LECTURE SHELL Λ: THE HAMBURGER MOMENT PROBLEM AND ITS APPLICATIONS TO NUMERICAL ANALYSIS: THE LANCZOS ALGORITHM

MARK GAGARINE

Abstract. Throughout this lecture we detail the integration of the classical moment problem with elementary linear algebra, focusing on its applications in numerical analysis, particularly with sparse Hermitian matrices. The Hamburger moment problem provides a basis for the spectral representation in operator theory, a connection we aim to make clear. We examine its role in achieving a computationally efficient spectral decomposition, as implemented by the Lanczos algorithm. This investigation bridges the remarkable connection between real analysis and the forefront of numerical linear algebra.

1. Introduction

In this lecture we revisit the foundational linear algebra concepts one would have previously encountered early in their math career, now through the notion of self-adjoint operators on a Hilbert space. This will work nicely throughout our discussion as we aim to work with *Sparse Hermitian* matrices, which can be elegantly characterized per the notion of operator theory.

Definition 1.1. Let H be a Hilbert space over $\mathbb{C}$, and consider the continuous linear map, $A : H \to H$. Then, the *adjoint* of A is the linear map, $A^* : H \to H$, satisfying

$$\langle Af, g\rangle = \langle f, A^*g\rangle, \forall f, g \in H$$

Recall that in the finite dimensional case, A represents the Hermitian adjoint of a matrix.

A matrix is sparse when the number of non-zero elements is far exceeded by the number of zero entries, this concept becomes useful once we begin looking into numerical algorithms. Recall that we can also meaningfully characterize matrices by their eigenvalues.

Definition 1.2. Let T be a linear operator on a Hilbert space H. If $\exists\lambda \in \mathbb{C}$ and some non-zero $f \in H$ such that $T(f) = \lambda f$, then λ is said to be an *eigenvalue* for T. If λ is an eigenvalue for T, then we get the *eigenspace* corresponding to λ,

$$E_\lambda := \{f \in H \mid T(f) = \lambda f\}$$

Definition 1.3. A linear operator T on a Hilbert space H is said to be *self-adjoint* when

$$\langle T(f), g\rangle = \langle f, T(g)\rangle, \forall f, g \in H$$

Eigenvalues have many important applications within theoretical and practical settings. The process of *diagonalizing* a matrix involves decomposing it into the product of orthonormal matrices composed of its eigenvectors and a diagonal matrix of its eigenvalues, hence the name. Its applications are endless, from quantum mechanics, and machine learning. Remarkably, this process arises from the *Spectral Theorem*, stated below.

Theorem 1.4. *Let T be a non-zero, continuous linear operator that's compact and self-adjoint on a Hilbert space H. Then, the set of non-zero eigenvalues of T can be arranged into the countable sequence, $(\lambda_i)_{i\in I}$, with*

$$|\lambda_1| \geq \cdots \geq |\lambda_n| \geq \cdots$$

Thus, we obtain the ***spectral decomposition of T****, $T = \sum_j \lambda_j P_{E_{\lambda_j}}$, where $P_{E_{\lambda_j}}$ is the projection of H onto E_{λ_j}.*

2. The Hamburger Moment Problem

The elegant representation of self-adjoint operators in Theorem 1.4 quickly becomes unappetizing once it's time to decompose, as one may recall upon their experience in a linear algebra course. However, the hunger for misery expressed by the spectral decomposition can be satiated with the Hamburger moment problem, which is central to developing of a computationally efficient spectral decomposition algorithm for a sparse Hermitian matrix.

Definition 2.1. Let $\mu : \mathcal{B}_{\mathbb{R}} \to [0,1]$ be a positive measure on $\mathbb{R}$. If $\int_{-\infty}^{\infty} |t|^k d\mu(t) < \infty$, then the value

$$\alpha_k := \int_{-\infty}^{\infty} t^k d\mu(t)$$

is called the *moment of order k* of μ.

Problem 2.2. **Hamburger Moment Problem**
Suppose we're given a sequence of moments, $(\alpha_n)_{n=1}^{\infty}$. The question is, does there exist a Borel measure μ on $\mathbb{R}$ for which all the moments exist? That is, when can you find a measure μ such that the following holds.

$$\alpha_k = \int_{-\infty}^{\infty} t^k d\mu(t), \forall k \in \mathbb{N}$$

We begin by developing the machinery for working with bounded self-adjoint operators, loosening the compactness condition. Along the way we encounter a concept fundamental to understanding and solving the moment problem.

