CMDA 4604: Intermediate Topics in Mathematical Modeling
Lecture 29: Stability of the Heat Discretization

The finite element solution of the heat equation $u_t(x,t) = u_{xx}(x,t)$ on $x \in [0,1]$ with Dirichlet boundary conditions seeks the solution

$$u_N(x,t) = \sum_{k=1}^{N} a_k(t)\phi_k(x),$$

where $\{\phi_k\}$ is the set of hat functions on the uniformly spaced grid $x_0, \ldots, x_{N+1}$, where $x_j = jh$ for grid size $h = 1/(N+1)$.

The Galerkin condition leads to a system of ordinary differential equations for the time-dependent coefficients $\{a_k(t)\}$:

$$\mathbf{M}\mathbf{a}'(t) = -\mathbf{K}\mathbf{a}(t), \tag{1}$$

where

$$\mathbf{a} = \begin{bmatrix} a_1(t) \\ a_2(t) \\ \vdots \\ a_N(t) \end{bmatrix}, \quad \mathbf{M} = h \begin{bmatrix} 2/3 & 1/6 & & \\ 1/6 & 2/3 & \ddots & \\ & \ddots & \ddots & 1/6 \\ & & 1/6 & 2/3 \end{bmatrix}, \quad \mathbf{K} = \frac{1}{h} \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix}.$$

Note that the mass matrix $\mathbf{M}$ and the stiffness matrix $\mathbf{K}$ are *tridiagonal* (zero side from the main diagonal, and the first super- and sub-diagonals), and *Toeplitz* (constant entries on the diagonals). These two properties are key to our analysis.

In this lecture we seek to relate the *spectral properties* (eigenvalues and eigenvectors) of $\mathbf{M}$ and $\mathbf{K}$ to understand the performance of the Forward Euler and Backward Euler methods for approximating the solution $\mathbf{a}(t)$.

1. The Famous Eigenvalues of a Special Matrix.

Remarkably, everything we need can be derived by understanding the single $N \times N$ matrix

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & & \\ 1 & 0 & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & 0 \end{bmatrix}.$$

Using basic identities from trigonometry, it is not hard to prove that the eigenvalues γ_j and corresponding eigenvectors $\mathbf{v}_j$ of $\mathbf{G}$ are given by

$$\gamma_k = 2\cos\Big(\frac{k\pi}{N+1}\Big), \qquad \mathbf{v}_k = \begin{bmatrix} \sin\left(\frac{k\pi}{N+1}\right) \\ \sin\left(\frac{2k\pi}{N+1}\right) \\ \vdots \\ \sin\left(\frac{Nk\pi}{N+1}\right) \end{bmatrix}.$$

Notice that $\gamma_k \in (-2, 2)$ for all k and N. It is also interesting to note that the entries of the vector $\mathbf{v}_k$ interpolate the function $\sin(k\pi x)$ at the grid points $x_1, \ldots, x_k$, so that the function

$$v_k(x) = \sum_{j=1}^{N} (\mathbf{v}_k)_j \, \phi_j(x)$$

interpolates the eigenfunction $\sin(k\pi x)$ of the Laplacian at the grid points $x_0, \ldots, x_{N+1}$. (We will come back to this point later.)

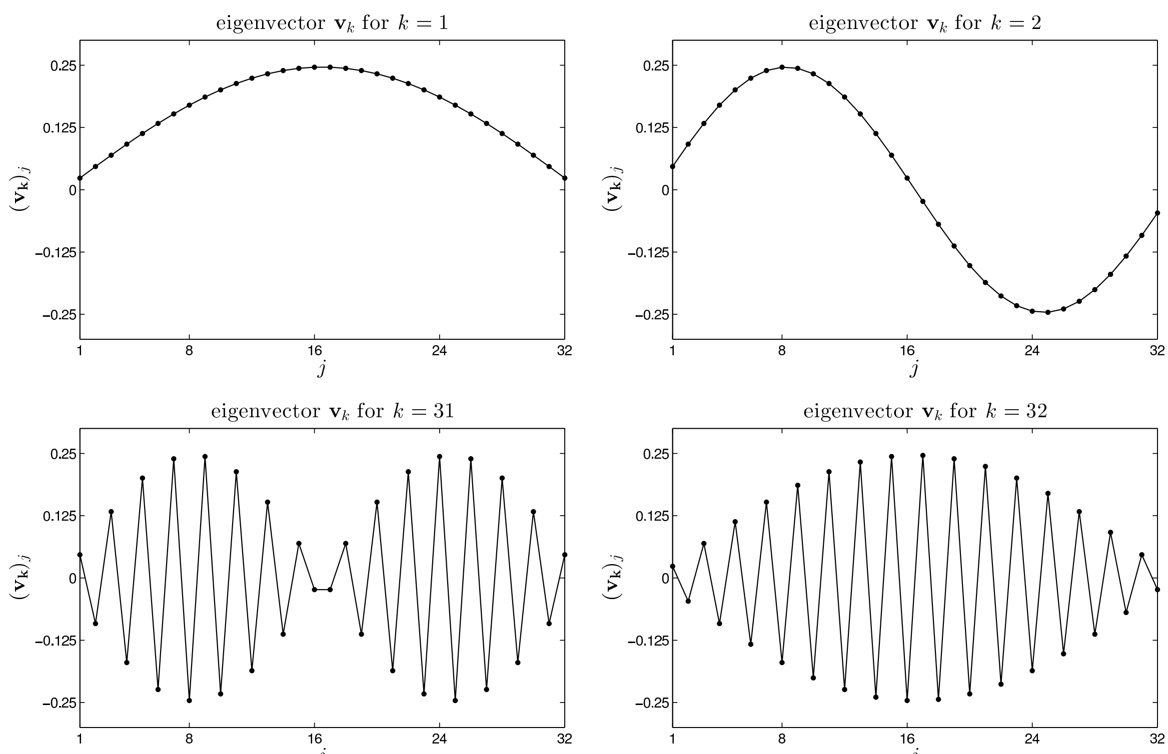
Since $\mathbf{G}$ is a symmetric matrix and the eigenvalues are distinct, the eigenvectors are orthogonal. If we instead scale them as

