Stakgold Chapter 3 Meta Notes PhL 3.17-21.09

History **1**
3.1 Introduction with Six Examples (191) **1**
3.2 The Neumann Series Method (206) **4**
3.3 The Spectrum of a self-adjoint H-S operator. (p 212) **5**
3.4 The solution of Ku = μu + f for K symmetric and H-S. (220) **7**
3.5 Extremal Principles (223) **8**
Courant minimax principle: 9
3.6 Approximation Methods Based on Extremal Principles (226) **9**
The Rayleigh-Ritz Procedure (228) 9
The Schwarz Iteration Procedure (p 231) 10
Upper Bounds to Eigenvalues 10
3.7 Continuity and Uniform Convergence: bilinear series and iterated kernels **11**
3.8 Approximation Methods for Solving Integral Equations **12**
3.9 Non-symmetric HS operators (250) **14**

History (verbatim**)**: Fredholm is best remembered for his work on integral equations and spectral theory. In fact much of this work was accomplished during the months of 1899 which Fredholm spent in Paris studying the Dirichlet problem with **Poincaré**, Emile **Picard**, and **Hadamard**. In 1900 a preliminary report on his theory of Fredholm integral equations was published as *Sur une nouvelle méthode pour la résolution du problème de Dirichlet*. **Volterra** had earlier studied some aspects of integral equations but before Fredholm little had been done. Of course **Riemann**, **Schwarz**, Carl **Neumann**, and Poincaré had all solved problems which now came under Fredholm's general case of an integral equation; this was an indication of how powerful his theory was. Fredholm's contributions quickly became well known to the world of mathematics when **Holmgren** lectured on Fredholm's theory at Göttingen in 1901. **Hilbert** immediately saw the importance of Fredholm's theory, and during the **first quarter of the 20th century the theory of integral equations became a major research topic**. Fredholm published a fuller version of his theory of integral equations in *Sur une classe d'équations fonctionelle* which appeared in Acta Mathematica in **1903**. Hilbert extended Fredholm's work to include a complete **eigenvalue theory** for the Fredholm integral equation. This work **led directly to the theory of Hilbert spaces.**

3.1 Introduction with Six Examples (191)

I think the only possible linear integral equations (with one integral) that you can have are **Fredholm Equations** and **Volterra Equations**, the only difference being in the nature of the upper integration endpoint. Stak starts talking Fredholm without explaining the why and wherefore of the term. The integration variable is always a single variable like "x", never anything like d^3r.

Since we don't have any real theory on *unbounded* integral equation operators, we won't care about them and we shall care only about **bounded** K operators.

Within the world of bounded linear integral equation operators, we shall only be interested in a subclass known as **Hilbert-Schmidt** operators which, by definition, have a finite double integral of the kernel, even though the kernel itself might not be bounded on the square. The reason we like this H-S

subclass is that such operators are **compact**, which is the modern term for Stakgold's **"completely continuous"**. This compact property of mapping a bounded function space into a compact function space is the key item, though it is a bit obscure to the practical mind. Recall that if you map a sequence in the domain to one in the range, the one in the range will always have a convergent subsequence if the operator is compact. This is a generalization of "boundedness" that is a necessary condition for the analogy between the finite dimensional matrix world to carry over well into the infinite dimensional matrix world of k(x,y). It is quite easy to understand this chapter without dwelling on this point.

Here, by the way, is the matrix analogy

$$\mu\, u(x) = \int dy\, k(x,y)u(y) \quad \Rightarrow \quad \mu\, u_{\mathbf{x}} = S_{\mathbf{x}}\, k_{\mathbf{x},\mathbf{y}}\, u_{\mathbf{y}} \quad \Rightarrow \quad \mu\, u = K\, u$$

The main thrust of this section is this: We first look at the notion of a **separable kernel**, which means that $k(x,y) = \Sigma^{\mathbf{n}}\, p_{\mathbf{i}}(x)\, q_{\mathbf{i}}(y)$ where the sum is *finite*, and the functions are L_2. He shows that such a kernel is **bounded and compact** and is in fact a HS kernel and $||K|| \leq$ double integral of k. He then shows that as you take the limit $n \to \infty$, two things happen: (1) the kernel remains compact; (2) any H-S kernel can be written in this "infinite sum separable" form. Therefore, a H-S operator is compact, just like a separable one. The step (2) follows basically from the fact that you can expand a function on the square in terms of a double complete basis expansion.

We then on page 195 look at three **variations of the Fredholm equation**. The "**second kind**" one has the form $Ku = \mu u + f$ and is the most general form, μ is a number, f is a function, you want to solve for u. If you set $\mu = 0$, you get the **first kind** $Ku = f$. Both the first and second kind are "**inhomogeneous**" because there is what I like to call a driving function f. The third equation has $f = 0$ so $Ku = \mu u$ and this is of course the **eigenvalue** problem equation, called **homogeneous**. The homo and inhomo sense is best understood if you think of $B = (K-\mu)$ as the operator, then inhomo2 says $Bu=f$ and homo says $Bu=0$.

The **Volterra** variant can be handled by adjusting the kernel and gets minor treatment in this book.

The rest of this introductory section looks at **6 detailed examples** of Fredholm equations. Each example considers a particular simple kernel, and at least one of the Fredholm equation forms. We try to solve for the eigenvalues in some cases, and see how many solutions there are, if any. We are told that the solutions for all eigenvalues form a complete set if k is symmetric. Sometimes there might be only a single solution for a given function f, or there might be more solutions by adding solutions to the homogeneous equation if μ is an eigenvalue. There might be no solutions to an equation. We are always talking about solutions u and functions f which are L_2. We are always interested in knowing how to classify the values of μ in terms of which spectral group they belong to (or the resolvent set).

Chapter 1 which I have not read recently talks about **Green's function** solutions to ODE's. You can recast an ODE as a Fredholm integral equation which incorporates the BC's of the ODE (since the Green's function does this for you). In this recast integral equation, **the kernel is just the Green's function of the ODE.** The operator in the ODE is unbounded, while that in the integral equation is bounded, and this is one reason one might work with the integral form instead of the differential form of the equation.