Definition 2.3. Let H be a Hilbert space over $\mathbb{C}$, and $(X, \mathcal{B})$ be a measurable space, where $\mathcal{B}$ is the Borel σ-algebra of X. A *spectral measure*, π, is a map from $\mathcal{B}$ to the set of orthogonal projections on H satisfying the following properties.

(1) $\pi(\emptyset) = 0, \pi(\mathbb{C}) = 1$
(2) If $(B_n)_{n=1}^{\infty} \in \mathcal{B}$ such that $B_i \cap B_j = \emptyset, \forall i \neq j$, then $\pi(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \pi(B_n)$

Remark 2.4. An interesting characteristic about spectral measures is they allow us to generalize the spectral theorem in a continuous sense. If we consider a bounded self-adjoint linear operator, $T : H \to H$, on the complex Hilbert space H, and take any $\xi \in H$, then there exists a positive linear map $f : H \to \langle \xi, f(T)\xi \rangle$. So, we can define the measure π_ξ on the *spectrum* of T, $\Lambda(T) := \{\lambda \in \mathbb{C} \mid T - \lambda I \text{ doesn't have a bounded linear inverse}\}$, with

$$\langle \xi, f(T)\xi \rangle = \int_{\Lambda(T)} \lambda f(T) d\pi_\xi(\lambda)$$

The measure π_ξ is called the *spectral measure associated with the vector* ξ, and is in fact unique, adding to the elegance of the moment problems solution. The details of its existence and uniqueness are a result of the spectral theorem, but well beyond the scope of this lecture, delving deep into the theory behind functional analysis.

Remark 2.5. Before we introduce the theorem which solves the moment problem, we note some properties of complex numbers used in providing a solution to Problem 2.2.

(1) Recall for any two $w, z \in \mathbb{C}$, the following properties of the conjugation holds.
$$\overline{w+z} = \overline{w} + \overline{z} \quad \overline{wz} = \overline{wz} \quad z\overline{z} = |z|^2$$

(2) The final result required to complete our discussion is part of von-Neumann's theorem for operators, which states if A is symmetric and there exists a conjugation, C, such that $C : H \to H$ and $AC = CA$ holds, then A has self-adjoint extensions.

3. The Sufficient Condition

The solution to the Hamburger moment problem has an elegant characterization by the following theorem. The proof itself is an excellent display of techniques covered throughout our discussions in PMATH450, implementing concepts and techniques we've encountered in almost every lecture.

Theorem 3.1. *Let μ be a positive measure on $\mathbb{R}$. It follows that the sequence, $(\alpha_n)_{n=1}^\infty \in \mathbb{R}$, are moments of μ if and only if for all $k \in \mathbb{N}$, and all $\beta_0, \cdots, \beta_k \in \mathbb{C}$, we have;*

$$(❄) \qquad \sum_{i,j=0}^{k} \alpha_{i+j}\beta_i\overline{\beta}_j \geq 0$$

Proof. The forward implication is quite easy to verify.

"$\Rightarrow$" Suppose μ is a positive measure on $\mathbb{R}$ and for some $k \in \mathbb{N}$, α_i is a moment of order i of μ, for $1 \leq i \leq k$. Thus,

$$\sum_{i,j=0}^{k} \alpha_{i+j}\beta_i\overline{\beta}_j = \sum_{i,j=0}^{k} \left(\int_{\mathbb{R}} x^{i+j}\, d\mu(x)\right) \beta_i\overline{\beta}_j = \int_{\mathbb{R}} \left(\sum_{i=0}^{k} \left(x^i\beta_i\right) \cdot \sum_{j=0}^{k} \overline{\left(x^j\beta_j\right)}\right) d\mu(x)$$
$$= \int_{\mathbb{R}} \left(\sum_{i=0}^{k} x^i\beta_i \cdot \overline{\sum_{j=0}^{k} x^j\beta_j}\right) d\mu(x) = \int_{\mathbb{R}} \left|\sum_{\ell=0}^{k} x^\ell\beta_\ell\right|^2 d\mu(x) \geq 0$$

This value satisfies the inequality due to the modulus always being positive, and μ being a positive measure.

