Lecture 4 Eigenvalue problems

Weinan E[1,2] and Tiejun Li[2]

[1] Department of Mathematics,
Princeton University,
weinan@princeton.edu

[2] School of Mathematical Sciences,
Peking University,
tieli@pku.edu.cn
No.1 Science Building, 1575

Outline

Eigenvalue problem

- Eigenvalue problem

  Find λ and $\boldsymbol{x}$ such that

  $$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}, \quad \boldsymbol{x} \neq \boldsymbol{0}.$$

  λ is called the eigenvalues of $\boldsymbol{A}$ which satisfies the eigenpolynomial

  $$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0,$$

  $\boldsymbol{x}$ is called the eigenvector corresponds to λ.

- The are n complex eigenvalues according to Fundamental Theorem of Algebra.

Eigenvalue problem for symmetric matrix

Theorem (For symmetric matrix)

The eigenvalue problem for real symmetric matrix has the properties

1. *The eigenvalues are real, i.e.* $\lambda_i \in \mathbb{R}, i = 1, \ldots, n$.
2. *The multiplicity of a eigenvalue to the eigenpolynomial = the number of linearly independent eigenvectors corresponding to this eigenvalue.*
3. *The linearly independent eigenvectors are orthogonal each other.*
4. $\boldsymbol{A}$ *has the following spectral decomposition*

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^T$$

where

$$\boldsymbol{Q} = (\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_n^T), \quad \Lambda = diag(\lambda_1, \cdots, \lambda_n).$$

Variational form for symmetric matrix

Theorem (Courant-Fisher Theorem)

Suppose $\boldsymbol{A}$ *is symmetric, and the eigenvalues* $\lambda_1 \geq \cdots \geq \lambda_n$*, if we define the Rayleigh quotient as*

$$R_{\boldsymbol{A}}(\boldsymbol{u}) = \frac{\boldsymbol{u}^T \boldsymbol{A} \boldsymbol{u}}{\boldsymbol{u}^T \boldsymbol{u}}$$

then we have,

$$\lambda_1 = \max R_{\boldsymbol{A}}(\boldsymbol{u}), \quad \lambda_n = \min R_{\boldsymbol{A}}(\boldsymbol{u}).$$

Jordan form for non-symmetric matrix

Theorem (Jordan form)

Suppose $\boldsymbol{A} \in \mathbb{C}^{n\times n}$*, if* $\boldsymbol{A}$ *has* r *different eigenvalues* $\lambda_1, \ldots, \lambda_r$ *with multiplicity* $n_1, \ldots, n_r$*, then there exists nonsingular* $\boldsymbol{P}$ *such that* $\boldsymbol{A}$ *has the following decomposition*

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}$$

where $\boldsymbol{J} = \mathrm{diag}(\boldsymbol{J}_1, \ldots, \boldsymbol{J}_r)$*, and*

$$\boldsymbol{J}_k = \begin{pmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{pmatrix}, \quad k = 1, \ldots, r$$

Gershgorin's disks theorem

Definition

Suppose that $n \geq 2$ and $\boldsymbol{A} \in \mathbb{C}^{n\times n}$. The Gershgorin discs D_i, $i = 1, 2, \ldots, n$, of the matrix $\boldsymbol{A}$ are defined as the closed circular regions

$$D_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq R_i\}$$

in the complex plane, where

$$R_i = \sum_{j=1,\ j\neq i}^{n} |a_{ij}|$$

is the radius of D_i.

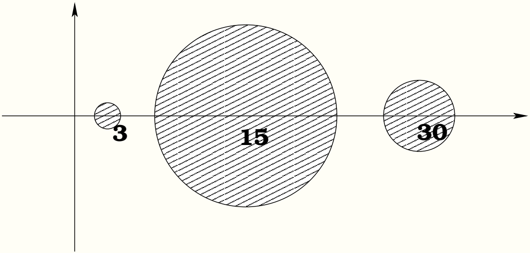
Theorem (Gershgorin theorem)

All eigenvalues of the matrix $\boldsymbol{A}$ lie in the region $D = \cup_{i=1}^{n} D_i$, where D_i are the Gershgorin discs of $\boldsymbol{A}$.

