Trefftz iterative method for three-dimensional electromagnetic waves propagation

Sébastien Pernet, Margot Sirdey, Sébastien Tordeux

Iterative Trefftz method for three-dimensional electromagnetic waves simulation

Sébastien Pernet[1], Margot Sirdey[1,2], Sébastien Tordeux[2]

[1] MACI LMA2S ONERA [2] EPI INRIA-Total Makutu, LMAP UMR CNRS 5142, Pau University

JCJC 2022
INRIA Sophia-Antipolis, France, 29^{th}, November, 2022

Adimensioned Maxwell problem

$$\left\{ \begin{array}{rcll} \nabla \times \mathbf{H} & = & ik\varepsilon_r \mathbf{E} & \\ \nabla \times \mathbf{E} & = & -ik\mu_r \mathbf{H} & \text{in } \Omega \subset \mathbb{R}^3, \end{array} \right.$$

where ε_r and μ_r are the relative permittivity and permeability.

Goal: achieve numerical computations on very large domains

$$\mathbf{200\lambda} \Longrightarrow \mathbf{416 \times 10^6} \text{ dof } (\lambda\text{-mesh}) \text{ or } \mathbf{3 \times 10^9} \text{ dof } (\tfrac{\lambda}{2}\text{-mesh})$$

The frigate Surcouf at sea near Toulon. Author: Franck Dubey. Source: netmarine.net.

Size of the computational domain large with respect to the wavelength $\Leftrightarrow$ $L >> \lambda$

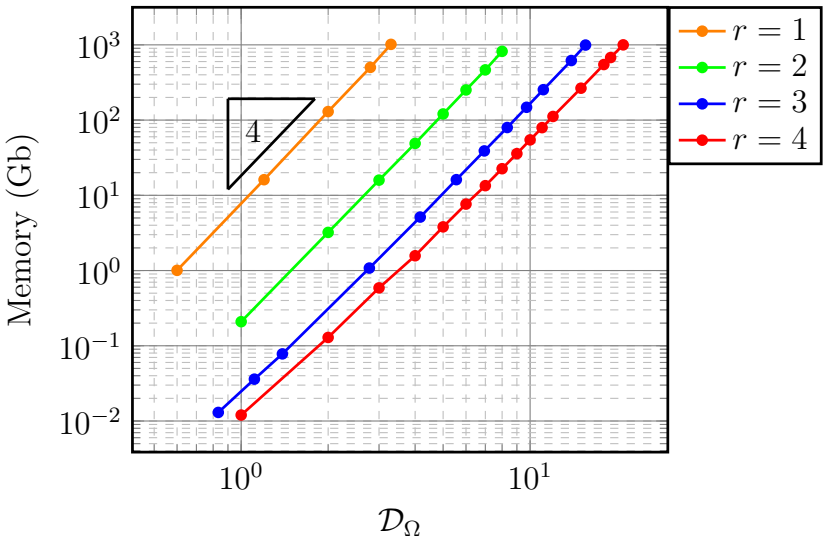
Numerical pollution problem[a]

[a] F. Ihlenburg and I. Babuska, *Finite element solution of the Helmholtz equation with high wave number part II: the hp version of the FEM, SIAM Journal on Numerical Analysis*, 34 (1), pp. 315–358 (1997).

Memory limit of a classic solver

$$\mathbf{A}[\mathbb{E}] = \mathbf{F}$$

- Second order Maxwell system and continuous finite elements leads to less degrees of freedom.
- The Finite Element method deals correctly with the geometry and evanescent modes.



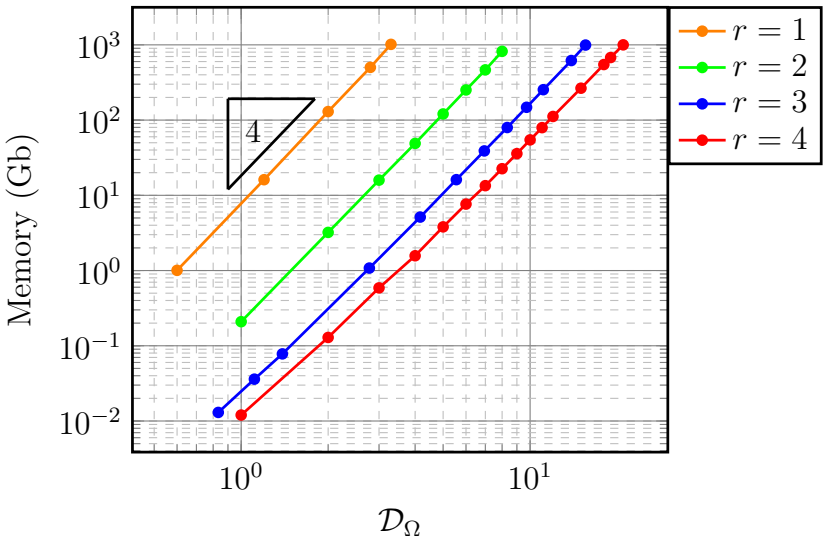
Memory cost of the Nédélec solver for different domain sizes $\mathcal{D}_\Omega$ and approximation orders r, memory $= (\mathcal{D}_\Omega)^4$.

