Selected Exercises {Main}

1 SF2524BlockA.A8

Prove the following.

Theorem: Suppose $(\lambda, x) \in \mathrm{R} \times \mathrm{R}^n$ is an eigenpair of $A \in \mathrm{R}^{n \times n}$ and assume $\|x\| = 1$. Let $v = x + \Delta$ be an eigenvector approximation. Then, for sufficiently small $\|\Delta\|$,

$$r(v) - \lambda = \begin{cases} \mathcal{O}(\|\Delta\|^2) & \text{if } A^T x = \lambda x \\ \mathcal{O}(\|\Delta\|) & \text{otherwise} \end{cases}$$

.

In particular, if $A^T = A$ then

$$r(v) - \lambda = \mathcal{O}(\|\Delta\|^2)$$

The proof was done in Lecture 1.

Edited by EJ, Anonymous

2 SF2524BlockA.A37

We know how to approximate the largest eigenvalue λ_1 with the power method. If λ_1 and x_1 are known, show that applying the Power Method to the starting vector $q_0 = (A - \lambda_1 I)q$ (with q an arbitrary vector) leads to an approximation of λ_2.

Edited by Anonymous

3 SF2524BlockA.A16

In the power method it is assumed that the eigenvalues are distinct in modulus (ASM1, lecture 1). What happens when this assumption is removed? For example, consider the case when the two largest eigenvalues are identical.

Edited by DE

4 SF2524BlockA.A23

A matrix has eigenvalues $\alpha, 10.1 + i, 10.2 - i, \ldots, 100 + i, 110$. To which eigenvalue will the power method in general converge if...

a) $\alpha = 20$?
b) $\alpha = -20$?
c) $\alpha = 200$?
d) $\alpha = 10.1 - i$?

This exercise is the same A27 but for the Arnoldi method.

Edited by EJ

5 SF2524BlockA.A3

Please show and discuss the following, assuming that A is a square matrix:

1. If (λ, x) is an eigenpair of A and A^{-1} exists, then $(1/\lambda, x)$ is an eigenpair of A^{-1}.
2. If (λ, x) is an eigenpair of A and $\mu \in \mathrm{C}$, then $(\lambda - \mu, x)$ is an eigenpair of $A - \mu I$.
3. Supposing that $A - \mu I$ is invertible, how are the eigenvalues of $(A - \mu I)^{-1}$ related to the eigenvalues of A?

Edited by kj, EJ

6 SF2524BlockA.A9

A matrix A has eigenvalues $1, 2, 3, 4, 5, 5 + 5i, 5 - 5i$. To which eigenvalue will inverse iteration in general converge if the shift is selected as

a) $\mu = 1.2$
b) $\mu = 100$

Edited by EJ

7 SF2524BlockA.A7

Consider the matrix $A = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}$.

How many iterations are needed to approximate the eigenvector corresponding to the largest eigenvalue up to a tolerance of 10^{-3}, i.e. $\|v^{(k)} - x_1\| \approx 10^{-3}$, using

- the power method
- inverse iteration with $\mu = 2.6$
- inverse iteration with $\mu = 2.9$?

Edited by LL, EJ

8 SF2524BlockA.A17

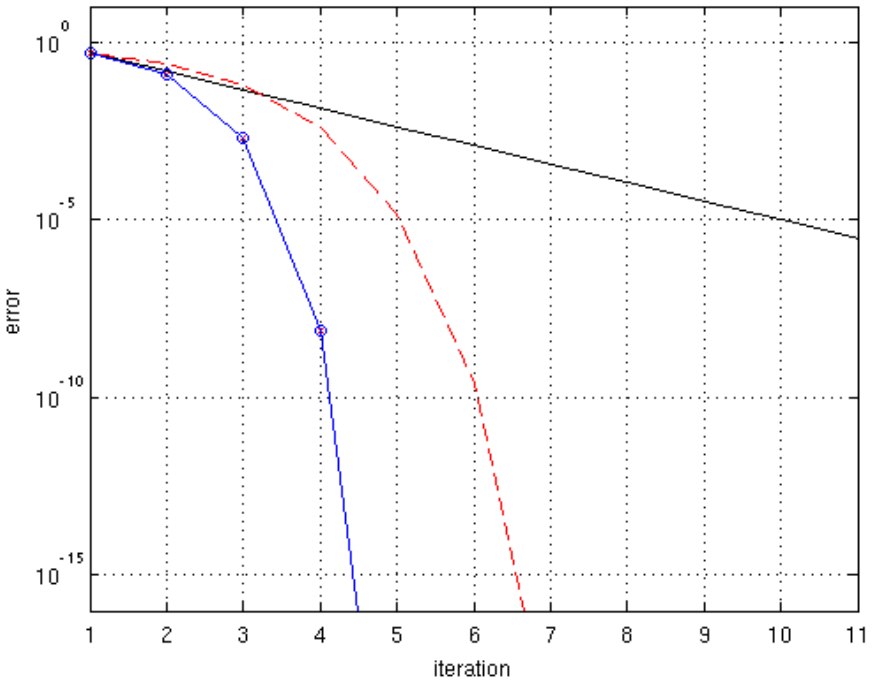
Match convergence predicted by convergence theory with appropriate curve for method...

1. Rayleigh quotient iteration for the matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

2. inverse iteration

3. Rayleigh quotient iteration for the matrix $A = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$

Justify your answer.



Edited by EJ

9 SF2524BlockA.A1

Describe the Rayleigh qoutient iteration.

Edited by EJ, SJ, EJ, Anonymous

10 SF2524BlockA.A4

What is the result of two steps of Rayleigh quotient iteration applied to the matrix

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

with the starting vector

$$x_0 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} ?$$

Edited by Anonymous

11 SF2524BlockA.A24

In what sense does the Rayleigh quotient iteration work better for symmetric matrices than for non-symmetric matrices?

Edited by Anonymous

12 SF2524BlockA.A5

Suppose $A = A^T$ and $v = x + \Delta$ where x is an eigenvector and $\|\Delta\|$ is of order of magnitude 0.2. How large is the error in $r(v)$ approximately?

Edited by EJ

13 SF2524BlockA.A11

The Rayleigh quotient iteration was shown to have a very fast convergence towards some eigenvector, in the symmetric case, $A^T = A$, we have

$$\|v^{(k+1)} \pm x_j\| = \mathcal{O}(\|v^{(k)} \pm x_j\|^3) \quad \text{and} \quad |\lambda^{(k+1)} - \lambda_j| = \mathcal{O}(|\lambda^{(k)} - \lambda_j|^3)$$

However, which eigenpair (λ_j, x_j) is approximated depends on the initial guess and is not fixed.

It is possible to obtain a second eigenpair of $A = A^T$ by applying the iteration to the matrix $A_1 = A - \lambda_j x_j x_j^T$. Why?

Edited by M, LL, EJ

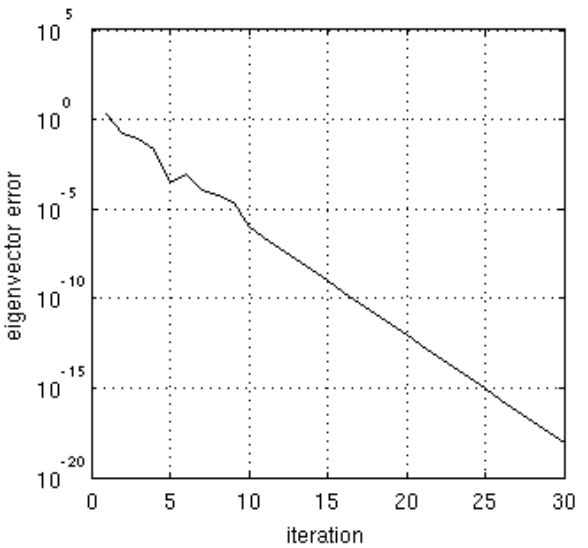
14 SF2524BlockA.A37

We know how to approximate the largest eigenvalue λ_1 with the power method. If λ_1 and x_1 are known, show that applying the Power Method to the starting vector $q_0 = (A - \lambda_1 I)q$ (with q an arbitrary vector) leads to an approximation of λ_2.

Edited by Anonymous

15 SF2524BlockA.A31

The power method applied to a specific matrix $A \in \mathbf{C}^{n \times n}$ gives the following convergence plot.



Let $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|$. The largest eigenvalue is $\lambda_1 = -4$. What is $|\lambda_2|$?

You may make the standard assumptions for the convergence theory of the power method are satisfied (such as $a_1 \neq 0$ and $a_2 \neq 0$).

Explain what information you extract from the figure.

Edited by EJ

16 SF2524BlockA.A47

Simplify the following expressions where c and x are vectors in R^n

a) $\frac{\partial}{\partial x}(c^T x)$

b) $\frac{\partial}{\partial x}(x^T x)$

where $\frac{\partial}{\partial x} = [\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}]$.

Edited by ML

17 SF2524BlockA.A62

Show that the Rayleigh quotient, for a given eigenvector approximation v, minimizes

$$\min_{\alpha \in \mathrm{R}} ||Av - \alpha v||_2 = ||Av - r(v)v||_2$$

Edited by mb

18 SF2524BlockA.A61

Provide support for the claim:
Any eigenvalue solver must be iterative

Edited by mb

19 SF2524BlockA.A15

We run the following matlab code.

```
>> Q=[1 -1; 1 1 ; 0 0]/sqrt(2);
>> w=[1;-1;1e-8];
>> h=Q'*w
>> y=w-Q*h
>> beta=norm(y);
>> qnew=y/beta;
>> Qnew=[Q,qnew]
>> norm(Qnew'*Qnew-eye(3))
ans =
 2.2888e-08
```

a) Which numerical method is this?
b) What do we expect "ans" to be if we would not have rounding errors?
c) What other methods have we learned about in this course that can solve this problem, and potentially be less sensitive to rounding errors?

Edited by EJ, Anonymous

20 SF2524BlockA.A12

Describe the modified Gram-Schmidt method.

Edited by EJ, Anonymous

21 SF2524BlockA.A63

Compare the efficiency of classical and modified Gram-Schmidt in terms of the number flops the respective algorithm require.

Edited by bo

22 SF2524BlockA.A13

Given a subspace $\mathrm{span}(q_1, \ldots, q_m)$ of which $q_1, \ldots, q_m$ is an orthonormal basis. How can the Rayleigh-Ritz approximation of eigenvalues of $A \in \mathrm{R}^{n \times n}$ corresponding to this subspace be computed?

Edited by EJ

23 SF2524BlockA.A14

Suppose the matrix Q_m and $\underline{H}_m$ satisfy the Arnoldi relation

$$AQ_m = Q_{m+1}\underline{H}_m$$

where $Q_{m+1} = [Q_m, q_{m+1}] \in \mathrm{C}^{n \times (m+1)}$ is an orthogonal matrix and $\underline{H}_m \in \mathrm{C}^{(m+1) \times m}$ a Hessenberg matrix.

Show that $H_m = Q_m^T A Q_m$ where $H_m \in \mathrm{C}^{m \times m}$ is the upper submatrix of $H_{\_m}$.

Proven during Lecture 3.

Edited by EJ

24 SF2524BlockA.A10

What is the Arnoldi factorization?

Edited by EJ

25 SF2524BlockA.A20

In the Arnoldi method, what can we say about the eigenvalues of H when A is nilpotent? (A nilpotent if $\exists k > 0 \mid A^n = 0$ for all $n > k$)

Edited by FI

26 SF2524BlockA.A21

Why has there been the need for the algorithm modifications of Gram-Schmidt, like Modified Gram-Schmidt and Repeated Gram-Schmidt?

Edited by FI

27 SF2524BlockA.A25

Prove the following (proven during Lecture 3).

Lemma: Suppose Q_m and H_m satisfy the Arnold relation and suppose that the Krylov matrix

$$K_m(A, b) := [b, Ab, \ldots, A^{m-1}b]$$

has full column rank $(\operatorname{rank}(K_m(A, b)) = m)$, then the columns of Q_m span a Krylov subspace, in other words

$$\operatorname{span}(Q_m) = \operatorname{span}(K_m(A, b)) = \mathcal{K}_m(A, b).$$

Edited by EJ

28 SF2524BlockA.A28

What is the relationship between the Arnoldi method and Lanczos in exact arithmetic?

EJ's addition: "In exact arithmetic" means that we assume that we do not have any rounding errors.

Edited by jr, EJ

29 SF2524BlockA.A29

Suppose $\lambda_1 = 1$ and $\lambda_2 = \lambda_3 = \cdots = \lambda_n = 3$. Give an explicit expression for the m-dependent coefficient in the convergence bound of Arnoldi's method for eigenvalue problems. In other words, find an explicit formula for

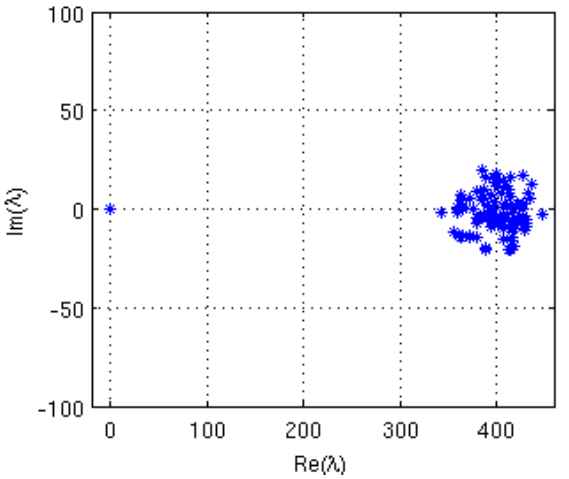
$$\epsilon_1^{(m)} = \min_{p \in P_{m-1}, p(\lambda_1)=1} \max\left(|p(\lambda_2)|, \ldots, |p(\lambda_n)|\right)$$

in this case

Edited by EJ

30 SF2524BlockA.A32

A matrix A has eigenvalues distributed as in the figure below.



Which eigenvalue/eigenvalues can be well approximated with...

a) The power method
b) The Arnoldi method applied to A
c) The Arnoldi method applied to $(A - \mu I)^{-1}$ where $\mu = 300$

Justify your answer.

Edited by EJ

31 SF2524BlockA.A34

Prove that if two rectangular matrices $W \in \mathrm{R}^{n \times m}$ and $V \in \mathrm{R}^{n \times m}$ with $n \geq m$ are related by: $W = VP$ for some non-singular matrix $P \in \mathrm{R}^{m \times m}$, then span($W$) = span($V$). This is used in lecture notes, 1.3 in Gram-Schmidt procedures.

Edited by Ni, Anonymous

32 SF2524BlockA.A35

In the convergence analysis for the Arnoldi method, the quantity $\|(I - QQ^T)x_i\|$ is considered as an indicator of error in eigenvector x_i.
Justify this indicator with a geometric reasoning.

Edited by M

33 SF2524BlockA.A30

Describe the Lanczos iteration.

a) How the upper Hessenberg matrix will look like.
b) How an algorithm is build.
c) For what type of matrices are Lanczos iteration used?

Edited by Anonymous

34 SF2524BlockA.A38

How are the eigenvalue approximations computed from the Arnoldi factorization in Arnoldi's method for eigenvalue problems?

Edited by EJ, Anonymous

35 SF2524BlockA.A40

Give an intuitive interpretation of the quantity ϵ_i^m which is defined in Theorem 1.4.2 (Lecture Notes).
Show two examples, one in which this quantity is small (favorable) and one in which this quantity is large (not-favorable).

Edited by M

36 SF2524BlockA.A41

Let the Arnoldi iteration (see [TB] page 252) yield $h_{k+1,k} = 0$ after k steps. Then $AQ_k = Q_k H_k$. Let (λ, x) be an eigenpair of the upper Hessenberg matrix H_k. Show that (λ, v) is an eigenpair of A and specify the eigenvector v.

Edited by st

37 SF2524BlockA.A43

Based on the arnoldi.m code on the web page, how do the Q and H-matrices differ when calling

```
[Q,H]=arnoldi(A,b,10)
```

and

```
[Q,H]=arnoldi(A,2*b,10)
```

Give an explanation (not only a matlab-code).

Edited by EJ

38 SF2524BlockA.A49

Suppose that $U \in \mathrm{C}^{n\times m}$ and $V \in \mathrm{C}^{n\times m}$ are two orthonormal basis of the same subspace. We know from question A34 that there exists a non-singular matrix $R \in \mathrm{C}^{m\times m}$ such that $UR = V$. Show that R is orthogonal, that is $R^*R = RR^* = I_m$.

Edited by Be, An

39 SF2524BlockA.A51

Give a general description of the eigenvalues that are favored by the power method and by the Arnoldi method, respectively.

Edited by ML

40 SF2524BlockA.A52

Theorem 1.4.2 in the lecture notes states that

$$||I - Q_m Q_m^T x_i|| \leq \xi_i \epsilon_i^m$$

In the proof of this theorem are the following identities used:

$$min_{z\in C^m} ||u - Qz||_2 = ||(I - QQ^T)u||$$

$$min_{z\in C^m} ||\alpha_i x_i - Qz|| = min_{y\in K_m(A,b)} ||\alpha_i x_i - y||$$

Explain why those two identities hold. The proof of theorem 1.4.2 was also done during the lecture of 8 of November.

Edited by Anonymous

41 SF2524BlockA.A53

Proof (show) the convergence of Shift and Invert Arnoldi and why this is faster than normal Arnold

Edited by Anonymous

42 SF2524BlockB.B1

What is the (minimization) definition of the iterates of GMRES?

Edited by EJ

43 SF2524BlockB.B4

Prove that the residual norm of the GMRES-iterates are non-increasing based on the definition of GMRES-iterates as defined in B1.

Edited by EE, EJ

44 SF2524BlockB.B5

In GMRES, the convergence can be estimated by a convergence factor as follows:

$$\frac{||Ax_n - b||}{||b||} \leq ||V|| \, ||V^{-1}|| \min_{p \in P_n^0} \max_{i=1,\ldots,m} |p(\lambda_i)|$$

Prove this and describe another way to compute the convergence.

Edited by Anonymous

45 SF2524BlockB.B7

Prove that for any matrix A and invertible matrix V we have

$$\mathcal{K}_m(V^{-1}AV, b) = V^{-1}\mathcal{K}_m(A, Vb).$$

Edited by EJ

46 SF2524BlockB.B10

Let Q_k be the orthogonal matrix of the Arnoldi method. Explain why $Q_k^* b = ||b|| e_1$

Edited by JW

47 SF2524BlockB.B11

Show that if D is a diagonal matrix and p any polynomial then:

$$||p(D)||_2 = \max_i |p(D_{i,i})|$$

You *may* find the following useful:

$$||p(D)||_2 = \max_{||x||_2 = 1} ||p(D)x||_2$$

(This equality is used in the convergence estimates)

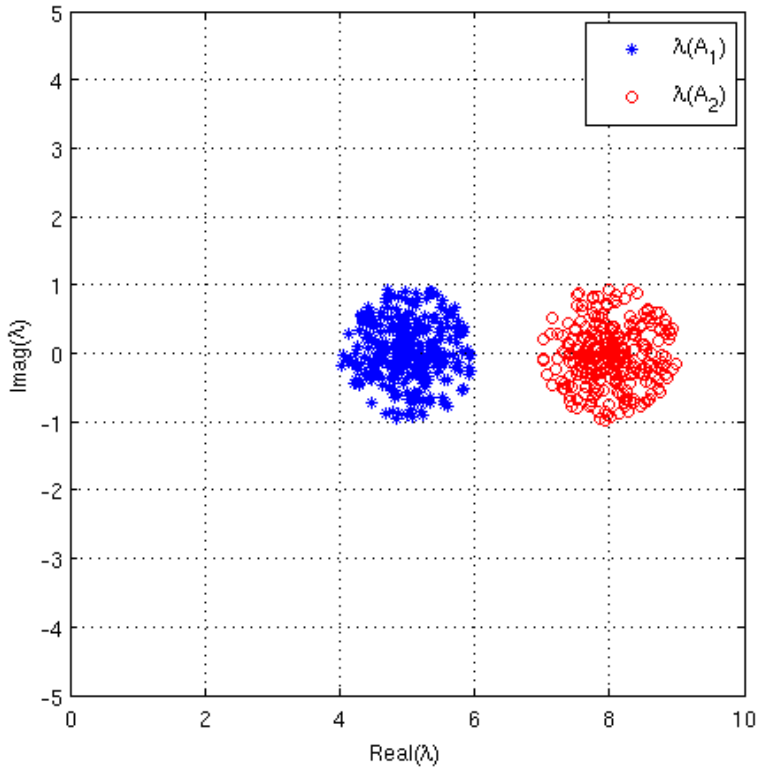
Edited by JW

48 SF2524BlockB.B12

Give bounds of

$$\min_{p \in P_n^0} \left(\max_i \left(|p(\lambda_i(A))| \right) \right)$$

for the matrices A_1 and A_2. Justify clearly what properties of the figure you extract.



Edited by EJ

49 SF2524BlockB.B13

Assume GMRES is applied to a symmetric positive definite matrix. How does the convergence bound (Main convergence theorem of GMRES) simplify? (Hint: Compare with the right-hand side of the CG bound TB Thm 38.3 and the following deductions)

Edited by LL

50 SF2524BlockB.B14

Prove that GMRES will terminate in a finite number of steps.

Edited by CR, An

51 SF2524BlockB.B15

Suppose two matrices are related by $B = -A$. Consider the linear systems $Ax = b$ and $Bz = b$. Prove that the approximation error of GMRES is the same for both problems.

Edited by An

52 SF2524BlockB.B16

Suppose the matrix $A \in \mathrm{R}^{m \times m}$ is diagonalizable, but that many eigenvalues are equal. More precisely, assume that it only has 10 different eigenvalues (and $m \gg 10$). Use the min-max bound for GMRES to show that GMRES returns the exact solution after 10 iterations. You may assume that it does not break down before that.

Edited by EJ

53 SF2524BlockB.B18

Does GMRES have disadvantages in terms of computation time? Explain your answer.

Edited by M

54 SF2524BlockB.B46

Prove that

$$x \in \mathcal{K}_n(A, b) \iff b - Ax = p(A)b \quad \text{for some } p \in P_n^0$$

Edited by M

55 SF2524BlockB.B37

Lemma 2.1.3 in the lecture notes states that for any $A \in C^{m \times m}$ and $b \in C^n$

$$\{b - Ax : x \in K_n(A, b)\} = \{p(A)b : p \in P_n^0\}$$

where $p(A)$ and P_n^0 are defined in the lecture notes.