Gershgorin's disks theorem

Geometrical interpretation of Gershgorin's disks theorem for

$$\boldsymbol{A} = \begin{pmatrix} 30 & 1 & 2 \\ 4 & 15 & -4 \\ -1 & 0 & 3 \end{pmatrix}$$



Outline

Basic idea of power method

- First suppose $\boldsymbol{A}$ is diagonizable, i.e.

$$\boldsymbol{A} = \boldsymbol{P}\Lambda\boldsymbol{P}^{-1}$$

and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$. We will assume

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$$

in the follows and assume $\boldsymbol{x}_i$ are the eigenvectors corresponding to λ_i.

- For any initial $\boldsymbol{u}_0 = \alpha_1\boldsymbol{x}_1 + \cdots + \alpha_n\boldsymbol{x}_n$, where $\alpha_k \in \mathbb{C}$. We have

$$\begin{aligned} \boldsymbol{A}^k\boldsymbol{u}_0 &= \sum_{j=1}^{n} \alpha_j \boldsymbol{A}^k \boldsymbol{x}_j = \sum_{j=1}^{n} \alpha_j \lambda_j^k \boldsymbol{x}_j \\ &= \lambda_1^k \Big(\alpha_1 \boldsymbol{x}_1 + \sum_{j=2}^{n} \alpha_j \big(\frac{\lambda_j}{\lambda_1}\big)^k \boldsymbol{x}_j\Big) \end{aligned}$$

Power method

- We have

$$\lim_{k\to\infty} \frac{\boldsymbol{A}^k \boldsymbol{u}_0}{\lambda_1^k} = \alpha_1 \boldsymbol{x}_1.$$

- Though λ_1 and α_1 is not known, the direction of $\boldsymbol{x}_1$ is enough!
- Power method

  1. Set up initial $\boldsymbol{u}_0$, $k = 1$;
  2. Perform a power step $\boldsymbol{y}_k = \boldsymbol{A}\boldsymbol{u}_{k-1}$;
  3. Find the maximal component for the absolute value of $\mu_k = \|\boldsymbol{y}_k\|_\infty$;
  4. Normalize $\boldsymbol{u}_k = \frac{1}{\mu_k}\boldsymbol{y}_k$ and repeat.

- We will have $\boldsymbol{u}_k \to \boldsymbol{x}_1,\ \mu_k \to \lambda_1$.

Power method: example

- Example 1: compute the eigenvalue with largest modulus for

$$A = \begin{pmatrix} 30 & 2 & 3 & 13 \\ 5 & 11 & 10 & 8 \\ 9 & 7 & 6 & 12 \\ 4 & 14 & 15 & 1 \end{pmatrix}$$

- Example 2: compute the eigenvalue with largest modulus for the second order ODE example (n=30)

Power method

Theorem (Convergence of power method)

If the eigenvalues of $\boldsymbol{A}$ has the order $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_p|$ (counting multiplicity), and the algebraic multiplicity of λ_1 is equal to the geometric multiplicity. Suppose the projection of $\boldsymbol{u}_0$ to the eigenspace of λ_1 is not 0, then the iterating sequence is convergent

$$\boldsymbol{u}_k \to \boldsymbol{x}_1, \ \mu_k \to \lambda_1,$$

and the convergence rate is decided by $\frac{|\lambda_2|}{|\lambda_1|}$.

