Theoretical and numerical study of a preconditioner for saddle-point systems

Anne Cassier

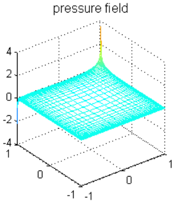
October 2, 2014

Supervisors : Serge Gratton, Xavier Vasseur, Selime Gürol

Outline

The Stokes problem

$$\begin{cases} -\nu \Delta u + \text{grad}\, p = f \\ \text{div}\, u = 0 \end{cases} \text{ in } \Omega$$



- u velocity, p pressure
- used in computational fluid dynamics, in climate forecasting, ...

When discretizing the equations :

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}$$

$A \in \mathbb{R}^{n \times n}$ symmetric positive definite, $B \in \mathbb{R}^{m \times n}$.

With a time-dependent problem, we have a sequence of saddle-point

How to solve this system ?

- Direct method : compute the exact solution by factorizing the matrix
  - need access to matrix coefficients
  - can be expensive when large-scale systems
- Iterative methods : build a sequence of vectors (x_k) converging to the exact solution of the systems
  - does not need the access to matrix coefficients
  - preconditioning to accelerate the convergence of the method

In our case, we consider a large-scale system solved by a preconditioned Krylov subspace method.

Preconditioning

To precondition a system we solve a system with the same solution, supposing this system is easier and quicker to solve than the initial one.

An efficient preconditioning matrix must be :

- low cost to compute and low cost in memory
- low cost to use
- accelerate the convergence of the method

The preconditioned system must have less distinct eigenvalues than the initial one.

It is difficult to preconditioned a saddle-point system because it is an indefinite system with eigenvalues in the complex plane.

Well-known preconditioners

$$\mathcal{A} = \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$$

- Block preconditioners ([Murphy, Golub, Wathen, 2000], [Klawonn, 1998], [Ruiz, Sartenaer, Tannier, 2014])

$$\mathcal{P}_d = \begin{pmatrix} A & 0 \\ 0 & S \end{pmatrix}, \mathcal{P}_t = \begin{pmatrix} A & B^T \\ 0 & S \end{pmatrix}$$

  $S = -B^T A^{-1} B$ the Schur complement.
- Krylov method converge in at most 3 or 2 iterations
- S can be expensive to compute
- Constraint preconditioner ([Keller, Gould, Wathen, 2000])

$$\mathcal{P}_c = \begin{pmatrix} G & B^T \\ B & 0 \end{pmatrix}$$

  $S = -BG^{-1}B^T$ the Schur complement of $\mathcal{P}_c$
- 1 is eigenvalue with multiplicity 2m

Second-level preconditioning

- We solve a saddle-point system with a preconditioned Krylov method (*first-level preconditioner*)
- We use an approximation of a well-known preconditioner
- The preconditioned system has a cluster of eigenvalues around isolated eigenvalues,
- other eigenvalues out of this cluster

We want to find another preconditioner (*second-level preconditioner*) to reduce the number of eigenvalues out of the cluster.

Case of a symmetric positive definite system [Gratton, Sartenaer, Tshimanga, 2011], [Morales, Nocedal, 2000]

- We want to solve the system $Ax = b$ with A symmetric positive definite
- We have secant equations $y_k = As_k$
- We have a preconditioner M (*first-level preconditioner*)
- We use the BFGS update formula to compute the preconditioner (*second-level preconditioner*) :

$$H_k = [I_n - S(S^T AS)^{-1} S^T A] M [I_n - AS(S^T AS)^{-1} S^T] + S(S^T AS)^{-1} S^T.$$

- the update can be seen as variational approach
- at least k eigenvalues equal to 1
- the other part of the spectrum of AH_k is not expanded

Objectives

- We want to solve the saddle-point system :

$$\underbrace{\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}}_{\mathcal{A}} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix}$$

  with a Krylov method.
- We can choose one of the well-known preconditioners according to the structure of the problem (*first-level preconditioner*)