Describe what is meant by that two sets are equal to each other, like the ones above.

Edited by Anonymous

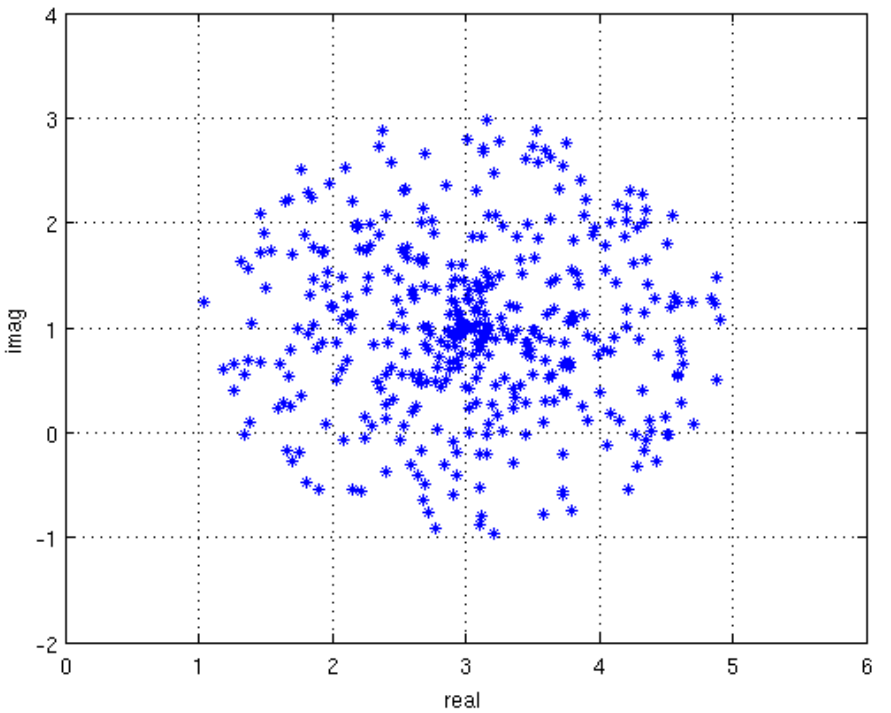
56 SF2524BlockB.B64

Let $\lambda_1, \ldots, \lambda_m$ be eigenvalues of a matrix A. The eigenvalues are given in the figure below.

a) Provide a bound for the quantity

$$\min_{p \in P_n^0} \max_{i=1, \ldots, m} |p(\lambda_i)|$$

b) How is this related to the convergence of GMRES?



Edited by EJ

57 SF2524BlockB.B65

Given $z \in \mathrm{R}^n$ and an orthogonal matrix $Q \in \mathrm{R}^{m \times n}$, with $m \geq n$, show that $||Qz||_2^2 = ||z||_2^2$.

Edited by ML, JW

58 SF2524BlockB.B2

Suppose A is symmetric positive definite.

What is the relationship between the approximations generated by GMRES and the

approximations generated by CG?

Edited by jr

59 SF2524BlockB.B3

State the minimization property of CG (sometimes also used as a definition of CG).

Edited by jr

60 SF2524BlockB.B44

Under certain conditions CG has in a sense a monotone convergence property. State sufficient conditions and prove the monotonic convergence property of CG.

Edited by DE, JW, EJ

61 SF2524BlockB.B21

In the justification of the CG-method it is claimed that the residual norm with respect to the A^{-1} -norm equals the error norm with respect to the A-norm. Formalize and prove this.

Edited by EJ

62 SF2524BlockB.B22

TB Thm 38.5 contains a characterization of convergence of CG with condition number K. From this result, derive an expression for the number of iterations required to reach a specified accuracy, for large condition numbers K.

A sketch of this was given in lecture 7.

Edited by EJ

63 SF2524BlockB.B23

We store the vectors generated in CG in matrix form:

$$X = [x_1, \ldots, x_n],\ R = [r_0, \ldots, r_{n-1}],\ P = [p_0, \ldots, p_{n-1}].$$

Derive three expressions between X, R and P from the individual steps of the CG-algorithm (in the lecture notes or in TB).

Edited by EJ

64 SF2524BlockB.B19

Prove that the residual vectors r_n of CG are orthogonal and the update vectors p_n of CG are A-orthogonal.

Edited by M

65 SF2524BlockB.B24

We usually say that the Lanczos method and CG-method are "short-term recurrence methods". What is the meaning of "short-term recurrence" and why is it important?

Edited by EJ

66 SF2524BlockB.B25

a) Why does $\|z\|_A = \sqrt{z^T A z}$ only define a norm if A is symmetric positive definite?

b) Why is $\|z\|_{A^{-1}}$ a norm if A is symmetric positive definite?

Edited by LL, Anonymous

67 SF2524BlockB.B26

In general, the condition number of a matrix A is defined as $\mathcal{K}(A) = \|A^{-1}\|\,\|A\|$. Show that for

normal matrices $\mathcal{K}(A) = \frac{|\lambda_{max}(A)|}{|\lambda_{min}(A)|}$.

Edited by LL, EJ

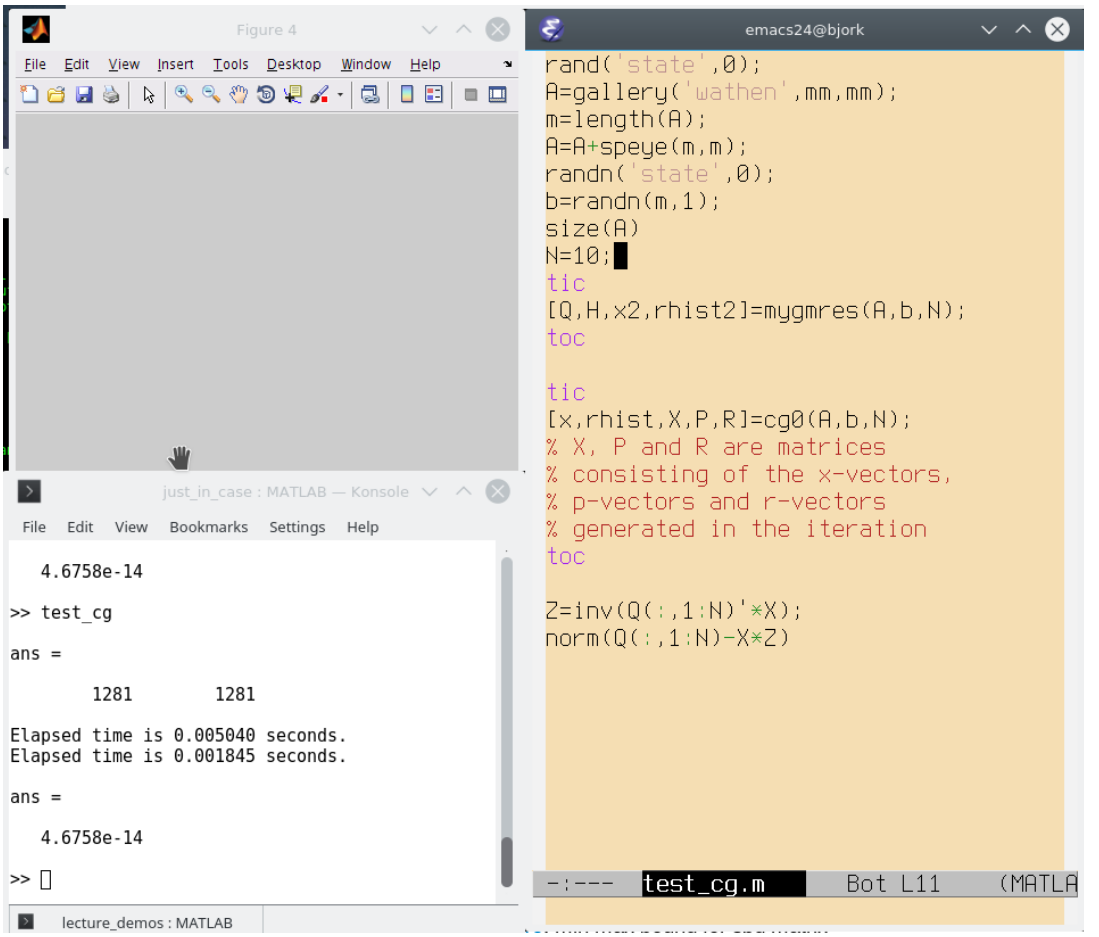
68 SF2524BlockB.B27

Let A be symmetric and real. Show that $\sigma(A) = |\lambda(A)|$.

Edited by st

69 SF2524BlockB.B28

In the screendump below, why is the result zero (or almost zero)? Justify your reasoning by relating it to a theorem in the course.



Edited by EJ, Anonymous

70 SF2524BlockB.B43

Application of CG. Solve by using a CG implementation the unconstrained quadratic optimization problem,

$$\min_{x \in \mathrm{R}^4} \frac{1}{2} x^T H x + c^T x,$$

where the Hessian, H, is symmetric positive definite and given by,

$$H = \begin{bmatrix} 5 & 2 & 3 & 2 \\ 2 & 3 & 1 & 1 \\ 3 & 1 & 4 & -1 \\ 2 & 1 & -1 & 7 \end{bmatrix}$$

and $c = \begin{bmatrix} -2 & 8 & -2 & 4 \end{bmatrix}^T$. Hint: Since the hessian is positive definite, the problem is convex and

thus the global optimal solution is given by the optimality conditions, $Hx + c = 0$.

Edited by DE

71 SF2524BlockB.B47

Let e_n be the error in step n of CG. Show that

$$||e_{n+1}||_A \leq ||e_n||_A$$

Edited by mb

72 SF2524BlockB.B48

In the Conjugated Gradient algorithm, what is the meaning of the matrices P and R? What is the relation between them? Under what hypothesis do they have the same span, and why?

Edited by FI

73 SF2524BlockB.B59

Below is a different formulation of an algorithm in this course (from the book Fundamentals of matrix computations [**1**]):
a) Which algorithm?
b) How are the variables related to the variables in the standard formulation of the algorithm (make references to the lecture notes PDFs)?

$r \leftarrow r - Ax$

$p \leftarrow r$

$\nu \leftarrow r^T r$

$k \leftarrow 0$

do until converged or $k = l$

$q \leftarrow Ap$

$\mu \leftarrow p^T q$

$\alpha \leftarrow \nu / \mu$

$x \leftarrow x + \alpha p$

$r \leftarrow r - \alpha q$

$\nu_+ \leftarrow r^T r$

$\beta \leftarrow \nu_+ / \nu$

$p \leftarrow r + \beta p$

$\nu \leftarrow \nu_+$

$k \leftarrow k + 1$

if not converged, set flag

Edited by EJ

74 SF2524BlockB.B62

Take $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ and $b = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$.

We solve $Au = b$.

Show that CG applied to this system should converge in 1 or 2 iterations.

Edited by EJ

75 SF2524BlockB.B30

What is the relationship between CG and CGN (sometimes referred to as CGNE)?

Edited by EJ

76 SF2524BlockB.B32

State the condition number bound for CGN. How is it different from the condition number bound for CG?

Edited by An

77 SF2524BlockB.B31

Prove that the iterates of CGN are minimizers of the residual with respect to the two-norm (similar to GMRES). Unlike GMRES, the mininimization is done over space X. What is X?

Edited by EJ

78 SF2524BlockB.B33

Explain a situation where CGN may be more computationally competitive than GMRES?

Edited by EJ

79 SF2524BlockB.B39

Consider the preconditioner $M^{-1} = LL^T$ for some non-singular matrix L.

a) Show that applying M^{-1} as a left-preconditioner to the CG method results in the following system:

$$L^T A L y = L^T b$$

where x is afterwards obtained from $x = Ly$.

b) Verify that the $L^T A L$ is symmetric positive definite if A is symmetric positive definite.

c) Provide new expressions for the recurrence formulas in the CG method after applying the preconditioner M^{-1}. Note that the formulas can be simplified slightly by introducing $\hat{\mathbf{p}}_n = L^{-1}\mathbf{p}_n$ and $\tilde{\mathbf{r}}_n = M^{-1}\mathbf{r}_n$.

Edited by mb

80 SF2524BlockB.B57

Let $x_0, \ldots$ be the sequence of vectors which are minimizers in the optimization problem

$$\min_{\spadesuit} \|\clubsuit\|_{\heartsuit} = \|Ax_n - b\|_{\diamondsuit}.$$

What are ♣, ♠, ◇ and ♡ for...
a) GMRES
b) CG
c) CGN

Edited by EJ

81 SF2524BlockB.B55

If the norm of $||x||_A$, with $A = I$, show this is equal to taking the 2-norm of x.

Edited by BO

82 SF2524BlockB.B58

This problem was posed on the exam March 2016.

Problem 7 (5p)

Let v_{k+1}, w_{k+1}, be generated by carrying out k steps of Algorithm X where $A \in \mathbb{R}^{n\times n}$.

(a) Prove that v_{k+1} and w_{k+1} are elements of certain Krylov subspaces? Which ones?

(b) Simplify the Algorithm X for the case A is symmetric. Under this symmetry assumption, Algorithm X is equivalent to an algorithm in this course. Which one?

(c) The iterates of the Algorithm X satisfy

$$\begin{aligned} AV_k &= V_{k+1}\underline{T}_k \\ A^T W_k &= W_{k+1}\underline{T}_k \end{aligned}$$

where $\underline{T}_k \in \mathbb{R}^{(k+1)\times k}$. Express V_k, W_k and $\underline{T}_k$ in terms of quantities in the algorithm.

Algorithm X:

1. $\tilde{v}_1 = b - Ax_0,\ v_1 = w_1 = \tilde{v}_1/\|\tilde{v}_1\|$
   for $k = 1, \ldots$ until converged
2. $\quad \tilde{v}_{k+1} = Av_k$
3. $\quad \tilde{w}_{k+1} = A^T w_k$
4. $\quad \alpha_k = w_k^T \tilde{v}_{k+1}$
5. $\quad \tilde{v}_{k+1} = \tilde{v}_{k+1} - \alpha_k v_k$
6. $\quad \tilde{w}_{k+1} = \tilde{w}_{k+1} - \alpha_k w_k$
7. $\quad$ if $k > 1$
8. $\quad\quad \tilde{v}_{k+1} = \tilde{v}_{k+1} - \beta_{k-1} v_k$
9. $\quad\quad \tilde{w}_{k+1} = \tilde{w}_{k+1} - \beta_{k-1} w_k$
10. $\quad \gamma_k = \|\tilde{v}_{k+1}\|,\ v_{k+1} = \tilde{v}_{k+1}/\gamma_k$
11. $\quad \beta_k = \|\tilde{w}_{k+1}\|,\ w_{k+1} = \tilde{w}_{k+1}/\beta_k$

    end

Edited by EJ

83 SF2524BlockC.C1

Describe the basic QR-method.

Edited by EJ

84 SF2524BlockC.Ca-11

Suppose the matrices A and B are related by

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

where V is a non-singular matrix. Prove that A and B have the same eigenvalues.

This is called a similarity transformation.

Edited by EJ

85 SF2524BlockC.C2

Why do we compute the Schur form and not the Jordan form in the context of QR method?

Edited by EJ

86 SF2524BlockC.C3

Prove that the iterates of the basic QR method have the same eigenvalues.

Edited by EJ, An

87 SF2524BlockC.C4

What is the difference between the QR factorization and the Schur factorization?

Edited by EJ

88 SF2524BlockC.C5

The QR factorization is not unique following the definition in the slides. Give a counter example.

Edited by EJ

89 SF2524BlockC.C6

Suppose

$$P = \begin{bmatrix} I & 0 \\ 0 & I - 2uu^T \end{bmatrix} \in \mathrm{R}^{n\times n},$$

where $I \in \mathrm{R}^{p\times p}$ is the identity matrix, $u \in \mathrm{R}^{n-p}$ and $\|u\| = 1$. Give a vector $v \in \mathrm{R}^n$ with $\|v\| = 1$

such that $P = I - 2vv^T$. In other words, show that P is a Householder reflector.

This type of Householder reflector is needed in the Hessenberg reduction (phase 1).

Edited by EJ

90 SF2524BlockC.Ca-7

In Lecture 10 we learned that we could work column by column and with a similarity transformation bring any matrix to Hessenberg form (by applying constructing several Householder reflectors of the type in C6). Can we use the same technique to bring the matrix to triangular form instead of Hessenberg? Why doesn't it work?

This (nice) question was asked by a student during lecture 10. Hint: An indication of what happens can be found in TB.

Edited by EJ

91 SF2524BlockC.Ca-8

Suppose

$$A = \begin{bmatrix} 1 & 4 & 6 \\ 1 & 1 & 1 \\ 0 & 2 & 4 \end{bmatrix}.$$

Let A_k be the iterates of the basic QR-method. What is ...

a) the element (3,1) of A_1 ?

b) the element (3,1) of A_{100}?

Relate to a lemma/theorem in the course.

Edited by EJ

92 SF2524BlockC.Ca-10

Let

$$P = I - 2uu^*$$

be the Householder reflector associated with some $u \in \mathrm{C}^n$ such that $||u|| = 1$. Show that

i) $Pu = -u$

ii) $P^* = P$

Edited by mb

93 SF2524BlockC.Ca-12

Find a Givens rotator $G \in \mathrm{R}^6$ such that

$$G^T \begin{bmatrix} -1 \\ 2 \\ 1 \\ 4 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ \star \\ 0 \\ 0 \end{bmatrix}.$$

What is $\star$?

Edited by EJ

94 SF2524BlockC.Ca-13

Let H be a Hessenberg matrix and let $H = QR$ be its QR factorization. Prove that RQ is again a Hessenberg matrix.

EJ's comment: Somewhat tricky since proof is not explicitly given in lecture notes. Hint for proof: Use theorem 2.2.6 in PDF lecture notes, and look at the structure of $Q = G_1 G_2 \cdots G_{n-1}$.

Hint for another proof: Use $RQ = RHR^{-1}$ where R and R^{-1} are upper-triangular.

Edited by EJ, Be

95 SF2524BlockC.Ca-14

Select ?? in this program such that A becomes upper triangular after the for-loop.

```
m=5;
% Create a Hessenberg matrix A
A=randn(m);  A=triu(A,-1);
for k=1:m-1
   r=??
   c=??
   s=??
   G=givens(i,i+1,c,s,m);
   A=G'*A;
end
```

Edited by EJ

96 SF2524BlockC.Ca-15

Prove that the QR-method preserves symmetry. In other words, show that if A is symmetric, then RQ is symmetric where Q, R corresponds to a QR-factorization of A.

Edited by EJ

97 SF2524BlockC.Ca-17

There are downsides with the basic QR-method and several improvements have been made to improve the performance of the method. What are the downsides of the basic QR-method and what are the ideas behind the improvements covered in the course?

Edited by DE

98 SF2524BlockC.Cb-2

Suppose a matrix has the following block structure

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}.$$

where A_{22} is upper triangular.

Prove that the shifted QR-method preserves the last block row. In other words, let B be the result of one step of the shifted QR-method. Prove that B has the structure

$$B = \begin{bmatrix} \star & \star \\ 0 & A_{22} \end{bmatrix}.$$

Edited by EJ

99 SF2524BlockC.Cb-3

In the lectures we claim that the shifted QR-method is a generalization of the basic QR-method. In what sense is it a generalization?

Edited by EJ

100 SF2524BlockC.Cb-4

Prove that if μ is an eigenvalue of A, then $A - \mu I$ is a singular matrix.

Edited by EJ

101 SF2524BlockC.Cb-5

Disadvantage 2 of the basic QR method, according to the lecture notes, is that it can be arbitrarily slow and is often slow in practice. Give examples for a matrix (its eigenvalues) where the QR method is very slow and one where it is not slow.

Edited by LL

102 SF2524BlockC.Cb-6

Prove that the shifted QR-method preserves symmetry. In other words, show that if A is symmetric, then $RQ + \lambda I$ is symmetric where Q, R corresponds to a QR-factorization of $A - \lambda I$ for some λ.

Edited by mb

103 SF2524BlockC.Cb-7

Given the following matrix

$$A = \begin{pmatrix} -6.05 & -2.15 & -0.51 \\ 36.6 & 11.8 & 1.86 \\ -16.0 & -3.41 & 3.19 \end{pmatrix}$$

with eigenvalues

$$\lambda_1 = 1, \lambda_2 = 3.95, \lambda_3 = 4$$

What is the expected number of iterations required for the basic QR method to converge?

Edited by mb

104 SF2524BlockD.Da-1

Give the Taylor definition of matrix functions.

Edited by EJ

105 SF2524BlockD.Da-2

State the Jordan form definition of matrix functions.

Edited by jr

106 SF2524BlockD.Da-4

Show that

$$\det \exp(A) = \exp(\operatorname{tr} A) .$$

Edited by Be

107 SF2524BlockD.Da-5

Which of the following matrix structures are preserved by matrix functions? Suppose A is

a) symmetric
b) tringular
c) hessenberg
d) anti-symmetric $A^T = -A$
e) diagonal
f) orthogonal

Does $f(A)$ have the same structure a,b,c,d,e?

Edited by EJ

108 SF2524BlockD.Da-6

Let $A \in \mathrm{R}^{n \times n}$ be a given matrix and define

$$p(\lambda) = \det(A - \lambda I) = a_0 + a_1\lambda + \ldots + a_n\lambda^n.$$

We now take the matrix function extension of the polynomial $p(\lambda)$. Derive a closed expression for

$$p(A)$$

Hint: proof of Cayley-Hamilton theorem (which is probably my favorite theorem)

Edited by EJ

109 SF2524BlockD.Da-7

a) Give an explicit formula for

$$f(A_\epsilon)$$

when

$$A_\epsilon = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda + \epsilon \end{bmatrix}.$$

b) What is $\lim_{\epsilon \to 0} f(A_\epsilon)$? Is this consistent with the Jordan definition?

Edited by EJ

110 SF2524BlockD.Da-8

State the Cauchy Integral definition of Matrix Functions.

Edited by Anonymous

111 SF2524BlockD.Da-9

Show that $f(A)f(B) = f(B)f(A)$ if $AB = BA$ for any definition.

Edited by An

112 SF2524BlockD.Da-10

In the first lecture on matrix functions we saw the following application of matrix functions. Derive the explicit formula for $y(t)$ when A is symmetric.

