Utrecht, 26 oktober 2016

Basic Iterative Methods II

Gerard Sleijpen

Universiteit Utrecht

Department of Mathematics

http://www.staff.science.uu.nl/∼sleij101/

Linear Problems

A is a given $n \times n$ matrix

Typically

- n is large ($10^5 - 10^8$)
- **A** has additional properties
  that have to be identified and exploited.
  Properties as
  - real
  - sparse (nearly sparse)
  - banded structure
  - Hermitian (near Hermitean)
  - Normal, positive definite, . . .
  - Tensor product
  - . . .

Linear Problems

A is a given $n \times n$ matrix

Linear equations.

For a given n-vector **b** solve $\quad \mathbf{Ax} = \mathbf{b} \quad$ for **x**.

Terminology and notation. The **spectrum** of **A**,

$$\Lambda(\mathbf{A}) \equiv \{\lambda \in \mathbb{C} \mid \mathbf{Av} = \lambda\, \mathbf{v} \text{ for some } n\text{-vector } \mathbf{v} \neq \mathbf{0}\},$$

is the set of all eigenvalues of **A**.

If $\mathbf{Av} = \lambda \mathbf{v}$, $\mathbf{v} \neq \mathbf{0}$, then **v** is an eigenvector *associated* to λ.

$$\Lambda(\mathbf{A}) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\} = \{\lambda_j(\mathbf{A})\}$$

Eigenvalues are counted according to multiplicity.

Linear Problems

$\mathbf{A}$ is a given $n \times n$ matrix

Linear equations.

For a given n-vector $\mathbf{b}$ solve $\quad \mathbf{A}\mathbf{x} = \mathbf{b} \quad$ for $\mathbf{x}$.

Eigenvalue problem. Solve $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$

for $\lambda = \lambda_0 = \lambda_{j_0}$ and associated $\mathbf{v} = \mathbf{v}_0 = \mathbf{v}_{j_0}$, where,

$\lambda_0 \in \Lambda(\mathbf{A})$, the "**wanted**" eigenvalue, has a property as

- $\lambda_0 = \text{argmax}\{|\lambda| \mid \lambda \in \Lambda(\mathbf{A})\}$ or $\text{argmin}\{|\lambda| \mid \lambda \in \Lambda(\mathbf{A})\}$
- $\lambda_0 = \text{argmax}\{\text{Re}(\lambda) \mid \lambda \in \Lambda(\mathbf{A})\}$
- For some **target value** $\tau \in \mathbb{C}$,
$$\lambda_0 = \text{argmin}\{|\lambda - \tau| \mid \lambda \in \Lambda(\mathbf{A})\}$$
- For some **target n-vector** $\mathbf{w}$,
$$\mathbf{v}_0 = \text{argmin}\{\angle(\mathbf{v}, \mathbf{w}) \mid \mathbf{A}\mathbf{v} = \lambda\mathbf{v} \text{ for some } \lambda\}$$
- ...

Solving Linear problems

Ideas/techniques for solving linear equations can often be translated to ideas/techniques for solving eigenvalue problems, and visa versa.

This observation may be explained by the following

Proposition. $\mathbf{x}$ solves

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

if and only if $\mathbf{v} \equiv \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix}$ with $\widetilde{\mathbf{A}} \equiv \begin{bmatrix} 0 & \mathbf{0}^* \\ -\mathbf{b} & \mathbf{A} \end{bmatrix}$ solves

$$\widetilde{\mathbf{A}}\mathbf{v} = 0\mathbf{v}.$$

Proposition. $\Lambda(\widetilde{\mathbf{A}}) = \Lambda(\mathbf{A}) \cup \{0\}$.

Program Lecture 5

- Power Method & Richardson
- Filtering
- Shift-and-Invert & Preconditioning
- Polynomial Iteration
- Selecting Parameters

  1) single parameter a) static
  b) dynamic

  2) Multiple parameters a) static (Chebyshev)
  b) dynamic (GCR)

Program Lecture 5

- Power Method & Richardson
- Filtering
- Shift-and-Invert & Preconditioning
- Polynomial Iteration
- Selecting Parameters

  1) single parameter a) static
  b) dynamic

  2) Multiple parameters a) static (Chebyshev)
  b) dynamic (GCR)

$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ **Shifted power**: $\tilde{\mathbf{u}}_k = (\mathbf{A} - \sigma\mathbf{I})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \tilde{\mathbf{u}}_k / \|\tilde{\mathbf{u}}_k\|_2$

Theorem.

The $\mathbf{u}_k$ converge to (a multiple of) $\mathbf{v}_{j_0}$ if

$|\lambda_{j_0} - \sigma| > |\lambda_j - \sigma|$ all $j \neq j_0$:

and $\mathbf{u}_0$ has a component in the direction of $\mathbf{v}_{j_0}$

$\mathbf{v}_{j_0}$ is the **dominant** eigenvector of $\mathbf{A} - \sigma\mathbf{I}$,
and $\lambda_{j_0} - \sigma$ is the **dominant** eigenvalue.

Eventual error reduction is $\rho \equiv \max_{j \neq j_0} \frac{|\lambda_j - \sigma|}{|\lambda_{j_0} - \sigma|}$

$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ **Shifted power**: $\tilde{\mathbf{u}}_k = (\mathbf{A} - \sigma\mathbf{I})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \tilde{\mathbf{u}}_k / \|\tilde{\mathbf{u}}_k\|_2$

Use the **Rayleigh quotient** to compute the eigenvalue

$$\lambda_1^{(k)} = \rho(\mathbf{u}_k) \equiv \frac{\mathbf{u}_k^* \mathbf{A}\mathbf{u}_k}{\mathbf{u}_k^* \mathbf{u}_k},$$

to compute the approximate eigenvalue

Exercise. If $\mathbf{u} \neq \mathbf{0}$, then

$$\|\mathbf{A}\mathbf{u} - \frac{\mathbf{u}^*\mathbf{A}\mathbf{u}}{\mathbf{u}^*\mathbf{u}}\mathbf{u}\|_2 \leq \|\mathbf{A}\mathbf{u} - \mu\mathbf{u}\|_2 \quad \forall \mu \in \mathbb{C}.$$

Given the approximate $\mathbf{u} = \mathbf{u}_k$, the Rayleigh quotient gives the smallest residual.

$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ **Shifted power**: $\widetilde{\mathbf{u}}_k = (\mathbf{A} - \sigma \mathbf{I})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \widetilde{\mathbf{u}}_k / \|\widetilde{\mathbf{u}}_k\|_2$

Alternative scalings: $\mathbf{u}_{k+1} = \widetilde{\mathbf{u}}_k / \mathbf{e}_1^* \widetilde{\mathbf{u}}_k$,

fix the first coordinate to 1

Use the Rayleigh quotient or the quotient

$$\lambda_1^{(k)} = \lambda_1^{(k)} = \frac{\mathbf{e}_1^* \mathbf{A} \mathbf{u}_k}{\mathbf{e}_1^* \mathbf{u}_k}$$

to compute the approximate eigenvalue

$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ **Shifted power**: $\widetilde{\mathbf{u}}_k = (\mathbf{I} - \alpha\mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \widetilde{\mathbf{u}}_k / \mathbf{e}_1^* \widetilde{\mathbf{u}}_k$

Different scaling (of $\mathbf{A} - \sigma\mathbf{I}$ or/and of $\widetilde{\mathbf{u}}_k$) does not affect the convergence of the power method.

$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ **Shifted power**: $\widetilde{\mathbf{u}}_k = (\mathbf{I} - \alpha \mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \widetilde{\mathbf{u}}_k / \mathbf{e}_1^* \widetilde{\mathbf{u}}_k$

$\mathbf{A}\mathbf{x} = \mathbf{b}$ **Richardson:** $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha(\mathbf{b} - \mathbf{A}\mathbf{x}_k)$

Note that

$$\begin{bmatrix} 1 \\ \mathbf{x}_{k+1} \end{bmatrix} = (\mathbf{I} - \alpha \widetilde{\mathbf{A}}) \begin{bmatrix} 1 \\ \mathbf{x}_k \end{bmatrix}, \text{ where } \widetilde{\mathbf{A}} \equiv \begin{bmatrix} 0 & \mathbf{0}^* \\ -\mathbf{b} & \mathbf{A} \end{bmatrix}$$

Observation.

Richardson for '$\mathbf{A}\mathbf{x} = \mathbf{b}$' with **relaxation parameter** α

$=$ the Shifted power method for '$\widetilde{\mathbf{A}}\mathbf{v} = \lambda \mathbf{v}$ for $\lambda = 0$'.

$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ **Shifted power**: $\tilde{\mathbf{u}}_k = (\mathbf{I} - \alpha \mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \tilde{\mathbf{u}}_k / \mathbf{e}_1^* \tilde{\mathbf{u}}_k$

$\mathbf{A}\mathbf{x} = \mathbf{b}$ **Richardson:** $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha(\mathbf{b} - \mathbf{A}\mathbf{x}_k)$

$$\mathbf{x} = \mathbf{x} + \alpha(\mathbf{b} - \mathbf{A}\mathbf{x})$$

$$\mathbf{e}_{k+1} = (\mathbf{I} - \alpha \mathbf{A})\mathbf{e}_k$$

Here, $\mathbf{e}_k = \mathbf{x} - \mathbf{x}_k$.

Converges iff

$$|1 - \alpha\lambda| < 1 \quad \text{for all } \lambda \in \Lambda(\mathbf{A})$$

$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ **Shifted power**:

$$\tilde{\mathbf{u}}_k = (\mathbf{I} - \alpha \mathbf{A})\mathbf{u}_k$$

Scale $\mathbf{u}_{k+1} = \tilde{\mathbf{u}}_k / \mathbf{e}_1^* \tilde{\mathbf{u}}_k$

$\mathbf{A}\mathbf{x} = \mathbf{b}$ **Richardson:**

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha(\mathbf{b} - \mathbf{A}\mathbf{x}_k)$$

$$\mathbf{x} = \mathbf{x} + \alpha(\mathbf{b} - \mathbf{A}\mathbf{x})$$

$$\mathbf{r}_{k+1} = (\mathbf{I} - \alpha \mathbf{A})\mathbf{r}_k$$

Here, $\mathbf{r}_k = \mathbf{A}\mathbf{e}_k = \mathbf{A}(\mathbf{x} - \mathbf{x}_k) = \mathbf{b} - \mathbf{A}\mathbf{x}_k$.

Converges iff

$$|1 - \alpha\lambda| < 1 \quad \text{for all } \lambda \in \Lambda(\mathbf{A})$$

$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ **Shifted power**: $\widetilde{\mathbf{u}}_k = (\mathbf{I} - \alpha \mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \widetilde{\mathbf{u}}_k / \mathbf{e}_1^* \widetilde{\mathbf{u}}_k$

Improvements are based on the fact that

$$f(\mathbf{A})\mathbf{v}_j = f(\lambda_j)\mathbf{v}_j.$$

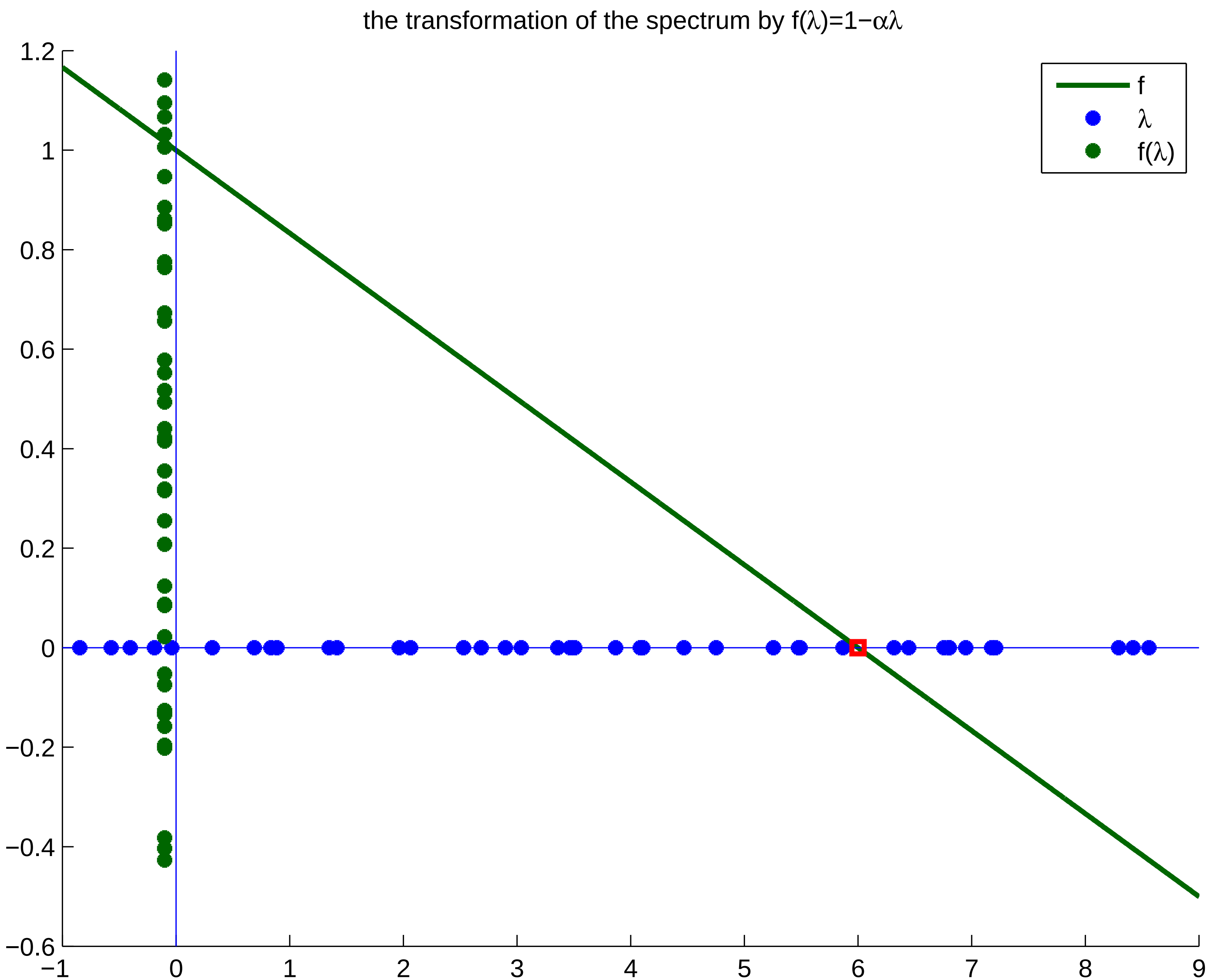
Examples. $f(\mathbf{A}) = (\mathbf{I} - \alpha \mathbf{A})$

Same eigenvectors, **better eigenvalue distribution!**

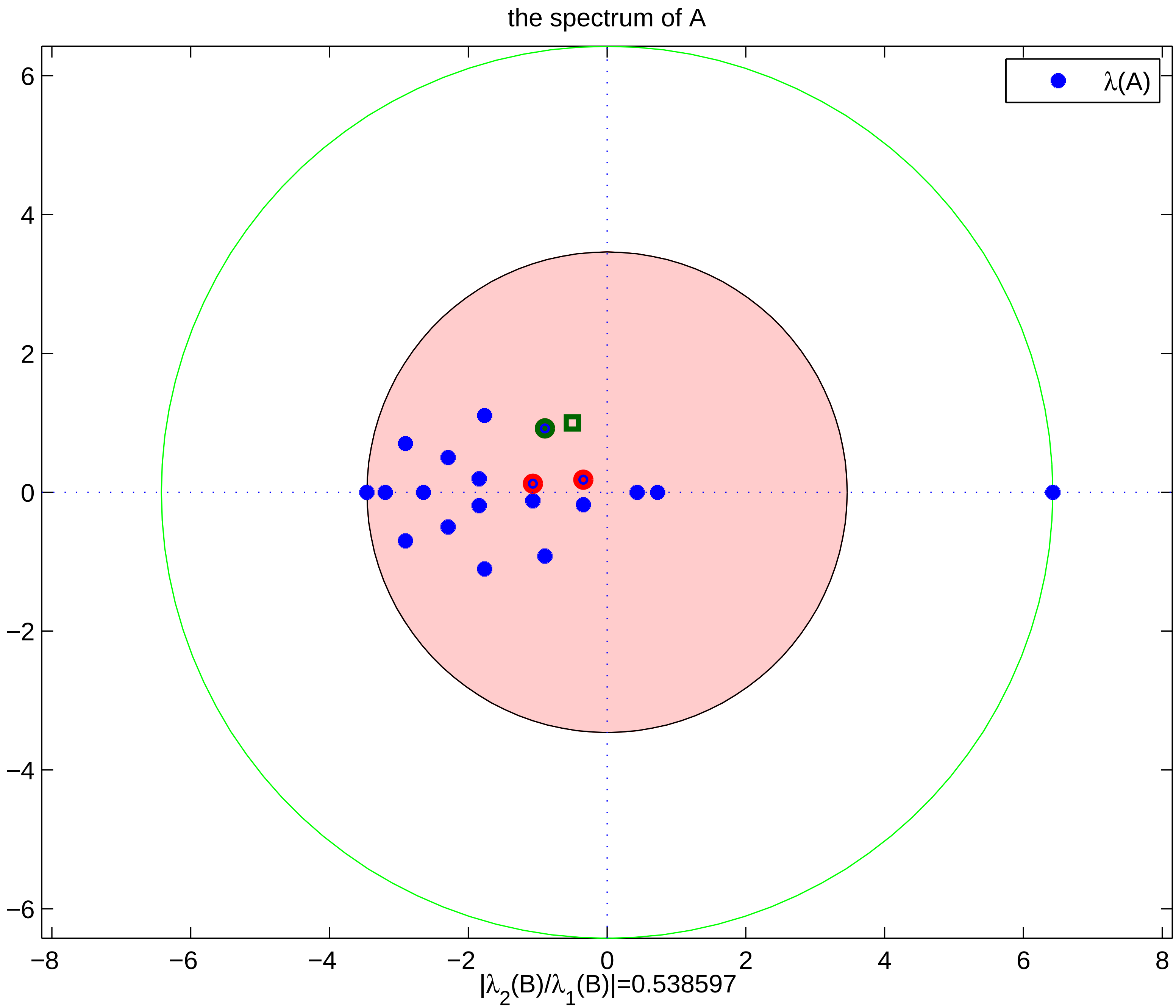
Find f such that $|f(\lambda_{j_0})|$ is relatively large, i.e., ρ small,

