



# Hierarchical Matrices and Inexact GMRES

---

Sudent:

GUIMARÃES Eduardo

Advisors:

FARIA Luiz

MARCHAND Pierre

Enseignant-Référent:

CHAPOUTOT Alexandre

Palaiseau, 13th May 2024

# Contents

<b>1</b>	<b>Iterative Methods and Krylov's Subspace</b>	<b>2</b>
1.1	Iterative Methods and motivation . . . . .	2
1.2	Krylov's Subspace . . . . .	3
1.3	Arnoldi's Method . . . . .	3
<b>2</b>	<b>GMRES</b>	<b>5</b>
2.1	Givens's Rotation . . . . .	5
2.2	Inexact GMRES . . . . .	6
<b>3</b>	<b>Hierarchical Matrices and ACA Method</b>	<b>9</b>
3.1	Low-rank Matrices . . . . .	9
3.2	ACA Method(Adaptative Cross Approximation) . . . . .	10
<b>4</b>	<b>First results</b>	<b>11</b>
4.1	Laplace's results . . . . .	12
4.2	Helmholtz's results . . . . .	12
	<b>Bibliography</b>	<b>14</b>

# Chapter 1

## Iterative Methods and Krylov's Subspace

### 1.1 Iterative Methods and motivation

Iterative methods appear as an alternative to exact solution methods, where the true solution is not desired and a good approximation is enough.

The idea is to find, after a certain number of iterations, a sequence  $x_k$  that converges to  $x$ , the correct solution of the problem 1.1.

$$x = \lim_{k \rightarrow \infty} x_k \quad (1.1)$$

The method stops after  $k$  iterations, where  $x_k$  is the first element of the sequence to satisfy the condition 1.2.

$$\frac{\|x_k - x\|}{\|x\|} \leq \epsilon \quad (1.2)$$

Where  $\epsilon$  is a tolerance defined by who is applying the algorithm.

Usually  $x$  isn't known, so 1.2 gets modified for 1.3, where  $A$  is the system's matrix and  $b$  is the RHS(right hand side).

$$\frac{\|Ax_k - b\|}{\|b\|} \leq \epsilon \quad (1.3)$$

The first iterative methods used a decomposition of  $A$  as a combination of two matrices 1.4, where  $A_1$  isn't singular, and each iteration is defined as 1.5.

$$A = A_1 - A_2 \quad (1.4)$$

$$A_1 x_{k+1} = b + A_2 x_k \quad (1.5)$$

With a substitution of the others  $x_k$ , 1.5 gives 1.6, which converges for every initial solution if and only lff  $\rho(A_2 A_1^{-1}) < 1$ , or  $\rho(X)$  is the spectral radius of  $X$  [5].

$$x_{k+1} = A_1^{-1}(b + A_2 x_k) = A_1^{-1}(b + A_2 A_1^{-1}(b + A_2 x_{k-1})) \dots = A_1^{-1} \left[ \sum_{i=0}^k (A_2 A_1^{-1})^i b \right] \quad (1.6)$$

If  $A_1 = I$  and  $A_2 = I - A$  in 1.4, the sequence found in 1.6 is:  $x_1 = b, x_2 = 2b - Ab, x_3 = 3b - 3Ab + A^2b, \dots$

Even if the condition  $\rho(A - I) \leq 1$  is strong [5], it shows that one approximation  $x_k$  could be represented as 1.7.

$$x_k \in \text{span}(b, Ab, A^2b, \dots, A^{k-1}b) \quad (1.7)$$

## 1.2 Krylov's Subspace

Be  $A \in \mathbb{K}^{n \times n}$  a matrix and  $b \in \mathbb{K}^n$ . To each  $k \leq n$  the Krylov's Subspace  $\mathcal{K}_k = \mathcal{K}_k(A, b)$  associated to A,b is defined as 1.8.

$$\mathcal{K}_k(A, b) = \text{span}(b, Ab, A^2b, \dots, A^{k-1}b) \quad (1.8)$$

These Subspaces also have the following property:  $k < l \rightarrow \mathcal{K}^k \subset \mathcal{K}^l$  [5].