Trigonometric matrix functions and square roots

Suppose $y(t) \in \mathbb{R}^n$ satisfies

$$y''(x) + Au(x) = 0 \;\; y(0) = y_0, \;\; y'(0) = y_0'.$$

The solution is explicitly given by

$$y(t) = \cos(\sqrt{A}t)y_0 + (\sqrt{A})^{-1}\sin(\sqrt{(A)}t)y_0'.$$

Edited by EJ

113 SF2524BlockD.Da-11

When is the Taylor definition of matrix functions a valid definition? Give a sufficient condition for which $f(A)$ is well defined through the Taylor-definition.

Edited by M

114 SF2524BlockD.Da-12

Compute explicitly $\exp(A)$ for

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

by using JCF-definition of matrix functions.

Edited by M

115 SF2524BlockD.Da-13

There are several ways to define matrix functions. Name three different definitions and at least one advantage and at least one consequence for each definition.

Edited by Anonymous

116 SF2524BlockD.Da-15

Prove the property $f\left(\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\right) = \begin{bmatrix} f(A) & 0 \\ 0 & f(B) \end{bmatrix}$, where A and B are $n \times n$ matrices, for a matrix function f.

Edited by st

117 SF2524BlockD.Db-1

Let T be a triangular matrix with distinct eigenvalues and let

$$f(T) = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{13} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}.$$

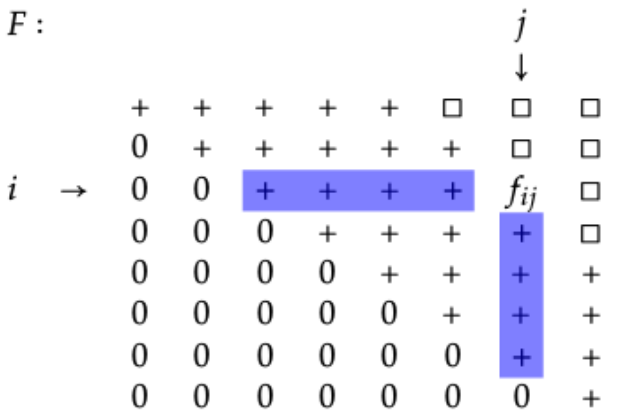
- a) Give the values for $f_{11}, f_{21}, f_{31}, f_{22}, f_{32}, f_{33}$ in term of the entries of T
- b) Derive an explicit formula for f_{12} involving only elements of T and the values computed in (a).

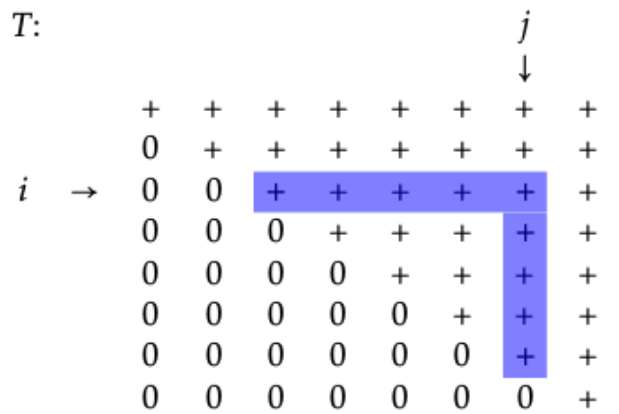
Edited by EJ

118 SF2524BlockD.Db-3

In the PDF-lecture notes, the following illustrates one step in the derivation the Schur-Parlett

method. Interpret the figure. What are the corresponding formulas?





Edited by EJ

119 SF2524BlockD.Db-2

Describe the Schur-Parlett method and the idea behind it.

Edited by DE

120 SF2524BlockD.Db-4

Algorithm 1 in Lecture Notes are an algorithm showing how to use the simplified Schur-Parlett method. This method can be unstable, in what situation does this unstability accur? What can be done to not obtain an unstable algorithm in this case?

Edited by Anonymous

121 SF2524BlockD.Dc-1

Use the interpolation property of matrix functions to derive an explicit formula for

$$\exp\begin{bmatrix} 1 & \alpha \\ 0 & 2 \end{bmatrix}.$$

Edited by jr

122 SF2524BlockD.Dc-4

a) Find two matrices A and B such that

$$e^A e^B \neq e^{A+B}$$

b) Prove that

$$e^{A/2} e^{A/2} = e^A$$

using any of the definitions of matrix functions we have used in the course.

Edited by EJ

123 SF2524BlockD.Dc-5

Use the following MATLAB output and compute an approximation of e^A, as the result of scaling-and-squaring where the initial approximation is computed with a truncated Taylor series. Solve the problem by only adding or subtracting certain matrices or multiplying matrices by scalar values.

```
>> B=eye(n)+A/(16)+A^2/32;
>> BB=B^2
BB=
 1.18380  -0.12917   0.10335
-0.19689   1.27569   0.26426
 0.16304  -0.15383   1.02728
>> BB=BB^2
BB=1.44367  -0.33358   0.19438
-0.44115   1.61216   0.58824
 0.39077  -0.37531   1.03150
>> BB=BB^2
 2.30731  -1.09233   0.28490
-1.11823   2.52545   1.46934
 1.13279  -1.12255   0.91917
>> BB=BB^2
BB =
 6.86790  -5.59880  -0.68577
-3.73966   5.94995   4.74273
 4.91020  -5.10415  -0.48180
>> BB=BB^2
BB =
 64.738  -68.264  -30.933
-24.647   32.132   28.499
 50.445  -55.402  -27.343
```

Edited by EJ

124 SF2524BlockD.Dd-2

In this youtube video from the Gene Golub summer school for PhD students at time-point 01:03:00 Nick Higham provides an example of a non-trivial square root of an identity matrix.

Prove that it is actually a square root of the identity matrix

The video is worthwhile to watch on its own.

Edited by EJ

125 SF2524BlockD.Dd-1

What is the Denman-Beavers iteration?

Edited by EJ

126 SF2524BlockD.De-1

The sign function (for scalar functions) is defined such that $f(x) = -1$ if $x < 0$ and $f(x) = 1$ if $x > 0$. Give a definition a matrix function definition of f consistent with this definition.

Edited by EJ

127 SF2524BlockD.De-2

State a quadratically convergent iterative method for the matrix sign function.

Edited by EJ

128 SF2524BlockD.De-3

Derive a definition for $S = \operatorname{sign}(A)$ function using the Jordan decomposition, and prove the following properties:

$$S^2 = I$$

$$S \text{ is diagonalizable and has eigenvalues } \pm 1$$

Edited by FI

129 SF2524BlockD.Df-1

Describe the method called Krylov method for matrix functions.

Edited by jr

130 SF2524BlockD.Df-2

What is a φ-function? For which values z is $\varphi(z)$ analytic?

Edited by EJ

131 SF2524BlockD.Df-3

Give an explicit solution to the ODE

$$y'(t) = Ay(t) + b$$

for a matrix A and a vector b in terms of φ-function.

Edited by EJ

132 SF2524BlockD.Df-4

What is the forward Euler exponential integrator?

Edited by EJ

133 SF2524BlockA.A8solution

For simplicity, assume $x \in \mathrm{R}^n$ and $\lambda \in \mathrm{R}$. Recall that

$$r(x) = \frac{x^T A x}{x^T x} = \lambda.$$

By Taylor expansion, we have

$$r(v) = r(x + \Delta) = r(x) + r'(x)\Delta + \mathcal{O}(||\Delta||^2),$$

where $r'(x) = \left[\frac{\partial r}{\partial x_1}, \frac{\partial r}{\partial x_2} \ldots, \frac{\partial r}{\partial x_n}\right]$.

Note 1. If $c \in \mathrm{R}^n$ is a constant vector, then

$$\frac{\partial}{\partial x} c^T x = \frac{\partial}{\partial x}\left(c_1 x_1 + c_2 x_2 + \ldots + c_n x_n\right) = \left[c_1, c_2, \ldots, c_n\right] = c^T.$$

Note 2.

$$\frac{\partial}{\partial x} x^T x = \{\text{Product rule}\} = \left(\frac{\partial}{\partial x} y^T x + \frac{\partial}{\partial x} x^T y\right)\Big|_{y=x} = 2x^T,$$

where we have used $\frac{\partial}{\partial x} y^T x = y^T$ and $x^T y = (x^T y)^T = y^T x$ from Note 1 for the first and the second term within the parenthesis, respectively.

Note 3.

$$\frac{\partial}{\partial x} x^T A x = \left(\frac{\partial}{\partial y} y^T A x + \frac{\partial}{\partial x} x^T A y \right) \Big|_{y=x} = \{(x^T A y)^T = y^T A^T x\}$$

$$= \left(\frac{\partial}{\partial x} y^T A x + \frac{\partial}{\partial x} y^T A^T x \right) \Big|_{y=x} = x^T A + x^T A^T.$$

By the quotient rule for differentiation, we have

$$r'(w) = \frac{\left(\frac{\partial}{\partial w} w^T A w \right) w^T w - \left(\frac{\partial}{\partial w} w^T w \right) w^T A w}{(w^T w)^2} =$$

$$\{\text{Note 2, 3}\} = \frac{w^T (A + A^T) w^T w - 2 w^T w^T A w}{(w^T w)^2}.$$

If $w = x$ where x is an eigenvector with $||x|| = 1$, it follows that

$$r'(x) = \frac{x^T (A + A^T) - 2 x^T x^T A x}{(1)^2} = x^T (A + A^T) - 2 x^T \lambda =$$

$$x^T A + (Ax)^T - 2\lambda x^T = \{(Ax)^T = \lambda x^T\} = x^T A - \lambda x^T = (A^T x - \lambda x)^T.$$

Hence, if $A = A^T$ (A is symmetric), then $Ax = \lambda x = A^T x$ and consequently $r'(x) = 0$ iff $A^T x = \lambda x$, Q.E.D.

Edited by jr, EJ

134 SF2524BlockA.A37solution

Let $A \in \mathrm{R}^{m \times m}$ and assume that the eigenvalues of A are distinct in modulus,

$$|\lambda_1| > |\lambda_2| > \ldots |\lambda_m|.$$

The eigenvectors $x_1, \ldots, x_m$ then span R^m.

If λ_1 and x_1 are known and the Power Method is applied to the starting vector $q_0 = (A - \lambda_1 I) q$ (with q an arbitrary vector), then

$$q^{(k)} = \frac{A^k q^{(0)}}{\| A^k q^{(0)} \|} = \frac{A^k (A - \lambda_1) q}{\| A^k (A - \lambda_1) q \|} = \frac{(A^{k+1} - \lambda_1 A^k) q}{\| (A^{k+1} - \lambda_1 A^k) q \|}$$

Expressing q in terms of the eigenvector basis we get

$$q^{(k)} = \frac{(A^{k+1} - \lambda_1 A^k)(a_1 x_1 + \ldots + a_m x_m)}{\| (A^{k+1} - \lambda_1 A^k)(a_1 x_1 + \ldots + a_m x_m) \|} = \frac{(A^{k+1} - \lambda_1 A^k)(a_1 x_1 + \ldots + a_m x_m)}{\| (A^{k+1} - \lambda_1 A^k)(a_1 x_1 + \ldots + a_m x_m) \|}$$

$$= \frac{a_1 \lambda_1^{k+1} x_1 + \ldots a_m \lambda_m^{k+1} x_m - (a_1 \lambda_1^{k+1} x_1 + a_2 \lambda_1 \lambda_2^k x_2 + \ldots + a_m \lambda_1 \lambda_m^k x_m)}{\| a_1 \lambda_1^{k+1} x_1 + \ldots a_m \lambda_m^{k+1} x_m - (a_1 \lambda_1^{k+1} x_1 + a_2 \lambda_1 \lambda_2^k x_2 + \ldots + a_m \lambda_1 \lambda_m^k x_m) \|} =$$

$$= \frac{a_2(\lambda_2-\lambda_1)\lambda_2^k x_2 + \ldots + a_m(\lambda_m-\lambda_1)\lambda_m^k x_m}{\|a_2(\lambda_2-\lambda_1)\lambda_2^k x_2 + \ldots + a_m(\lambda_m-\lambda_1)\lambda_m^k x_m\|} = \frac{a_2\lambda_2^k}{|a_2|\,|\lambda_2|^k} \cdot \frac{(\lambda_2-\lambda_1)x_2 + \frac{a_3(\lambda_1-\lambda_3)\lambda_3^k}{a_2\lambda_2^k}x_3 \ldots + \frac{a_m(\lambda_m-\lambda_1)\lambda_m^k}{a_2\lambda_2^k}x_m}{\|(\lambda_2-\lambda_1)x_2 + \frac{a_3(\lambda_1-\lambda_3)\lambda_3^k}{a_2\lambda_2^k}x_3 \ldots + \frac{a_m(\lambda_m-\lambda_1)\lambda_m^k}{a_2\lambda_2^k}x_m\|}$$

As $k \to \infty$, $(|\frac{\lambda_j}{\lambda_2}|)^k \to 0$ for $j > 2$ as $|\lambda_2| > |\lambda_j|$ and thus

$$q^{(k)} \to \frac{a_2\lambda_2^k}{|a_2|\,|\lambda_2|^k} \cdot \frac{(\lambda_2-\lambda_1)x_2}{\|(\lambda_2-\lambda_1)x_2\|}.$$

Edited by LL

135 SF2524BlockA.A16solution

If $\lambda_1 = \lambda_2$, $|\lambda_1| \geq |\lambda_3| \geq |\lambda_4| \geq \cdots \geq |\lambda_n|$, and the geometric multiplicity of λ_1 is 2, then using the usual procedure: at the iteration k, if we write our starting vector in the basis of the eigenvectors (assuming A diagonalizable),

$$v^{(0)} = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$$

where x_1 and x_2 form a basis for the eigenspace corresponding to the eigenvalue λ_1. We get that the iterate $v^{(k)}$ has the form

$$v^{(k)} = c_k A^k v^{(0)} = c_k A^k(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n) =$$

$$= c_k(\alpha_1\lambda_1^k x_1 + \alpha_2\lambda_1^k x_2 + \alpha_3\lambda_3^k x_3 + \cdots + \alpha_n\lambda_n^k x_n) =$$

$$= c_k\lambda_1^k(\alpha_1 x_1 + \alpha_2 x_2) + c_k\left(\alpha_3\frac{\lambda_3^k}{\lambda_1^k}x_3 + \cdots + \alpha_n\frac{\lambda_n^k}{\lambda_1^k}x_n\right)$$

from which we get the estimate

$$\|v^{(k)} \pm (\beta_1 x_1 + \beta_2 x_2)\| = O\left(\left|\frac{\lambda_3}{\lambda_1}\right|^k\right)$$

and because x_1 and x_2 form a basis for the eigenspace, $\beta_1 x_1 + \beta_2 x_2$ is an eigenvector of λ_1, so PM converges nonetheless.

If the first two eigenvalues have identical module but are not equal ($\lambda_2 = -\lambda_1$), then there is no convergence: using the same equations as before, we get

$$\|v^{(k)} \pm (\beta_1 x_1 + (-1)^k\beta_2 x_2)\| = O\left(\left|\frac{\lambda_3}{\lambda_1}\right|^k\right)$$

so the iterate oscillates between two vectors in the $span(x_1, x_2)$.

If instead the geometric multiplicity of λ_1 is 1, then A is not diagonalizable; using the Jordan decomposition, $A = VJV^{-1}$, where V is the vector of the generalized eigenvectors and J is a block diagonal matrix:

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_m \end{bmatrix}.$$

In our case

$$J_1 = \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{bmatrix}.$$

$$v^{(0)} = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$$

where x_1 is an eigenvector corresponding to the eigenvalue λ_1, and x_2 is not an eigenvector but a so-called generalized eigenvector, which comes out from V in the Jordan decomposition of A. Let's consider the simple case

$$A = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

We get that the iterate has the form

$$v^{(k)} = \frac{A^k v^{(0)}}{\|A^k v^{(0)}\|} = \frac{c_1\lambda^k e_1 + c_2 k\lambda^k e_2 + c_2\lambda^k e_2}{\|c_1\lambda^k e_1 + c_2 k\lambda^{k-1} e_2 + c_2\lambda^k e_2\|} =$$

$$= \frac{\lambda^k}{|\lambda|^k} \frac{\left(c_1 + c_2 \frac{k}{\lambda}\right) e_1 + c_2 e_2}{\|\left(c_1 + c_2 \frac{k}{\lambda}\right) e_1 + c_2 e_2\|}$$

So for $k \to \infty$ $v^{(k)}$ converges in the eigenspace of λ, monotonically or oscillating depending on the sign of λ. The convergence as we can see is very slow, algebraic instead of geometric.
For the general case the reasoning is absolutely analogous, because what matters is the first block of the Jordan decomposition, while the others go to zero when divided by λ_1 for $k \to \infty$.

Edited by be, FI

136 SF2524BlockA.A23solution

The power method in general converges to the largest eigenvalue (in absolute value) when the absolute value of the largest eigenvalue and the absolute value of the next largest eigenvalue are different. This is the case for all α. In a),b),d) the largest eigenvalue is 110 so we will have convergence to 110. In c) the largest eigenvalue is 200, so we will have convergence to 200.

Edited by CR, EJ

137 SF2524BlockA.A3solution

Let $A \in \mathrm{C}^{n\times n}$, (λ, x) an eigenpair of A and $\mu \in \mathrm{C}$.
To question 1.)
Suppose that A is regular. Therefore every eigenvalue of A is nonzero. Since (λ, x) is an eigenpair, we have $Ax = \lambda x$. Thus

$$x = Ix = A^{-1}Ax = A^{-1}\lambda x.$$

Multiplying this equation by $\frac{1}{\lambda}$ yields $A^{-1}x = \frac{1}{\lambda}x$. This means that $(\frac{1}{\lambda}, x)$ is an eigenpair of A^{-1}.
To question 2.)

$$(A - \mu I)x = Ax - \mu x = \lambda x - \mu x = (\lambda - \mu)x$$

yields that $(\lambda - \mu, x)$ is an eigenpair of $A - \mu I$.
To question 3.)
We combine both results from above and suppose that $A - \mu I$ is regular. Thus

$$x = Ix = (A - \mu I)^{-1}(A - \mu I)x = (A - \mu I)^{-1}(\lambda - \mu)x.$$

As seen before, multiplying this equation with $\frac{1}{\lambda-\mu}$, which is possible since $\lambda - \mu$ is not equal to zero, gives $\frac{1}{\lambda-\mu}x = (A - \mu I)^{-1}x$. We have found the eigenpair $(\frac{1}{\lambda-\mu}, x)$ of $(A - \mu I)^{-1}$.

Edited by st

138 SF2524BlockA.A9solution

We have seen during lectures that the inverse iteration will in general converge to the eigenvector corresponding to the eigenvalue that is closest to μ, if there is only one which is the closest.

a) $\mu = 1.2$, the closest eigenvalue is 1
b) $\mu = 100$, the closest eigenvalue is 5

Edited by M, EJ

139 SF2524BlockA.A7solution

We can use the convergence result saying that the error of the power estimates decreases of the order $(\lambda_1/\lambda_2)^k$. The eigenvalues of the matrix are 3, 2, 1 and 0.5, and so $\lambda_1, \lambda_2 = 3, 2$. Solving $(\lambda_1/\lambda_2)^k = 10^-3$ for k gives $k = 17.0366$.

For both μ specified, the closest eigenvalue is 3. Doing the same analysis, replacing λ_1, λ_2 with $1/(\lambda_1 - \mu), 1/(\lambda_2 - \mu)$ yields $k = 17.0366$ for $\mu = 2.6$ and $k = 3.1439$ for $\mu = 2.9$.

An implementation in matlab for the power iteration with initial guess (0.1, 0.1, 0.1, 0.1) looks like the following:

```
%Power method
A=[[3,1,0,0],
    [0,2,0,1],
    [0,0,1,0.2],
    [0,0,0,0.5]]

eig(A)

vap=transpose([0.1,0.1,0.1,0.1])
i=0
while norm(vap-transpose([1,0,0,0]))> 10^-3
    i=i+1
    vap=A*vap/(norm(A*vap))
end
```

Running the script stops at i=17.

Edited by C, M, CR, EJ

140 SF2524BlockA.A17solution

It is known that Rayleigh quotient iteration has cubic convergence when the matrix is symmetric and it has in general quadratic convergence otherwise. Whereas inverse iteration has linear convergence, assuming the closest and the second closest eigenvalue to μ are distinct and not at the same distance from μ.
Therefore
1. Rayleigh quotient iteration with symmetric matrix: blue line
2. Inverse iteration: black line
3. Rayleigh quotient iteration with non-symmetric matrix: red line

Edited by M, EJ

141 SF2524BlockA.A1solution

The Rayleigh quotient iteration is a method for large eigenvalue problems. Given an initial approximation $x_0 \in \mathrm{R}^n$ new approximations $x_1, x_2, \ldots \in \mathrm{R}^n$ are computed by first solving the linear system

$$z_{i+1} = (A - r(x_i)I)^{-1} x_i$$

and subsequently normalizing the vector

$$x_{i+1} = \frac{z_{i+1}}{\|z_{i+1}\|}$$

where $r(x) := \frac{x^T A x}{x^T x}$ is the *Rayleigh quotient*.

Example in MATLAB:

```
n=100;
A=sprandn(n,n,0.1);
x=randn(n,1);
x=x/norm(x);
for k=1:10
  p=x'*A*x;
  x=(A-p*speye(n,n))\x;
  x=x/norm(x);
end
```

Edited by SJ, EJ, EJ, Anonymous

142 SF2524BlockA.A4solution

What is the result of two steps of Rayleigh quotient iteration applied to the matrix

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

with the starting vector

$$x_0 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} ?$$

Solution:

The Rayleigh quotient iteration is of the form, given x_0, then

$$x_{i+1} = \frac{(A - \mu_i I)^{-1} x_i}{\|(A - \mu_i I)^{-1} x_i\|}$$

where $\mu_i = x_i * A x_i / x_i^* x_i$.

EJ's comment: Therefore, $\mu_0 = ?$ and $x_1 = ?$

Thus, $\mu_0 = 3$ and $x_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

Edited by kj, EJ, PM

143 SF2524BlockA.A24solution

If the matrix for which we search eigenpairs A is symmetric, i.e $A^T = A$, then the convergence towards the true eigen vector x_i for the RQI updates v^k can be stated in terms $|v^{k+1} \pm x_i| = O(|v^k \pm x_i|^p)$ where $p = 3$. If the matrix is not symmetric we have the same result but for $p = 2$.