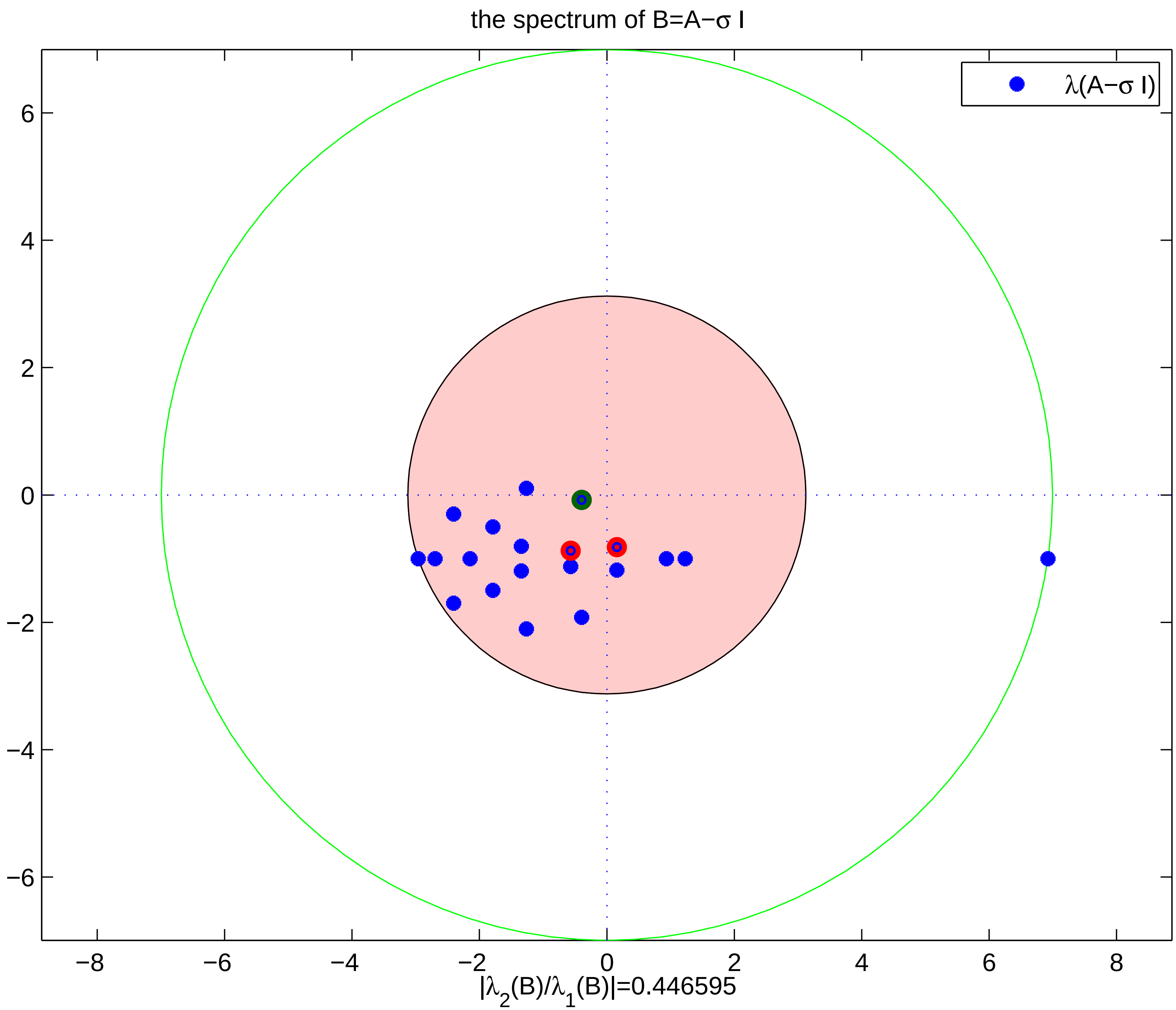
where $$\rho \equiv \max_{j \neq j_0} \frac{|f(\lambda_j)|}{|f(\lambda_{j_0})|}$$

($\rho < 1$ for convergence): ρ is the asymptotic error reduction (i.e., for $k \to \infty$) for the power method with $f(\mathbf{A})$.



the transformation of the spectrum by f(λ)=1−αλ
f
λ
f(λ)
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−1
0
1
2
3
4
5
6
7
8
9


the spectrum of A
λ(A)
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4
6
8
$|\lambda_2(B)/\lambda_1(B)|$=0.538597


the spectrum of B=A−σ I
λ(A−σ I)
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4
6
8
|λ2(B)/λ1(B)|=0.446595

Program Lecture 5

- Power Method & Richardson
- Filtering
- Shift-and-Invert & Preconditioning
- Polynomial Iteration
- Selecting Parameters

  1) single parameter a) static
b) dynamic

  2) Multiple parameters a) static (Chebyshev)
b) dynamic (GCR)

$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ **Shifted power**: $\tilde{\mathbf{u}}_k = (\mathbf{I} - \alpha \mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \tilde{\mathbf{u}}_k / \mathbf{e}_1^* \tilde{\mathbf{u}}_k$

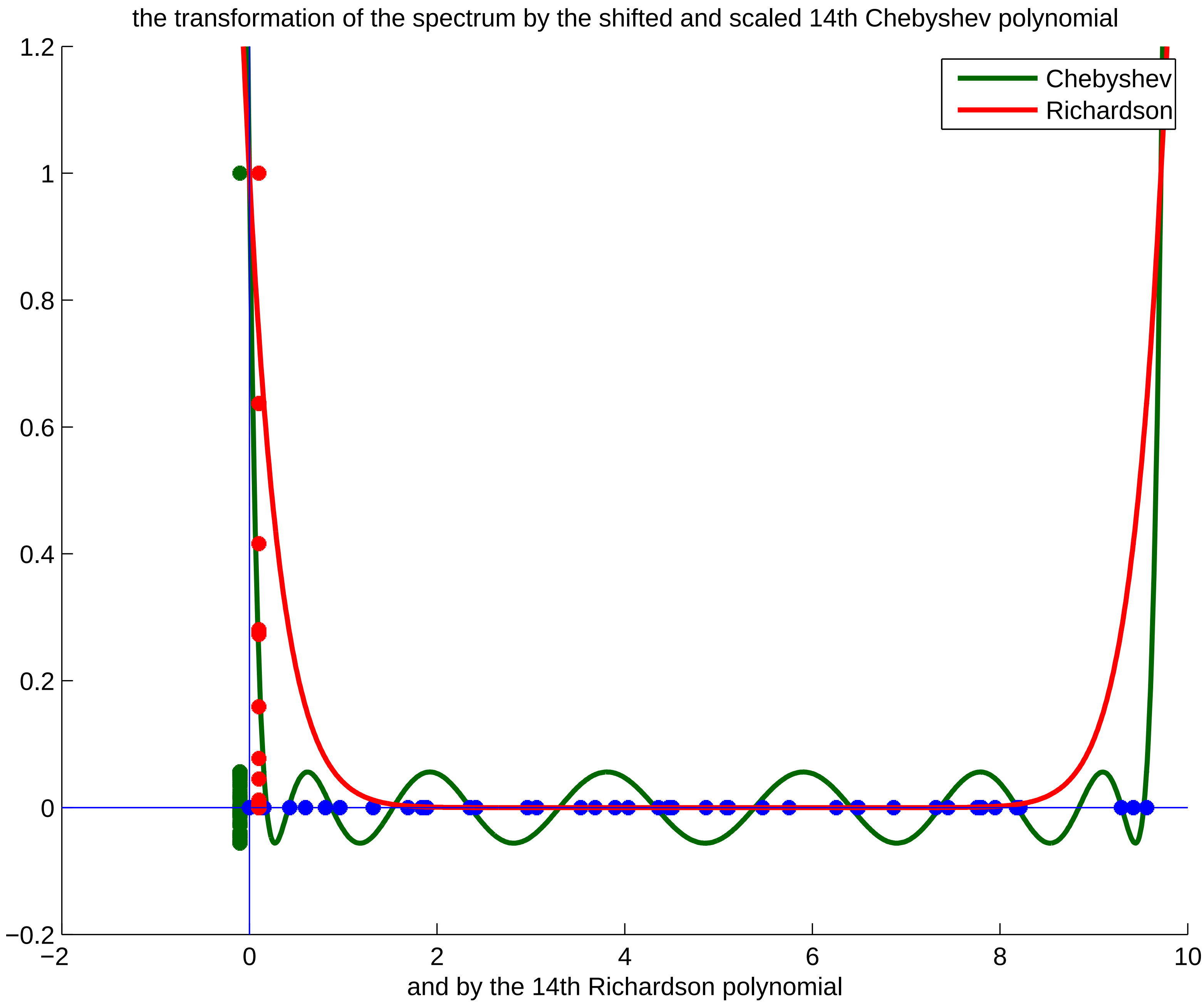
Improvements are based on the fact that

$$f(\mathbf{A})\mathbf{v}_j = f(\lambda_j)\mathbf{v}_j.$$

Examples. $f(\mathbf{A}) = (\mathbf{I} - \alpha \mathbf{A})$

$f(\mathbf{A}) = \mathbf{I} + \gamma_1 \mathbf{A} + \ldots + \gamma_\ell \mathbf{A}^\ell = (\mathbf{I} - \alpha_1 \mathbf{A}) \ldots (\mathbf{I} - \alpha_\ell \mathbf{A})$

Diminish unwanted components



the transformation of the spectrum by the shifted and scaled 14th Chebyshev polynomial
Chebyshev
Richardson
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
−2
0
2
4
6
8
10
and by the 14th Richardson polynomial

$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ **Shifted power**: $\tilde{\mathbf{u}}_k = (\mathbf{I} - \alpha \mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \tilde{\mathbf{u}}_k / \mathbf{e}_1^* \tilde{\mathbf{u}}_k$

Improvements are based on the fact that

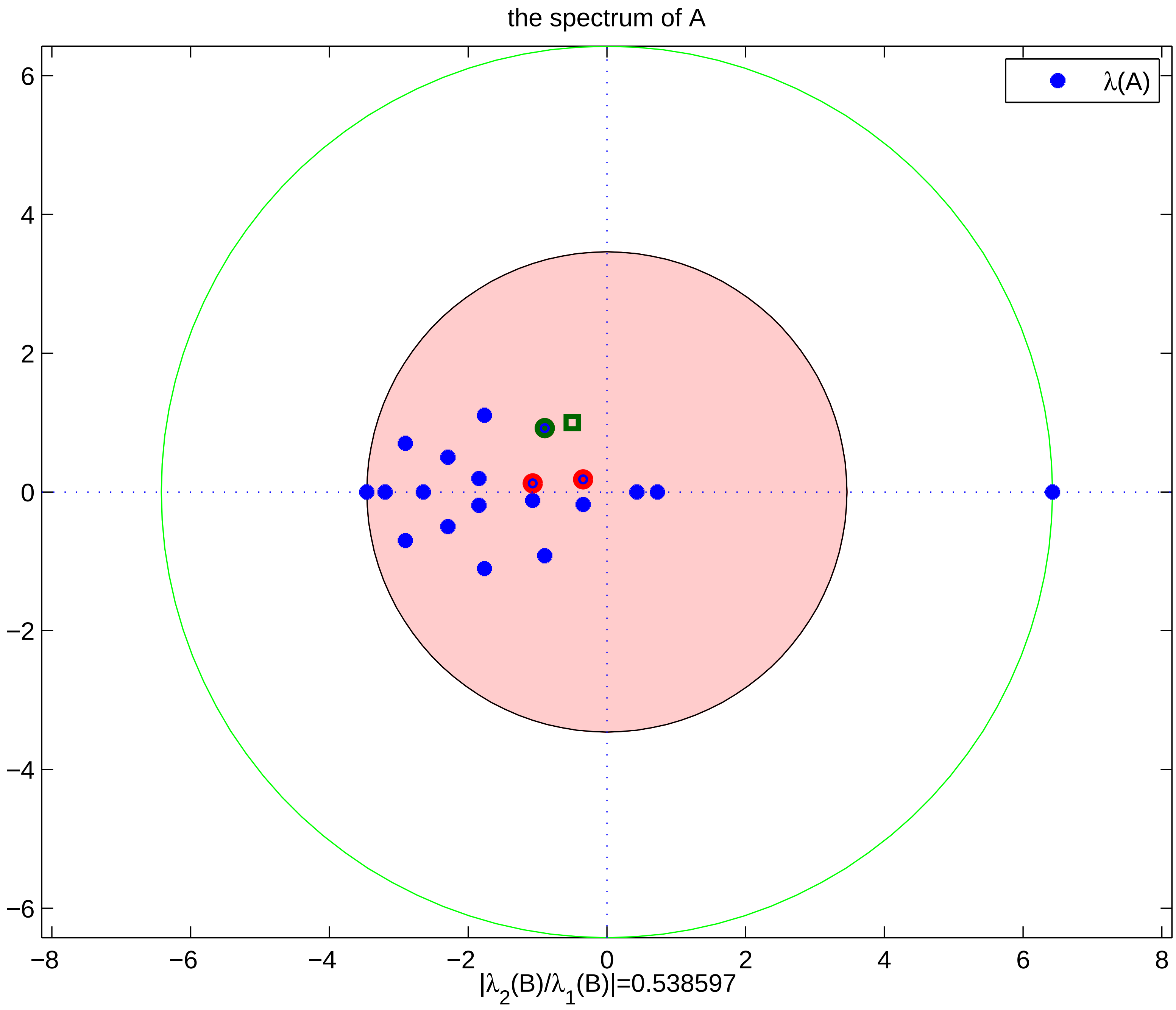
$$f(\mathbf{A})\mathbf{v}_j = f(\lambda_j)\mathbf{v}_j.$$

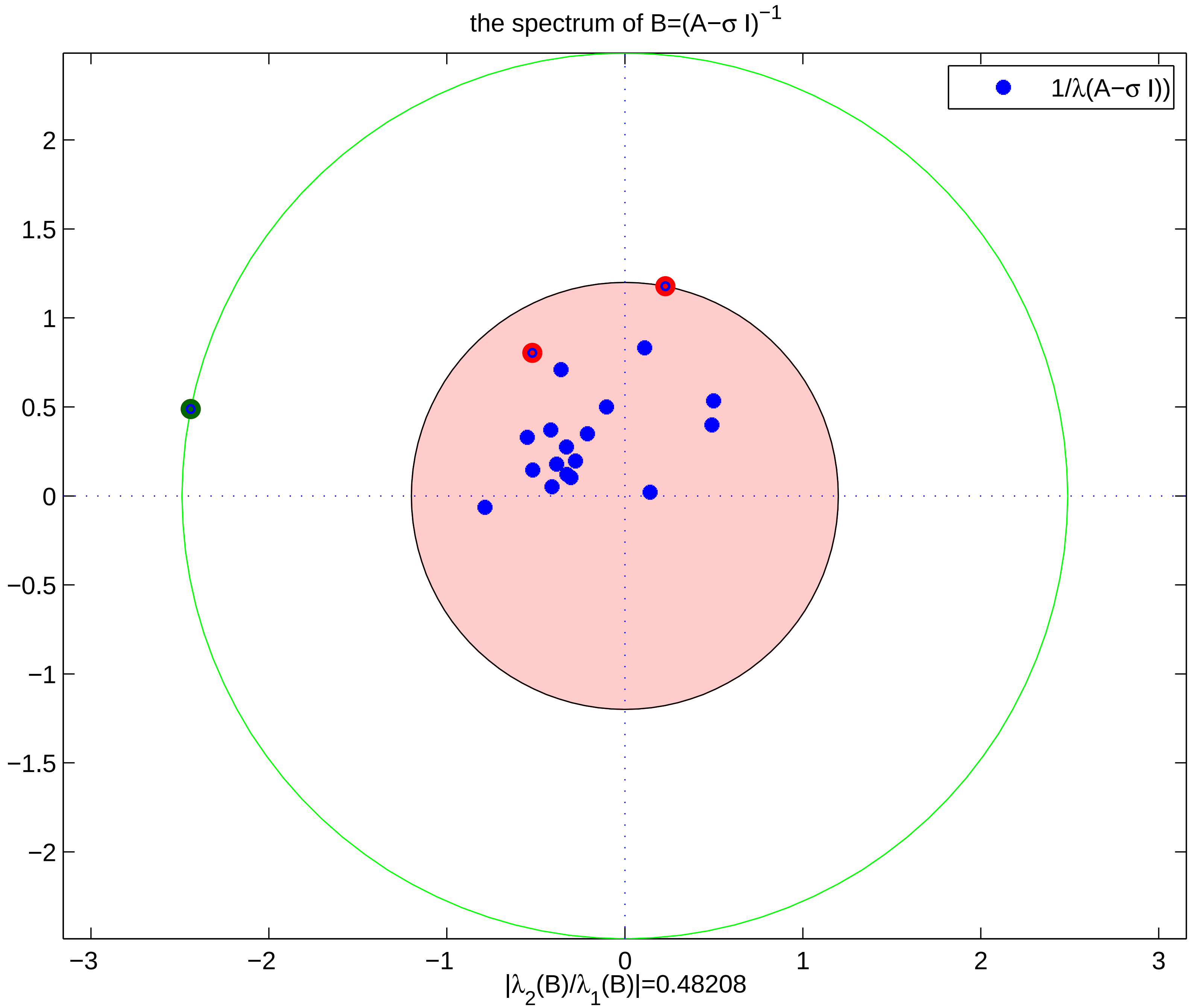
Examples. $f(\mathbf{A}) = (\mathbf{I} - \alpha \mathbf{A})$

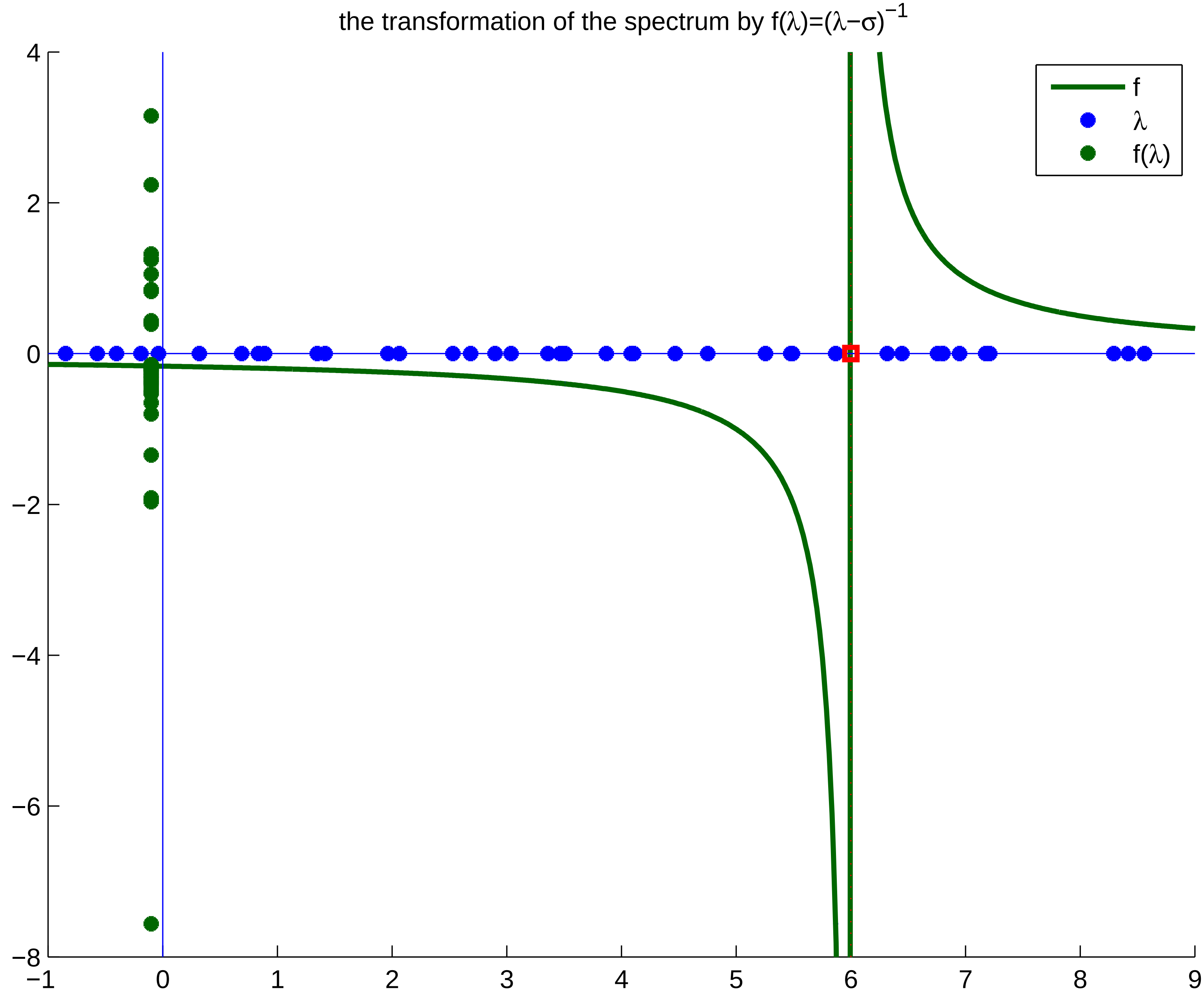
$f(\mathbf{A}) = \mathbf{I} + \gamma_1 \mathbf{A} + \ldots + \gamma_\ell \mathbf{A}^\ell = (\mathbf{I} - \alpha_1 \mathbf{A}) \ldots (\mathbf{I} - \alpha_\ell \mathbf{A})$

$f(\mathbf{A}) = (\mathbf{A} - \sigma \mathbf{I})^{-1}$

Amplify wanted components


the spectrum of A
λ(A)
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4
6
8
|λ2(B)/λ1(B)|=0.538597


the spectrum of B=(A−σ I)^{−1}
1/λ(A−σ I))
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−3
−2
−1
0
1
2
3
|λ_2(B)/λ_1(B)|=0.48208



the transformation of the spectrum by f(λ)=(λ−σ)^−1
f
λ
f(λ)
4
2
0
−2
−4
−6
−8
−1
0
1
2
3
4
5
6
7
8
9