Objectives

- We want to solve the saddle-point system :

$$\underbrace{\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}}_{\mathcal{A}} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix}$$

  with a Krylov method.
- We can choose one of the well-known preconditioners according to the structure of the problem (*first-level preconditioner*)
- We want to modify the existing preconditioner from secant equations (*second-level preconditioner*)

Objectives

- We want to solve the saddle-point system :

$$\underbrace{\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}}_{\mathcal{A}} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix}$$

  with a Krylov method.
- We can choose one of the well-known preconditioners according to the structure of the problem (*first-level preconditioner*)
- We want to modify the existing preconditioner from secant equations (*second-level preconditioner*)

Objectives :

Find a second-level preconditioner using a variational approach with a Frobenius norm

Outline

Second-level preconditioning

We have the matrix-vector product :

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix}.$$

Second-level preconditioning

We have the matrix-vector product :

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix}.$$

$\begin{bmatrix} \tilde{x}^T & \tilde{y}^T \end{bmatrix}^T$:

- can be an eigenvector of $\mathcal{A}$,
- can satisfy the Arnoldi relation,
- can be an Harmonic-Ritz vector of $\mathcal{A}$.

Second-level preconditioning

We have the matrix-vector product :

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix}.$$

$\begin{bmatrix} \tilde{x}^T & \tilde{y}^T \end{bmatrix}^T$:

- can be an eigenvector of $\mathcal{A}$,
- can satisfy the Arnoldi relation,
- can be an Harmonic-Ritz vector of $\mathcal{A}$.

We have the first-level preconditioner $\mathcal{P}_0 = \begin{pmatrix} \tilde{A} & \tilde{B}^T \\ \tilde{B} & 0 \end{pmatrix}$. We want to find the second-level preconditioner $\mathcal{P}_1$ such that

$$\underbrace{\begin{pmatrix} \tilde{A} & \tilde{B}^T \\ \tilde{B} & 0 \end{pmatrix} + \begin{pmatrix} \Delta A & \Delta B^T \\ \Delta B & 0 \end{pmatrix}}_{\mathcal{P}_1} \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix}.$$

How to choose $\Delta A, \Delta B$?

Second-level preconditioning

$$\begin{pmatrix} \tilde{A} + \Delta A & \tilde{B}^T + \Delta B^T \\ \tilde{B} + \Delta B & 0 \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} \tilde{b} \\ \tilde{c} \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \Delta A \tilde{x} + \Delta B^T \tilde{y} = b - \tilde{A}\tilde{x} - \tilde{B}^T \tilde{y} \\ \Delta B \tilde{x} = c - \tilde{B}\tilde{x} \end{pmatrix}.$$

We define r_b, r_c, w as :

$$\begin{aligned} r_b &= b - \tilde{A}\tilde{x} - \tilde{B}^T \tilde{y}, \\ r_c &= c - \tilde{B}\tilde{x}, \\ w &= r_b - \Delta B^T \tilde{y} \end{aligned}$$

We have :

$$\begin{aligned} \Delta A \tilde{x} + \Delta B^T \tilde{y} &= r_b \\ \Delta B \tilde{x} &= r_c \end{aligned}$$

Second-level preconditioning

$$\begin{pmatrix} \tilde{A} + \Delta A & \tilde{B}^T + \Delta B^T \\ \tilde{B} + \Delta B & 0 \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} \tilde{b} \\ \tilde{c} \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \Delta A \tilde{x} + \Delta B^T \tilde{y} = b - \tilde{A}\tilde{x} - \tilde{B}^T \tilde{y} \\ \Delta B \tilde{x} = c - \tilde{B}\tilde{x} \end{pmatrix}.$$

We define r_b, r_c, w as :

$$r_b = b - \tilde{A}\tilde{x} - \tilde{B}^T \tilde{y},$$
$$r_c = c - \tilde{B}\tilde{x},$$
$$w = r_b - \Delta B^T \tilde{y}$$

We have :

$$\Delta A \tilde{x} + \Delta B^T \tilde{y} = r_b$$
$$\Delta B \tilde{x} = r_c$$

$\rightarrow$ We want to find the least Frobenius norm update!

