) = K - K^2/2 + K^3/3 - \ldots$, by definition. The identity is trivial at $K = 0$ ($C = I$). Let ΔK be a tiny variation of K. Then, with a self-evident notation,

$$
\begin{aligned}
\Delta \ln \det(I + K) &= \ln\left[\det(I + K + \Delta K)/\det(I + K)\right] \\
&= \ln \det\left(I + (I + K)^{-1}\Delta K\right) \qquad \text{by Rules 2 and 3} \\
&\simeq \ln\left[1 + \mathrm{sp}\,(I + K)^{-1}\Delta K\right] \qquad \text{why?} \\
&\simeq \mathrm{sp}\,(I + K)^{-1}\Delta K = \mathrm{sp}\left(I - K + K^2 - K^3 + \ldots\right)\Delta K \\
&= \mathrm{sp}\left[\Delta K - \frac{1}{2}(K\Delta K + \Delta K K) + \frac{1}{3}\left(K^2\Delta K + K\Delta K K + \Delta K K^2\right) - \ldots\right] \\
&= \mathrm{sp}[\Delta \ln(I + K)] = \Delta \mathrm{sp} \ln(I + K).
\end{aligned}
$$

The identity $\mathrm{sp}(AB) = \mathrm{sp}(BA)$ is used in line 5. I leave the rest to you.

1.6. Cramer's rule

The determinant of M being linear in each of its columns,

$$
\frac{\partial \det M}{\partial m_{11}} = \sum_{\pi \in S_d} \chi(\pi) \prod_{k=2}^{d} m_{k\pi(k)} \times \begin{Bmatrix} 1 \text{ if } \pi(1) = 1 \\ 0 \text{ if } \pi(1) \neq 1 \end{Bmatrix} = \sum_{\pi' \in S_{d-1}} \chi(\pi') \prod_{k=2}^{d} m_{k\pi'(k)}
$$

where π' permutes $2, \ldots, d$ leaving 1 fixed. This is nothing but the 11 cofactor of M. To evaluate $\partial \det M/\partial m_{ij}$, put the ith row/jth column in first place, forming a new matrix M' with $m'_{11} = m_{ij}$. This can be done by permutations of rows and columns requiring $i + j$ transpositions, counting modulo 2: for example, m_{23} is brought to the place of m_{11} by transposing a) columns 3 and 2, b) columns 1 and 2, and c) rows 1 and 2, requiring $5 = (2 + 3)$ transpositions. It follows that $\det M' = (-1)^{i+j} \det M$, so the rule is

$$
\frac{\partial \det M}{\partial m_{ij}} = (-1)^{i+j} \times [\text{the } ij \text{ cofactor of } M].
$$

Now look at

$$\sum_{k=1}^{d} \frac{\partial \det M}{\partial m_{ki}} \times m_{kj},$$

in which you see $\partial \det M/\partial m$ transposed. This represents the signed volume associated to M with its ith column replaced by its jth column; as such, it is just $\det M$ if $i = j$, as you will see by Observation 2 of Section 1.3, but the parallelepiped collapses if $i \neq j$ and you get 0, *i.e.*, with an abbreviated notation,

$$\left(\frac{\partial \det M}{\partial m} \right)^{\dagger} M = \det M \times [\text{the identity}].$$

That is Cramer's rule, expressing M^{-1} as $(\det M)^{-1} \times (\partial \det M/\partial m)^{\dagger}$.

1.7. Fredholm's idea

To extend this machinery from $d = 1, 2, 3$, etc. to $d = \infty$ dimensions, it is clear that M cannot be too far from the identity, *i.e.* you must have $M = I + K$ with K "small". For example, the $\infty \times \infty$ determinant

$$\det \begin{bmatrix} 1+1 & 0 & 0 & 0 & \ldots \\ 0 & 1+1/2 & 0 & 0 & \ldots \\ 0 & 0 & 1+1/3 & 0 & \ldots \\ 0 & 0 & 0 & 1+1/4 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix} = \prod_{n=1}^{\infty} \left(1 + \frac{1}{n} \right) = +\infty$$

makes no sense, but

$$\det \begin{bmatrix} 1 & 1/2 & 0 & 0 & 0 & \ldots \\ 1/2 & 1 & 1/4 & 0 & 0 & \ldots \\ 0 & 1/4 & 1 & 1/8 & 0 & \ldots \\ 0 & 0 & 1/8 & 1 & 1/16 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

should be OK: after all the determinant as volume cannot be more than the product of the lengths of the individual columns, so the square of the determinant should be less than

$$\left(1 + \frac{1}{4} \right) \left(1 + \frac{1}{4} + \frac{1}{16} \right) \left(1 + \frac{1}{16} + \frac{1}{64} \right) \cdot \ldots = \frac{5}{4} \left(1 + \frac{5}{16} \right) \left(1 + \frac{5}{64} \right) \cdot \ldots < \frac{5}{4} \exp \left(5 \sum_{n=2}^{\infty} 4^{-n} \right) = \frac{5}{4} \exp \frac{5}{12}.$$

The task is now to re-express $\det(I + K)$ in such a way that the smallness of K can be exploited when $d \uparrow \infty$. Following Fredholm, this is done by introducing a parameter λ, for book-keeping, as in $\Delta(\lambda) = \det(I + \lambda K)$, expanding in powers of λ:

$$\Delta(\lambda) = \Delta_0 + \lambda \Delta_1 + \lambda^2 \Delta_2 + \cdots + \lambda^d \Delta_d,$$

and putting λ back equal 1. Here, $\Delta_0 = \Delta(0) = 1$, $\Delta_d = \det K$ as you will check, and the rest is not hard to come by: For $1 \leq p \leq d$ and $1 \leq i_1 < i_2 < \ldots < i_p \leq d$, write $\mathbf{i} = (i_1, \ldots, i_p)$, and $|\mathbf{i}| = p$. Then, with $[e_{ij}] = I$, the identity matrix,

$$\det(I + \lambda K) = \sum_{\pi} \chi(\pi) \prod_{i=1}^{d} \left(e_{i\pi(i)} + \lambda K_{i\pi(i)} \right) = \sum_{p=1}^{d} \lambda^p \sum_{|\mathbf{i}|=p} \sum_{\pi \text{ fixing } \mathbf{i}} \chi(\pi) \prod_{i \in \mathbf{i}} K_{i\pi(i)} = \sum_{p=1}^{d} \lambda^p \sum_{|\mathbf{i}|=p} \det[K_{ij}]_{i,j\in\mathbf{i}},$$

i.e.

$$\det(I + \lambda K) = 1 + \lambda \sum_{1 \leq i \leq d} K_{ii} + \lambda^2 \sum_{1 \leq i < j \leq d} \det \begin{bmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{bmatrix} + \lambda^3 \sum_{1 \leq i < j < k \leq d} \det \begin{bmatrix} K_{ii} & K_{ij} & K_{ik} \\ K_{ji} & K_{jj} & K_{jk} \\ K_{ki} & K_{kj} & K_{kk} \end{bmatrix} + \ldots + \lambda^d \det K$$

$$= 1 + \lambda \Delta_1 + \lambda^2 \Delta_2 + \lambda^3 \Delta_3 + \ldots + \lambda^d \Delta_d.$$

Cramer's rule is to be recast in this format with a view to inverting $I + \lambda K$ if $\Delta(\lambda) \neq 0$. Taking λ into account, the recipe is

$$(I + \lambda K)^{-1} = \Delta^{-1} \times \frac{1}{\lambda} \times \left(\frac{\partial \Delta}{\partial K}\right)^{\dagger},$$

but I want to spell it out.

Let $\mathbf{i}$ be any *un*ordered collection of p distinct indices $1 \leq i_1, \ldots, i_p \leq d$. $[K_{ij}]_{1 \leq i, j \leq p}$ is abbreviated as $K_{\mathbf{ij}}$. Then

$$\Delta_p = \frac{1}{p!} \sum_{\substack{|\mathbf{i}|=p \\ \mathbf{i}=\mathbf{j}}} \det K_{\mathbf{ij}}$$

with a self-evident notation. Here, the extension of the sum to unordered indices merely produces $p!$ copies of each sub-determinant in Δ_p, whence the factor $1/p!$. Now look at $\partial \Delta_{p+1}/\partial K_{ij}$ on diagonal: with further self-evident notations,

$$\frac{\partial \Delta_{p+1}}{\partial K_{ii}} = \frac{1}{(p+1)!} \times \left[\text{the sum over } \mathbf{i} = \mathbf{j} \ni i \text{ with } |\mathbf{i}| = p+1 \text{ of the } ii \text{ cofactor of } K_{\mathbf{ij}}\right]$$

$$= \frac{1}{(p+1)!} \times \left[\text{the } p+1 \text{ ways of placing } i \text{ in } \mathbf{i}\right] \times \sum_{\substack{\mathbf{i}'=\mathbf{j}'\neq i \\ |\mathbf{i}'|=p}} \det K_{\mathbf{i}'\mathbf{j}'}$$

$$= \frac{1}{p!} \times [\text{the unrestricted sum over } \mathbf{i}' - \text{ the restricted sum with } \mathbf{i}' \ni i]$$

$$= \Delta_p - \frac{1}{p!} \times \left[\text{the } p \text{ ways of placing } i \text{ in } \mathbf{i}'\right] \times \sum_{\substack{\mathbf{i}=\mathbf{j}\neq i \\ |\mathbf{i}|=p-1}} \det \begin{bmatrix} K_{ii} & K_{\mathbf{ij}} \\ K_{\mathbf{ii}} & K_{\mathbf{ij}} \end{bmatrix} = \Delta_p - \frac{1}{(p-1)!} \sum_{\substack{\mathbf{i}=\mathbf{j}\ni i \\ |\mathbf{i}|=p-1}} \det \begin{bmatrix} K_{ii} & K_{\mathbf{ij}} \\ K_{\mathbf{ii}} & K_{\mathbf{ij}} \end{bmatrix},$$

from which the restriction $\mathbf{i} \not\ni i$ can be removed since the determinant vanishes if $\mathbf{i} \ni i$. The off-diagonal is much the same: with $i < j$ for definiteness,

$$\left(\frac{\partial \Delta_{p+1}}{\partial K_{ij}}\right)^{\dagger} = \frac{1}{(p+1)!} \times \left[\text{the sum over } \mathbf{i} \ni \text{ both } i \text{ and } j \text{ of } (-1)^{i+j}\right] \times \left[\text{the } ji \text{ cofactor of } K_{\mathbf{ij}}\right]$$

$$= \frac{1}{(p+1)!} \sum_{\mathbf{i}=\mathbf{j}\ni i,j} (-1)^{i+j} \times [\text{the determinant of the matrix displayed below}]$$

in which you keep $ij(\bullet)$ and remove $ji(\circ)$ together with its attendant row and column. Let's move ij to the upper left-hand corner, keeping the indices $k \neq i$ or j in their original order. This takes $i - 1$ transpositions for i and $j - 2$ for j, the intervening ith column having been crossed out, for a total of $i - 1 + j - 2 \equiv i + j + 1 \pmod 2$ changes of sign. Taking the factor $(-1)^{i+j}$ into account leaves just one change of sign, *net*, and if now you also reflect that there are $(p+1) \times p$ ways of placing i and j in $\mathbf{i}$, you will obtain

$$\left(\frac{\partial \Delta_{p+1}}{\partial K_{ij}}\right)^{\dagger} = -\frac{1}{(p-1)!} \sum_{\substack{\mathbf{i}=\mathbf{j}\neq i,j \\ |\mathbf{i}|=p-1}} \det \begin{bmatrix} K_{ij} & K_{\mathbf{ij}} \\ K_{\mathbf{ij}} & K_{\mathbf{ij}} \end{bmatrix},$$

just as *on* diagonal, only now the term Δ_p is missing. I leave it to you to bring all this into the form

$$\frac{1}{\lambda}\left(\frac{\partial\Delta}{\partial K}\right)^{\dagger} = \Delta - \text{the matrix} \left[\sum_{p=0}^{d-1}\lambda^p \sum_{\substack{i=j\\|i|=p}}\det\begin{bmatrix}K_{ij} & K_{ij}\\K_{ij} & K_{ij}\end{bmatrix}\right]_{1\leq i,j\leq d}.$$

It is instructive to see what these objects are when expressed more compactly. You know that

$$(I+\lambda K)\lambda^{-1}\left(\frac{\partial\Delta}{\partial K}\right)^{\dagger} = \Delta$$

always, by Cramer's rule, and also that

$$(I+\lambda K)^{-1} = \sum_{p=0}^{\infty}(-\lambda K)^p$$

if λ is small, so

$$\sum_{p=0}^{\infty}\lambda^p\left(\frac{\partial\Delta_{p+1}}{\partial K}\right)^{\dagger} = \frac{1}{\lambda}\left(\frac{\partial\Delta}{\partial K}\right)^{\dagger} = \Delta\times(I+\lambda K)^{-1} = \sum_{p=0}^{\infty}\lambda^p\Delta_p\times\sum_{q=0}^{\infty}(-\lambda K)^q = \sum_{n=0}^{\infty}\lambda^n\sum_{p+q=n}\Delta_p(-K)^q,$$

i.e.

$$\left(\frac{\partial\Delta_{p+1}}{\partial K}\right)^{\dagger} = \sum_{q=0}^{p}\Delta_{p-q}(-K)^q.$$

It is this kind of book-keeping that is the key to Fredholm's success for $d = \infty$, as you will see in the next chapter.

2. Fredholm in function space

2.1. Where to start

I take the simplest case presenting the least technical difficulties; variants will readily suggest themselves and will be used in Chapter 3 without further comment. I propose to replace $\mathbb{R}^d$ by the ∞-dimensional space $C[0,1]$ and K by the integral operator

$$K : f \mapsto \int_0^1 K(x,y)\,f(y)\,dy$$

with kernel $K(x,y)$, $0 \leq x, y \leq 1$, of class $C[0,1]^2$. Here and below, I do not distinguish between the operator and the kernel by means of which it is expressed: K means what it has to mean, by the context.

2.2. Aside on compactness

It is always some kind of compactness that makes things manageable for $d = \infty$; in effect, it reduces $d = \infty$ to finite d. Two kinds of compactness are needed here. The proofs are relegated to an appendix.

The space $C[0,1]$.

This is the space of (real) continuous functions f on the unit interval, provided with the norm $\|f\|_{\infty} = \max\{|f(x)| : 0 \leq x \leq 1\}$. A figure $F \subset C$ is compact relative to the natural distance $d(f_1, f_2) = \|f_1 - f_2\|_{\infty}$ if it is 1) closed, 2) bounded, and 3) $|f(b) - f(a)| \leq \omega(|b - a|)$ for every $f \in F$, with one and the same, positive, increasing function $\omega(h)$, $h > 0$, subject to $\omega(0+) = 0$. Below, $f_n \to f_{\infty}$ means this type of convergence.

The space $L^2[0, 1]$.

This is the space of (real) Lebesgue–measurable functions subject to $\int_0^1 f^2\, dx < \infty$, with the Pythagorean norm $\|f\|_2 = \sqrt{\int_0^1 f^2\, dx}$ and the associated inner product $(f_1, f_2) = \int_0^1 f_1 f_2\, dx$. Here, the unit ball $B = \{\|f\|_2 \leq 1\}$ is not compact relative to the distance $d(f_1, f_2) = \|f_1 - f_2\|_2$: in fact, B contains the mutually perpendicular functions $e_n(x) = \sin n\pi x$, $n \geq 1$, and $d(e_i, e_j) = 1$ for any $i \neq j$. Fortunately, B is *weakly compact* and this will be enough. It means that if $e_n \in B$ for $n \geq 1$, then you can pick indices $n_1 < n_2 < n_3 <$ etc. and a function $e_\infty \in B$ so as to make (e_n, f) tend to (e_∞, f) as $n = n_1, n_2, \ldots \uparrow \infty$, for every $f \in L^2$ simultaneously. Below, $e_n \rightsquigarrow e_\infty$ means *this* type of convergence. Take it as an exercise to check that $\sin n\pi x \rightsquigarrow 0$.

Example.

Obviously, K maps L^2 into C. Let's prove that $F = KB$ is compact in C by verifying 1), 2), 3) in reverse order: 3), 2), 1).