$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ **Shifted power**: $\tilde{\mathbf{u}}_k = (\mathbf{I} - \alpha\mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \tilde{\mathbf{u}}_k / \mathbf{e}_1^* \tilde{\mathbf{u}}_k$

Improvements are based on the fact that

$$f(\mathbf{A})\mathbf{v}_j = f(\lambda_j)\mathbf{v}_j.$$

Examples. $f(\mathbf{A}) = (\mathbf{I} - \alpha\mathbf{A})$

$f(\mathbf{A}) = \mathbf{I} + \gamma_1\mathbf{A} + \ldots + \gamma_\ell\mathbf{A}^\ell = (\mathbf{I} - \alpha_1\mathbf{A})\ldots(\mathbf{I} - \alpha_\ell\mathbf{A})$

$f(\mathbf{A}) = (\mathbf{A} - \sigma\mathbf{I})^{-1}$

$f(\mathbf{A}) = \exp(-\alpha\mathbf{A}), \ldots$

$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ **Shifted power**: $\tilde{\mathbf{u}}_k = (\mathbf{I} - \alpha \mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \tilde{\mathbf{u}}_k / \mathbf{e}_1^* \tilde{\mathbf{u}}_k$

Improvements are based on the fact that

$$f(\mathbf{A})\mathbf{v}_j = f(\lambda_j)\mathbf{v}_j.$$

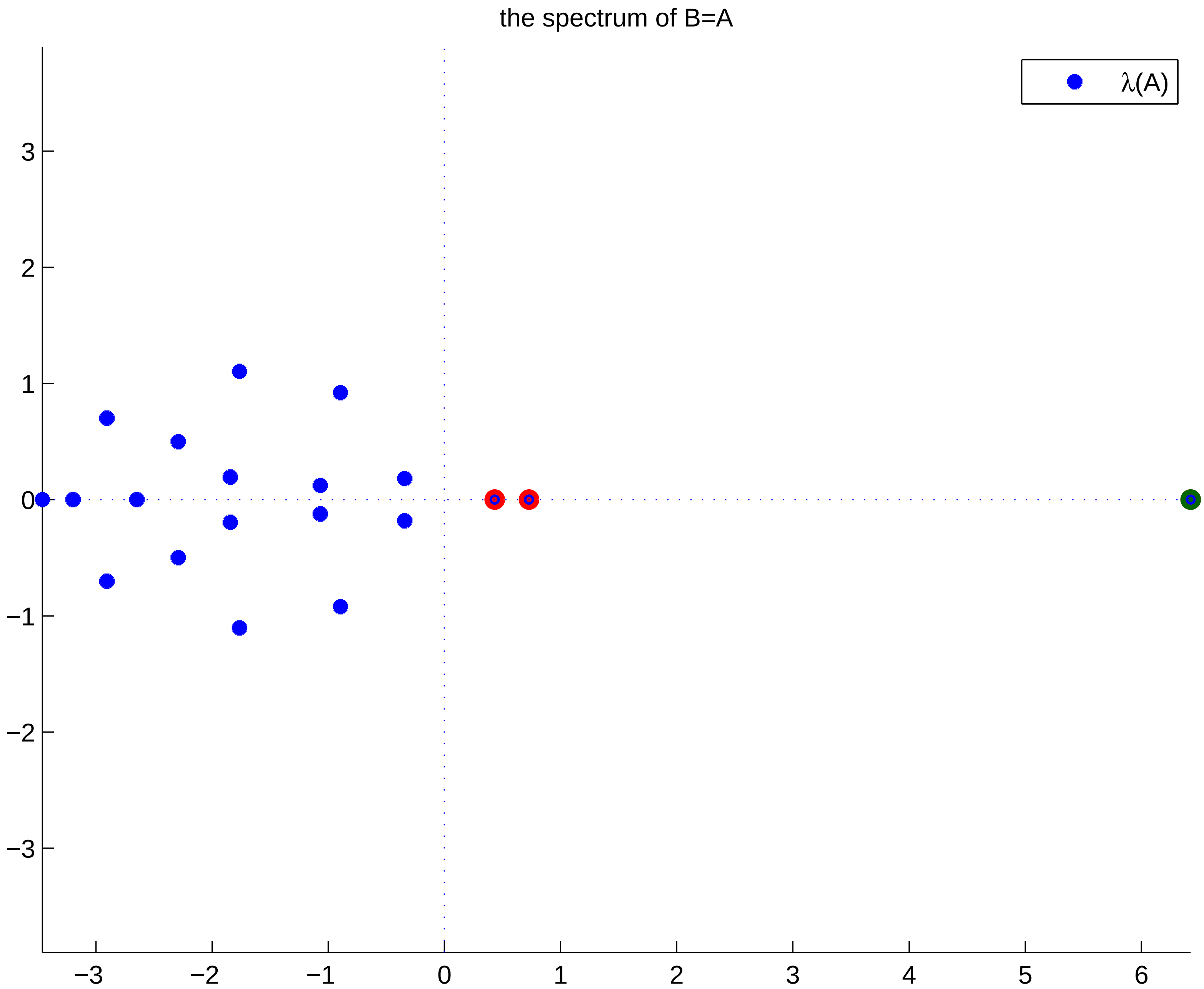
Examples. $f(\mathbf{A}) = (\mathbf{I} - \alpha \mathbf{A})$

$f(\mathbf{A}) = \mathbf{I} + \gamma_1 \mathbf{A} + \ldots + \gamma_\ell \mathbf{A}^\ell = (\mathbf{I} - \alpha_1 \mathbf{A}) \ldots (\mathbf{I} - \alpha_\ell \mathbf{A})$

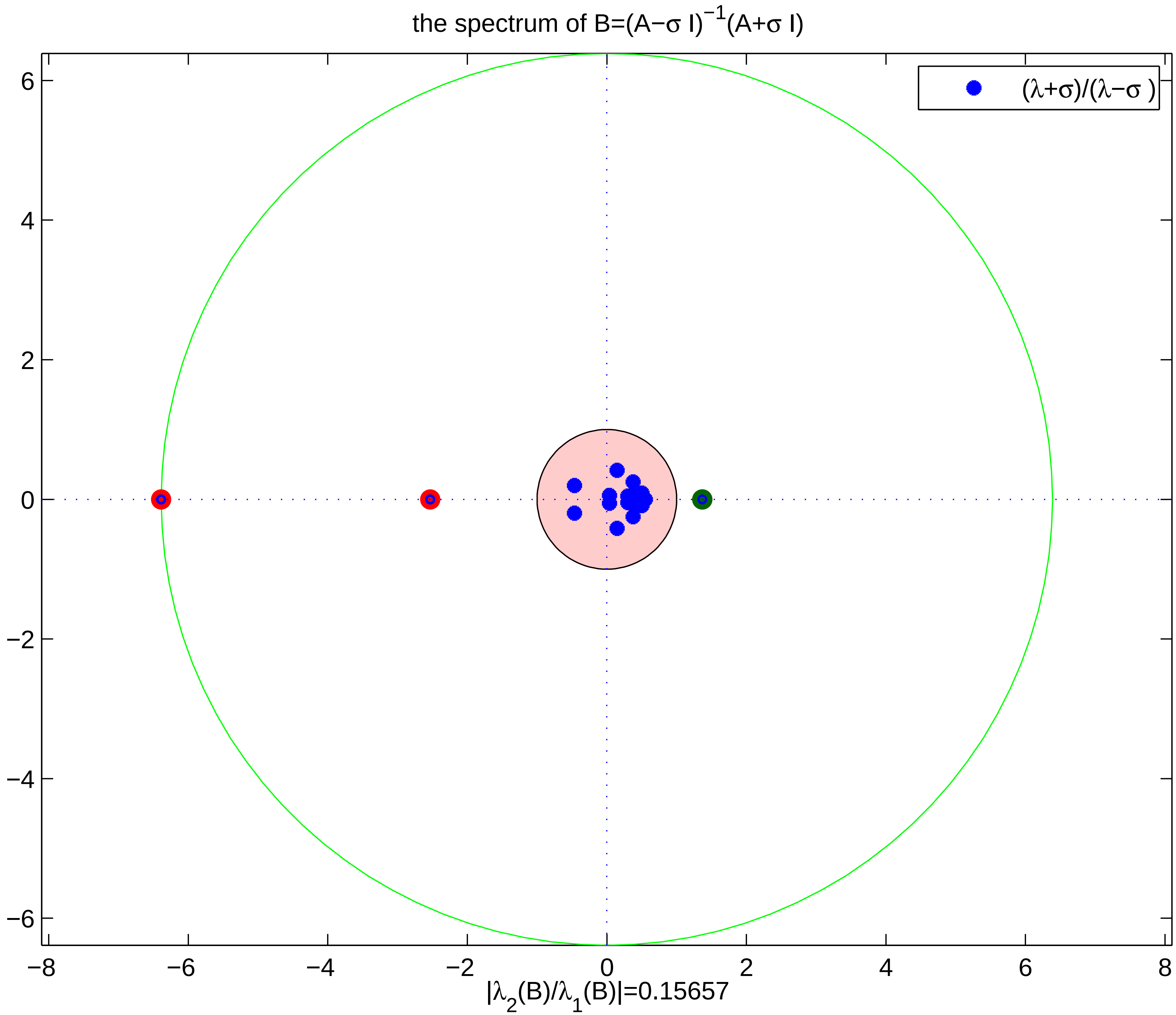
$f(\mathbf{A}) = (\mathbf{A} - \sigma \mathbf{I})^{-1}$

Combination. Cayley transform:

$$f(\mathbf{A}) = (\mathbf{A} - \mathbf{I})^{-1}(\mathbf{I} + \mathbf{A})$$



the spectrum of B=A
λ(A)
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3
4
5
6


the spectrum of B=(A−σ I)^−1(A+σ I)
(λ+σ)/(λ−σ)
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4
6
8
|λ2(B)/λ1(B)|=0.15657

$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ **Shifted power**: $\tilde{\mathbf{u}}_k = (\mathbf{I} - \alpha \mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \tilde{\mathbf{u}}_k / \mathbf{e}_1^* \tilde{\mathbf{u}}_k$

Improvements are based on the fact that

$$f(\mathbf{A})\mathbf{v}_j = f(\lambda_j)\mathbf{v}_j.$$

Examples. $f(\mathbf{A}) = (\mathbf{I} - \alpha \mathbf{A})$

$f(\mathbf{A}) = \mathbf{I} + \gamma_1 \mathbf{A} + \ldots + \gamma_\ell \mathbf{A}^\ell = (\mathbf{I} - \alpha_1 \mathbf{A}) \ldots (\mathbf{I} - \alpha_\ell \mathbf{A})$

$f(\mathbf{A}) = (\mathbf{A} - \sigma \mathbf{I})^{-1}$

Combination. **Cayley transform:**

$$f(\mathbf{A}) = (\mathbf{A} - \sigma \mathbf{I})^{-1}(\mathbf{I} - \alpha \mathbf{A})$$

$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ **Shifted power**: $\widetilde{\mathbf{u}}_k = (\mathbf{I} - \alpha \mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \widetilde{\mathbf{u}}_k / \mathbf{e}_1^* \widetilde{\mathbf{u}}_k$

Improvements. Apply power method with

$f(\mathbf{A}) = (\mathbf{I} - \alpha_1 \mathbf{A}) \ldots (\mathbf{I} - \alpha_\ell \mathbf{A})$ or $f(\mathbf{A}) = (\mathbf{A} - \sigma \mathbf{I})^{-1} (\mathbf{I} - \alpha \mathbf{A})$

Equivalent interpretations.

1. Diminish unwanted components. **Filtering.**
2. Amplify wanted components
3. Improve distribution eigenvalues. **Preconditioning.**

Program Lecture 5

- Power Method & Richardson
- Filtering
- Shift-and-Invert & Preconditioning
- Polynomial Iteration
- Selecting Parameters

  1) single parameter a) static
  b) dynamic

  2) Multiple parameters a) static (Chebyshev)
  b) dynamic (GCR)

Preconditioning

Purpose. To improve the distribution of the eigenvalues in order to speed up convergence.

For eigenvalue computation:

make the wanted eigenvector (strongly) dominant.

Shift & Invert can be a feasible strategy

For linear systems: cluster the eigenvalues round 1.

Precondition with a matrix $\mathbf{M}$ for which

- $\Lambda(\mathbf{M}^{-1}\mathbf{A})$ clusters 'better' round 1 than $\Lambda(\mathbf{A})$
- the system $\mathbf{M}\mathbf{u} = \mathbf{r}$ can efficiently be solved for $\mathbf{u}$.

For eigenvalue computation:
$\mathbf{A}$ and $\mathbf{M}^{-1}\mathbf{A}$ generally do not have the same eigenvectors.

Program Lecture 5

- Power Method & Richardson
- Filtering
- Shift-and-Invert & Preconditioning
- Polynomial Iteration
- Selecting Parameters

  1) single parameter a) static
  b) dynamic

  2) Multiple parameters a) static (Chebyshev)
  b) dynamic (GCR)

$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ **Shifted power**: $\widetilde{\mathbf{u}}_k = (\mathbf{I} - \alpha\mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \widetilde{\mathbf{u}}_k / \mathbf{e}_1^* \widetilde{\mathbf{u}}_k$

Improvements. Apply power method with

$f(\mathbf{A}) = (\mathbf{I} - \alpha_1\mathbf{A})\ldots(\mathbf{I} - \alpha_\ell\mathbf{A})$ or $f(\mathbf{A}) = (\mathbf{A} - \sigma\mathbf{I})^{-1}(\mathbf{I} - \alpha\mathbf{A})$

How to implement the shifts strategy

in a solver for $\mathbf{A}\mathbf{x} = \mathbf{b}$?

$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ **Shifted power**: $\quad \widetilde{\mathbf{u}}_k = (\mathbf{I} - \alpha \mathbf{A})\mathbf{u}_k$

Scale $\mathbf{u}_{k+1} = \widetilde{\mathbf{u}}_k / \mathbf{e}_1^* \widetilde{\mathbf{u}}_k$

Improvements. Apply power method with

$f(\mathbf{A}) = (\mathbf{I} - \alpha_1 \mathbf{A}) \ldots (\mathbf{I} - \alpha_\ell \mathbf{A})$ or $f(\mathbf{A}) = (\mathbf{A} - \sigma \mathbf{I})^{-1} (\mathbf{I} - \alpha \mathbf{A})$

$\mathbf{A}\mathbf{x} = \mathbf{b}$ **Richardson:** $\quad \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha(\mathbf{b} - \mathbf{A}\mathbf{x}_k)$

Polynomial version: Select α_k per step.

Purpose: Diminish all components ‘equally’ well.

Richardson (with relax. par.)

```
Select x_0, α, tol, k_max
Compute r_0 = b − A x_0
for k = 0, 1, 2, ..., k_max do
    If ||r|| ≤ tol, break, end if
    u_k = r_k
    c_k = A u_k
    x_{k+1} = x_k + α u_k
    r_{k+1} = r_k − α c_k
end do
```

$\mathbf{u}_k$ **search direction** (for the approximate)

Note. Update $\mathbf{r}_k$ of the form $\mathbf{A}\mathbf{u}_k$ with $\mathbf{u}_k$ update $\mathbf{x}_k$.

Richardson (with relax. par.)

```
Select x, α, tol, k_max
Compute r = b − Ax

for k = 0, 1, 2, ..., k_max do
    If ||r|| ≤ tol, break, end if
    u = r
    c = Au
    x ← x + αu
    r ← r − αc
end do
```

This is a 'memory friendly' version.

$\leftarrow$: new value replaces old one.

Polynomial iteration

```
Select x, α_1, ..., α_ℓ, tol, k_max
Compute r = b − Ax

for k = 0, 1, 2, ..., k_max do
    If ‖r‖ ≤ tol, break, end if
    u = r
    c = Au
    j = k mod ℓ, α = α_{j+1}
    x ← x + α u
    r ← r − α c
end do
```

Polynomial iteration

With $\alpha_1, \ldots, \alpha_\ell$ as in the polynomial iteration algorithm, let

$$f(\mathbf{A}) \equiv (\mathbf{I} - \alpha_1 \mathbf{A}) \ldots (\mathbf{I} - \alpha_\ell \mathbf{A}).$$

Then

Proposition. $\quad \mathbf{r}_{j\ell} = f(\mathbf{A})^j \, \mathbf{r}_0 \quad$ for all j.

For $k = j\ell$ large

$$\|\mathbf{r}_{k+\ell}\|_2 \approx \rho \, \|\mathbf{r}_k\|_2 \quad \text{with} \quad \rho \equiv \max\{|f(\lambda)| \mid \lambda \in \Lambda(\mathbf{A})\}$$

General remarks for linear systems.

- **The preconditioned system**.

For ease of discussion assume no preconditioning:
if preconditioner replace $\mathbf{A}$ by $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{b}$ by $\mathbf{M}^{-1}\mathbf{b}$.

- **Consistent updates.**

We update $\mathbf{r}$ and $\mathbf{x}$ consistently:
update $\mathbf{r}$ by vectors $-\mathbf{c}$ of the form $\mathbf{c} = \mathbf{A}\mathbf{u}$ with $\mathbf{u}$ explicitly avaliable and update $\mathbf{x}$ by $\mathbf{u}$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{u}_k, \qquad \mathbf{c}_k = \mathbf{A}\mathbf{u}_k, \qquad \mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{c}_k$$

- **The shifted system**.

Assume $\mathbf{x}_0 = \mathbf{0}$.
If $\mathbf{x}_0 \neq \mathbf{0}$, solve $\mathbf{A}\mathbf{x} = \mathbf{r}_0 \equiv \mathbf{b} - \mathbf{A}\mathbf{x}_0$.

Program Lecture 5

- Power Method & Richardson
- Filtering
- Shift-and-Invert & Preconditioning
- Polynomial Iteration
- Selecting Parameters

  1) single parameter a) static
  b) dynamic

  2) Multiple parameters a) static (Chebyshev)
  b) dynamic (GCR)

How to select the α_j and σ?

Static.
Select parameter(s) before starting the iteration.
Base selection on pre-knowledge of the spectrum.

Dynamic.
Let the computational process determine the parameter(s).
Computation based on information that becomes available during the iteration.

Static. Single parameter

Examples. $\mathbf{A}\mathbf{v}_0 = \lambda_0 \mathbf{v}_0$, $\lambda_0 \in \Lambda(\mathbf{A})$ wanted eigenvalue.

- If $|\lambda_0 - \mu| > |\lambda - \mu|$ for all other $\lambda \in \Lambda(\mathbf{A})$:

$$f(\mathbf{A}) = \mathbf{A} - \mu \mathbf{I}.$$

Shifted power method.

- If λ_0 closest to some target value τ is wanted:

$$f(\mathbf{A}) = (\mathbf{A} - \sigma \mathbf{I})^{-1} \text{ with } \sigma = \tau.$$

Inverse iteration or **Wielandt iteration**.

Static. Single parameter

Examples. $\mathbf{Ax} = \mathbf{b}$.

- If all λ_j eigenvalues $\mathbf{A}$ in $[\lambda_-, \lambda_+] = [\mu - \rho, \mu + \rho] \subset (0, \infty)$:

$\mu = (\lambda_+ + \lambda_-)/2, \qquad \rho = (\lambda_+ - \lambda_-)/2.$

$f(\mathbf{A}) = \mathbf{I} - \alpha_{\text{opt}} \mathbf{A}$ with $\alpha_{\text{opt}} \equiv 1/\mu$,

$$\max |f(\lambda_j)| \leq \frac{\lambda_+ - \lambda_-}{\lambda_+ + \lambda_-} = \frac{1 - \frac{1}{\mathcal{C}}}{1 + \frac{1}{\mathcal{C}}} \leq e^{-\frac{2}{\mathcal{C}}}, \quad \text{where} \quad \mathcal{C} \equiv \frac{\lambda_+}{\lambda_-}$$

Therefore, for Richardson with $\alpha = \alpha_{\text{opt}}$,

$$\|\mathbf{r}_{k+1}^{\text{Rich}}\| \lesssim \exp\left(-\frac{2}{\mathcal{C}}\right) \|\mathbf{r}_k^{\text{Rich}}\| \qquad k \text{ large}.$$

Dynamic. Single parameter

Examples. $\mathbf{A}\mathbf{v}_0 = \lambda_0\mathbf{v}_0$, $\lambda_0 \in \Lambda(\mathbf{A})$ wanted eigenvalue.