Note added 4.22.09. I have now read Chapter 1. In the above paragraph, if we have $Lu = \mu u$ with some BC's, and if $Lu=0$ has no non-trivial solutions, we know that $Lg=\delta$ has a unique solution g that meets the BC's and then $Lu=f$ with the same BC's has the solution $u = Gf$ where G is the integral operator of kernel g. If L is self-adjoint, then g is symmetric. If we take $f = \mu 1$, then $Lu = \mu u$ becomes $u = \mu Gu$ which is then an integral EV equation for u. The solutions u of this equation respect the BC's. This is why one says that the ODE eigenvalue equation with BC's can be cast as an integral equation eigenvalue problem which

"builds in" the same BC's. Since the integral operator is "nicer" (bounded), maybe best to study the EV problem there instead of as an ODE, at least for certain purposes such as obtaining the general nature of the EV's.

The examples are so important, that I feel each needs some comment here. Examples 5 and 6 involve the Green's function stuff. Later I think we learn that only EV $\mu=0$ can have an infinite multiplicity. Range restrictions are called **consistency conditions**. Symmetric kernels yield complete EF sets.

Example 1: $k(x,y) = xy$ (symmetric) on (0,1). Illustration of the Alternative Theorem:

(a) $Ku = \mu u$. There are only two eigenvalues. The one $\mu=1/3$ has one normalized eigenfunction Cx and so is a simple eigenvalue. The other $\mu=0$ has an infinite multiplicity and the eigenfunction is anything orthogonal to x. Point spectrum has 2 points. The eigenfunctions form a complete set.

(b,c) ($K-\mu)u = f$. For $\mu\neq 1/3,0$, the single solution is constructed for any f, so full range, no nullspace. For $\mu=1/3$, solution only if $\langle f,x\rangle=0$, range is restricted, 1D nullspace exists, alternative theorem! For $\mu=0$, range now hugely restricted to $f = \alpha x$, nullspace ∞D, solution is now $u = 2\alpha$ + any u such that $\langle u,x\rangle=0$.

Example 2: $k(x,y) = \sin(x)\sin(y) + \theta\cos(x)\cos(y)$ (symmetric) on $(0,\pi)$. $Ku = \mu u$. The eigenvalue $\mu=0$ has an infinite multiplicity, while $\mu = \pi/2$ is simple and $\mu=\theta\pi/2$ is also simple. If $\theta=1$, these last two merge and then $\mu=\pi/2$ has degeneracy 2.

Example 3: $k(x,y) = \sin(x)\cos(y)$ (non-symmetric) on $(0,\pi)$. $Ku = \mu u$. Only EV is $\mu=0$, infinite multiplicity, but cos(x) missing from set, so EF set is not complete (due to non-symmetric). If $k(x,y) = p(x)\,q(y)$ then $\mu = \int dy\, q(y)\, p(y)$. If this $\mu\neq 0$, then $C\,p(x)$ is an EF for this μ EV. And $\mu=0$ is also an EV. If q=p, k is symmetric, and $\mu=0$ will I think have an ∞ multiplicity and you get complete set. See next.

Example 4: $k(x,y)$ = separable $= \Sigma_{i=1}^{n}\, p_i(x)\, q_i^*(y)$. $Ku = \mu u$. For $\mu \neq 0$, the eigenvalue problem converts to an nxn *matrix* eigenvalue problem shown on the left side of 3.8. At once we assume that k is symmetric so $q_i = p_i$. There will be n eigenvalues μ_i (not necessarily all different, but none 0), and each will have an independent eigenvector c_i which corresponds to a solution of the original eigenvalue problem $Ku = \mu u$. So this gives us n eigensolutions of our problem. For $\mu = 0$, the solutions u(x) are all functions which are orthogonal to ALL the n functions p_i^* . If I think of the first n Legendres, say, I then have u(x) being all Legendres with ℓ>n. So in this problem, we first have the n basis functions associated with $\mu_i \neq 0$, and these are linear combinations of the $p_i(x)$. Then we have sort of ∞-n eigenfunctions associated with $\mu = 0$, and combined, we have a full basis. Again, this seems always to happen when the kernel is symmetric.

Example 5: $k(x,y)$ = sym of $x(1-y)$ on (0,1), $Ku = \mu u$. This is the string ODE problem $L = -D^2$ and k is the Green's for this string, symmetric. We know the EV's and EF's from the ODE, and these apply also to the associated integral equation problem, so we know everything. Now look at $Ku = \mu u + f$. For general μ, there is a unique solution obtainable by Fourier method. For μ = EV, range is restricted to $\langle f,EF\rangle = 0$. If $\mu = 0$, there are complicated restrictions on the range interpretable in terms of the string problem.

Example 6: $k(x,y) = e^{-|x-y|}/2$ = Green's for $L = -D^2 + 1$ on $(-\infty,\infty)$ (symmetric, but not HS). I turns out K is still bounded ($k \leq 1$) , so here is an example of what I was wondering above! Due to expo problem, the EV equation has no eigenvalues at all, so point spectrum is empty. Looking then at $Ku = \mu u + f$, as long as μ not on (0,1), a solution exists for any f, and such μ are regular and in the resolvent set. It turns out that the little range (0,1) is the continuous spectrum for this problem, again relating an ODE to integral equation.

3.2 The Neumann Series Method (206)

We want to solve our inhomo Fred 2 $Ku = \mu u + F$. Divide through by μ to get $(1/\mu)\, Ku = u + F/\mu$. Redefine the last term to be -f, so have $(1/\mu)\, Ku = u - f$. Set $\lambda \equiv 1/\mu$ so have $\lambda Ku = u - f$. This then says that $f = u - \lambda Ku = (1-\lambda K)u$. If $(1-\lambda K)^{-1}$ exists, then we can solve to get $u = (1-\lambda K)^{-1} f$. This type of solution is often used in scattering and is the basis of all of QED, and really all scattering theory, it is pretty major.