- Any solution of the equations, ΔA and ΔB can be written as :

$$\Delta B = r_c \tilde{x}^\dagger + Z P_{\tilde{x}}^\perp,$$
$$\Delta A = w \tilde{x}^\dagger + (w \tilde{x}^\dagger)^T - (\tilde{x}^\dagger)^T w^T P_{\tilde{x}} + P_{\tilde{x}}^\perp T P_{\tilde{x}}^\perp$$

  with $T \in \mathbb{R}^{n\times n}, Z \in \mathbb{R}^{m\times m}$.
- We want to find Z, T that minimize

$$\min_{Z,T} \|\Delta A\|_F^2 + \|\Delta B\|_F^2$$

- The least-norm solution is :

$$T = 0, Z = 2P_{\tilde{x}}^\perp (r_b \tilde{x}^\dagger)(\tilde{y}\tilde{x}^\dagger)^T (I_m + 2(\tilde{y}\tilde{x}^\dagger)(\tilde{y}\tilde{x}^\dagger)^T)^{-1}$$

The second-level preconditioner can be written as

$$\mathcal{P}_1 = \mathcal{P}_0 + FG^T, F, G^T \in \mathbb{R}^{(n+m)\times 5}$$

Using the Sherman-Morrison formula we have :

$$\mathcal{P}_1^{-1} = \mathcal{P}_0^{-1} - \mathcal{P}_0^{-1} F \left(I_5 + G^T \mathcal{P}_0^{-1} F \right)^{-1} G^T \mathcal{P}_0^{-1}.$$

Outline

Spectral analysis

A clustered spectrum often results in rapid convergence when using GMRES.

We define :

- $\mathcal{A} = \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$, A symmetric positive definite, rank $B = m$,
- $\lambda^A_{min}, \lambda^A_{max}$ the minimal and maximal eigenvalues of A,
- $\sigma_{min}, \sigma_{max}$ the smallest non-zero and the maximum singular values of B.

We know that [Rusten, Winther, 1992] the eigenvalues of $\mathcal{A}$ are contained in $\mathcal{I}_0^- \bigcup \mathcal{I}_0^+ \subset \mathbb{R}$:

$$\mathcal{I}_0^- = \left[\frac{1}{2}\left(\lambda^A_{min} - \sqrt{(\lambda^A_{min})^2 + 4\sigma^2_{max}}\right), \frac{1}{2}\left(\lambda^A_{max} - \sqrt{(\lambda^A_{max})^2 + 4\sigma^2_{min}}\right)\right],$$

$$\mathcal{I}_0^+ = \left[\lambda^A_{min}, \frac{1}{2}\left(\lambda^A_{max} + \sqrt{(\lambda^A_{max})^2 + 4\sigma^2_{max}}\right)\right].$$

Spectral analysis

We suppose :

- $\mathcal{P}_0 = I_{n+m}$
- $\mathcal{P}_1$ is block diagonal symmetric,
- $\begin{bmatrix}\tilde{x}^T & \tilde{y}^T\end{bmatrix}^T = \begin{bmatrix}u^T & v^T\end{bmatrix}^T$ is an eigenvector of $\mathcal{A}$ associated to the eigenvalue λ.

We define P_u, P_v the orthogonal projectors onto the range of u and v. We can deduce that :

$$\mathcal{P}_1 = \begin{pmatrix} I_n + (\lambda - 1)P_u & 0 \\ 0 & I_m + (\lambda - 1)P_v \end{pmatrix}.$$

Spectral analysis

Let's denote μ an eigenvalue of $\mathcal{P}_1^{-1}\mathcal{A}$ and $(x^T y^T)^T$ its associated eigenvector.