Shifted power method

- Shifted power method:
  Since the convergence rate is decided by $\frac{|\lambda_2|}{|\lambda_1|}$, if $\frac{|\lambda_2|}{|\lambda_1|} \lesssim 1$, the convergence will be slow. An idea to overcome this issue is to "shift" the eigenvalues, i.e. to apply power method to $\boldsymbol{B} = \boldsymbol{A} - \mu \boldsymbol{I}$ (μ is suitably chosen) such that

$$\frac{|\lambda_2(\boldsymbol{B})|}{|\lambda_1(\boldsymbol{B})|} = \frac{|\lambda_2 - \mu|}{|\lambda_1 - \mu|} \ll 1$$

  the eigenvalue with largest modulus keeps invariant.

- Shifted Power method
  1. Set up initial $\boldsymbol{u}_0$, $k = 1$;
  2. Perform a power step $\boldsymbol{y}_k = (\boldsymbol{A} - \mu \boldsymbol{I})\boldsymbol{u}_{k-1}$;
  3. Find the maximal component for the absolute value of $a_k = \|\boldsymbol{y}_k\|_\infty$;
  4. Normalize $\boldsymbol{u}_k = \frac{1}{a_k}\boldsymbol{y}_k$ and repeat.
  5. $\lambda_{\max}(\boldsymbol{A}) = \lambda_{\max}(\boldsymbol{A} - \mu \boldsymbol{I}) + \mu$ (under suitable shift).

- Example 1: Shifted power method $\mu = ?$

Inverse power method

- How to obtain the smallest eigenvalue of $\boldsymbol{A}$?

  This is closely related to computing the ground state energy E_0 for Schrödinger operator in quantum mechanics:

$$\Big(-\frac{\hbar^2}{2\mu}\nabla^2 + U(\boldsymbol{r})\Big)\psi = E_0\psi$$

  where ψ is the wave function.

- Inverse power method: applying power method to $\boldsymbol{A}^{-1}$.

  The inverse of the largest eigenvalue (modulus) of $\boldsymbol{A}^{-1}$ corresponds to the smallest eigenvalue of $\boldsymbol{A}$.

- Just change the step $\boldsymbol{y}_k = \boldsymbol{A}\boldsymbol{u}_{k-1}$ in power method into $\boldsymbol{A}\boldsymbol{y}_k = \boldsymbol{u}_{k-1}$

- Compute the smallest eigenvalue of Example 2 (n=30).

Inverse power method

- Sometimes inverse power method is cooperated with shifting to obtain the eigenvalue and eigenvector corresponding to some λ^* if we already have an approximate $\tilde{\lambda} \approx \lambda^*$, then the power step

$$(\boldsymbol{A} - \tilde{\lambda}\boldsymbol{I})\boldsymbol{y}_k = \boldsymbol{u}_{k-1}$$

  Notice since $\tilde{\lambda} \approx \lambda^*$, we have

$$\lambda_{\max}(\boldsymbol{A} - \tilde{\lambda}\boldsymbol{I}) = \frac{1}{|\tilde{\lambda} - \lambda^*|} \gg 1$$

  The convergence will be very fast.

- Compute the eigenvalue closest to 0.000 for Example 2 (n=30).

Rayleigh quotient accelerating

- When do we need Rayleigh quotient accelerating?
  If $\boldsymbol{A}$ is symmetric and we already have an approximate eigenvector $\boldsymbol{u}_0$, we want to refine this eigenvector and corresponding eigenvalue λ.
- Rayleigh quotient iteration: (Inverse power method + shift)
  1. Choose initial $\boldsymbol{u}_0,\ k=1$;
  2. Compute Rayleigh quotient $\mu_k = \boldsymbol{R}_{\boldsymbol{A}}(\boldsymbol{u}_{k-1})$;
  3. Solve equation for $\boldsymbol{u}_k$, $(\boldsymbol{A}-\mu_k\boldsymbol{I})\boldsymbol{y}_k = \boldsymbol{u}_{k-1}$;
  4. Normalize $\boldsymbol{u}_k = \frac{1}{\|\boldsymbol{y}_k\|_\infty}\boldsymbol{y}_k$ and repeat.
- Remark on Rayleigh quotient iteration and inverse power method.