We can write a formal solution as

$$u = (1 - \lambda K)^{-1} f = (1 + \lambda K + \lambda^2 K^2 + \ldots.)\, f \qquad 3.25 \qquad k_2(x,z) = \int dz'\, k(x,z')k(z',z) \text{ etc}$$

typical term in series: $u(x) = u_x = \lambda^2 [K^2 f]_x = \lambda^2 \int dz\, [K^2]_{xz} f_z = \lambda^2 \int dz\, k_2(x,z) f(z)$

So, to get a solution using this series, you first have to compute all the objects $[K^n]_{xz} = k_n(x,z)$ which means doing a bunch of integrals only using the kernel k(x,z), and second you have to integrate each of these iterated kernels against the driving function f(z), then add up all the terms. It is a prescription and it gives a definite and unique solution u.

Notice that if λ is an inverse eigenvalue, then $Ku = \mu u \Rightarrow \lambda Ku = u \Rightarrow (1-\lambda K)u = 0$, where u is an eigenfunction. Now suppose the series for $(1-\lambda K)^{-1}$ shown above converges. Then if λ is an inverse eigenvalue, we get $u = (1-\lambda K)^{-1} 0 = 0$. But this is a contradiction because u is supposed to be an eigenfunction. Thus, the series must not converge if λ is an inverse eigenvalue, or λ EV => not converge. Therefore series converges $\Rightarrow$ λ not EV.

The series above *is* the Neumann Series and, when it converges, it is the unique solution. In fact, it always converges when $|\lambda| < 1/\|K\|$, a disc in the λ plane (again, K is bounded we assume). According to the last sentence of the last paragraph, there can be no eigenvalues inside this circle! This translates into saying that $Ku = \mu u$ must have its eigenvalues inside the disc $|\mu| \leq \|K\|$. This is consistent with our idea later that EV's are bounded by $\|K\|$. Neumann downside: often not easy to compute the K^n.

Note that the formal operator $(1 - \lambda K)^{-1}$ is called the **resolvent**. (Stak no mention).

Example 1: $k(x,y) = xy$. We find $k^n(x,y) = xy(1/3)^n$ and $\|K\| = 1/3$ so $|\lambda| < 3$ is disc of convergence for the series. Recall from example 1 earlier that $\mu = 1/3, 0$ which are both inside $|\mu| \leq 1/3$. The formal series 3.26 can be summed to give 3.27 which then provides an analytic continuation of the series outside the disc $|\lambda| < 3$, except for the single point $\lambda = 3$ (a typical situation with analytic continuation).

Example 2: The Volterra Form. This is the only section in this book on this subject I think. If K is HS, the first thing we learn is that $Ku = \mu u$ has no solutions, but Stak only proves this for $|k| < M$. The proof involves inequalities which make use of the Volterra endpoint such as $\int_a^x \xi^{n-1}\, d\xi = (x-a)^{n-1}/(n-1)!$ as

you see in p 209A. In the special case $\mu=0$, again there are no EF's unless $k(x,x) = 0$ and Stak says mysteriously that this is related to the subject of singular points of the corresponding ODE.

Now he moves on to $Ku = \mu u + f$. Again because of the Volterra upper endpoint, it now turns out that the Neumann series converges for $|\lambda| < \infty$, whereas a normal Fredholm case has $|\lambda| < 1/\|K\|$, so the Volterra Form is very happy with a Neumann series solution. Here is an application:

Exercise 3.10. Stak starts with a completely general inhomo second order ODE $Lu=f$ with the usual BC's at one point $x=a$ (initial-value). He is then able to recast this as a Volterra Form integral equation for $u''(x)$. Since this always has a unique convergent Neumann series solution (no problem with convergence radius as found above), there is a unique solution $u''(x)$. Then integrate twice and use the two BC's and get unique $u(x)$. Thus he *proves* that such an $n=2$ ODE system has a unique solution! [An example of how you can "do something" in the integral equation world that gives an answer to a question in the ODE world. Recall from Chapter 1 that there is a big theorem about n^{th} order ODE having n solutions, but Stak never proved that theorem in Chap 1. Maybe this is that proof if you extend to general n.]

3.3 The Spectrum of a self-adjoint H-S operator. (p 212)

Luckily, if $k(x,y) = \overline{k}(y,x)$ [ie, k is Hermitian = a "math-symmetric" matrix] then K is self-adjoint with no further complications, and this directly says $\langle Ku,u\rangle$ = real. Stak then shows for symmetric K:

(1) The eigenvalues of K are all real
(2) the EF's of different EV's are orthogonal
(3) the multiplicity of any *non-zero* EV μ is finite.
(4) if number of EV's is ∞, then $\mu=0$ is the only possible limit point.

but the last two items also require that K be HS.

At this point we enter a long saga which concerns upper bounds on eigenvalues. It is trivial to show for a bounded K that $|\mu_i| \le \|K\|$ (see raw notes). But since the HS K is compact, we get this extra fact:

Theorem 6 of page 190 Chap 2: If A is symmetric *and* completely continuous (compact), then:
(1) At least one of these numbers *is* an eigenvalue: $\|A\|$ or $-\|A\|$.
(2) There is no eigenvalue μ for which $|\mu| > \|A\|$.

So compactness of K says we actually *hit* $\|K\|$ with the largest eigenvalue (in abs value). This then is why we are interested in K being H-S, since H-S implies K is compact.

Now, recall that in my notation,

$$\| K \|^2 = \max_u \langle Ku,Ku\rangle/\langle u,u\rangle \quad \Rightarrow \quad \| K \| = \max_u \|Ku\| / \|u\|$$

$$||| K |||^2 = \max_u \langle Ku,u\rangle/\langle u,u\rangle$$

When the dust settles in my raw notes, we learn using Theorem 6 that for a symmetric HS operator,

$$\| K \| = ||| K ||| = |\mu_1| \qquad \text{// this is } \textbf{Theorem 1} \text{ on page 215}$$

where again we *hit* the max eigenvalue μ_1. This says there are two variational problems you could work on to try to approach up to this eigenvalue from below, and the one $|||\ K\ |||\ = |\mu_1|$ is computationally simplest. This also tells you that the symmetric HS EV problem must have at least one EV, namely, μ_1.

