Numerical Linear Algebra
Chap. 3: Eigenvalue Problems

Heinrich Voss

voss@tu-harburg.de

Hamburg University of Technology
Institute of Numerical Simulation

TUHH
Technische Universität Hamburg-Harburg

Eigenvalues

$\lambda \in \mathbb{C}$ is an eigenvalue of $A \in \mathbb{C}^{n\times n}$ if the homogeneous linear system of equations

$$Ax = \lambda x$$

has a nontrivial solution $x \in \mathbb{C}^n \setminus \{0\}$. Then, x is called an eigenvector of A corresponding to λ.

The set of all eigenvalues of A is called the spectrum of A and is denoted by $\sigma(A)$.

λ is an eigenvalue of A if and only if

$$\det(A - \lambda I) = 0.$$

$\chi(\lambda) := \det(A - \lambda I)$ is a polynomial of degree n, the characteristic polynomial of A.

Eigenvalues ct.

If $\tilde{\lambda}$ is a root of χ of multiplicity k (i.e. the poynomial $\chi(\lambda)$ is divisable by $(\lambda - \tilde{\lambda})^k$ but not by $(\lambda - \tilde{\lambda})^{k+1}$) then k is called the algebraic multiplicity of $\tilde{\lambda}$. The algebraic multiplicity of $\tilde{\lambda}$ is denoted by $\alpha(\tilde{\lambda})$.

For $A \in \mathbb{C}^{n\times n}$ its characterictic polynomial χ has degree n. Hence, the sum of all algebraic multiplicities of eigenvalues equals n.

If λ is an eigenvalue of A then

$$E_\lambda := \{x \in \mathbb{C}^n \,:\, (A - \lambda I)x = 0\}$$

is a subspace of $\mathbb{C}^n$, which is called the eigenspace of A corresponding to λ.

$\gamma(\lambda) :=$ dim E_λ is the geometric multiplicity of an eigenvalue λ of A.

It can be shown that $\gamma(\lambda) \leq \alpha(\lambda)$ for every eigenvalue λ.

Similar matrices

Let $X \in \mathbb{C}^{n\times n}$ be nonsingular. Then

$$A \quad \text{and} \quad B := X^{-1}AX$$

are called similar matrices. $A \mapsto X^{-1}AX$ is called similarity transformation.

Since

$$\begin{aligned} \det(B-\lambda I) &= \det(X^{-1}(A-\lambda I)X) \\ &= \det(X^{-1})\det(A-\lambda I)\det(X) = \det(A-\lambda I), \end{aligned}$$

similar matrices have the same eigenvalues including their algebraic multiplicities.

It can be shown that the geometric multiplicities coincide as well.

Diagonalizable matrix

Let $Ax^j = \lambda_j x^j$, $j = 1, \ldots, k$ where $\lambda_i \neq \lambda_j$ for $i \neq j$. Then the set $\{x^1, \ldots, x^k\}$ is linearly independent.

Let $x = \sum_{j=1}^{k} \alpha_j x^j = 0$. For $j \in \{1, \ldots, k\}$ it follows

$$(A - \lambda_1 I) \cdot \cdots \cdot (A - \lambda_{j-1} I)(A - \lambda_{j+1} I) \cdot \cdots \cdot (A - \lambda_k I)x = \alpha_j \prod_{i=1, i\neq j}^{k} (\lambda_j - \lambda_i)x^j = 0,$$

and therefore $\alpha_j = 0$.

In particular, if A has n different eigenvalues λ_j with eigenvectors x^j, then $X := (x^1, \ldots, x^n)$ is nonsingular, and it holds

$$AX = (Ax^1, \ldots, Ax^n) = (\lambda_1 x^1, \ldots, \lambda_n x^n) = X\Lambda \quad \Longleftrightarrow \quad X^{-1}AX = \Lambda$$

where $\Lambda =: \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ denotes a diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$.

Hence, A is diagonalizable, i.e. similar to a diagonal matrix.

Diagonalizable matrix ct.

More generally, if for all eigenvalues λ_j, $j = 1, \ldots, k$ of A the algebraic and geometric multiplies coincide ($\alpha(\lambda_j) = \gamma(\lambda_j)$), then choosing in each of the eigenspaces E_{λ_j} a basis $x^{j,1}, \ldots, x^{j,\alpha(\lambda_j)}$, the matrix

$$X = (x^{1,1}, \ldots, x^{1,\alpha(\lambda_1)}, x^{2,1}, \ldots, x^{k,\alpha(\lambda_k)})$$

is nonsingular, and it digonalizes A.