Edited by CR

144 SF2524BlockA.A5solution

Since A is symmetric, we may utilize the findings in A8 to conclude that the error in $r(v)$ is given by

$$\text{error} = |r(v) - r(x)| = \mathcal{O}(||\Delta||^2) \approx 0.2^2 = 0.04.$$

Edited by jr, Anonymous

145 SF2524BlockA.A11solution

Consider the matrix $A - \lambda_j x_j x_j^T$, by the assumptions made in Trefethen and Bau, we know that the matrix A has a complete set of orthogonal eigenvectors. If (λ_i, x_i) is another eigenpair of the matrix and $\lambda_i \neq \lambda_j$, then

$$(A - \lambda_j x_j x_j^T)x_i = \lambda_i x_i \implies Ax_i = \lambda_i x_i$$

since the two eigenvectors are perpendicular. Therefore, by applying the iteration on $A - \lambda_j x_j x_j^T$, it is possible to generate a second eigenpair of A.

Edited by Anonymous

146 SF2524BlockA.A37solution

Let $A \in \mathrm{R}^{m \times m}$ and assume that the eigenvalues of A are distinct in modulus,

$$|\lambda_1| > |\lambda_2| > \ldots |\lambda_m|.$$

The eigenvectors $x_1, \ldots, x_m$ then span R^m.

If λ_1 and x_1 are known and the Power Method is applied to the starting vector $q_0 = (A - \lambda_1 I)q$ (with q an arbitrary vector), then

$$q^{(k)} = \frac{A^k q^{(0)}}{\|A^k q^{(0)}\|} = \frac{A^k(A - \lambda_1)q}{\|A^k(A - \lambda_1)q\|} = \frac{(A^{k+1} - \lambda_1 A^k)q}{\|(A^{k+1} - \lambda_1 A^k)q\|}$$

Expressing q in terms of the eigenvector basis we get

$$q^{(k)} = \frac{(A^{k+1} - \lambda_1 A^k)(a_1 x_1 + \ldots + a_m x_m)}{\|(A^{k+1} - \lambda_1 A^k)(a_1 x_1 + \ldots + a_m x_m)\|} = \frac{(A^{k+1} - \lambda_1 A^k)(a_1 x_1 + \ldots + a_m x_m)}{\|(A^{k+1} - \lambda_1 A^k)(a_1 x_1 + \ldots + a_m x_m)\|}$$

$$= \frac{a_1\lambda_1^{k+1}x_1 + \ldots a_m\lambda_m^{k+1}x_m - (a_1\lambda_1^{k+1}x_1 + a_2\lambda_1\lambda_2^k x_2 + \ldots + a_m\lambda_1\lambda_m^k x_m)}{\|a_1\lambda_1^{k+1}x_1 + \ldots a_m\lambda_m^{k+1}x_m - (a_1\lambda_1^{k+1}x_1 + a_2\lambda_1\lambda_2^k x_2 + \ldots + a_m\lambda_1\lambda_m^k x_m)\|} =$$

$$= \frac{a_2(\lambda_2 - \lambda_1)\lambda_2^k x_2 + \ldots + a_m(\lambda_m - \lambda_1)\lambda_m^k x_m}{\|a_2(\lambda_2 - \lambda_1)\lambda_2^k x_2 + \ldots + a_m(\lambda_m - \lambda_1)\lambda_m^k x_m\|} = \frac{a_2\lambda_2^k}{|a_2|\,|\lambda_2|^k} \cdot \frac{(\lambda_2 - \lambda_1)x_2 + \frac{a_3(\lambda_1 - \lambda_3)\lambda_3^k}{a_2\lambda_2^k}x_3 \ldots + \frac{a_m(\lambda_m - \lambda_1)\lambda_m^k}{a_2\lambda_2^k}x_m}{\|(\lambda_2 - \lambda_1)x_2 + \frac{a_3(\lambda_1 - \lambda_3)\lambda_3^k}{a_2\lambda_2^k}x_3 \ldots + \frac{a_m(\lambda_m - \lambda_1)\lambda_m^k}{a_2\lambda_2^k}x_m\|}$$

As $k \to \infty$, $(|\frac{\lambda_j}{\lambda_2}|)^k \to 0$ for $j > 2$ as $|\lambda_2| > |\lambda_j|$ and thus

$$q^{(k)} \to \frac{a_2 \lambda_2^k}{|a_2|\,|\lambda_2|^k} \cdot \frac{(\lambda_2 - \lambda_1)x_2}{\|(\lambda_2 - \lambda_1)x_2\|}.$$

Edited by LL

147 SF2524BlockA.A31solution

Assuming that all assumptions made under theorem 27.1(Trefethen and Bau) hold, we have that

$$\|v^{(k)} - (\pm q_1)\| = \mathcal{O}(\left|\frac{\lambda_2}{\lambda_1}\right|^k)$$

This gives us that the error is at most a constant times $\left|\frac{\lambda_2}{\lambda_1}\right|^k$. If we let $\epsilon^{(k)}$ represent the error at kth iteration, we arrive at

$$\log \epsilon^{(k)} \propto k \cdot \log \left|\frac{\lambda_2}{\lambda_1}\right|$$

Thus, by calculating the slope of the error curve (say m), we get $|\lambda_2| = |\lambda_1| \cdot 10^m$. From the graph, the slope is approximately -0.6, and therefore, $|\lambda_2| \approx 1$.

Edited by Anonymous

148 SF2524BlockA.A47solution

a) $\frac{\partial}{\partial x} c^T x = \frac{\partial}{\partial x}(c_1 x_1 + \cdots + c_n x_n = [c_1, c_2, \cdots, c_n] = c^T$

b)

$$\frac{\partial}{\partial x} x^T x = \{\text{product rule}\} = \left(\frac{\partial}{\partial x} y^T x + \frac{\partial}{\partial x} x^T y\right)_{y=x} = \{\text{we that } (x^T y^T) = y^T x \text{ since scalar, and using} a)\} = \left(y^T + y^T\right)_{y=x} = 2x^T$$

Edited by bo

149 SF2524BlockA.A62solution

In order to minimize the two-norm, we can interpret the minimization problem as a linear least squares problem in one unknown. The solution to linear least squares problems are given by the normal equations, in our case $x = \alpha$ is the unknown and

$$Av \approx \alpha v$$

(with approximation interpreted in a least squares sense) has the normal equation

$$v^T A v = \alpha v^T v.$$

This implies that

$$\alpha = \frac{v^T A v}{v^T v},$$

which we recognize as the rayleigh quotient.

Edited by JW, EJ

150 SF2524BlockA.A61solution

Finding the eigenvalue of $A \in \mathrm{C}^{n \times n}$ is equivalent to solving $p_A(\lambda) = det(A - \lambda I) = 0$. $p_A(\lambda)$ is a polynomial of degree n, it is known that there can not be an explicit formula for findings the roots of a polynomial of degree larger than 4. Thus the methods must be iterative.

EJ's comment (beyond the scope of the course): The above description is a correct interpretation of what is written in TB. I think TB's reasoning is somewhat incomplete. Galois theory states that we cannot find an explicit formula for roots of polynomials of high degree. In the context of Galois theory "explicit formula" refers to a finite fixed number of operations involving addition, subtraction, multiplication, division, exponentiation to a (fixed) rational number. This interpretation of "explicit formula" is not usually what we mean by explicit formula in numerics, where we have a bigger set of fundamental operations, defined by the IEEE-standard of floating point operations. Moreover, Galois-theory does not imply that there is no formula which gives a solution up to machine precision. A recent class of eigenvalue methods (based on numerical quadrature) have in a certain sense this property. Howeover, I agree with TB in the sense that iterative methods are much more important for eigenvalue problems than they are for linear systems.

Edited by JW, EJ

151 SF2524BlockA.A15solution

a) This is the single Graham-Schmidt orthogonalization algorithm.
b) Q is already exactly orthogonal. With this algorithm, we add to Q the vector q_{new} which should be orthogonal to the space spanned by the columns of Q and also q_{new} should be normalized. Hence, Q_{new} should be an orthogonal matrix as well so that $Q_{\text{new}}^T Q_{\text{new}}$ should be the identity matrix and thus $\| Q_{\text{new}}^T Q_{\text{new}} - I_{3\times 3} \|$ should be zero.
c) The modified Graham-Scmidt algorithm is traditionally seen as less sensitive to rounding errors.

Edited by be

152 SF2524BlockA.A12solution

The modified Gram-Schmidt method looks almost similar to the classical Gram-Schmidt method, but the classical Gram Schmidt method is sensitive to rounding erros. We compute the orthonormalized vectors $q_1, q_2, \ldots . q_k \in \mathcal{R}^{\mathrm{n}}$ in order to get $Q_k = [q_1, \ldots . q_k]$ and orthogonalize w against the columns of Q_k and obtain h. Then we subtract $Q_k h$ from w, but due to rounding errors, the columns of Q_k are not completely orthogonal. So we subtract also the rounding errors.

In the case of Modified Gram-Schmidt we orthogonalize w against q_1 to obtain h and subtract $[q_1 * h]$ from w. Then we orthogonalize the new w against q_2 and so on, so we orthogonalize on the new version of w everytime. Than finally w is more orthogonal to Q_k than in the CGS method.

Note: the Matlab file is part of the homework, so I don't think it should be given here?

Edited by Anonymous

153 SF2524BlockA.A63solution

Let Q be m by n. Then, orthogonalizing b with the classical GS. requires m flops more than the modified G.S. Let us compare:

G.S:

```
h=Q'*b;             Step 1
z=b-Q*h;           Step 2
beta=norm(z);  Step 3
q=z/beta;           Step 4
```

Modified:

```
z=b;
for i=1:n
    h(i)=Q(:,i)'*b;    Step 1
     z=z-h(i)*Q(:,i);  Step 2
 end
 beta=norm(z);       Step 3
 z = z/beta;             Step 4
```

Step 1 requires the same number of calculations in both algorithms, so does step 3 and 4. Step 2 in G.S. requires m*n multiplications and n*m+m additions, i.e. 2*m*n + m flops. Step 2 in modified G.S. requires m multiplications and m subtractions for each loop count. This adds up to 2m*n flops. The difference is thus, n flops.

Edited by JW, Anonymous

154 SF2524BlockA.A13solution

Given a subspace $span(q_1, \ldots, q_m)$ of which $q_1, \ldots, q_m$ is an orthonormal basis. Let $Q = [q_1, \ldots, q_m] \in \mathrm{R}^{n \times m}$. To compute the Rayleigh-Ritz approximation of eigenvalues of $A \in \mathrm{R}^{n \times n}$ one computes the matrix $H = Q^T A Q \in \mathrm{R}^{m \times m}$. (This is what one in general uses the Arnoldi method for.) Then one computes the eigenvalues of H, i.e. one solves the eigenvalue problem

$$Hz = \mu z.$$

These μ are approximations of the eigenvalues of A corresponding to the subspace.

Let $Ax = \lambda x$ and assume that $x \in span(q_1, \ldots, q_m)$, i.e. it holds that $x = \sum_{i=1}^{m} z_i q_i = Qz$. Thus

$$AQz = \lambda Qz$$

As Q is orthogonal, $Q^T Q = I$ and thus

$$Q^T A Q z = \lambda z$$

Hence for $x \in span(q_1, \ldots, q_m)$ the eigenvalues of $H = Q^T A Q$ are the same as the eigenvalues of A that correspond to the subspace. For eigenvectors x with $x \approx \hat{x} \in span(q_1, \ldots, q_m)$ one gets a "good" approximation. ($\rightarrow$ as the quality of the approximation depends on how well $\hat{x} \in span(q_1, \ldots, q_m)$ approximates x it is important to choose a "good" subspace $span(Q)$.)

Edited by LL

155 SF2524BlockA.A14solution

Multiplying both sides by Q_m^T, we get

$$Q_m^T A Q_m = Q_m^T Q_{m+1} \underline{H}_m = [Q_m^T Q_m, Q_m^T q_{m+1}] \underline{H}_m .$$

Since Q_m is orthogonal and q_{m+1} is orthogonal to the $\mathrm{span}(Q_m)$,

$$[Q_m^T Q_m, Q_m^T q_{m+1}] \underline{H}_m = [I_{m \times m}, 0_{m \times 1}] \underline{H}_m = H_m.$$

Edited by be

156 SF2524BlockA.A10solution

The Arnoldi factorization is the factorization

$$AQ_m = Q_{m+1}\tilde{H}_m$$

where $QQ^* = I$ and $\tilde{H}_m$ is the $(m+1) \times m$ upper-left section of the upper Hessenberg matrix H

Edited by bo

157 SF2524BlockA.A20solution

Suppose that $A^n = 0$ for every $n > k$. Then, we have that

$$\mathcal{K}_k = \text{span}(b, Ab, \ldots, A^{k-1}b) = \mathcal{K}_{k+1} = \mathcal{K}_{k+2} = \ldots = K_n$$

for every $n > k$. Suppose that Q_k is the orthogonal matrix generated at the k-th step of the Arnoldi algorithm. Then, let $z = Aq_k$. By hypothesis, $z \in \mathcal{K}_k$ so that $z_\perp = 0$ and thus $h_{k+1,k} = \|z_\perp\| = 0$. Hence, using the *a posteriori* theorem from the lecture notes, we have that

$$\|Av - \mu v\| = |h_{k+1,k} e_k^T z| = 0$$

for each Ritz pair $H_k z = \mu z$ and $v = Q_k z$. Hence, the eigenvalues of H_k are eigenvalues of A.

Edited by be, Be

158 SF2524BlockA.A21solution

The Gram-Schmidt algorithm is numerically unstable which means that the algorithm is more sensitive to round-off errors. If the algorithm is used on a computer it will the accumulated effects of rounding errors will be worse for modified Gram-Schmidt. To improve this algorithm it needs to be modified and therefore we have the Modified Gram-Schmidt and Repeated Gram-Schmidt.

Edited by EJ, An

159 SF2524BlockA.A25solution

We use induction.

If $m = 1$, it is trivial that $\text{span}(b) = \text{span}(q_1)$.

Inductive step: suppose the Lemma holds for $m-1$, that is $\text{span}(Q_{m-1}) = \text{span}(K_{m-1}(A, b))$. We now want to prove it for m.

First notice that
$\text{span}(K_m(A, b)) = \text{span}(b, AK_{m-1}(A, b)) = \text{span}(b, A\text{span}(K_{m-1}(A, b)))$

by the inductive hypothesis we have
$\text{span}(b, A\text{span}(K_{m-1}(A, b))) = \text{span}(b, A\text{span}(Q_{m-1})) = \text{span}(b, AQ_{m-1})$

and since Q_m and H_m satisfy the Arnoldi factorization, we get

$$\text{span}(b, AQ_{m-1}) = \text{span}(b, Q_m H_m)$$

furthermore, we have
$\text{span}(b, Q_m H_m) \subseteq \text{span}(Q_m)$

So we have proven that
$\text{span}(K_m(A, b)) \subseteq \text{span}(Q_m)$

We also know that the rank of $K_m(A, b)$ is m by assumption and since Q_m is orthogonal it also has rank m. Since the two subspaces have the same dimension, the previous inclusion must be an

equality. Q.E.D.

Edited by M, EJ

160 SF2524BlockA.A28solution

For symmetric matrices, $A = A^T$, the Lanczos algorithm is in exact arithmetic equivalent to the Arnoldi method.

Consider the Arnoldi relation, $AQ_n = Q_{n+1}\underline{H}_n$, and thus $H_n = Q_n^T A Q_n$. Consider now the tranposed of H, $H^T = Q_n^T A^T Q_n$. Hence for symmetric matrics, $A = A^T$, $H^T = H$. As we know that H_n is by construction a Hessenberg matrix this means that H_n has to be tridiagonal.

In the Lanczos method this tridiagonality of H_n (which means that when applying the "full" Arnoldi method many become zero) is used to create an algorithm that requires less operations and memory than the Arnoldi method. It is obtained by "rearranging" the resulting terms and is thus equivalent in exact arithmetic.

This was not covered in the lectures but explained in the online video: https://people.kth.se/~eliasj/lanczos_method_derivation.mp4

Edited by LL, EJ

161 SF2524BlockA.A29solution

For $m = 1$ the first degree polynomial

$$p(x) = -\frac{x}{2} + \frac{3}{2}$$

has value in $p(\lambda_1) = p(1) = 1$, and value $p(\lambda_j) = p(3) = 0, j = 2, \ldots, n$, so

$$\varepsilon_1^{(1)} = 0$$

from which we conclude that at the first step the method finds the eigenvalue exactly.

Edited by FI

162 SF2524BlockA.A32solution

The power method is used to approximate the highest (in absolute sense) eigenvalue. This works in the following way: We multipy a vector by a matrix, then the contribution of the eigenvector corresponding to the largest eigenvalue increased more than other eigenvectors. So if we multiply it many times, then the contribution of the eigenvector shall dominate. Then we can approximate this eigenvector and with it the eigenvalue.

The Arnoldi method to approximate the most outlying eigenvalue (ie the one on the left) and the shift invert Arnoldi method is good in approxomating the eigenvalues close to a given shift. So if you select a good shift, then you can approximate very well the clustered eigenvalues.

Edited by LL, Anonymous

163 SF2524BlockA.A34solution

In the context of Gram-Schmidt procedures.

Not that from Lemma 1.2.2 we have that, $q_{m+1} = z/\|z\|$ and $z = b - Qh$.

Select P as,

$$\begin{bmatrix} I & h \\ 0 & \|z\| \end{bmatrix}.$$

Obviously P is nonsingular since all the diagonal entries are positive, thus all the eigenvalues are positive and hence $\det(P) \neq 0$.

Then $W = VP$ is satisfied with $V = \begin{bmatrix} Q & q_{m+1} \end{bmatrix}$ and $W = \begin{bmatrix} Q & b \end{bmatrix}$, since

$$VP = \begin{bmatrix} Q & q_{m+1} \end{bmatrix} \begin{bmatrix} I & h \\ 0 & \|z\| \end{bmatrix} = \begin{bmatrix} Q+0 & Qh + \|z\| q_{m+1} \end{bmatrix} = \begin{bmatrix} Q & Qh+z \end{bmatrix} = \begin{bmatrix} Q & b \end{bmatrix} = W.$$

Thus span(W)=span(V).

For the general case: Since P is a nonsingular linear transformation it is bijective and bijections perserve bases. I will not give the proof of this since I believe it is not in the scope of this course. Thus span(W)=span(V).

Edited by DE

164 SF2524BlockA.A35solution

The error indicator $||(I - QQ^*)x_i||$ is the norm of the difference of the eigenvector x_i and its projection on the column space of Q. It can be seen as (the sine of) the "angle" between the Krylov subspace and the eigenvector x_i in a R^n dimensions geometry.

Edited by ML, EJ

165 SF2524BlockA.A30solution

a) The Hessenberg matrix is a tridiagonal matrix

b) The algorithm can be derived from the tridiagonal structure of the hessenberg matrix. The Gram-Schmidt process can be simplified since we do not need to orthogonalize against all previous basis vectors.

c) It is designed for symmetric matrices.

Edited by EJ

166 SF2524BlockA.A38solution

The input of the Arnoldi method is a matrix $A \in R^{n \times n}$ and a vector $b \in R^{n \times 1}$ and the output is a matrix $Q_m \in R^{n \times m}$ and $Q_{m+1} = [Q_m, q_{m+1}] \in R^{n \times m+1}$. The columns of Q_{m+1} is an orthonormal basis to the Krylov subspace generated by A and b and the Arnoldi factorization holds, namely

$$AQ_m = Q_{m+1}H_m$$

where H_m is a Hessenberg matrix. The eigenvalues of A that are farthest away from the region in the complex plane where most of A:s eigenvalues are, are approximated by the eigenvalues of the Hessenberg matrix H_m.

Edited by Anonymous

167 SF2524BlockA.A40solution

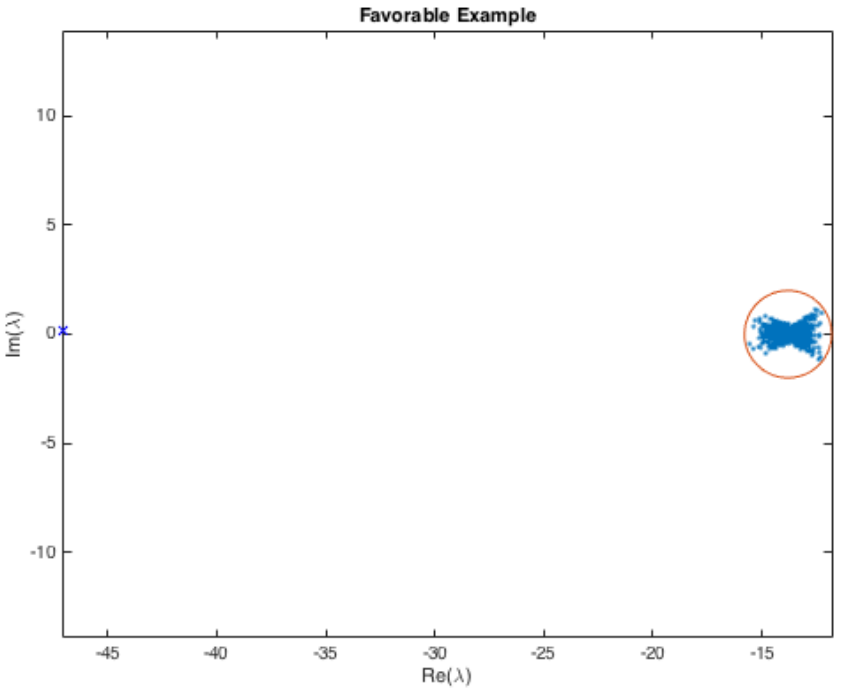
An intuitive interpretation of ε_i^m is as a measure of how "difficult" it is to push down a polynomial in points λ_j for all $j \neq i$ and maintain $p(\lambda_i) = 1$.