Outline

QR method

- Suppose $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, then QR method is to apply iterations as follows

$$\boldsymbol{A}_{m-1} = \boldsymbol{Q}_m \boldsymbol{R}_m$$

$$\boldsymbol{A}_m = \boldsymbol{R}_m \boldsymbol{Q}_m$$

  where $\boldsymbol{Q}_m$ is a orthogonal matrix, $\boldsymbol{R}_m$ is an upper triangular matrix.

- Finally $\boldsymbol{R}_m$ will tend to

$$\begin{pmatrix} \lambda_1 & * & \cdots & * \\ & \lambda_2 & \ddots & * \\ & & \ddots & * \\ & & & \lambda_n \end{pmatrix}.$$

  We find all of the eigenvalues of $\boldsymbol{A}$!

- How to find matrix $\boldsymbol{Q}$ and $\boldsymbol{R}$ efficiently to perform QR factorization?

Simplest example

- Vector

$$\boldsymbol{x} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

  Try to eliminate the second component of $\boldsymbol{x}$ to 0.

- Define $\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{x}$,

$$\boldsymbol{Q} = \begin{pmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} 5 \\ 0 \end{pmatrix},$$

Givens transformation

- Suppose

$$x = \begin{pmatrix} a \\ b \end{pmatrix}$$

- Define rotation matrix

$$G = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}$$

where $c = \frac{a}{\sqrt{a^2+b^2}} = \cos\theta$, $s = \frac{b}{\sqrt{a^2+b^2}} = \sin\theta$. It's quite clear that G is a orthogonal matrix.

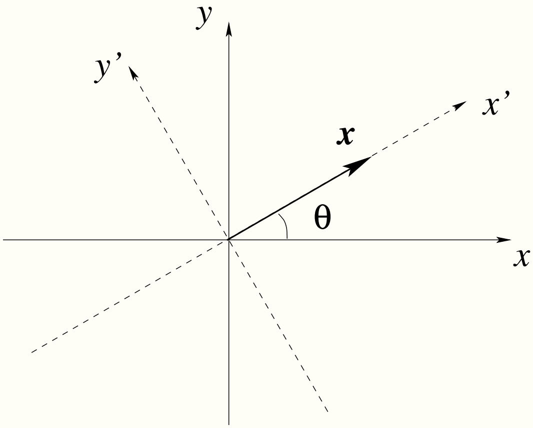
- We have

$$Gx = y = \begin{pmatrix} \sqrt{a^2+b^2} \\ 0 \end{pmatrix}$$

- This rotation is called Givens transformation.

Givens transformation

- Geometrical interpretation of Givens transformation



General Givens transformation

- Define Givens matrix

$$
G(i,k;\theta) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & c & \cdots & s & & \\ & & \vdots & & \vdots & & \\ & & -s & \cdots & c & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ \leftarrow i\text{-th row} \\ \\ \leftarrow k\text{-th row} \\ \\ \\ \end{matrix}
$$

  where $c = \cos\theta, s = \sin\theta$.

- Geometrical interpretation:

  Rotation with θ angle in $i - k$ plane.

Properties of Givens transformation

- Suppose the vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ and we want to eliminate x_k to 0 with x_i.
- Define

$$c = \frac{x_i}{\sqrt{x_i^2 + x_k^2}}, \quad s = \frac{x_k}{\sqrt{x_i^2 + x_k^2}}$$

and $\boldsymbol{y} = \boldsymbol{G}(i, k; \theta)\boldsymbol{x}$, then we have

$$y_i = \sqrt{x_i^2 + x_k^2}, \quad y_k = 0$$

Householder transformation

- Definition. Suppose $\boldsymbol{w} \in \mathbb{R}^n$ and $\|\boldsymbol{w}\|_2 = 1$, define $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ as

$$\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^T.$$

  $\boldsymbol{H}$ is called a Householder transformation.

