The power method

Siltanen/Railo/Kaarnioja

Spring 2018
Applications of matrix computations

Introduction

The spectral properties of operators give valuable insight on various physical phenomena.

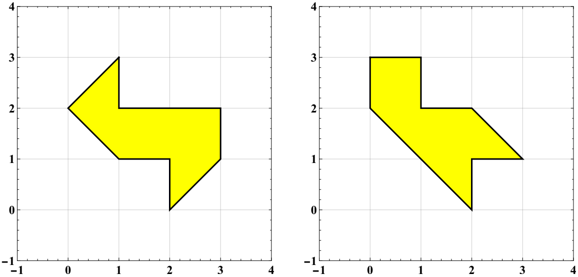


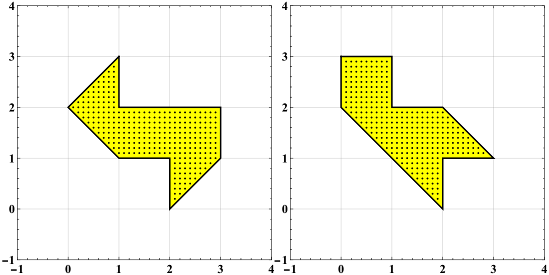
Figure: Isospectral drum shapes in 2D [Gordon, Webb, and Wolpert].

Two drums with clamped boundaries give the same sound if they have the same set of (Dirichlet) eigenvalues λ satisfying

$$-\Delta u = \lambda u \text{ in } D, \; u|_{\partial D} = 0, \quad D \subset \mathbb{R}^2 \text{ bounded domain.}$$

Discretization (FDM)

For a collection $\{x_i\}_{i=1}^N$ of collocation points within the domain D, form the discretized solution vector $\mathbf{u} = [u(x_1), \ldots, u(x_N)]^{\mathrm{T}}$.



It is possible to discretize the Laplacian (incl. boundary conditions) as $\Delta u \approx A\mathbf{u}$.

In consequence, the Dirichlet eigenvalue problem can be approximated by the matrix eigenvalue problem $-A\mathbf{u} = \lambda\mathbf{u}$.

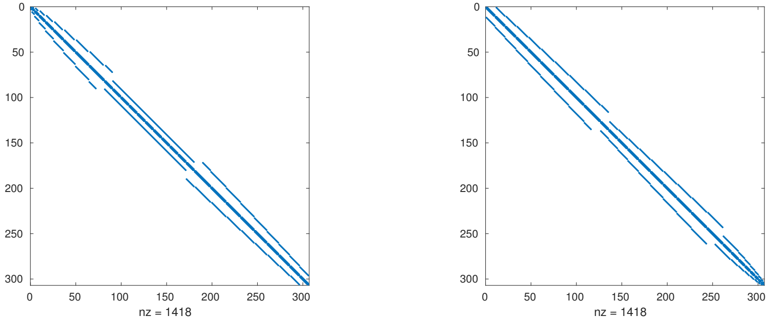


Figure: Left: the FDM matrix A of the first drum. Right: the FDM matrix B of the second drum. Both matrices have dimensions 306×306.

```
>> norm(eig(A)-eig(B))
ans =

   8.8575e-12
```

Eigenvalues of matrices

Let A be an $n \times n$ matrix. Suppose that the pair $(\lambda, \mathbf{v}) \in \mathbb{C} \times (\mathbb{C}^n \setminus \{\mathbf{0}\})$ satisfies

$$A\mathbf{v} = \lambda\mathbf{v}.$$

Then

- λ is called an *eigenvalue* of matrix A.
- $\mathbf{v}$ is called an *eigenvector* of matrix A.

You should be familiar with the algebraic approach to solving the eigenvalues of A: by finding the roots of the characteristic polynomial $p(\lambda) = \det(A - \lambda I)$.

The eigenvector(s) corresponding to λ can be determined by solving the basis vectors spanning $\mathrm{Ker}(A - \lambda I)$ (usually by Gaussian elimination when computing by hand).

Definition

The matrix A is called *diagonalizable* if it can be written as

$$A = PDP^{-1}$$

for some invertible $n \times n$ matrix P and some diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

By writing the columns of P as $P = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ the connection to eigenvalues becomes apparent:

$$\begin{aligned} A = PDP^{-1} \quad &\Leftrightarrow \quad AP = PD \\ &\Leftrightarrow \quad A[\mathbf{v}_1, \ldots, \mathbf{v}_n] = [\mathbf{v}_1, \ldots, \mathbf{v}_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \\ &\Leftrightarrow \quad [A\mathbf{v}_1, \ldots, A\mathbf{v}_n] = [\lambda_1 \mathbf{v}_1, \ldots, \lambda_n \mathbf{v}_n] \\ &\Leftrightarrow \quad A\mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i \in \{1, \ldots, n\}. \end{aligned}$$

Some things to keep in mind:

- All real symmetric matrices $A = A^{\mathrm{T}}$ are diagonalizable (and, in fact, their eigenvalues and eigenvectors are real).
- The eigenvectors of a diagonalizable matrix form a basis for $\mathbb{R}^n$.
- The eigenvectors of a real symmetric matrix form an orthogonal basis for $\mathbb{R}^n$.
- Even if the matrix A is not diagonalizable, it still has a Jordan canonical form.