At this point we run through a set of theorems 2, 2A, 3, 3A. In his proof method of these theorems, he constructs new operators with kernels $k^{(n+1)} = k(x,y) - \Sigma_{i=1}^{n}\mu_i\varphi_i(x)\bar{\varphi}_i(y)$ where he is thus subtracting off products of "the higher solutions" while working on the nth solution. These new kernels are symmetric of course, and allow him to sort of "recur" his theorems. The upshot of all this is really stated in the last theorem, Theorem 3A

Theorem 3A: Subject to $\langle u,\varphi_i\rangle=0$ for i=1..n, $\max_u [\ \langle Ku,u\rangle/\langle u,u\rangle] = |\mu_{n+1}|$ and gives φ_{n+1}.

This suggests an iterative process using an idealized variation method. You first vary u with no conditions to find $|\mu_1|$ and then, in a perfect world, you know μ_1 and φ_1. Then you vary u subject to $\langle u,\varphi_1\rangle = 0$ and this gives you μ_2 and φ_2 , and you just keep going. You need some kind of "perfect varying machine" to make this go, and I suspect the computer now provides such a machine!

If our varying machine is imperfect, we get some value $a_k \le |\mu_k|$ because we did not reach the perfect limit. In this case, the thing we get a_k is a **"lower bound"** for $|\mu_k|$, the thing we are seeking.

For non-zero degenerate eigenvalues, our "machine" must somehow be able to find the finite number of degenerate eigenfunctions.

How does this iterative process end? If $\mu=0$ is a eigenvalue limit point, you go on forever not getting there. If $\mu=0$ is an eigenvalue, then the sequence $k^{(n+1)}$ no longer generates new kernels and your "machine" somehow has to "sweep out" all the nullspace eigenfunctions.

Stak then claims that (1) this procedure will find all non-zero EV's and EF's. (2) if you combine these with the EF's of $\mu=0$, you get a complete set.

We now arrive at this strange intermediate theorem (always talking symmetric and HS)

Theorem 4: If you take the EF's you get from the above procedure but *only include* the EF's from non-zero EV's, the set of EF's you then have is a complete basis for any function g that can be written g = Kf. In other words, this set of EF's spans the range of K [f of course is L_2 as usual]. So this explains how you could have a "basis" that might only have a finite number of EF's -- if the range of K is finite dimensional. // This is pretty obvious since the zero-eigenvalue is associated with the nullspace.

We then come to the main result:

Theorem 5: The full set of EF's of a symmetric HS kernel is complete on L_2.

Here you include the EF's above *and* the EF's of the nullspace $\mu=0$. He writes general f = h + g where g is in the range of K and h is in the nullspace of K. He writes g as a Fourier sum on the non-zero EV EF's which is calls φ_i.

Theorem 5A: Obviously, if $\mu = 0$ is not an EV, then the Theorem 4 set of EF's is complete on L_2.

Now why have we been pursuing these obscure theorems? Here is the reason: Stak takes an arbitrary self-adjoint ODE $Lu = \mu u$. Note that self-adjoint ODE's are going to be the central topic of Chapter 4! He first recasts this ODE into a new form $My = \theta y$ which is guaranteed NOT to have $\theta=0$ be an EV. He then recasts this ODE into an integral equation where, as usual, the kernel is the Green's function of the ODE including the BC's. He claims this kernel is both symmetric and H-S. We now have an integral equation which says $Ku = \mu u$ where $\mu = 1/\theta$. We know that $\mu=\infty$ cannot be an eigenvalue so things are reasonable and bounded as expected. But then he shows that $\mu=0$ is also not an eigenvalue. Therefore, the EF's of $Ku = \mu u$ (for non zero eigenvalues) form a complete set. Therefore he has proven that the eigenfunctions of any self-adjoint ODE form a complete basis on L_2. I think this is a major result and justifies our usual claims of "completeness" or "closure" of things like the Legendre polynomials. This claim is the content of **Theorem 6**.

Catch 22: Stak has never shown that a Green's Function kernel is H-S! He is sneaking this in.

3.4 The solution of Ku = μu + f for K symmetric and H-S. (220)

This section discusses solving $Ku = \mu u + f$ under the assumption that you have already solved the equation $K\varphi_i = \mu_i\varphi_i$ for the non-zero eigenvalues and eigenfunctions. The inhomo solutions here are generally simple Fourier series type solutions some of which always converge. The real problem I think is solving the eigenvalue problem!

Stak sets $u = \Sigma_i a_i\varphi_i$, $f = \Sigma_i b_i\varphi_i$ and we have $Ku = \Sigma_i a_i K\varphi_i = \Sigma_i a_i\mu_i\varphi_i$ where the φ_i are only those of non-zero EV's of the EV equation. Although the u series may be incomplete, the Ku result is complete as I show in the raw notes. We have Ku = g and the series for g we know is also complete because it is in the range of K, and its series is $\Sigma_i(\mu a_i + b_i)\varphi_i$. Setting these equal gives:

$$a_i(\mu_i - \mu) = b_i$$

where the b_i describe f, and then the a_i describe our solution u. Stak now breaks things into 4 cases:

Case 1: $\mu \neq 0$ and $\mu \neq$ EV. Here we are fine with $a_i = b_i/(\mu_i - \mu)$ and we get a constructed solution. Stak then shows using various useful past theorems that this series for u converges. Since $\mu \neq$ EV, we cannot add to this constructed solution any nullspace solutions of $(K-\mu)u = 0$, and thus the solution is unique. This is the content of **Theorem 1** on page 221.

Case 2a: $\mu \neq 0$ and $\mu = \mu_m$ a *simple* EV. If $b_m \neq 0$, then $a_m = \infty$ and does not exist so $Ku = \mu_m u+f$ has no solution. But if $b_m = 0$, which means $\langle f,\varphi_m\rangle = 0$, then a_m can be anything and we get

$$u = -f/\mu_m + \{ a_m\varphi_m + \Sigma_{i\neq m} b_i \mu_i /\mu_m(\mu_i - \mu)\varphi_i \} \qquad a_m = \text{arbitrary} \quad \langle f,\varphi_m\rangle = 0$$

which is particular solution plus nullspace solution of $(K-\mu_m)u = 0$. As usual, non-trivial nullspace goes with a restriction on the range, here $\langle f,\varphi_m\rangle = 0$, a consistency condition.

