Introduction to Arnoldi method

SF2524 - Matrix Computations for Large-scale Systems

Main eigenvalue algorithms in this course

- Fundamental eigenvalue techniques (Lecture 1)
- Arnoldi method (Lecture 2-3).
- QR-method (Lecture 9-10).

Main eigenvalue algorithms in this course

- Fundamental eigenvalue techniques (Lecture 1)
- Arnoldi method (Lecture 2-3).
  Typically suitable when
  - we are interested in a small number of eigenvalues,
  - the matrix is large and sparse
  - Solvable size on current desktop $m \sim 10^6$ (depending on structure)
- QR-method (Lecture 9-10).
  Typically suitable when
  - we want to compute all eigenvalues,
  - the matrix does not have any particular easy structure.
  - Solvable size on current desktop $m \sim 1000$.

Agenda lecture 2

- Introduction to Arnoldi method
- Gram-Schmidt - efficiency and roundoff errors
- Derivation of Arnoldi method
- (Next lecture: Convergence characterization)

Idea of Arnoldi method (slide 1/3)

Eigenvalue problem

$$Ax = \lambda x.$$

Idea of Arnoldi method (slide 1/3)

Eigenvalue problem

$$Ax = \lambda x.$$

Suppose $Q = [q_1, \ldots, q_m] \in \mathbb{R}^{n \times m}$ and

$$x \in \mathrm{span}(q_1, \ldots, q_m)$$

Idea of Arnoldi method (slide 1/3)

Eigenvalue problem

$$Ax = \lambda x.$$

Suppose $Q = [q_1, \ldots, q_m] \in \mathbb{R}^{n \times m}$ and

$$x \in \operatorname{span}(q_1, \ldots, q_m)$$

$\Rightarrow$ There exists $z \in \mathbb{R}^m$ such that

$$AQz = \lambda Qz$$

Idea of Arnoldi method (slide 1/3)

Eigenvalue problem

$$Ax = \lambda x.$$

Suppose $Q = [q_1, \ldots, q_m] \in \mathbb{R}^{n \times m}$ and

$$x \in \text{span}(q_1, \ldots, q_m)$$

$\Rightarrow$ There exists $z \in \mathbb{R}^m$ such that

$$AQz = \lambda Qz$$

and

$$Q^T AQz = \lambda Q^T Qz$$

Idea of Arnoldi method (slide 1/3)

Eigenvalue problem

$$Ax = \lambda x.$$

Suppose $Q = [q_1, \ldots, q_m] \in \mathbb{R}^{n\times m}$ and

$$x \in \mathrm{span}(q_1, \ldots, q_m)$$

$\Rightarrow$ There exists $z \in \mathbb{R}^m$ such that

$$AQz = \lambda Qz$$

and Q orthogonal

$$Q^T AQz = \lambda Q^T Qz = \lambda z$$

Idea of Arnoldi method (slide 2/3)

Rayleigh-Ritz procedure

Idea of Arnoldi method (slide 2/3)

Rayleigh-Ritz procedure

Let $Q \in \mathbb{R}^{n \times m}$ be orthogonal basis of some subspace.

Idea of Arnoldi method (slide 2/3)

Rayleigh-Ritz procedure

Let $Q \in \mathbb{R}^{n \times m}$ be orthogonal basis of some subspace. Solutions (μ, z) to the eigenvalue problem

$$Q^T AQz = \mu z$$

are called *Ritz pairs*.