3) For any function $f = Ke$ in KB,

$$|f(b) - f(a)| \leq \int_0^1 |K(b, y) - K(a, y)|\,|e(y)|\, dy \leq \|e\|_2 \sqrt{\int_0^1 |K(b, y) - K(a, y)|^2\, dy} \leq \omega(|b - a|)$$

with

$$\omega(h) = \max\left\{|K(b, y) - K(a, y)| : 0 \leq a, b \leq 1, |b - a| \leq h, 0 \leq y \leq 1\right\}.$$

3) follows from the fact that $\omega(0+) = 0$. *Why so?*

2) More simply,

$$|f(x)| \leq \sqrt{\int_0^1 |K(x, y)|^2\, dy} \leq \max\left\{|K(x, y)| : 0 \leq x, y \leq 1\right\} \equiv m < \infty,$$

so KB is bounded.

1) KB is closed: in fact, if $KB \ni f_n = Ke_n \to f_\infty$ and if you assume, as you may, that $e_n \rightsquigarrow e_\infty$, then you will have $f_\infty = Ke_\infty$, as required.

2.3. Some geometry

Now to business, recapitulating Section 1.1. Let $R/R^\dagger$ be the range and $N/N^\dagger$ the null space of $I + K/I + K^\dagger$, respectively, in the space $C[0, 1]$.

Item 1.

N, and likewise $N^\dagger$, is finite–dimensional; in particular, the associated (L^2) projections map L^2 into C.

Proof.   Pick e_k, $1 \leq k \leq n$, from N, each of unit length:

$$\|e_k\|_2^2 = \int_0^1 e_k^2 = 1$$

and mutually perpendicular:

$$(e_i, e_j) = \int_0^1 e_i e_j = 0, \qquad i \neq j.$$
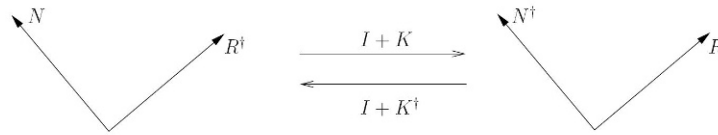
Then $Ke_k = -e_k$, so

$$\sum_{k=1}^{n} |e_k(x)|^2 = \sum_{k=1}^{n} \|\text{the projection of } K(x, \cdot) \text{ upon } e_k\|_2^2 \leq \int_0^1 (K(x,y))^2 \, dy,$$

this being an instance of Bessel's inequality, expressive of the Pythagorean rule. Now integrate to obtain

$$n = \sum_{k=1}^{n} \int_0^1 |e_k|^2 \, dx \leq \int_0^1 \int_0^1 |K(x,y)|^2 \, dxdy. \qquad \square$$

Item 2.

$N^\dagger = R^o$ and $N = (R^\dagger)^o$ as in the picture, just as for $d < \infty$; in particular, $(I + K) : R^\dagger \to R$ is one-to-one and onto.

Item 3.

R, and so also $R^\dagger$, is closed as it stands; in particular, C is the perpendicular sum of closed subspaces $R^\dagger \oplus N$ or $R \oplus N^\dagger$.

Proof.   Let $R \ni f_n = (I + K) e_n \to f_\infty$. If $\|e_n\|_2 \uparrow \infty$, then

$$f_n' = \frac{f_n}{\|e_n\|_2} = (I + K) e_n' \qquad \text{with} \qquad e_n' = \frac{e_n}{\|e_n\|_2},$$

and you may suppose that $e_n' \rightsquigarrow e_\infty'$. Now $f_n' \to 0$, so $e_n' = f_n' - Ke_n' \to -Ke_\infty'$ on the one hand and, on the other hand, to e_∞' itself, the upshot being that e_∞', which cannot vanish as it is of unit length, belongs to N. But you may also suppose, from the start, that $e_n \in N^o$, whence $e_\infty' \in N^o$ too, and this is contradictory. If now $\liminf \|e_n\|_2 < \infty$ you may suppose $e_n \rightsquigarrow e_\infty$. Then $Ke_n \to Ke_\infty$ and $f_n \to f_\infty$ force $e_n \to e_\infty$ and so also $f_\infty = (I + K) e_\infty$, as wanted. $\square$

This was all plain sailing with the help of compactness. What is not clear is whether $\dim N = \dim N^\dagger$ is still true. Now you can't just count dimensions as for $d < \infty$ when you could see at once that $\dim R = \dim R^\dagger$. Both $\dim R$ and $\dim R^\dagger$ are infinite and it is not obvious what to do.

Item 4.

$n = \dim N$ and $n^\dagger = \dim N^\dagger$ are one and the same.

Proof.   For conversation's sake let $n^\dagger$ be larger than n, let e_k, $1 \leq k \leq n$, and $e_k^\dagger$, $1 \leq k \leq n^\dagger$, be unit mutually perpendicular functions spanning N and $N^\dagger$ respectively, and replace K by

$$K' = K + \sum_{k=1}^{n} \left[e_k^\dagger \otimes e_k \equiv e_k^\dagger(x) e_n(y) \right].$$

Then $I + K'$ has no null space: It maps C one-to-one onto $R' = R \oplus \text{span} \{ e_n^\dagger : 1 \leq k \leq n \}$, permitting a reduction to the simplified case: $n = 0 < n^\dagger$. Pick $e^\dagger$ from $N^\dagger$ and discretize $(I + K)^\dagger e^\dagger = 0$ so:

$$e^\dagger \left(\frac{i}{d} \right) + \sum_{j=1}^{d} K \left(\frac{j}{d}, \frac{i}{d} \right) \frac{1}{d} e^\dagger \left(\frac{j}{d} \right) \equiv f \left(\frac{i}{d} \right) \qquad \text{for} \quad 1 \leq i \leq d$$

with

$$\sum_{k=1}^{d} \left| e^\dagger \left(\frac{k}{d} \right) \right|^2 \frac{1}{d} = 1$$

as you may do and

$$\max_i \left| f \left(\frac{i}{d} \right) \right| = o(1) \qquad \text{as} \quad d \uparrow \infty.$$

This can be rewritten:

$$e^\dagger \left(\frac{i}{d} \right) + \sum_{j=1}^{d} \left[K \left(\frac{j}{d}, \frac{i}{d} \right) - f \left(\frac{i}{d} \right) e^\dagger \left(\frac{j}{d} \right) \right] \frac{1}{d} e^\dagger \left(\frac{j}{d} \right) = 0.$$

Now for $d < \infty$, it is guaranteed that the transposed problem also has a solution, e:

$$e \left(\frac{i}{d} \right) + \sum_{j=1}^{d} \left[K \left(\frac{i}{d}, \frac{j}{d} \right) - e^\dagger \left(\frac{i}{d} \right) f \left(\frac{j}{d} \right) \right] \frac{1}{d} e \left(\frac{j}{d} \right) = 0$$

with

$$\sum_{k=1}^{d} \left| e \left(\frac{k}{d} \right) \right|^2 \frac{1}{d} = 1$$

if you want, and the idea is to produce a (contradictory) null function $e \in N$ by making $d \uparrow \infty$ here. This is easy:

$$\left| \sum_{j=1}^{d} f \left(\frac{j}{d} \right) e \left(\frac{j}{d} \right) \frac{1}{d} \right|^2 \leq \sum_{j=1}^{d} \left| f \left(\frac{j}{d} \right) \right|^2 \frac{1}{d}.$$

is negligible, so

$$e \left(\frac{i}{d} \right) + \sum_{j=1}^{d} K \left(\frac{i}{d}, \frac{j}{d} \right) e \left(\frac{j}{d} \right) \frac{1}{d} = o(1)$$

in which the error $o(1)$ is small independently of $1 \leq i \leq d$. Let $f_d(x)$ be the function with values $e(i/d)$ for $(i - 1)/d \leq x < i/d$, and suppose, as you may, that $f_d \rightsquigarrow f_\infty$, $\| f_d \|_2$ being unity by display no. 3 above. Then

$$f_d(x) + \int_0^1 K(x, y) f_d(y) \, dy = 0$$

up to a like error, small independently of $0 \leq x \leq 1$, and if you now make $d \uparrow \infty$, you will find 1) $K f_d \to K f_\infty$, 2) $f_d \to f_\infty$ with $\| f_\infty \|_2 = 1$, and 3) $(I + K) f_\infty = 0$, with the implication $f_\infty \in C$, i.e. $n > 0$ which is contradictory. $\qquad \square$

2.4. Fredholm's determinant

To make a determinant for $I + K$ in the ∞–dimensional space $C[0, 1]$, you do the obvious thing: Replace the function $f(x)$, $0 \leq x \leq 1$, by the d-dimensional vector $f(k/d)$, $1 \leq k \leq d$, and the operator K by the $d \times d$ matrix $\left[K(i/d, j/d)/d\right]_{1 \leq i, j \leq d}$ so that

$$\sum_{j=1}^{d} K\left(\frac{i}{d}, \frac{j}{d}\right) \frac{1}{d} f\left(\frac{j}{d}\right) \qquad \text{approximates} \qquad \int_0^1 K(x, y) f(y) \, dy,$$

write out Fredholm's series for the determinant of $\left[I + K(i/d, j/d)/d\right]_{1 \leq i, j \leq d}$ as in

$$1 + \sum_{p=1}^{d} \frac{1}{p!} \sum_{|\mathbf{k}|=p} \det\left[K\left(\frac{i}{d}, \frac{j}{d}\right) \frac{1}{d}\right]_{i,j \in \mathbf{k}},$$

and make $d \uparrow \infty$ to produce

$$\det(I + K) = 1 + \sum_{p=1}^{\infty} \frac{1}{p!} \int_0^1 \cdots \int_0^1 \det\left[K(x_i, x_j)\right]_{1 \leq i, j \leq p} d^p x.$$

The convergence is obvious term–wise, only you must have an estimate to be sure the tail of the sum is small. This is easy: a finite-dimensional determinant is the signed volume of the parallelepiped spanned by the rows/columns of the matrix in hand so with $|\mathbf{k}| = p$,

$$\det\left[K\left(\frac{i}{d}, \frac{j}{d}\right) \frac{1}{d}\right]_{i,j \in \mathbf{k}} \leq \prod_{i \in \mathbf{k}} \sqrt{\sum_{j \in \mathbf{k}} \left|K\left(\frac{i}{d}, \frac{j}{d}\right)\right|^2 \times d^{-2}} \leq p^{p/2} \left(\frac{m}{d}\right)^p,$$

in which $m = \max\left\{|K(x, y)| : 0 \leq x, y \leq 1\right\}$ and

$$\frac{1}{p!} \sum_{|\mathbf{k}|=p} \det\left[K\left(\frac{i}{d}, \frac{j}{d}\right) \frac{1}{d}\right]_{i,j \in \mathbf{k}}$$

is dominated by

$$\frac{1}{p!} d \, (d-1) \, \cdots \, (d-p+1) \, p^{p/2} \left(\frac{m}{d}\right)^p \lesssim \frac{p^{p/2} m^p}{p^{p + \frac{1}{2}} e^{-p} \sqrt{2\pi}} \qquad \text{by Stirling's approximation}$$

$$< (me)^p p^{-p/2}$$

independently of d. This is plenty.

Without further ado, the main rules (1,2,3) of Section 1.5 carry over:

Rule 1. $\det(I + K) = \det\left(I + K^\dagger\right)$.

Rule 2. $\det\left[(I + K)(I + G)\right] = \det(I + K) \times \det(I + G)$.

Rule 3 applies when $\det(I + K)$ does not vanish. Then, as will be seen in Section 2.5, $I + K$ has a nice inverse $I + G$ of the same type as itself, and Rule 2 implies $\det\left[(I + K)^{-1}\right] = \left[\det(I + K)\right]^{-1}$, as it should be.

Rule 4 is left aside; it's analogue is not needed.

Rule 5. $\ln \det(I + K) = \operatorname{sp} \ln(I + K)$ if everything makes sense. This is proved as in Section 2.4 for small K. Take it as an exercise. Here and below $\operatorname{sp} G$ means $\int_0^1 G(x, x) \, dx$.

Rule 6. If K depends upon a parameter h and $K' = dK/dh$ is nice, then $[\ln \det(I + K)]' = \text{sp}\left[(I + K)^{-1}K'\right]$. I leave it to you to see what is needed and make a proof.

The function $\Delta(\lambda) = \det(I + \lambda K)$ of $\lambda \in \mathbb{C}$ is central to the rest of the discussion: it is an integral function of order ≤ 2 in view of the estimate

$$|\Delta(\lambda)| \leq 1 + \sum_{p=1}^{\infty} \frac{(\text{constant} \times r)^p}{(p/2)!} \leq A \exp\left(Br^2\right)$$

with $r = |\lambda|$ and suitable constants A and B. Take this also as an exercise. The estimate can be much improved if $K(x, y)$ is smooth: for example, if it is of class C^1, then it pays, in the estimation of $\det[K(x_i, x_j)]_{1 \leq i,j \leq p}$, to subtract, from each row of index $i \geq 2$, the row before. This does not change the determinant and allows a better estimate:

$$\left| \det[K(x_i, x_j)]_{1 \leq i,j \leq p} \right| \leq [\text{an unimportant geometrical factor}] \times p^{p/2} \qquad \text{very roughly}$$
$$\times (x_2 - x_1)(x_3 - x_2) \ldots (x_p - x_{p-n})$$

for $x_1 < x_2 < \ldots < x_p$. Here, line 2 is largest when the points are equally spaced (*why?*), leading to the estimate

$$\left| \frac{1}{p!} \int \det[K(x_i, x_j)] \, d^p x \right| \leq \int_{x_1 < x_2 < \ldots < x_p} |\det K(x_i, x_j)| \, d^p x \leq p^{p/2} \times \left(\frac{1}{p}\right)^p \times \frac{1}{p!} \simeq \frac{1}{p^{3p/2}}, \qquad \text{roughly}$$

whence $|\Delta(\lambda)| \leq A \exp\left(Br^{2/3}\right)$, i.e. Δ is of order $2/3$ as the phrase is. Note in this connection, the implication of Hadamard's theorem (Ahlfors [1, p. 208]): that such a function of order < 1 with $\Delta(0) = 1$ is the product $\Pi(1 - \lambda/\lambda_n)$ over its roots λ_1, λ_2, *etc.* counted according to multiplicity.

Example.

Here $K(x, y) = x(1 - y)$ if $x \leq y$ and $y(1 - x)$ if $x > y$. Let's compute Δ explicitly: $\det[K(x_i, x_j)]_{1 \leq i,j \leq p}$ is a polynomial in $x_1, x_2, \ldots, x_p$ of degree 2 in each variable: for example with $p = 3$ and $x_1 < x_2 < x_3$,

$$\det \begin{bmatrix} x_1(1 - x_1) & x_1(1 - x_2) & x_1(1 - x_3) \\ x_1(1 - x_2) & x_2(1 - x_2) & x_2(1 - x_3) \\ x_1(1 - x_3) & x_2(1 - x_3) & x_3(1 - x_3) \end{bmatrix} = x_1(1 - x_3) \times \det \begin{bmatrix} 1 - x_1 & 1 - x_2 & 1 - x_3 \\ x_1(1 - x_2) & x_2(1 - x_2) & x_2(1 - x_3) \\ x_1 & x_2 & x_3 \end{bmatrix}$$

$$= x_1(1 - x_3) \times \det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_2 \\ x_1 & x_2 & x_3 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & 0 \\ x_1 & x_2 - x_1 & 0 \\ x_1 & x_2 - x_1 & x_3 - x_2 \end{bmatrix} = x_1(x_2 - x_1)(x_3 - x_2)(1 - x_3).$$

Now in general, $\det[K(x_i, x_j)]_{1 \leq i,j \leq p}$ vanishes: in respect to x_1 at 0 and x_2; in respect to x_2 at x_1 and x_3; in respect to x_3 at x_2 and x_4; etc., so what could it be but a constant multiple of

$$x_1(x_2 - x_1)(x_3 - x_2) \ldots (x_p - x_{p-1})(1 - x_p).$$

This constant is unity, as you will check, so

$$\frac{1}{p!} \int \det[K(x_i, x_j)]_{1 \leq i,j \leq p} \, d^p x = \int_0^1 (1 - x_p) \, dx_p \ldots \int_0^{x_3} (x_3 - x_2) \, dx_2 \int_0^{x_2} (x_2 - x_1) x_1 \, dx_1 = \frac{1}{(2p + 1)!}.$$

In short,

$$\Delta(\lambda) = \sum_{p=0}^{\infty} \frac{\lambda^p}{(2p + 1)!} = \frac{1}{\sqrt{\lambda}} \sum_{p=0}^{\infty} \frac{(\sqrt{\lambda})^{2p+1}}{(2p + 1)!} = \frac{\text{sh}\,\sqrt{\lambda}}{\sqrt{\lambda}},$$

which is of order $1/2$.