$$\mathbf{v}_k = \sqrt{\tfrac{2}{N+1}} \begin{bmatrix} \sin\left(\frac{k\pi}{N+1}\right) \\ \sin\left(\frac{2k\pi}{N+1}\right) \\ \vdots \\ \sin\left(\frac{Nk\pi}{N+1}\right) \end{bmatrix},$$

the eigenvectors will normalized, $\|\mathbf{v}_k\| = 1$. From now on we assume this scaling, so that we can write $\mathbf{G}$ in its spectral representation (diagonalization)

$$\mathbf{G} = \sum_{k=1}^{N} \gamma_k \mathbf{v}_j \mathbf{v}_j^T = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_N \end{bmatrix} \begin{bmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_N \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_1^T \end{bmatrix} = \mathbf{V}\boldsymbol{\Gamma}\mathbf{V}^T.$$

The plots below show eigenvectors $\mathbf{v}_k$ for $k = 1, 2, 31, 32$ for $N = 32$.



2. $\mathbf{G}$ is the key that unlocks $\mathbf{M}$, $\mathbf{K}$, and $\mathbf{M}^{-1}\mathbf{K}$.

Notice that the matrices $\mathbf{M}$ and $\mathbf{K}$ can be written as simple transformations of $\mathbf{G}$:

$$\mathbf{M} = \frac{2h}{3}\mathbf{I} + \frac{h}{6}\mathbf{G}, \qquad \mathbf{K} = \frac{2}{h}\mathbf{I} - \frac{1}{h}\mathbf{G}.$$

These transformed matrices have the same eigenvectors as $\mathbf{G}$, and the eigenvalues follow the same map:

$$\begin{aligned}
\mathbf{M}\mathbf{v}_k &= \Big(\frac{2h}{3}\mathbf{I} + \frac{h}{6}\mathbf{G}\Big)\mathbf{v}_k = \frac{2h}{3}\mathbf{v}_k + \frac{h}{6}\mathbf{G}\mathbf{v}_k = \frac{2h}{3}\mathbf{v}_k + \frac{h}{6}\gamma_k\mathbf{v}_k = \Big(\frac{2h}{3} + \frac{h}{6}\gamma_k\Big)\mathbf{v}_k \\
\mathbf{K}\mathbf{v}_k &= \Big(\frac{2}{h}\mathbf{I} - \frac{1}{h}\mathbf{G}\Big)\mathbf{v}_k = \frac{2}{h}\mathbf{v}_k - \frac{1}{h}\mathbf{G}\mathbf{v}_k = \frac{2}{h}\mathbf{v}_k + \frac{1}{h}\gamma_k\mathbf{v}_k = \Big(\frac{2}{h} - \frac{1}{h}\gamma_k\Big)\mathbf{v}_k.
\end{aligned}$$

Thus we have:

$$\begin{array}{llll}
\text{eigenvalues of } \mathbf{M}: & \mu_k = \frac{2h}{3} + \frac{h}{6}\gamma_k; & \text{eigenvectors of } \mathbf{M}: & \mathbf{v}_k \\
\text{eigenvalues of } \mathbf{K}: & \kappa_k = \frac{2}{h} - \frac{1}{h}\gamma_k; & \text{eigenvectors of } \mathbf{K}: & \mathbf{v}_k
\end{array}$$