According to Corollary 1.4.3 in the lecture notes,

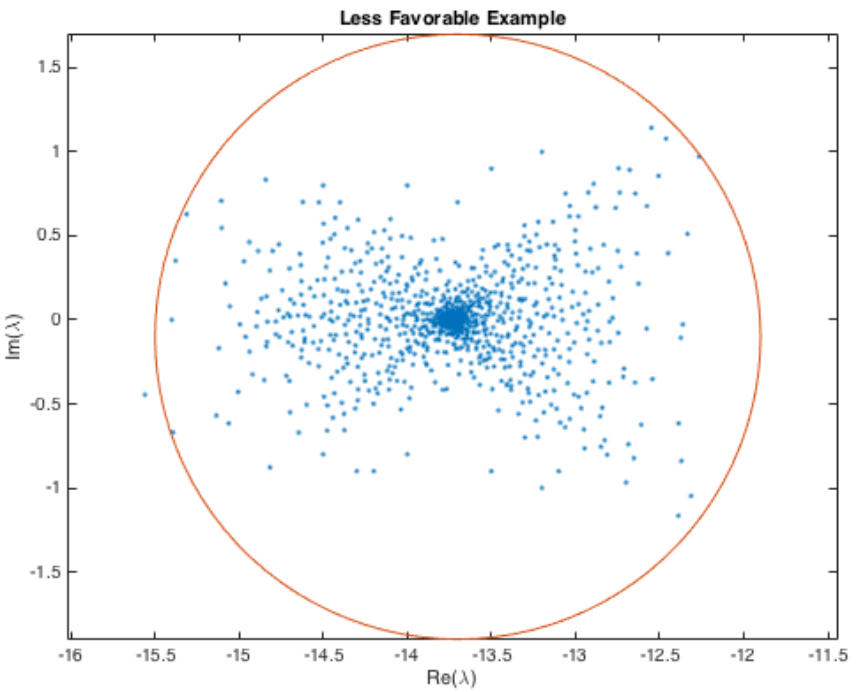
$$\varepsilon_i^m \leq \left(\frac{\rho}{|\lambda_i - c|}\right)^{m-1},$$

where ρ is the radius, λ_i the i'th eigenvalue and c the center of the disk.

Examples



$$\varepsilon_i^m \leq \left(\frac{\rho}{|\lambda_i - c|}\right)^{m-1} \approx \left(\frac{2}{33}\right)^{m-1},$$



$$\varepsilon_i^m \leq \left(\frac{\rho}{|\lambda_i - c|}\right)^{m-1} \approx \left(\frac{1.8}{1.94}\right)^{m-1}.$$

Edited by DE

168 SF2524BlockA.A41solution

Right multiply by the eigenvector x of H_k:

$$AQ_k x = Q_k H_k x = Q_k(\lambda x) = \lambda Q_k x$$

Now, introduce $v = Q_k x$, so that the above expression becomes

$$Av = \lambda v$$

That is, v is an eigenvector to A with eigenvalue λ

Edited by mb

169 SF2524BlockA.A43solution

The start vector is normalized and set as the first column in the Q matrix. That means that it does not matter if you input b or 2*b, since they are normalized to the same vector. Q and H will therefore not differ between the function calls.

Edited by ML

170 SF2524BlockA.A49solution

Let $U, V \in \mathrm{C}^{n\times m}$ be orthonormal bases of the subspace $\mathcal{K} \subseteq \mathrm{C}^n$. The solution of question A34 proves that $\exists R \in \mathrm{C}^{m\times m}$ non-singular such that $UR = V$. Thus multiplication with the adjoint of U yields $R = U^*V$ since U and V are orthonormal bases and therefore also unitary matrices, i.e. $U^*U = I_m$ and $V^*V = I_m$. We can conclude that

$$I_m = V^*V = (UR)^*UR = R^*U^*UR = R^*R.$$

Since R is square and regular, the uniqueness of the inverse in $GL_m(\mathrm{C})$ yields $R^* = R^{-1}$, which implies $RR^* = RR^{-1} = I_m$.

Edited by st, EJ

171 SF2524BlockA.A51solution

The power method favors the eigenvalue largest in modulus.

The Arnoldi method favors eigenvalues at the exterior of the spectrum.

Edited by mb

172 SF2524BlockA.A52solution

The left hand side

$$\min_{z\in\mathrm{C}^m} ||u - Qz||_2$$

is a linear least squares problem, with known solution given by the projection

$$z^* = (Q^TQ)^{-1}Q^Tu$$

which since Q is orthogonal ($Q^TQ = I$) implies $z^* = Q^Tu$. Hence,

$$\min_{z\in\mathrm{C}^m} ||u - Qz||_2 = ||(I - QQ^T)u||$$

In the Arnoldi method, Q_m is constructed so that its columns form an orthonormal basis of the Krylov subspace $\mathcal{K}_m(A, b)$. Hence, for a given $z \in \mathrm{C}^m$, it holds that Qz is a linear combination of the basis vectors of $\mathcal{K}_m(A, b)$, so that $y = Qz \in \mathcal{K}_m(A, b)$. Therefore,

$$\min_{z\in\mathrm{C}^m} ||\alpha_i x_i - Qz|| = \min_{y\in K_m(A,b)} ||\alpha_i x_i - y||$$

Edited by mb

173 SF2524BlockA.A53solution

Assume that $Ax_i = \lambda_i x_i$. Applying shift and invert Arnoldi means applying Arnoldi's method to the matrix $(A - \sigma I)^{-1}$ where σ is supposed to lie close to the eigenvalues of interest. The eigenvalues of $(A - \sigma I)^{-1}$ are known to be $\lambda((A - \sigma I)^{-1}) = \frac{1}{\lambda(A) - \sigma}$.

Assume that $(A - \sigma I)^{-1} Q_m = Q_{m+1} \underline{H}_m$ and consider the convergence indicator $\|(I - QQ^*)y_i\|$ where $(A - \sigma I)^{-1} y_i = \frac{1}{\lambda_i - \sigma} y_i$.

The first steps of the convergence proof are identical to the one in the lecture notes, with $(A - \sigma I)^{-1}$ instead of A. For the final step one gets

$$\|(I - QQ^*)\alpha_i y_i\| = \min_{p \in \mathrm{P}_{m-1}} \|\alpha_i y_i - p((A - \sigma I)^{-1}) \sum_{j=1}^{n} \alpha_j x_j\| \leq \min_{p \in \mathrm{P}_{m-1}, p(\frac{1}{\lambda_i - \sigma}) = 1} \|\alpha_i y_i - \sum_{j=1}^{n} \alpha_j p(\frac{1}{\lambda_j - \sigma}) x_j\| =$$

$$= \min_{p \in \mathrm{P}_{m-1}, p(\frac{1}{\lambda_i - \sigma}) = 1} \| \sum_{j=1, j \neq i}^{n} \alpha_j p(\frac{1}{\lambda_j - \sigma}) x_j\| \leq \left(\sum_{j=1, j \neq i}^{n} |\alpha_j| \right) \min_{p \in \mathrm{P}_{m-1}, p(\frac{1}{\lambda_i - \sigma}) = 1} \max_{j \neq i} |p(\frac{1}{\lambda_j - \sigma})|$$

Then dividing by $|\alpha_i|$ yields

$$\|(I - QQ^*)y_i\| \leq \left(\sum_{j=1, j \neq i}^{n} \frac{|\alpha_j|}{|\alpha_i|} \right) \min_{p \in \mathrm{P}_{m-1}, p(\frac{1}{\lambda_i - \sigma}) = 1} \max_{j \neq i} |p(\frac{1}{\lambda_i - \sigma})| = \xi_i \epsilon_i^{(m)}$$

Now applying the rule of thumb as in the lecture notes one finds that given a disk centered at $c \in \mathrm{C}$ that contains all eigenvalues $\mu_j = \frac{1}{\lambda_j - \sigma}$ except for μ_i, then

$$\epsilon_i^{(m)} \leq \max_{i \neq j} \frac{|\frac{1}{\lambda_j - \sigma} - c|^{(m-1)}}{|\frac{1}{\lambda_i - \sigma} - c|^{(m-1)}}$$

While we observed for the "classical" that it favors "extreme", isolated eigenvalues (which lie far away from a disk containing all other eigenvalues", with the shift and invert Arnoldi method it is possible to "select" specific eigenvalues by choosing σ such that $\frac{1}{\lambda_i - \sigma}$ is "extreme" compared to the other $\frac{1}{\lambda_j - \sigma}$ for $j \neq i$. Especially for eigenvalues which are not "extreme" to begin with, faster convergence can thus be achieved by choosing an appropriate σ.

Edited by LL

174 SF2524BlockB.B1solution

The GMRES-iterates are defined as the minimizers:

$$\min_{x \in \mathcal{K}_n(A,b)} \|Ax - b\|_2 = \|Ax_n - b\|_2$$

They are computed as follows: Suppose Q_n and $\underline{H}_n$ satisfy the Arnoldi relation and $q_1 = b/\|b\|$. Then,

$$\min_{x \in \mathcal{K}_n(A,b)} \|Ax - b\|_2 = \min_{z \in \mathrm{C}^n} \|\underline{H}_n z - \|b\| e_1\|_2,$$

where e_1 is the unit vector corresponding to the first component. The approximations x_n are obtained from solving the right hand side and then computing $x_n = Q_n z$.

Edited by EJ, An

175 SF2524BlockB.B4solution

This follows directly from the fact that the Krylov subspace at step m is included in the Krylov subspace at step $m + 1$.
Let r_m be the residual vector at step m, then

$$\|r_{m+1}\|_2 = \|Ax_{m+1} - b\|_2 = \min_{x \in \mathcal{K}_{m+1}(A,b)} \|Ax - b\|_2 \leq$$

$$\leq \min_{x \in \mathcal{K}_m(A,b)} \|Ax - b\|_2 = \|Ax_m - b\|_2 = \|r_m\|_2$$

which means the norm of the residual vector is not increasing.

Edited by M, An

176 SF2524BlockB.B5solution

The solution is sought in $\mathcal{K}_n(A, b)$. Therefore $x_n = q(A)b$ where q is polynomial of degree $n - 1$. The residual:

$$r_n = b - Ax_n = (I - Aq(A))b = p_n(A)b$$

where p_n is a polynomial of degree n for which $p_n(0) = 1$. This space is denoted P_n^0.

We now see that the relative residual is:

$$\frac{||r_n||}{||b||} \leq \inf_{p_n \in P_n^0} ||p_n(A)|| \leq \{if\ A\ diagonalizable\} \leq \inf_{p \in P_n^0} ||V||\ ||V^{-1}||\ ||p(\Lambda)|| \leq ||V||\ ||V^{-1}|| \inf_{p \in P_n^0} \max_i p(\lambda_i)$$

A way to characterise, i.e. bound, this is to find a circle $B_\rho(c)$ enclosing all the eigenvalues. Then the estimate becomes:

$$\frac{||r_n||}{||b||} \leq ||V||\ ||V^{-1}||\ (\frac{\rho}{|c|})^n$$

Edited by JW

177 SF2524BlockB.B7solution

We know that,

$$\mathcal{K}_m(A, b) = \text{span}(b, Ab, A^2b, \ldots, A^{m-1}b).$$

Therefore,

$$\mathcal{K}_m(V^{-1}AV, b) = \text{span}(b, V^{-1}AVb, (V^{-1}AV)^2b, \ldots, (V^{-1}AV)^{m-1}b)$$
$$= \text{span}(b, V^{-1}AVb, V^{-1}A^2Vb, \ldots, V^{-1}A^{m-1}Vb).$$

Similarly,

$$V^{-1}\mathcal{K}_m(A, Vb) = V^{-1}(\text{span}(Vb, AVb, A^2Vb, \ldots, A^{m-1}Vb))$$
$$= \text{span}(b, V^{-1}AVb, V^{-1}A^2Vb, \ldots, V^{-1}A^{m-1}Vb).$$

Edited by Anonymous

178 SF2524BlockB.B10solution

Since b is set to be the first column of orthogonal matrix $Q_k = (b, q_1, \ldots, q_{k-1})$, it follows
$Q_k^* b = (b^* Q_k)^* = (b^* b, b^* q_1, \ldots, b^* q_{k-1})^* = (|b|, 0, \ldots, 0)^* = (|b| e_1^*)^* = |b| e_1,$

Edited by CR

179 SF2524BlockB.B11solution

The polynomial $p(D)$ of a diagonal matrix is again a diagonal matrix D'. We want to prove that the maximum x, call it x', in the definition of the norm fulfills $x' = (0, \ldots, 1, \ldots, 0)$ where the 1 is found at position i. We can write the norm for a Diagonal matrix as $(x_1'^2 d_1^2 + \ldots + x_n'^2 d_n^2)^{1/2}$. For any x of norm 1, it holds that
$(x_1^2 d_1^2 + \ldots + x_n^2 d_n^2) \leq x_1^2 max(d_i^2) + \ldots + x_n^2 max(d_i^2) = max(d_i^2)$

since $(x_1^2 + \ldots + x_n^2) = 1$. Thus $x' = (0, \ldots, 1, \ldots, 0)$ with 1 in position i indeed.

Edited by CR

180 SF2524BlockB.B12solution

We observe that the eigenvalues of A_1 fit inside a circle in the complex plane centered at $c(A_1) = 5$ and with radius $r(A_1) = 1$. Similarly, the eigenvalues of A_2 fit inside a circle in the complex plane centered at $c(A_2) = 8$ with radius $r(A_2) = r(A_1) = 1$.

We consider the polynomial

$$q(z) := \frac{(c - z)^n}{c^n},$$

known as *the Zarantonello polynomial*. It is the minimizing polynomial over a disk in the sense that

$$\min_{p \in P_n^0} \max_{z \in \bar{C}(c,r)} |p(z)| = \max_{z \in \bar{C}(c,r)} |q(z)| = \left(\frac{r}{|c|} \right)^n,$$

where $\bar{C}(c, r)$ is the disc with center at c and radius r, $P_n = \{\text{polynomials of degree at most } n\}$ and $P_n^0 = \{p \in P_n : p(0) = 1\}$. Since $q \in P_n^0$, we have

$$\min_{p \in P_n^0} \left(\max_i \left(|p(\lambda_i(A_k))| \right) \right) \leq \max_i |q(\lambda_i(A_k))| = \max_i \frac{|c(A_k) - \lambda_i(A_k)|^n}{|c(A_k)|^n} \leq \frac{r(A_k)^n}{|c(A_k)|^n},$$

where $k = 1, 2$. It follows that

$$\min_{p \in P_n^0} \left(\max_i \left(|p(\lambda_i(A_1))| \right) \right) \leq \frac{1^n}{5^n} = \frac{1}{5^n}$$

and

$$\min_{p \in P_n^0} \left(\max_i \left(|p(\lambda_i(A_2))| \right) \right) \leq \frac{1^n}{8^n} = \frac{1}{8^n}$$

which are the sought bounds. Since all eigenvalues are strictly in the right half-plane for both matrices, they are invertible and hence diagonalizable and the bounds may be utilized together with the main convergence theorem of GMRES to give a bound for the relative residual error.

Edited by jr

181 SF2524BlockB.B13solution

Since A is symmetric positive definite,

$$A = U\Lambda U^*$$

for some U orthogonal and Λ diagonal with real positive elements on its diagonal.

Hence, the general bound

$$\frac{\|r_n\|}{\|b\|} \leq \kappa(U) \min_{p_n \in P_n^0} \max_{\lambda \in \mathrm{spec}(A)} |p_n(\lambda)|$$

simplifies as

$$\frac{\|r_n\|}{\|b\|} \leq \min_{p_n \in P_n^0} \max_{\lambda \in \mathrm{spec}(A)} p_n(\lambda)$$

since $\kappa(U) = 1$.

Following the same reasoning as for the CG convergence, we can also find

$$\frac{\|r_n\|}{\|b\|} \leq 2\left(\frac{\sqrt{\kappa(A)}+1}{\sqrt{\kappa(A)}-1}\right)^n.$$

Edited by Be, LL

182 SF2524BlockB.B14solution

Let m be the size of A. Hence, A has at most m eigenvalues. We denote $\{\lambda_i\}_{i=1}^m$ the eigenvalues of A. Since A is assumed invertible, the λ_i's must be non-zero.

It is always possible to find a polynomial p_m of degree m such that $p_m(\lambda_i) = 0$ for every $1 \leq i \leq m$ and $p_m(0) = 1$. Hence, the algorithm will terminate after at most m steps.

Edited by Be

183 SF2524BlockB.B15solution

The approximation x_n of the solution of $Ax = b$ after n steps is defined as minimizer of the residual norm, i.e. x_n such that

$$\|Ax_n - b\|_2 = \min_{x \in \mathcal{K}_n(A,b)} \|Ax - b\|_2$$

similarly, the approximation z_n of the solution of $Bz = b$ is such that

$$\|Bz_n - b\|_2 = \min_{z \in \mathcal{K}_n(A,b)} \|Bz - b\|_2$$

If $A = -B$ we have that

$$\|Ax_n - b\|_2 = \min_{x \in \mathcal{K}_n(A,b)} \|Ax - b\|_2 = \min_{x \in \mathcal{K}_n(A,b)} \|B(-x) - b\|_2 =$$

$$\min_{-x \in \mathcal{K}_n(A,b)} \|B(-x) - b\|_2 = \min_{z \in \mathcal{K}_n(A,b)} \|Bz - b\|_2 = \|Bz_n - b\|_2$$

which implies $x_n = -z_n$.
Furthermore, we have the same relationship between the exact solution x^* of $Ax = b$ and the exact solution z^* of $Bz = b$, that is $x^* = -z^*$.
Therefore, the approximation error is the same for both problems:

$$\|x_n - x^*\| = \|-(x_n - x^*)\| = \|z_n - z^*\|$$

Edited by M, An

184 SF2524BlockB.B16solution

Consider the min-max bound for GMRES,

$$\frac{\|Ax_n - b\|}{\|b\|} \leq \|V\|\|V^{-1}\| \min_{p \in P_n^0} \max_{i=1,\ldots,m} |p(\lambda_i)|$$

Assume that A has only 10 different eigenvalues. After 10 iterations $p \in P_1 0^0$, i.e. it has 10 roots and thus $\min_{p \in P_n^0} \max_{i=1,\ldots,m} |p(\lambda_i)| = 0$ as one can "put" a root at every λ_i, specifically $p(z) = \Pi_{i=1}^{10}(1 - \frac{x}{\lambda_i})$.

Then $\|Ax_1 0 - b\| \leq 0$ and hence $Ax_1 0 = b$, the exact solution is returned after 10 iterations.

Edited by LL

185 SF2524BlockB.B18solution

The computation time per iteration increases with iterations in GMRES since the number of vectors in Q increases, leading to a more and more computationally heavy Gram-Schmidt orthogonalization. This means that running many iterations of GMRES can be infeasible.

Edited by ML

186 SF2524BlockB.B46solution

It holds that

$$\mathcal{K}(A, b) = \{\alpha_0 b + \alpha_1 Ab + \ldots \alpha_n A^{n-1} b : \alpha_0, \alpha_1, \ldots, \alpha_n \in \mathrm{C}\} = \{q(A)b : q \in P_{n-1}\}$$

where P_{n-1} is the set of polynomials of degree less than or equal to $n - 1$. Hence,

$$\{b - Ax : x \in \mathcal{K}(A, b)\} = \{b - Aq(A)b : q \in P_{n-1}\}$$

Now, note that polynomials of the form $p(z) = 1 - zq(z)$ for $q(z) \in P_{n-1}$ have the properties

$$p(0) = 1$$

and

$$p(z) \in P_n$$

i.e.

$$p(z) \in P_n^0$$

It now holds that $b - Aq(A)b = p(A)b$ for some $p \in P_n^0$. In conclusion,

$$x \in \mathcal{K}(A, b) \Leftrightarrow b - Ax = p(A)b$$

for some $p \in P_n^0$

Edited by mb

187 SF2524BlockB.B37solution

Set equality between two sets A and B means that each point in A is found in B, and each point in B is found in A. In the context of the exercise, a point corresponds to to a residual vector. The equality of the sets describe that there are two ways of representing residual vectors, one indexed by vectors x of a Krylov space, the other indexed by polynomials of a certain type.

Edited by CR

188 SF2524BlockB.B64solution

To a)
To use Corollary 2.1.5 (Single localization disk) we draw a disk $D(c = 3 + i, \rho = 2)$ which contains all eigenvalues of A. Thus we obtain a bound

$$\min_{p \in P_n^0 i=1, \dots, m} \max |p(\lambda_i)| \leq \frac{\rho^n}{|c|^n} = \left(\frac{2}{\sqrt{10}}\right)^n.$$

To b)
This also gives a convergence bound for GMRES, since we can estimate the residual norm in each iteration step n with

$$\frac{\|Ax_n - b\|}{\|b\|} \leq \|V\| \|V^{-1}\| \left(\frac{2}{\sqrt{10}}\right)^n.$$

Edited by st

189 SF2524BlockB.B65solution

$$||Qx||_2 = \sqrt{(Qx)^T(Qx)} = \sqrt{x^T Q^T Q x} = \sqrt{x^T I_n x} = \sqrt{x^T x} = ||x||_2$$

Edited by JW

190 SF2524BlockB.B2solution

Both the GMRES and the CG method generate approximate solutions x_n to the linear system

$Ax = b$

after n iterations of the methods. Both the solutions $x_n^{(CG)}$ and $x_n^{(GMRES)}$ are calculated to minimize the residual

$r = Ax_n - b$

but since the residual is measured in different norms for the two different methods, the approximate solutions vectors x_n^{CG} and x_n^{GMRES} wont be equal to each other. They will however lie in the same Krylov subspace $\mathcal{K}_n(A, b)$.

Edited by Anonymous

191 SF2524BlockB.B3solution

The Conjugate Gradient (CG) iterates for a matrix A are defined as the minimizers of $\|Ax - b\|_{A^{-1}}$, over the nth Krylov Subspace. That is, the CG iterates $x_1, x_2, \dots x_n$ satisfy

$$\min_{x \in \mathcal{K}_n(A,b)} \|Ax - b\|_{A^{-1}} = \|Ax_n - b\|_{A^{-1}}, \quad n = 1, 2, \dots$$

Equivalently,

$$\min_{x \in \mathcal{K}_n(A,b)} \|x - x_*\|_A = \|x_n - x_*\|_A, \quad n = 1, 2, \dots$$

where $x_* = A^{-1}b$.