2.5. Fredholm's formulas

These implement Cramer's rule in dimension $d = \infty$. They follow at once from Section 1.7 in the style of Section 2.4, as I ask you to check. $[K(x_i, x_j)]_{1 \le i, j \le p}$ is abbreviated as $K(\mathbf{i}, \mathbf{j})$; the value of p will be obvious from the context. Now

$$\Delta(\lambda) = 1 + \sum_{p=1}^{\infty} \Delta_p \lambda^p \qquad \text{with} \qquad \Delta_p = \frac{1}{p!} \int_0^1 \cdots \int_0^1 \det K(\mathbf{i}, \mathbf{j}) \, d^p x,$$

and

$$\left(\frac{\partial \Delta_{p+1}}{\partial K(x, y)} \right)^\dagger = \Delta_p - \frac{1}{(p-1)!} \int_0^1 \cdots \int_0^1 \det \begin{bmatrix} K(x, y) & K(x, \mathbf{j}) \\ K(\mathbf{i}, y) & K(\mathbf{i}, \mathbf{j}) \end{bmatrix} d^{p-1} x$$

with a self-evident notation, a factor of $1/d$ having been dropped from both sides in comparison to Section 1.7: symbol-ically,

$$\left(\frac{\partial \Delta_{p+1}}{\partial K} \right)^\dagger = \sum_{q=0}^{p} \Delta_{p-q} (-K)^q.$$

Also,

$$\frac{1}{\lambda} \left(\frac{\partial \Delta}{\partial K} \right)^\dagger = \Delta - \lambda \sum_{p=0}^{\infty} \frac{\lambda^p}{p!} \int_0^1 \cdots \int_0^1 \det \begin{bmatrix} K(x, y) & K(x, \mathbf{j}) \\ K(\mathbf{i}, y) & K(\mathbf{i}, \mathbf{j}) \end{bmatrix} d^p x,$$

and

$$(I + \lambda K) \frac{1}{\lambda} \left(\frac{\partial \Delta}{\partial K} \right)^\dagger = \Delta(\lambda) \times [\text{the identity}].$$

I leave it to you to check this line by line from Section 1.7.

The last display shows that $I + \lambda K$ is invertible if $\Delta(\lambda) \ne 0$: Then $(I + \lambda K)(I + G/\Delta) = I$ with $\Delta + G = (\partial \Delta / \partial K)^\dagger / \lambda$, and the same applies to $I + \lambda K^\dagger$ so that

$$\left(I + \frac{G}{\Delta} \right) (I + \lambda K) = \left[(I + \lambda K^\dagger) \left(I + \frac{G^\dagger}{\Delta} \right) \right]^\dagger = I,$$

as it should be. But what if $\Delta(\lambda) = 0$? Then $I + \lambda K$ kills $G(\cdot, y)$ for every $0 \le y \le 1$ and it is desired to conclude that $1 + \lambda K$ is not invertible, i.e. $N \ne 0$. This is an automatic if $G(x, y)$ does not vanish identically, and even if it did, there is an easy way out: For large d,

$$\det \left[I + \lambda(d) K \left(\frac{i}{d}, \frac{j}{d} \right) \frac{1}{d} \right]_{1 \le i, j \le d}$$

vanishes for some $\lambda(d)$ close to λ, by Hurwitz's theorem (Ahlfors [1, p. 178]), so there must be an associated null vector $e(k/d)$, $1 \le k \le d$, which may be adjusted so that $\sum |e(k/d)|^2 / d = 1$. Now, upon making $d \uparrow \infty$, the sort of compactness employed in Section 2.3, item 4 produces, from

$$e \left(\frac{i}{d} \right) + \lambda(d) \sum_{j=1}^{d} K \left(\frac{i}{d}, \frac{j}{d} \right) \frac{1}{d} e \left(\frac{j}{d} \right) \equiv 0,$$

a genuine null function $e(x)$, $0 \le x \le 1$, of $I + \lambda K$ with $\int_0^1 e^2 = 1$, and all is well.

Here Fredholm's strategy comes into focus: If λ is small, the series for $\Delta + G$ reduces to $\Delta \times [\text{the standard Neumann series for } (I + \lambda K)^{-1}]$, as in

$$\Delta + G = \sum_{p=0}^{\infty} \lambda^p \sum_{q=0}^{p} \Delta_{p-q} (-K)^q = \sum_{p=0}^{\infty} \Delta_p \lambda^p \sum_{q=0}^{\infty} (-\lambda K)^q = \Delta (I + \lambda K)^{-1}.$$

But, at best, the sum $\sum(-\lambda K)^q$ represents the inverse only out to the first root of $\Delta(\lambda) = 0$ where it fails to exist. Contrariwise, Fredholm's recipe, expressing the inverse as a "rational" function $(\Delta + G)/\Delta$, works in the large, provided only that you avoid roots of Δ. The trick is that the coefficients $G_p = \sum_{q=0}^{p} \Delta_{p-q}(-K)^q$ of the series for $\Delta + G$ may be expressed by determinants of size $[(p/2)!]^{-1}$, more or less, providing a technical control not at all apparent at the start. That is the key to Fredholm's success.

Aside. The notation $\partial\Delta/\partial K$ is not fanciful. Recall that for smooth functions $F : \mathbb{R}^d \to \mathbb{R}$, you have

$$F(x + hy) = F(x) + h \times \big[\text{the inner product } \partial F/\partial x \bullet y\big] + o(h),$$

which you may take as the definition of grad $F = \partial F/\partial x$. It is the same in dimension $d = \infty$: A reasonable function $F : C[0,1] \to \mathbb{R}$ ought to satisfy

$$F(f + hg) = F(f) + h \times \big[\text{the inner product of something with } g\big] + o(h),$$

and you declare the "something" to be grad $F = \partial F/\partial f(x)$, as in

$$F(f + hg) = F(f) + h \int_0^1 \frac{\partial F}{\partial f} \, g \, dx + o(h).$$

The idea extends in an obvious way to Δ which is a function on the space $C[0,1]^2$ where the kernel $K(x, y)$ lives. You may like to check that with this understanding, the formula for $\partial\Delta/\partial K$ is quite correct.

Some examples.

1. K is convolution by x, i.e. $Kf = x * f \equiv \int_0^x (x - y) f(y) \, dy$ with kernel $K(x, y) = (x - y)^+$ vanishing for $x \leq y$. Then $\Delta \equiv 1$ and $\Delta + G = \sum \lambda^p \sum_{q=0}^{p} \Delta_{p-q}(-K)^q$ reduces to the Neumann series $\sum(-\lambda K)^q$. This must be OK as it stands: in fact, from $x * \cdots * x$ p-fold $= x^{2p-1}/(2p - 1)!$, it develops that $(I + \lambda K)^{-1}$ is the convolution operator $I - \sqrt{\lambda} \sin \sqrt{\lambda} x \, *$, as you will readily check.

2. K is now Volterra's operator $f \mapsto 1 * f(x) = \int_0^x f(y) \, dy$. This is outside the present rules since $K(x, y)$ jumps down from 1 to 0 at the diagonal, but never mind: just take $K(x, x) = 0$ and hope for the best. Then $\Delta \equiv 1$, and

$$\int_{x_1 < x_2 < \ldots < x_p} \det \begin{bmatrix} K(x, y) & K(x, \mathbf{j}) \\ K(\mathbf{i}, y) & K(\mathbf{i}, \mathbf{j}) \end{bmatrix} d^p x = \frac{(-1)^p}{p!}(x - y)^p \text{ or } 0 \qquad \text{according as } x > y \text{ or not.}$$

For example, if $p = 2$ and $x_1 < x_2$, the determinant vanishes except on the figure $x > x_2 > x_1 > y$ where it reduces to

$$\det \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} = 1,$$

and the integral is just the volume of that figure: $(x - y)^2/2!$. Try it for $p = 3$. The outcome is

$$\sum_0^{\infty} \lambda^p \int_{x_1 < \ldots < x_p} \det \begin{bmatrix} K(x, y) & K(x, \mathbf{j}) \\ K(\mathbf{i}, y) & K(\mathbf{i}, \mathbf{j}) \end{bmatrix} = e^{-\lambda(x-y)} \text{ or } 0 \qquad \text{according as } x > y \text{ or not,}$$

i.e. $(I + \lambda K)^{-1} = I - \lambda e^{-\lambda x} *$. Check this by hand.

3. It is amusing to observe that if K is the even more singular operator

$$f \mapsto \int_0^x \frac{f(y)}{\sqrt{x-y}}\, dy$$

of Abel, then $(I + \lambda K)^{-1}$ may be found by a trick due to himself: $K^2 f = \pi \int_0^x f$, as you will check, so from $(I + \lambda K) f = g$ you find $Kf + \lambda \pi \int_0^x f = Kg$, i.e. $f - \lambda^2 \pi \int_0^x f = (I - \lambda K)\, g$, to which the inversion $I + \lambda^2 \pi \exp(\lambda^2 \pi x)*$ may be applied to recover f. Check it out. What is Δ now?

4. $K(x, y) = K(y, x) = x\,(1 - y)$ if $x \le y$, as in the Example of Section 2.4. I leave it as an exercise to verify the formula

$$(I + \lambda K)^{-1} f = f - \Delta^{-1} \times \left\{ \text{sh } \sqrt{\lambda}\,(1 - x) \int_0^x \text{sh } \sqrt{\lambda} y\, f(y)\, dy + \text{sh } \sqrt{\lambda} x \int_x^1 \text{sh } \sqrt{\lambda}\,(1 - y)\, f(y)\, dy \right\},$$

in which Δ is the determinant $\det(I + \lambda K) = (\text{sh } \sqrt{\lambda})/\sqrt{\lambda}$ computed in Section 2.4. Here, Fredholm's integrals are formidable: even for $p = 3$ you must evaluate

$$\int_{x_1 < x_2 < x_3} \det \begin{bmatrix} K(x, y) & K(x, x_1) & K(x, x_2) & K(x, x_3) \\ K(x_1, y) & x_1(1 - x_1) & x_1(1 - x_2) & x_1(1 - x_3) \\ K(x_2, y) & x_1(1 - x_2) & x_2(1 - x_2) & x_2(1 - x_3) \\ K(x_3, y) & x_1(1 - x_3) & x_2(1 - x_3) & x_3(1 - x_3) \end{bmatrix} dx_1\, dx_2\, dx_3$$

and the integral has to be split into 16 pieces according to the placement of x and y among the 5 points $0 < x_1 < x_2 < x_3 < 1$! Let's try $p = 1$ only. Then, with $x < y$,

$$\int_0^1 \det \begin{bmatrix} x(1 - y) & K(x, z) \\ K(z, y) & z(1 - z) \end{bmatrix} dz = \frac{x(1 - y)}{6} \int_0^x z(1 - x)z(1 - y)\, dz = \frac{x^3}{3}\,(1 - x)(1 - y)$$

$$- \int_x^y x(1 - z)z(1 - y)\, dz = x(1 - y)\left(\frac{y^2}{2} - \frac{y^3}{3} - \frac{x^2}{2} + \frac{x^3}{3} \right)$$

$$- \int_y^1 x(1 - z)y(1 - z)\, dz = xy\, \frac{(1 - y)^3}{3}$$

$$= \frac{1}{6} x^3(1 - y) + \frac{1}{6} x(1 - y)^3,$$

which is, indeed, the coefficient of λ^2 in $\text{sh } \sqrt{\lambda} x\, \text{sh } \sqrt{\lambda}\,(1 - y)$, God be thanked!

2.6. A question of multiplicities

It is natural to ask what if any is the relation between n the dimension of the null space N of $I + \lambda K$ and m the multiplicity of the root of $\Delta(\lambda) = 0$. So far, it is known only that $n = 0$ if and only if $m = 0$. The general rule is $n \le m$ with $n = 0$ only if $m = 0$. Skip this at a first reading if you like.

Example.

To illustrate the possibility of $n < m$, consider $K = f_1 \otimes e_1 + f_2 \otimes e_2$. Taking

$$G = \left[\int_0^1 f_i e_j \right]_{1 \le i, j \le 2} = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

the question is quickly reduced to the dimension n of the null space of $I + \lambda G$ vs. the multiplicity m of the root, if any, of $\det(1 + \lambda G)$. Take $a = c = d = 1$ and $b = 0$. Then $\Delta = (1 + \lambda)^2$ has a double root at -1 but

$$I - G = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}$$

has only the one null vector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Proof of $n \leq m$ (adapted from Kato [13, p. 40]). If $n = 0$, there is nothing to prove, so take $1 \leq n$ and let Δ vanish at $\lambda = 1$, say, with multiplicity $m \geq 1$. Obviously $\Delta(\lambda) \neq 0$ for nearby $\lambda \neq 1$. $(I + \lambda K)$ is then invertible, and you can form the operator

$$P = -\frac{1}{2\pi\sqrt{-1}} \cdot \oint (I + \lambda K)^{-1} \frac{d\lambda}{\lambda},$$
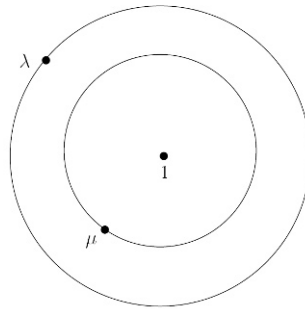
the integral being taken about a small circle enclosing $\lambda = 1$.

Item 1.

P is a projection, i.e. $P^2 = P$. To confirm this, express P^2 as a double integral:

$$P^2 = \frac{1}{2\pi\sqrt{-1}} \cdot \oint \frac{d\lambda}{\lambda} \frac{1}{2\pi\sqrt{-1}} \cdot \oint \frac{d\mu}{\mu}(I + \lambda K)^{-1}(I + \mu K)^{-1}$$

$$= \frac{1}{2\pi\sqrt{-1}} \cdot \oint d\lambda \frac{1}{2\pi\sqrt{-1}} \cdot \oint d\mu \frac{1}{\lambda - \mu} \left(\frac{\lambda}{I + \lambda K} - \frac{\mu}{I + \mu K} \right) \frac{1}{\lambda\mu},$$

where the first/second integral is taken about the outer/inner circle seen in the picture.

Do the inner integral first to wash out the contribution of $(I + \lambda K)^{-1}$ (λ is outside). Then do the outer integral of what remains to produce

$$-\frac{1}{2\pi\sqrt{-1}} \cdot \oint (I + \mu K)^{-1} \frac{d\mu}{\mu} = P.$$

Item 2.

P acts as the identity on the null space N of $I + K$: indeed, $Ke = -e$ for $e \in N$, so

$$Pe = -\frac{1}{2\pi\sqrt{-1}} \oint (1 - \lambda)^{-1} \frac{d\lambda}{\lambda} e = e,$$

as advertised.

Item 3.

$P = -\left(2\pi\sqrt{-1}\right)^{-1} \oint (G/\Delta) \, d\lambda/\lambda$ inherits from G a nice kernel of class $C[0,1]^2$ and so is compact, meaning that the image PB of the unit null $B = \left\{ f : \int_0^1 f^2 \leq 1 \right\}$ is compact in $L^2[0,1]$; as such, its range Q is of some finite dimension d. (*Why?*) Now P commutes with K, so you may write $K = PKP + (I-P)K(I-P)$ and infer that

$$\Delta(\lambda) = \det(I + \lambda K) = \det(I + \lambda PKP) \times \det\left(I + \lambda(I-P)K(I-P)\right).$$

Here, the second factor cannot vanish at $\lambda = 1$ since $\left(I + (I-P)K(I-P)\right) f = 0$ implies $Pf = 0$ and so also $(I+K)f = 0$, i.e. $f \in N$, and $f = Pf = 0$, by Item 2.

Item 4.

$(I+K)^m P = 0$, m being the multiplicity of the root of Δ: indeed,

$$(I+K)P = -\frac{1}{2\pi\sqrt{-1}} \cdot \oint \left(\frac{I+K}{I+\lambda K} - I \right) \frac{d\lambda}{\lambda} = -\frac{1}{2\pi\sqrt{-1}} \cdot \oint \frac{(1-\lambda)K}{I+\lambda K} \frac{d\lambda}{\lambda}$$

$$(I+K)^2 P = -\frac{1}{2\pi\sqrt{-1}} \cdot \oint (1-\lambda) K \left(\frac{I+K}{I+\lambda K} - I \right) \frac{d\lambda}{\lambda} = -\frac{1}{2\pi\sqrt{-1}} \cdot \oint \frac{(1-\lambda)^2 K^2}{I+\lambda K} \frac{d\lambda}{\lambda}$$

and so forth, up to

$$(I+K)^m P = -\frac{1}{2\pi\sqrt{-1}} \cdot \oint \frac{(1-\lambda)^m K^m}{I+\lambda K} \frac{d\lambda}{\lambda} = -\frac{1}{2\pi\sqrt{-1}} \cdot \oint (1-\lambda)^m K^m \frac{G}{\Delta} \frac{d\lambda}{\lambda}.$$

This vanishes as the integrand has no pole anymore.