It can be shown that A is diagonalizable if and only if $\alpha(\lambda_j) = \gamma(\lambda_j)$ for every eigenvalue λ_j of A.

For

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$\alpha(0) = 2 \neq 1 = \gamma(0)$, and therefore not every matrix is diagonalizable.

Jordan's canonical form

Let $A \in \mathbb{C}^{n\times n}$ with distict eigenvalues $\lambda_1, \ldots, \lambda_k$. Then there exists a nonsingular matrix X such that

$$X^{-1}AX = \operatorname{diag}(J_1, \ldots, J_k) := \begin{pmatrix} J_1 & O & \ldots & O \\ O & J_2 & \ldots, O & \\ & & \ddots & \\ O & O & \ldots & J_k \end{pmatrix}$$

is a block diagonal matrix.

Each of the diagonal blocks $J_j = \operatorname{diag}(J_{j,1}, \ldots, J_{j,\gamma(\lambda_j)})$ is a block diagonal matrix of dimension $\alpha(\lambda_j)$ with $\gamma(\lambda_j)$ blocks where

$$J_{j,i} = \begin{pmatrix} \lambda_j & 1 & & \ldots & 0 \\ 0 & \lambda_j & \ddots & & \vdots \\ & \ddots & \ddots & \ddots & \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \ldots & & 0 & \lambda_j \end{pmatrix}$$

Hermitian matrices

$A \in \mathbb{R}^{n\times n}$ is symmetric if $A = A^T$. More generally, $A \in \mathbb{C}^{n\times n}$ is a Hermitian matrix if $A^H := \overline{A}^T = A$, where $\overline{A}$ denotes the matrix obtained from A by replacing each of its entries by its conjugate complex.

All eigenvalues of a Hermitian matrix are real: for $Ax = \lambda x$ it holds

$$x^H Ax = x^H(\lambda x) = \lambda x^H x \quad \text{and} \quad x^H Ax = (A^H x)^H x = (Ax)^H x = (\lambda x)^H x = \overline{\lambda} x^H x$$

from which we get $\lambda = \overline{\lambda}$, i.e. $\lambda \in \mathbb{R}$.

Eigenvectors of a Hermitian matrix correponding to distinct eigenvalues are orthogonal: for $Ax = \lambda x$, $Ay = \mu y$ and $\lambda \neq \mu$ it holds

$$y^H Ax = \lambda y^H x \quad \text{and} \quad y^H Ax = (A^H y)^H x = (Ay)^H x = \mu y^H x.$$

Hence, $(\lambda - \mu) y^H x = 0$, and $\lambda \neq \mu$ implies $y^H x = 0$.

Invariant subspace

A subspace V of $\mathbb{C}^n$ is an invariant subspace of A if $Ax \in V$ for every $x \in V$.

Every invariant subspace of A contains an eigenvector of A.

Let $x^1, \ldots, x^k \in \mathbb{C}^n$ be a basis of V. Then for $j = 1, \ldots, k$ there exists $b_{ij} \in \mathbb{C}$ such that $Ax^j = \sum_{i=1}^k b_{ij} x^i$.

Let λ be an eigenvalue of $B = (b_{ij}) \in \mathbb{C}^{k \times k}$ with eigenvector $\xi = (\xi_1, \ldots, \xi_k)^T$, and let $x := \sum_{i=1}^k \xi_i x^i \neq 0$. Then

$$Ax = \sum_{j=1}^k \xi_j Ax^j = \sum_{j=1}^k \sum_{i=1}^k \xi_j b_{ij} x^i = \sum_{i=1}^k (\sum_{j=1}^k b_{ij} \xi_j) x^i = \sum_{i=1}^k \lambda \xi_i x^i = \lambda x.$$

Hermitian matrices are diagonalizable

Let A be a Hermitian matrix. Then there exists a unitary matrix $U \in \mathbb{C}^{n\times n}$ (i.e. $U^H U = I$) such that

$$U^H A U = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

Let x^1 be an eigenvector of A such that $Ax^1 = \lambda_1 x^1$ and $(x^1)^H x^1 = 1$. Then for $x \in \mathbb{C}^n$ such that $x^H x^1 = 0$ it holds

$$(Ax)^H x^1 = x^H A^H x^1 = x^H (Ax^1) = \lambda_1 x^H x^1 = 0.$$

Hence, $V_1 := \{x \in \mathbb{C}^n \,:\, x^H x^1 = 0\}$ is an invariant subspace of A, and therefore it contains an eigenvector x^2 which can be normalized such that $(x^2)^H x^2 = 1$.