"$\Leftarrow$" Assume for every $n \in \mathbb{N}$, and $\beta_1, \cdots, \beta_n \in \mathbb{C}$, that (❄) holds, and define an inner product on the set of polynomials over $\mathbb{R}$ with complex coefficients, $\mathcal{P}$, by

$$\left\langle \sum_{i=0}^{k} \beta_i x^i, \sum_{j=0}^{\ell} \gamma_j x^j \right\rangle = \sum_{i=0}^{k} \sum_{j=0}^{\ell} \alpha_{i+j} \beta_i \overline{\gamma}_j$$

Let $\mathcal{N} := \{\psi \in \mathcal{P} \mid \langle \psi, \psi \rangle = 0\}$, and take the quotient space, $\mathcal{Q} := \mathcal{P}/\mathcal{N}$. Then, one has the the *hat placing map*, $A \to \hat{A}$, from $\mathcal{P} \to \mathcal{Q}$, thus completing the space $\mathcal{P}$ over the inner product above. We take this to be our Hilbert space, H.

Consider the the shift map, $A : \mathcal{P} \to \mathcal{P}$, where $A : \sum_{i=0}^{k} \beta_i x^i = \sum_{i=0}^{k} \beta_i x^{i+1}$. Observe that A symmetric, since for polynomials $p, q \in P$,

$$\langle Ap, q \rangle = \left\langle A \left(\sum_{i=0}^{k} \beta_i x^i \right), \sum_{j=0}^{\ell} \gamma_j x^j \right\rangle = \sum_{i=0}^{k} \sum_{j=0}^{\ell} \left(\int_{\mathbb{R}} x^{i+j+1} d\mu(x) \right) \beta_i \overline{\gamma}_j$$

$$= \left\langle \sum_{i=0}^{k} \beta_i x^i, \sum_{j=0}^{\ell} \gamma_j x^{j+1} \right\rangle = \langle p, Aq \rangle$$

Also note $A : \mathcal{Q} \to \mathcal{Q}$, since by Cauchy-Schwarz in combination with A's symmetry, we get $0 \leq \langle A\psi, A\psi \rangle = |\langle A^2\psi, \psi \rangle| \leq \sqrt{\|A^2\psi\|} \cdot \sqrt{\|\psi\|} = 0$. Thus, A maps to an operator $\hat{A}$ on H.

Now, consider the usual complex conjugate on $\mathcal{P}$, denoted as C. Then, C also maps to an operator $\hat{C} : \mathcal{Q} \to \mathcal{Q}$. Indeed, $\hat{C}$ remains a conjugation on H, since

$$\langle \hat{C}p, \hat{C}q \rangle = \left\langle \sum_{i=0}^{k} \overline{\beta}_i x^i, \sum_{j=0}^{\ell} \overline{\gamma}_j x^j \right\rangle = \sum_{i=0}^{k} \sum_{j=0}^{\ell} \alpha_{i+j} \gamma_j \overline{\beta}_i = \langle q, p \rangle, \forall p, q \in H$$

It can be seen that $\hat{C}\hat{A} = \hat{A}\hat{C}$ holds, since for any $p \in \mathcal{Q}$, one has

$$\hat{C}\hat{A}p = \hat{C} \left(\sum_{i=0}^{k} \beta_i x^{i+1} \right) = \sum_{i=0}^{k} \overline{\beta}_i x^{i+1} = \hat{A} \left(\sum_{i=0}^{k} \overline{\beta}_i x^i \right) = \hat{A}\hat{C}$$

So, by von-Neumann's theorem it follows that $\hat{A}$ has some self-adjoint extension, say $\tilde{A}$. Now, take μ to be the spectral measure associated with the vector $1 \in \mathcal{P}$. Then,

$$\int_{\Lambda(\tilde{A})} x^n d\mu(x) = \langle 1, \tilde{A}^n 1 \rangle = \langle 1, x^n \rangle = \alpha_n$$

This completes the proof of Theorem (3.1), and the solution suggests that μ is the spectral measure for the operator $\tilde{A}$ associated with the vector $1 \in \mathcal{P}$. $\square$

4. (Introduction)^2

Now that we have covered most of the real analysis needed to make the connection between the Hamburger moment problem and the Lanczos algorithm, we introduce the topics in numerical analysis which provoke this interesting link. Two foundational ideas in numerical analysis that to allow computers to approximate integrals and find solutions to systems of linear equations are quadratures and iterative methods.