- $f(\mathbf{A}) = (\mathbf{A} - \sigma\mathbf{I})^{-1}$, with $\sigma = \sigma_k = \rho(\mathbf{u}_k) \equiv \dfrac{\mathbf{u}_k^*\mathbf{A}\mathbf{u}_k}{\mathbf{u}_k^*\mathbf{u}_k}$.

Rayleigh Quotient Iteration

The Rayleigh quotient $\rho(\mathbf{u}_k)$ is the 'best' available approximate eigenvalue at step k.

If RQI converges, it converges quadratically eventually.
For Hermitian $\mathbf{A}$, the asymptotic convergence is even cubic.

"If converges": **Example.** $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. $\mathbf{v}_0 = e_1$.

RQI:

+ Fast convergence (if convergence).
+ Can detect eigenvalues in the interior of the spectrum.
− No controle on what eigenvalue is going to be detected.
  Remedy: First a few steps of Wielandt iteration.
− The linear systems to be solved require
  a new LU-decompostion in each step.

Wielandt Iteration:

− Linear convergence.
+ Can detect eigenvalues in the interior of the spectrum.
+ Finds eigenvalue close to the shift.
+ The same LU-decomposition can used in each step.

Note. The fact that linear systems have to be solved may make the methods not feasible for huge n.

Dynamic. Single parameter

Examples. $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Select $f(\mathbf{A}) = \mathbf{I} - \alpha_k \mathbf{A}$ with α_k to minimize:

- **Minimal Residual**: $\|\mathbf{r}_{k+1}\|_2 = \|\mathbf{r}_k - \alpha_k \mathbf{c}_k\|_2$ minimal
- If $\mathbf{A}$ is positive definite

  Steepest descent: $\|\mathbf{x} - \mathbf{x}_{k+1}\|_A$ minimal

Convergence if $\mathrm{Re}(\lambda_j) > 0$ for all eigenvalues λ_j of $\mathbf{A}$.

Local Minimal Residuals

```
Select x, α, tol, k_max
Compute r = b − Ax

for k = 0, 1, 2, ..., k_max do
    If ‖r‖ ≤ tol, break, end if
    u = r
    c = Au
    σ = c*c,  ρ = c*r,  α = ρ/σ
    x ← x + αu
    r ← r − αc
end do
```

Steepest Descent

Select $\mathbf{x}$, α, tol, $k_{\max}$

Compute $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$

for $k = 0, 1, 2, \ldots, k_{\max}$ do

  If $\|\mathbf{r}\| \leq tol$, break, end if

  $\mathbf{u} = \mathbf{r}$

  $\mathbf{c} = \mathbf{A}\mathbf{u}$

  $\sigma = \mathbf{u}^*\mathbf{c}$, $\rho = \mathbf{u}^*\mathbf{r}$, $\alpha = \rho/\sigma$

  $\mathbf{x} \leftarrow \mathbf{x} + \alpha\mathbf{u}$

  $\mathbf{r} \leftarrow \mathbf{r} - \alpha\mathbf{c}$

end do

Program Lecture 5

- Power Method & Richardson
- Filtering
- Shift-and-Invert & Preconditioning
- Polynomial Iteration
- Selecting Parameters

  1) single parameter a) static
  b) dynamic

  2) Multiple parameters a) static (Chebyshev)
  b) dynamic (GCR)

Static. Multiple parameter

Examples. $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Suppose we have a set $\mathcal{E} \subset \mathbb{C}$ that contains all λ_i.

Select $f(\mathbf{A}) = (\mathbf{I} - \alpha_1 \mathbf{A}) \cdot \ldots \cdot (\mathbf{I} - \alpha_\ell \mathbf{A})$, i.e., α_j, such that

$$\nu \equiv \max\{|f(\zeta)| = |(1 - \alpha_1 \zeta) \cdot \ldots \cdot (1 - \alpha_\ell \zeta)| \mid \zeta \in \mathcal{E}\}$$

is as small as possible. If $\nu < 1$ then one application of f (i.e., one sweep of ℓ steps) gives an error (and residual) reduction by at least a factor ν: for $k = j\ell$, large

$$\|\mathbf{x} - \mathbf{x}_{k+\ell}\|_2 \lesssim \nu \|\mathbf{x} - \mathbf{x}_k\|_2, \qquad \|\mathbf{r}_{k+\ell}\|_2 \lesssim \nu \|\mathbf{r}_k\|_2$$

Note. $\max\{|f(\lambda)| \mid \lambda \in \Lambda(\mathbf{A})\} \leq \nu$

Static. Multiple parameter

Examples. $\mathbf{Ax} = \mathbf{b}$.

Suppose we have a set $\mathcal{E} \subset \mathbb{C}$ that contains all λ_i.

Select $f(\mathbf{A}) = (\mathbf{I} - \alpha_1 \mathbf{A}) \cdot \ldots \cdot (\mathbf{I} - \alpha_\ell \mathbf{A})$, i.e., α_j, such that

$$\nu \equiv \max\{|f(\zeta)| = |(1 - \alpha_1 \zeta) \cdot \ldots \cdot (1 - \alpha_\ell \zeta)| \mid \zeta \in \mathcal{E}\}$$

is as small as possible.

Notation.

$\mathcal{P}_\ell$ is the set of all polynomials of degree at most ℓ.

$$\mathcal{P}_\ell^0 \equiv \{p \in \mathcal{P}_\ell \mid p(0) = 1\}$$

Observation. $p \in \mathcal{P}_\ell$

$$p(0) = 1 \qquad \Leftrightarrow \qquad p(\zeta) = (1 - \alpha_1 \zeta) \cdot \ldots \cdot (1 - \alpha_\ell \zeta).$$

Static. Multiple parameter

Examples. $\mathbf{Ax} = \mathbf{b}$.
Suppose we have a set $\mathcal{E} \subset \mathbb{C}$ that contains all λ_i.

Select $f(\mathbf{A}) = (\mathbf{I} - \alpha_1 \mathbf{A}) \cdot \ldots \cdot (\mathbf{I} - \alpha_\ell \mathbf{A})$, i.e., α_j, such that

$$\nu \equiv \max\{|f(\zeta)| = |(1 - \alpha_1 \zeta) \cdot \ldots \cdot (1 - \alpha_\ell \zeta)| \mid \zeta \in \mathcal{E}\}$$

is as small as possible.

Notation.
$\mathcal{P}_\ell$ is the set of all polynomials of degree at most ℓ.

$$\mathcal{P}_\ell^0 \equiv \{p \in \mathcal{P}_\ell \mid p(0) = 1\}$$

Observation. $p \in \mathcal{P}_\ell$

$$p(0) = 1 \quad \Leftrightarrow \quad p(\zeta) = 1 - \zeta q(\zeta) \text{ for some } q \in \mathcal{P}_{\ell-1}.$$

Static. Multiple parameter

Examples. $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Suppose we have a set $\mathcal{E} \subset \mathbb{C}$ that contains all λ_i.

Select $f(\mathbf{A}) = (\mathbf{I} - \alpha_1 \mathbf{A}) \cdot \ldots \cdot (\mathbf{I} - \alpha_\ell \mathbf{A})$, i.e., α_j, such that

$$\nu \equiv \max\{|f(\zeta)| = |(1 - \alpha_1 \zeta) \cdot \ldots \cdot (1 - \alpha_\ell \zeta)| \mid \zeta \in \mathcal{E}\}$$

is as small as possible.

Observation. Consider a vector $\mathbf{r}$ for which

$$\mathbf{r} = p(\mathbf{A})\mathbf{b} \quad \text{for some} \quad p \in \mathcal{P}_\ell$$

$\mathbf{r}$ is a residual $\Leftrightarrow$ $p(0) = 1$, i.e., $p(\zeta) = 1 - \zeta q(\zeta)$:

$$\mathbf{r} = \mathbf{b} - \mathbf{A}\tilde{\mathbf{x}} \quad \text{with} \quad \tilde{\mathbf{x}} = q(\mathbf{A})\mathbf{b}.$$

Polynomials in $\mathcal{P}_\ell^0$ are called **residual polynomials**.

Static. Multiple parameter

Examples. $\mathbf{Ax} = \mathbf{b}$.
Suppose we have a set $\mathcal{E} \subset \mathbb{C}$ that contains all λ_i.

Select $f(\mathbf{A}) = (\mathbf{I} - \alpha_1 \mathbf{A}) \cdot \ldots \cdot (\mathbf{I} - \alpha_\ell \mathbf{A})$, i.e., α_j, such that

$$\nu \equiv \max\{|f(\zeta)| = |(1 - \alpha_1 \zeta) \cdot \ldots \cdot (1 - \alpha_\ell \zeta)| \mid \zeta \in \mathcal{E}\}$$

is as small as possible.

This is a problem from approximation theory:

Find a polynomial in $\mathcal{P}_\ell^0$ that is as small as possible on $\mathcal{E}$.

Solutions for $\mathcal{E} = [\lambda_-, \lambda_+] \subset (0, \infty)$ (**Chebyshev pols**)

Approximate solutions for ellipses (Cheb.), polygons (Faber pols).

Static. Multiple parameter

Examples. $\mathbf{Av} = \lambda\mathbf{v}$.

Suppose we have a set $\mathcal{E} \subset \mathbb{C}$ that contains all λ_i, except for the wanted eigenvalue $\lambda_0 \in \Lambda(\mathbf{A})$.

Select $f(\mathbf{A}) = (\mathbf{I}-\alpha_1\mathbf{A})\cdot\ldots\cdot(\mathbf{I}-\alpha_\ell\mathbf{A})$ such that with

$$\nu \equiv \max_{\zeta\in\mathcal{E}} |(1-\alpha_1\zeta)\cdot\ldots\cdot(1-\alpha_\ell\zeta)|$$

$\nu/|f(\lambda_0)|$ is as small as possible.

Chebyshev polynomials

$$T_\ell(x) \equiv \frac{1}{2}(\zeta^\ell + \zeta^{-\ell}), \quad \text{where} \quad x \equiv \frac{1}{2}(\zeta + \zeta^{-1}) \quad (\zeta \in \mathbb{C}).$$

Exercise. For all $x \in \mathbb{C}$ we have

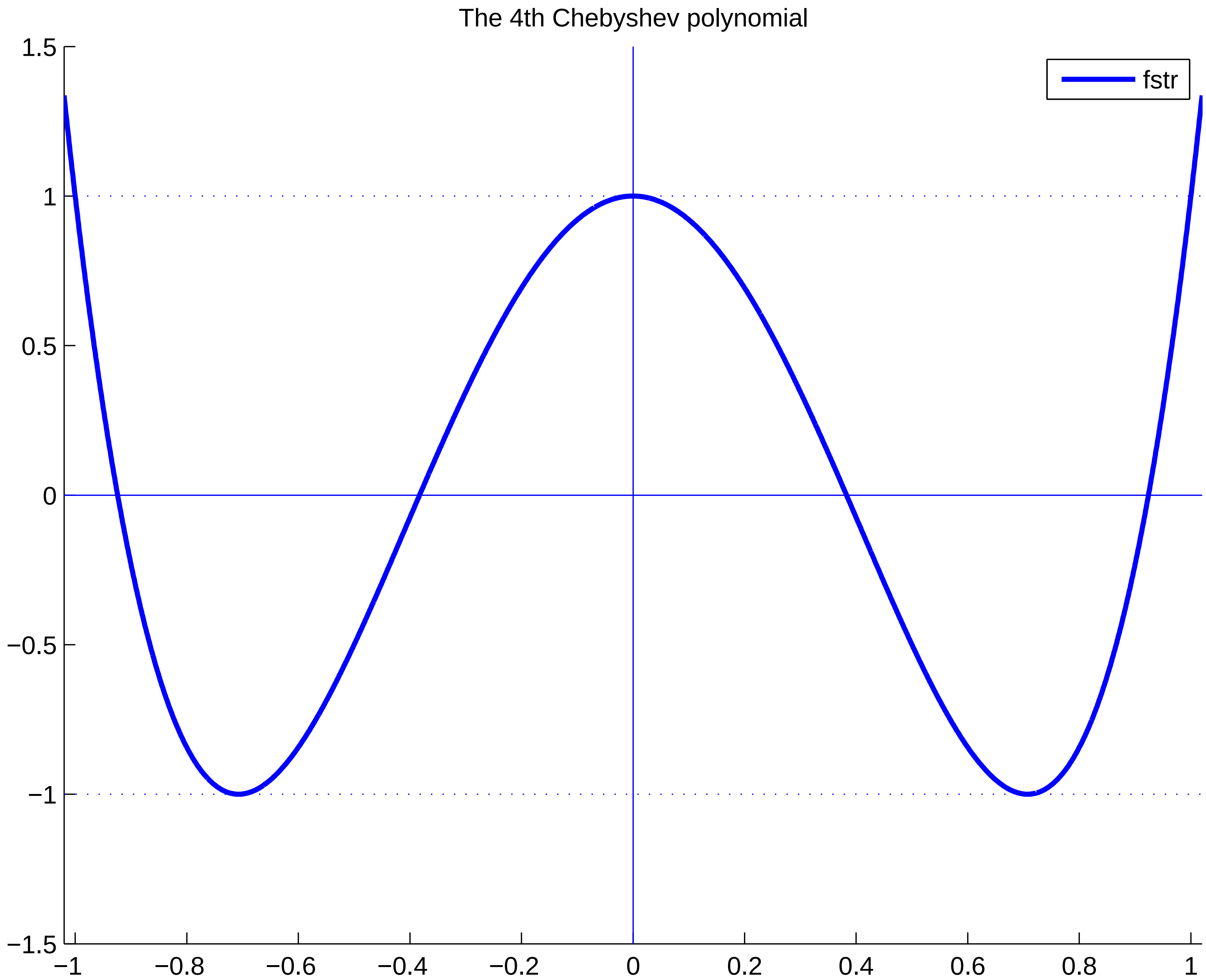
$$\begin{cases} T_0(x) = 1, \quad T_1(x) = x, \\ T_{k+1}(x) = 2\,x\,T_k(x) - T_{k-1}(x) \ \text{ for } \ k = 1, 2, \ldots. \end{cases}$$

$T_k(x) = 2^{k-1}x^k +$ lower degree terms.

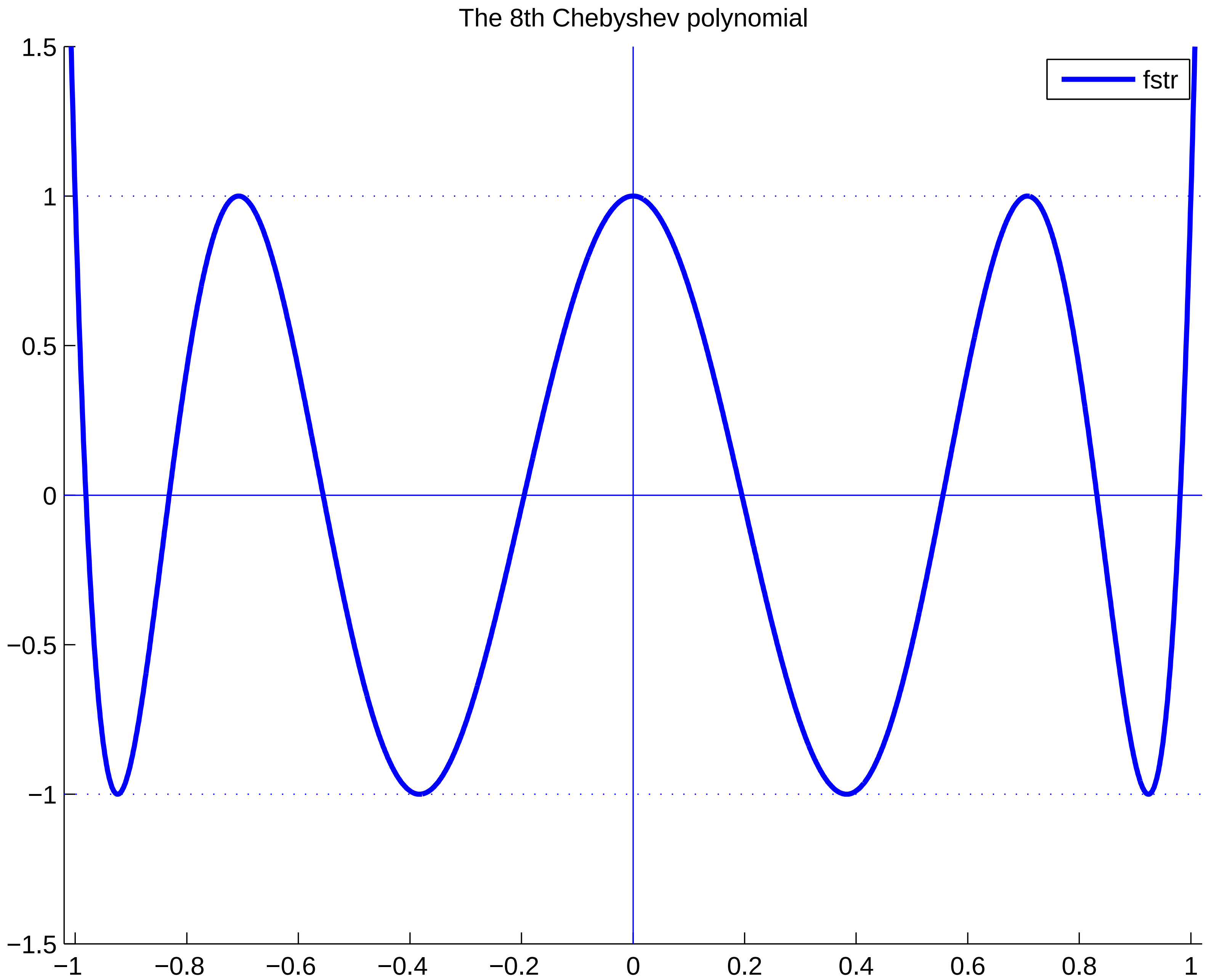
T_k is the kth **Chebyshev polynomial**

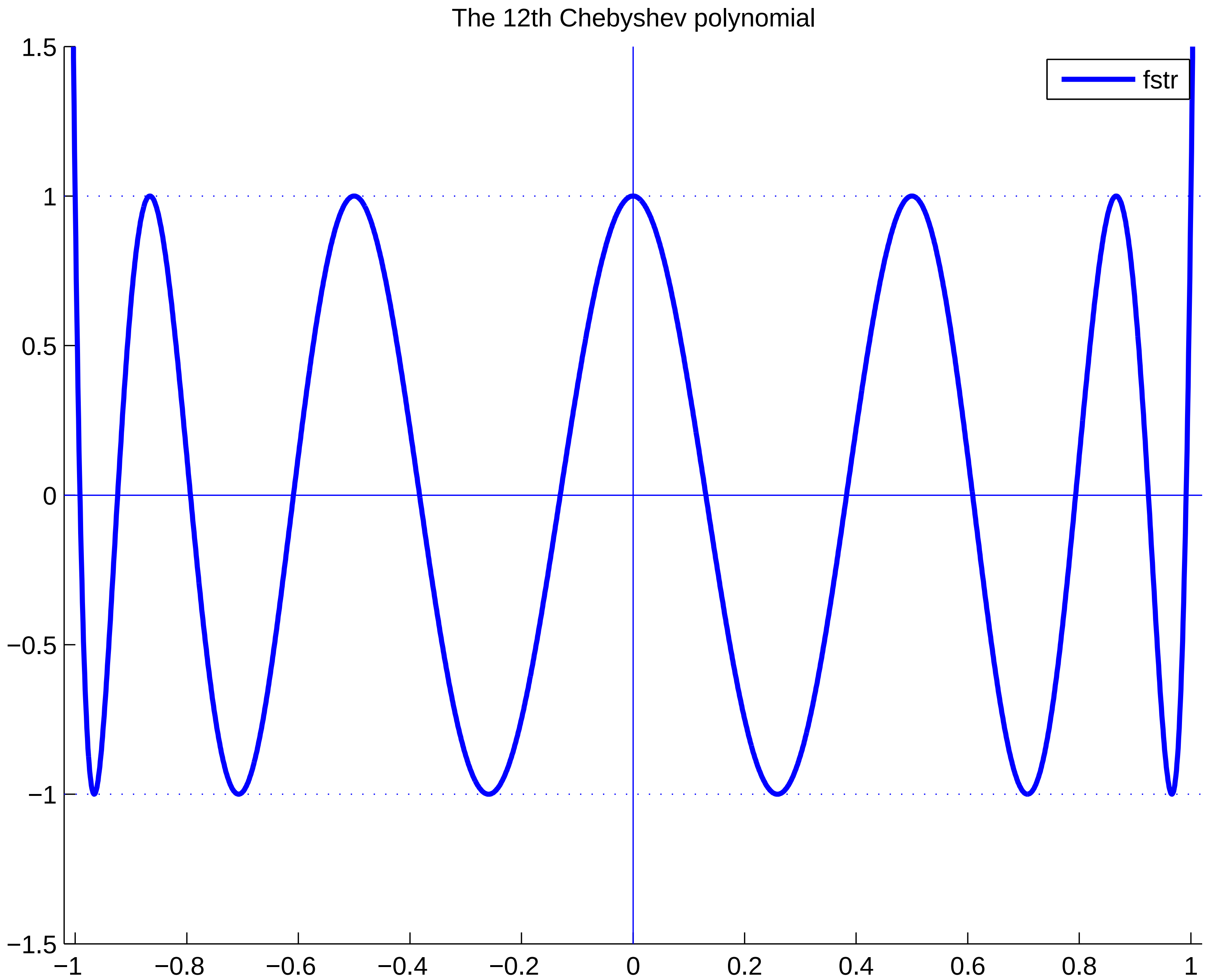
$T_k(\cos(\phi)) = \cos(k\phi) \qquad (\phi \in \mathbb{R}))$Regel With $\zeta = e^{i\phi}$ we have that $x = \cos(\phi) \in [-1, 1]$.