If $x^1, \ldots, x^j$ are j orthogonal eigenvectors of A, then in the same way as before

$$V_j := \{x^1, \ldots, x^j\}^\perp = \{x \in \mathbb{C}^n \,:\, x^H x^i = 0,\ i = 1, \ldots, j\}$$

is an invariant subspace of A, and hence there exists an eigenvector x^{j+1} which is orthogonal to $x^1, \ldots, x^j$.
$U = (x^1, \ldots, x^n)$ renders the desired property.

Rayleigh's principle

Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. Then for $x \neq 0$

$$R_A(x) := \frac{x^H A x}{x^H x}$$

is called Rayleigh quotient of A at x.

Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of A, and let $x^1, \ldots, x^n$ be a set of corresponding orthogonalized eigenvectors. Then it holds

$$\lambda_1 = \min_{x\neq 0} R_A(x) \quad \text{and} \quad \lambda_n = \max_{x\neq 0} R_A(x).$$

for $i = 1, 2, \ldots, n$ it holds

$$\begin{aligned} \lambda_i &= \min\{R_A(x) : x \in \mathbb{C}^n,\ x^H x^j = 0,\ j = 1, \ldots, i-1\} \\ &= \max\{R_A(x) : x \in \mathbb{C}^n,\ x^H x^j = 0,\ j = i+1, \ldots, n\} \end{aligned}$$

Proof of Rayleigh's principle

Let $x^1, \ldots, x^n$ be an orthonormal system of eigenvectors of $A \in \mathbb{C}^{n\times n}$ where $Ax^j = \lambda_j x^j$.

For $x \in \mathbb{C}^n$, $x \neq 0$ let $x = \sum_{j=1}^n \xi_j x^j$.

$$\begin{aligned} x^H x &= \Big(\sum_{j=1}^n \xi_j x^j\Big)^H \Big(\sum_{k=1}^n \xi_k x^k\Big) = \sum_{j,k=1}^n \overline{\xi_j}\xi_k (x^j)^H x^k = \sum_{j=1}^n |\xi_j|^2 \\ x^H A x &= \Big(\sum_{j=1}^n \xi_j x^j\Big)^H A\Big(\sum_{k=1}^n \xi_k x^k\Big) = \Big(\sum_{j=1}^n \xi_j x^j\Big)^H \Big(\sum_{k=1}^n \xi_k A x^k\Big) \\ &= \Big(\sum_{j=1}^n \xi_j x^j\Big)^H \Big(\sum_{k=1}^n \xi_k \lambda_k x^k\Big) = \sum_{j=1}^n \lambda_j |\xi_j|^2 \end{aligned}$$

Hence,

$$R_A(x) = \sum_{j=1}^n \alpha_j \lambda_j, \quad \text{with } \alpha_j = \frac{|\xi_j|^2}{\sum_{k=1}^n |\xi_k|^2}$$

Proof of Rayleigh's principle

From $0 \le \alpha_j \le 1$ and $\sum_{j=1}^n \alpha_j = 1$ one obtains

$$\lambda_1 = \sum_{j=1}^n \alpha_j \lambda_1 \le \sum_{j=1}^n \alpha_j \lambda_j \le \sum_{j=1}^n \alpha_j \lambda_n = \lambda_n.$$

$$\lambda_1 = R_A(x^1), \quad \lambda_n = R_A(x^n).$$

$$\begin{aligned} \lambda_i &= \min\{R_A(x) : x \in \mathbb{C}^n,\ x^H x^j = 0,\ j = 1, \ldots, i-1\} \\ &= \max\{R_A(x) : x \in \mathbb{C}^n,\ x^H x^j = 0,\ j = i+1, \ldots, n\} \end{aligned}$$

follow in a similar way since $\xi_1 = \cdots = \xi_{i-1} = 0$ if $x^H x^j = 0$ for $j = 1, \ldots, i-1$.

Numerical methods

Linear systems of equations $Ax = b$ can be solved by a finite algorithm (i.e. a finite number of operations) like Gauss elimination.