Theorem 1

Let's denote $x = x_u + x_{u^\perp}, y = y_v + y_{v^\perp}$,
$x_u \in \text{Span}(u), x_{u^\perp} \in \text{Span}(u)^\perp, y_v \in \text{Span}(v), y_{v^\perp} \in \text{Span}(v)^\perp$. The following equation is satisfied :

$$\mu^2 \left(\|x_{u^\perp}\|_2^2 + \lambda \|x_u\|_2^2 \right) - \mu x^T A x - \|y_{v^\perp}\|_2^2 - \frac{1}{\lambda} \|y_v\|_2^2 = 0$$

Particular case (assuming $x_{u^\perp} = 0, y_{v^\perp} = 0$)

1 and $-\frac{\|v\|_2^2}{\|u\|_2^2}$ are eigenvalues of $\mathcal{P}_1^{-1}\mathcal{A}$.

Spectral analysis

We suppose $\lambda > 0$. Let's denote

- $B\begin{pmatrix} Y & Z \end{pmatrix} = \begin{pmatrix} L & 0 \end{pmatrix}$ an LQ factorization of B, with Z a basis of the nullspace of B.
- $\lambda_{min}^{Y}, \lambda_{max}^{Y}$ the minimal and maximal eigenvalues of $Y^T A Y$.
- $\sigma_{min}^{Y}, \sigma_{max}^{Y}$ the smallest non-zero and the maximum singular values of Y.

Theorem 2

The eigenvalues of $\mathcal{P}_1^{-1}\mathcal{A}$ are contained in $\mathcal{I}_1^- \bigcup \mathcal{I}_1^+ \subset \mathbb{R}$:

$$\mathcal{I}_1^- = \left[\frac{\lambda_{min}^A - \sqrt{(\lambda_{min}^A)^2 + 4(\sigma_{max}^B)^2}}{2\min(1,\lambda)}, \frac{\lambda_{max}^Y - \sqrt{(\lambda_{max}^Y)^2 + 4(\sigma_{min}^B)^2(\sigma_{max}^Y)^2}}{2\max(1,\lambda)(\sigma_{max}^Y)^2} \right],$$

$$\mathcal{I}_1^+ = \left[\frac{\lambda_{min}^A}{\max(1,\lambda)}, \frac{\lambda_{max}^A + \sqrt{(\lambda_{max}^A)^2 + 4(\sigma_{max}^B)^2}}{2\min(1,\lambda)} \right].$$

Numerical illustration

- $A = \begin{pmatrix} 17 & -2 & 0 \\ -2 & 16 & 0 \\ 0 & 0 & 0.8 \end{pmatrix}, B^T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}$
- $I_0^- = [-0.6770; -0.0537]$, $I_0^+ = [0.8; 18.6153]$
- $\mathrm{Sp}(\mathcal{A}) = \{-0.0689; -0.0537; 0.8; 14.5074; 18.6153\}$.

λ	$\mathcal{I}_1^-$	$\mathcal{I}_1^+$	$\mathrm{Sp}(\mathcal{A}\mathcal{P}_1^{-1})$
18.6153			
14.5074			
0.8			
-0.0537			
-0.0689			

Numerical illustration

- $A = \begin{pmatrix} 17 & -2 & 0 \\ -2 & 16 & 0 \\ 0 & 0 & 0.8 \end{pmatrix}, B^T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}$
- $I_0^- = [-0.6770; -0.0537]$, $I_0^+ = [0.8; 18.6153]$
- $\mathrm{Sp}(\mathcal{A}) = \{-0.0689; -0.0537; 0.8; 14.5074; 18.6153\}$.