	Memory	Sockets	$\mathcal{D}_\Omega$	Power	Cost
Laptop	32 Gb	1	10λ	200 W	0.034 €/h
// Computation	1 Tb	16	24λ	3.2 KW	0.544 €/h
HPC	320 Tb	5000	100λ	1 MW	170 €/h
HPC+	5120 Tb	80000	200λ	16 MW	2720 €/h

Memory limit of a classic solver

$$\mathbf{A}[\mathbb{E}] = \mathbf{F}$$

- Even efficient optimizations cannot overcome the **memory issue**.
- An iterative method ?



Memory cost of the Nédélec solver for different domain sizes $\mathcal{D}_\Omega$ and approximation orders r, memory $= (\mathcal{D}_\Omega)^4$.

	Memory	Sockets	$\mathcal{D}_\Omega$	Power	Cost
Laptop	32 Gb	1	10λ	200 W	0.034 €/h
// Computation	1 Tb	16	24λ	3.2 KW	0.544 €/h
HPC	320 Tb	5000	100λ	1 MW	170 €/h
HPC+	5120 Tb	80000	200λ	16 MW	2720 €/h

Example of GMRES solver with classic methods

Let $\Omega := [0, \mathcal{D}_\Omega]^3$. We solve

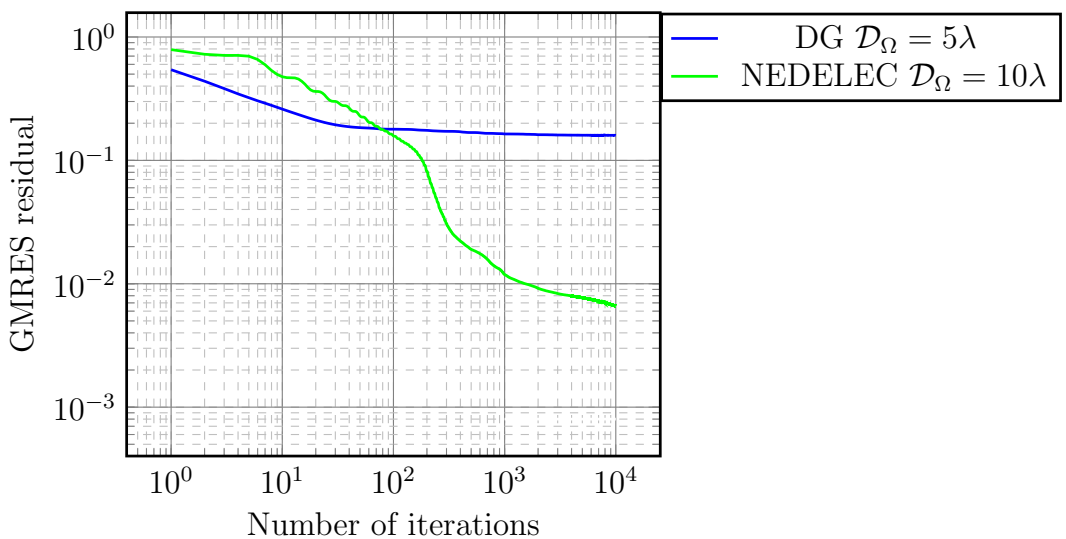
$$\left\{ \begin{array}{rcll} \nabla \times \mathbf{H} & = & ik\mathbf{E}, & \text{in } \Omega, \\ \nabla \times \mathbf{E} & = & -ik\mathbf{H}, & \text{in } \Omega, \end{array} \right.$$

equipped with the impedance boundary condition

$$\gamma_t \mathbf{E} + Z_{\partial\Omega}\, \mathbf{n}_{\partial\Omega} \times \gamma_t \mathbf{H} = \mathbf{g} \quad \text{on } \partial\Omega, \qquad \text{with} \quad \gamma_t \mathbf{E} = \mathbf{E} - (\mathbf{E} \cdot \mathbf{n})\mathbf{n}.$$

GMRES solver is not convergent

High order Nedelec Finite element on hexahedra (NEDELEC)
High order Dicontinuous Galerkin on tetrahedra (DG)



Is the Trefftz method an alternative ?