Remember that for a self-adjoint operator, we don't have to worry about the distinction between A and A* and the (∞ dim) alternative theorem says that $\overline{R_A} = (N_{A^*})^{\perp}$ and $(R_A)^{\perp} = N_{A^*}$ so as the nullspace grows, the range shrinks, even though the sum of the nullity and rank is not a finite number as it is in the matrix case.

Case 2b: $\mu \neq 0$ and $\mu = \mu_m$ an *N degenerate* EV. If any $b_m^{(i)} = 0$, no solution as above. If all $b_m^{(i)} = 0$, then similar solution to the above. So now we have this solution and N consistency conditions:

$$u = -f/\mu_m + \{ \Sigma_i \; a_m^{(i)} \varphi_m^{(i)} + \Sigma_{i \neq m} b_i \mu_i / \mu_m(\mu_i - \mu)\varphi_i \} \quad a_m^{(i)} = \text{arbitrary} \quad \langle f, \varphi_m^{(i)} \rangle = 0$$

Now the range is N constricted and the nullspace $(K-\mu_m)u = 0$ has N solutions. Alternative theorem!

Case 3: $\mu = 0 \neq$ EV. Then $a_i = b_i/\mu_i$ and $u = \Sigma_i (b_i/\mu_i)\varphi_i$ but it may not converge, especially if $\mu=0$ is a limit point, since we have μ_i in the denominator!

Case 4: $\mu = 0 =$ EV. Now we have $Ku = f$ and we know there are some $K\varphi_o^{(i)} = 0$ which are EF's for $\mu=0$. We again have $a_i = b_i/\mu_i$ for other eigenvalues $u_i \neq 0$, and $u = \Sigma_{i \neq 0} (b_i/\mu_i)\varphi_i$. As above, we must have $b_0^{(i)} = 0$ for all $\varphi_o^{(i)}$ in order to even have a solution, so we require $\langle f, \varphi_0^{(i)} \rangle = 0$. I think the solution now in this case is this:

$$u = \Sigma_i \; a_0^{(i)} \varphi_0^{(i)} + \Sigma_{i \neq 0} (b_i/\mu_i)\varphi_i \qquad a_0^{(i)} = \text{arbitrary} \quad \langle f, \varphi_0^{(i)} \rangle = 0$$

and as in Case 3 we are likely to have series convergence problems.

3.5 Extremal Principles (223)

Back in Section 3.3 we developed the following two ideas:

$\| K \| = ||| K ||| = |\mu_1|$

Theorem 3A: Subject to $\langle u, \varphi_i \rangle = 0$ for i=1..n, $\max_u [\langle Ku,u \rangle / \langle u,u \rangle] = |\mu_{n+1}|$ and gives φ_{n+1}.

In this section, we are going to see what these things look like if we get rid of the absolute values around the eigenvalues. We start by listing the eigenvalues in increasing order, negatives on the left:

$$\mu_1^- \; \mu_2^- \quad \ldots\ldots\ldots (0) \quad \ldots\ldots \mu_2^+ \; \mu_1^+$$

The above two ideas then translate into the following:

$||| K ||| = \mu_1^+$ // 3.47

Subject to $\langle u, \varphi^+_i \rangle = 0$ for i=1..n, $\max_u [\langle Ku,u \rangle / \langle u,u \rangle] = \mu_{n+1}^+$ and gives φ^+_{n+1}. ` // 3.48

Subject to $\langle u, \varphi^-_i \rangle = 0$ for i=1..n, $\min_u [\langle Ku,u \rangle / \langle u,u \rangle] = \mu_{n+1}^-$ and gives φ^-_{n+1}. // 3.50

When you work on the negative side, the max turns into min as you see.

Comment: You can see generally how the above works. If you have computer scan through the entire L_2 function space, your <Ku,u> will eventually hit the max value μ_1^+ when it finds φ_1^+. And if you then scan orthogonal to this, the max in that case will be φ_2^+ which will hit μ_2^+ , and so on. On the negative side, a scan with no conditions looking to minimize <Ku,u> will hit μ_1^- when it finds φ_1^- , and so on. It really is quite simple and reasonable, and of course suggests a search method for finding eigenvalues and eigenfunctions. The trick is to select a smart trial function parameterization.

We then have some definitions for **qualities of K**: +, non-, -, non+, indefinite, real

Def: if all eigenvalues are non-negative, then <Ku,u> ≥ 0 and K is said to be **non-negative.**
Def: if all eigenvalues are positive, then <Ku,u> > 0 and K is said to be **positive**.
Def: if all eigenvalues are non-positive, then <Ku,u> ≤ 0 and K is said to be **non-positive**.
Def: if all eigenvalues are negative, then <Ku,u> > 0 and K is said to be **negative**.
Def: Otherwise, the operator K is said to be **indefinite**.
Def: Given u = real, if Ku = real, then K is said to be **real**.
Fact: For integral operator, K is real iff k is real.

Courant minimax principle: Assume K is non-negative, real, and symmetric. You pick a set of n *arbitrary* functions ψ_n against which you force your trial functions u to be orthogonal. Compute $\max_u$ [<Ku,u>/<u,u>] and record the result. Then go pick 1,000 other sets of ψ_n and do the same thing and record all the results. The theorem says that $\min_\psi$ { $\max_u$ [<Ku,u>] } = μ_{n+1} which is quite amazing. The only use of "n" is that n is the number of functions in each ψ set. We do already know that if you pick "the right functions" for ψ, then $\max_u$ [<Ku,u>/<u,u>] = μ_{n+1} if K is non-negative (in which case abs value not needed). So when you pick a "wrong set", you get $\max_u$ [<Ku,u>] > μ_{n+1}. The proof is pretty simple and appears in the raw notes.

3.6 Approximation Methods Based on Extremal Principles (226)

We work in this Section only with K that are: real, symmetric, H-S and non-negative. The non-negative part lets us deal directly with eigenvalues instead of absolute values of eigenvalues.