Item 5.

Put $(I+K)P = J$ and keep it in mind that P commutes with K so that $J^m = 0$, by Item 4. Then

$$I + \lambda PKP = (1-\lambda)(I + \Lambda J) \qquad \text{with} \qquad \Lambda = \lambda(1-\lambda)^{-1}$$

is a $d \times d$ matrix, d being the dimension of Q the range of P. It follows that the root–bearing factor $\det(I + \lambda PKP)$ of $\Delta(\lambda)$ from Item 3 can be expressed as

$$\det(1-\lambda)(I + \Lambda J) = (1-\lambda)^d \det(I + \Lambda J) \qquad \text{for} \quad \lambda \neq 1$$

But $\det(I + \Lambda J) = 1$ being a polynomial in Λ without roots: in fact, a root entails a null function, and $(I + \Lambda J) f = 0$ only if $f = -\Lambda J f = (-\Lambda)^m J^m f = 0$. This means that $\Delta = (1-\lambda)^d \times$ [a non-vanishing factor], so $d = m$, and since $P = I$ on N, by Item 2, so $Q \supset N$ and $m = \dim Q \geq n$, as wanted. $\qquad \square$

2.7. Symmetric K

This is a very special case, but it's important for all sorts of applications and so worthy to be spelled out. For simplicity and because it is all I need, I take K non–negative in the sense that the quadratic form $Q[f] = \int_0^1 fKf$ is always ≥ 0.

Existence of eigenfunctions.

For $d < \infty$, the eigenvalue problem $Ke = \lambda e$ has just enough solutions to span the range of K and it is the same here. In keeping with the tone of this chapter, I will obtain the existence of enough eigenfunctions by the simplest means. The quadratic form vanishes identically only if $K \equiv 0$ (*Why?*). Otherwise, it is capable of positive values, as I now suppose. Q is controlled by $\int_0^1 f^2 \times \max\{|K(x,y)| : 0 \le x, y \le 1\}$, so $\sup\{Q[f] : \|f\|_2 \le 1\} = \lambda$ is finite and positive. Replace K by the $d \times d$ matrix $[K(i/d, j/d)/d]_{1 \le i,j \le d}$. The top value of *its* quadratic form in the ball of vectors $e(k/d)$, $1 \le k \le d$, with $\sum_{k=1}^d |e(k/d)|^2/d \le 1$ is its top eigenvalue λ_d, which is close to λ if d is large, as I invite you to check; moreover, you can make the associated eigenvector e_d approximate a function $e \in C[0,1]$ by choice $d = d_1 < d_2 < \ldots \uparrow \infty$, by the now familiar type of compactness employed in Section 2.3. Then for $d = \infty$, you have $Ke = \lambda e$, i.e. e is an eigenfunction of K with eigenvalue λ and unit length: $\int_0^1 e^2 = 1$. Now write $\lambda = \lambda_1$, $e = e_1$, and look at the reduced operator $K - \lambda_1 e_1 \otimes e_1$ acting on the annihilator of e_1. It may be that *its* quadratic form is incapable of positive values. Then $Q[f] = \lambda_1(e_1, f)^2$, $K = \lambda_1 e_1 \otimes e_1$, and you stop. Otherwise, Q is capable of positive values on the annihilator of e_1, in which case you can repeat the procedure to produce a new unit eigenfunction e_2, perpendicular to e_1, with eigenvalue $0 < \lambda_2 \le \lambda_1$, and so on. If the quadratic form of $K - \sum_{n=1}^d \lambda_n e_n \otimes e_n$ vanishes on the annihilator of e_n, $n \le d$, the procedure stops at stage d, $K \equiv \sum_1^d \lambda_n e_n \otimes e_n$, and "enough" eigenfunctions are in hand. Otherwise, it never stops, producing an infinite number of mutually perpendicular eigenfunctions e_n, $n \ge 1$, with diminishing eigenvalues $\lambda_1 \ge \lambda_2 \ge \ldots > 0$.

Mercer's Theorem.

Mercer's Theorem states that $K(x,y) = \sum_{n=1}^\infty \lambda_n e_n \otimes e_n$ with uniform convergence of the sum on the square $[0,1]^2$; in particular, $\sum_{n=1}^\infty e_n(e_n, f)$ converges to f in $C[0,1]$ for every function f in the range of K: that is what "enough" eigenfunctions means. The case of finite rank when the procedure stops is left aside.

Proof. Fix $d < \infty$. The reduced operator $K - \sum_{n=1}^d \lambda_n e_n \otimes e_n$ is non-negative, i.e.

$$\int_0^1 \int_0^1 f(x) \left[K(x,y) - \sum_1^d \lambda_n e_n \otimes e_n \right] f(y) \, dx \, dy \ge 0,$$

from which you learn the following.

1) $\sum_{n=1}^d \lambda_n e_n^2(x) \le K(x,x)$ by making f approximate $h^{-1} \times$ [the indicator of $|x - y| \le h/2$] and taking h down to $0+$.

2) $\sum_{n=1}^\infty \lambda_n \le \int K(x,x) dx < \infty$, by integration.

3) $\sum_{n=1}^\infty \lambda_n e_n \otimes e_n$ is now seen to converge in the whole square and, indeed, uniformly in x/y for fixed y/x in view of 1) and the tail estimate

$$\left| \sum_{n \ge d} \lambda_n e_n(x) e_n(y) \right|^2 \le \sum_{n \ge d} \lambda_n e_n^2(x) \times \sum_{n \ge d} \lambda_n e_n^2(y) \le \quad \text{e.g.} \quad K(x,x) \sum_{n \ge d} \lambda_n e_n^2(y).$$

4) $\sum \lambda_n e_n \otimes e_n \, f = \sum \lambda_n e_n(e_n, f)$ is likewise uniformly convergent by the tail estimate:

$$\left| \sum_{n \ge d} \lambda_n e_n(e_n, f) \right|^2 \le \sum_{n \ge d} \lambda_n^2 e_n^2 \sum_{n \ge d} (e_n, f)^2 \le \lambda_d K(x,x) \int_0^1 f^2 \, dx$$

and the fact that $\lambda_d \downarrow 0$ by 2); alternatively, 1) and the decrease of $\sum_{n \ge d} (e_n, f)^2$ to 0 do the trick; see Dini's theorem in the appendix.

5) Now the continuous function $f_0 = Kf - \sum_{n=1}^{\infty} \lambda_n e_n(e_n, f)$ is perpendicular to e_n, $n \geq 1$, and since $Q[f_0] \leq \lambda_d |f_0|_2^2 \downarrow 0$ as $d \uparrow \infty$ by 2), so $Q[f_0] = 0$. Therefore, $Q[f_0 + hf]$ is minimum (0) at $h = 0$, and

$$\frac{d}{dh} Q[f_0 + hf] = 2 \int f_0 Kf = 2 \int_0^1 f_0^2 = 0,$$

i.e. $f_0 \equiv 0$. It follows that

$$Kf(x) = \int_0^1 \sum_{n=1}^{\infty} \lambda_n e_n(x) e_n(y) f(y) \, dy,$$

for every $f \in C[0, 1]$ and every $0 \leq x \leq 1$, whence $K(x, y) = \sum_{n=1}^{\infty} \lambda_n e_n(x) e_n(y)$ at all points $0 \leq x, y \leq 1$, by 3).

6) Take $y = x$. Then $\sum_{n \leq d} \lambda_n e_n^2(x)$ increases with d to the continuous function $K(x, x)$. Dini's theorem now tells you that this convergence is uniform. Therefore, the convergence of the sum $\sum \lambda_n e_n \otimes e_n$ for $K(x, y)$ must be uniform, too. Check it out.

Item 6) plus 5) is the content of Mercer's theorem. □

Expressing $\triangle$ as a product.

Because $K = \sum \lambda_n e_n \otimes e_n$, with uniform convergence in the square,

$$\det(I + \lambda K) = \lim_{d \uparrow \infty} \det \prod_{n=1}^{d} \left[I + \lambda \left(\lambda_n e_n \otimes e_n \right) \right] = \lim_{d \uparrow \infty} \prod_{n=1}^{d} \det \left[I + \lambda \left(\lambda_n e_n \otimes e_n \right) \right] = \prod_{n=1}^{\infty} (1 + \lambda \lambda_n)$$

since all the little $p \times p$ determinants entering into the series for $\det(I + \lambda e \otimes e)$ vanish for $p \geq 2$. Here, there is no question as to the convergence of the product in view of 2). Besides $|\Delta(\lambda)| \leq \exp\left(r \sum \lambda_n \right)$ for $r = |\lambda|$, so Δ is of order 1, only.

Example.

I go back to the symmetric kernel of Section 2.4: $K(x, y) = x(1 - y)$ if $x \leq y$. This is positive, being the inverse of $-D^2$ acting on the domain $C^2[0, 1] \cap \{f : f(0) = 0 = f(1)\}$, with eigenfunctions $e_n(x) = \sqrt{2} \sin \pi n x$ and eigenvalues $1/n^2 \pi^2$. But you know the determinant of $1 + \lambda K$ from Section 2.4 already, and its present evaluation may be looked upon as a proof of the standard product

$$\frac{\sin \lambda}{\lambda} = \prod_{n=1}^{\infty} \left(1 - \frac{\lambda^2}{n^2 \pi^2} \right) \qquad \text{in the present form} \qquad \frac{\text{sh} \sqrt{\lambda}}{\sqrt{\lambda}} = \prod_{n=1}^{\infty} \left(1 + \frac{\lambda}{n^2 \pi^2} \right).$$

3. Fredholm applied

I collect here three applications of Fredholm determinants to: 1) Brownian motion, 2) long surface waves in shallow water, and 3) unitary matrices in high dimension. Quite a variety, no?

3.1. Brownian motion: preliminaries

The Brownian motion describes the path of, e.g. a (visible) dust note in a sun beam, impelled unpredictably, this way and that, by collisions with the (invisible) molecules of the air. Brown [2] studied it first, in Nature; later, Einstein [9] related it to diffusion and the underlying heat equation $\partial p / \partial t = \frac{1}{2} \partial^2 p / \partial x^2$.

The mathematical set-up required can be reduced to the unit interval, equipped with its standard Lebesgue measure. I suppose you know something about that. Only the language changes: a measurable set $A \subset [0,1]$ is now an "event"; the measure of A is its "probability" $P(A)$; a measurable function $\mathfrak{x} : [0,1] \to \mathbb{R}$ is a "random variable"; the integral $\int_0^1 \mathfrak{x} \equiv E(\mathfrak{x})$ is its "mean-value" or "expectation".

With this lingo in place, a collection of random variables $\mathfrak{x}$ is said to be a Gaussian family if any combination $k \bullet \mathfrak{x} = k_1 \mathfrak{x}_1 + \ldots + k_n \mathfrak{x}_n$ with $(k_1, \ldots, k_n) \in \mathbb{R}^n - 0$ is Gauss-distributed with mean 0 and positive root-mean-square σ, i.e.

$$P(a \leq k \bullet \mathfrak{x} < b) = \int_a^b \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \, dx \qquad \text{for any} \quad a < b.$$

Here, σ^2 is a quadratic form in k:

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \, dx = E(k \bullet \mathfrak{x})^2 = \sum_{1 \leq i,j \leq n} k_i k_j E(\mathfrak{x}_i \mathfrak{x}_j) \equiv k \bullet Qk,$$

Q being the (symmetric, positive-definite) matrix of "correlations" $Q_{ij} = [E(\mathfrak{x}_i \mathfrak{x}_j)]_{1 \leq i,j \leq n}$. Now

$$E\left(e^{\sqrt{-1}k \bullet \mathfrak{x}}\right) = \int_{-\infty}^{+\infty} e^{\sqrt{-1}k \bullet x} \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \, dx = e^{-\sigma^2/2} = e^{-k \bullet Qk/2}$$

is the Fourier transform of the joint distribution of $\mathfrak{x}_1, \ldots, \mathfrak{x}_n$. This may be inverted to obtain their joint probability density as function of $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$:

$$p(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-\sqrt{-1}x \bullet k} e^{-k \bullet Qk/2} \, d^n k = \frac{e^{-x \bullet Q^{-1} x/2}}{(2\pi)^{n/2}\sqrt{\det Q}}.$$

You see how the determinant enters in and can imagine where all this could lead. I spell out the computation: Q can be brought to principal axes by a rotation $k \to k'$ of $\mathbb{R}^n$. This does not change the volume element $d^n k$ but simplifies the quadratic form as in $k \bullet Qk = \sum_{i=1}^n \lambda_i (k_i')^2$, the λ's being the (positive) eigenvalues of Q. Then, with the same rotation $x \to x'$,

$$p(x) = \frac{1}{(2\pi)^n} \int e^{-\sqrt{-1}x' \bullet k'} e^{-\sum \lambda_i (k_i')^2/2} \, d^n k' = \frac{1}{(2\pi)^n} \int e^{-\sqrt{-1}x' \bullet k} e^{-\sum \lambda_i (k_i)^2/2} \, d^n k.$$

The joint density is now seen to be the product of

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{\sqrt{-1}x_i' k} e^{-\lambda_i k^2/2} \, dk$$

taken over $1 \leq i \leq n$, i.e.

$$p(x) = \prod_{i=1}^n \frac{e^{-(x_i')^2/2\lambda_i}}{\sqrt{2\pi\lambda_i}} = \frac{e^{-x \bullet Q^{-1} x/2}}{(2\pi)^{n/2}\sqrt{\det Q}}$$

as advertised. The formula itself is nice, and you learn from it two important things: 1) that the statistics of a Gaussian family are completely determined by its correlations, and 2) that two such families are statistically independent if and only if they are uncorrelated, one with the other: indeed, if $\mathfrak{x}_1, \ldots, \mathfrak{x}_p$ and $\mathfrak{x}_{p+1}, \ldots, \mathfrak{x}_{p+q}$ are uncorrelated, then their joint correlation Q breaks up into blocks as in the figure,

$p \times p$	0
0	$q \times q$

Q^{-1} does likewise, and the joint density splits, which is what independence is all about: if A is any event concerning $\mathfrak{x}_i$, $i \leq p$, and B any event concerning the other $\mathfrak{x}_j$, $j > p$, then $P(A \cap B) = P(A) \times P(B)$.

A little trick will help to convince you that the simple space $[0, 1]$, equipped with its Lebesgue measure, really is rich enough to support such families. Let $x \in [0, 1]$ be expanded in binary notation: $x = .x_1 x_2 \ldots$ The variables x_n, $n \geq 1$, are independent (*check this!*) with common distribution $P(x_n = 0) = P(x_n = 1) = 1/2$.

Define

$$\mathfrak{x}_1 = . x_1 x_3 x_6 x_{10} \ldots ,$$
$$\mathfrak{x}_2 = . x_2 x_5 x_9 \ldots ,$$
$$\mathfrak{x}_3 = . x_4 x_8 \ldots ,$$
$$\mathfrak{x}_4 = . x_7 \ldots ,$$

and so on.

You see the pattern. These new variables are independent, with common distribution $P(\mathfrak{x} \leq x) = x$, $0 \leq x \leq 1$. Then the modified variables $\mathfrak{x}'_n$ defined by $\mathfrak{x}_n = \int_{-\infty}^{\mathfrak{x}'_n} (2\pi)^{-1/2} e^{-x^2/2} dx$ are also independent with common probability density $(2\pi)^{-1/2} e^{-x^2/2}$. Now just one more step: if Q is any $n \times n$, symmetric, positive matrix and $\sqrt{Q}$ its symmetric positive root, then the family $\mathfrak{x}''_i = \sum_{j=1}^{n} (\sqrt{Q})_{ij} \mathfrak{x}'_j$, $1 \leq i \leq n$, is Gaussian with correlation Q.

The generalities now are ended and I turn to the (standard, 1–dimensional) Brownian motion itself. This is the Gaussian family $\mathfrak{x}(t)$, $t \geq 0$, with correlation $E[\mathfrak{x}(t_1)\mathfrak{x}(t_2)] =$ the smaller of t_1 and t_2.