In practice, large-scale linear systems are usually sparse. As a result, methods that directly solve the system $Ax = b$, which typically have a time complexity between $\mathcal{O}(n^{1.5}))$ and $\mathcal{O}(n^3)$, can be significantly outperformed by methods which approximate the system by iterating towards a solution. This increase in performance comes from exploiting the sparse nature of these matrices. One class of these *iterative methods* are the *Krylov subspace methods.*

Definition 4.1. Suppose we're given an $n \times n$ matrix, A, and an n dimensional vector, b, . If we successively apply A to b, the resulting sequence, $\{b, Ab, A^2b, \cdots, A^{r-1}b\}$, is the *Krylov* sequence. Then, the *Krylov subspace* generated by A and b is the span of the Krylov sequence,

$$\mathcal{K}_r(A, b) := \operatorname{span}\{b, Ab, A^2b, \cdots, A^{r-1}b\}$$

The Krylov sequence resulting from iteratively applying A is typically chosen to be an orthonormal sequence, resulting in the Krylov subspace forming an orthonormal basis. An approximation to the solution of a linear system is formed by minimizing the residual, $r - A^k b$, over the generated subspace.

Another foundational topic in numerical analysis is the numerical integration of a "nice" function. We can achieve this result by computing the integral for some simpler polynomial which approximates our function. This numerical approximation is called a *quadrature* and is classified by the choice of the approximating polynomial.

Definition 4.2. Let f be a continuous function over $\mathbb{R}$, and suppose we approximate f by choosing p to be an orthogonal polynomial. Then, the *gaussian quadrature rule* approximates

$$\int f(x)dx \approx \sum_{i=0}^{n} w_i f(x_i)dx$$

Here, $(w_i)_{i=0}^n$, and $(x_i)_{i=0}^n$ are the optimal weights and nodes chosen such that x_i is the i-th root of p.

5. The Lanczos Algorithm

We now discuss the much anticipated algorithm for finding the eigenvalues of a sparse hermitian matrix and its connection to the Hamburger moment problem. Suppose we wish to find an approximate solution to the the quadratic forms, $v^*(\lambda_i I - A)^{-1}v$, $\forall 1 \leq i \leq m$, where $\lambda_i \in \mathbb{R}$, and $\lambda_i I - A$ is assumed to be invertible. Begin constructing an orthonormal basis for $\mathcal{K}_r(A, v)$ by starting with $v_1 = v/\|v\|$, and use Gram-Schmidt to orthogonalize Av_k against the previous vectors, $V_{k-1} = [v_1 \ \cdots \ v_{k-1}]$, where V_k is the matrix of columns forming an orthonormal basis of the Krylov subspace. Then, we obtain the *Lanczos decomposition* of A, $AV_k = V_{k+1}T_{k+1,k}$, with

$$T_{k+1,k} = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & \cdots & 0 & 0 \\ \beta_1 & \alpha_2 & \beta_2 & \cdots & 0 & 0 \\ 0 & \beta_2 & \alpha_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{k-1} & \beta_{k-1} \\ 0 & 0 & 0 & \cdots & \beta_{k-1} & \alpha_k \\ 0 & 0 & 0 & \cdots & 0 & \beta_k \end{bmatrix} = \begin{bmatrix} T_{k,k} \\ \beta_k e_k^T \end{bmatrix}$$

This process results in the *Jacobi matrix*, which elegantly demonstrates how our construction of the Krylov subspace mirrors the 3-term-recurrence structure used to construct orthogonal basis vectors, rather than applying the full Gram-Schmidt process, allowing for a much faster computation. In this matrix, the diagonal elements, $\alpha_i = v_i^* A v_i$, represent the projections of A onto the Krylov subspace, and the β_i's normalize the projected vector.