Edited by EJ, An

192 SF2524BlockB.B44solution

If A is symmetric and positive definite then the convergence is monotonic in the error measured in

the $\|\cdot\|_A$-norm. First we note that x_n is the minimiser in $\mathcal{K}_n$, consider $\mathcal{K}_n \ni x = x_n + \Delta x$ which implies that $e = x_* - x = e_n + \Delta x$

$$||e_n||_A^2 = (e_n + \Delta x)^T A(e_n + \Delta x) = e_n^T A e_n + \Delta x^T A \Delta x + 2e_n^T A \Delta x$$

$2e_n^T A\Delta x = 2r_n^T \Delta x = 0$. (Why?, see thm. 38.1)

Since A is positive definite $\Delta x \cdot (A\Delta x)$ is positive. Thus x_n is the a minimiser in $\mathcal{K}_n$. Now since $\mathcal{K}_n \subset \mathcal{K}_{n+1}$ it follows that $||e_n||_A \geq ||e_{n+1}||_A$. (if $\mathcal{K}_n = \mathcal{K}_{n+1}$ then $||e_n||_A = ||e_{n+1}||_A$

Edited by JW, EJ

193 SF2524BlockB.B21solution

As in CG, assume that A is symmetric positive definite and hence invertible. The residual norm with respect to A^{-1} squared is in CG defined as,

$$\|Ax - b\|_{A^{-1}}^2 = (Ax-b)^T A^{-1}(Ax-b) = (Ax-b)^T(x - A^{-1}b) = \left[A^{-1}b = x^*\right] = \left(A(x - A^{-1}b\right)^T(x-x^*) = (x-x^*)^T A(x-x^*) = \|x - x^*\|_A^2.$$

By taking the square root,

$$\|Ax - b\|_{A^{-1}} = \|x - x^*\|_A.$$

Thus the residual norm with respect to A^{-1} equals the error norm with respect to A.

Edited by Anonymous

194 SF2524BlockB.B22solution

Theorem 38.5 in T&B states that if we have a symmetric positive definite matrix problem $Ax = b$, where A has 2-norm condition number κ. Then the A-norms of the errors satisfy

$$\frac{||e_n||_A}{||e_0||_A} \leq 2\left[\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^n + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^{-n}\right]^{-1} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n.$$

By Taylor expansion, we have

$$\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} = 1 - \frac{2}{\sqrt{\kappa}} + \mathcal{O}\left(\frac{1}{\kappa}\right).$$

Combining these two results, we get

$$\frac{||e_n||_A}{||e_0||_A} \leq 2\left(1 - \frac{2}{\sqrt{\kappa}} + \mathcal{O}\left(\frac{1}{\kappa}\right)\right)^n.$$

We want to determine the numbers of iterations n required to reach a specific accuracy ε. For large κ, we may neglect the remainder in the Taylor expansion and we obtain

$$2\left(1 - \frac{2}{\sqrt{\kappa}}\right)^n \approx \varepsilon \Rightarrow n \approx \frac{\ln(\varepsilon/2)}{\ln\left(1 - \frac{2}{\sqrt{\kappa}}\right)} \approx \frac{\sqrt{\kappa}}{2}|\ln(\varepsilon/2)|$$

where in the last approximate equality, we have once again made a Taylor expansion, which is valid for large κ.

Edited by LL, jr

195 SF2524BlockB.B23solution

The algorithm for CG includes the following steps:

(1) $\alpha_n = \frac{r_{n-1}^T r_{n-1}}{p_{n-1}^T A p_{n-1}}$

(2) $x_n = x_{n-1} + \alpha_n p_{n-1}$

(3) $r_n = r_{n-1} - \alpha A p_{n-1}$

(4) $\beta_n = \frac{r_n^T r_n}{r_{n-1}^T r_{n-1}}$

(5) $p = r_n + \beta_n p_{n-1}$

Now, lets define transformation matrices with dimension nxn:

$$T := \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \\ & & & & 1 \end{bmatrix}$$

$$B := \begin{bmatrix} 1 & -\beta_1 & & & \\ & 1 & -\beta_2 & & \\ & & \ddots & \ddots & \\ & & & 1 & -\beta_{n-1} \\ & & & & 1 \end{bmatrix}$$

$$D := \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{bmatrix}$$

With these matrices it is possible to express the three expressions between X, R and P.

From (2) it is possible to write:

$$x_n = x_{n-1} + \alpha_n p_{n-1}$$

$$\iff x_n - x_{n-1} = \alpha_n p_{n-1}$$

It is here possible to write that, for $i = 1, . . , n$, $x_n - x_{n-1} = XT$ and for $\alpha_n p_{n-1} = PD$ which gives

$$x_n - x_{n-1} = \alpha_n p_{n-1}$$

$$\iff XT = PD$$

That is the first expression.
From (5) it is possible to write:

$$p_n = r_n + \beta_n p_{n-1}$$

$$\iff p_n - \beta_n p_{n-1} = r_n$$

It is here possible to write that, for $i = 1, . . , n$, $p_n - \beta_n p_{n-1} = PB$ and it is known that $\beta_n = B$ which gives

$$p_n - \beta_n p_{n-1} = r_n$$

$$\iff PB = R$$

This is the second expression.
From (3) it is possible to write:

$$r_n = r_{n-1} - \alpha A p_{n-1}$$

$$\iff r_n - r_{n-1} = \alpha A p_{n-1}$$

It is here possible to write that, for $i = 1, . . , n$, $\alpha A p_{n-1} = APD$ and it is possible to express $r_n - r_{n-1} = RT^T - r_n e_n^T$ which gives

$$r_n - r_{n-1} = \alpha A p_{n-1}$$

$$\iff APD = RT^T - r_n e_n^T$$

This is the third expression.

Edited by An

196 SF2524BlockB.B19solution

In CG, x_n is defined as the minimizer to

$$\min_{x \in \mathcal{K}_n(A,b)} \|Ax - b\|_{A^{-1}} = \|Ax_n - b\|_{A^{-1}}, \qquad n = 1, 2, \ldots,$$

and thus r_n satisfy the minimization property that $r_n^T Q = 0$, where $r_n = b - Ax_n$ and Q is a matrix such that $span(Q) = \mathcal{K}_n(A, b)$. Since $span(Q) = \mathcal{K}_n(A, b) = span(r_0, \ldots, r_{n-1})$ it also holds that $r_n^T r_i = 0$ for $i < n$.

For the second part. Note that in CG, $r_{n+1} = r_n - \alpha_{n+1} A p_n$ and $r_n^T p_i = 0$ for $i < n$ since all old search directions lie in $\mathcal{K}_n(A, b)$, i.e. $span(Q) = \mathcal{K}_n(A, b) = span(p_0, \ldots, p_{n-1})$.

By premultiplying the residual equation by $p_i^T,\ i < n$

$$p_i^T r_{n+1} = p_i^T r_n - \alpha_{n+1} p_i^T A p_n, \quad i < n, 0 = 0 - \alpha_{n+1} p_i^T A p_n, \quad i < n, 0 = p_i^T A p_n, \quad i < n.$$

Thus the search directions $p_0, \ldots, p_n$ are A-orthogonal.

Edited by EJ, An

197 SF2524BlockB.B24solution

Short-term recurrence methods only need to save a few vectors instead of big sequences of vectors. The CG method is a short-term method since only three vectors are stored at each iteration, in contrast to for example GMRES where the matrix Q_n increases in every iteration. The GMRES is therefore not a short-term recurrence method.

The short-term recurrence methods are preferable since they require less computer memory.

Edited by Anonymous

198 SF2524BlockB.B25solution

a) Because the norm has to be a positive value (thus $z^T A z$ has to be positive), and this is not generally the case if A is not SPD.

b) Since the inverse of a positive definite matrix is positive definite.

Edited by CR, BO

199 SF2524BlockB.B26solution

Let $A \in \mathrm{C}^{n \times n}$ be normal. Therefore, the spectral theorem gives $\sigma_i = |\lambda_i| \ \forall i = 1, \ldots, n$, where $\sigma_1 \geq \ldots \geq \sigma_n$ are the singular values of A and $|\lambda_1| \geq \ldots \geq |\lambda_n|$ are its eigenvalues. We conclude

$$\kappa(A) = ||A|| \cdot ||A^{-1}|| = \sigma_{\max}(A)\sigma_{\max}(A^{-1}) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} = \frac{|\lambda_{\max}(A)|}{|\lambda_{\min}(A)|}.$$

Edited by st

200 SF2524BlockB.B27solution

$A^T = A$, then

$$\sigma(A) = \sqrt{\lambda(A^T A)} = \sqrt{\lambda(A^2)} = \sqrt{\lambda(A)^2} = |\lambda(A)|$$

where we have used $\lambda(A^2) = \lambda(A)^2$. This can be proved by considering that A is diagonalizable, $A = V^{-1}DV$ with D having the same eigenvalues of A. Therefore $A^2 = V^{-1}D^2V$ has the same eigenvalues of D^2 which are the squared eigenvalues of A.

Edited by M

201 SF2524BlockB.B28solution

In exact arithmetic we expect zero. We know from theory of CG and GMRES that

- For the Arnoldi method we know that the column span of the Q-matrix is the Krylov subspace $\mathcal{K}_N(A, b)$.
- For CG we know that the column span of $x_1, \ldots, x_N$ is the Krylov subspace $\mathcal{K}_N(A, b)$ (See theorem about CG in lecture notes).

We therefore know that there exists a matrix $Z \in \mathrm{R}^{N \times N}$ such that

$$Q = XZ.$$

That is

$$QZ^{-1} = X.$$

By multiplication from the left with Q^T and using that it is an orthogonal matrix we have

$$Q^T X = Q^T Q Z^{-1} = Z^{-1}$$

which is computed in the first formula in the source code

```
Z=inv(Q'*X).
```

The final formula

```
norm(Q(:,1:N)-X*Z)
```

confirms the result.

Edited by EJ

202 SF2524BlockB.B43solution

As mentioned as a hint, the solution can be found with $Hx + c$. The following MATLAB code solves this equation with CG (H is SPD):

```
H = [5 2 3 2;
2 3 1 1;
3 1 4 -1;
2 1 -1 7];
c = [-2 8 -2 4]'

x = cg(H, -c, -c)
norm(H*x+c)
```

The norm has to be zero, it was 2.0515e-14 for me.

```
x =
   3.0000
  -4.0000
  -1.0000
  -1.0000
```

Edited by BO

203 SF2524BlockB.B47solution

EJ's comment: The (nice) solution below is based on the CG-algorithm, another solution is based on the definition of the CG-iterates as minimizers

Use that $x_{n+1} - x_n = \alpha_{n+1} p_n$, i.e. $x_n = x_{n+1} - \alpha_{n+1} p_n$:

$$\|e_n\|_A = \|x_* - x_n\|_A = \|x_* - x_{n+1} + \alpha_{n+1} p_n\|_A = \|e_{n+1} + \alpha_{n+1} p_n\|_A =$$

$$(e_{n+1} + \alpha_{n+1} p_n)^T A (e_{n+1} + \alpha_{n+1} p_n) = e_{n+1}^T A e_{n+1} + 2 e_{n+1}^T A \alpha_{n+1} p_n + (\alpha_{n+1} p_n)^T A \alpha_{n+1} p_n =$$

$$\|e_{n+1}\|_A + 2 e_{n+1}^T A \alpha_{n+1} p_n + \|\alpha_{n+1} p_n\|_A$$

Now consider $2\alpha_{n+1} e_{n+1}^T A p_n$:

$$e_{n+1}^T A p_n = (x_* - x_{n+1})^T A p_n = r_{n+1}^T p_n$$

Use that

$$r_{n+1} = b - A x_{n+1} = b - A(x_n + \alpha_{n+1} p_n) = r_n - \alpha_{n+1} A p_n,$$

thus

$$r_{n+1}^T p_n = (r_n - \alpha_{n+1} A p_n)^T p_n = r_n^T p_n - \alpha_{n+1} p_n^T A p = r_n^T p_n - r_n^T r_n$$

using that $\alpha_{n+1} = \frac{r_n^T r_n}{p_n^T A p_n}$. Now use that $p_n = r_n + \beta_n p_{n-1}$, hence

$$r_n^T p_n - r_n^T r_n = r_n^T (r_n + \beta_n p_{n-1}) - r_n^T r_n = \beta_n r_n^T p_{n-1}.$$

Summarizing,

$$r_{n+1}^T p_n = \beta_n r_n^T p_{n-1}$$

Using this recursively we obtain that

$$r_{n+1}^T p_n = \beta_n \cdot \ldots \beta_1 r_1^T p_0$$

Using the initial conditions

$$r_1^T p_0 = r_1^T r_0 = (r_0 - \alpha_1 A r_0)^T r_0 = r_0^T r_0 - \frac{r_0^T r_0}{r_0^T A r_0} r_0^T A r_0 = 0$$

Hence,

$$r_{n+1}^T p_n = 0$$

and thus

$$\|e_n\|_A = \|e_{n+1}\|_A + \|\alpha_{n+1} p_n\|_A$$

Edited by LL, EJ

204 SF2524BlockB.B48solution

The matrices are defined a

$$R := [r_0, \ldots, r_n - 1] \quad P := [p_0, \ldots, p_{n-1}]$$

The p_i vectors denote the correction direction at each iteration,

$$x_i - x_{i-1} = \alpha_i p_{i-1}$$

where α_i is a scaling factor (the "step size").

The vectors r_i denote the residual at each iteration,

$$r_i := b - A x_i = b - A(x_{i-1} + \alpha_i p_{i-1}) = r_{i-1} - \alpha_i A p_{i-1}$$

Once the residual vector for iteration i is obtained it can be used to update

$$p_i = \beta_i p_{i-1} + r_i .$$

From this follows the relation between the matrices P and R, as $p_i - \beta_i p_{i-1} = r_i$,

$$P \begin{bmatrix} 1 & -\beta_i & & \\ & \ddots & \ddots & \\ & & \ddots & -\beta_{n-1} \\ & & & 1 \end{bmatrix} = R$$

As stated in Lemma 2.2.4 in the lecture notes, P and B span the same subspace, specifically $\mathcal{K}_n(A, b)$, if α_i, β_i for $i = i, \ldots, n$ are non-zero.

Edited by LL

205 SF2524BlockB.B59solution

a) Different Fomulation of the CG algorithm

b) Comparison will be done to the CG formulation in algorithm 2

v and v_+ are used for storing the products $r_0^T \cdot r_0$ and $r_{n-1}^T \cdot r_{n-1}$ to avoid double calculation of $r^T \cdot r$

q stores the result of the matrix-vector-multiplication $A \cdot p_{n-1}$ to calculate this only once per iteration step

α, x, r, β respresents the same as in the lecture notes

k is a loop counter

Edited by CN

206 SF2524BlockB.B62solution

Solution idea: Use that the min-max bound of CG implies that it should require at most as many iterations as the number of (different) eigenvalues. In this case we have two different eigenvalues, so at most two iterations.

Edited by EJ

207 SF2524BlockB.B30solution

CG stands for Conjugate Gradient while CGN stands for CG Applied to the Normal Equation. The relationship between these two iterative methods is that both are using the same algorithm to solve a problem, but applied to different matrices. CGN is CG applied to $A^TAx = A^Tb$.

Edited by An

208 SF2524BlockB.B32solution

The bound for CGN is given by

$$\frac{||Ax-b||_2}{||b||_2} \leq 2\left(\frac{\kappa-1}{\kappa+1}\right)^n$$

which could be compared to the analogous bound for CG

$$\frac{||Ax-b||_{A^{-1}}}{||b||_{A^{-1}}} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n$$

Apart from the square-root of κ, the bounds are also measured in different norms. In brief, CGN is generally worse at the iterations grows as $\mathcal{O}(\kappa)$, in contrast to $\mathcal{O}(\sqrt{\kappa})$ for CG.

Edited by mb

209 SF2524BlockB.B31solution

From lecture 8:

From definition we have that

$$\min_{x \in \mathcal{K}_n(B,c)} ||Bx-c||_{B^{-1}} = ||Bx_n - c||_{B^{-1}}$$

Reformulating the r.h.s of the above equation gives

$$||Bx-c||^2_{B^{-1}} = ||x-x^*||_B = (x-x^*)^T B(x-x^*)^T = (x-x^*)^T A^T A(x-x^*)^T = (Ax-b)(Ax-b)^T = ||Ax-b||_2^2$$

Such that the expression may be expanded as

$$\min_{x \in \mathcal{K}_n(B,c)} ||Bx-c||_{B^{-1}} = ||Bx_n - c||_{B^{-1}} = \min_{x \in \mathcal{K}_n(B,c)} ||Ax-b||_2$$

Thus, the iterates are minimized over $\mathcal{K}_n(B,c)$ instead of $\mathcal{K}_n(A,b)$ as for GMRES

Edited by bo

210 SF2524BlockB.B33solution

The convergence of CGN depends on the singular values of the matrix, while the convergence of GMRES depends on the eigenvalues of the matrix. Hence, in a situation where the eigenvalues are more ill-conditioned than the singular values, CGN is favored over GMRES. For example, a

well-conditioned matrix (i.e. the singular values are well-behaved) with eigenvalues close to the origin would favor the CGN method.

Edited by mb

211 SF2524BlockB.B39solution

a) Applying a left preconditioner means we multiply from the left so the linear system becomes

$$Ax = b$$

$$M^{-1}Ax = M^{-1}b$$

$$LL^TAx = LL^Tb$$

Moreover, L is invertible and (by problem formulation) $x = Ly$ such that we have

$$L^TALy = L^Tb$$

which proves the statement.
b) Since A is symmetric positive definite we have by definition that

$$z^TAz > 0$$

for any $z \neq 0$. If we set $z = Lw$, we have

$$w^TL^TALw = (Lw)^TALw = z^TAz > 0$$

where we used that A is positive definite in the last step. Hence L^TAL is symmetric positive definite.
c) The recurrence relation for the transformed problem is

$$\alpha_n = \frac{r_{n-1}^T r_{n-1}}{p_{n-1}^T L^TALp_{n-1}}$$

$$x_n = x_{n-1} + \alpha_n p_{n-1}$$

$$r_n = r_{n-1} - \alpha_n L^TALp_{n-1}$$

$$\beta_n = \frac{r_n^T r_n}{r_{n-1}^T r_{n-1}}$$

We define (slightly different from suggested in exercise)

$$\hat{p}_n = Lp_n$$

$$\hat{x}_n = Lx_n$$

$$\hat{r}_n = L^{-T}r_n.$$

Moreover, we can simplify the formulas if we introduce the additional vector

$$z_n = M^{-1}\hat{r}_n$$

The recursion formulas are changed as follows:

$$\alpha_n = \frac{\hat{r}_{n-1}^T LL^T\hat{r}_{n-1}}{\hat{p}_{n-1}A\hat{p}_{n-1}} = \frac{\hat{r}_{n-1}^T M^{-1}\hat{r}_{n-1}}{\hat{p}_{n-1}A\hat{p}_{n-1}} = \frac{\hat{r}_n^T z_n}{\hat{p}_{n-1}A\hat{p}_{n-1}}$$

$$\hat{r}_n = L^{-T}r_n = L^{-T}r_{n-1} + \alpha_n L^{-T}L^TALp_{n-1} = \hat{r}_{n-1} - \alpha_n A\hat{p}_n$$

$$\hat{x}_n = Lx_n = \hat{x}_{n-1} + \alpha\hat{p}_{n-1}$$

$$\hat{p}_n = Lp_n = Lr_n + \beta_n Lp_{n-1} = LL^T\hat{r}_n + \beta_n\hat{p}_{n-1} = M^{-1}\hat{r}_n + \beta_n\hat{p}_{n-1} = z_n + \beta_n\hat{p}_{n-1}.$$

These formulas form an algorithm which is identical to what is called preconditioned conjugate gradient method on wikipedia.

Edited by EJ

212 SF2524BlockB.B57solution

A) ♣: $Ax - b$ ♠: $x \in \mathcal{K}_n(A, b)$ ◇: 2 ♡: 2

B) ♣: $Ax - b$ ♠: $x \in \mathcal{K}_n(A, b)$ ◇: A^{-1} ♡: A^{-1}

C) Two alternative solutions with $B = A^T A$ and $c = A^T b$

Alt 1.: ♣: $Bx - c$ ♠: $x \in \mathcal{K}_n(B, c)$ ◇: B^{-1} ♡: B^{-1}

Alt 2.: ♣: $Ax - b$ ♠: $x \in \mathcal{K}_n(B, c)$ ◇: 2 ♡: 2

Edited by bo, EJ

213 SF2524BlockB.B55solution

As I is SPD this **is** a norm and it is defined as $||x||_I = \sqrt{x^T \cdot I \cdot x}$

but as

$x^T \cdot I = x^T$

we find

$||x||_I = \sqrt{x^T \cdot I \cdot x} = \sqrt{x^T x}$

which is the definition of the euclidean norm

$||x||_2 = \sqrt{x^T x}$

Edited by CN

214 SF2524BlockB.B58solution

a) To carry out the proof for v_{k+1} we do induction and use steps 2,5 and 8 and 10 and the initialization in step 1: Initialization step we set $v_1 = (b - Ax_0)/\|b - Ax_0\|$. Use induction hypothesis, $v_k \in \mathcal{K}_k(A, v_1)$. From step 2 we find that $\tilde{v}_{k+1} \in \mathcal{K}_{k+1}(A, v_1)$. In the operations 5,8,10 we let v_{k+1} be linear combination of $\tilde{v}_{k+1}$, v_k such that $v_{k+1} \in \mathcal{K}_{k+1}(A, v_1)$. The proof is analogous but with a transpose $w_{k+1} \in \mathcal{K}_{k+1}(A^T, v_1)$.

b) If A is symmetric $w_k = v_k$ for all k and the algorithm reduces to Lanczos.

c) $V_k = [v_1, \ldots, v_k]$, $W_k = [w_1, \ldots, w_k]$. The matrix $\underset{-}{T}_k$ is given by

$$\underset{-}{T}_k = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ \gamma_1 & \ddots & \ddots & \\ & \ddots & \ddots & \beta_{k-1} \\ & & \ddots & \alpha_k \\ & & & \gamma_k \end{bmatrix}$$

Edited by EJ

215 SF2524BlockC.C1solution

Suppose that A is a matrix with dimension mxm. Then it is possible to express A as $A = QR$ where

R is an upper triangular matrix such that $R \in C^{mxm}$ and Q is an unitary matrix such that $Q \in C^{mxm}$.

The basic QR-algorithm is given by:

Let $A_0 = A$ and iterate:

- Compute the QR-factorization of $A_k = QR$.
- Set $A_{k+1} = RQ$.

This is done until the matrix A is an upper triangular matrix.

Observe that k is the iteration count

Edited by EJ, An

216 SF2524BlockC.Ca-11solution

Equality is shown by separation into two parts. Assuming that (λ, v) is an eigenpair A and showing a corresponding eigenpair of B, and then showing the converse.

- 1) Suppose (λ, v) is an eigenpair of A. We have

$$\lambda v = Av$$

If we define w as $w = V^{-1}v$ we have

$$\lambda V^{-1}w = AVw$$

By multiplication of V^{-1} from the left we obtain

$$\lambda w = V^{-1}AVw = Bw$$

which shows that (λ, w) is an eigenpair of B.