Particularity of Trefftz methods: basis functions are local solutions of the problem.

- O. Cessenat, and B. Després. "Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem." SIAM journal on numerical analysis. 1998.
- P. Monk, and D. Q. Wang. A least-squares method for the Helmholtz equation. Computer methods in applied mechanics and engineering. 1999.
- A. Moiola. Trefftz-discontinuous Galerkin methods for time-harmonic wave problems. Diss. ETH Zurich. 2011.
- R. Hiptmair, A. Moiola, and I. Perugia. "A survey of Trefftz methods for the Helmholtz equation." Building bridges: connections and challenges in modern approaches to numerical partial differential equations. Springer, Cham, 2016.
- H. Barucq, A. Bendali, M. Fares, V. Mattesi, and S. Tordeux. A symmetric Trefftz-DG formulation based on a local boundary element method for the solution of the Helmholtz equation. J.C.P. 2017.

General problematic

Are Trefftz methods an efficient alternative to study large computational domains ?

Addressed issues in this presentation

1. Is the **Trefftz direct** method an alternative ?

   → Choice of the variational formulation
   → Limits of a LU factorisation

2. Is it possible to define a **Trefftz iterative** method ?

   → Ultra-Weak Variational Formulation (UWVF)
   → Krylov-type solver

3. Is it possible to **optimize** this method to reach one **billion degrees of freedom** ?

   → Use of a reduced space
   → Use of preconditioners

General problematic

Are Trefftz methods an efficient alternative to study large computational domains ?

Code developed from scratch

FORTRAN code, **GoTEM3**, linked to the GMRES of CERFACS© [a]

- **Go** → **G**iga**o**ctet (**G**iga**b**yte in french),
- **T** → **T**refftz method,
- **EM** → **E**lectro**M**agnetic waves,
- **3** → **3**D

[a] V. Frayssé, L. Giraud, S. Gratton, and J. Langou (2005). Algorithm 842: A set of GMRES routines for real and complex arithmetics on high performance computers. ACM Transactions on Mathematical Software (TOMS), 31(2), 228-238.

Steps to define a Trefftz method

1. The solution and the test functions satisfy the Maxwell problem in every element T

$$\nabla \times \mathbf{H}^T = ik_0\varepsilon_r \mathbf{E}^T \quad \text{and} \quad \nabla \times \mathbf{E}^T = -ik_0\mu_r \mathbf{H}^T.$$

2. Virtual work principle in an element T: $W_T = -ik_0 \int_T \varepsilon_r \mathbf{E}^T \cdot \overline{\mathbf{E}'^T} + \mu_r \mathbf{H}^T \cdot \overline{\mathbf{H}'^T}$, can be written as

$$\begin{aligned} W_T &= -\int_T \underbrace{\nabla \times \mathbf{H}^T}_{ik_0\varepsilon_r \mathbf{E}^T} \cdot \overline{\mathbf{E}'^K} - \underbrace{\nabla \times \mathbf{E}^T}_{-ik_0\mu_r \mathbf{H}^T} \cdot \overline{\mathbf{H}'^T}, \\ W_T &= \int_T \mathbf{E}^T \cdot \underbrace{\overline{\nabla \times \mathbf{H}'^T}}_{ik_0\varepsilon_r \mathbf{E}'^T} - \mathbf{H}^T \cdot \underbrace{\overline{\nabla \times \mathbf{E}'^T}}_{-ik_0\mu_r \mathbf{H}'^T}. \end{aligned}$$

3. Stokes formula leads to the **reciprocity formula**, defined on boundaries

$$\sum_T \int_{\partial T} \left(\mathbf{n}_T \times \gamma_t \mathbf{H}^T\right) \cdot \overline{\gamma_t \mathbf{E}'^T} + \gamma_t \mathbf{E}^T \cdot \left(\mathbf{n}_T \times \overline{\gamma_t \mathbf{H}'^T}\right) = 0.$$