Power method

Let A be a diagonalizable matrix such that $A = PDP^{-1}$ for $P = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ and $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$. This of course means that

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i \in \{1, \dots, n\},$$

and the eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ form a basis for $\mathbb{R}^n$. Let us assume that the eigenvalues are ordered $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$, i.e., the largest eigenvalue in modulus is a simple eigenvalue.

Let us investigate the simple power iteration, where we begin by initializing a random vector $\mathbf{x}^0 \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and proceed to compute the subsequent iterates as

$$\begin{aligned}
\mathbf{x}^1 &= A\mathbf{x}^0 \\
\mathbf{x}^2 &= A\mathbf{x}^1 = A^2\mathbf{x}^0 \\
\mathbf{x}^3 &= A\mathbf{x}^2 = A^3\mathbf{x}^0 \\
&\vdots \\
\mathbf{x}^k &= A\mathbf{x}^{k-1} = A^k\mathbf{x}^0.
\end{aligned}$$

Let us write the arbitrary initial guess $\mathbf{x}^0 \in \mathbb{R}^n$ using the eigenbasis of A as $\mathbf{x}^0 = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. In addition, we assume that $c_1 \neq 0$, that is, the initial guess contains a nonzero component in the direction of the dominant eigenvalue.[1]

Now $A^k = PD^kP^{-1}$, where $D^k = \mathrm{diag}(\lambda_1^k, \ldots, \lambda_n^k)$.

[1] This is not a restricting assumption since in practice the initial guess is generated randomly. In particular, this means that the probability that $c_1 = 0$ is zero.

Then

$$\begin{aligned} A^k\mathbf{x}^0 &= PD^kP^{-1}(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) && (P^{-1}\mathbf{v}_i = \mathbf{e}_i) \\ &= PD^k(c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n) = P(c_1\lambda_1^k\mathbf{e}_1 + \cdots + c_n\lambda_n^k\mathbf{e}_n) && (P\mathbf{e}_i = \mathbf{v}_i) \\ &= \sum_{i=1}^n c_i\lambda_i^k\mathbf{v}_i = c_1\lambda_1^k\mathbf{v}_1 + \sum_{i=2}^n c_i\lambda_i^k\mathbf{v}_i = \lambda_1^k\left(c_1\mathbf{v}_1 + \sum_{i=2}^n c_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\mathbf{v}_i\right). \end{aligned}$$

Hence

$$\frac{A^k\mathbf{x}^0}{\lambda_1^k} = c_1\mathbf{v}_1 + \sum_{i=2}^n c_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\mathbf{v}_i \xrightarrow{k\to\infty} c_1\mathbf{v}_1$$

with convergence rate $\mathcal{O}(|\lambda_2/\lambda_1|^k)$ as $k \to \infty$.[2]

Note that once the eigenvector $\mathbf{v}_1$ is (approximately) known, the corresponding eigenvalue can be computed as

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1 \quad \Rightarrow \quad \lambda_1 = \frac{\mathbf{v}_1^{\mathrm{T}}A\mathbf{v}_1}{\mathbf{v}_1^{\mathrm{T}}\mathbf{v}_1}.$$

[2] Landau's big-O notation: $f(x) = \mathcal{O}(g(x))$ (as $x \to \infty$) $\Leftrightarrow$ for some constant $C > 0$, $|f(x)| \le C|g(x)|$ for sufficiently large $x \gg 0$. What is constant C here? :)

Power method

Algorithm

Start with an initial guess $\mathbf{x}^0 \in \mathbb{R}^n \setminus \{\mathbf{0}\}$.

for $k = 1, 2, \ldots$ **do**

    Compute $\mathbf{y} = A\mathbf{x}^{k-1}$;

    Set $\mathbf{x}^k = \mathbf{y}/\|\mathbf{y}\|$;

    Set $\lambda_k = (\mathbf{x}^k)^{\mathrm{T}} A\mathbf{x}^k$;

end for

The algorithm can be terminated once, e.g., $\|A\mathbf{x}^k - \lambda_k \mathbf{x}^k\| < \text{threshold}$.

If *the algorithm converges,* then $\mathbf{x} = \lim_{k\to\infty} \mathbf{x}_k$ *and* $\lambda = \lim_{k\to\infty} \lambda_k$ *satisfy* $A\mathbf{x} = \lambda\mathbf{x}$, *where* λ *is the largest eigenvalue of matrix* A *in modulus.*

In applications involving discrete time Markov chains, the dominant eigenvector has a natural interpretation as a stationary probability distribution.

In part II, we will discuss Markov chains and their properties.

Bibliography

- C. Gordon, D. Webb, and S. Wolpert. Isospectral plane domains and surfaces via Riemannian orbifolds. *Inventiones Mathematicae* **110**(1):1–22, 1992. DOI:10.1007/BF01231320
- G. Golub and C. Van Loan. *Matrix Computations*, 2nd edition, The Johns Hopkins University Press, 1989.
- R. A. Horn and C. R. Johnson. *Matrix Analysis*, 1st paperback edition, Cambridge University Press, 1985.
- D. Poole. *Linear Algebra: A Modern Introduction*. Thomson Brooks/Cole, 2005.