The subspace  $\mathcal{K}_k(A, b)$  is also the subspace of all the vectors from  $\mathbb{R}^m$  which could be written as  $x = p(A)b$ , where  $p(A)$  is a polynom of degree less than  $k - 1$  which  $p(0) = 1$ .

The problem with using  $A^k b, k \in 0, 1, 2, \dots$  as a base comes from the fact that successive products of  $A$  make vectors that are *approximately colinears*, since those are really close of the eigenvector with the biggest eigenvalue of  $A$ .

## 1.3 Arnoldi's Method

With the task of obtaining an orthonormal basis to  $\mathcal{K}_k(A, b)$ , the method searches for a unitary matrix  $Q$  for which the expression 1.9 is valid.  $H_k = h_{ij}$  is an Hessenberg's matrix.

$$AQ_k = Q_{k+1}H_k \quad (1.9)$$

For each column-vector of  $Q$ ,  $q_i$ , 1.9 could be written as 1.10, where the representation of  $\mathcal{K}_k(A, b)$  with an orthonormal basis becomes more evident. In a pratical application,  $Q$  est initialized with  $q_1 = \frac{b}{\|b\|}$ .

$$Aq_m = h_{1m}q_1 + h_{2m}q_2 + \dots h_{m+1,m}q_{m+1} \quad (1.10)$$

An algorithm for the method can be found in 1.

---

**Algorithm 1** Arnoldi's iteration
 

---

```

1:  $A \in \mathbb{K}^{n \times n}$  et  $b \in \mathbb{K}^n$ 
2:  $x = 0, \beta = \|b\|, q_1 = \frac{b}{\beta}$ 
3: for  $j = 1, 2, \dots, k$  do
4:    $q_{j+1} = Aq_j$ 
5:   for  $i = 1, 2, \dots, j$  do
6:      $h_{ij} = q_{j+1}^t q_i$ 
7:      $q_{j+1} = q_{j+1} - h_{ij}q_i$ 
8:   end for
9:    $h_{j+1,j} = \|q_{j+1}\|$ 
10:   $q_{j+1} = \frac{q_{j+1}}{h_{j+1,j}}$ 
11: end for

```

---

## Chapter 2

# GMRES

A projection in  $\mathcal{K}_k(A, b)$ , where the different approximations are taken as in 2.1, where  $Q_m$  is the vector in 1.9.

$$x = x_0 + Q_m y \quad (2.1)$$

With 2.1 and 1.9 the residue becomes 2.2, where  $x_0 = 0$ ,  $\beta = \|b\|$  and  $Q_{m+1}^t b = (\|b\| \ 0 \ 0 \dots)^t$  since the columns of  $Q_{m+1}$  are orthonormal vectors and  $q_1 = \frac{b}{\|b\|}$ .

$$\begin{aligned} r(y) &= \|b - Ax\| \\ &= \|b - A(Q_m y)\| \\ &= \|b - Q_{m+1} H_m y\| \\ &= \|Q_{m+1} (Q_{m+1}^t b - H_m y)\| \\ &= \|\beta e_1 - H_m y\| \end{aligned} \quad (2.2)$$

Thus,  $y$  which appears in 2.1, is found as the solution of the residual's minimisation problem in 2.2.

$$y = \min_y \|\beta e_1 - H_m y\| \quad (2.3)$$

An initial version of the GMRES is in 2. The lines 4 to 12 bring the Arnoldi's Method presented in 1.

However, 2 doesn't bring an efficient way of finding the residual in each iteration. To solve this problem and also to find a more efficient way of solving the least squares problem in 2.3, a transformation is applied to  $H_m$ , turning it into a triangular matrix.

### 2.1 Givens's Rotation

Givens's operator,  $G(i, i+1)$ , is a unitary matrix such that the column vector  $a = Gb$  has the elements  $a(i) = r \in \mathbb{R}$  and  $a(i+1) = 0$ . It has a structure as in 2.4. The coefficients  $c_i, s_i$  only appear in the rows  $i$  et  $i+1$ .