4. Find $\mathbb{E} := (\mathbf{E}, \mathbf{H}) \in \mathbb{X}$, such that $\forall \mathbb{E}' := (\mathbf{E}', \mathbf{H}') \in \mathbb{X}$, we have $\mathbf{a}(\mathbb{E}, \mathbb{E}') = \ell(\mathbb{E}')$, with $\mathbb{X}$ a **discontinuous space of local Maxwell solutions** and

$$\mathbf{a}(\mathbb{E}, \mathbb{E}') = \sum_T \int_{\partial T} \widehat{\gamma_\times^T \mathbf{H}^T} \cdot \gamma_t \overline{\mathbf{E}'^T} + \widehat{\gamma_t \mathbf{E}^T} \cdot \gamma_\times^T \overline{\mathbf{H}'^T}, \quad \text{with } \gamma_\times^T \mathbf{H}^T := \mathbf{n}_T \times \gamma_t \mathbf{H}^T.$$

Numerical traces choices

Clever choices of the numerical traces $\widehat{\gamma_t \mathbf{E}^T}$ and $\widehat{\gamma_\times^T \mathbf{H}^T}$ ensure the **coercivity**.

Numerical traces are defined on interior and boundary faces F.

Consistance: coefficients are chosen so that the exact solution satisfies

$$\begin{aligned} \left(\widehat{\gamma_t \mathbf{E}}\right)_{|F} &= \gamma_t \mathbf{E}^T = \gamma_t \mathbf{E}^K = \gamma_t \mathbf{E}, \\ \left(\widehat{\gamma_\times^K \mathbf{H}}\right)_{|F} &= \gamma_\times^T \mathbf{H}^T = -\gamma_\times^K \mathbf{H}^K = \gamma_\times^T \mathbf{H}. \end{aligned}$$

Equivalent approaches to compute numerical traces

- Riemann solver,
- Upwind fluxes,
- UWVF method.

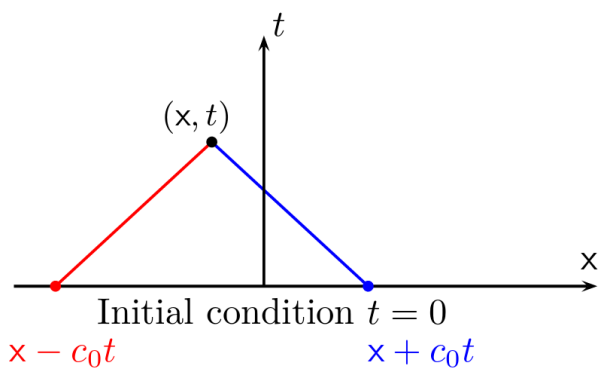
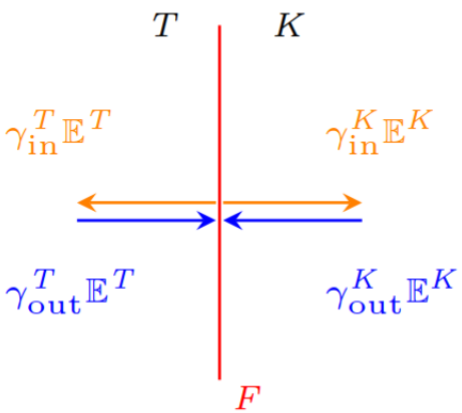


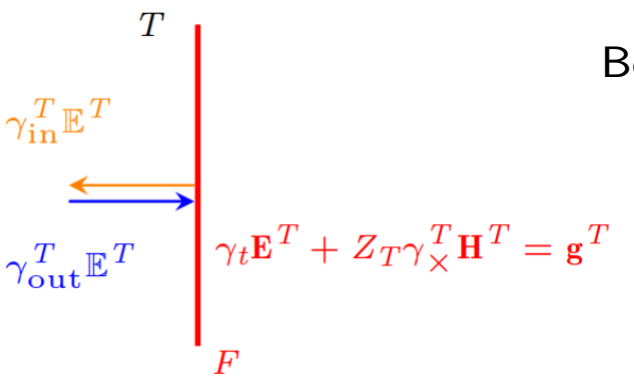
Illustration for a Riemann solver.