Note that $x = \dfrac{1+\delta}{1-\delta}$ iff $\zeta = \dfrac{1+\sqrt{\delta}}{1-\sqrt{\delta}} \qquad (\delta > 0).$



The 4th Chebyshev polynomial
fstr
1.5
1
0.5
0
−0.5
−1
−1.5
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1


The 8th Chebyshev polynomial
fstr
1.5
1
0.5
0
−0.5
−1
−1.5
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1


The 12th Chebyshev polynomial
fstr
1.5
1
0.5
0
−0.5
−1
−1.5
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

Chebyshev polynomials

$$T_\ell(x) \equiv \frac{1}{2}(\zeta^\ell + \zeta^{-\ell}), \quad \text{where} \quad x \equiv \frac{1}{2}(\zeta + \zeta^{-1}) \quad (\zeta \in \mathbb{C}).$$

Exercise. For all $x \in \mathbb{C}$ we have

$$\begin{cases} T_0(x) = 1, \quad T_1(x) = x, \\ T_{k+1}(x) = 2\,x\,T_k(x) - T_{k-1}(x) \ \text{ for } \ k = 1, 2, \ldots. \end{cases}$$

Properties. At the interval $[-1, +1]$,

• the kth Chebyshev polynomial takes its extremal values at $k+1$ points,

• these extremal values have alternating signs with values equal to $+1$ and -1.

Chebyshev polynomials

$$T_\ell(x) \equiv \frac{1}{2}(\zeta^\ell + \zeta^{-\ell}), \quad \text{where} \quad x \equiv \frac{1}{2}(\zeta + \zeta^{-1}) \quad (\zeta \in \mathbb{C}).$$

Exercise. For all $x \in \mathbb{C}$ we have

$$\begin{cases} T_0(x) = 1, \quad T_1(x) = x, \\ T_{k+1}(x) = 2\,x\,T_k(x) - T_{k-1}(x) \ \text{ for } \ k = 1, 2, \ldots. \end{cases}$$

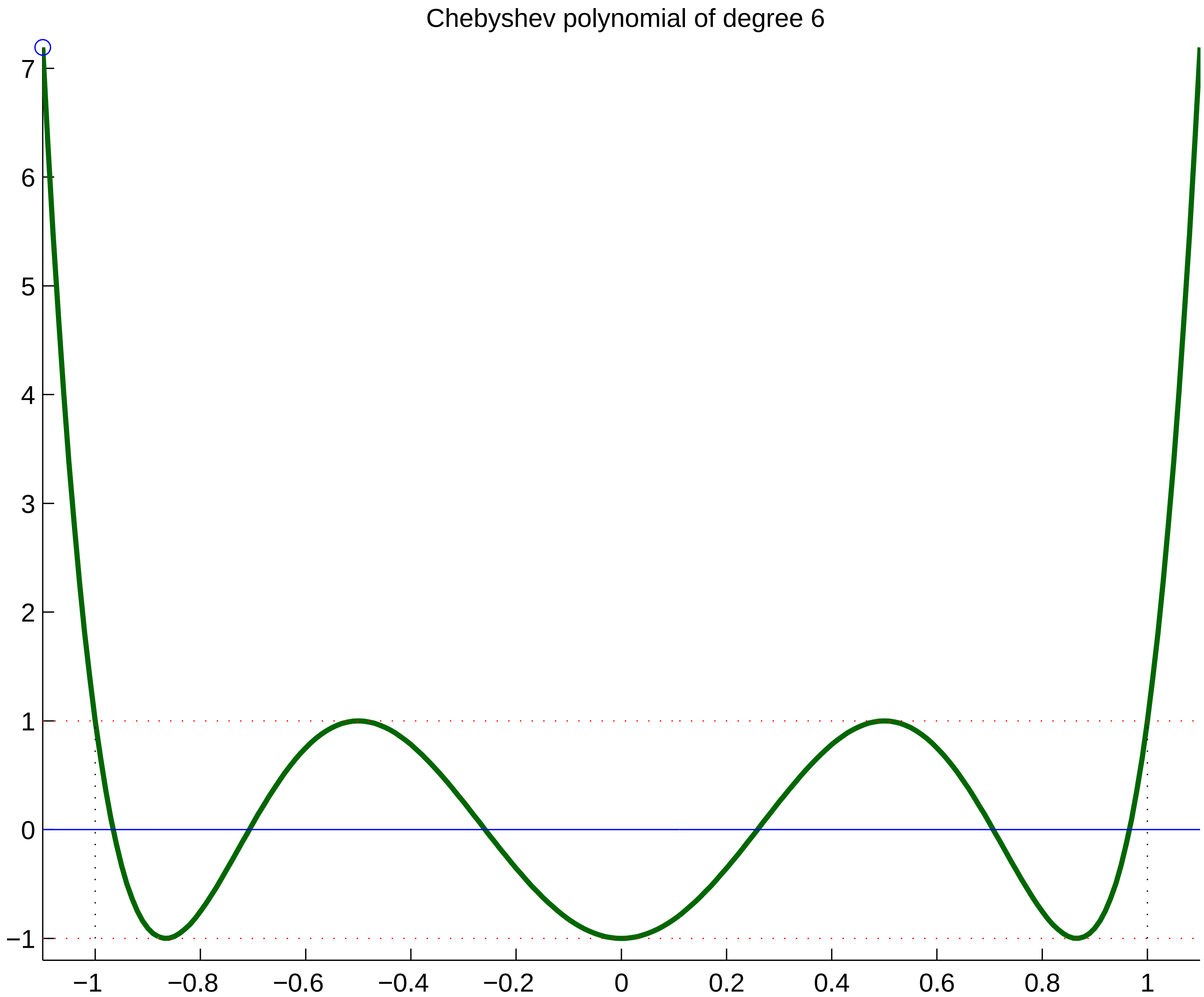
Assume $[\lambda_-, \lambda_+] = [\mu - \rho, \mu + \rho] \subset (0, \infty)$.

Theorem. With $x \equiv (\mu - \lambda)/\rho$

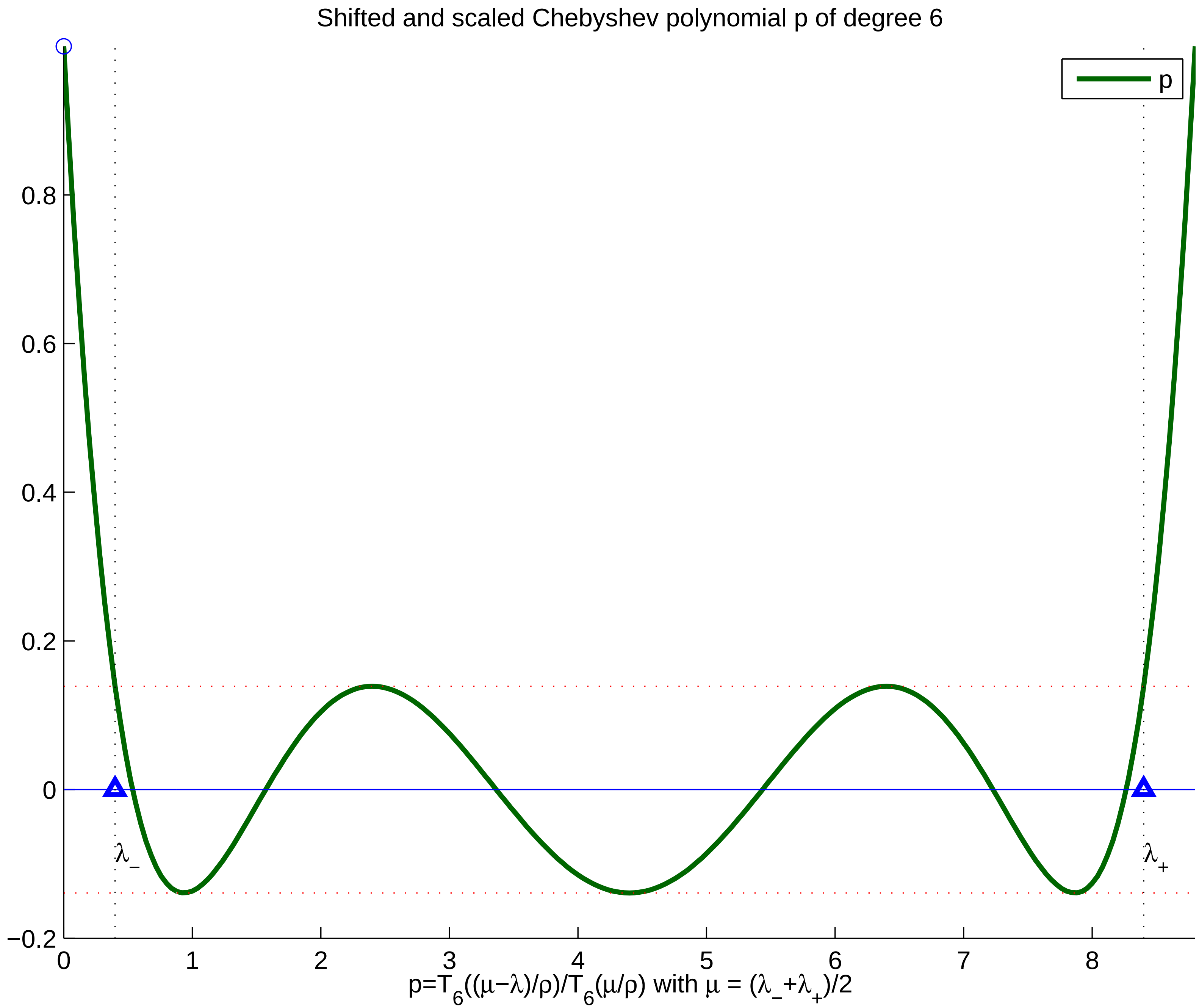
x runs between -1 and $+1$ iff λ runs between λ_+ and λ_-.

x depends linearly on λ $\quad\quad$ $\lambda \rightsquigarrow x$ preserves degree pol.

At $\lambda = 0$, x equals $\dfrac{\lambda_+ + \lambda_-}{\lambda_+ - \lambda_-} = \dfrac{1 + 1/\mathcal{C}}{1 - 1/\mathcal{C}}$, where $\mathcal{C} \equiv \frac{\lambda_+}{\lambda_-}$



Chebyshev polynomial of degree 6
7
6
5
4
3
2
1
0
−1
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1



Shifted and scaled Chebyshev polynomial p of degree 6
p
0.8
0.6
0.4
0.2
0
−0.2
0
1
2
3
4
5
6
7
8
λ_−
λ_+
p=T_6((μ−λ)/ρ)/T_6(μ/ρ) with μ = (λ_−+λ_+)/2

Chebyshev polynomials

$$T_\ell(x) \equiv \frac{1}{2}(\zeta^\ell + \zeta^{-\ell}), \quad \text{where} \quad x \equiv \frac{1}{2}(\zeta + \zeta^{-1}) \quad (\zeta \in \mathbb{C}).$$

Exercise. For all $x \in \mathbb{C}$ we have

$$\begin{cases} T_0(x) = 1, \quad T_1(x) = x, \\ T_{k+1}(x) = 2\,x\,T_k(x) - T_{k-1}(x) \text{ for } k = 1, 2, \ldots. \end{cases}$$

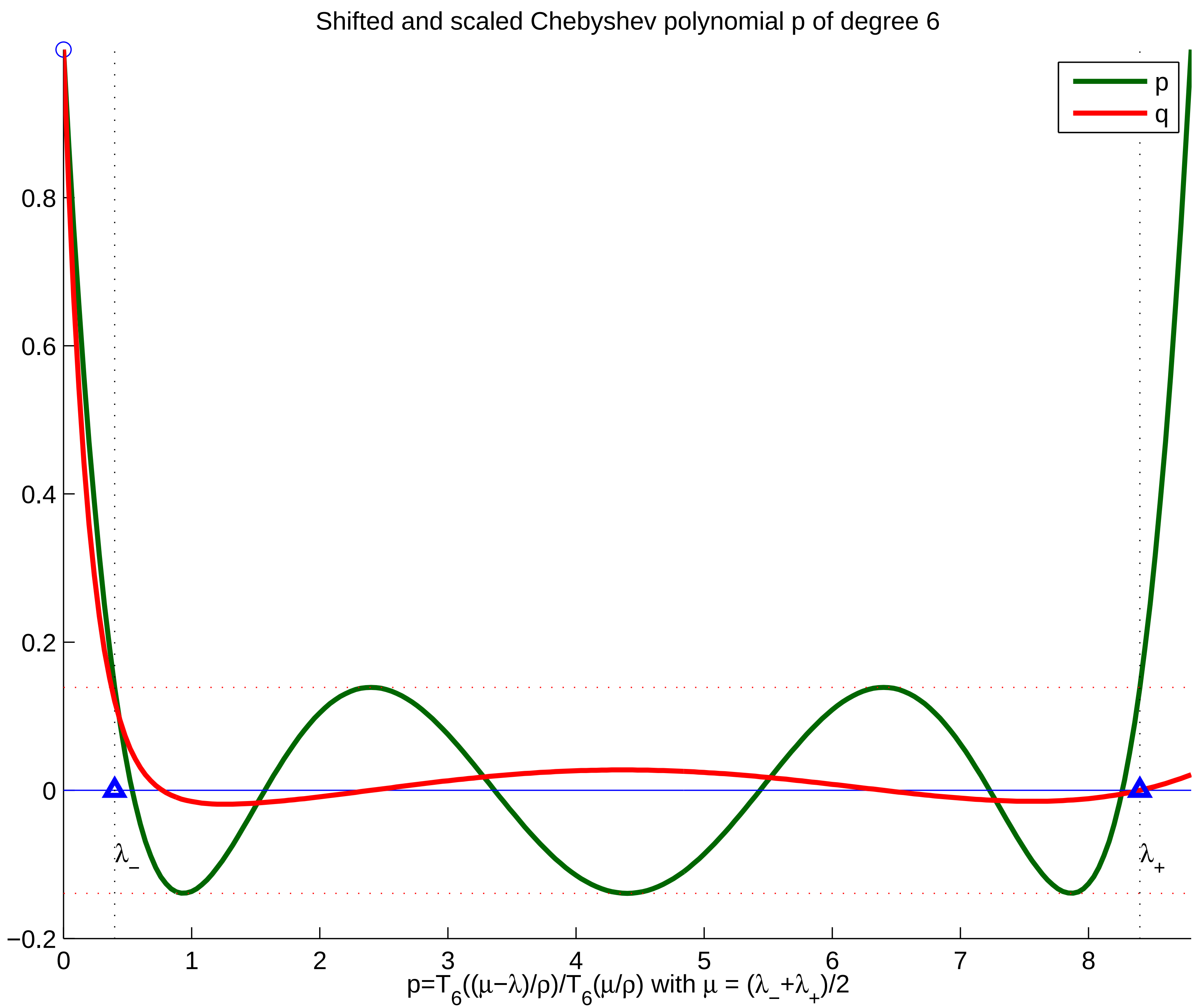
Assume $[\lambda_-, \lambda_+] = [\mu - \rho, \mu + \rho] \subset (0, \infty)$.

Theorem. With $x \equiv (\mu - \lambda)/\rho$ and $p_{\text{Cheb}}(\lambda) \equiv \dfrac{T_\ell(x)}{T_\ell(\mu/\rho)}$,

we have that $p_{\text{Cheb}} \in \mathcal{P}_\ell^0$ and for any $q \in \mathcal{P}_\ell^0$,

$$\max |p_{\text{Cheb}}(\lambda)| \leq \max |q(\lambda)|,$$

where the maxima are taken over all $\lambda \in [\lambda_-, \lambda_+]$.



Shifted and scaled Chebyshev polynomial p of degree 6
p
q
0.8
0.6
0.4
0.2
0
−0.2
λ_
λ+
0
1
2
3
4
5
6
7
8
p=T6((μ−λ)/ρ)/T6(μ/ρ) with μ = (λ_+λ+)/2

Chebyshev polynomials

$$T_\ell(x) \equiv \frac{1}{2}(\zeta^\ell + \zeta^{-\ell}), \quad \text{where} \quad x \equiv \frac{1}{2}(\zeta + \zeta^{-1}) \quad (\zeta \in \mathbb{C}).$$

Exercise. For all $x \in \mathbb{C}$ we have

$$\begin{cases} T_0(x) = 1, \quad T_1(x) = x, \\ T_{k+1}(x) = 2\,x\,T_k(x) - T_{k-1}(x) \text{ for } k = 1, 2, \ldots. \end{cases}$$

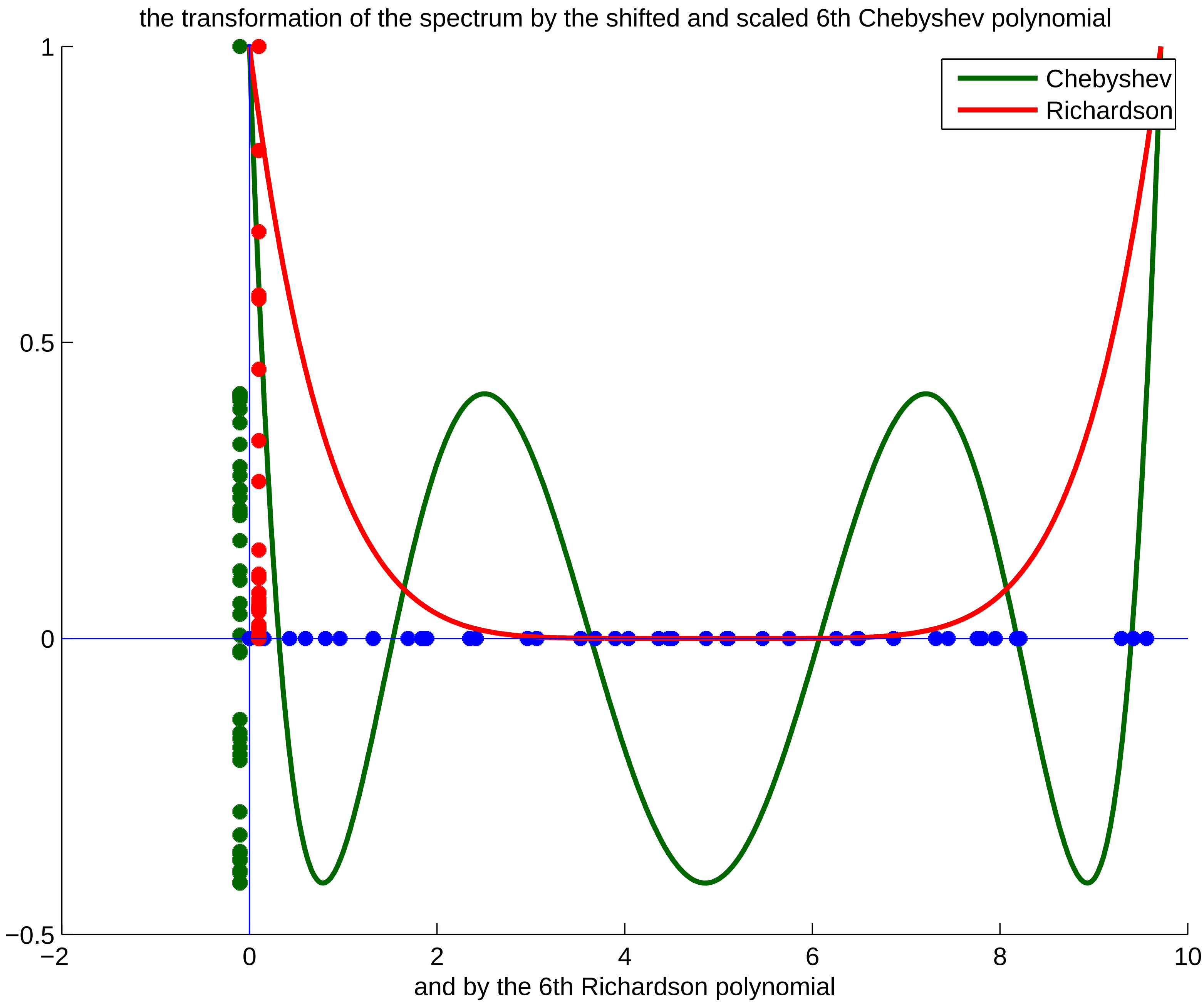
Assume $[\lambda_-, \lambda_+] = [\mu - \rho, \mu + \rho] \subset (0, \infty)$.