Determining an eigenvalue of a matrix $A \in \mathbb{R}^{n\times n}$ is equivalent to finding a root of the characteristic polynomial

$$\chi(\lambda) := \det(A - \lambda I) = 0.$$

It is known (Theorem of Abel) that for $n \geq 5$ there is no formula for solving

$$\det(A - \lambda I) = 0$$

for λ. Hence, the eigenvalue problem $Ax = \lambda x$ usually can be solved only by iterative methods.

Example

$$A = \begin{pmatrix} 0.2 & 0.3 & 0.4 \\ 0.6 & 0.2 & 0.5 \\ 0.2 & 0.5 & 0.1 \end{pmatrix}$$

Choose any vector $x^0 \in \mathbb{R}^3$ and compute the sequence

$$x^k := Ax^{k-1}, \; k = 1, 2, 3, \ldots$$

After a small number of steps (≈ 10) we obtain

$$x^k = \begin{pmatrix} 0.5122 \\ 0.6974 \\ 0.5013 \end{pmatrix} \quad \text{and} \quad \|Ax^k - x^k\| \text{ small.}$$

x^k seems to be an eigenvector corresponding to the eigenvalue $\lambda = 1$.
Is this a miracle?

A is stochastic

All elements of A are nonnegative, and every column of A adds to 1. Matrices with these properties are called stochastic. They describe the behavior of Markov chaines.

If A is stochastic, then every row of A^T adds to 1, and therefore $(1, 1, \ldots, 1)^T$ is an eigenvector of A^T corresponding to the eigenvalue 1.

$$\det(A - \lambda I) = \det(A^T - \lambda I)$$

implies that the eigenvalues of A and A^T coincide. Hence, every stochastic matrix has one eigenvalue $\lambda = 1$.

Power method

Assume that A is diagonizable, i.e. there exist n linearly independent eigenvectors $u^1, \ldots, u^n$ of A, and assume that λ_1 is a dominant eigenvalue

$$|\lambda_1| > |\lambda_2|, |\lambda_3|, \ldots, |\lambda_n|.$$

The initial vector x^0 can be representeted as

$$x^0 = \sum_{j=1}^{n} \alpha_j u^j$$

$$Ax^0 = A\Big(\sum_{j=1}^{n} \alpha_j u^j\Big) = \sum_{j=1}^{n} \alpha_j Au^j = \sum_{j=1}^{n} \alpha_j \lambda_j u^j$$

Power method ct.

$$A^2x^0 = A\Big(\sum_{j=1}^{n}\alpha_j\lambda_j u^j\Big) = \sum_{j=1}^{n}\alpha_j\lambda_j Au^j = \sum_{j=1}^{n}\alpha_j\lambda_j^2 u^j$$

By induction it follows

$$A^m x^0 = \sum_{j=1}^{n}\alpha_j\lambda_j^m u^j = \lambda_1^m\Big(\alpha_1 u^1 + \sum_{j=2}^{n}\alpha_j\Big(\frac{\lambda_j}{\lambda_1}\Big)^m u^j\Big).$$

From $|\lambda_j|/|\lambda_1| < 1$ it follows that $(\lambda_j/\lambda_1)^m \to 0$. Hence, if $\alpha_1 \neq 0$, then the sequence

$$\lambda_1^{-m}A^m x^0 = \alpha_1 u^1 + \sum_{j=2}^{n}\alpha_j\Big(\frac{\lambda_j}{\lambda_1}\Big)^m u^j$$

converges to an eigenvector corresponding to λ_1.

Power method ct.

If $|\lambda_1| \neq 1$, then for increasing m one obtains overflow or underflow.

Apply the method to

$$B = \begin{pmatrix} 0.2 & 0.3 & 0.4 \\ 0.6 & -0.1 & 0.5 \\ 0.2 & 0.5 & 0.1 \end{pmatrix}$$

The sequence x^m converges to the null vector. The largest eigenvalue of B in modulus seems to be smaller than 1.

Power method

Normalize x^m in each step to avoid underflow or overflow.

Power method

1: Given initial vector x^0
2: **for** $m = 0, 1, 2, \ldots$ until convergence **do**
3: $\quad y^{m+1} = Ax^m$;
4: $\quad k_{m+1} = \|y^{m+1}\|$
5: $\quad x^{m+1} = y^{m+1}/k_{m+1}$
6: **end for**

With this modification the power method converges in a reasonable number of steps to an eigenvector corresponding to the dominant eigenvalue $\lambda_1 = 0.9304$.