Masked behind the intricate steps of the process we defined is what the Lanczos algorithm actually achieves. By projecting the original model given by A and initial vector v onto a reduced model, $T_{k,k}$, we are approximating A via a process called *moment matching*. To see this, we outline the following process. Suppose we're given a sequence of moments, $(\alpha_n)_{i=1}^{\infty}$, and we wish to find a distribution for which each point of increase is such that we get the following integral.

$$\alpha_i = \int_{\mathbb{R}} \lambda^i d\mu(\lambda), \forall i \in \mathbb{N}$$

This is precisely the Hamburger moment problem! Thanks to the theory we developed throughout this lecture, we now know exactly how to tackle this problem. Recall that we found the measure satisfying this property to be the spectral measure associated with the 1 vector on the self-adjoint extension of A lifted to the quotient space. Thus, we can define the distribution function,

$$\mu(\lambda) = \begin{cases} 0 & \text{If } \lambda < \lambda_1 \\ \sum_{j=1}^{i} w_j & \text{If } \lambda_i \leq \lambda < \lambda_{i+1} \\ \sum_{j=1}^{i} w_j = 1 & \text{If } \lambda \geq \lambda_n \end{cases}, \ \forall 1 \leq i \leq n-1$$

with weights, $w_j = |x_j^* v/\|v\||^2$, $1 \leq j \leq n$, where $\lambda_1 < \cdots < \lambda_n$ are the eigenvalues corresponding to A, and x_j, $1 \leq j \leq n$, are the corresponding eigenvectors. Hence, μ is connected with the eigenvalues of A. Notice that we can now represent the moment by the Gaussian quadrature,

$$\alpha_i = \int_{\mathbb{R}} \lambda^i \mu(\lambda) = \sum_{j=1}^{n} w_i \lambda^i = v^* A^i v$$

Now, define the distribution function

$$\mu^{(k)}(\lambda) = \begin{cases} 0 & \text{If } \lambda < \lambda_1^{(k)} \\ \sum_{j=1}^{i} w_j^{(k)} & \text{If } \lambda_i^{(k)} \leq \lambda < \lambda_{i+1}^{(k)} \\ \sum_{j=1}^{i} w_j^{(k)} = 1 & \text{If } \lambda \geq \lambda_k^{(k)} \end{cases}, \ \forall 1 \leq i \leq k-1$$

with weights, $w_j^{(k)} = |(x_j^{(k)})^* e^T|^2$, $1 \leq j \leq k$, where $\lambda_1^{(k)} < \cdots < \lambda_k^{(k)}$ are the eigenvalues corresponding to $T_{k,k}$, and $x_j^{(k)}$, $1 \leq j \leq k$, are the corresponding eigenvectors. We can again represent the moment by the Gaussian quadrature,

$$\int_{\mathbb{R}} \lambda^i \mu^{(k)}(\lambda) = \sum_{j=1}^{n} w_i^{(k)} (\lambda^{(k)})^i = e_1^T T_{k,k}^i e^T$$

Because the Gaussian quadrature is exact for polynomials up to degree $2k-1$, the first $2k$ moments match.

$$v^* A^i v = (v^* v) e_1^T T_{k,k}^i e_1, \ \forall 0 \leq i \leq 2k-1$$

Finally, we can "shift" the systems via the transfer function $e_1^*(\lambda_i I - T_{k,k})^{-1}$ for the dynamical system

$$\hat{x}'(t) = T_{k,k}\hat{x}(t) + e_1 u(t),\ \hat{y}(t) = e_1^*\hat{x}(t)$$

which matches the first $2k$ moments of the transfer function $v^*(\lambda_i I - A)^{-1}v$ for the n dimensional system

$$x'(t) = Ax(t) + vu(t),\ y(t) = v^*x(t)$$

This method leverages the *shift property* of Krylov subspaces, where $\mathcal{K}_k(\lambda_i I - A, v) = \mathcal{K}_k(A, v)$. As a result, our theory developed throughout the lecture provides us with a derivation of the Lanczos method used to compute the eigenvalues of a sparse hermitian matrix, by approximating the eigenvalues projected onto a reduced space.

References

[1] Andrew M. Bruckner, Judith B. Bruckner, Brian S. Thomson. *Real Analysis*. Prentice-Hall Publishers, 1997.

[2] A. Nica. Lecture shells for PMath 450 lectures in Summer Term 2024, available on the Learn web-site of the course.

[3] Jorg Liesen, Zdenek Strakos. *Krylov Subspace Methods: Principles and Analysis*. Oxford University Press, 2013.

[4] Michael Reed; Barry Simon. *Functional Analysis, Methods of modern mathematical physics, vol. 2*. Academic Press, 1980. pp. 223.

[5] Michael Reed; Barry Simon. *Fourier Analysis, Self-Adjointness, Methods of modern mathematical physics, vol. 2*. Academic Press, 1975. pp. 145.

Faculty of Mathematics,University of Waterloo
Email address: magagarine@uwaterloo.ca