**Algorithm 2** Initial GMRES

---

```

1:  $A \in \mathbb{K}^{n \times n}$  and  $b \in \mathbb{K}^n$ 
2:  $x = 0, \beta = \|b\|, q_1 = \frac{b}{\beta}$ 
3: for  $k = 1, 2, \dots$  do
4:   for  $j = 1, 2, \dots, k$  do
5:      $q_{j+1} = Aq_j$ 
6:     for  $i = 1, 2, \dots, j$  do
7:        $h_{ij} = q_{j+1}^t q_i$ 
8:        $q_{j+1} = q_{j+1} - h_{ij} q_i$ 
9:     end for
10:     $h_{j+1,j} = \|q_{j+1}\|$ 
11:     $q_{j+1} = \frac{q_{j+1}}{h_{j+1,j}}$ 
12:  end for
13:  Find  $y = \min_y \|\beta e_1 - H_m y\|$ 
14:   $x = Q_k y$ 
15:  Stop if the residual is smaller than the tolerance
16: end for

```

---

$$G(i, i+1) = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c_i & s_i & & \\ & & & -s_i & c_i & & \\ & & & & & 1 & \\ & & & & & & \ddots \\ & & & & & & & 1 \end{bmatrix} \quad (2.4)$$

This operator offers a way to transform the columns in  $H_m$ , *zeroing* the elements outside the main diagonal. Since a product of unitary operators is still unitary, 2.3 can be written as 2.5, where  $R_m$  and  $g_m$  are the results from the application of multiple Givens's operators to  $H_m$  and  $\beta e_1$ .

$$y = \min_y \|\beta e_1 - H_m y\| = \min_y \|g_m - R_m y\| \quad (2.5)$$

It can be shown that  $g_m$  contains the residual of each iteration [7].

Thus, the new problem 2.5 can be solved with a simple backwards substitution.

(écrire le nouvel algorithme )

## 2.2 Inexact GMRES

The heaviest part in the code is in the matrix-vector product 1, line 4. Therefore, one approach to accelerate the iterations involves an approximation of  $Aq$ , instead of using the exact answer, as shown in 2.6.

$$\mathcal{A}q = (A + E)q \quad (2.6)$$

Where  $E$  in 2.6 is a *perturbation matrix* that changes with each iteration and will be written as  $E_k$

for iteration  $k$ .

When the inexact matrix-vector product is the one being made, the left side of 1.9 must be changed by 2.7.

$$\begin{aligned} [(A + E_1)q_1, (A + E_2)q_2, \dots, (A + E_k)q_k] &= Q_{k+1}H_k \\ (A + \mathcal{E}_k)Q_k &= Q_{k+1}H_k, \quad \mathcal{E}_k = \sum_{i=1}^k E_i q_i q_i^t \\ \mathcal{A}Q_k &= W_k \end{aligned} \quad (2.7)$$

Where  $W_m = Q_{m+1}H_m$  from this point forward.

Now the subspace spanned by the vectors of  $Q_k$  is not the Krylov's subspace  $\mathcal{K}_k(A, b)$ , but these are still orthonormal. The expression 2.7 also shows that  $Q_k$  becomes a basis for a new Krylov's subspace,  $\mathcal{K}_k(A + \mathcal{E}_k, b)$ , made by a big perturbation in  $A$ , that gets updated in each iteration.

A new distinction should also be made between the two types of residues appearing in the process:  $r_k$ , the exact residue of an iteration, and  $\tilde{r}_k$ , the one that will really be calculated. A detailed definition for both and a measure of how distant they are is in 2.8.

$$\begin{aligned} r_k &= r_0 - AQ_k y_k \\ &= r_0 - (Q_{k+1}H_k - [E_1 q_1, \dots, E_k q_k])y_k \\ &= \tilde{r}_k + [E_1 q_1, \dots, E_k q_k]y_k \\ \rightarrow \delta_k &= \|r_k - \tilde{r}_k\| = \|[E_1 q_1, \dots, E_k q_k]y_k\| \end{aligned} \quad (2.8)$$

Considering  $y_k = [\eta_1^{(k)} \dots \eta_n^{(k)}]$ , upper index to clarify the iteration, an upper bound for  $\delta_k$  can be found [8], as shown in 2.9

$$\delta_k = \|r_k - \tilde{r}_k\| \leq \sum_{i=1}^k \|E_i\| \|\eta_i^{(k)}\| \quad (2.9)$$

2.9 tells us that in order to keep both residues close, either the perturbation of  $A$ , somewhat measured by  $\|E_i\|$ , or the elements of  $y_i$  should be kept small. Since we expect to use more *relaxed* approximations of  $A$  as the iterations go on, a greater tolerance in  $E_k$  could be compensated with a sufficiently small  $y_k$ .