- Properties of Householder transformation
  1. Symmetric $\boldsymbol{H}^T = \boldsymbol{H}$;
  2. Orthogonal $\boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{I}$;
  3. Reflection (Go on to the next page! :-))

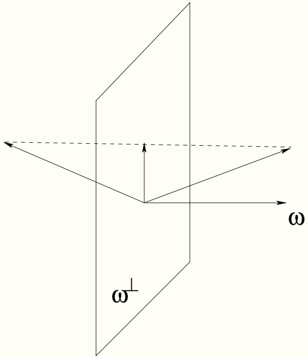
Householder transformation

- For any $\boldsymbol{x} \in \mathbb{R}^n$,

$$\boldsymbol{H}\boldsymbol{x} = \boldsymbol{x} - 2(\boldsymbol{w}^T\boldsymbol{x})\boldsymbol{w}$$

which is the mirror image of $\boldsymbol{x}$ w.r.t. the plane perpendicular to $\boldsymbol{w}$.

- Geometrical interpretation



Application of Householder transformation

- For arbitrary $\boldsymbol{x} \in \mathbb{R}^n$, there exists $\boldsymbol{w}$ such that

$$\boldsymbol{H}\boldsymbol{x} = \alpha \boldsymbol{e}_1$$

where $\alpha = \pm\|\boldsymbol{x}\|_2$. Taking

$$\boldsymbol{w} = \frac{\boldsymbol{x} - \alpha\boldsymbol{e}_1}{\|\boldsymbol{x} - \alpha\boldsymbol{e}_1\|_2}$$

is OK.

- Proof: Define $\beta = \|\boldsymbol{x} - \alpha\boldsymbol{e}_1\|_2$, then

$$\begin{aligned}
\boldsymbol{H}\boldsymbol{x} &= \boldsymbol{x} - 2(\boldsymbol{w}^T\boldsymbol{x})\boldsymbol{w} \\
&= \boldsymbol{x} - \frac{2}{\beta^2}(\alpha^2 - \alpha\boldsymbol{e}_1^T \cdot \boldsymbol{x})(\boldsymbol{x} - \alpha\boldsymbol{e}_1) \\
&= \boldsymbol{x} - \frac{2}{2\alpha^2 - 2\alpha\boldsymbol{e}_1^T \cdot \boldsymbol{x}}(\alpha^2 - \alpha\boldsymbol{e}_1^T \cdot \boldsymbol{x})(\boldsymbol{x} - \alpha\boldsymbol{e}_1) \\
&= \boldsymbol{x} - (\boldsymbol{x} - \alpha\boldsymbol{e}_1) \\
&= \alpha\boldsymbol{e}_1
\end{aligned}$$

Application of Householder transformation

- If define $\boldsymbol{x}' = (x_2, \ldots, x_n)^T$, there exists $\boldsymbol{H}' \in \mathbb{R}^{(n-1)\times(n-1)}$ such that

$$\boldsymbol{H}'\boldsymbol{x}' = \alpha \boldsymbol{e}_1'$$

Define

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 \\ 0 & \boldsymbol{H}' \end{pmatrix}$$

Then we have the last $n-2$ entries of $\boldsymbol{H}\boldsymbol{x}$ will be 0. i.e.

$$\boldsymbol{H}\boldsymbol{x} = (x_1, \sqrt{x_2^2 + \ldots + x_n^2}, 0, \ldots, 0)$$

Upper Hessenberg form and QR method

- Upper Hessenberg form

  Upper Hessenberg matrix $\boldsymbol{A}$ with entry $a_{ij} = 0, j \leq i - 2$, i.e. with the following form

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \ddots & \ddots & \vdots \\ & & a_{n-1,n} & a_{nn} \end{pmatrix}$$