Observations

$$\lambda_1^{-m} A^m x^0 = \alpha_1 u^1 + \sum_{j=2}^{n} \alpha_j \left(\frac{\lambda_j}{\lambda_1}\right)^m u^j$$

demonstrates that the speed of convergence depends on

$$q := \max_{j=2,\ldots,m} \frac{|\lambda_j|}{|\lambda_1|}.$$

The smaller q is, the faster is the convergence of the power method.

If the initial vector x^0 has no component of the eigenvector corresponding to the dominant eigenvalue (i.e. $\alpha_1 = 0$), then in the course of the algorithm rounding errors usually produce a component of u^1 which is amplified in further iterations until convergence.

Starting the power method for A with a linear combination of eigenvectors corresponding to λ_2 and λ_3 one obtains a reasonable approximation to an eigenvector corresponding to λ_1 after 40 iterations.

Observations ct.

If λ_1 is a multiple dominant eigenvalue of A

$$\lambda_1 = \lambda_2 = \cdots = \lambda_p, \quad |\lambda_1| > |\lambda_j| \text{ for } j = p+1, \ldots, n,$$

and A is diagonalizable, then all considerations above stay true.

For

$$|\lambda_1| = |\lambda_2| > |\lambda_j| \text{ for } j = 3, \ldots, n, \quad \text{and} \quad \lambda_1 \neq \lambda_2$$

one does not obtain convergence of the power method.

In steps 4 and 5 of the power method the normalization can be replaced by a scaling

$$k_{m+1} = \ell^T y^{m+1}$$

where $\ell \in \mathbb{R}^n$ is a vector which is not orthogonal to the eigenvector u^1 corresponding to the dominant eigenvector.

Inverse iteration

Applying the power method to the inverse matrix A^{-1} one can determine the smallest eigenvalue in modulus.

Inverse iteration

Given initial vector x^0
for $m = 0, 1, 2, \ldots$ until convergence **do**
Solve $Ay^{m+1} = x^m$ for y^{m+1}
$k_{m+1} = \|y^{m+1}\|$
$x^{m+1} = y^{m+1}/k_{m+1}$
end for

Applying inverse iteration to the matrix B one gets fast convergence to an eigenvector corresponding to the smallest eigenvalue $\lambda_3 = -0.2111$. For A the convergence is very slow. What is the difference?

Inverse iteration ct.

The shifted matrix $A - \tilde{\lambda} I$ has eigenvalues $\lambda_j - \tilde{\lambda}$, if λ_j are the eigenvalues of A.

If $\tilde{\lambda}$ is not an eigenvalue of A, then $(A - \tilde{\lambda} I)^{-1}$ has eigenvalues $\frac{1}{\lambda_j - \tilde{\lambda}}$.

If $|\lambda_p - \tilde{\lambda}| < |\lambda_j - \tilde{\lambda}|$ for $j = 1, \ldots, n$, $j \neq p$ then
Inverse iteration with fixed shift

Given initial vector x^0
for $m = 0, 1, 2, \ldots$ until convergence **do**
Solve $(A - \tilde{\lambda} I) y^{m+1} = x^m$ for y^{m+1}
$k_{m+1} = \ell^T y^{m+1}$
$x^{m+1} = y^{m+1} / k_{m+1}$
end for

converges to an eigenvector corresponding to λ_p. The rate of convergence is

$$q = \max_{j \neq k} \frac{|\lambda_k - \tilde{\lambda}|}{|\lambda_j - \tilde{\lambda}|}.$$

Inverse iteration with variable shifts

For large m it holds that x^m is an approximate eigenvector corresponding to λ_p and $\ell^T x^m = 1$. Hence,

$$k_{m+1} = \ell^T y^{m+1} = \ell^T((A - \tilde{\lambda} I)^{-1} x^m) \approx \frac{1}{\lambda_p - \tilde{\lambda}} \ell^T x^m = \frac{1}{\lambda_p - \tilde{\lambda}}.$$

This observations suggests to iterate the shift as well:

$$k_{m+1} \approx \frac{1}{\lambda_{m+1} - \lambda_m} \quad \Longrightarrow \quad \lambda_{m+1} := \lambda_m + 1/k_{m+1}$$

Inverse iteration with variable shifts

Given initial vector x^0 and initial approximation λ_0
for $m = 0, 1, 2, \ldots$ until convergence **do**
Solve $(A - \lambda_m I) y^{m+1} = x^m$ for y^{m+1}
$k_{m+1} = \ell^T y^{m+1}$
$x^{m+1} = y^{m+1}/k_{m+1}$
$\lambda_{m+1} = \lambda_m + 1/k_{m+1}$
end for

Quadratic convergence

Let $\tilde{\lambda}$ be an algebraically simple eigenvalue of A (i.e. $\tilde{\lambda}$ is a simple root of $\det(A-\lambda I)=0$), let $\tilde{u}$ be a corresponding eigenvector such that $\ell^T\tilde{u}=1$.