Idea of Arnoldi method (slide 2/3)

Rayleigh-Ritz procedure

Let $Q \in \mathbb{R}^{n\times m}$ be orthogonal basis of some subspace. Solutions (μ, z) to the eigenvalue problem

$$Q^T AQz = \mu z$$

are called *Ritz pairs*. The procedure (Rayleigh-Ritz)

- returns an exact eigenvalue if $x \in \operatorname{span}(q_1, \ldots, q_m)$,

Idea of Arnoldi method (slide 2/3)

Rayleigh-Ritz procedure

Let $Q \in \mathbb{R}^{n \times m}$ be orthogonal basis of some subspace. Solutions (μ, z) to the eigenvalue problem

$$Q^T AQz = \mu z$$

are called *Ritz pairs*. The procedure (Rayleigh-Ritz)

- returns an exact eigenvalue if $x \in \operatorname{span}(q_1, \ldots, q_m)$, and
- approximates eigenvalues of A "well" if

$$x \approx \tilde{x} \in \operatorname{span}(q_1, \ldots, q_m).$$

Idea of Arnoldi method (slide 2/3)

Rayleigh-Ritz procedure

Let $Q \in \mathbb{R}^{n\times m}$ be orthogonal basis of some subspace. Solutions (μ, z) to the eigenvalue problem

$$Q^T AQz = \mu z$$

are called *Ritz pairs*. The procedure (Rayleigh-Ritz)

- returns an exact eigenvalue if $x \in \operatorname{span}(q_1, \ldots, q_m)$, and
- approximates eigenvalues of A "well" if

$$x \approx \tilde{x} \in \operatorname{span}(q_1, \ldots, q_m).$$

What is a good subspace $\operatorname{span}(Q)$?

Idea of Arnoldi method (slide 2/3)

Rayleigh-Ritz procedure

Let $Q \in \mathbb{R}^{n\times m}$ be orthogonal basis of some subspace. Solutions (μ, z) to the eigenvalue problem

$$Q^T AQz = \mu z$$

are called *Ritz pairs*. The procedure (Rayleigh-Ritz)

- returns an exact eigenvalue if $x \in \mathrm{span}(q_1, \ldots, q_m)$, and
- approximates eigenvalues of A "well" if

$$x \approx \tilde{x} \in \mathrm{span}(q_1, \ldots, q_m).$$

What is a good subspace $\mathrm{span}(Q)$?

Idea of Arnoldi method (slide 3/3)

Recall: Power method approximates the largest eigenvalue well.

Idea of Arnoldi method (slide 3/3)

Recall: Power method approximates the largest eigenvalue well.

Definition: Krylov matrix / subspace

$$K_m(A, q_1) := (q_1, Aq_1, \dots A^{m-1}q_1)$$

$$\mathcal{K}_m(A, q_1) := \text{span}(q_1, Aq_1, \dots A^{m-1}q_1).$$

Idea of Arnoldi method (slide 3/3)

Recall: Power method approximates the largest eigenvalue well.

Definition: Krylov matrix / subspace

$$K_m(A, q_1) := (q_1, Aq_1, \dots A^{m-1}q_1)$$

$$\mathcal{K}_m(A, q_1) := \operatorname{span}(q_1, Aq_1, \dots A^{m-1}q_1).$$

Justification of Arnoldi method

- Use Rayleigh-Ritz on $Q = (q_1, \dots, q_m)$ and $Q^T Q = I$, where

$$\operatorname{span}(q_1, \dots, q_m) = \mathcal{K}_m(A, q_1)$$

Idea of Arnoldi method (slide 3/3)

Recall: Power method approximates the largest eigenvalue well.

Definition: Krylov matrix / subspace

$$K_m(A, q_1) := (q_1, Aq_1, \ldots A^{m-1}q_1)$$

$$\mathcal{K}_m(A, q_1) := \mathrm{span}(q_1, Aq_1, \ldots A^{m-1}q_1).$$

Justification of Arnoldi method

- Use Rayleigh-Ritz on $Q = (q_1, \ldots, q_m)$ and $Q^T Q = I$, where

$$\mathrm{span}(q_1, \ldots, q_m) = \mathcal{K}_m(A, q_1)$$

- Arnoldi method is a “clever” procedure to construct $H_m = Q^T AQ$.

Idea of Arnoldi method (slide 3/3)

Recall: Power method approximates the largest eigenvalue well.