Theorem. With $x \equiv (\mu - \lambda)/\rho$ and $p_{\text{Cheb}}(\lambda) \equiv \dfrac{T_\ell(x)}{T_\ell(\mu/\rho)}$,

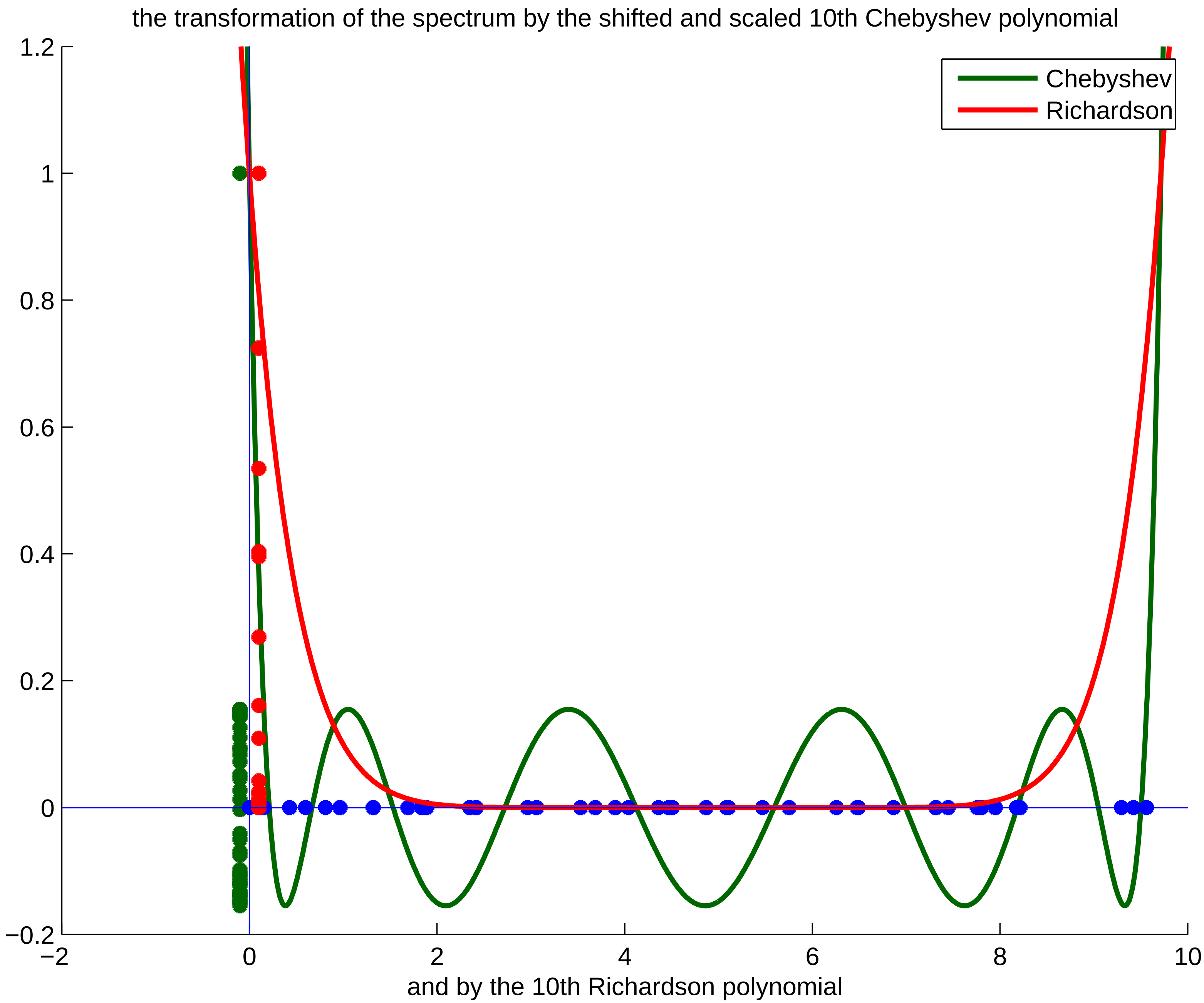
we have that $p_{\text{Cheb}} \in \mathcal{P}_\ell^0$ and

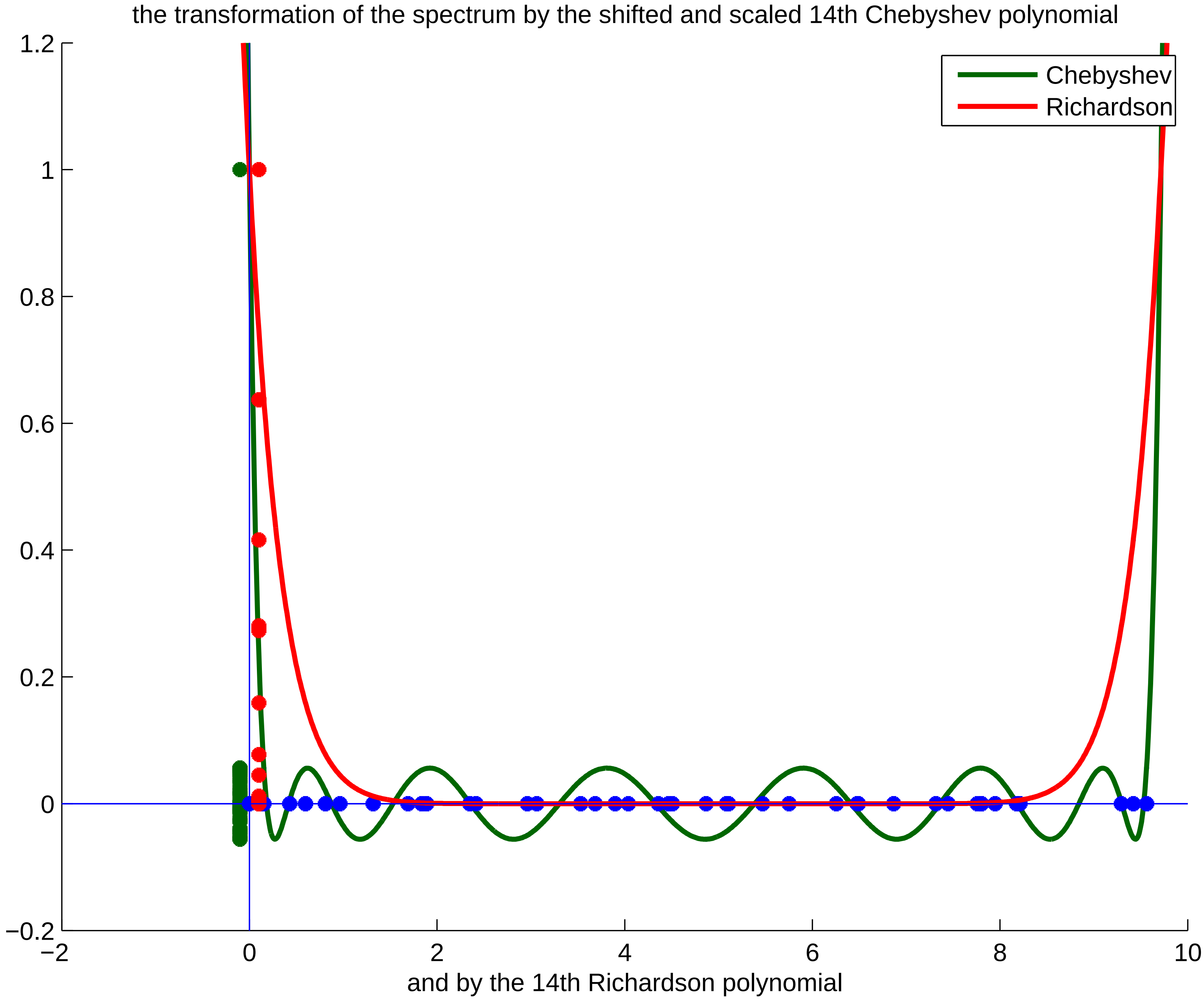
$$\max |p_{\text{Cheb}}(\lambda)| = \frac{1}{|T_\ell(\mu/\rho)|} \leq 2 \exp\left(-\frac{2\ell}{\sqrt{\mathcal{C}}}\right),$$

where the max. is taken over all $\lambda \in [\lambda_-, \lambda_+]$ and $\mathcal{C} \equiv \frac{\lambda_+}{\lambda_-}$.



the transformation of the spectrum by the shifted and scaled 6th Chebyshev polynomial
Chebyshev
Richardson
1
0.5
0
−0.5
−2
0
2
4
6
8
10
and by the 6th Richardson polynomial


the transformation of the spectrum by the shifted and scaled 10th Chebyshev polynomial
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
−2
0
2
4
6
8
10
Chebyshev
Richardson
and by the 10th Richardson polynomial


the transformation of the spectrum by the shifted and scaled 14th Chebyshev polynomial
Chebyshev
Richardson
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
−2
0
2
4
6
8
10
and by the 14th Richardson polynomial

Chebyshev versus Richardson

Error reduction for spectrum in $[\lambda_-,\lambda_+]\subset(0,\infty)$.
Put $\mathcal{C}\equiv\frac{\lambda_+}{\lambda_-}$.

- Degree ℓ Chebychev.

$$\|\mathbf{r}_{k+\ell}^{\text{Cheb}(\ell)}\|_2 \lesssim 2\exp\left(-\frac{2\ell}{\sqrt{\mathcal{C}}}\right)\|\mathbf{r}_k^{\text{Cheb}(\ell)}\|_2 \qquad k \text{ large}$$

- Richardson with optimal α.

$$\|\mathbf{r}_{k+\ell}^{\text{Rich}}\|_2 \lesssim \exp\left(-\frac{2\ell}{\mathcal{C}}\right)\|\mathbf{r}_k^{\text{Rich}}\|_2 \qquad k \text{ large}$$

Remark. We actually have

$$\|\mathbf{r}_{k+\ell}^{\text{Cheb}(\ell)}\|_2 \lesssim \frac{1}{\cosh(2\ell/\sqrt{\mathcal{C}})}\|\mathbf{r}_k^{\text{Cheb}(\ell)}\|_2 \qquad k \text{ large.}$$

Since $\cosh(x)>1$ if $x>1$, this estimate is better than the above one in case $\exp(2\ell/\sqrt{\mathcal{C}})<2$. They are $\approx$ the same if $\exp(2\ell/\sqrt{\mathcal{C}})\gg 2$.

Chebyshev versus Richardson

Error reduction for spectrum in $[\lambda_-,\lambda_+]\subset(0,\infty)$.
Put $\mathcal{C}\equiv\frac{\lambda_+}{\lambda_-}$.

- Degree ℓ Chebychev.

$$\|\mathbf{r}_{k+\ell}^{\mathsf{Cheb}(\ell)}\|_2 \lesssim 2\exp\left(-\frac{2\ell}{\sqrt{\mathcal{C}}}\right)\|\mathbf{r}_k^{\mathsf{Cheb}(\ell)}\|_2 \qquad k \text{ large}$$

- Richardson with optimal α.

$$\|\mathbf{r}_{k+\ell}^{\mathsf{Rich}}\|_2 \lesssim \exp\left(-\frac{2\ell}{\mathcal{C}}\right)\|\mathbf{r}_k^{\mathsf{Rich}}\|_2 \qquad k \text{ large}$$

Note. Chebyshev iteration is designed for spectra in intervals, but works well also for (narrow) ellipses around an interval.

Chebyshev

The zeros of the shifted Chebyshev polynomial determine the parameters in the iteration.

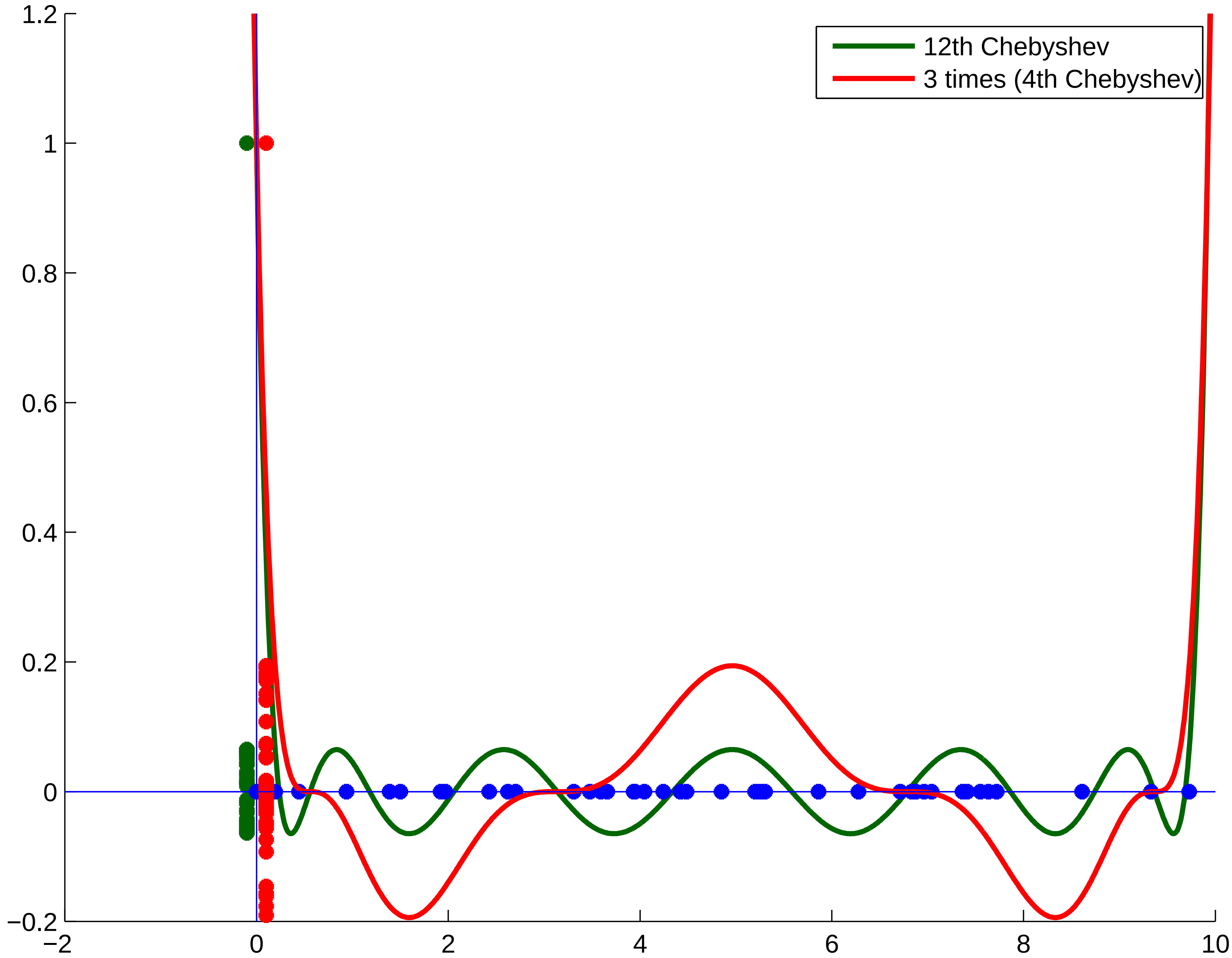
For fixed ℓ and increasing j, take $k = j\ell$.
The polynomial

$$p_k(\lambda) \equiv \frac{T_{j\ell}(\frac{\mu-\lambda}{\rho})}{T_{j\ell}(\frac{\mu}{\rho})}$$

is better than the polynomial

$$q_k(\lambda) \equiv \left(\frac{T_{\ell}(\frac{\mu-\lambda}{\rho})}{T_{\ell}(\frac{\mu}{\rho})}\right)^j$$

Can we repeatedly increase the degree without restarting?


12th Chebyshev
3 times (4th Chebyshev)
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
−2
0
2
4
6
8
10

Chebyshev

With $\mu \equiv \frac{\lambda_+ + \lambda_-}{2}$ and $\rho \equiv \frac{\lambda_+ - \lambda_-}{2}$ we have that

$$\mathbf{r}_k = \frac{\widetilde{\mathbf{r}}_k}{\gamma_k} \quad \text{with} \quad \widetilde{\mathbf{r}}_k \equiv T_k(\tfrac{1}{\rho}(\mu\mathbf{I} - \mathbf{A}))\mathbf{r}_0, \qquad \gamma_k \equiv T_k(\tfrac{\mu}{\rho})$$

$T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)$ implies that

$$\gamma_{k+1} = 2\tfrac{\mu}{\rho}\gamma_k - \gamma_{k-1} \quad \text{and} \quad \widetilde{\mathbf{r}}_{k+1} = \tfrac{2\mu}{\rho}\widetilde{\mathbf{r}}_k - \tfrac{2}{\rho}\mathbf{A}\widetilde{\mathbf{r}}_k - \widetilde{\mathbf{r}}_{k-1}.$$

Hence,

$$\mathbf{r}_{k+1} = \frac{2\mu\gamma_k}{\rho\gamma_{k+1}}\,\mathbf{r}_k - \frac{2\gamma_k}{\rho\,\gamma_{k+1}}\,\mathbf{A}\mathbf{r}_k - \frac{\gamma_{k-1}}{\gamma_{k+1}}\,\mathbf{r}_{k-1}$$

$$\mathbf{x}_{k+1} = \frac{2\mu\gamma_k}{\nu\gamma_{k+1}}\,\mathbf{x}_k + \frac{2\gamma_k}{\rho\gamma_{k+1}}\,\mathbf{r}_k - \frac{\gamma_{k-1}}{\gamma_{k+1}}\,\mathbf{x}_{k-1}$$

Chebyshev

With $\mu \equiv \frac{\lambda_+ + \lambda_-}{2}$ and $\rho \equiv \frac{\lambda_+ - \lambda_-}{2}$ we have that

$$\mathbf{r}_k = \frac{\widetilde{\mathbf{r}}_k}{\gamma_k} \quad \text{with} \quad \widetilde{\mathbf{r}}_k \equiv T_k(\tfrac{1}{\rho}(\mu\mathbf{I} - \mathbf{A}))\mathbf{r}_0, \qquad \gamma_k \equiv T_k(\tfrac{\mu}{\rho})$$

$T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)$ implies that

$$\gamma_{k+1} = 2\tfrac{\mu}{\rho}\gamma_k - \gamma_{k-1} \quad \text{and} \quad \widetilde{\mathbf{r}}_{k+1} = \tfrac{2\mu}{\rho}\widetilde{\mathbf{r}}_k - \tfrac{2}{\rho}\mathbf{A}\widetilde{\mathbf{r}}_k - \widetilde{\mathbf{r}}_{k-1}.$$

Hence,

$$\mathbf{r}_{k+1} = \frac{2\mu\gamma_k}{\rho\gamma_{k+1}}\mathbf{r}_k - \frac{2\gamma_k}{\rho\,\gamma_{k+1}}\mathbf{A}\mathbf{r}_k - \frac{\gamma_{k-1}}{\gamma_{k+1}}\mathbf{r}_{k-1}$$

$$\mathbf{x}_{k+1} = \frac{2\mu\gamma_k}{\nu\gamma_{k+1}}\mathbf{x}_k + \frac{2\gamma_k}{\rho\gamma_{k+1}}\mathbf{r}_k - \frac{\gamma_{k-1}}{\gamma_{k+1}}\mathbf{x}_{k-1}$$

Note that the update of the residual also uses an additional 'older' residual.