- 2) Suppose (λ, w) is an eigenpair of B. We have

$$\lambda w = Bw$$

If we define $v = Vw$ we have

$$\lambda Vv = BV^{-1}v.$$

By multiplication of V from the left we obtain

$$\lambda v = VBV^{-1}v = Av$$

which shows that (λ, v) is an eigenpair of A.

Edited by EJ

217 SF2524BlockC.C2solution

The reason why the Schur form is used instead of the Jordan form is because the Schur form is numerically stable while the Jordan form is often not numerically stable.

Edited by Anonymous

218 SF2524BlockC.C3solution

The iterates of the QR method are matrices $A_k = R_kQ_k$, where $Q_k^*Q_k = Q_kQ_k^* = I$. We shall show that all A_k have the same eigenvalues.

Proof alternative 1. Note that $A_{k+1} = Q_k^TA_kQ_k$ for an orthogonal matrix Q_k. This is a similarity transformation and similarity transformations do not change the eigenvalues (as we proved in

SF2524BlockC.Ca-11).

Proof alternative 2. Let λ be an eigenvalue of A, and let $A_0 = A$. Then

$$0 = \det(A - I\lambda) = \det(Q_0 A_1 Q_0^* - \lambda(Q_0 Q_0^*)) = \det(Q_0(A_1 - \lambda I)Q_0^*) = \det(Q_0)\det(A_1 - \lambda I)\det(Q_0^*) =$$

$$= \det(Q_0)\det(A_1 - \lambda I)\det(Q_0^*) = \left\{ \det(Q_0^*) = \det(Q_0^{-1}) = \frac{1}{\det(Q_0)} \right\} =$$

$$= \det(Q_0)\det(A_1 - \lambda I)\frac{1}{\det(Q_0)} = \det(A_1 - \lambda I).$$

Since all Q_k are orthogonal, it follows, by induction, that all A_k have the same eigenvalues as A, Q.E.D.

Edited by jr, EJ

219 SF2524BlockC.C4solution

The difference between the Schur factorization and QR-factorization is that QR-factorization can be computed with a finite number of operations while Schur factorization will directly give us the eigenvalues. In equations, the QR-factorization is factorization of A as

$$A = QR$$

whereas the Schur factorization is the factorization

$$A = Q^T R Q$$

where in both cases Q is an orthogonal matrix and R an upper triangular matrix.

EJ's addition: In this course we learn procedures to compute the Schur factorization but not the QR-factorization.

Edited by An, EJ

220 SF2524BlockC.C5solution

Consider the 1×1 matrix $A = [1]$. Then the following are QR factorizations of A:

$$A = [1][1] = [-1][-1].$$

Edited by Be

221 SF2524BlockC.C6solution

Let $v = [0, u]^T$, where $0 \in \mathrm{R}^p$. Then, we have

$$vv^T = \begin{bmatrix} 0 \\ u \end{bmatrix} \begin{bmatrix} 0^T & u^T \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & uu^T \end{bmatrix}.$$

Therefore,

$$I - 2vv^T = \begin{bmatrix} I & 0 \\ 0 & I - 2uu^T \end{bmatrix} = P.$$

By construction, $||v|| = 1$ and thus, P is a householder reflector.

Edited by Anonymous

222 SF2524BlockC.Ca-7solution

The Householder reflectors Q_i must be applied to the matrix A first multiplied from the left in its

Hermitian conjugate form $Q_i^\star$ and then multiplied from the right:

$$A \rightarrow Q_1^\star A \rightarrow Q_1^\star A Q_1$$

If with the first reflector $Q_1^\star$ the whole first column were changed to a multiple of e_1, that means that all the rows of $A\ a_1, a_2. \ldots, a_m$ are used in a linear combination to introduce zeros in the first elements of $a_2. \ldots, a_m$. Consequently, when Q_1 is applied, all the columns of $A\ a^1, a^2. \ldots, a^m$ are used in a linear combination, so the first columns goes from being a multiple of e_1 to a linear combination of the other columns, losing the zeros it had gained in the preceding step.
This does not happen if we make the matrix Hessenberg because by not taking into consideration the first row with the first reflector, the first column will not be considered at the second multiplication, and the zeros will be preserved.

Edited by FI

223 SF2524BlockC.Ca-8solution

a)

$$A_1 = \begin{bmatrix} 3.5 & -2.3 & -4.63 \\ -2.06 & 2.15 & 5.28 \\ 0 & 3.32 & 3.53 \end{bmatrix}$$

b)

$$A_100 = \begin{bmatrix} 5.37 & -2.028 & 6.33 \\ 0 & 1 & -1.33 \\ 0 & 0 & -3.72 \end{bmatrix}$$

Edited by EJ

224 SF2524BlockC.Ca-10solution

$$Pu = u - 2u(u^*u) = u - 2\|u\|^2 u = u - 2u = -u$$

$$P^* = (I - 2uu^*)^* = I - 2(uu^*)^* = I - 2(u^*)^* u^* = I - 2uu^* = P$$

Edited by FI

225 SF2524BlockC.Ca-12solution

We know that the Givens Rotation of a vector x is given by,

$$G(i,j,c,s)x = \begin{bmatrix} x_1 \\ \vdots \\ x_{i-1} \\ cx_i - sx_j \\ x_{i+1} \\ \vdots \\ x_{j-1} \\ sx_i + cx_j \\ x_{j+1} \\ \vdots \\ x_n \end{bmatrix}$$

Therefore,

$$G(i,j,c,s)^T x = \begin{bmatrix} x_1 \\ \vdots \\ x_{i-1} \\ cx_i + sx_j \\ x_{i+1} \\ \vdots \\ x_{j-1} \\ -sx_i + cx_j \\ x_{j+1} \\ \vdots \\ x_n \end{bmatrix}$$

The Givens Rotation of a vector therefore only changes two elements of the vector. In the given problem, notice that the 4th and 5th elements of the vector appear to be changed. Therefore,

$$G(4,5,c,s)^T[-1,2,1,4,3,0]^T = [-1,2,1,c(4)+s(3),\ -s(4)+c(3),0]^T$$

Solving for $-s(4)+c(3)=0$ (given) and $c^2+s^2=1$ (condition for Givens Rotator), we get that $c=\frac{4}{5}$ and $s=\frac{3}{5}$. Therefore, the Givens Rotator is $G(4,5,\frac{4}{5},\frac{3}{5})$, and the element $\star$ is $c(4)+s(3)=5$.

Edited by An, EJ

226 SF2524BlockC.Ca-13solution

First note that we know from theorem in PDF lecture notes that the QR-factorization of a Hessenberg matrix has a particular structure such that it can be expressed with Givens rotators

$$H = QR = G_1 G_2 \cdots G_{n-1} R$$

where G_i are Givens rotators. More precisely,

$$G_i = G(i, i+1, c_i, s_i)$$

for some values c_i, s_i, and $G(i, i+1, \cdot, \cdot)$ is the Givens rotator acting on row and column i and $i+1$

One step of basic QR is therefore,

$$RQ = RG_1 \cdots G_{n-1}$$

We note that the structure of the first product becomes

$$RG_1 = \begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \\ & & & \times \end{bmatrix} \begin{bmatrix} \times & \times & & \\ \times & \times & & \\ & & 1 & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & & \times & \times \\ & & & \times \end{bmatrix}$$

Similarly,

$$RG_1G_2 = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & & \times & \times \\ & & & \times \end{bmatrix} \begin{bmatrix} 1 & & & \\ & \times & \times & \\ & \times & \times & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & \times & \times & \times \\ & & & \times \end{bmatrix}$$

By a more formal reasoning (with induction) one can show that

$$RG_1 \cdots G_{n-1} = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \end{bmatrix}$$

which is a Hessenberg matrix. Q.E.D.

Edited by EJ

227 SF2524BlockC.Ca-14solution

```
m = 5;
% Create a Hessenberg matrix A
A = randn(m);
A = triu(A,-1);
for i = 1:m-1
    r = norm(A(i:i+1,i));
    c = A(i,i)/r;
    s = A(i+1,i)/r;
    G = givens(i,i+1,c,s,m);
    A = G'*A;
end
```

with the Givens function

```
function G = givens(i,j,c,s,n)
    G = eye(n);
    G(i,i) = c;
    G(i,j) = -s;
    G(j,i) = s;
    G(j,j) = c;
end
```

Edited by Be

228 SF2524BlockC.Ca-15solution

First, it holds that

$$RQ = Q^T A Q$$

and if A is symmetric, it holds that

$$(Q^T A Q)^T = Q^T A^T Q = Q^T A Q$$

so that RQ is symmetric.

Edited by mb

229 SF2524BlockC.Ca-17solution

There are several downsides of the QR method mentioned. These are two of them: It takes quite a lot of iterations before convergence is reached and the second one is that the computation of one step is expensive.

Improvements that are mentioned are Shifted QR method and the two-phase approach. The shifted QR method is best used to reduce the number of iterations and so the convergence rate. This is, because the convergence rate depends on the ratio between 2 eigenvalues and this ratio is less when you subtract the shift from both eigenvalues and divide the two numbers then. The two-phase approach reduces the computation time of one step, because it never explicitly forms the matrix, only the parameters needed.

Edited by Anonymous

230 SF2524BlockC.Cb-2solution

This was covered in a lecture.

Proof of deflation:

Let $H = \begin{bmatrix} H_0 & H_1 \\ 0 & H_3 \end{bmatrix}$, $H_3 = [\,\square\,]$

- we can explicitly write down a QR-decomposition:

$$\underbrace{\begin{bmatrix} Q_{11} & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} R & Z \\ 0 & H_3 \end{bmatrix}}_{\text{Ansatz}} = \begin{bmatrix} H_0 & H_1 \\ 0 & H_3 \end{bmatrix}$$

(1) (2)

$$Q_{11} R = H_0 \quad \leftarrow \text{QR-decomp. of } H_0$$

$$Q_{11} Z = H_1 \Rightarrow$$

$$Z = Q_{11}^{-1} H_1$$

(Check that ansatz is correct:

$$\begin{bmatrix} Q_{11} & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} R & Q_{11}^{-1} H_1 \\ 0 & H_3 \end{bmatrix} = \underbrace{\begin{bmatrix} H_0 & H_1 \\ 0 & H_3 \end{bmatrix}}_{=H}$$

ok)

One step of QR:

$$\bar{H} = \begin{bmatrix} R & Q_{11}^{-1} H_1 \\ 0 & H_3 \end{bmatrix} \begin{bmatrix} Q_{11} & 0 \\ 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} RQ_{11} & Q_{11}^{-1} H_1 \\ 0 & H_3 \end{bmatrix}$$

One step of QR-method on H_0.

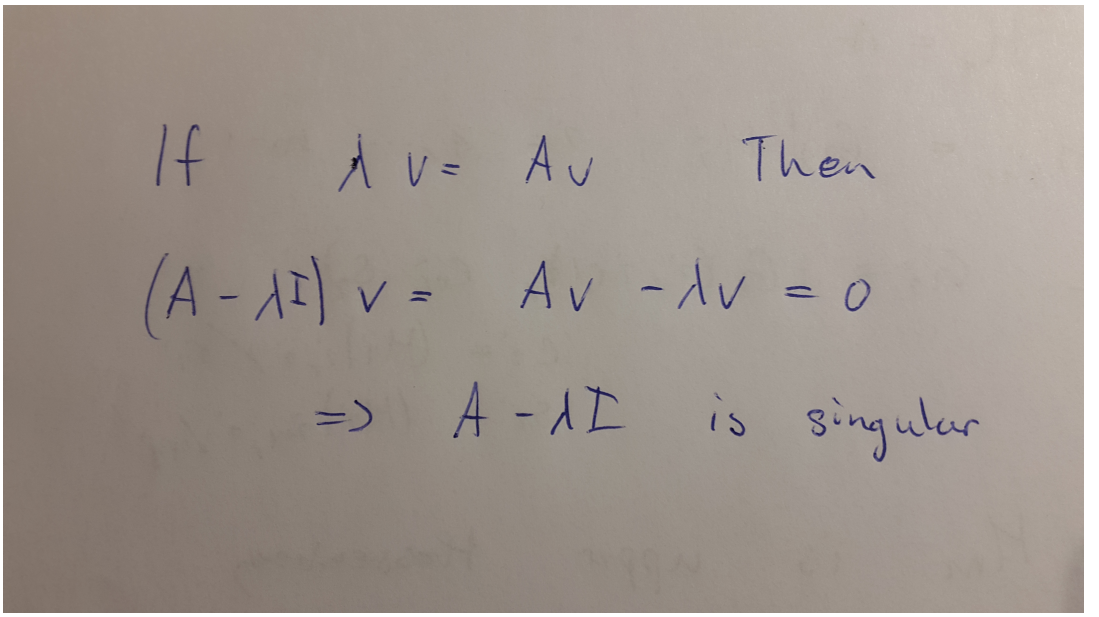
Edited by EJ

231 SF2524BlockC.Cb-3solution

If we perform a shift with $\lambda = 0$, we get back the original QR-method. This is seen instantly from looking at the definition of shifted QR in the lecture notes

Edited by CR

232 SF2524BlockC.Cb-4solution



Edited by EJ

233 SF2524BlockC.Cb-5solution

Since convergence is governed by the quantity

$$\max_{i=1,\dots,m-1} \frac{|\lambda_i|}{|\lambda_{i+1}|}$$

fast convergence requires the eigenvalues to be far apart.

a) Hence, an example of a bad collection of eigenvalues could for example be

$$\lambda_1 = 1.95, \lambda_2 = 2, \lambda_3 = 2.05.$$

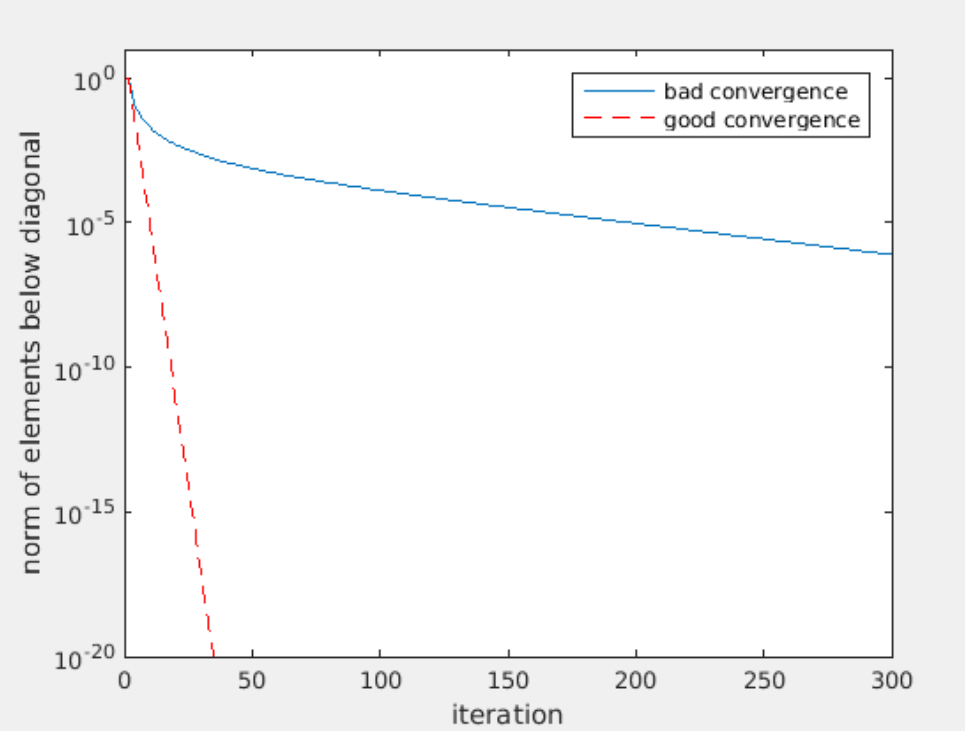
Corresponding matlab simulation

```
  >> A=[0 1 0; 0 0 1; 1599/200, -4799/400, 6]
     A=
         0    1.0000         0
         0         0    1.0000
    7.9950  -11.9975    6.0000
  >> eig(A)
  ans =
    1.9500
    2.0000
    2.0500
  >> errvec=[];
  >> for k=1:1000; [Q,R]=qr(A); A=R*Q; errvec=[errvec;norm(tril(A,-1))];
end
  >> semilogy(errvec);
```

b) A more well posed problem could have eigenvalues

$$\lambda_1 = 2, \lambda_2 = 8, \lambda_3 = 20$$

```
>> A=[0 1 0; 0 0 1; 210, -2016, 210]
A =
          0           1           0
          0           0           1
        210       -2016         210
>> eig(A)
ans =
  200.0000
    2.0000
    8.0000
```



Edited by mb, EJ

234 SF2524BlockC.Cb-6solution

Similarity transformations with an orthogonal matrix preserve symmetry. Let $A^T = A$ and $B = Q^T A Q$ where Q is an orthogonal matrix. Then

$$B^T = (Q^T A Q)^T = Q^T A^T Q = Q^T A Q = B$$

.

In the QR method we have $A_k = Q_k R_k$, $A_{k+1} = R_k Q_k = Q_k^T A_k Q_k$ with $Q^T Q = I$. Hence if $A_0 = A = A^T$, then $A_k = A_k^T$, the QR method preserves symmetry.

The $-\lambda I$ and $+\lambda I$ in the shifted QR method do not influence the symmetry property of A. If $A = A^T$ then $(A - \lambda I)^T = A - \lambda I$. Thus following the argument for the "normal" QR method above, RQ in the shifted QR will be a symmetric matrix and hence also $RQ + \lambda I$.

Edited by LL

235 SF2524BlockC.Cb-7solution

We know that the convergence of the basic QR method is governed by the quantity

$$\beta = \max_{i=1,\,\ldots\,,m-1} \frac{|\lambda_i|}{|\lambda_{i+1}|}$$

, given the eigenvalues are ordered as $|\lambda_i| < |\lambda_{i+1}|$. By considering that an error of 10^{-15} is what is practically achievable, we can see that the expected number of iterations is given by $\epsilon_k = \beta^k$. Therefore,

$$10^{-15} = \beta^k \implies -15 = k\log(3.95/4) \implies k \approx 2750$$

Edited by Anonymous

236 SF2524BlockD.Da-1solution

Let the Taylor series for $f\colon \mathrm{C} \to \mathrm{C}$ be convergent for the expansion point $\mu \in \mathrm{C}$, i.e.

$$f(z) = \sum_{k=0}^{\infty} \frac{f^k(\mu)}{k!}(z-\mu)^k.$$

Let A be a matrix. The Taylor definition of the matrix function associated with $f(z)$ (with expansion point $\mu \in \mathrm{C}$) is then given by

$$f(A) = \sum_{k=0}^{\infty} \frac{f^k(\mu)}{k!}(A - I\mu)^k.$$

Edited by jr

237 SF2524BlockD.Da-2solution

Suppose $A \in \mathrm{C}^{n\times n}$ and let

$$A = X\,\mathrm{diag}(J_1, \ldots, J_q)\,X^{-1},$$

be the Jordan canonical form with

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix} \in \mathrm{C}^{n_i\times n_i}.$$

The JCF-definition of the matrix function $f(A)$ is then given by

$$f(A) := X\,\mathrm{diag}(F_1, \ldots, F_q)\,X^{-1},$$

where

$$F_i = f(J_i) = \begin{bmatrix} f(\lambda_i) & \frac{f'(\lambda_i)}{1!} & \cdots & \frac{f^{(n_i-1)}(\lambda_i)}{(n_i-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \frac{f'(\lambda_i)}{1!} \\ & & & f(\lambda_i) \end{bmatrix} \in \mathrm{C}^{n_i\times n_i}.$$

Edited by Anonymous

238 SF2524BlockD.Da-4solution

Let $A \in C^{n\times n}$. We use the Schur factorization $A = Q^{\star}TQ$, where Q is unitary and T is upper triangular, to obtain

$$\exp(A) = Q^{\star}\exp(T)Q.$$

The Schur factorization is a similarity transformation of A to T triangular, such that the eigenvalues $\lambda_1, \ldots, \lambda_m$ of A are located on the diagonal of T. Thus $\exp(T)$ has the diagonal values $\exp(\lambda_1), \ldots, \exp(\lambda_m)$. Furthermore, the determinant of a unitary matrix Q fulfills $|det(Q)| = 1$ and the determinant of a triangular matrix equals the product of its diagonal entries such that

$$\det(\exp(A)) = \det(Q^{\star}) \cdot \det(\exp(T)) \cdot \det(Q) = \det(\exp(T)) = \exp(\lambda_1) \cdot \ldots \cdot \exp(\lambda_m) = \exp(\lambda_1 + \ldots + \lambda_m) = \exp(\mathrm{tr}(A)).$$

Edited by st

239 SF2524BlockD.Da-5solution

Suppose A is

a) symmetric, then $f(A) = [f(A)]^T$.
Jordan canonical form of a symmetric matrix: Let $A = X\Lambda X^{-1}$, then $A^T = (X\Lambda X^{-1})^T = (X^{-1})^T\Lambda X^T = X\Lambda X^{-1} = A$ where X is a unitary matrix, $X^TX = I$. Thus from the Jordan form definition we get

$$[f(A)]^T = [Xf(\Lambda)X^{-1}]^T = (X^{-1})^Tf(\Lambda)X^T = Xf(\Lambda)X^{-1} = f(A)$$

b) triangular, $f(A)$ is also triangular
This is e.g. shown in the derivation of the Schur-Parlett method, see Thm 4.2.1, or problem Db-1
[**2**]

c) hessenberg, then $f(A)$ is not hessenberg in general.

Consider a simple counter-example, $f(A) = A^2$ and $A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 3 & 1 \end{pmatrix}$. Then $A^2 = \begin{pmatrix} 7 & 13 & 10 \\ 6 & 13 & 13 \\ 9 & 6 & 7 \end{pmatrix}$ which

is not a Hessenberg marix.

d) anti-symmetric $A^T = -A$, then $f(A)^T \neq -f(A)$ in general.