Stak shows that, although you can construct many functionals F(u) which have the property F(φ_1) = μ_1, R(u) = <Ku,u>/<u,u> is better than most because it is *stationary* when u = φ_1 as shown in the page 227 figure. This is of course just our ||| K ||| operation. Here the ratio is finally given a name :
R(u) = **The Rayleigh Quotient**.

The Rayleigh-Ritz Procedure (228)

The goal is to try to solve the EV problem $K\varphi_i = \mu_i\varphi_i$. We don't even know how many eigenvalues there are. But we pick n arbitrary but independent functions and call them v_i and we make trial eigensolutions $u = \Sigma c_i v_i$ with some real coefficients **c**. R(u) is called R*. We then vary the c_i so as to make stationary R(u) which we know should get us close to μ_1. The procedure yields a little nxn matrix equation A**c** = 0

where **c** are our coefficients. There can be a solution if detA = 0, and this is a polynomial of degree n in R*. We then use the largest root R^*_1 as our resulting estimate (or lower bound) for μ_1, knowing $R^*_1 \leq \mu_1$, and the resulting **c** as our estimate for $\varphi_1 = \Sigma c_i v_i$. It is also claimed that $R^*_2 \leq \mu_2$ and so on, but these lower solutions are far less accurate. Stak then has several remarks: (1) accuracy of μ_1 will be much better than accuracy of φ_1, due to the stationary idea. (2) accuracy goes down rapidly for lower solutions like μ_2, μ_3 ...; (3) larger n gives more accuracy; (4) good to select v_i based on symmetry. (5) Things are simpler if you use orthonormal v_i, but you don't have to. (6) if you could use $n = \infty$, you would get exact results since your v_i would be a complete basis, but of course this is not practical.

Stak then gives a single good example of using the R-R procedure on page 231.

In the nuts and bolts of R-R, you first "precompute" the integrals $K_{ik} = \langle Kv_i, v_k \rangle$ and $\alpha_{ik} = \langle v_i, v_k \rangle$. For some complicated K, this can only mean numerical integration, so we are talking computer work here, not analytic stuff. Then you compute the n x n determinant shown in 3.55, and then you have to find the roots of your polynomial. If you pick orthonormals, then $\alpha_{ik} = \delta_{ik}$ and the determinant is the much simpler one shown in 3.57. Matrix K will be symmetric.

Basically, we are making an n x n matrix approximation to our ∞ x ∞ matrix problem.

The Schwarz Iteration Procedure (p 231)

Again, the goal is to solve the EV problem $K\varphi_i = \mu_i\varphi_i$. Recall that we used iterated kernels K^n in the Neumann series method for solving the inhomo equation written as $u = f + \lambda Ku$ and we found that the series was $u = (1 - \lambda K)^{-1}f = (1 + \lambda K + \lambda^2K^2 + ...) f$. In the Schwarz method, we again make use of these iterated kernels K^n, but they are used differently and of course for the homo equation. The idea is to start off with some seed function f_0 and then compute $f_n = K^n f$. If you are lucky, this sequence will become stable after a while and you get $f_{n+1} \approx \alpha f_n$ which says $f_{n+1} = Kf_n \approx \alpha f_n$ and thus this stabilized limiting function f_n is *an* eigenfunction (or an approximation for one), eigenvalue α. That is the first idea.

The second part of Schwarz is that you define $a_n \equiv \langle f_o, K^n f_0 \rangle$ and consider the sequence for large n. You define the ratio $\theta_n = a_{n+1}/a_n$ (Schwarz quotient) . Basically the claim is that $\lim \theta_n \leq \mu_1$.

Now the discussion talks a lot about the exact solutions φ_i and -- if we knew what they were -- we could talk about $f_0 = \Sigma c_i \varphi_i$ and then we get more specific θ limits for lower eigenvalues if we somehow know that the leading c_i are 0. But this seems of little use to me since we don't know what the φ_i are so we can hardly know what the c_i are. Well, I guess the idea is this: use the method and hope you get a good estimate for μ_1 and φ_1. Then maybe select a *new* starting f_0 that is orthogonal to this φ_1, and then your θ limit will be a lower bound for μ_2, and so on. He says it really only works well for μ_1.

***Upper* Bounds to Eigenvalues**

Our usual variation methods like $R(u) = \langle Ku,u \rangle / \langle u,u \rangle$ always give lower bounds for eigenvalues, it would be nice to have some *upper* bounds so you could bracket (enclose) the eigenvalue to know how accurate you might be. The ideas here are based on the notion that if you define

$$\sigma \equiv \langle Kv,v \rangle$$

$$\alpha^2 \equiv \| Kv - \sigma v \|^2 = \|Kv\|^2 - \sigma^2$$

if v is an exact normalized eigenfunction, then $\alpha = 0$ and $\sigma = \mu$. If v is only an approximate normalized eigenvalue, then you find by the WBFS Theorem (four names, no web hits so no date) that your eigenvalue is bracketed in this way, reminding me of a mean and standard deviation idea,

$$\sigma - \alpha \leq \mu \leq \sigma + \alpha$$

There are technical issues of making sure only one eigenvalue is in the bracketed region. A variation of this method was done by Kohn and Kato.

3.7 Continuity and Uniform Convergence: bilinear series and iterated kernels

One often discusses a sequence of functions $f_n \rightarrow f$. One kind of sequence is $f_n = \Sigma_{i=1}^{n} q_i(x)$ where the f_n in this case is a *partial sum* as shown, and the limit is of course $n \rightarrow \infty$. In an interval of x, a sequence can converge pointwise, uniformly, or in the mean (to pick a few). In raw notes, I give a trivial example which converges pointwise and in the mean, but not uniformly. Stak does not explain to us why it is so important that things converge uniformly, but I do understand that this is a more rigorous form of convergence than "in the mean" (which means in the usual L_2 sense).

Through this book up to this point, we have always used convergence in the mean, which is a weaker form of convergence. In this section, he wants to up the ante and show that certain things converge uniformly. One benefit of uniform convergence is this: if the f_n are continuous, so is the limit f.