Chebyshev

```
Select x_0, tol, kmax, μ, ρ
Compute r_0 = b − A x_0
Set ν_0 = μ, r_1 = r_0 − (1/μ) A r_0, x_1 = x_0 + (1/μ) r_0
for k = 1, ..., kmax do
    If ||r|| ≤ tol, break, end if
    ν_k = 2μ − ρ^2/ν_{k−1}
    α_k = 2μ/ν_k, β_k = 2/ν_k, γ_k = ρ^2/(ν_{k−1} ν_k),
    r_{k+1} = α_k r_k − β_k A r_k − γ_k r_{k−1}
    x_{k+1} = α_k x_k + β_k r_k − γ_k x_{k−1}
end for
```

Select $\mathbf{x}_0$, tol, kmax, μ, ρ

Compute $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$

Set $\nu_0 = \mu$, $\mathbf{r}_1 = \mathbf{r}_0 - \frac{1}{\mu}\mathbf{A}\mathbf{r}_0$, $\mathbf{x}_1 = \mathbf{x}_0 + \frac{1}{\mu}\mathbf{r}_0$

for $k = 1, \ldots,$ kmax do

If $\|\mathbf{r}\| \leq tol$, break, end if

$$\nu_k = 2\mu - \rho^2/\nu_{k-1}$$

$$\alpha_k = \frac{2\mu}{\nu_k},\ \beta_k = \frac{2}{\nu_k},\ \gamma_k = \frac{\rho^2}{\nu_{k-1}\nu_k},$$

$$\mathbf{r}_{k+1} = \alpha_k \mathbf{r}_k - \beta_k \mathbf{A}\mathbf{r}_k - \gamma_k \mathbf{r}_{k-1}$$

$$\mathbf{x}_{k+1} = \alpha_k \mathbf{x}_k + \beta_k \mathbf{r}_k - \gamma_k \mathbf{x}_{k-1}$$

end for

With $\mu, \rho \in \mathbb{R}, \rho > 0$ such that

$$\Lambda(\mathbf{A}) \subset [\mu - \rho, \mu + \rho] \subset (0, \infty).$$

Chebyshev

```
Select x_0, tol, kmax, μ, ρ
Compute r_0 = b − A x_0
Set ν = μ, r = r_0 − (1/μ) A r_0, x = x_0 + (1/μ) r_0
for k = 1, ..., kmax do
    Break if ||r||_2 < tol
    γ = ρ^2/ν, ν = 2μ − γ
    α = 2μ/ν, β = 2/ν, γ ← γ/ν,
    r_2 = α r − β A r − γ r_0, r_0 = r, r = r_2
    x_2 = α x + β r_0 − γ x_0, x_0 = x, x = x_2
end for
```

Select $\mathbf{x}_0$, tol, kmax, μ, ρ

Compute $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$

Set $\nu = \mu$, $\mathbf{r} = \mathbf{r}_0 - \frac{1}{\mu}\mathbf{A}\mathbf{r}_0$, $\mathbf{x} = \mathbf{x}_0 + \frac{1}{\mu}\mathbf{r}_0$

for $k = 1, \ldots,$ kmax do

Break if $\|\mathbf{r}\|_2 < tol$

$\gamma = \rho^2/\nu$, $\nu = 2\mu - \gamma$

$\alpha = \frac{2\mu}{\nu}$, $\beta = \frac{2}{\nu}$, $\gamma \leftarrow \frac{\gamma}{\nu}$,

$\mathbf{r}_2 = \alpha\,\mathbf{r} - \beta\,\mathbf{A}\mathbf{r} - \gamma\,\mathbf{r}_0$, $\mathbf{r}_0 = \mathbf{r}$, $\mathbf{r} = \mathbf{r}_2$

$\mathbf{x}_2 = \alpha\,\mathbf{x} + \beta\,\mathbf{r}_0 - \gamma\,\mathbf{x}_0$, $\mathbf{x}_0 = \mathbf{x}$, $\mathbf{x} = \mathbf{x}_2$

end for

With $\mu, \rho \in \mathbb{R}, \rho > 0$ such that

$$\Lambda(\mathbf{A}) \subset [\mu - \rho, \mu + \rho] \subset (0, \infty).$$

Degree ℓ Chebyshev versus Chebyshev

Error reduction for spectrum in $[\lambda_-, \lambda_+] \subset (0, \infty)$.
Put $\mathcal{C} \equiv \frac{\lambda_+}{\lambda_-}$.

- Degree ℓ Chebychev.

$$\|\mathbf{r}_{j\ell}^{\mathrm{Cheb}(\ell)}\|_2 \leq \mathcal{C}_E\, 2^j \exp\left(-\frac{2j\ell}{\sqrt{\mathcal{C}}}\right) \|\mathbf{r}_0\|_2 \qquad k \text{ large}$$

- Chebyshev

$$\|\mathbf{r}_{j\ell}^{\mathrm{Cheb}}\|_2 \leq \mathcal{C}_E\, 2 \exp\left(-\frac{2j\ell}{\sqrt{\mathcal{C}}}\right) \|\mathbf{r}_0\|_2 \qquad k \text{ large}$$

Here, $\mathcal{C}_E$ some constant like $\mathcal{C}_E = \|\mathbf{V}\|_2 \|\mathbf{V}^{-1}\|\|_2$,
the conditioning of the basis of eigenvectors.

Degree ℓ Chebyshev versus Chebyshev

Error reduction for spectrum in $[\lambda_-, \lambda_+] \subset (0, \infty)$.
Put $\mathcal{C} \equiv \frac{\lambda_+}{\lambda_-}$.

- Degree ℓ Chebychev.

$$\|\mathbf{r}_{j\ell}^{\text{Cheb}(\ell)}\|_2 \leq \mathcal{C}_E\, 2^j \exp\left(-\frac{2j\ell}{\sqrt{\mathcal{C}}}\right) \|\mathbf{r}_0\|_2 \qquad k \text{ large}$$

- Chebyshev

$$\|\mathbf{r}_{j\ell}^{\text{Cheb}}\|_2 \leq \mathcal{C}_E\, 2 \exp\left(-\frac{2j\ell}{\sqrt{\mathcal{C}}}\right) \|\mathbf{r}_0\|_2 \qquad k \text{ large}$$

Remark. The PM argument that for large k

$$\|\mathbf{r}_{k+\ell}^{\text{Cheb}(\ell)}\|_2 \lesssim 2 \exp\left(-2\ell/\sqrt{\mathcal{C}}\right) \|\mathbf{r}_k^{\text{Cheb}(\ell)}\|_2$$

is not applicable for Chebyshev, since we do not iterate with a fixed polynomial.

$\mathbf{A}\mathbf{x} = \mathbf{b}$

Summary.

- $\mathbf{r}_k$ is of the form $p_k(\mathbf{A})\mathbf{r}_0$ with $p_k \in \mathcal{P}_k^0$.

Examples.

$p_k(x) = (1 - \alpha x)^k$	Richardson,
$p_{m\ell}(x) = (\prod_{j=1}^{\ell}(1 - \alpha_j x))^m$	Polynomial,
$p_k(x) = T_k(\frac{\mu - x}{\rho})/T_k(\frac{\mu}{\rho})$	Chebyshev,. . .

- Since $p_k(0) = 1$ we have that $p_k(x) = 1 - x q_{k-1}(x)$ for some polynomial q_{k-1} of degree $k - 1$ and

$$\mathbf{r}_k = \mathbf{r}_0 - \mathbf{A} q_{k-1}(\mathbf{A})\mathbf{r}_0, \qquad \mathbf{x}_k = \mathbf{x}_0 + q_{k-1}(\mathbf{A})\mathbf{r}_0.$$

- Consistent update of $\mathbf{r}_k$ and $\mathbf{x}_k$,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{u}_k, \quad \mathbf{c}_k = \mathbf{A}\mathbf{u}_k, \quad \mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{c}_k$$

i.e., no need to gather explicit information on q_{k-1}.

$\mathbf{Ax} = \mathbf{b}$

Summary.
Let $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$ be the **Krylov subspace of order** k generated by **A** and $\mathbf{r}_0$:

$$\begin{aligned} \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0) &\equiv \text{span}(\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{k-1}\mathbf{r}_0) \\ &= \{q(\mathbf{A})\mathbf{r}_0 \mid q \in \mathcal{P}_{k-1}\}. \end{aligned} .$$

Then

$$\mathbf{r}_k \in \mathbf{r}_0 + \mathbf{A}\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0) \subset \mathcal{K}_{k+1}(\mathbf{A}, \mathbf{r}_0),$$
$$\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0).$$

Dynamic. Multiple parameter

Find the residual in the Krylov subspace $\mathcal{K}_{k+1}(\mathbf{A}, \mathbf{r}_0)$ with ‘smallest’ norm. Use also ’older’ residuals in the update process.

Program Lecture 5

- Power Method & Richardson
- Filtering
- Shift-and-Invert & Preconditioning
- Polynomial Iteration
- Selecting Parameters

  1) single parameter a) static
  b) dynamic

  2) Multiple parameters a) static (Chebyshev)
  b) dynamic (GCR)

Generalized Conjugate Residuals

Recall (local minimal residuals)

$$\mathbf{c}_k = \mathbf{A}\mathbf{r}_k, \quad \mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{c}_k \text{ with } \alpha_k \text{ st } \mathbf{r}_{k+1} \perp \mathbf{c}_k.$$

Idea. $$\mathbf{r}_{k+1} = \mathbf{r}_k - (\alpha_k \mathbf{c}_k + \alpha_{k-1}\mathbf{c}_{k-1} + \ldots + \alpha_0 \mathbf{c}_0)$$

with α_j st $\|\mathbf{r}_{k+1}\|_2$ smallest ($\Leftrightarrow$ $\mathbf{r}_{k+1} \perp \mathbf{c}_j$ all $j \leq k$).

How to compute the α_j efficiently?

Generalized Conjugate Residuals

Recall (local minimal residuals)

$$\mathbf{c}_k = \mathbf{A}\mathbf{r}_k, \quad \mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{c}_k \text{ with } \alpha_k \text{ st } \mathbf{r}_{k+1} \perp \mathbf{c}_k.$$

Idea. $\mathbf{r}_{k+1} = \mathbf{r}_k - (\alpha_k \mathbf{c}_k + \alpha_{k-1}\mathbf{c}_{k-1} + \ldots + \alpha_0 \mathbf{c}_0)$

with α_j st $\|\mathbf{r}_{k+1}\|_2$ smallest ($\Leftrightarrow$ $\mathbf{r}_{k+1} \perp \mathbf{c}_j$ all $j \leq k$).

- Solve $\mathbf{C}_{k+1}\vec{\alpha}_{k+1} \equiv [\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_k](\alpha_0, \ldots, \alpha_k)^\mathsf{T} = \mathbf{r}_k$ in least square sense
- 'Orthogonalise' $\mathbf{c}_0, \ldots, \mathbf{c}_k$ first.

Exercise. Suppose $\mathbf{c}_0, \ldots, \mathbf{c}_{k-1}$ orthogonal

$$\left.\begin{array}{l} \mathbf{r}_k \perp \mathbf{c}_0, \ldots, \mathbf{c}_{k-1} \\ \mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{c}_k \perp \mathbf{c}_k \end{array}\right\} \Rightarrow \mathbf{r}_{k+1} \perp \mathbf{c}_0, \ldots, \mathbf{c}_k$$

Local Minimal Residuals

```
Select x, α, tol, k_max
Compute r = b − Ax

for k = 0, 1, 2, ..., k_max do
    If ||r|| ≤ tol, break, end if
    u = r
    c = Au
    σ = c*c, ρ = c*r, α = ρ/σ
    x ← x + αu
    r ← r − αc
end do
```

Generalized Conjugate Residuals

```
Select x_0, k_max, tol
Compute r_0 = b - A x_0
for k = 0, 1, ..., k_max do
    break if ||r_k||_2 <= tol
    u_k = r_k, c_k = A u_k
    for j = 0, ..., k-1 do
        beta_j = c_j^* c_k / sigma_j
        u_k <- u_k - beta_j u_j
        c_k <- c_k - beta_j c_j
    end for
    sigma_k = c_k^* c_k,   alpha_k = c_k^* r_k / sigma_k
    x_{k+1} = x_k + alpha_k u_k
    r_{k+1} = r_k - alpha_k c_k
end for
```

Select $\mathbf{x}_0$, $k_{\max}$, *tol*

Compute $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$

for $k = 0, 1, \ldots, k_{\max}$ do

  break if $\|\mathbf{r}_k\|_2 \le \mathit{tol}$

  $\mathbf{u}_k = \mathbf{r}_k$, $\mathbf{c}_k = \mathbf{A}\mathbf{u}_k$

  for $j = 0, \ldots, k-1$ do

    $\beta_j = \mathbf{c}_j^* \mathbf{c}_k / \sigma_j$

    $\mathbf{u}_k \leftarrow \mathbf{u}_k - \beta_j \mathbf{u}_j$

    $\mathbf{c}_k \leftarrow \mathbf{c}_k - \beta_j \mathbf{c}_j$

  end for

  $\sigma_k = \mathbf{c}_k^* \mathbf{c}_k$, $\alpha_k = \mathbf{c}_k^* \mathbf{r}_k / \sigma_k$

  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{u}_k$

  $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{c}_k$

end for

Generalized Conjugate Residuals

```
Select x_0, k_max, tol
Compute r_0 = b - A x_0
for k = 0, 1, ..., k_max do
    break if ||r_k||_2 <= tol
    u_k = r_k, c_k = A u_k
    for j = 0, ..., k-1 do
        beta_j = c_j^* c_k / sigma_j
        u_k <- u_k - beta_j u_j
        c_k <- c_k - beta_j c_j
    end for
    sigma_k = c_k^* c_k,  alpha_k = c_k^* r_k / sigma_k
    x_{k+1} = x_k + alpha_k u_k
    r_{k+1} = r_k - alpha_k c_k
end for
```

Select $\mathbf{x}_0$, $k_{\max}$, tol

Compute $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$

for $k = 0, 1, \ldots, k_{\max}$ do

  break if $\|\mathbf{r}_k\|_2 \leq tol$

  $\mathbf{u}_k = \mathbf{r}_k$, $\mathbf{c}_k = \mathbf{A}\mathbf{u}_k$

  for $j = 0, \ldots, k-1$ do

    $\beta_j = \mathbf{c}_j^* \mathbf{c}_k / \sigma_j$

    $\mathbf{u}_k \leftarrow \mathbf{u}_k - \beta_j \mathbf{u}_j$

    $\mathbf{c}_k \leftarrow \mathbf{c}_k - \beta_j \mathbf{c}_j$

  end for

  $\sigma_k = \mathbf{c}_k^* \mathbf{c}_k$, $\alpha_k = \mathbf{c}_k^* \mathbf{r}_k / \sigma_k$

  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{u}_k$

  $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{c}_k$

end for

Generalized Conjugate Residuals

Select $\mathbf{x}_0$, $k_{\max}$, tol

Compute $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$

for $k = 0, 1, \ldots, k_{\max}$ do

  break if $\|\mathbf{r}_k\|_2 \leq tol$

  $\mathbf{u}_k = \mathbf{r}_k$, $\mathbf{c}_k = \mathbf{A}\mathbf{u}_k$

  for $j = 0, \ldots, k-1$ do

    $\beta_j = \mathbf{c}_j^* \mathbf{c}_k / \sigma_j$

    $\mathbf{u}_k \leftarrow \mathbf{u}_k - \beta_j \mathbf{u}_j$

    $\mathbf{c}_k \leftarrow \mathbf{c}_k - \beta_j \mathbf{c}_j$

  end for

  $\sigma_k = \mathbf{c}_k^* \mathbf{c}_k$, $\alpha_k = \mathbf{c}_k^* \mathbf{r}_k / \sigma_k$

  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{u}_k$

  $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{c}_k$

end for

Generalized Conjugate Residuals

Select $\mathbf{x}_0$, $k_{\max}$, *tol*

Compute $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$

for $k = 0, 1, \ldots, k_{\max}$ do

    break if $\|\mathbf{r}_k\|_2 \le tol$

    $\mathbf{u}_k = \mathbf{r}_k$, $\mathbf{c}_k = \mathbf{A}\mathbf{u}_k$

    for $j = 0, \ldots, k-1$ do

        $\beta_j = \mathbf{c}_j^* \mathbf{c}_k / \sigma_j$

        $\mathbf{u}_k \leftarrow \mathbf{u}_k - \beta_j \mathbf{u}_j$

        $\mathbf{c}_k \leftarrow \mathbf{c}_k - \beta_j \mathbf{c}_j$

    end for

    $\sigma_k = \mathbf{c}_k^* \mathbf{c}_k$, $\alpha_k = \mathbf{c}_k^* \mathbf{r}_k / \sigma_k$

    $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{u}_k$

    $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{c}_k$

end for

Generalized Conjugate Residuals

```
Select x, k_max, tol
Compute r = b − Ax
for k = 0, 1, ..., k_max do
    break if ||r||_2 ≤ tol
    u_k = r,   c_k = A u_k
    for j = 0, ..., k − 1 do
        β = c_j^* c_k / σ_j
        u_k ← u_k − β u_j
        c_k ← c_k − β c_j
    end for
    σ_k = c_k^* c_k,   α = c_k^* r / σ_k
    x ← x + α u_k
    r ← r − α c_k
end for
```

Generalized Conjugate Residuals

Exercise. Before the orthogonalisation loop we have

$$\widetilde{\mathbf{c}}_k = \mathbf{A}\widetilde{\mathbf{u}}_k \text{ by definition.}$$

Prove that after the orthogonalisation loop we still have

$$\mathbf{c}_k = \mathbf{A}\mathbf{u}_k.$$

Exercise.

$$\begin{aligned}
\text{span}(\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_k) &= \mathcal{K}_{k+1}(\mathbf{A}, \mathbf{r}_0), \\
\text{span}(\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{k-1}) &= \mathbf{A}\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0), \\
\mathbf{r}_k \in \text{span}(\mathbf{r}_0) \oplus \mathbf{A}\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0) &= \mathcal{K}_{k+1}(\mathbf{A}, \mathbf{r}_0).
\end{aligned}$$

Generalized Conjugate Residuals

GCR is an optimal Krylov subspace solver:

Theorem. Assume $\mathbf{x}_0 = \mathbf{0}$: $\mathbf{r}_0 = \mathbf{b}$.
The GCR approximate solution $\mathbf{x}_k$ at step k is the vector in $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$ with smallest residual norm:

$$\|\mathbf{r}_k\|_2 = \|\mathbf{r}_0 - \mathbf{A}\mathbf{x}_k\|_2 \leq \|\mathbf{r}_0 - \mathbf{A}\widetilde{\mathbf{x}}\|_2 \quad (\widetilde{\mathbf{x}} \in \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)).$$

In particular, $\|\mathbf{r}_k^{\text{GCR}}\|_2 \leq \|\mathbf{r}_k^{\text{Cheb}}\|_2$.