λ	$\mathcal{I}_1^-$	$\mathcal{I}_1^+$	$\mathrm{Sp}(\mathcal{A}\mathcal{P}_1^{-1})$
18.6153	$[-0.6770; -0.0029]$	$[0.0430; 18.6153]$	-0.0689 ; -0.0029 ; 0.8 ; 1 ; 14.5074
14.5074	$[-0.6770; -0.0037]$	$[0.0551; 18.6153]$	-0.0537 ; -0.0048 ; 0.8 ; 1 ; 18.6153
0.8	$[-0.0826; -0.0537]$	$[0.8; 23.2691]$	-0.0826 ; -0.0560 ; 1 ; 14.5190 ; 18.6196
-0.0537			
-0.0689			

Numerical illustration

- $A = \begin{pmatrix} 17 & -2 & 0 \\ -2 & 16 & 0 \\ 0 & 0 & 0.8 \end{pmatrix}, B^T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}$

- $I_0^- = [-0.6770; -0.0537]$, $I_0^+ = [0.8; 18.6153]$

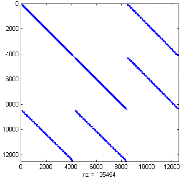
- $\mathrm{Sp}(\mathcal{A}) = \{-0.0689; -0.0537; 0.8; 14.5074; 18.6153\}$.

λ	$\mathcal{I}_1^-$	$\mathcal{I}_1^+$	$\mathrm{Sp}(\mathcal{A}\mathcal{P}_1^{-1})$
18.6153	$[-0.6770; -0.0029]$	$[0.0430; 18.6153]$	-0.0689 ; -0.0029 ; 0.8 ; 1 ; 14.5074
14.5074	$[-0.6770; -0.0037]$	$[0.0551; 18.6153]$	-0.0537 ; -0.0048 ; 0.8 ; 1 ; 18.6153
0.8	$[-0.0826; -0.0537]$	$[0.8; 23.2691]$	-0.0826 ; -0.0560 ; 1 ; 14.5190 ; 18.6196
-0.0537			-346.52 ; -0.0689 ; 0.8 ; 1 ; 14.507
-0.0689			-210.464 ; -0.0537 ; 0.8 ; 1 ; 18.6152

Outline

The Stokes problem

- Stokes problem from Ifiss library
- channel domain
- natural boundary outflow
- stabilization parameter : 1/4
- uniform streamline



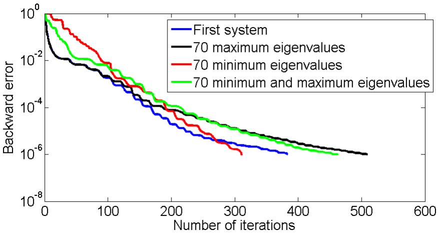
We solve the first system with the preconditioner $\mathcal{P}_0$: $\mathcal{A}_0 = \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$.

We solve a second system with the preconditioner $\mathcal{P}_1$:

$$\mathcal{A}_1 = \begin{pmatrix} A + \tilde{A} & B^T \\ B & 0 \end{pmatrix}.$$

Both systems are solved using a GMRES method with a 10^{-8} tolerance.

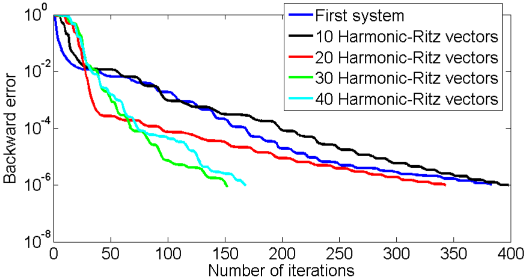
- grid : 32*32 points
- $\mathcal{P}_0 = I_{n+m}$
- $\mathcal{P}_1$ computed from eigenvectors of $\mathcal{A}$



→ using eigenvectors associated to the minimum modulus eigenvalues may accelerate the GMRES convergence but

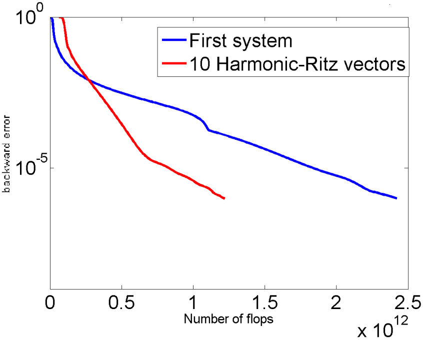
- eigenvectors are expensive to compute
- expensive in memory
- can not be applied with larger grid

- grid : 32*32 points
- $\mathcal{P}_0 = I_{n+m}$
- $\mathcal{P}_1$ computed from Harmonic-Ritz vectors of $\mathcal{A}$



$\rightarrow$ Using 30 or more Harmonic-Ritz vectors of $\mathcal{A}$ accelerates the GMRES convergence but the grid is too small to gain flops.