Definition: Krylov matrix / subspace

$$K_m(A, q_1) := (q_1, Aq_1, \dots A^{m-1}q_1)$$

$$\mathcal{K}_m(A, q_1) := \mathrm{span}(q_1, Aq_1, \dots A^{m-1}q_1).$$

Justification of Arnoldi method

- Use Rayleigh-Ritz on $Q = (q_1, \dots, q_m)$ and $Q^T Q = I$, where

$$\mathrm{span}(q_1, \dots, q_m) = \mathcal{K}_m(A, q_1)$$

- Arnoldi method is a "clever" procedure to construct $H_m = Q^T AQ$.
- "Clever": We expand Q with one row in each iteration
  $\Rightarrow$ Iterate until we are happy.

Arnoldi method graphically

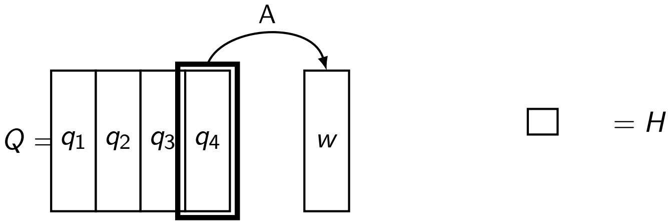
Graphical illustration of algorithm:

$$Q = \begin{bmatrix} q_1 & q_2 & q_3 & q_4 \end{bmatrix}$$

$$\boxed{} \qquad = H$$

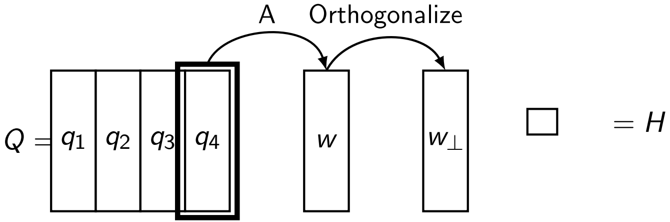
Arnoldi method graphically

Graphical illustration of algorithm:



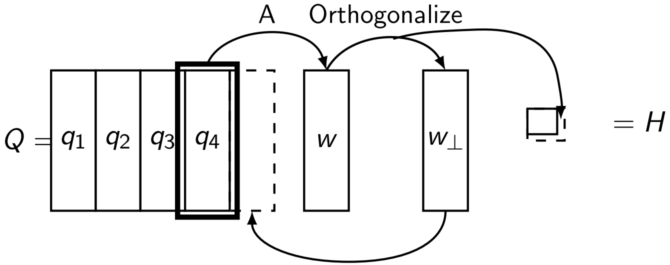
Arnoldi method graphically

Graphical illustration of algorithm:



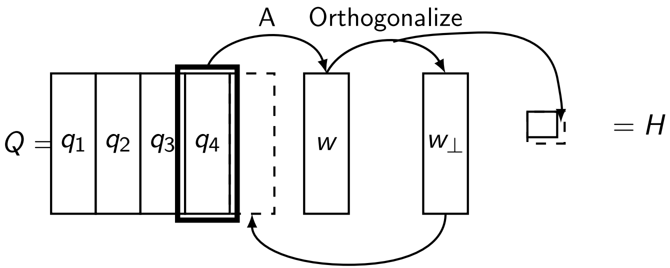
Arnoldi method graphically

Graphical illustration of algorithm:



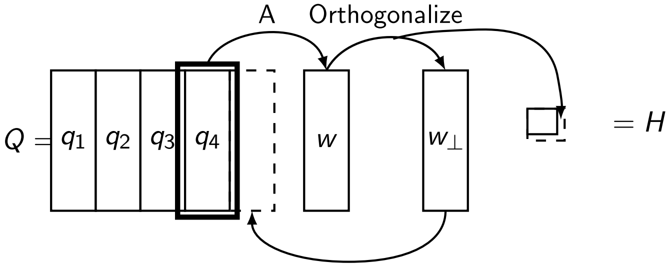
Arnoldi method graphically

Graphical illustration of algorithm:



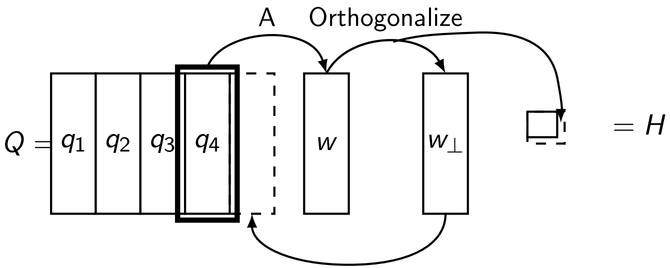
Arnoldi method graphically

Graphical illustration of algorithm:



Arnoldi method graphically

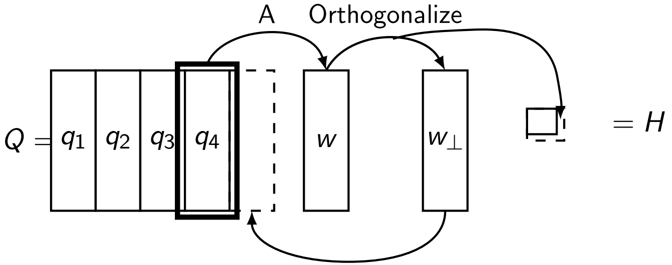
Graphical illustration of algorithm:



After iteration: Take eigenvalues of H as approximate eigenvalues.

Arnoldi method graphically

Graphical illustration of algorithm:

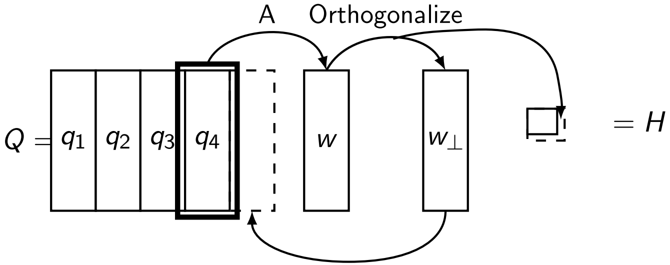


After iteration: Take eigenvalues of H as approximate eigenvalues.

* show arnoldi.m and Hessenberg matrix in matlab *

Arnoldi method graphically

Graphical illustration of algorithm:



After iteration: Take eigenvalues of H as approximate eigenvalues.

* show arnoldi.m and Hessenberg matrix in matlab *

We will now...

(1) derive a good orthogonalization procedure: variants of Gram-Schmidt,

(2) show that Arnoldi generates a Rayleigh-Ritz approximation,

(3) characterize the convergence (next lecture).

Gram-Schmidt methods (for numerical computations)

in particular for the Arnoldi method

Problem ("easier" problem)

Given:

- $Q_m \in \mathbb{R}^{n \times m}$ orthogonal matrix
- $w \in \mathbb{R}^n$ satisfying $w \notin \operatorname{span}(Q_m)$

Problem ("easier" problem)

Given:

- $Q_m \in \mathbb{R}^{n \times m}$ orthogonal matrix
- $w \in \mathbb{R}^n$ satisfying $w \notin \text{span}(Q_m)$

Compute

- $h_1, \ldots, h_m, \beta \in \mathbb{R}$
- $q_{m+1} \in \mathbb{R}^n$

Problem ("easier" problem)

Given:

- $Q_m \in \mathbb{R}^{n\times m}$ orthogonal matrix
- $w \in \mathbb{R}^n$ satisfying $w \notin \operatorname{span}(Q_m)$

Compute

- $h_1, \ldots, h_m, \beta \in \mathbb{R}$
- $q_{m+1} \in \mathbb{R}^n$

such that

(a) $Q_{m+1} = [Q_m, q_{m+1}]$ is orthogonal

(b) $\operatorname{span}(q_1, \ldots, q_{m+1}) = \operatorname{span}(q_1, \ldots, q_m, w)$