It is immediate that 1) $P[\mathfrak{x}(0) = 0] = 1$; 2) for $t_1 \leq t_2$, the increment $\mathfrak{x}(t_2) - \mathfrak{x}(t_1)$ has mean–square $t_2 - t_1$; 3) increments taken over disjoint intervals of time are uncorrelated and so also statistically independent. It follows that for $0 < t_1 < t_2 < \ldots < t_n$, the joint probability density of the "observations" $\mathfrak{x}(t_1), \mathfrak{x}(t_2), \ldots, \mathfrak{x}(t_n)$ is

$$p(x) = \frac{e^{-x_1^2/2t_1}}{\sqrt{2\pi t_1}} \frac{e^{-(x_2-x_1)^2/2(t_2-t_1)}}{\sqrt{2\pi(t_2-t_1)}} \cdots \frac{e^{-(x_n-x_{n-1})^2/2(t_n-t_{n-1})}}{\sqrt{2\pi(t_n - t_{n-1})}} .$$

Wiener [26] proved that the Brownian motion has continuous paths. I explain what he meant, technically. The present machinery cannot deal directly with the event "$\mathfrak{x}(\cdot)$ is continuous" since it involves an uncountable number of observations of the Brownian path $t \mapsto \mathfrak{x}(t)$. What is needed is something like this: with probability 1, $|\mathfrak{x}(i/n) - \mathfrak{x}(j/n)| \leq \sqrt[3]{i/n - j/n}$ for all $0 \leq j/n < i/n \leq 1$ with sufficiently small separation $\max |i/n - j/n| \leq h$, independently of $n \geq 1$. This permits you to complete the *countable* Gaussian family $\mathfrak{x}(k/n)$, $k \leq n$, $n \geq 1$, to the *uncountable* family $\mathfrak{x}(t) = \lim_{k/n \downarrow t} \mathfrak{x}(k/n)$, $0 \leq t \leq 1$, with the correct joint distributions for any finite number of observations and guaranteed continuity of paths. The proof takes some work; see McKean [18] for Ciesielski's slick version.

It is natural to write the measure induced from $[0, 1]$ onto path–space, symbolically:

$$dP = \frac{e^{-\int_0^\infty [\mathfrak{x}^\bullet(t)]^2/2\, dt}}{(2\pi 0+)^{\infty/2}} \, d^\infty \mathfrak{x}.$$

Here nothing is kosher: with probability 1, $\mathfrak{x}^\bullet(t)$ does not exist for *any* $t \geq 0$; $(2\pi 0+)^{\infty/2}$ is silly; and the flat volume element $d^\infty \mathfrak{x}$ is a bogus, too. Nevertheless, the whole reminds you of what is happening and is even a reliable guide to computation.

A variant of this "free" Brownian motion figures below. This is the "tied" Brownian motion, alias the free motion, considered for times $t \leq 1$ and conditioned so that $\mathfrak{x}(1) = 0$. The conditioning does not change the Gaussian character of the family, so only the correlation needs to be found.

Here is a neat way to do that. Look at the picture.
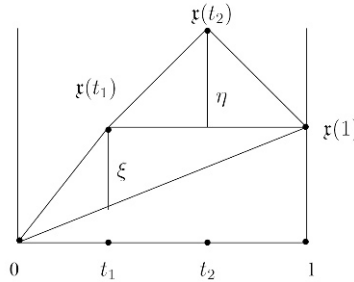
The variables
$$\xi = \mathfrak{x}(t_1) - t_1 \mathfrak{x}(1), \qquad \eta = \mathfrak{x}(t_2) - \mathfrak{x}(t_1) - \frac{\mathfrak{x}(1) - \mathfrak{x}(t_1)}{1 - t_1} \times (t_2 - t_1), \qquad \text{and} \qquad \mathfrak{x}(1)$$

are uncorrelated, as you will check, and so independent. Now condition by $\mathfrak{x}(1) = 0$. The independence of ξ and η is not spoiled, whence the (conditional) correlation between $\mathfrak{x}(t_1) = \xi$ and $\mathfrak{x}(t_2) = \xi(1 - t_2)/(1 - t_1) + \eta$ is just the unconditional correlation $E(\xi^2)(1 - t_2)/(1 - t_1) = t_1(1 - t_2)$.

A nice model of the tied Brownian motion is obtained from the free Brownian motion in the form $\mathfrak{x}(t) - t\mathfrak{x}(1)$, $0 \leq t \leq 1$. You have only to check (in one line) that the correlations are what they should be.

3.2. Brownian motion: some applications

Simplest examples.

I want to compute the mean-value of $F = \exp\left[-(\lambda/2) \int_0^1 \mathfrak{x}^2(t)\, dt \right]$ for the free Brownian motion. The task is simplified by the continuity of the Brownian path, permitting the reduction of F to $F_n = \exp\left[-(\lambda/2) \sum_{k=1}^n \mathfrak{x}^2(k/n)/n \right]$ with a subsequent passage to the limit $n \uparrow \infty$. Write the joint density for $\mathfrak{x}(k/n)$, $1 \leq k \leq n$, as

$$\frac{\exp\left(-x \bullet Q^{-1}x/2 \right)}{(2\pi)^{n/2}\sqrt{\det Q}} \qquad \text{with} \qquad Q_{ij} = \min\left(\frac{i}{n}, \frac{j}{n} \right), \quad 1 \leq i, j \leq n.$$

Then

$$E(F_n) = \int_{\mathbb{R}^n} \frac{\exp\left(-x \bullet ((\lambda/n)I + Q^{-1})x/2 \right)}{(2\pi)^{n/2}\sqrt{\det Q}}\, d^n x = \sqrt{\frac{\det\left((\lambda/n)I + Q^{-1} \right)^{-1}}{\det Q}} = \left[\det(I + \lambda Q/n) \right]^{-1/2}$$

so that

$$E \exp\left(-\frac{\lambda}{2} \int_0^1 \mathfrak{x}^2\, dt \right) = \frac{1}{\sqrt{\det(I + \lambda K)}}$$

with Fredholm's determinant and the (symmetric) operator

$$K : f \mapsto \int_0^1 \min(x, y) f(y)\, dy.$$

The determinant is easy to compute: K is the (symmetric) inverse to $-D^2$ acting on $C^2[0,1] \cap \{f : f(0) = f'(1) = 0\}$, with eigenfunctions $\sqrt{2}\sin\pi(n+1/2)x$ and eigenvalues $[\pi(n+1/2)]^{-2}$, $n \geq 0$, so

$$\det(I + \lambda K) = \prod_{n=0}^{\infty}\left[1 + \frac{\lambda}{\pi^2(n+1/2)^2}\right] = \mathrm{ch}\sqrt{\lambda}$$

and

$$E\exp\left(-\frac{\lambda}{2}\int_0^1 \mathfrak{x}^2\,dt\right) = \left(\mathrm{ch}\sqrt{\lambda}\right)^{-1/2}.$$

The computation for the tied Brownian motion is much the same: Now $Q_{ij} = (i/n)(i - j/n)$ for $i \leq j$; $K(x,y)$ is the symmetric kernel $x(1-y)$ if $x \leq y$; the associated operator inverts $-D^2$ acting on $C^2[0,1] \cap \{f : f(0) = f(1) = 0\}$; the eigenvalues are $(\pi n)^{-2}$, $n \geq 1$; and $\det(I + \lambda K) = \lambda^{-1/2}\mathrm{sh}\sqrt{\lambda}$, as in Section 2.7, so now

$$E\exp\left(-\frac{\lambda}{2}\int_0^1 \mathfrak{x}^2\,dt\right) = \left(\sqrt{\lambda}\,\mathrm{sh}\sqrt{\lambda}\right)^{-1/2}.$$

P. Lévy's area.

Let $\mathfrak{x} : t \in [0,1] \to \mathbb{R}^2$ be the plane motion of two independent copies $\mathfrak{x}_1$ & $\mathfrak{x}_2$ of the tied Brownian motion and consider P. Lévy's area [17, p. 254]

$$A = \frac{1}{2}\int_0^1 (\mathfrak{x}_1 d\mathfrak{x}_2 - \mathfrak{x}_2 d\mathfrak{x}_1) = \int_0^1 \mathfrak{x}_1 d\mathfrak{x}_2 = \lim_{n\uparrow\infty} A_n$$

with

$$A_n = \sum_{k=0}^{n-1} \mathfrak{x}_1\left(\frac{k}{n}\right)\left[\mathfrak{x}_2\left(\frac{k+1}{n}\right) - \mathfrak{x}_2\left(\frac{k}{n}\right)\right].$$

The Brownian path is too irregular for this to be any kind of classical integral, but the limit exists with probability 1 and is declared to be the area. I ask you to believe this and proceed to the evaluation of the Fourier transform $E\exp(\sqrt{-1}\lambda A)$ of the distribution of A. The individual Brownian motions $\mathfrak{x}_1$ and $\mathfrak{x}_2$ are independent, so their joint probabilities split and you may begin by fixing $\mathfrak{x}_1$ and integrating out $\mathfrak{x}_2$ with the help of the general rule

$$\int_{\mathbb{R}^n} e^{x\bullet y}\frac{e^{-x\bullet P^{-1}x/2}\,d^n x}{(2\pi)^{n/2}\sqrt{\det P}} = e^{y\bullet Py/2}.$$

Here, P is to be taken as the correlation of the increments of $\mathfrak{x}_2$ figuring in A_n, and $\mathfrak{y}$ is the vector $\mathfrak{x}_1(k/n)$, $0 \leq k < n$. You have $P_{ij} = -1/n^2$ plus an extra $1/n$ on diagonal, as you will check, so with $\mathfrak{x}_1$ fixed, you have

$$E\left[e^{\sqrt{-1}\lambda A_n}\,\middle|\,\mathfrak{x}_1(t) : 0 \leq t \leq 1\right] = \exp\left\{-\frac{\lambda^2}{2}\left[\sum_{k=0}^{n-1}\mathfrak{x}_1^2\left(\frac{k}{n}\right)\frac{1}{n} - \left(\sum_{k=0}^{n-1}\mathfrak{x}_1\left(\frac{k}{n}\right)\frac{1}{n}\right)^2\right]\right\} = \exp\left(-\frac{\lambda^2}{2}\mathfrak{y}\bullet P\mathfrak{y}\right).$$

It is convenient to shift the summations above from $0 \leq k < n$ to $0 \leq k \leq n$; this makes no difference to the limit. Then, with the correlations

$$Q_{ij} = \frac{i}{n}\left(1 - \frac{j}{n}\right), \quad i \leq j, \quad \text{for} \quad \mathfrak{x}_1\left(\frac{k}{n}\right), \quad 1 \leq k \leq n, \quad \text{alias } \mathfrak{y},$$

you find

$$E\left(e^{\sqrt{-1}\lambda A_n}\right) = \int e^{-\lambda^2 x \bullet Px/2} \frac{e^{-x\bullet Q^{-1}x/2}}{(2\pi)^{n/2}\sqrt{\det Q}}\, d^n x = \sqrt{\frac{\det(\lambda^2 P + Q^{-1})^{-1}}{\det Q}} = \frac{1}{\sqrt{\det(I + \lambda^2 QP)}},$$

in which $QP = Q/n - (Q1/n^2) \otimes 1$. This matrix approximates the operator with the familiar kernel $K(x, y) = x(1 - y)$ minus $e1 \otimes 1$ with $e = K1$, so $E \exp(\sqrt{-1}\lambda A)$ is 1 over the square root of Fredholm's determinant

$$\det\left[I + \lambda^2(K - e\otimes 1)\right] = \det\left(I + \lambda^2 K\right) \times \det\left(I - \left(I + \lambda^2 K\right)^{-1} e\otimes 1\right) = \lambda^{-1}\operatorname{sh}\lambda \times \left[1 - \operatorname{sp}\left(I + \lambda^2 K\right)^{-1} e\otimes 1\right].$$

It remains to evaluate the trace (sp) which is not hard: $K = \sum_{n=1}^{\infty} e_n \otimes e_n/n^2\pi^2$ with $e_n = \sqrt{2}\sin n\pi x$ for $n \geq 1$, so

$$\operatorname{sp}\left(I + \lambda^2 K\right)^{-1} e \otimes 1 = \int_0^1 1\left(I + \lambda^2 K\right)^{-1} e\, dx = \sum_{n=1}^{\infty} \frac{1}{\lambda^2 + n^2\pi^2}\left(\int_0^1 e_n\right)^2 = 8\sum_{\text{odd } n} \frac{1}{\lambda^2 + n^2\pi^2} \times \frac{1}{n^2\pi^2} = 1 - \frac{2}{\lambda}\frac{\operatorname{sh}\lambda/2}{\operatorname{ch}\lambda/2}.$$

The last expression comes from the product $\lambda^{-1}\operatorname{sh}\lambda = \prod_{n=1}^{\infty}\left(1 + \lambda^2/n^2\pi^2\right)$, by logarithmic differentiation and a bit of manipulation which I leave to you. Then

$$\det\left[I + \lambda^2(K - e\otimes 1)\right] \qquad \text{reduces to} \qquad \left[\frac{\operatorname{sh}(\lambda/2)}{\lambda/2}\right]^2$$

for the final evaluation:

$$E\left(e^{\sqrt{-1}\lambda A}\right) = \frac{\lambda/2}{\operatorname{sh}(\lambda/2)}.$$

Inverting.

This is all very well, but what is really wanted in these examples is the inverse transform, i.e. the probability density of the variable in hand. P. Lévy's area A is the simplest in this regard, the density being $(\pi/2) \times \operatorname{ch}^{-2}(\pi x)$, as is easily checked. I leave it to you: Just compute the inverse transform $(2\pi)^{-1}\int e^{-\sqrt{-1}\lambda x}(\lambda/2)[\operatorname{sh}(\lambda/2)]^{-1}\, d\lambda$ for $x < 0$ by residues in the upper half-plane.

The other inversions: of 1) $\left(\operatorname{ch}\sqrt{\lambda}\right)^{-1/2}$ and 2) $\left(\lambda^{-1/2}\operatorname{sh}\sqrt{\lambda}\right)^{-1/2}$ are more difficult. 1) was done by Cameron and Martin [3, pp. 195–209]. The present computation is more economical:

$$\frac{1}{\sqrt{\operatorname{ch}\sqrt{\lambda}}} = \frac{\sqrt{2}e^{-\sqrt{\lambda}/2}}{\sqrt{1 + e^{-2\sqrt{\lambda}}}} = \sqrt{2}\sum_{n=0}^{\infty}(-1)^n \frac{1\cdot 3\cdot 5\cdot\ldots\cdot(2n - 1)}{2\cdot 4\cdot 6\cdot\ldots\cdot 2n} e^{-(2n+1/2)\sqrt{\lambda}}$$

$$= \sum_{n=0}^{\infty}(-1)^n \frac{1\cdot 3\cdot 5\cdot\ldots\cdot(2n - 1)}{2\cdot 4\cdot 6\cdot\ldots\cdot 2n}\left(n + \frac{1}{4}\right)\int_0^{\infty}\frac{e^{-(n+1/4)^2/x}}{\sqrt{\pi x^3}} e^{-\lambda x}\, dx$$

where

$$\frac{1}{\sqrt{1 - x}} = \sum_{n=0}^{\infty}\frac{1\cdot 3\cdot 5\cdot\ldots\cdot(2n - 1)}{2\cdot 4\cdot 6\cdot\ldots\cdot 2n} x^n$$

is used for the second equality and

$$\int_0^{\infty} x\frac{e^{-x^2/2t}}{\sqrt{2\pi t^3}} e^{-\lambda t}\, dt = e^{-x\sqrt{2\lambda}}$$

for the third one; the latter is easily deduced from the fact that $p = (2\pi t)^{-1/2}e^{-x^2/2t}$ solves the heat equation $\partial p/\partial t = \partial^2 p/2\partial x^2$. The probability density of $\int_0^1 \mathfrak{x}^2(t)\, dt$ for the free Brownian motion may now be read off:

$$\sum_{n=0}^{\infty}(-1)^n \frac{1\cdot 3\cdot 5\cdot\ldots\cdot(2n - 1)}{2\cdot 4\cdot 6\cdot\ldots\cdot 2n} \times \left(n + \frac{1}{4}\right)\frac{e^{-(n+1/4)^2 x}}{\sqrt{\pi x^3}}.$$

It is related to one of Jacobi's theta functions. Not very attractive, but there it is. The inversion of 2) $\left(\lambda^{-1/2}\operatorname{sh}\sqrt{\lambda}\right)^{-1/2}$ is worse, so I leave the matter here.