The problem is  $y_k$  is only found after the construction of  $E_k$ , so an upper bound must be also found for its value. It can be shown, lemma 5.1 in [8], that for  $i = 1, 2, \dots, k$ , where  $k$  is the iteration number, the bound is given by 2.10.

$$\|\eta_i^{(k)}\| \leq \frac{1}{\sigma_k(H_k)} \|\tilde{r}_{i-1}\| \quad (2.10)$$

Putting 2.10 in 2.9 gives the results 2.11. Setting  $\delta_k \leq \epsilon$  and determining a bound for each  $\|E_i\|$  gets us 2.12.



$$\delta_k \leq \sum_{i=1}^k \frac{\|E_i\|}{\sigma_k(H_k)} \|\tilde{r}_{i-1}\| \quad (2.11)$$

$$\|E_i\| \leq \frac{\sigma_k(H_k)\epsilon}{k \|\tilde{r}_{i-1}\|} \quad (2.12)$$

Since  $H_k$  is also one of the matrices being constructed throughout the method, a workaround is necessary to apply find these bounds in a practical situation. Either using an estimation of  $\sigma_k(H_k)$  with the singular values of  $A$  or grouping all uncalculated terms in a  $\ell_k$  that will be estimated empirically [8], obtaining 2.13.

$$\|E_i\| \leq \ell_k \frac{1}{\|\tilde{r}_{i-1}\|} \epsilon \quad (2.13)$$

## Chapter 3

# Hierarchical Matrices and ACA Method

### 3.1 Low-rank Matrices

In reality, most matrices are big, so storing each element is not efficient, or even possible. If  $A \in \mathbb{C}^{n \times m}$  has a rank  $k$  such that  $k \leq m$  and  $k(n+m) < n * m$  ( $A$  is low-rank),  $A$  can be written in outer product form, as a product between the matrices  $U \in \mathbb{C}^{n \times k}$  and  $V \in \mathbb{C}^{m \times k}$ , which can be seen in 3.1, where  $u_i, v_i$  are the column vectors of  $U$  and  $V$ .

$$A = UV^H = \sum_{i=1}^k u_i v_i^* \quad (3.1)$$

Therefore, storing  $k(n+m)$  elements to write  $A$ , and not  $n \times m$ . A matrix  $A$  that can be represented as 3.1 is an element of  $\mathbb{C}_k^{n \times m}$ .

The representation in 3.1 also facilitates other operations with  $A$ , like matrix-vector products  $Ab$  that are always present in methods like GMRES [4] and different kinds of norms, like  $\|A\|_F, \|A\|_2$  [4].

However, even full rank matrices can be approximated by matrices with lower rank. A theorem [4] establishes that the closest matrix from  $\mathbb{C}_k^{n \times m}$  of a matrix from  $\mathbb{C}^{n \times m}$  can be obtained from the SVD  $A = U\Sigma V^H$ , where  $\Sigma$  contains the singular values  $\sigma_1 \geq \sigma_2 \dots \sigma_m \geq 0$  and  $U, V$  are unitary. If  $A_k$  is the approximation obtained after taking the first  $k$  elements of  $\Sigma$  (creating the matrix  $\Sigma_k$ ), the error between  $A$  and  $A_k$  is 3.2.

$$\|A - A_k\| = \left\| U\Sigma V^H - U'\Sigma_k V_H \right\| = \|\Sigma - \Sigma_k\| \quad (3.2)$$

If the spectral norm,  $\|\cdot\|_2$  is used instead, the error in 3.2 is given by  $\sigma_{k+1}$ . For Frobenius's norm,  $\|\cdot\|_F$ , the error becomes  $\sum_{l=k+1}^n \sigma_l^2$ .