(c) $w = h_1 q_1 + \cdots + h_m q_m + \beta q_{m+1}$; and

Problem ("easier" problem)

Given:

- $Q_m \in \mathbb{R}^{n\times m}$ orthogonal matrix
- $w \in \mathbb{R}^n$ satisfying $w \notin \mathrm{span}(Q_m)$

Compute

- $h_1, \ldots, h_m, \beta \in \mathbb{R}$
- $q_{m+1} \in \mathbb{R}^n$

such that

(a) $Q_{m+1} = [Q_m, q_{m+1}]$ is orthogonal

(b) $\mathrm{span}(q_1, \ldots, q_{m+1}) = \mathrm{span}(q_1, \ldots, q_m, w)$

(c) $w = h_1 q_1 + \cdots + h_m q_m + \beta q_{m+1}$; and

Solution:

1. Compute a vector y which is orthogonal to Q_k
2. Normalize vector y

- 1. Element of $\text{span}(q_1, \ldots, q_k)$ can be expressed as Qh:

- 1. Element of $\text{span}(q_1, \ldots, q_k)$ can be expressed as Qh:

$$y = w - Qh.$$

- 1. Element of $\text{span}(q_1, \ldots, q_k)$ can be expressed as Qh:

$$y = w - Qh. \tag{*}$$

$\Rightarrow \text{span}(q_1, \ldots, q_k, w) = \text{span}(q_1, \ldots, q_k, y)$

- 1. Element of $\mathrm{span}(q_1, \ldots, q_k)$ can be expressed as Qh:

$$y = w - Qh. \qquad (*)$$

$\Rightarrow \mathrm{span}(q_1, \ldots, q_k, w) = \mathrm{span}(q_1, \ldots, q_k, y)$
Select h such that y orthogonal to $q_1, \ldots, q_k$:

- 1. Element of $\mathrm{span}(q_1, \ldots, q_k)$ can be expressed as Qh:

$$y = w - Qh. \qquad (*)$$

$\Rightarrow \mathrm{span}(q_1, \ldots, q_k, w) = \mathrm{span}(q_1, \ldots, q_k, y)$
Select h such that y orthogonal to $q_1, \ldots, q_k$:

$$\begin{aligned} 0 &= q_1^T y \\ &\vdots \\ 0 &= q_k^T y \end{aligned}$$

- 1. Element of $\text{span}(q_1, \dots, q_k)$ can be expressed as Qh:

$$y = w - Qh. \tag{*}$$

$\Rightarrow \text{span}(q_1, \dots, q_k, w) = \text{span}(q_1, \dots, q_k, y)$
Select h such that y orthogonal to $q_1, \dots, q_k$:

$$\begin{aligned} 0 &= q_1^T y \\ &\vdots \\ 0 &= q_k^T y \end{aligned}$$

can be expressed in vectors as

$$0 = Q^T y$$

- 1. Element of $\mathrm{span}(q_1, \ldots, q_k)$ can be expressed as Qh:

$$y = w - Qh. \tag{*}$$

$\Rightarrow \mathrm{span}(q_1, \ldots, q_k, w) = \mathrm{span}(q_1, \ldots, q_k, y)$
Select h such that y orthogonal to $q_1, \ldots, q_k$:

$$\begin{aligned} 0 &= q_1^T y \\ &\vdots \\ 0 &= q_k^T y \end{aligned}$$

can be expressed in vectors as

$$0 = Q^T y = Q^T (w - Qh)$$

- 1. Element of $\text{span}(q_1, \dots, q_k)$ can be expressed as Qh:

$$y = w - Qh. \tag{*}$$

$\Rightarrow \text{span}(q_1, \dots, q_k, w) = \text{span}(q_1, \dots, q_k, y)$
Select h such that y orthogonal to $q_1, \dots, q_k$:

$$\begin{aligned} 0 &= q_1^T y \\ &\vdots \\ 0 &= q_k^T y \end{aligned}$$

can be expressed in vectors as

$$0 = Q^T y = Q^T (w - Qh) = Q^T w - Q^T Qh = Q^T w - h$$

- 1. Element of $\mathrm{span}(q_1, \ldots, q_k)$ can be expressed as Qh:

$$y = w - Qh. \tag{*}$$

$\Rightarrow \mathrm{span}(q_1, \ldots, q_k, w) = \mathrm{span}(q_1, \ldots, q_k, y)$
Select h such that y orthogonal to $q_1, \ldots, q_k$:

$$\begin{aligned} 0 &= q_1^T y \\ &\vdots \\ 0 &= q_k^T y \end{aligned}$$

can be expressed in vectors as

$$0 = Q^T y = Q^T (w - Qh) = Q^T w - Q^T Qh = Q^T w - h$$

$\Rightarrow h = Q^T w.$

- 1. Element of $\mathrm{span}(q_1, \ldots, q_k)$ can be expressed as Qh:

$$y = w - Qh. \tag{*}$$

$\Rightarrow \mathrm{span}(q_1, \ldots, q_k, w) = \mathrm{span}(q_1, \ldots, q_k, y)$
Select h such that y orthogonal to $q_1, \ldots, q_k$:

$$\begin{aligned} 0 &= q_1^T y \\ &\vdots \\ 0 &= q_k^T y \end{aligned}$$

can be expressed in vectors as

$$0 = Q^T y = Q^T(w - Qh) = Q^T w - Q^T Qh = Q^T w - h$$

$\Rightarrow h = Q^T w.$

2. Let $\beta = \|y\|$ and set $q_{k+1} = y/\beta$

- 1. Element of $\text{span}(q_1, \ldots, q_k)$ can be expressed as Qh:

$$y = w - Qh. \tag{*}$$

$\Rightarrow \text{span}(q_1, \ldots, q_k, w) = \text{span}(q_1, \ldots, q_k, y)$
Select h such that y orthogonal to $q_1, \ldots, q_k$:

$$\begin{aligned} 0 &= q_1^T y \\ &\vdots \\ 0 &= q_k^T y \end{aligned}$$

can be expressed in vectors as

$$0 = Q^T y = Q^T (w - Qh) = Q^T w - Q^T Qh = Q^T w - h$$

$\Rightarrow h = Q^T w.$

2. Let $\beta = \|y\|$ and set $q_{k+1} = y/\beta$

The construction implies that (*) reduces to

$$y = w - Qh$$

- 1. Element of $\text{span}(q_1, \ldots, q_k)$ can be expressed as Qh:

$$y = w - Qh. \qquad (*)$$

$\Rightarrow \text{span}(q_1, \ldots, q_k, w) = \text{span}(q_1, \ldots, q_k, y)$

Select h such that y orthogonal to $q_1, \ldots, q_k$:

$$\begin{aligned} 0 &= q_1^T y \\ &\vdots \\ 0 &= q_k^T y \end{aligned}$$

can be expressed in vectors as

$$0 = Q^T y = Q^T(w - Qh) = Q^T w - Q^T Qh = Q^T w - h$$

$\Rightarrow h = Q^T w.$

2. Let $\beta = \|y\|$ and set $q_{k+1} = y/\beta$

The construction implies that (*) reduces to

$$\begin{aligned} y &= w - Qh \\ w &= Qh + y \end{aligned}$$

- 1. Element of $\operatorname{span}(q_1, \ldots, q_k)$ can be expressed as Qh:

$$y = w - Qh. \tag{*}$$

$\Rightarrow \operatorname{span}(q_1, \ldots, q_k, w) = \operatorname{span}(q_1, \ldots, q_k, y)$
Select h such that y orthogonal to $q_1, \ldots, q_k$:

$$\begin{aligned} 0 &= q_1^T y \\ &\vdots \\ 0 &= q_k^T y \end{aligned}$$

can be expressed in vectors as

$$0 = Q^T y = Q^T (w - Qh) = Q^T w - Q^T Q h = Q^T w - h$$

$\Rightarrow h = Q^T w.$

2. Let $\beta = \|y\|$ and set $q_{k+1} = y/\beta$

The construction implies that (*) reduces to

$$\begin{aligned} y &= w - Qh \\ w &= Qh + y = h_1 q_1 + \cdots h_k q_k + \beta q_{k+1} \end{aligned}$$