3.3. Robt. Brown's Brownian motion

I come to the subject by indirection.

Let's modify the symbolic expression of the free Brownian measure for "short" paths $\mathfrak{x}(t)$, $0 \leq t \leq 1$, by the introduction of a density function, as in

$$e^{m/2} \times \exp\left(-\frac{m}{2}\,\mathfrak{x}^2(1) - \frac{m^2}{2}\int_0^1 \mathfrak{x}^2\right) \times \frac{1}{(2\pi 0+)^{\infty/2}} \exp\left(-\frac{1}{2}\int_0^1 (\mathfrak{x}^\bullet)^2\right) d^\infty \mathfrak{x}.$$

The exponent is of degree 2 in $\mathfrak{x}$, so the measure is still Gaussian, up to a normalizer to make the total mass be 1. That is the role of the factor $e^{m/2}$: The total mass of the rest, alias the free Brownian mean-value of $\exp\left[-m\mathfrak{x}^2(1)/2 - (m^2/2)\int_0^1 \mathfrak{x}^2\right]$, may be reduced in the now familiar way, to the reciprocal square root of $\Delta = \det\left(I + m^2 K + mKe\otimes e\right)$ with $K(x,y) = \min(x,y)$ and e the unit mass at $x = 1$, i.e. $\int_0^1 fe \equiv f(1)$. The symbol $e \otimes e$ is not quite kosher but don't worry: it is only of rank 1 and will take care of itself. The computation of this determinant follows:

$$\Delta = \det\left(I + m^2 K\right) \times \left(I + \left(I + m^2 K\right)^{-1} mKe\otimes e\right) = \operatorname{ch} m \times \left(1 + \operatorname{sp}\left(I + m^2 K\right)^{-1} mKe\otimes e\right)$$

$$= \operatorname{ch} m \times \left[1 + \left(I + m^2 K\right)^{-1} mK \quad \text{evaluated at } (1,1)\right].$$

Now you need to know $\left(I + m^2 K\right)^{-1} \equiv I + G$. Here, K is inverse to $-D^2$ acting on $C^2[0,1] \cap \{f : f(0) = 0 = f'(1)\}$, so G is inverse to $D^2/m^2 - I$ acting on the same domain, and an easy computation produces its (symmetric) kernel

$$G(x,y) = \frac{-m}{\operatorname{ch} m}\,\operatorname{sh} mx\,\operatorname{ch} m(1-y) \qquad \text{if} \quad x \leq y.$$

Then

$$\left(I + m^2 K\right)^{-1} K = \left(I + m^2 K\right)^{-1} \times \frac{1}{m^2}\left(I + m^2 K - I\right) = -\frac{G}{m^2},$$

so $\Delta = \operatorname{ch} m \times (1 - G(1,1)/m) = e^m$, as wanted, i.e. the measure is normalized.

But what Gaussian measure is this anyhow? The correlation

$$Q = e^{m/2} E\left[\exp\left(-\frac{m}{2}\,\mathfrak{x}^2(1) - \frac{m^2}{2}\int_0^1 \mathfrak{x}^2\right)\mathfrak{x}(t_1)\,\mathfrak{x}(t_2)\right]$$

will tell, or, what is just as good, the associated quadratic form

$$(f, Qf) = e^{m/2} E\left[\exp\left(-\frac{m}{2}\,\mathfrak{x}^2(1) - \frac{m^2}{2}\int_0^1 \mathfrak{x}^2\right)\left(\int_0^1 f\mathfrak{x}\right)^2\right],$$

which may be obtained from $Z = e^{m/2} E\left[\exp\left(-m\mathfrak{x}^2(1)/2 - m^2\int_0^1 \mathfrak{x}^2/2\right)\exp\left(-\lambda\left(\int_0^1 f\mathfrak{x}\right)^2/2\right)\right]$ by taking $-2\partial/\partial\lambda$ at $\lambda = 0$. The computation follows the pattern for the normalizer: Z is reduced to

$$e^{m/2}\left[\det\left(I + m^2 K + mKe\otimes e + \lambda Kf\otimes f\right)\right]^{-1/2}$$

from which rule 6 of Section 2.4 produces

$$(f, Qf) = -2e^{m/2} \times -\frac{1}{2}\left[\det\left(I + m^2 K + mKe\otimes e\right)\right]^{-3/2} \times \det\left(I + m^2 K + mKe\otimes e\right) \times \operatorname{sp}\left[\left(I + m^2 K + mKe\otimes e\right)^{-1} Kf\otimes f\right].$$

Now $\det\left(I + m^2K + mKe\otimes e\right) = e^m$, as you already know, so line 1 is cancelled. Only line 2 remains, which is to say

$$Q = \left(I + m^2K + mKe\otimes e\right)^{-1}K = \left(I + \left(I + m^2K\right)^{-1}mKe\otimes e\right)^{-1}\left(I + m^2K\right)^{-1}K.$$

Here, $\left(I + m^2K\right)^{-1}K = -G/m^2$ as before, and the simple rule

$$(I + a\otimes b)^{-1} = I - (1 + (a, b))^{-1}a\otimes b \qquad \text{with} \qquad a = -Ge/m \quad \text{and} \quad b = e$$

produces

$$Q = \left[I - \left(1 - \frac{1}{m}G(1,1)\right)^{-1}\left(-\frac{1}{m}Ge\otimes e\right)\right] \times -\frac{1}{m^2}G$$

$$= -\frac{1}{m^2}G(x,y) - \frac{1}{m}e^{-m}\,\text{ch}\,m \times \left(-\frac{1}{m}G(x,1)\right) \times \left(-\frac{1}{m}G(y,1)\right) = \frac{1}{m}\,\text{sh}\,m\,x\,e^{-my} \qquad \text{for} \quad x \le y,$$

as you will check, only now you should put $x = t_1$ and $y = t_2$.

OK. But, still, what is that? Something more tangible is wanted. A symbolic calculation will guide us: Let $F(\mathfrak{x})$ be any nice function of the short path $\mathfrak{x}(t)$, $0 \le t \le 1$. Then

$$e^{m/2}\int \frac{F(\mathfrak{x})}{(2\pi 0+)^{\infty/2}}\,\exp\left(-\frac{m}{2}\mathfrak{x}^2(1) - \frac{m^2}{2}\int_0^1 \mathfrak{x}^2\right)\exp\left(-\frac{1}{2}\int(\mathfrak{x}^\bullet)^2\right)d^\infty\mathfrak{x} = e^{m/2}\int \frac{F(\mathfrak{x})}{(2\pi 0+)^{\infty/2}}\,\exp\left(-\frac{1}{2}\int(\mathfrak{x}^\bullet + m\mathfrak{x})^2\right)d^\infty\mathfrak{x}.$$

Put $\mathfrak{x}^\bullet + m\mathfrak{x} = \mathfrak{y}^\bullet$, or what is the same for $\mathfrak{x}(0) = \mathfrak{y}(0) = 0$,

$$\mathfrak{x}(t) = e^{-mt} * \mathfrak{y}^\bullet = \mathfrak{y}(t) - me^{-mt} * \mathfrak{y},$$

where $*$ stands for the convolution. The prior display now takes the form

$$e^{m/2}\int \frac{F(e^{-mt} * \mathfrak{y}^\bullet)}{(2\pi 0+)^{\infty/2}}\,\exp\left(-\frac{1}{2}\int_0^1(\mathfrak{y}^\bullet)^2\right)\det\left[\frac{\partial\mathfrak{x}}{\partial\mathfrak{y}}\right]d^\infty\mathfrak{y}$$

with the formal Jacobian determinant of $\mathfrak{x}$ in respect to $\mathfrak{y}$:

$$\frac{\partial\mathfrak{x}(t_1)}{\partial\mathfrak{y}(t_2)} = \begin{cases} [\text{the identity}] - m \times e^{m(t_2 - t_1)} & \text{if } t_2 < t_1, \\ 0 & \text{otherwise.} \end{cases}$$

This is ambiguous on diagonal, but if you take it to be $-m/2$ by way of compromise, then the Jacobian determinant is $e^{-m/2}$, and the expectation assumes its final form:

$$e^{m/2}E\left[F(\mathfrak{x})\,\exp\left(-\frac{m}{2}\mathfrak{x}^2(1) - \frac{m^2}{2}\int_0^1 \mathfrak{x}^2\right)\right] = E\left[F(e^{-mt} * \mathfrak{y}^\bullet)\right]$$

with the free Brownian mean–value to the right. The formula originates with Cameron and Martin [3]; it suggests that the Gaussian probabilities

$$e^{m/2}\,\exp\left(-\frac{m}{2}\mathfrak{x}^2(1) - \frac{m^2}{2}\int_0^1 \mathfrak{x}^2\right)\frac{1}{(2\pi 0+)^{\infty/2}}\,\exp\left(-\frac{1}{2}\int_0^1(\mathfrak{x}^\bullet)^2\right)d^\infty\mathfrak{x}$$

are one and the same as what the free Brownian $\mathfrak{y}$ induces on path space by the transformation $\mathfrak{y}(t) \to \mathfrak{x}(t) = e^{-mt} * \mathfrak{y}^\bullet$; in fact, for these paths $\mathfrak{x}$,

$$E\left[\mathfrak{x}(t_1)\mathfrak{x}(t_2)\right] = e^{-mt_1} \exp\left(-mt_2 \int_0^{t_1} e^{2mt'}\, dt'\right) = \frac{1}{m}\, \mathrm{sh}\,(mt_1)\, e^{-mt_2} \qquad \text{for} \quad t_1 \le t_2,$$

as it ought to be. Take that as an exercise.

Finally, I come to the "real" Brownian motion: Ornstein–Uhlenbeck [23] advocated $\mathfrak{x}$ as a model of velocity of the Brownian motion of Robt. Brown. I change notation accordingly, from $\mathfrak{x}$ to $\mathfrak{v}$. Then the position of the Brownian particle, starting from rest at 0, is

$$\mathfrak{x}(t) = \int_0^t \mathfrak{v}(t')\, dt'$$

and you have, symbolically, Newton's law:

$$[\text{mass 1}] \times [\text{acceleration}] = \frac{d^2\mathfrak{x}}{dt^2} = \frac{d\mathfrak{v}}{dt} = -m\mathfrak{v} + \frac{d\mathfrak{y}}{dt} = [\text{force}]$$

in which $-m\mathfrak{v}$ is descriptive of air resistance (drag) and $d\mathfrak{y}/dt$ of the (thermal) agitation of the air molecules. Now intensify this agitation by a factor m, so that drag and agitation are comparable, and make $m \uparrow \infty$. Then

$$\frac{d^2\mathfrak{x}}{dt^2} = -m\left(\mathfrak{v}\,\frac{d\mathfrak{x}}{dt}\right) + m\,\frac{d\mathfrak{y}}{dt},$$

and with $\mathfrak{x}(0) = \mathfrak{v}(0) = 0$, you have

$$\mathfrak{x}(t) = \mathfrak{y} + [\text{a negligible error of root-mean-square}] = \sqrt{\frac{1}{2m}\,(1 - e^{-2mt})}.$$

In this way, Ornstein–Ulhenbeck reconciled their "real" Brownian motion $\mathfrak{x}$ with that of Einstein [9] who took it to be the free Brownian motion $\mathfrak{y}$, plain.

3.4.   Shallow water: background and preliminaries

The equation of Korteweg–de Vries (1895) is descriptive of the leading edge of long surface waves in shallow water; in convenient units, it reads: $\partial v/\partial t + v\partial v/\partial x = \partial^3 v/\partial x^3$, $v(t, x)$ being the velocity of the wave, taken to be independent of y as in the figure.
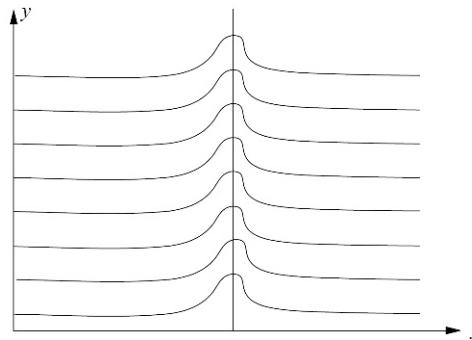
Here, you have a competition between 1) $\partial v/\partial t + v\,\partial v/\partial x = 0$ and 2) $\partial v/\partial t = \partial^3 v/\partial x^3$. 1) exhibits "shocks". 2) is "dispersive". I explain: 1) states that points at height h on the graph of v move to the right at speed h, so if $f = v(0+, \cdot)$ is increasing nothing bad happens, but if it is decreasing, then high values overtake low values and pass beyond, so that $v(t, \cdot)$ ceases to be a single-valued function. The inception of this turning over is a "shock". As to 2), $v(t, x) = \sin k\big(x - k^2 t\big)$ is a solution, showing that waves of different frequencies (k) move at different speeds (k^2). This is "dispersion". In KdV, dispersion overcomes the shocks and the solution is perfectly nice for all times $t \geq 0$ and, indeed, for $t \leq 0$, too.

The explicit solution of KdV by Gardiner–Greene–Kruskal–Miura [11] makes a fascinating story. I cannot tell it here, but see Lamb [14] for the best introduction to what has come to be called "KdV and all that". But one remarkable aspect of it *can* be told quite simply, following a beautiful computation of Ch. Pöppe [21].

Take a solution $w = w(t, x)$ of KdV with the non-linearity $v\partial v/\partial x$ crossed out, i.e. $\partial w/\partial t = \partial^3 w/\partial x^3$, and use it to make an operator W on the half-line $[0, \infty)$, depending upon $0 \leq t < \infty$ and $-\infty < x < \infty$, as in

$$W : f \in C[0, \infty) \mapsto \int_0^\infty w(t, x + \xi + \eta)\, f(\eta)\, d\eta.$$

Now let $\Delta = \Delta(t, x)$ be the determinant $\det(I + W)$, introduced into the subject by Dyson [8]. Then, miraculously, $v(t, x) = -(1/12)\,\partial^2 \ln \Delta/\partial x^2$ solves KdV, which is to say that this machinery $w \to \Delta \to v$ converts the (easy) linear flow $\partial w/\partial t = \partial^3 w/\partial x^3$ into the (hard) non-linear KdV flow $\partial v/\partial t + v\,\partial v/\partial x = \partial^3 v/\partial x^3$, but more of this below.

Of course, you must take care that v makes sense: The function little w had better decay at $x = +\infty$ so that Δ is kosher as a Fredholm determinant, as would be the case if $w(0+, \cdot)$ were smooth and rapidly vanishing at $\pm\infty$ and you took

$$w(t, x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\sqrt{-1}/kx}\, e^{\sqrt{-1}\,k^3 t}\, \widehat{w}(0+, k)\, dk,$$

in which $\widehat{}$ signifies the transform $f \mapsto \int e^{\sqrt{-1}kx} f(x)\, dx$. Then $\Delta = 1$ at $x = +\infty$, but you still need to keep Δ positive elsewhere so that $\ln \Delta$ is nice, which is to say that $I + W$ has never any null-space. Now if $f \in C[0, \infty)$ is such a (real) null-function, then the rapid vanishing of w at $+\infty$ imparts the same decay to f itself, and with the understanding that $f(\xi)$, vanishes for $\xi \leq 0$, you may conclude from

$$f(\xi) + \frac{1}{2\pi} \int e^{-\sqrt{-1}\,k\xi} e^{-\sqrt{-1}\,kx} e^{\sqrt{-1}\,k^3 t} \widehat{w}(0+, k)\widehat{f}(-k)\, dk = 0$$

that

$$\int_0^\infty f^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\widehat{f}|^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\widehat{w}|^2 |\widehat{f}|^2\, dk$$

in view of $\widehat{f}(k)^* = \widehat{f}(-k)$, f being real. This cannot be maintained if $|\widehat{w}| < 1$ unless $\widehat{f}$, and so also f, vanishes: in fact, $|\widehat{w}| \leq 1$ suffices and is necessary, as well, but that is a little longer story which I omit. This ends the preparations.