- grid : 256*256 points
- $\mathcal{P}_0 = \begin{pmatrix} I_n & B^T \\ B & 0 \end{pmatrix}$



→ gain 50% of flops when using 10 Harmonic-Ritz vectors associated to values of minimum modulus, low cost in memory

Outline

Conclusions

- We have derived a second-level preconditioner using a low-rank least Frobenius norm update for a saddle-point system

$$\mathcal{A} = \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$$

- We have derived a second-level preconditioner using a low-rank least Frobenius norm update for a generalized saddle-point system

$$\mathcal{A} = \begin{pmatrix} A & E^T \\ F & C \end{pmatrix}$$

- We have provided a spectral analysis for a specific case where A is symmetric positive definite, B^T is full-rank.
- We have analyzed the performance of the second-level preconditioner by using a Stokes problem.

Perspectives

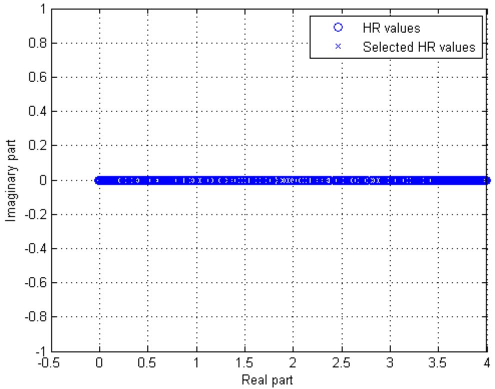
- Develop a methodology to analyze the spectrum of $\mathcal{P}_1^{-1}\mathcal{A}$ with other first-level preconditioner and other vectors to compute $\mathcal{P}_1$.

- Study the sharpness of the bounds given by the spectral analysis

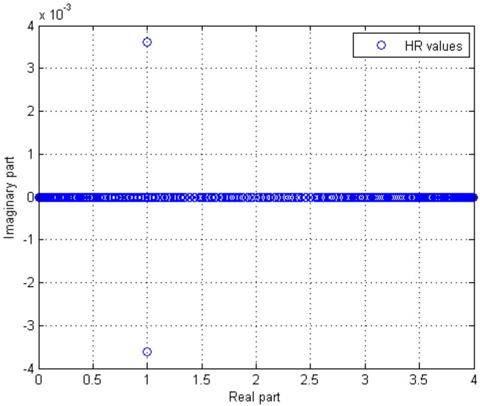
- Solve a sequence of system by using a time-dependent Stokes problem

Thank you for your attention !

Harmonic-Riz values of $\mathcal{A}$, grid of size 32*32



Harmonic-Ritz values of $\mathcal{A}$, grid of size 256*256



The Arnoldi relation

We want to solve the system $Ax = b, A \in \mathbb{R}^{n \times n}$.

Let's denote $V_m \in \mathbb{R}^{n \times m}$ the matrix containing the vectors $v_1, \ldots, v_m$ computed by the Arnoldi procedure. H_m the Hessenberg matrix of size $(m+1) \times m$.

The Arnoldi relation

$$AV_m = V_{m+1} H_m.$$

If we use the right preconditioner P, let's denote $Z_m = P^{-1} V_m$ and we have :

The Arnoldi relation

$$AZ_m = V_{m+1} H_m.$$

Harmonic-Ritz vector

Definition

A pair (λ, x) is called an Harmonic-Riz pair of A with respect to the subspace $\mathcal{L}$ if $x \in \mathcal{L}$ and $(Ax - \lambda x) \perp A\mathcal{L}$.

In our case $A = \mathcal{A} Z_m V_m^{\dagger}$ and $\mathcal{L} = \text{Span}(V_m)$.