We can thus write

$$\begin{aligned}
\mathbf{M} &= \sum_{k=1}^{N} \mu_k \mathbf{v}_j\mathbf{v}_j^T = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_N \end{bmatrix} \begin{bmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_N \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_1^T \end{bmatrix} \\
\mathbf{K} &= \sum_{k=1}^{N} \kappa_k \mathbf{v}_j\mathbf{v}_j^T = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_N \end{bmatrix} \begin{bmatrix} \kappa_1 & & \\ & \ddots & \\ & & \kappa_N \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_1^T \end{bmatrix}.
\end{aligned}$$

To understand the solution of the finite element discretization of the heat equation, rewrite (1) as

$$\mathbf{a}'(t) = -\mathbf{M}^{-1}\mathbf{K}\mathbf{a}(t). \tag{2}$$

We need to investigate $\mathbf{M}^{-1}\mathbf{K}$. Since we know the eigenvalues and eigenvectors of $\mathbf{M}$, we can easily write the inverse:

$$\mathbf{M}^{-1} = \sum_{k=1}^{N} \frac{1}{\mu_k}\mathbf{v}_j\mathbf{v}_j^T = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_N \end{bmatrix} \begin{bmatrix} 1/\mu_1 & & \\ & \ddots & \\ & & 1/\mu_N \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_1^T \end{bmatrix}.$$

Notice that $\mathbf{M}^{-1}$ has the same eigenvectors too, and the eigenvalues are just the reciprocals of the eigenvalues of $\mathbf{M}$.

Since $\mathbf{M}^{-1}$ and $\mathbf{K}$ have the same eigenvectors, their product also has a simple form:

$$\begin{aligned}
\mathbf{M}^{-1}\mathbf{K} &= \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_N \end{bmatrix} \begin{bmatrix} 1/\mu_1 & & \\ & \ddots & \\ & & 1/\mu_N \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_1^T \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_N \end{bmatrix} \begin{bmatrix} \kappa_1 & & \\ & \ddots & \\ & & \kappa_N \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_1^T \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_N \end{bmatrix} \begin{bmatrix} \kappa_1/\mu_1 & & \\ & \ddots & \\ & & \kappa_N/\mu_N \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_1^T \end{bmatrix} = \sum_{k=1}^{N} \frac{\kappa_k}{\mu_k}\mathbf{v}_k\mathbf{v}_k^T.
\end{aligned}$$

Thus, $\mathbf{M}^{-1}\mathbf{K}$ has the same eigenvectors, and eigenvalues

$$\lambda_k = \frac{\kappa_k}{\mu_k} = \frac{\frac{2}{h} - \frac{1}{h}\gamma_k}{\frac{2h}{3} + \frac{h}{6}\gamma_k} = \frac{6}{h^2}\left(\frac{2-\gamma_k}{4+\gamma_k}\right).$$

$$\text{eigenvalues of } \mathbf{M}^{-1}\mathbf{K}: \quad \lambda_k = \frac{6}{h^2}\left(\frac{2-\gamma_k}{4+\gamma_k}\right); \quad \text{eigenvectors of } \mathbf{M}^{-1}\mathbf{K}: \quad \mathbf{v}_k$$

Recall that $\gamma_k = 2\cos(k\pi/(N+1))$. Hence, the eigenvalues of $\mathbf{M}^{-1}\mathbf{K}$ are

$$\lambda_k = \frac{6}{h^2}\left(\frac{2-2\cos\theta_k}{4+2\cos\theta_k}\right)$$

for $\theta_k = k\pi/(N+1) \in (0,\pi)$. We can thus bound the eigenvalues λ_k by finding the extrema of the function

$$f(\theta) = \frac{6}{h^2}\left(\frac{2-2\cos\theta}{4+2\cos\theta}\right)$$

over $\theta \in [0,\pi]$, with $\lambda_k = f(\theta_k)$. By computing

$$f'(\theta) = \frac{18\sin\theta}{h^2(2+\cos\theta)^2},$$

one can see that the only candidate extrema of f over $\theta \in [0,\pi]$ occur at $\theta = 0$ and $\theta = \pi$. Note that $f(0) = 0$ and

$$f(\pi) = \frac{12}{h^2}.$$

Moreover, since $f'(\theta) > 0$ for all $\theta \in (0,\pi)$, the eigenvalues λ_k are an increasing function of k. We conclude that

$$\lambda_k \in \left(0, \frac{12}{h^2}\right),$$

with $\lambda_1 \approx 0$ the smallest eigenvalue of $\mathbf{M}^{-1}\mathbf{K}$, and $\lambda_N \approx 12/h^2$ the largest. The eigenvalues of $\mathbf{M}^{-1}\mathbf{K}$ for $N = 32$ are shown in the plot below.