3.5. Pöppe's computation

Pöppe showed, by direct manipulation of determinants, that $-(1/12)(\ln \Delta)''$ solves KdV. The simpler function $V \equiv -(\ln \Delta)'$ is better suited to the proof. It is $-\operatorname{sp}(1 + W)^{-1}DW$ with $D = d/dx$, by rule 6 of Section 2.4, and may be reduced to $\big[(I + W)^{-1}W\big](0, 0)/2$ by means of the "Hankel" character of W, by which I mean that its kernel is a function of $\xi + \eta$ alone: in fact, the kernel of $(I + W)^{-1}DW$ is $(\partial/\partial\eta)\big[(I + W)^{-1}W\big](\xi, \eta)$, so

$$V = -\operatorname{sp}(I + W)^{-1}DW = -\frac{1}{2}\operatorname{sp}\big[DW(I + W)^{-1} + (I + W)^{-1}DW\big]$$

$$= -\frac{1}{2}\operatorname{sp}\big[D(I + W)^{-1}W + (I + W)^{-1}DW\big] = -\frac{1}{2}\int_0^\infty \frac{d}{d\xi}\big[(I + W)^{-1}W\big](\xi, \xi)\, d\xi = \frac{1}{2}\big[(I + W)^{-1}W\big](0, 0).$$

I will need $V', V'', V''', (V')^2$, and also $V^\bullet$. The first three are easy, as is the last: with the notation $W' = DW$ etc. and abbreviating $[\ ](0,0)$ by $[\ ]$ plain, you find

$$V' = \frac{1}{2}\Big[-(I+W)^{-1}W'(I+W)^{-1}W + (I+W)^{-1}W'\Big] = \frac{1}{2}\Big[(I+W)^{-1}W'(I+W)^{-1}\Big],$$

$$V'' = -\Big[(I+W)^{-1}W'(I+W)^{-1}W'(I+W)^{-1}\Big] + \frac{1}{2}\Big[(I+W)^{-1}W''(I+W)^{-1}\Big],$$

$$V''' = 3\Big[(I+W)^{-1}W'(I+W)^{-1}W'(I+W)^{-1}W'(I+W)^{-1}\Big] - 3\Big[(I+W)^{-1}W''(I+W)^{-1}W'(I+W)^{-1}\Big]$$
$$+ \frac{1}{2}\Big[(I+W)^{-1}W'''(I+W)^{-1}\Big],$$

and

$$V^\bullet = \frac{1}{2}\Big[(I+W)^{-1}W'''(I+W)^{-1}\Big].$$

The reduction of $(V')^2$ to like form employs the Hankel trick again: Let the operators A, B, C and D have nice kernels and good decay at $+\infty$, abandoning temporarily the prior meaning of D, and let B and C be Hankel besides. Then, with an obvious notation,

$$[AB](0,0)[CD](0,0) = -\int_0^\infty \frac{d}{d\xi}[AB](0,\xi) \times [CD](\xi,0)\,d\xi - \int_0^\infty [AB](0,\xi) \times \frac{d}{d\xi}[CD](\xi,0)\,d\xi$$

$$= -\int_0^\infty [AB'](0,\xi)[CD](\xi,0)\,d\xi - \int_0^\infty [AB](0,\xi)[C'D](\xi,0)\,d\xi = -[A(B'C + BC')D](0,0).$$

Now write $2V'$ in its initial form:

$$2V' = \Big[(I+W)^{-1}W'\Big] - \Big[(I+W)^{-1}W'(1+W)^{-1}W\Big],$$

dropping the $(0,0)$ as before, and square:

$$4(V')^2 = \Big[(I+W)^{-1}W'\Big]\Big[(W'(I+W)^{-1}\Big] - 2\Big[(I+W)^{-1}W'\Big]\Big[W(I+W)^{-1}W'(I+W)^{-1}\Big]$$
$$+ \Big[(I+W)^{-1}W'(I+W)^{-1}W\Big]\Big[W(I+W)^{-1}W'(I+W)^{-1}\Big],$$

in which the symmetric character of W, W', and $(I+W)^{-1}$ is used to manipulate the second factors one by one. Then apply the $ABCD$ rule to produce

$$4(V')^2 = -\Big[(I+W)^{-1}\big(W''W' + W'W''\big)(I+W)^{-1}\Big] + 2\Big[(I+W)^{-1}\big(W''W + (W')^2\big)(I+W)^{-1}W'(I+W)^{-1}\Big]$$
$$- \Big[(I+W)^{-1}W'(I+W)^{-1}\big(W'W + WW'\big)(I+W)^{-1}W'(I+W)^{-1}\Big]$$

and write out $V^\bullet - V''' + 6(V')^2$ in the present notation:

$$V^\bullet - V''' + 6(V')^2$$

$$= \frac{1}{2}\Big[(I+W)^{-1}W'''(I+W)^{-1}\Big] \tag{1}$$

$$- 3\Big[(I+W)^{-1}W'(I+W)^{-1}W'(I+W)^{-1}W'(I+W)^{-1}\Big] \tag{-(2)}$$

$$+ 3\Big[(I+W)^{-1}W''(I+W)^{-1}W'(I+W)^{-1}\Big] \tag{3}$$

$$- \frac{1}{2}\Big[(I+W)^{-1}W'''(I+W)^{-1}\Big] \tag{-(4)}$$

$$- 3\Big[(I+W)^{-1}W''W'(I+W)^{-1}\Big] \tag{-(5)}$$

$$+ 3\Big[(I+W)^{-1}W''\big(I-(I+W)^{-1}\big)W'(I+W)^{-1}\Big] \tag{(6)-(7)}$$

$$+ 3\Big[(I+W)^{-1}(W')^2(I+W)^{-1}W'(I+W)^{-1}\Big] \tag{8}$$

$$- 3\Big[(I+W)^{-1}W'(I+W)^{-1}W'\big(I-(I+W)^{-1}\big)W'(I+W)^{-1}\Big] \tag{-(9)+(10)}$$

in which (1) and (4) cancel, also (5) and (6), (3) and (7), (8) and (9), (2) and (10), i.e. the whole thing vanishes. Remarkable! Now take $v = V'$. Then $v^\bullet - v''' + 12vv' = 0$ which is KdV aside from the nuisance factor 12, and this is removed by changing v into $v/12$.

Example: solitons.

Take, as solution of $\partial w/\partial t = \partial^3 w/\partial x^3$, the function

$$w(t,x) = \sum_{i=1}^{n} m_i \exp\left(-k_i x - k_i^3 t\right)$$

with positive numbers m and distinct positive k's. Then $v = -(1/12)(\ln \Delta)''$ solves KdV. Here, Δ can be expressed pretty explicitly. Let's look for eigenfunctions of W at $t = 0$:

$$W : f \mapsto \sum m_i\, e^{-k_i \xi} \int_0^\infty e^{-k_i \eta} f(\eta)\, d\eta$$

so any such eigenfunction is of the form $f(\xi) = \sum_{j=1}^{n} f_j\, e^{-k_j \xi}$, and a self–evident computation shows that $Wf = \lambda f$ is the same as to say

$$\lambda f_i = \sum_{j=1}^{n} \frac{m_i}{k_i + k_j}\, f_j \qquad \text{for} \quad 1 \le i \le n,$$

i.e. $\Delta = \det(I + W)$ is the ordinary determinant

$$\det\left[I + \frac{m_i}{k_i + k_j}\right]_{1 \le i,j \le n},$$

which may be spelled out in Fredholm's way as

$$1 + \sum_{p=1}^{n} \sum_{|\mathbf{n}|=p} \left(\frac{\mathbf{m}}{2\mathbf{k}}\right)^{\mathbf{n}} \prod_{\substack{i<j \\ i,j \in \mathbf{n}}} \left(\frac{k_i - k_j}{k_i + k_j}\right)^2,$$

in which $\mathbf{n}$ is a collection of indices $1 \le n_1 < n_2 < \cdots < n_p \le n$ and $(\mathbf{m}/2\mathbf{k})^{\mathbf{n}}$ means $(m_1/2k_1)^{n_1} \times (m_2/2k_2)^{n_2} \times$ etc. I leave this to you as an exercise with a hint to be verified first:

$$\det\left[\frac{1}{k_i + k_j}\right]_{1 \le i,j \le n} = \prod_{i<j}(k_i - k_j)^2 \bigg/ \prod_{i,j}(k_i + k_j)^2.$$

Omitting the corrective 1/12, the corresponding solution of KdV is then $v(t,x) = -(\ln \Delta)''$ with

$$\Delta = \det\left[I + \frac{m_i}{k_i + k_j}\, e^{-k_i x} e^{-k_i^3 t}\right]_{1 \le i,j \le n}.$$

For $n = 1$, Δ is simply $1 + (m/2k)\exp(-kx - k^3 t)$, and

$$v(t,x) = -\frac{k^2}{4}\, \text{ch}^{-2}\left[\frac{kx}{2} + \frac{k^3 t}{2} + \frac{1}{2}\ln\frac{2k}{m}\right],$$

as you will check. This is the famous "soliton" of "KdV and all that", representing, e.g. a solitary wave (tsunami) in the open ocean, traveling east to west; compare Lamb [14]. For $n \ge 2$, v is a non–linear combination of such solitons. Here is why: The soliton has amplitude $k^2/4$ and speed k^2, i.e. big solitons go fast, small solitons go slow. Now you can start off a solution of KdV by placing widely–spaced solitons in order of diminishing amplitude/speed, right to left. They hardly feel each other: the individual soliton dies off fast from its peak value, so the non–linearity vv' does not count for much, and you can simply add the n solitons together. Now turn on KdV: The big solitons catch up to the little ones, and a complicated (non–linear) interaction takes place, but by and by you see the original solitons emerging to the left, moving off to $-\infty$ in order of increasing amplitude/speed, left to right, with a little evanescent "radiation" in between. It is this evolution that $v = -(\ln \Delta)''$ describes. I omit the computation, but try it for $n = 2$.

Coda.

Now the only trouble is that you don't know how to pick $w(0+, \cdot)$ to produce the initial velocity profile $v(0+, \cdot)$ you want. That is a pretty long story, how it's done. Here, I am after determinants and just refer you to Lamb [14] and Dyson [8] for the full story. It involves an elaborate detour via a sort of non-linear Fourier transform ("scattering"), quite remarkable in itself but, to my mind, a little wrong-headed. If only you could invert the map $w(0+, \cdot) \to v(0+, \cdot)$ directly, in real space so to say, then all would be well. I like to make this caricature: To solve $x^\bullet = 1$ is easy even if you know nothing; $x^\bullet = \sqrt{1 - x^2}$ looks harder, but if you have the wit to introduce the special function $y = \sin^{-1} x = \int_0^x (1 - r^2)^{-1/2} \, dr$, then $y^\bullet = 1$, $x = \sin y$, and you're back in business. The map $v \to w$ is just a variant of this device, i.e. it is an ∞-dimensional "special function", inverse to $w \to v = -(\ln \Delta)''$. How bad can it be? The question merits a direct answer.

## 3.6. The unitary ensemble: background and preparations

The idea to model atomic energy levels by the spectra of random matrices is due to Wigner [27]. The subject has been much developed since then; see Mehta [19] for the best general account. Curiously, there even seems to be a mysterious connection to Gauss's prime number theorem, $\#\{\text{primes} \leq n\} \simeq n / \lg n$, mediated by Riemann's zeta function, for which see Conrey [4]. The group G of $N \times N$ unitary matrices provides the simplest example of what people do. G is equipped with its (one and only) translation invariant volume element, normalized to make the total volume equal 1. That is the only natural way to introduce probabilities into the picture, and it is the statistics of the eigenvalues $e^{\sqrt{-1}\theta}$ of a "typical" unitary matrix which is sought, especially in high dimensions $N \uparrow \infty$. The pretty fact is that these statistics can be expressed by means of Fredholm determinants: for example, for $N < \infty$, the probability that no eigenvalue falls in the circular arc $\{e^{\sqrt{-1}\theta} : \alpha \leq \theta \leq \beta\}$ is

$$\det\left[I - \frac{1}{2\pi} \frac{\sin N(\theta - \varphi)/2}{\sin(\theta - \varphi)/2} \right]_{\alpha \leq \theta, \varphi \leq \beta}.$$

The discussion is adopted from Tracy–Widom [22], more or less.

The invariant volume.

I follow Weyl [24, pp. 194–198]. Let $D \subset G$ be the commutative subgroup of diagonal unitaries $e = e^{\sqrt{-1}\theta}$, by which I mean $e^{\sqrt{-1}\theta_n}$, $1 \leq n \leq N$, on diagonal.
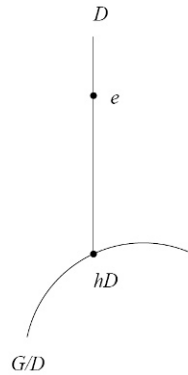
The figure depicts G fibered by D over the base G/D. To understand what is meant, take a general element $g \in G$ and bring it, as you may, to diagonal form by a unitary similarity: $g = heh^{-1}$ with $e \in D$ and $h \in G$. Here, h can be multiplied from the right by any $e' \in D$ without changing g, so it is only the coset hD that counts. In fact, for g in "general position", both this coset and the covering point e figuring in $g = heh^{-1}$ are fully determined in the small by the requirement that e not change much in response to a little change in $g - e$ is made from the eigenvalues of g which can only be permuted, and for g in general position, such a permutation is a *big* change, so if now h is determined at

the base, then also e is known, and likewise in a small patch about $g = heh^{-1}$, with e and h changing only a little, by de and dh in response to a little change dg in g. Obviously

$$g^{-1} dg = he^{-1}h^{-1}\left(dh\, eh^{-1} + h\, de\, h^{-1} - heh^{-1}\, dh\, h^{-1}\right) = h\left(e^{-1}h^{-1}\, dh\, e - h^{-1}\, dh + \sqrt{-1}\, d\theta\right) h^{-1},$$

in which you may reduce the apparent supernumerary (real) degrees of freedom (N^2 for dh plus N more for $d\theta$) to the correct number (N^2) by the observation that $h^{-1}dh$ is pure imaginary on diagonal, permitting the restriction $h^{-1}dh = \sqrt{-1}\, d\theta$ there. The volume element Ω is now produced by wedging the several 1–forms $g^{-1}dg$ up to top dimension. Each of these is insensitive to left translations, so Ω inherits this feature. Besides, the associated line element

$$\mathrm{sp}\left[(g^{-1}dg)^{\dagger *} g^{-1} dg\right] = \mathrm{sp}\left[(dg)^{\dagger *} dg\right] \equiv |dg|^2$$

is insensitive to the presence of h to the left and h^{-1} to the right, so these can be dropped and the bare 1–forms $e^{-1}h^{-1}dh\, e - h^{-1}d = h + \sqrt{-1}\, d\theta$ wedged up to top dimension to produce the volume element:

$$\Omega = \prod_{i \neq j}\left(\frac{e_j}{e_i} - 1\right) d\theta_1\, d\theta_2 \ldots \times [\text{a volume element } \omega \text{ on } G/D],$$

or, taking the factors of the product by pairs,

$$\Omega = \prod_{i<j} |e_i - e_j|^2 d\theta \times \omega.$$

The spectral density.

The form of ω does not matter anymore. What is wanted is the volume element induced upon the eigenvalues:

$$Z^{-1} \times \prod_{i<j} \left|e^{\sqrt{-1}\,\theta_i} - e^{\sqrt{-1}\,\theta_j}\right|^2 \prod d\theta_k \equiv p(\theta)\, d\theta,$$

in which Z is a normalizing factor, to be determined presently, fixing the total volume at 1 so that these are probabilities. To put them in a better form, introduce the matrix $Q = \left[e^{\sqrt{-1}\,(i-1)\,\theta_j}\right]_{1 \leq i,j \leq N}$ with the Vandermonde–type determinant:

$$\prod_{i<j}\left(e^{\sqrt{-1}\,\theta_j} - e^{\sqrt{-1}\,\theta_i}\right).$$

Then

$$\prod_{i<j}\left|e^{\sqrt{-1}\,\theta_j} - e^{\sqrt{-1}\,\theta_i}\right|^2 = \det Q^{\dagger} \times \det Q^* = \det\left(Q^{\dagger} Q^*\right),$$

and from

$$\left(Q^{\dagger} Q^*\right)_{ij} = \sum_{k=0}^{N-1} e^{\sqrt{-1}\,(\theta_i - \theta_j)k} = \frac{e^{\sqrt{-1}\,(\theta_i - \theta_j)N} - 1}{e^{\sqrt{-1}(\theta_i - \theta_j)} - 1} = \frac{e^{\sqrt{-1}\,\theta_i N/2} \sin\frac{N}{2}(\theta_i - \theta_j)\, e^{-\sqrt{-1}\,\theta_j N/2}}{e^{\sqrt{-1}\,\theta_i/2} \sin\frac{1}{2}(\theta_i - \theta_j)\, e^{\sqrt{-1}\,\theta_j/2}},$$

you find

$$p(\theta) = Z^{-1} \det\left[K(\theta_i - \theta_j)\right]_{1 \leq i,j \leq N}$$

with

$$K(\theta, \varphi) = \frac{1}{2\pi} \frac{\sin N(\theta - \varphi)/2}{\sin(\theta - \varphi)/2}$$

and the modified normalizer

$$Z = \int \det\left[\frac{1}{2\pi} \sum_{k=0}^{N-1} e^{\sqrt{-1}\,(\theta_i - \theta_j)k}\right]_{1 \leq i,j \leq N} d\theta.$$

Now

$$\det\left[\sum_{k=0}^{N-1} e^{\sqrt{-1}(\theta_i-\theta_j)k}\right]_{1\le i,j\le N} = \sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\dots \det\left[e^{\sqrt{-1}(\theta_i-\theta_j)k_i}\right]_{1\le i,j\le N} = \sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\dots \det\left[e^{\sqrt{-1}\theta_i(k_i-k_j)}\right]_{1\le i,j\le N},$$

and as the latter determinant vanishes unless the k's are distinct, so Z reduces to the sum over such k's of

$$\sum_{\pi}\frac{\chi(\pi)}{(2\pi)^N}\int\prod_{i=0}^{N}e^{\sqrt{-1}\,\theta_i(k_i-k_{\pi i})}\,d\theta.$$

But all these integrals vanish unless π is the identity, so the whole reduces to the number of choices of N distinct k's from $0,1,\dots,N-1$, i.e. $Z = N!$.