Instead of approximating big matrices entirely, it's better to think in approximations made to each of their blocks. Blocks that appear after the discretization of elliptic operators also have the possibility of being approximated by matrices that decay exponentially with  $k$ ,  $S_k$ , as in 3.3.

$$\|A - S_k\|_2 < q^k \|A\|_2 \quad (3.3)$$

That way, the rank and the precision are related in a logarithmic manner, and the rank required by a certain  $\epsilon$  is 3.4.

$$k(\epsilon) = \min\{k \in \mathbb{N} : \sigma_{k+1} < \epsilon \sigma_1\} \quad (3.4)$$

## 3.2 ACA Method(Adaptative Cross Approximation)

As shown in the last section, the SVD methods gives us an approximation of  $A$  given a certain  $\epsilon$ , through the relation in 3.2. Nevertheless, this is an expensive method, where the complexity becomes too big for some calculations.

The algorithm for the method is in 3, where  $a_{ij}$  are the elements of a matrix  $A \in \mathbb{R}^{n \times m}$ . The main objective is to approximate  $A$  as  $A = S_k + R_k$ ,  $S_k = \sum_{l=1}^k u_l v_l^t$  and  $R_k$  is the residue.

---

### Algorithm 3 ACA Method

---

```

1:  $k = 1$  et  $\mathbf{Z} = \emptyset$ 
2: repeat
3:   TFind  $i_k$ 
4:    $\hat{v}_k = a_{i_k, 1:m}$ 
5:   for  $l = 1, \dots, k-1$  do
6:      $\hat{v}_k = \hat{v}_k - (u_l)_{i_k} v_l$ 
7:   end for
8:    $Z = Z \cup \{i_k\}$ 
9:   if  $\hat{v}_k$  doesn't disappear then
10:     $j_k = \operatorname{argmax}_j |(\hat{v}_k)_j|$  ;  $v_k = (\hat{v}_k)_{j_k}^{-1} \hat{v}_k$ 
11:     $u_k = a_{1:n, j_k}$ 
12:    for  $l = 1, \dots, k-1$  do
13:       $u_k = u_k - (v_l)_{j_k} u_l$ 
14:    end for
15:     $k = k + 1$ 
16:  end if
17: until  $\|u_k\| \|v_k\| \leq \epsilon$ 

```

---

Considering  $I, J \in \mathbb{N}$  the index set of a given matrix and  $\mathbf{T}_{I \times J}$  the cluster block tree that contains an admissible partition  $P$  of  $I \times J$  in its leaves,  $\mathfrak{L}(\mathbf{T}_{I \times J})$ . The set of hierarchical matrices in  $\mathbf{T}_{I \times J}$  rank  $k$  for each block  $A_b$  defined in 3.5.

$$\mathfrak{H}(\mathbf{T}_{I \times J}, k) = \{A \in \mathbb{C}^{I \times J} : \operatorname{rank} A_b \leq k, \forall b \in P\} \quad (3.5)$$

## Chapter 4

### First results

Before using the inexact product in big problems, simpler examples are used to validate the approach and fix minor parameters in the scheme.

The two firsts tests evaluate the speedup in the product of a Hierarchical Matrix and a vector and an execution of the Inexact GMRES with few iterations, using the operators obtained through 2nd type Equations of Laplace and Helmholtz 4.1, where the last one is a scattering problem, where  $\Delta$  is the Laplace operator and everything is supposed to be solved in two dimensions. The last test will use a cavity problem to test the speedup of the algorithm in a situation with more iterations.

$$\begin{aligned}\Delta u &= 0 \\ \Delta u + k^2 u &= 0\end{aligned}\tag{4.1}$$

Reformulating both equations as a Boundary Integral Equation, the simple *direct* formulation is used to write the solution as 4.2, where  $\Gamma$  is boundary of the domain,  $S$  and  $D$  are the single and double layer operators, defined as 4.3, and  $G(x, y)$  is the fundamental solution of the desired PDE.

$$-\frac{u(x)}{2} + D[u](x) = S[\partial_\nu u](x), \quad x \in \Gamma\tag{4.2}$$

$$\begin{aligned}S[\sigma](x) &= \int_{\Gamma} G(x, y) \sigma(y) ds(y) \\ D[\sigma](x) &= \int_{\Gamma} \frac{\partial G}{\partial \nu_y}(x, y) \sigma(y) ds(y)\end{aligned}\tag{4.3}$$

For the first two examples, a unit circle around the origin is used as the boundary to generate the operators, with the mesh being created with the *Inti* library [3].