Why upper Hessenberg form

- Why take upper Hessenberg form?
  It can be proved that if $\boldsymbol{A}_{m-1}$ is in upper Hessenberg form then
  $$\boldsymbol{A}_{m-1} = \boldsymbol{Q}_m \boldsymbol{R}_m, \quad \boldsymbol{A}_m = \boldsymbol{R}_m \boldsymbol{Q}_m.$$
  $\boldsymbol{A}_m$ will be in upper Hessenberg form, too.
- The computational effort for QR factorization of upper Hessenberg form will be small.
- Example 3: A QR-factorization step for matrix
  $$\begin{pmatrix} 3 & 1 & 4 \\ 2 & 4 & 3 \\ 0 & 3 & 5 \end{pmatrix}$$

QR method for upper Hessenberg form

- How to transform upper Hessenberg form into QR form?
  The approach is to apply Givens transformation to $\boldsymbol{A}$ column by column to eliminate the sub-diagonal entries.
- Suppose

$$\boldsymbol{A} = \begin{pmatrix} d_1 & * & \cdots & * \\ b_1 & d_2 & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & b_{n-1} & d_n \end{pmatrix}$$

  Apply Givens transformation $\boldsymbol{G}(1,2;\theta_1)$, where $\cos\theta_1 = \frac{d_1}{\sqrt{d_1^2+b_1^2}}$, $\sin\theta_1 = \frac{b_1}{\sqrt{d_1^2+b_1^2}}$, then we have

$$\boldsymbol{A} = \begin{pmatrix} d_1 & * & \cdots & * \\ 0 & d_2' & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & b_{n-1} & d_n \end{pmatrix}$$

QR method for upper Hessenberg form

- Now

$$A = \begin{pmatrix} d_1 & * & \cdots & * \\ 0 & d_2' & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & b_{n-1} & d_n \end{pmatrix}$$

Apply Givens transformation $\boldsymbol{G}(2,3;\theta_2)$, where $\cos\theta_2 = \frac{d_2'}{\sqrt{{d_2'}^2 + b_2^2}}$, $\sin\theta_2 = \frac{b_2}{\sqrt{{d_2'}^2 + b_2^2}}$. We would zero out the entry a_{32}.

- Applying this procedure successively, we obtain

$$A = R = \begin{pmatrix} d_1 & * & \cdots & * \\ 0 & d_2' & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & 0 & d_n' \end{pmatrix}$$

Transformation to upper Hessenberg form

- How to transform a matrix into upper Hessenberg form?
  The approach is to apply Householder transformation to $\boldsymbol{A}$ column by column.

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

- Suitably choose Householder matrix $\boldsymbol{H}_1$ such that

$$\boldsymbol{H}_1 \cdot \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix} = \begin{pmatrix} a'_{11} \\ a'_{21} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Transformation to upper Hessenberg form

- Now we have

$$A_1 = H_1 A H_1 = \begin{pmatrix} a'_{11} & a'_{12} & \cdots & a'_{1n} \\ a'_{21} & a'_{22} & \cdots & a'_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & a'_{n2} & \cdots & a'_{nn} \end{pmatrix}$$

- Suitably choose Householder matrix H_2 such that

$$H_2 \cdot \begin{pmatrix} a'_{12} \\ a'_{22} \\ a'_{32} \\ a'_{42} \\ \vdots \\ a'_{n2} \end{pmatrix} = \begin{pmatrix} a'_{12} \\ a'_{22} \\ a'_{32} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Transformation to upper Hessenberg form

- Apply the Householder transformation $\boldsymbol{A}_2 = \boldsymbol{H}_2 \boldsymbol{A}_1 \boldsymbol{H}_2$, ... successively, we will have the upper Hessenberg form

$$\boldsymbol{B} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \ddots & \ddots & \vdots \\ & & a_{n-1,n} & a_{nn} \end{pmatrix}$$

- $\boldsymbol{B}$ has the same eigenvalues as $\boldsymbol{A}$ because of similarity transformation.

Transformation to upper Hessenberg form

Compute all the eigenvalues of Example 2 (second order ODE, n=5) with QR method.

Homework assignment 4

1. Compute all the eigenvalues of Example 2 (second order ODE, n=20) with QR method.