Then inverse iteration with variable shifts converges locally and quadratically to $(\tilde{\lambda},\tilde{u})$: There exists some positive constant $C>0$ such that, if λ_0 is sufficiently close to $\tilde{\lambda}$ and x^0 is sufficiently close to $\tilde{u}$, then it holds

$$|\tilde{\lambda}-\lambda_{m+1}|\le C|\tilde{\lambda}-\lambda_m|^2 \quad \text{and} \quad \|\tilde{u}-x^{m+1}\|\le C\|\tilde{u}-x^m\|^2.$$

Deflation

Assume that we have already obtained the largest (smallest, closest to a given shift) eigenvalue $\tilde{\lambda}$ and corresponding eigenvector $\tilde{u}$.
How can we compute further eigenpairs by the power method?

Let $\tilde{y}$ be a left eigenvector of A corresponding to some eigenvalue $\tilde{\mu} \neq \tilde{\lambda}$, i.e. $\tilde{y}^T A = \tilde{\mu}\tilde{y}$.

Then it holds

$$\tilde{\mu}\tilde{y}^T\tilde{u} = (\tilde{y}^T A)\tilde{u} = \tilde{y}^T(A\tilde{u}) = \tilde{\lambda}\tilde{y}^T\tilde{u} \quad \Longrightarrow \quad \tilde{y}^T\tilde{u} = 0.$$

Deflation ct.

Let $B := A - \tilde{u}w^T$, where $w \in \mathbb{R}^n$ satisfies $w^T\tilde{u} \neq 0$

$$B\tilde{u} = A\tilde{u} - \tilde{u}w^T\tilde{u} = (\tilde{\lambda} - w^T\tilde{u})\tilde{u},$$

i.e. $\tilde{u}$ is an eigenvector of B corresponding to the eigenvalue $\tilde{\lambda} - w^T\tilde{u}$.

With eigenvalue $\tilde{\mu} \neq \tilde{\lambda}$ of A and its corresponding left eigenvector y, it holds

$$y^T B = y^T A - y^T\tilde{u}w^T = \tilde{\lambda}y^T.$$

Hence, all eigenvalues of A are kept (only the right eigenvectors can change), whereas the eigenvalue $\tilde{\lambda} - w^T\tilde{u}$ can be moved anywhere by the choice of w (for instance to 0 to compute the second largest eigenvalue of A in modulus).

Symmetric matrices

Let $A = A^T \in \mathbb{R}^{n\times n}$ be a symmetric matrix, $\tilde{\lambda}$ an eigenvalue of A, and $\tilde{u}$ a corresponding eigenvalue such that $\|\tilde{u}\| = 1$.

Let

$$B = A - \tilde{\lambda}\tilde{u}\tilde{u}^T$$

If $v \in \mathbb{R}^n$ is an eigenvector of A ($Av = \mu v$) such that $v^T\tilde{u} = 0$ then

$$Bv = Av - \tilde{u}\tilde{u}^T v = Av = \mu v$$

Hence, all eigenvalues of A which are different from $\tilde{\lambda}$ are eigenvalues of B as well. 0 is an eigenvalue of B replacing $\tilde{\lambda}$. If $\tilde{\lambda}$ is a multiple eigenvalue of A, then $\tilde{\lambda}$ is an eigenvalue of B, but the multiplicity is reduced by 1.

QR Algorithm

QR algorithm

$A_0 := A$
for $m = 0, 1, 2, \ldots$ until convergence **do**
  Factorize $A_m = Q_m R_m$
  $A_{m+1} = R_m Q_m$
end for

$$A_{m+1} = R_m Q_m = Q_m^T (Q_m R_m) Q_m = Q_m^T A_m Q_m$$

Hence, all A_m are (orthogonally) similar, and therefore they have the same eigenvalues.

QR algorithm ct.