Consider a simple counter-example, $f(A) = A^2$ and $A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ -1 & 0 & 3 & 1 \\ -2 & -3 & 0 & 2 \\ -3 & -1 & -2 & 0 \end{pmatrix}$. Then

$A^2 = \begin{pmatrix} -14 & -9 & -3 & 5 \\ -9 & -11 & -4 & 3 \\ -3 & -4 & -17 & -9 \\ 5 & 3 & -9 & -14 \end{pmatrix}$ which is not anti-symmetric. For this example $f(A)^T = f(A)$, but that

is not true in general.

e) diagonal, then $f(A)$ also is diagonal
This follows trivially from Taylor/Jordan from definition.

f) orthogonal, then $f(A)$ is not in generally orthogonal

Example: consider $Q = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Then $Q' * Q = Q * Q = I$, but $f(Q) = Q^2 + Q$ is not an

orthogonal matrix, $(Q^2 + Q)' * (Q^2 + Q) = Q^4 + Q'Q + Q^2Q + Q'Q^2 = 2 * I + Q + Q' \neq I$

Edited by LL

240 SF2524BlockD.Da-6solution

p is the characteristic polynomial of the matrix A so it can be written as

$$p(\lambda) = \prod_1^q (\lambda - \lambda_i)^{n_i}$$

where λ_i, n_i are the eigenvalues with their multiplicities.

From this expression of p it is clear that $p(\lambda_i) = 0, p'(\lambda_i) = 0, \ldots, p^{n_i-1}(\lambda_i) = 0$

Let $A = XJX^{-1}$ be the Jordan decomposition of the matrix A. By JCF-definition of matrix functions, $p(A) = Xp(J)X^{-1}$, where $p(J) = diag(p(J_1), \ldots, p(J_q))$.

Note that $p(J_i) = 0$ since $p(\lambda_i) = 0, p'(\lambda_i) = 0, \ldots, p^{n_i-1}(\lambda_i) = 0$. So $p(J)$ is the zero matrix and consequently

$$p(A) = 0$$

Edited by M

241 SF2524BlockD.Da-7solution

For this matrix A_ε, if a polynomial p interpolates a function f in the eigenvalues, then is it valid

$$p(A_\varepsilon) = p(X\mathrm{diag}(\lambda, \lambda+\varepsilon)X^{-1}) = X\mathrm{diag}(p(\lambda), p(\lambda+\varepsilon))X^{-1} = X\mathrm{diag}(f(\lambda), f(\lambda+\varepsilon))X^{-1} = f(X\mathrm{diag}(\lambda, \lambda+\varepsilon)X^{-1}) = f(A_\varepsilon)$$

A first degree polynomial interpolating $f(\lambda), f(\lambda+\varepsilon)$ is

$$p(z) = f(\lambda) - \frac{f(\lambda+\varepsilon) - f(\lambda)}{\varepsilon}\lambda + \frac{z}{\varepsilon}(f(\lambda+\varepsilon) - f(\lambda))$$

Then we can calculate $f(A_\varepsilon) = p(A_\varepsilon)$, so the limit is

$$\lim_{\varepsilon\to 0} f(A_\varepsilon) = \lim_{\varepsilon\to 0} p(A_\varepsilon) = (f(\lambda) - \lambda f'(\lambda))I + f'(\lambda)A_0 = \begin{bmatrix} f(\lambda) & f'(\lambda) \\ 0 & f(\lambda) \end{bmatrix}$$

And

$$f(A_0) = f\left(\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}\right) = \begin{bmatrix} f(\lambda) & f'(\lambda) \\ 0 & f(\lambda) \end{bmatrix}$$

because A_0 is a Jordan block. Hence,

$$\lim_{\varepsilon\to 0} f(A_\varepsilon) = f(A_0)$$

so we have continuity of the function for this matrix.

Edited by FI

242 SF2524BlockD.Da-8solution

Suppose f is analytic inside and on a simple closed piecewise-smooth curve Γ, which encloses the eigenvalues of A once counter-clockwise. The Cauchy integral definition of matrix functions is then given by

$$f(A) := \frac{1}{2i\pi}\oint_\Gamma f(z)(zI - A)^{-1}dz.$$

Edited by Anonymous

243 SF2524BlockD.Da-9solution

Let $f(A) = \frac{1}{2\pi i}\oint_\Gamma f(z)(Iz - A)^{-1}dz$ and $f(B) = \frac{1}{2\pi i}\oint_\Gamma f(y)(Iy - B)^{-1}dy$. Then

$$f(A)f(B) = \frac{1}{2\pi i}\oint_\Gamma f(z)(Iz - A)^{-1}dz\frac{1}{2\pi i}\oint_\Gamma f(y)(Iy - A)^{-1}dy = -\frac{1}{4\pi^2}\oint_\Gamma f(z)(Iz - A)^{-1}\oint_\Gamma f(y)(Iy - B)^{-1}dydz$$

$$= \oint_\Gamma\oint_\Gamma f(z)f(y)(Iz - A)^{-1}(Iy - B)^{-1}dydz = \oint_\Gamma\oint_\Gamma f(z)f(y)((Iz - A)(Iy - B))^{-1}dydz$$

$$= \oint_\Gamma\oint_\Gamma f(z)f(y)((Izy - Ay - Bz + AB))^{-1}dydz$$

If $AB = BA$, then

$$\oint_\Gamma\oint_\Gamma f(z)f(y)((Izy - Ay - Bz + AB))^{-1}dydz = \oint_\Gamma\oint_\Gamma f(z)f(y)(Izy - Bz - Ay + BA)^{-1}dydz$$

$$= \oint_\Gamma\oint_\Gamma f(z)f(y)((Iy - B)(Iz - A))^{-1}dydz = \oint_\Gamma\oint_\Gamma f(y)f(z)(Iy - B)^{-1}(Iz - A)^{-1}dydz$$

$$= \oint_\Gamma f(y)(Iy - B)^{-1}\oint_\Gamma f(z)(Iz - A)^{-1}dzdy = \oint_\Gamma f(y)(Iy - B)^{-1}dy\oint_\Gamma f(z)(Iz - A)^{-1}dz = f(B)f(A)$$

Edited by LL

244 SF2524BlockD.Da-10solution

For simplicity, suppose A is diagonalizable: $A = VDV^{-1}$.

$$y'' + VDV^{-1}y = 0$$

set $u = V^{-1}$, then this becomes:

$$u'' + Du = 0$$

This is a non-coupled equation in the components of u, with solution:

$$u_1 = c_1e^{i\sqrt{(}\lambda_1)x} + c_2e^{-i\sqrt{(}\lambda_1)x} = c_1^1cos(\sqrt{(}\lambda_1)x) + c_2^1sin(\sqrt{(}\lambda_1)x)$$

Now we change coordinates back to y:

$$y(t) = Vcos(\sqrt{(}D)x)c_1 + Vsin(\sqrt{(}D)x)$$

The 1st initial condition yields $c_1 = V^{-1}y_0$, then set $c_2 = V^{-1}c$:

$$y = cos(\sqrt{(}A))y_0 + sin(\sqrt{(}A))c$$

Now the second initial condition yields:

$$y = cos(\sqrt{(}A))y_0 + \sqrt{(}A)^{-1}sin(\sqrt{(}A))y_0'$$

Edited by JW

245 SF2524BlockD.Da-11solution

The definition is valid for matrix A if:

$$f(z) \text{ is analytic in } \bar{B}(\mu, r)$$

and

$$\frac{||A - \mu I||}{r} < 1$$

Edited by JW

246 SF2524BlockD.Da-12solution

$$A = \begin{bmatrix} 1 & 1/\sqrt{(2)} \\ 0 & 1/\sqrt{(2)} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & \sqrt{(2)} \end{bmatrix}$$

Thus

$$\exp(A) = \begin{bmatrix} 1 & 1/\sqrt{(2)} \\ 0 & 1/\sqrt{(2)} \end{bmatrix} \begin{bmatrix} e^2 & 0 \\ 0 & e \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & \sqrt{(2)} \end{bmatrix}$$

Edited by JW

247 SF2524BlockD.Da-13solution

We have the Taylor-definition (4.1.1 LN), the Jordan definition (4.1.2 LN) and the Caucy Integral Definition (4.1.3 LN).

Benefit Taylor: It is the most intuitive approach. We want the definitions to agree with matrix multiplication and addition so we define a matrix function in terms of a matrix polynomial corresponding to the taylor series for the considered function. Consequence: Hard to know how many terms is needed for a specific convergence. Not very general in the sense it requires convergent Taylor series.

Benefit Jordan: This a more general definition, as it does not require all derivatives and convergence of the taylor series. It also permits a simple approach to exact computation. Consequence: Not very intuitive, and requires the Jordan decomposition to be computed.

Benefit Cauchy: Allows for application of Krylov methods, which can be very effective. Consequence: Need to compute eigenvalues to be sure we encircle them with our chosen contour.

Edited by CR

248 SF2524BlockD.Da-15solution

We use the Jordan decomposition of A and B

$$A = X_A J_A X_A^{-1} \quad B = X_B J_B X_B^{-1}$$

Then the Jordan decomposition of our block matrix is

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} X_A & 0 \\ 0 & X_B \end{pmatrix} \begin{pmatrix} J_A & 0 \\ 0 & J_B \end{pmatrix} \begin{pmatrix} X_A & 0 \\ 0 & X_B \end{pmatrix}^{-1}$$

Now use the JCF-definition of matrix functions

$$f\left(\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\right) = \begin{pmatrix} X_A & 0 \\ 0 & X_B \end{pmatrix} f\left(\begin{pmatrix} J_A & 0 \\ 0 & J_B \end{pmatrix}\right) \begin{pmatrix} X_A & 0 \\ 0 & X_B \end{pmatrix}^{-1} = \begin{pmatrix} X_A & 0 \\ 0 & X_B \end{pmatrix} \begin{pmatrix} f(J_A) & 0 \\ 0 & f(J_B) \end{pmatrix} \begin{pmatrix} X_A & 0 \\ 0 & X_B \end{pmatrix}^{-1} =$$

$$= \begin{pmatrix} X_A f(J_A) X_A^{-1} & 0 \\ 0 & X_B f(J_B) X_B^{-1} \end{pmatrix} = \begin{pmatrix} f(A) & 0 \\ 0 & f(B) \end{pmatrix}$$

Edited by M

249 SF2524BlockD.Db-1solution

Let T be a triangular matrix with distinct eigenvalues and let

$$f(T) = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{13} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}.$$

- a) Give the values for $f_{11}, f_{21}, f_{31}, f_{22}, f_{32}, f_{33}$ in term of the entries of T
  - Consider the Taylor definition of a matrix function:

$$f(A) = \sum_{i=0}^{\infty} \frac{f^{(i)}(\mu)}{i!}(A - \mu I)^i$$

A and thus $A - \mu I$ are triangular matrices, and so are all their powers. Thus $f(A)$ is a sum of (scaled) triangular matrices and therefore also triangular, thus $f_{21} = f_{31} = f_{32} = 0$.

Furthermore, the diagonal entries of the triangular matrix powers are the powers of the entries of the original matrix

$$((A - \mu I)^i)_{jj} = (A_{jj} - \mu)^i$$

. Hence

$$(f(A))_{jj} = (\sum_{i=0}^{\infty} \frac{f^{(i)}(\mu)}{i!}(A - \mu I)^i)_{jj} = \sum_{i=0}^{\infty} \frac{f^{(i)}(\mu)}{i!}(A - \mu I)^i_{jj} = \sum_{i=0}^{\infty} \frac{f^{(i)}(\mu)}{i!}(A_{jj} - \mu)^i = f(A_jj)$$

and for the example $f_{11} = f(T_{11}), f_{22} = f(T_{22}), f_{33} = f(T_{33})$, i.e. we have

$$f(T) = \begin{bmatrix} f(T_{11}) & f_{12} & f_{13} \\ 0 & f(T_{22}) & f_{13} \\ 0 & 0 & f(T_{33}) \end{bmatrix}.$$

- b) Derive an explicit formula for f_{12} involving only elements of T and the values computed in (a).

All matrix function definitions satisfy

$$Tf(T) = f(T)T$$

T and $f(T)$ commute.

$$Tf(T) - f(T)T = 0$$

$$\begin{bmatrix} t_{11}f_{11} & t_{11}f_{12} + t_{12}f_{22} & \star \\ 0 & t_{22}f_{22} & \star \\ 0 & 0 & \star \end{bmatrix} - \begin{bmatrix} f_{11}t_{11} & f_{11}t_{12} + f_{12}t_{22} & \star \\ 0 & f_{22}t_{22} & \star \\ 0 & 0 & \star \end{bmatrix} = 0$$

$$t_{11}f_{12} + t_{12}f_{22} - f_{11}t_{12} - f_{12}t_{22} = 0$$

$$f_{12}(t_{11} - t_{22}) = (f_{11} - f_{22})t_{12}$$

$$f_{12} = t_{12}\frac{f(T_{11}) - f(T_{22})}{T_{11} - T_{22}}$$

Edited by LL

Theorem 3.2.1 gives the formula

$$f_{ij} = \frac{s}{t_{jj} - t_{ii}},$$

where

$$s = t_{ij}(f_{jj} - f_{ii}) + \sum_{k=i+1}^{j-1} f_{ik}t_{kj} - t_{ik}f_{kj}.$$

The non-zero elements of F and T are denoted by $+$ for already computed and $\square$ for unknowns.

In the first figure all elements of F which are required in the formula above to compute f_{ij} are marked, i.e. f_{kj}, f_{ik} for $k = i + 1, \ldots, j - 1$ as well as f_{ii} and f_{jj}.

In the second figure all elements of T which are required in the formula above to compute f_{ij} are marked, i.e. t_{ik}, t_{kj} for $k = i + 1, \ldots, j - 1$ as well as t_{ij}, t_{ii} and t_{jj}.

It is notable that the formula (or rather the algorithm, we can define with it) computes F recursively subdiagonal-wise. It starts at the first upper subdiagonal (since we know the diagonal of F), calculates it element-wise downwards and continues with the first element of the next upper subdiagonal. We will finally end up, computing the top right element of F by using the other elements in its first row and its last column.

Edited by st

251 SF2524BlockD.Db-2solution

The Schur-Parlett method computes $f(A)$ for a given matrix function f. Using the Schur factorization $A = Q^\star TQ$, where Q is orthogonal and T is upper triangular, and the fact that the theory for matrix functions agrees with similarity transformations, we obtain

$$f(A) = Q^\star f(T)Q.$$

Thus we compute Q, $Q^\star$, $f(T)$ and plug them into this formula, instead of directly approaching $f(A)$. The matrix Q stems from the Schur factorization and Theorem 3.2.1 in the lecture notes provides an algorithm to obtain $f(T)$ based on its triangular form.

Edited by st, An

252 SF2524BlockD.Db-4solution

If the eigenvalues are very close, we will end up with numerical cancellation effects in the algorithm, resulting in numerical instability. This happens in the line reading $f_{ij} = s/(t_{jj} - t_{ii})$ in the algorithm. We may resolve this by using a blocks in the Schur form.

Example. If we apply the Schur-Parlett method to the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 + t \end{bmatrix}$$

with the matrix exponential we get

$$\exp(A) = \begin{bmatrix} e^1 & 2\frac{e^{1+t} - e^1}{t} \\ 0 & e^{1+t} \end{bmatrix}$$

The "problem" with this expression is that the term $\frac{e^{1+t}-e^1}{t}$ will suffer from numerical cancellation when t is small (for the same reason that finite difference approximation is inaccurate for small steplengths). For illustration purposes we can now for this specific example reformulate the expression such that we avoid cancellation:

$$\exp(A) = \begin{bmatrix} e^1 & 4e^{1+t/2}\sinh(t/2)/t \\ 0 & e^{1+t} \end{bmatrix}.$$

The MATLAB-code below shows that Schur-Parlett is inaccurate (due to round-off errors) when the eigenvalues are close, which corresponds to small t.

```
>> A=@(t) [1 2 ; 0 1+t];
>> expm_SP=@(t) [exp(1), 2*(exp(1+t)-exp(1))/t; 0 exp(1+t)];
>> expm_exact=@(t) [exp(1),4*exp(1+t/2)*sinh(t/2)/t ; 0 exp(1+t)];
>> norm(expm_SP(0.1)-expm_exact(0.1))
ans =
     0
>>  norm(expm_SP(0.0001)-expm_exact(0.0001))
ans =
    5.1363e-12
>>  norm(expm_SP(1e-6)-expm_exact(1e-6))
ans =
     1.2278e-09
>>  norm(expm_SP(1e-9)-expm_exact(1e-9))
ans =
      4.6001e-07
>> norm(expm_SP(eps)-expm_exact(eps))
ans =
       1.4366
```

Edited by CR, EJ

253 SF2524BlockD.Dc-1solution

Let

$$A = \begin{bmatrix} 1 & \alpha \\ 0 & 2 \end{bmatrix}.$$

The eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = 2$. Let

$$p(z) = \exp(1) + (\exp(2) - \exp(1))(z - 1)\,.$$

Hence, p interpolates $\exp$ at $z = \lambda_1$ and $z = \lambda_2$. We have thus that

$$\exp(A) = p(A) = \exp(1)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + (\exp(2) - \exp(1))\begin{bmatrix} 0 & \alpha \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \exp(1) & \alpha(\exp(2) - \exp(1)) \\ 0 & \exp(2) \end{bmatrix}.$$

Edited by Be

254 SF2524BlockD.Dc-4solution

a) The identity does not hold in general. We can for instance take

```
>> A=-[1 2; 3 4];
>> B=-[1 3; 3 4];
>> expm(A)*expm(B)
ans =
    2.4086   -1.4881
   -1.6521    1.0207
>> expm(A+B)
ans =
    2.5706   -1.3903
   -1.6683    0.9023
```

It does however hold if the matrices commute: $AB = BA$. So if we take $A = I$ we get equality:

```
>> expm(A)*expm(B)
ans =
   4.6231   -2.8524
  -2.8524    1.7707
>> expm(A+B)
ans =
   4.6231   -2.8524
  -2.8524    1.7707
```

b) From the basic properties of matrix functions we know that if $f(z) = g(h(z))$, then

$$f(A) = g(h(A)).$$

If we select $f(z) = \exp(z) = \exp(z/2)\exp(z/2)$ we can decompose it in a trivial way by setting $h(z) = \exp(z/2)$ and $g(x) = x * x$. Hence,

$$\exp(A) = f(A) = g(h(A)) = \exp(A/2)\exp(A/2).$$

Edited by EJ

255 SF2524BlockD.Dc-5solution

So we truncate the series for $e^{A/16} = I + A/16 + A^2/36 + \ldots$

Then we note that $e^A = (e^{A/16})^{2^4}$. That is, we need to square it 4 times. Therefore we arrive at:

```
  6.86790  -5.59880  -0.68577
 -3.73966   5.94995   4.74273
  4.91020  -5.10415  -0.48180
```

Edited by JW

256 SF2524BlockD.Dd-2solution

Can be directly verified with matrix-matrix multiplication

```
>> A=[1 1 1 1 ; 0 -1 -2 -3 ; 0 0 1 3; 0 0 0 -1];
>> A*A
ans =
  1     0     0     0
  0     1     0     0
  0     0     1     0
  0     0     0     1
```

Edited by EJ

257 SF2524BlockD.Dd-1solution

The Denman-Beavers iteration is an iterative method for the matrix square root.

It is defined by the iteration

$$Y_{k+1} = \frac{1}{2}(X_k + Y_k^{-1})$$

$$X_{k+1} = \frac{1}{2}(X_k^{-1} + Y_k)$$

In exact arithmetic it is equivalent with Newton-SQRT. It has very different properties in finite arithmetic. It is in generally more numerically stable than Newton-SQRT.

Edited by EJ

258 SF2524BlockD.De-1solution

Note first that the scalar sign function can be expressed as

$$\mathrm{sign}(x) = \frac{|x|}{x} = \frac{\sqrt{x^2}}{x}$$

Hence, a natural generalization to matrices is given by

$$\mathrm{sign}(A) = A^{-1}\sqrt{A^2}$$

Edited by mb

259 SF2524BlockD.De-2solution

The matrix sign function has been proven to be quadratically convergent. It is given by the iteration

$$S_{k+1} = \frac{1}{2}(S_k + S_k^{-1})$$

Edited by EJ

260 SF2524BlockD.De-3solution

We first note that if we define the sign as

$$\mathrm{s}(z) = \sqrt{z^2}/z$$

the derivative is

$$\mathrm{s}'(z) = 1/\sqrt{z^2} - \sqrt{z^2}/z^2 = 0.$$

By induction one can show that $\mathrm{s}^{(k)}(z) = 0$.

Therefore the Jordan definition becomes

$$\mathrm{sign}(A) = V\mathrm{diag}(s(\lambda_1), \ldots, s(\lambda_n))V^{-1}$$

The diagonal matrix has elements ± 1.

Edited by EJ

261 SF2524BlockD.Df-1solution

The Krylov approximation of $f(A)b$ is

$$f(A)b \approx Q_m f(H_m)e_1 \|b\|,$$

where Q_m and $\underline{H}_m$ correspond to an Arnoldi factorization

$$AQ_m = Q_{m+1}\underline{H}_m$$

where $Q_{m+1} = [Q_m, q_{m+1}] \in \mathrm{R}^{n\times(m+1)}$ is orthogonal and $\underline{H}_m \in \mathrm{R}^{(m+1)\times m}$ is upper Hessenberg and $H_m \in \mathrm{R}^{m\times m}$ is the upper submatrix of $\underline{H}_m$. Moreover, the starting vector of the Arnoldi factorization is $q_1 = b/\|b\|$.

Edited by EJ

262 SF2524BlockD.Df-2solution

The φ function is a function defined by

$$\varphi(x) := \begin{cases} \frac{\exp(x)-1}{x} & x \neq 0 \\ 1 & x = 0 \end{cases}$$

It is an entire function (analytic everywhere) since the special case $x = 0$ corresponds to to analytic continuation at a removable pole. We have

$$\lim_{x\to 0} \varphi(x) = 1.$$

Edited by EJ

263 SF2524BlockD.Df-3solution

The ODE

$$y'(t) = Ay(t) + b$$

with initial condition $y(0) = y_0$ has solution

$$y(t) = y_0 + t\varphi(tA)g_1(y_0)$$

where $g_1(t) = Ay(t) + b$ and $\varphi(z) = \frac{e^z-1}{z}$

Edited by M

264 SF2524BlockD.Df-4solution

If you have a differential equation of the following form: $y'(t) = -Ay(t) + \mathcal{N}(t, y)$ and $y(0) = y_0$.

Then we can approximate the solution with the Forward Euler Exponential integrator:

$y_{n+1} = e^{-Ah}y_n + A^{-1}(1 - e^{-Ah})\mathcal{N}(t_n, y(t_n))$ with h the timestep. ($h = t_{n+1} - t_n$).

Edited by Anonymous