For the last test, the mesh is made from a cavity *.geo* file available in [1]. A view of the figure can be seen in 4.1. The incident wave's angle is chosen to be  $\frac{\pi}{4}$  rad.

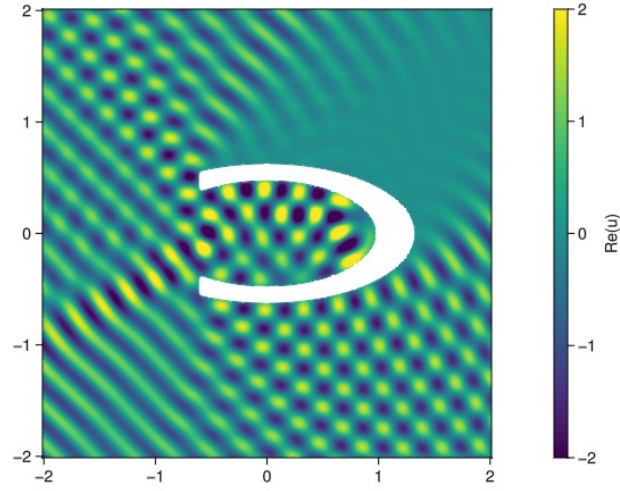


Figure 4.1: Geometry used in the test.

## 4.1 Laplace's results

All results are contained in 4.2, showing the evolution of the residual with the product tolerance aswell as the speedup to each of these values.

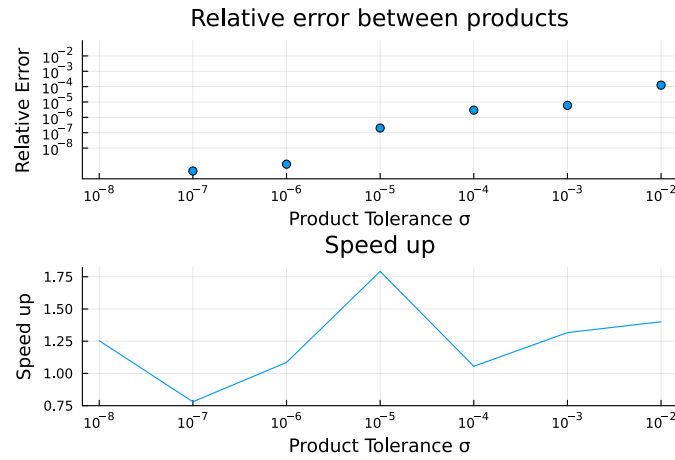


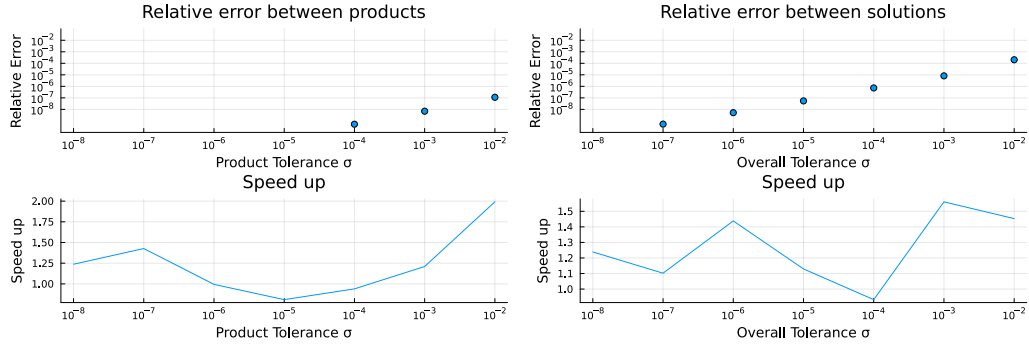
Figure 4.2: Speedup and residual evolution for the product between a 8000x8000 HMatrix and a vector.