If the eigenvalues of A are pairwise different of each other in modulus,

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

and if a further technical condition is satisfied, then the QR algorithm converges in the following sense:

If $(A_m)_{jk} = a_{jk}^{(m)}$, then

$$\lim_{m\to\infty} a_{jk}^{(m)} = 0 \quad \text{for} \quad j > k$$

$$\lim_{m\to\infty} a_{jj}^{(m)} = \lambda_j \quad \text{for} \quad j = 1, \ldots, n$$

QR algorithm and power method

With

$$U_m = Q_1 Q_2 \cdot \cdots \cdot Q_m, \;\; S_m = R_m R_{m-1} \cdot \cdots \cdot R_1$$

it holds

$$A^m = U_m S_m. \qquad (*)$$

For $m = 1$ the statement is trivial: $A = Q_1 R_1 = U_1 S_1$.

$A_{m+1} = R_m Q_m = Q_m^T A_m Q_m$ yields by induction $A_{m+1} = U_m^T A U_m$.

QR algorithm and power method ct.

If $(*)$ is valid for some $m-1$, then it follows from the definition of A_{m+1}

$$R_m = A_{m+1} Q_m^T = U_m^T A U_m Q_m^T = U_m^T A U_{m-1}$$

Multiplying by S_{m-1} from the right und by U_m from the left we obtain

$$U_m S_m = A U_{m-1} S_{m-1} = A^m$$

which is the proposition for m.

From $(*)$ we obtain for the first unit vector e^1 and $\rho = (R_m)_{(1,1)}$

$$A^n e^1 = U_m R_m e^1 = \rho U_m e^1 .$$

Hence, the first column has the same direction as the m-th iterate of the power method with initial vector e^1, and it is not surprising that r_{11} converges to the largest eigenvalue of A in modulus and the first column to a corresponding eigenvector.

Examples

For

$$A = \begin{pmatrix} 1 & -1 & -1 \\ 4 & 6 & 3 \\ -4 & -4 & -1 \end{pmatrix}$$

the upper triangular form appears after approximately 10 steps, and the diagonal elements are in the right order.

For

$$B = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 3 & -1 \\ -2 & -2 & 2 \end{pmatrix}$$

the upper triangular form is arrived after approximately 20 steps, but the diagonal elements are not ordered by magnitude (So, the technical condition of the last Theorem is not satisfied).
After further 50 steps the diagonal elements are ordered by magnitude.

QR algorithm with shifts

QR algorithm with shifts

$A_0 := A$
for $m = 0, 1, 2, \ldots$ until convergence **do**
  Choose a suitable shift κ_m
  Factorize $A_m - \kappa_m I = Q_m R_m$
  $A_{m+1} = R_m Q_m + \kappa_m I$
end for

Again all matrices A_m are similar

$$\begin{aligned} A_{m+1} &= R_m Q_m + \kappa_m I = Q_m^T (Q_m R_m) Q_m + \kappa_m I \\ &= Q_m^T (A_m - \kappa_m I) Q_m + \kappa_m I = Q_m^T A_m Q_m. \end{aligned}$$

and therefore all eigenvalues of the matrices A_m coincide.

Choice of shifts

Let Q_j and R_j be the orthogonal and upper triangular matrices obtained in the QR algorithm with shifts κ_j, and let

$$U_m = Q_1 Q_2 \cdot \cdots \cdot Q_m, \; S_m = R_m R_{m-1} \cdot \cdots \cdot R_1.$$

Then

$$U_m S_m = (A - \kappa_m I)(A - \kappa_{m-1} I) \cdot \cdots \cdot (A - \kappa_1 I). \qquad (+)$$

From $A_{m+1} = Q_m^T A_m Q_m$ it follows immediately by induction $A_{m+1} = U_m^H A U_m$.

For $m = 1$ equation (+) reads

$$U_1 S_1 = Q_1 R_1 = A - \kappa_1 I$$

which is the decomposition in the first step of the QR algorithm with shifts.

Choice of shifts ct.

Assume that (+) holds for some $m-1$. From the definition of A_{m+1} follows

$$R_m = (A_{m+1} - \kappa_m I)Q_m^T = U_m^T(A - \kappa_m I)U_m Q_m^T = U_m^T(A - \kappa_m I)U_{m-1}.$$

Multiplying with S_{m-1} from the right and U_m from the left one obtains

$$U_m S_m = (A - \kappa_m I)U_{m-1}S_{m-1} = (A - \kappa_m I)(A - \kappa_{m-1} I) \cdot \cdots \cdot (A - \kappa_1 I).$$

Choice of shifts ct.