Classical Gram-Schmidt

```
>> h=Q'*w
>> y=w-Q*h
>> beta=norm(y)
>> qnew=y/beta
>> Qnew=[Q,qnew]
```

Classical Gram-Schmidt

```
>> h=Q'*w
>> y=w-Q*h
>> beta=norm(y)
>> qnew=y/beta
>> Qnew=[Q,qnew]
```

* Show that it often works *

* Show a case where it doesn't work *

* Modified GS on black board *

* Double GS on black board *

```matlab
function [Q,H]=arnoldi(A,b,m)
% [Q,H]=arnoldi(A,b,m)
% A simple implementation of the Arnoldi method.
% The algorithm will return an Arnoldi "factorization":
%   Q*H(1:m+1,1:m)-A*Q(:,1:m)=0
% where Q is an orthogonal basis of the Krylov subspace
% and H a Hessenberg matrix.
%
    n=length(b);;
    Q=zeros(n,m+1);
    Q(:,1)=b/norm(b);
    for k=1:m
        w=A*(Q(:,k)); % Matrix-vector product
                      % with last element
        %%% Orthogonalize w against columns of Q
        % correct sol of HW1.2b
        [h,beta,worth]=hw1_good_gs(Q,w,k);
        %%% Put Gram-Schmidt coefficients into H
        H(1:(k+1),k)=[h;beta];
        %%% normalize
        Q(:,k+1)=worth/beta;
    end
end
```

-:--- arnoldi.m   All L8   (MATLAB +1 yas Abbrev Fill)

Convergence theory of Arnoldi's method for eigenvalue problems

Example of convergence theory of the Arnoldi method for eigenvalue problems:

Theorem (Jia, SIAM J. Matrix. Anal. Appl. 1995)

Let Q_m and H_m be generated by the Arnoldi method and suppose $\lambda_i^{(m)}$ is an eigenvalue of H_m. Assume that $\ell_i = 1$ and the associated value $\|(I - Q_m Q_m^T)x_i\|$ is sufficiently small. Let $P_i^{(m)}$ be the spectral projector associated with $\lambda_i^{(m)}$. Then,

$$|\lambda_i^{(m)} - \lambda_i| \leq \|P_i^{(m)}\| \gamma_m \frac{\|(I - Q_m Q_m^T)x_i\|}{\|Q_m Q_m^T x_i\|} + \mathcal{O}\left(\frac{\|(I - Q_m Q_m^T)x_i\|^2}{\|Q_m Q_m^T x_i\|^2}\right)$$

Example of convergence theory of the Arnoldi method for eigenvalue problems:

Theorem (Jia, SIAM J. Matrix. Anal. Appl. 1995)

Let Q_m and H_m be generated by the Arnoldi method and suppose $\lambda_i^{(m)}$ is an eigenvalue of H_m. Assume that $\ell_i = 1$ and the associated value $\|(I - Q_m Q_m^T)x_i\|$ is sufficiently small. Let $P_i^{(m)}$ be the spectral projector associated with $\lambda_i^{(m)}$. Then,

$$|\lambda_i^{(m)} - \lambda_i| \leq \|P_i^{(m)}\|\gamma_m \frac{\|(I - Q_m Q_m^T)x_i\|}{\|Q_m Q_m^T x_i\|} + \mathcal{O}\left(\frac{\|(I - Q_m Q_m^T)x_i\|^2}{\|Q_m Q_m^T x_i\|^2}\right)$$

The theorem is not a part of the course. In this course we will gain qualitative understanding by bounding

$$\|(I - Q_m Q_m^T)x_i\|.$$