We focus on our old Theorem 4 page 218 which said that a function in the range of K can be written as a Fourier series on the non-zero eigenfunctions, and we showed such a series converges in the mean:

$$g = Kf = \Sigma_i \langle f, \mu_i \varphi_i \rangle \varphi_i = \Sigma_i \langle f, K\varphi_i \rangle \varphi_i = \Sigma_i \langle Kf, \varphi_i \rangle \varphi_i$$

So here we have some partial sums where we have taken the limit. So he now proves:

Theorem 2: The above series associated with Theorem 4 page 218 are also uniformly convergent.

So I am quite happy with Theorem 2, and the proof is clear, although it does require an extra condition on K which is that the single integral of $|k|^2$ be finite which is not the HS condition. The theorem implies that various series we have encountered in earlier sections of Chapter 3 are uniformly convergent.

Notice that I have skipped Theorem 1. I don't see really where it fits into the program here exactly. I will state it, and then comment on it:

Theorem 1: The φ_i for non-zero eigenvalues μ_i are continuous functions.

This theorem and maybe Theorem 2 as well require that the k_2 iterated kernel be continuous and bounded on the square. Maybe the reason for Theorem 1 is this: in our partial sums, we always have these φ_i . So if the φ_i are continuous, and if the series is uniformly convergent, then the infinite series sum must represent a continuous function, a little theorem I quoted about 6" above.

Now we come to Theorem 3 for which I have a large paragraph of raw notes. Again, I will just state it and comment:

Theorem 3: Under suitable conditions, if you expand a function on the eigenfunctions of a second order ODE, that infinite series expansion is a function which converges uniformly on the interval.

This is of course shown based on our analysis of the corresponding integral equation. The key idea is that a solution of the ODE Lu=f will be in the range of an integral operator, u = Kf, and is thus an example of our uniformly convergent series on eigenfunctions.

Stak next talks about expansions of things of the form $f(x,\xi) = \Sigma_i C_i\varphi_i(x)\bar{\varphi}_i(\xi)$ which we would call a **bilinear expansion** of a function of two variables in the eigenfunctions of some K. He starts out with the case $f = k_2$ but we might as well just jump to the general case of $f = k_n$ with $n \geq 2$ and state:

Theorem 4: $k_n(x,\xi) = \Sigma_i \mu_i^n \varphi_i(x)\,\bar{\varphi}_i(\xi)$, and this converges uniformly on the square, $n \geq 2$.

Theorem 4A: $\int dx\, k_n(x,x) = \Sigma_i \mu_i^n$ // an interesting Parseval like sum rule $n \geq 2$

Theorem 4B: $[\int dx\, k_n(x,x)]^{1/n} \geq \mu_1$ if K is non-negative, an *upper* bound for μ_1 $n \geq 2$

Theorem 4C: $K^n\varphi_i = \mu_i^n\varphi_i$ // which is trivial to show. $n \geq 1$

Apart from the last item 4C, the previous theorems are only true for $n \geq 2$. But we have

Mercer's Theorem: For suitable extra conditions on K, the above things are also true for n = 1.

Fact: Even without these conditions, $k(x,\xi) = \Sigma_i \mu_i\varphi_i(x)\bar{\varphi}_i(\xi)$ converges in the mean.

3.8 Approximation Methods for Solving Integral Equations

For gory details see **Kantorovich and Krylov** 1958 English translation. In this section we discuss 5 different plans for finding approximate solutions to our inhomo2 integral (Fred second kind inhomo) equation. For the first 3 of these sections, my meta notes are verbatim from the raw notes.

1. Approximate area as little strips, then do lattice integral (Trap and Simpson versions allowed).
2. Approximate the kernel k as a finite separable sum of independent functions.
3. Approximate the solution $u(x) = \Sigma_i c_i v_i(x)$, install in the IE, then do one of two things: (a) minimize the L_2 error of the two sides of the IE (Best in the Mean); (b) require that both sides have the same projection on some finite subspace spanned by some w_i (Galerkin's Method).
4. Two variational methods are presented, theory only: maximize either I(u) or J(u).
5. Carry out the maximize I(u) plan, get the Rayleigh-Ritz Equations, same as a Galerkin case.

1. Lattice Integration (242).

Replace the integral with an n-fold sum, that is, put the integral equation "on a lattice". The integral is then the sum of n terms. We now talk about the n numbers u_i or f_i. The integral equation becomes an n x

n matrix equation 3.70 and we can solve for the u_i in the usual way, such as Cramer's rule. We can then somehow interpolate between the u_i to get u(x). One method of interpolation is provided by the integral equation itself, written as p 242 A.

Here we are just dividing our "area" into thin vertical rectangles where the height is the average of adjacent function values. Recall from Scheid that you get better results if you use thin trapezoids (doing a linear fit within each strip, Trapezoidal Rule) or if you use little quadratic fits (doing a quadratic fit within each strip's region, Simpson's Rule).

Stak comments that Fredholm originally (perhaps 1903) did things in this way when he developed his theory of integral equations. He did the n x n matrix approach on the lattice, then took the limit $n \to \infty$ to get the theory. This approach gives the right results, but has been replaced by H-S theory (perhaps 1912), and so Fredholm's approach is only of historical interest.

Notice that this section is still describing an analytic method. We are not just blindly doing numeric integration of anything as we might do in a variational method.

2. Approximating the kernel as a finite separable sum.

We have some kernel $k(x,\xi)$ and we expand it on some arbitrary finite set of n orthonormal functions $q_i(\xi)$ so that we then have $k(x,\xi) = \Sigma_{i=1}^{n} p_i(x)\, q_i(\xi)$ where $p_i(x)$ are the "coefficients", but you can then compute $p_i(x) = \int d\xi\, k(x,\xi)\, q_i(\xi)$, so now you know both the $q_i(\xi)$ and $p_i(x)$, and now we have a finite separable representation of k. Since we did not use an *infinite* set, this is of course just an *approximation* to $k(x,\xi)$. We then write (3.71) our candidate solutions as $\mu u(x) = \Sigma_i c_i p_i(x) - f(x)$ where we then seek to find the c_i coefficients. We shuffle things and are not surprised to obtain in p343 B an n x n matrix equation for the c_i of the form $A\mathbf{c} = \mathbf{d}$ where the d_i are integrals of f(x) against the q_i. We solve this for the c_i and have our approximate solution for u(x) ! I would add that it would be good to use the symmetry of the underlying physical problem when selecting a good finite set of basis functions. Obviously larger n gives better results. [Caveat: this method requires that the $p_i(x)$ be independent, not clear how one can cause this to be true for an arbitrary k. In general, each set p_i and q_i needs to be an independent set.]