From (+) one gets for the last unit vector e^n

$$(A^T - \kappa_m I)^{-1} \cdot \ldots \cdot (A^T - \kappa_1 I)^{-1} e^n = U_m (S_m^T)^{-1} e^n$$

Since S_m^T and $(S_m^T)^{-1}$ are lower triangular matrices, it holds that

$$U_m (S_m^T)^{-1} e^n = \sigma U_m e^n \quad \text{for some } \sigma.$$

Hence

$$(A^T - \kappa_m I)^{-1} \cdot \ldots \cdot (A^T - \kappa_1 I)^{-1} e^n = \sigma U_m e^n$$

and the last column of U_m can be interpreted as the result of m steps of inverse iteration with shifts $\kappa_1, \ldots, \kappa_m$ and initial vector e^n
This suggests to choose $\kappa_m = a_{n,n}^{(m)}$ which is expected to converge to λ_n.

Reducing the cost

The most expensive part in the QR algorithm (shifted or not) is the computation of the QR factorization in every step.

This cost can be reduced considerably, if the matrix is transformed to upper Hessenberg form first:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2,n-1} & a_{2n} \\ 0 & a_{32} & a_{33} & \dots & a_{3,n-1} & a_{3n} \\ \vdots & & \ddots & \ddots & & \\ \vdots & & & \ddots & \ddots & \\ 0 & 0 & 0 & \dots & a_{n.n-1} & a_{nn} \end{pmatrix}$$

A has upper Hessenberg form, if $a_{jk} = 0$ for $j > k1$.

Reducing the cost ct.

Assume that A_m has upper Hessenberg form. Then a QR decomposition can be obtained in the following way:

Multiply A_m from the left by a rotation in the plane spanned by the first two unit vectors e^1 and e^2, i.e. by a matrix

$$U_{12} = \begin{pmatrix} \cos\theta & \sin\theta & 0 & 0 & \dots & 0 \\ -\sin\theta & \cos\theta & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Then $U_{12}A_m$ contains in its first two rows linear combinatiosn of the first two rows of A, and the rows $3, \dots, n$ are the same as in A_m. The rotation angle can be chosen such that the element in the position $(2,1)$ is annihilated.

Reducing the cost ct.

Multiplying $U_{12}A_m$ from the left by a rotation matrix U_{23} corresponding to rows 2 and 3, we annihilate the element in position $(3,2)$, which does not change the element 0 in the $(2,1)$ position.

Continuing that way we annihilate the elements in positions $(i+1,1)$ by a rotation $U_{i,i+1}$ in the plane spanned by e^i and e^{i+1}.

We finally arrive at

$$U_{n-1,n} \cdot \ldots \cdot U_{23} U_{12} A_m = R, \quad \text{i.e. } A_m = QR,\ Q = U_{12}^T \cdot \ldots \cdot U_{n-1,n}^T.$$

Reducing the cost ct.

$$A_{m+1} = RQ = RU_{12}^T \cdot \cdots \cdot U_{n-1,n}^T$$

Multiplying R by U_{12}^T combines the first two columns of R and leaves the other columns unchanged. Multiplying by U_{23}^T combines columns 2 and 3 and leaves the other ones unchanged, etc.

Obviously

$$A_{m+1} = RU_{12}^T \cdot \cdots \cdot U_{n-1,n}^T$$

becomes an upper Hessenberg matrix.

Reduction to Hessenberg form

A given matrix can be transformed to upper Hessenberg form using Householder matrices.
For

$$A = \begin{pmatrix} a_{11} & c^T \\ b & B \end{pmatrix}, \; B \in \mathbb{R}^{(n-1)\times(n-1)}, \; b, c \in \mathbb{R}^{n-1}$$

let $w \in \mathbb{R}^{n-1}$, $\|w\| = 1$ such that the Householder matrix $Q_1 = I - 2ww^T$ maps b to a multiple of the first unit vector in $\mathbb{R}^{n-1}$.
Then with $P_1 = \begin{pmatrix} 1 & 0 \\ 0 & Q_1 \end{pmatrix}$ we get

$$A_1 := P_1 A P_1 = \begin{pmatrix} a_{11} & c^T Q_1 \\ k & \\ 0 & \\ \vdots & Q_1 B Q_1 \\ 0 & \end{pmatrix}$$

and the first column already has obtained the desired form.
The following columns can be tranformed in a similar way.