## 4.2 Helmholtz's results

For the unitary circle boundary the results can be seen in 4.3.

For the cavity, 4.4.

An evolution of the number of iterations in face of the different tolerances passed to the algorithm is in 4.5.



(a) Results for the product of a 70000x 70000 HMatrix and a vector. (b) Results for an initial application of the Inexact GMRES algorithm.

Figure 4.3: Results for the application of the Inexact GMRES algorithm with a 70000x70000 HMatrix.

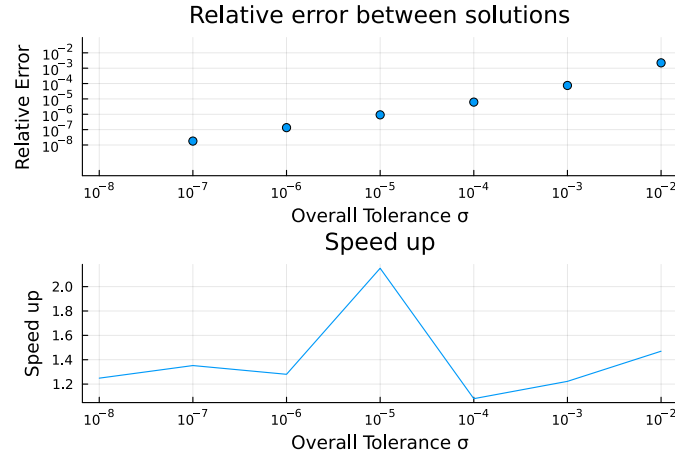


Figure 4.4: Speedup witnessed in the application of the Inexact GMRES in a 50000x50000 matrix.

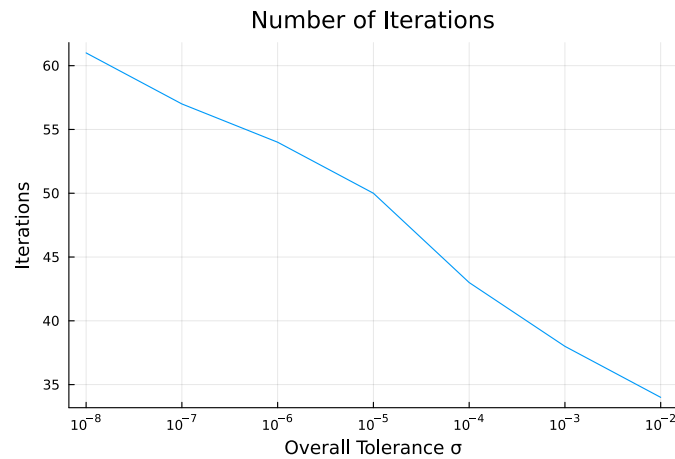


Figure 4.5: Evolution of the quantity of iterations needed for convergence and overall tolerance passed as an argument.

# Bibliography

- [1] Dépôt github. <https://github.com/DuduGuima/InexactGMRES.git>.
- [2] Hmatrices.jl. <https://github.com/IntegralEquations/HMatrices.jl>.
- [3] Inti.jl. <https://github.com/IntegralEquations/Inti.jl>.
- [4] Mario Bebendorf. *Hierarchical matrices*. Springer, 2008.
- [5] Marc Bonnet. *Lecture notes on numerical linear algebra*. ENSTA Paris - POEMS Group, 2021.
- [6] Eric Darve and Mary Wootters. *Numerical linear algebra with Julia*. SIAM, 2021.
- [7] Yousef Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.
- [8] Valeria Simoncini and Daniel B Szyld. Theory of inexact krylov subspace methods and applications to scientific computing. *SIAM Journal on Scientific Computing*, 25(2):454–477, 2003.
- [9] Lloyd N Trefethen, David Bau III, and Ricardo D Fierro. Numerical linear algebra. *SIAM Review*, 40(3):735–738, 1998.