Hence, if $\Lambda(\mathbf{A}) \subset [\lambda_-, \lambda_+] \subset (0, \infty)$, then, with $\mathcal{C} \equiv \frac{\lambda_+}{\lambda_-}$,

$$\|\mathbf{r}_k^{\text{GCR}}\|_2 \leq \mathcal{C}_E\, 2 \exp\left(-\frac{2k}{\sqrt{\mathcal{C}}}\right) \|\mathbf{r}_0\|_2.$$

Here, $\mathcal{C}_E$ some constant like $\mathcal{C}_E = \|\mathbf{V}\|_2 \|\mathbf{V}^{-1}\|\|_2$,

the conditioning of the basis of eigenvectors.

Chebyshev versus GCR

Chebyshev.

\+ No inner products

\+ Short recurrences (three term recurrences)

− Not the smallest residuals with appr. sol. from $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$.

− Sensitive to the estimate of the hull of the spectrum.

− Only effective if spectrum in
a narrow ellipse in a half plane as $\mathbb{C}^+$.

GCR.

\+ Smallest residual with appr. sol. from $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$.

\+ Flexible (any information can be used for $\mathbf{u}_k$)

\+ Stable

− **Growing recurrences with increasing step number k: increasing computational costs, increasing storage demands.**

Flexible GCR

In the preceding transparancies, GCR has been constructed as an **optimal Krylov subspace solver**.

However, GCR can be turned into a **supspace solver!**:

If

$$\mathbf{u}_k = \mathbf{r}_k$$

is replaced by

$$\texttt{Solve approximately } \mathbf{A}\mathbf{u}_k = \mathbf{r}_k \texttt{ for } \mathbf{u}_k$$

then we search for an approximate solution in the search subspace span$(\mathbf{u}_0, \ldots, \mathbf{u}_{k-1})$ and GCR finds the one with smallest residual.

Exercise. Exact solve of $\mathbf{A}\mathbf{u}_k = \mathbf{r}_k$ leads to $\mathbf{x}_{k+1} = \mathbf{x}$.

GCR

```
Select x, k_max, tol
Compute r = b − Ax
for k = 0, 1, ..., k_max do
    break if ||r||_2 ≤ tol
    u_k = r
    c_k = A u_k
    for j = 0, ..., k − 1 do
        β = c_j^* c_k / σ_j
        u_k ← u_k − β u_j
        c_k ← c_k − β c_j
    end for
    σ_k = c_k^* c_k ,   α = c_k^* r / σ_k
    x ← x + α u_k
    r ← r − α c_k
end for
```

Select $\mathbf{x}$, $k_{\max}$, tol

Compute $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$

for $k = 0, 1, \ldots, k_{\max}$ do

  break if $\|\mathbf{r}\|_2 \leq tol$

  $\mathbf{u}_k = \mathbf{r}$

  $\mathbf{c}_k = \mathbf{A}\mathbf{u}_k$

  for $j = 0, \ldots, k-1$ do

    $\beta = \mathbf{c}_j^* \mathbf{c}_k / \sigma_j$

    $\mathbf{u}_k \leftarrow \mathbf{u}_k - \beta\, \mathbf{u}_j$

    $\mathbf{c}_k \leftarrow \mathbf{c}_k - \beta\, \mathbf{c}_j$

  end for

  $\sigma_k = \mathbf{c}_k^* \mathbf{c}_k$, $\alpha = \mathbf{c}_k^* \mathbf{r} / \sigma_k$

  $\mathbf{x} \leftarrow \mathbf{x} + \alpha\, \mathbf{u}_k$

  $\mathbf{r} \leftarrow \mathbf{r} - \alpha\, \mathbf{c}_k$

end for

Flexible GCR

```
Select x, k_max, tol
Compute r = b − Ax
for k = 0, 1, ..., k_max do
    break if ||r||_2 ≤ tol
    Approximately solve Au_k = r for u_k
    c_k = Au_k
    for j = 0, ..., k − 1 do
        β = c_j^* c_k / σ_j
        u_k ← u_k − β u_j
        c_k ← c_k − β c_j
    end for
    σ_k = c_k^* c_k,   α = c_k^* r / σ_k
    x ← x + α u_k
    r ← r − α c_k
end for
```

Flexible GCR

Solve approximately $\mathbf{A}\mathbf{u}_k = \mathbf{r}_k$ for $\mathbf{u}_k$

Examples.

- $\mathbf{u}_k = \mathbf{r}_k$: standard GCR searches $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$
- Solve $\mathbf{M}\mathbf{u}_k = \mathbf{r}_k$ for $\mathbf{u}_k$: preconditioned GCR
  searches the Krylov subspace $\mathbf{M}^{-1}\mathcal{K}_k(\mathbf{A}\mathbf{M}^{-1}, \mathbf{r}_0)$.
- Use ℓ steps of GCR to solve $\mathbf{A}\mathbf{u}_k = \mathbf{r}_k$: nested GCR
  solution in $\mathcal{K}_{\ell k}(\mathbf{A}, \mathbf{r}_0)$
- Use GCR to solve $\mathbf{A}\mathbf{u}_k = \mathbf{r}_k$ to rel. res. acc. 0.1
- At step $k = 0, 1, \ldots, \ell$ use information on the solution (as $\mathbf{u}_k$ representing singularities, etc.)
- At step $k = 0, \ldots, \ell$ use a '$\mathbf{u}_j$' from GCR run for $\mathbf{A}\mathbf{x} = \widetilde{\mathbf{b}}$.

⋮

GCR and Krylov subspace solvers

GCR is a subspace solver

Pros

- Flexible (any information can be exploited)

Cons

- Higher computational costs per step

Krylov subspace solvers

Pros

- Krylov subspace structure can be exploited to save computational costs per step [to be implemented $*$)].

- Polynomial approximation theory provides insight in convergence behaviour

Cons

- Sensitive to rounding errors [if $*$)].

- Not flexible (only fixed preconditioners are allowed).

Exploiting the Krylov subspace structure

to save computational costs.

Exercise. For GCR prove that:

$$\text{span}(\mathbf{r}_0, \ldots, \mathbf{r}_{k-1}) = \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0),$$

$$\text{span}(\mathbf{c}_0, \ldots, \mathbf{c}_{k-1}) = \mathbf{A}\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0),$$

$$\mathbf{r}_k \perp \text{span}(\mathbf{c}_0, \ldots, \mathbf{c}_{k-1}),$$

$$\mathbf{r}_k \perp \mathbf{A}\mathbf{r}_j \quad \forall j = 0, 1, \ldots, k-1:$$ **Conjugate residuals**.

$\mathbf{A}^* = \mathbf{A}$ Conjugate Residuals

Now, suppose $\mathbf{A}^* = \mathbf{A}$.

$$\mathbf{A}\mathbf{r}_k \perp \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0) \supset \mathbf{A}\mathcal{K}_{k-1}(\mathbf{A}, \mathbf{r}_0).$$

Let β_{k-1} be such that $\widetilde{\mathbf{c}}_k \equiv \mathbf{A}\mathbf{r}_k - \beta_{k-1}\mathbf{c}_{k-1} \perp \mathbf{c}_{k-1}$.

$$\Rightarrow \quad \widetilde{\mathbf{c}}_k \perp \text{span}(\mathbf{c}_0, \ldots, \mathbf{c}_{k-1}) \quad \Rightarrow \quad \mathbf{c}_k = \widetilde{\mathbf{c}}_k.$$

$\mathbf{A}^* = \mathbf{A}$ Conjugate Residuals

```
Select x_0, k_max, tol
Compute r_0 = b - A x_0
for k = 0, 1, ..., k_max do
    break if ||r_k||_2 <= tol
    u_k = r_k,   c_k = A u_k
    beta_{k-1} = c*_{k-1} c_k / sigma_{k-1}
    u_k <- u_k - beta_{k-1} u_{k-1}
    c_k <- c_k - beta_{k-1} c_{k-1}
    sigma_k = c*_k c_k,   alpha_k = c*_k r_k / sigma_k
    x_{k+1} = x_k + alpha_k u_k
    r_{k+1} = r_k - alpha_k c_k
end for
```

Select $\mathbf{x}_0$, $k_{\max}$, *tol*

Compute $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$

for $k = 0, 1, \ldots, k_{\max}$ do

  break if $\|\mathbf{r}_k\|_2 \leq tol$

  $\mathbf{u}_k = \mathbf{r}_k, \quad \mathbf{c}_k = \mathbf{A}\mathbf{u}_k$

  $\beta_{k-1} = \mathbf{c}_{k-1}^* \mathbf{c}_k / \sigma_{k-1}$

  $\mathbf{u}_k \leftarrow \mathbf{u}_k - \beta_{k-1}\mathbf{u}_{k-1}$

  $\mathbf{c}_k \leftarrow \mathbf{c}_k - \beta_{k-1}\mathbf{c}_{k-1}$

  $\sigma_k = \mathbf{c}_k^* \mathbf{c}_k, \quad \alpha_k = \mathbf{c}_k^* \mathbf{r}_k / \sigma_k$

  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{u}_k$

  $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{c}_k$

end for

$\mathbf{A}^* = \mathbf{A}$ Conjugate Residuals

```
Select x_0, k_max, tol
Compute r_0 = b - A x_0
for k = 0, 1, ..., k_max do
    break if ||r_k||_2 <= tol
    u_k = r_k,   c_k = A u_k
    β_{k-1} = c*_{k-1} c_k / σ_{k-1}
    u_k <- u_k - β_{k-1} u_{k-1}
    c_k <- c_k - β_{k-1} c_{k-1}
    σ_k = c*_k c_k,   α_k = c*_k r_k / σ_k
    x_{k+1} = x_k + α_k u_k
    r_{k+1} = r_k - α_k c_k
end for
```

Select $\mathbf{x}_0$, $k_{\max}$, *tol*

Compute $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$

for $k = 0, 1, \ldots, k_{\max}$ do

break if $\|\mathbf{r}_k\|_2 \leq \mathit{tol}$

$\mathbf{u}_k = \mathbf{r}_k$, $\mathbf{c}_k = \mathbf{A}\mathbf{u}_k$

$\beta_{k-1} = \mathbf{c}_{k-1}^* \mathbf{c}_k / \sigma_{k-1}$

$\mathbf{u}_k \leftarrow \mathbf{u}_k - \beta_{k-1}\mathbf{u}_{k-1}$

$\mathbf{c}_k \leftarrow \mathbf{c}_k - \beta_{k-1}\mathbf{c}_{k-1}$

$\sigma_k = \mathbf{c}_k^* \mathbf{c}_k$, $\alpha_k = \mathbf{c}_k^* \mathbf{r}_k / \sigma_k$

$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{u}_k$

$\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{c}_k$

end for

$\mathbf{A}^* = \mathbf{A}$ Conjugate Residuals

3 DOTs:

$$\beta_{k-1} = \frac{\mathbf{c}_{k-1}^* \mathbf{A} \mathbf{r}_k}{\sigma_{k-1}}, \qquad \sigma_k = \mathbf{c}_k^* \mathbf{c}_k, \qquad \rho_k \equiv \mathbf{c}_k^* \mathbf{r}_k, \qquad \alpha_k = \frac{\rho_k}{\sigma_k}$$

$\mathbf{A}^* = \mathbf{A}$ Conjugate Residuals

3 DOTs:

$$\beta_{k-1} = \frac{\mathbf{c}_{k-1}^*\mathbf{A}\mathbf{r}_k}{\sigma_{k-1}}, \qquad \sigma_k = \mathbf{c}_k^*\mathbf{c}_k, \qquad \rho_k \equiv \mathbf{c}_k^*\mathbf{r}_k, \qquad \alpha_k = \frac{\rho_k}{\sigma_k}$$

Save 1 DOT:

$$\beta_{k-1} = \frac{\mathbf{c}_{k-1}^*\mathbf{A}\mathbf{r}_k}{\mathbf{c}_{k-1}^*\mathbf{c}_{k-1}} = \frac{[\mathbf{r}_k - \mathbf{r}_{k-1}]^*\mathbf{A}\mathbf{r}_k}{[\mathbf{r}_k - \mathbf{r}_{k-1}]^*\mathbf{c}_{k-1}} = -\frac{\mathbf{r}_k^*\mathbf{A}\mathbf{r}_k}{\mathbf{r}_{k-1}^*\mathbf{c}_{k-1}} = -\frac{\rho_k}{\rho_{k-1}}$$

Here we used that

$$\alpha_{k-1}\mathbf{c}_{k-1} = \mathbf{r}_k - \mathbf{r}_{k-1}$$

$$\mathbf{r}_k \perp \mathbf{c}_{k-1}, \quad \mathbf{r}_k \perp \mathbf{A}\mathbf{r}_{k-1}$$

$$\mathbf{c}_k = \mathbf{A}\mathbf{r}_k - \beta_{k-1}\mathbf{c}_{k-1}$$

$$\mathbf{c}_k \perp \mathbf{c}_{k-1}$$

Exercise. $\sigma_k = \mathbf{c}_k^*\mathbf{A}\mathbf{r}_k, \quad \rho_k = \mathbf{r}_k^*\mathbf{A}\mathbf{r}_k \in \mathbb{R}.$

$\mathbf{A}^* = \mathbf{A}$ Conjugate Residuals

```
Select x, k_max, tol
Compute r = b − Ax
Set u = 0, c = 0, ρ = 1
for k = 0, 1, ..., k_max do
    break if ||r||_2 ≤ tol
    u_1 = r,  c_1 = A u_1
    σ = −ρ,  ρ = r* c_1,  β = ρ/σ
    u ← u_1 − β u
    c ← c_1 − β c
    σ = c* c,  α = ρ/σ
    x ← x + α u
    r ← r − α c
end for
```

$\mathbf{A}^* = \mathbf{A}$ Conjugate Residuals

```
Select x, k_max, tol
Compute r = b - Ax
Set u = 0, c = 0, ρ = 1
for k = 0, 1, ..., k_max do
    break if ||r||_2 ≤ tol
    c_1 = Ar
    σ = -ρ, ρ = r*c_1, β = ρ/σ
    u ← r - β u
    c ← c_1 - β c
    σ = c*c, α = ρ/σ
    x ← x + α u
    r ← r - α c
end for
```

$\mathbf{A}^* = \mathbf{A} > 0$ Conjugate Gradient

Suppose $\mathbf{A}$ is **positive definite**, i.e., $\mathbf{A}^* = \mathbf{A} > 0$.

Property. $(\mathbf{x}, \mathbf{y}) \equiv \mathbf{y}^* \mathbf{A}^{-1} \mathbf{x}$ is an inner product:
the $\mathbf{A}^{-1}$ **inner product**.

Replace standard inner product by the $\mathbf{A}^{-1}$ inner product.

$\mathbf{r}^* \mathbf{c}_1 \rightsquigarrow \mathbf{r}^* \mathbf{A}^{-1} \mathbf{c}_1 = \mathbf{r}^* \mathbf{r} = \|\mathbf{r}\|_2^2$ **Norm r comes for free!**

$\mathbf{c}^* \mathbf{c} \rightsquigarrow \mathbf{c}^* \mathbf{A}^{-1} \mathbf{c} = \mathbf{c}^* \mathbf{u}$ **No $\mathbf{A}^{-1}$ needed!**

$\mathbf{r}_k \perp \mathbf{A}\mathbf{r}_j \rightsquigarrow \mathbf{r}_k \perp_{\mathbf{A}^{-1}} \mathbf{A}\mathbf{r}_j \Leftrightarrow \mathbf{r}_k \perp \mathbf{r}_j$: **orthogonal residuals**.

$\mathbf{A}^* = \mathbf{A} > 0$ Conjugate Gradient

Suppose $\mathbf{A}$ is **positive definite**, i.e., $\mathbf{A}^* = \mathbf{A} > 0$.

Property. $(\mathbf{x}, \mathbf{y}) \equiv \mathbf{y}^* \mathbf{A}^{-1} \mathbf{x}$ is an inner product:
the $\mathbf{A}^{-1}$ **inner product**.

Replace standard inner product by the $\mathbf{A}^{-1}$ inner product.

$\mathbf{r}^* \mathbf{c}_1 \rightsquigarrow \mathbf{r}^* \mathbf{A}^{-1} \mathbf{c}_1 = \mathbf{r}^* \mathbf{r} = \|\mathbf{r}\|_2^2$ **Norm r comes for free!**

$\mathbf{c}^* \mathbf{c} \rightsquigarrow \mathbf{c}^* \mathbf{A}^{-1} \mathbf{c} = \mathbf{c}^* \mathbf{u}$ **No $\mathbf{A}^{-1}$ needed!**

$\mathbf{r}_k \perp \mathbf{A}\mathbf{r}_j \rightsquigarrow \mathbf{r}_k \perp_{\mathbf{A}^{-1}} \mathbf{A}\mathbf{r}_j \Leftrightarrow \mathbf{r}_k \perp \mathbf{r}_j$: **orthogonal residuals**.

Additional saving of

1 DOT (norm $\mathbf{r}$ for free) and 1 AXPY $\rightsquigarrow$ CG

$$\mathbf{A}^* = \mathbf{A} > 0$$

```
Select x, k_max, tol
Compute r = b - Ax
Set u = 0, c = 0, ρ = 1
for k = 0, 1, ..., k_max do
    break if ||r||_2 <= tol
    c_1 = Ar
    σ = -ρ, ρ = r*r, β = ρ/σ
    u ← r - β u
    c ← c_1 - β c = Au
    σ = c*u,  α = ρ/σ
    x ← x + α u
    r ← r - α c
end for
```

Select $\mathbf{x}$, $k_{\max}$, tol

Compute $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$

Set $\mathbf{u} = \mathbf{0}$, $\mathbf{c} = \mathbf{0}$, $\rho = 1$

for $k = 0, 1, \ldots, k_{\max}$ do

  break if $\|\mathbf{r}\|_2 \leq tol$

  $\mathbf{c}_1 = \mathbf{A}\mathbf{r}$

  $\sigma = -\rho$, $\rho = \mathbf{r}^*\mathbf{r}$, $\beta = \rho/\sigma$

  $\mathbf{u} \leftarrow \mathbf{r} - \beta\,\mathbf{u}$

  $\mathbf{c} \leftarrow \mathbf{c}_1 - \beta\,\mathbf{c} = \mathbf{A}\mathbf{u}$

  $\sigma = \mathbf{c}^*\mathbf{u}$, $\alpha = \rho/\sigma$

  $\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}$

  $\mathbf{r} \leftarrow \mathbf{r} - \alpha\,\mathbf{c}$

end for

$\mathbf{A}^* = \mathbf{A} > 0$

```
Select x, k_max, tol
Compute r = b - Ax
Set u = 0, c = 0, ρ = 1
for k = 0, 1, ..., k_max do
    break if ||r||_2 ≤ tol

    σ = -ρ, ρ = r*r, β = ρ/σ
    u ← r - β u
    c = Au
    σ = c*u, α = ρ/σ
    x ← x + α u
    r ← r - α c
end for
```

[Hestenes Stieffel 52]

$\mathbf{A}^* = \mathbf{A} > 0$ Conjugate Gradient

```
Select x, k_max, tol
Compute r = b - Ax
Set u = 0, ρ = 1
for k = 0, 1, ..., k_max do
    break if ||r||_2 ≤ tol
    σ = -ρ, ρ = r*r, β = ρ/σ
    u ← r - β u
    c = Au
    σ = c*u, α = ρ/σ
    x ← x + α u
    r ← r - α c
end for
```

Select $\mathbf{x}$, $k_{\max}$, tol

Compute $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$

Set $\mathbf{u} = \mathbf{0}$, $\rho = 1$

for $k = 0, 1, \ldots, k_{\max}$ do

break if $\|\mathbf{r}\|_2 \le tol$

$\sigma = -\rho$, $\rho = \mathbf{r}^*\mathbf{r}$, $\beta = \rho/\sigma$

$\mathbf{u} \leftarrow \mathbf{r} - \beta\,\mathbf{u}$

$\mathbf{c} = \mathbf{A}\mathbf{u}$

$\sigma = \mathbf{c}^*\mathbf{u}$, $\alpha = \rho/\sigma$

$\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}$

$\mathbf{r} \leftarrow \mathbf{r} - \alpha\,\mathbf{c}$

end for