3. Approximating u(x) in terms of a finite set of independent functions $v_i(x)$.

In 2 above we approximated the ξ-part of the $k(x,\xi)$ with such a set. Here we directly approximate the solution u(x). To the extent that $u(x) = \Sigma_i c_i v_i(x)$ is good, 3.72 is good (just $Ku - \mu u = f$ with u replaced twice by this sum). In 3.72, we combine the two sums by defining $g_i(x)$ as shown, and these are n functions that we can compute from k and v_i. The question is now: how should we pick the c_i to make 3.73 ($\Sigma c_i g_i = f$) most true?

(1) Best in the Mean. One obvious thing to do is pick the c_i to minimize the L_2 difference in the two sides, and that gives the n x n matrix problem $A\mathbf{c} = \mathbf{d}$ shown in 3.74.

(2) Galerkin's Method (Russia ~1915). His method is to require that both sides of $\Sigma c_i g_i = f$ (see p 243 3.72,3 for g_i) have the same projections onto some *new* set of n functions $w_i(x)$, as shown in 3.75. That is to say, you pick a subspace and make both sides equal there. These n equations are called Galerkin's equations and they form a matrix equation $A\mathbf{c} = \mathbf{d}$ which you then solve for **c**.

Example 1: try $w_i(x) = g_i(x)$. This replicates method 1, best in the mean.

Example 2: try $w_i(x) = v_i(x)$. This gives 3.77 and 3.78.

Example 3: same as Example 2, but set f = 0 so we have the EV problem. This gives p 245 A which is the same as the set of equations from Rayleigh-Ritz, 3.54 on page 229, with μ now in place of R*. Recall in R-R how the determinant of this thing gave R* values which were "close to the eigenvalues μ_i" and we make exactly that claim for the Galerkin Method. It is interesting that we arrive at the R-R result without having to set any variation derivatives to 0.

Stak claims that the best in the mean method is optical. Web says Galerkin had a tough life, died in 1945, his method also applies to ODE's and other problems, and is widely used in the world.

4. A Variational Principle (245)

Define the functional $I(v) \equiv 2\langle f,v\rangle - \langle Av,v\rangle$. A theorem then shows that when v = u, the solution of the equation Au=f, I(v) is maximized and takes the value $I(u) = \langle f,u\rangle$ -- provided A is real, symmetric and positive. So this gives you a variational method for finding the solution u (other variational methods usually were for the max EV of an *EV equation* and its corresponding EF.) That is, parameterize v, and then adjust parameters to maximize I(v) and you might get close to the true solution u.

In Remark 3a Stak shows that the theory here can be compactly stated if you define a new inner product $[f,g] = \langle f,Ag\rangle$ which you can show is an inner product since A is positive.

Remark 3b: An alternative to maximizing I(v) is to maximize the J(v) shown in 3.82a. This has the possible advantage that it is maximized not just by v = u, but by v = Cu where C is any constant, so you can focus on shape and ignore scale while pondering trial functions v.

Remark 3c: If you set A = K - μI, then this method solves our general inhomo equation for μ < 0. I think he uses A here because this method applies also to ODE's and maybe other equations as well.

5. The Rayleigh-Ritz Equations (248)

Here we carry out the I(v) variational plan of subsection 4 above: Parameterize v as $v = \Sigma_{i=1}^{n} c_i v_i$ and then maximize I(v) by adjusting parameters c_i. Setting $dI/dc_i = 0$ gives the Rayleigh-Ritz Equations which are an n x n matrix equation A**c** = **d** for the c_i,

$$\langle f,v_i\rangle = \Sigma_j^n \langle Av_i, v_j\rangle c_j \qquad // \text{ or } \qquad \langle f,v_i\rangle = c_i \text{ if } [v_i,v_j] = \delta_{ij}$$

This should not be confused with the Rayleigh-Ritz *Procedure* we used on pp 228-231 in our approximate method to find the EVs and EF solutions of Ku = μu.

Set A = K – μI and the R-R equations are then just the Galerkin equations with $w_i = v_i$ as in 3.78.

If you could magically select your v_i such that $[v_i,v_j] = \delta_{ij}$, you get the right above in which case you have the c_i directly and your solution is then just $v = \Sigma_{i=1}^{n}\langle f,v_i\rangle v_i$ which is the start of an infinite series which would give the exact solution if the v_i are complete in [f,g].

Exercise 3.23. Discusses the **extended (generalized) eigenvalue problem** Nu = λMu.

3.9 Non-symmetric (but still HS) operators (250)

For non-symmetric K, it turns out that you can still do pretty much in terms of solving an integral equation of the form Ky = f, but you cannot make use of the φ_i EF's of K because these are not in general

a complete set, as a very simple example shows. *However*, the EF's v_n and u_n of the left and right *iterated* kernels K_L = KK* and K_R= K*K can be used instead! Operators $K_{L,R}$ are both HS (if K is), are symmetric, have the same eigenvalues, and are non-negative, so they have all the properties we like! For example, either set u_n or v_n is complete. The upshot is that you can solve Ky = h as long as h meets a few restrictions.

Along the way, we learn that the v_n span the perp space of K, while the u_n the perp space of K*. These facts lead to a final solution in the usual form we expect such as $y = y_0 + \Sigma_i [(h.u_i)/\mu_i] v_i$. Notice that both the u_i and v_i functions are involved in this formula which solves Ky=h.

Everything studied in this section is illustrated in the simple Example which ends the section.

I am reminded of "matrix theory" where we talk about $M = UU^\dagger$, but cannot find a reference. Cholesky is related but not it. This would be a theory of matrix equation solutions for non-Hermitian matrices. It would repeat the content of this section but in matrix notation.