The same style of computation produces a nice formula for the marginal density $p(\theta_1,\dots,\theta_n)$ with the other angles θ_k, $N \ge k > n$, integrated out:

$$p(\theta_1,\dots,\theta_n) = \frac{(N-n)!}{N!}\det\left[K(\theta_i-\theta_j)\right]_{1\le i,j\le n}.$$

I leave that to you as an exercise.

3.7. Occupation numbers, determinants and the scaling limit

Now Fredholm determinants enter the story. Let $\#$ be the number of angles θ that fall within an interval $J \subset [0,2\pi)$. It is desired to evaluate the probabilities $P(\# = n)$ for $n \ge 0$. Everything is symmetric in the angles and there are "N choose n" ways of specifying which particular angles fall in J, so

$$P(\# = n) = \binom{N}{n}P\left[\bigcap_{i=0}^{n}E_i\bigcap_{j=n+1}^{N}E_j'\right]$$

in which E_i is the event "$\theta_i \in J$" and E_j' the compliment of E_j, i.e. "$\theta_j \notin J$". This yields to the principle of inclusion–exclusion: with $\bigcap_{i=0}^{n}E_i = E$ for short, you have

$$P\left(E\cap\bigcap_{j=n+1}^{N}E_j'\right) = P(E) - P\left(E\cap\bigcup_{j=n+1}^{N}E_j\right)$$

$$= P(E) - \sum_{i>n}P(E\cap E_i) + \sum_{j>i>n}P(E\cap E_i\cap E_j) - \sum_{k>j>i>n}P(E\cap E_i\cap E_j\cap E_k) + \dots$$

$$= \sum_{p=n}^{N}(-1)^{p-n}\binom{N-n}{p-n}P\left(E\cap E_{n+1}\cap\dots\cap E_p\right) \qquad\text{by symmetry}$$

$$= \sum_{p=n}^{N}(-1)^{p-n}\binom{N-n}{p-n}\int_{J^p}\frac{(N-p)!}{N!}\det\left[K(\theta_i-\theta_j)\right]_{1\le i,j\le p}\,d\theta,$$

which reduces, after multiplication by "N chose n", to

$$P(\# = n) = \sum_{p=n}^{N}\frac{(-1)^{p-n}}{n!(p-n)!}\int_{J^p}\det\left[K(\theta_i-\theta_j)\right]_{1\le i,j\le p}\,d\theta.$$

Here is the punch-line: multiply by $(1-\lambda)^n$, sum from $n=0$ to N, and reverse the two summations to produce

$$\sum_{n=0}^{N}(1-\lambda)^n P(\#=n) = \sum_{p=0}^{N}\left[\sum_{n=0}^{p}(1-\lambda)^n(-1)^{p-n}\frac{p!}{n!(p-n)!}\right]\frac{1}{p!}\int_{J^p}\det\left[K(\theta_i-\theta_j)\right]_{1\le i,j\le p}d\theta$$

$$=\sum_{p=0}^{N}\frac{(-\lambda)^p}{p!}\int_{J^p}\det\left[K(\theta_i-\theta_j)\right]_{1\le i,j\le p}d\theta = \det\left[I-\lambda\left(K \text{ restricted to } J\times J\right)\right],$$

K being of rank N, i.e. $P(\#=n)$ is the coefficient of $(1-\lambda)^n$ in $\det\left[I-\lambda\left(K \text{ restricted to } J\times J\right)\right]$.

Now let $0 < \theta_1^* < \theta_2^* < \ldots$ be the angles taken in order. Obviously, the mean-value of $\theta_2^* - \theta_1^*$ is $2\pi/N$, by symmetry, so if you want to see what happens for infinite dimensions, i.e. for $N\uparrow\infty$, you'd better scale the angles up by the factor $N/2\pi$, as in

$$P\left(\#[0,2\pi L/N) = n\right) = \text{ the coefficient of } (1-\lambda)^n \text{ in}$$

$$\sum_{p=0}^{N}\frac{(-\lambda)^p}{p!}\int_0^{2\pi L/N}\cdots\int_0^{2\pi L/N}\det\left[\frac{1}{2\pi}\frac{\sin N(\theta_i-\theta_j)/2}{\sin(\theta_i-\theta_j)/2}\right]_{1\le i,j\le p}d\theta$$

$$=\sum_{p=0}^{N}\frac{(-\lambda)^p}{p!}\int_0^{L}\cdots\int_0^{L}\det\left[\frac{1}{2\pi(N/2\pi)}\frac{\sin\pi(x_i-x_j)}{\sin\pi(x_i-x_j)/N}\right]_{1\le i,j\le p}dx$$

$$\simeq\sum_{p=0}^{\infty}\frac{(-\lambda)^p}{p!}\int_0^{L}\cdots\int_0^{L}\det\left[\frac{\sin\pi(x_i-x_j)}{\pi(x_i-x_j)}\right]_{1\le i,j\le p}dx \qquad \text{for large } N$$

$$=\det\left[I-\lambda\frac{\sin\pi(x-y)}{\pi(x-y)}\right]_{0\le x,y\le L} = \det(I-\lambda K)$$

with

$$K(x,y) = \frac{\sin\pi(x-y)}{\pi(x-y)} \qquad \text{restricted to} \quad 0\le x,y\le L.$$

It follows that

$$\lim_{N\uparrow\infty}P\left(\#[0,2\pi L/N) = n\right)$$

exists for each $n\ge 0$, and it is a source of satisfaction that these numbers add to unity, i.e. they represent a genuine probability distribution. This is because $\det(I-\lambda K)$ reduces to 1 at $\lambda=0$.

Jimbo *et al.* [12] discovered the surprising fact that $\Delta = \det(I-\lambda K)$, as a function of L, solves a differential equation – and a famous one too, called Painlevé 5 – but that lies outside the scope and purpose of these notes. A nice account may be found in Tracy–Widom [22]. Here I note only one particular consequence: For $\lambda=1$, $\Delta(L)$ is just the probability that no (scaled) angle θ is found in the interval between 0 and L, and this has a Gaussian tail, *to wit*, $\Delta(L)\simeq\exp\left(-\pi^2 L^2/8\right)$.

Appendix on compactness

The space $C[0,1]$.

The statement is that the figure $F\subset C$ is compact relative to the distance $d(f_1,f_2)=\|f_1-f_2\|_\infty$ if (and only if) it is 1) closed, 2) bounded, and 3) $|f(b)-f(a)|\le\omega(|b-a|)$ for every $f\in F$ with some universal, positive, increasing function $\omega(h)$, $h>0$, subject to $\omega(0+)=0$. I sketch the proof.

Sufficiency.

Let $r_1, r_2, \ldots$ be a list of the (countably many) rational points $0 \le r \le 1$. The numbers $f_n(r_m)$ are bounded for fixed $m \ge 1$, so you can pick 1) a subsequence $\mathbf{n}_1 = (n_1, n_2, \ldots)$ of $\mathbf{n} = (1, 2, 3, \ldots)$ to make $f_n(r_1)$, $n \in \mathbf{n}_1$, tend to some number $f_\infty(r_1)$, 2) a further subsequence $\mathbf{n}_2 = (n_{21}, n_{22}, \ldots) \subset \mathbf{n}_1$, to make $f_n(r_2)$, $n \in \mathbf{n}_2$, tend to some number $f_\infty(r_2)$, and so on, and if you now pass to the diagonal sequence $\mathbf{d} = (n_{11}, n_{22}, n_{33}, \ldots)$, then $f_d(r_m)$ will tend to $f_\infty(r_m)$ for every $m = 1, 2, 3, \ldots$ simultaneously. 3) does the rest.

Necessity.

3) states that

$$\sup \left\{ |f(b) - f(a)| : f \in F, \, 0 \le a < b \le 1, \, (b - a) \le h \right\} \downarrow 0 \qquad \text{with } h.$$

Its negation asserts the existence of a number $\omega > 0$ of this nature: that for every $n \ge 1$, you can find $f_n \in F$ and $0 \le a_n < b_n \le 1$ with $b_n - a_n \le 1/n$ so as to make $|f_n(b_n) - f_n(a_n)| \ge \omega$. But if F were compact, you could weed out to make $f_n \to f_\infty$ and, at the same time, make a_n & b_n tend to a common value c, $[0, 1]$ being compact. Then $0 < \omega \le |f_n(b_n) - f_n(a_n)|$ tends to 0, which is contradictory.

Dini's Theorem, used twice in Section 2.7, belongs to the same circle of ideas. It states that if $f_n \in C[0, 1]$ decreases pointwise to 0, then $f_n \to 0$, *i.e.* $\|f_n\|_\infty \downarrow 0$.

Otherwise, for each $n \ge 1$, $f_n(x_n)$ exceeds some fixed number $\omega > 0$ for some $0 \le x_n \le 1$. Then you can weed out to make x_n tend to x_∞ as $n \uparrow \infty$, with the contradictory result that

$$f_m(x_\infty) = \lim_{n \uparrow \infty} f_m(x_n) \ge \lim_{n \uparrow \infty} f_n(x_n) \ge \omega,$$

i.e. $f_m(x_\infty)$ does *not* decrease to 0.

The space $L^2[0, 1]$.

Now comes the "weak" compactness of the unit ball $B = \left\{ e : \int_0^1 e^2 \le 1 \right\}$. The key to this is the fact that L^2 contains a countable dense family of functions f_n, $n \ge 1$, such as the broken lines with corners at rational places and rational slopes between. Take e_n, $n \ge 1$, from B. The diagonal method employed above permits you to weed out so as to make (e_n, f_m) tend, as $n \uparrow \infty$, to some number $L(f_m)$ for every $m = 1, 2, 3, \ldots$, simultaneously. Obviously, $|L(f_i) - L(f_j)| \le \|f_i - f_j\|_2$, whence L extends to a bounded linear map from $L^2[0, 1]$ to $\mathbb{R}$ and (e_n, f) tends to the extended $L(f)$ for every $f \in L^2$, simultaneously. The final step employs the fact that any such L is of the form $L(f) = (e_\infty, f)$ with a fixed function $e_\infty \in L^2[0, 1]$. The proof is easy. Let N be the null space of L and N^o its annihilator. Then either $L(f) \equiv 0$, i.e. $e_n \rightsquigarrow 0$ or else $L(e) \ne 0$ for any non-vanishing $e \in N^o$. Pick such an $e = e_1$. Then for any other $e_2 \in N^o$, L sends $L(e_1)e_2 - L(e_2)e_1$ to 0 from which you infer that dim $N^o = 1$, i.e, $N^o = \mathbb{R}e$ with $\|e\|_2 = 1$ if you want. Now split f into its projection $e(e, f)$ onto N^o plus the rest, viz. $f - e(e, f) \in N$. Then $L(f) = L(e)(e, f) \equiv$ which is to say $e_n \rightsquigarrow L(e) \times e$. This is the weak compactness of the unit ball.

That's all you need to know about compactness, but I will add a simple

Exercise.

The functions $e_n = \sqrt{2} \sin n\pi x$, $n \ge 1$, form a unit perpendicular base of $L^2[0, 1]$, i.e. every $f \in L^2$ can be expanded as $f = \sum \hat{f}(n) \, e_n$ with $\hat{f}(n) = (e_n, f)$. Prove that the ellipsoid $E = \left\{ f : \sum |\hat{f}(n)|^2 / l_n^2 \le 1 \right\}$ is (truly) compact in $L^2[0, 1]$ if and only if the positive numbers l_n^2 tend to 0 as $n \uparrow \infty$.

References

[1] Ahlfors L.V., Complex Analysis, Internat. Ser. Pure Appl. Math., 3rd ed., McGraw-Hill, New York, 1979

[2] Brown R., A brief account of microscopical observations made in the months of June, July, and August, 1827, on the particles contained in the pollen of plants, etc., Philosophical Magazine, 1828, 4, 161–173

[3] Cameron R.H., Martin W.T., The Wiener measure of Hilbert neighborhoods in the space of real continuous functions, Journal of Mathematics and Physics Mass. Inst. Tech., 1944, 23, 195–209

[4] Conrey J.B., The Riemann hypothesis, Notices Amer. Math. Soc., 2003, 50(3), 341–353

[5] Courant R., Differential and Integral Calculus, vol. 2, Wiley Classics Lib., John Wiley & Sons, New York, 1988

[6] Courant R., Hilbert D., Methoden der Mathematischen Physik, Springer, Berlin, 1931

[7] Dyson F.J., Statistical theory of the energy levels of complex systems. I, J. Mathematical Phys., 1962, 3, 140–156

[8] Dyson F.J., Fredholm determinants and inverse scattering problems, Comm. Math. Phys., 1976, 47(2), 171–183

[9] Einstein A., Über die von der molekularkinetischen Theorie der Wärme gefordete Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, Ann. Phys., 1905, 17, 549–560; reprinted in: Investigations on the Theory of the Brownian Movement, Dover, New York, 1956

[10] Fredholm I., Sur une classe d'équations fonctionelles, Acta Math., 1903, 27(1), 365–390

[11] Gardner C.S., Greene J.M., Kruskal M.D., Miura R.M., Methods for solving the Korteweg–deVries equation, Phys. Rev. Lett., 19(19), 1967, 1095–1097

[12] Jimbo M., Miwa T., Môri Y., Sato M., Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent, Phys. D, 1980, 1(1), 80–158

[13] Kato T., A Short Introduction to Perturbation Theory for Linear Operators, Springer, New York–Berlin, 1982

[14] Lamb G.L., Jr., Elements of Soliton Theory, Pure Appl. Math. (N.Y.), John Wiley & Sons, New York, 1980

[15] Lax P.D., Linear Algebra, Pure Appl. Math. (N.Y.), John Wiley & Sons, New York, 1997

[16] Lax P.D., Functional Analysis, Pure Appl. Math. (N.Y.), John Wiley & Sons, New York, 2002

[17] Lévy P., Le Mouvement Brownien, Mémoir. Sci. Math., 126, Gauthier–Villars, Paris, 1954

[18] McKean H.P., Jr., Stochastic Integrals, Probab. Math. Statist., Academic Press, New York–London, 1969

[19] Mehta M.L., Random Matrices, 2nd ed., Academic Press, Boston, 1991

[20] Munroe M.E., Introduction to Measure and Integration, Addison–Wesley, Cambridge, 1953

[21] Pöppe Ch., The Fredholm determinant method for the KdV equations, Phys. D, 1984, 13(1-2), 137–160

[22] Tracy C.A., Widom H., Introduction to random matrices, In: Geometric and Quantum Aspects of Integrable Systems, Scheveningen, 1992, Lecture Notes in Phys., 424, Springer, Berlin, 1993, 103–130

[23] Uhlenbeck G.E., Ornstein L.S., On the theory of the Brownian motion, Phys. Rev., 1930, 36, 823–841

[24] Weyl H., The Classical Groups, Princeton University Press, Princeton, 1939

[25] Whittaker E.T., Watson G.N., A Course of Modern Analysis, 4th ed., Cambridge University Press, New York, 1962

[26] Wiener N., Differential space, Journal of Mathematics and Physics Mass. Inst. Tech., 1923, 2, 131–174

[27] Wigner E., On the distribution of the roots of certain symmetric matrices, Ann. of Math., 1958, 67, 325–326