Upwind numerical traces

The incoming and outgoing traces

$$\gamma_{\text{in}}^T \mathbb{E}^T := \gamma_t \mathbf{E}^T + Z_T \gamma_\times^T \mathbf{H}^T \quad \text{and} \quad \gamma_{\text{out}}^T \mathbb{E}^T := \gamma_t \mathbf{E}^T - Z_T \gamma_\times^T \mathbf{H}^T.$$





Coefficients are chosen so that the exact solution satisfies

$$\begin{aligned} \widehat{(\gamma_t \mathbf{E})}_{|F} &= \gamma_t \mathbf{E}^T = \gamma_t \mathbf{E}^K = \gamma_t \mathbf{E}, \\ \widehat{(\gamma_\times^K \mathbf{H})}_{|F} &= \gamma_\times^T \mathbf{H}^T = -\gamma_\times^K \mathbf{H}^K = \gamma_\times^T \mathbf{H}. \end{aligned}$$

Interior numerical traces

$$\begin{aligned} \widehat{(\gamma_t \mathbf{E})}_{|F} &= \alpha^T \gamma_{\text{out}}^T \mathbb{E}^T + \alpha^K \gamma_{\text{out}}^K \mathbb{E}^K, \\ \widehat{(\gamma_\times^T \mathbf{H})}_{|F} &= \beta^T \gamma_{\text{out}}^T \mathbb{E}^T + \beta^K \gamma_{\text{out}}^T \mathbb{E}^K. \end{aligned}$$

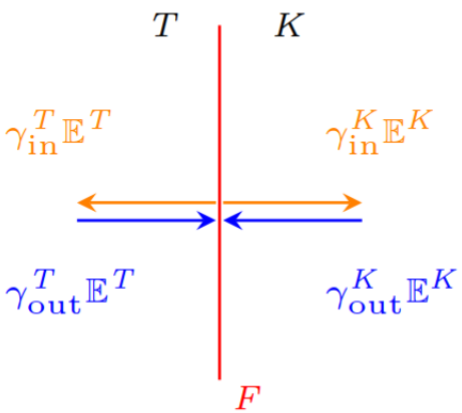
Boundary numerical traces

$$\begin{aligned} \widehat{(\gamma_t \mathbf{E})}_{|F} &= \alpha^T \gamma_{\text{out}}^T \mathbb{E}^T + \alpha^K \mathbf{g}^T, \\ \widehat{(\gamma_\times^T \mathbf{H})}_{|F} &= \beta^T \gamma_{\text{out}}^T \mathbb{E}^T + \beta^K \mathbf{g}^T. \end{aligned}$$

Upwind numerical traces

The incoming and outgoing traces

$$\gamma_{\text{in}}^T \mathbb{E}^T := \gamma_t \mathbf{E}^T + Z_T \gamma_\times^T \mathbf{H}^T \quad \text{and} \quad \gamma_{\text{out}}^T \mathbb{E}^T := \gamma_t \mathbf{E}^T - Z_T \gamma_\times^T \mathbf{H}^T.$$

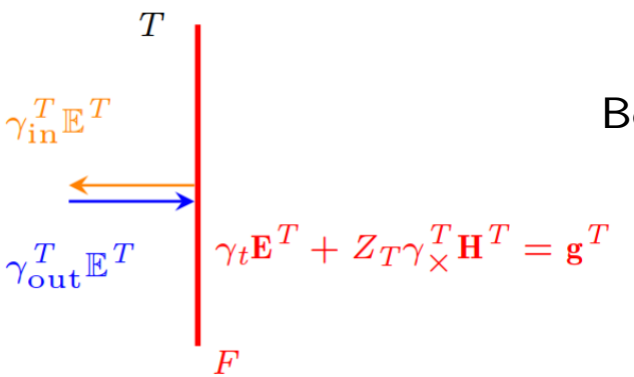


Coefficients are chosen so that the exact solution satisfies

$$\begin{aligned} \widehat{(\gamma_t \mathbf{E})}_{|F} &= \gamma_t \mathbf{E}^T = \gamma_t \mathbf{E}^K = \gamma_t \mathbf{E}, \\ \widehat{(\gamma_\times^K \mathbf{H})}_{|F} &= \gamma_\times^T \mathbf{H}^T = -\gamma_\times^K \mathbf{H}^K = \gamma_\times^T \mathbf{H}. \end{aligned}$$

Interior numerical traces

$$\begin{aligned} \widehat{(\gamma_t \mathbf{E})}_{|F} &= \frac{Z_K}{Z_T + Z_K} \gamma_{\text{out}}^T \mathbb{E}^T + \frac{Z_T}{Z_T + Z_K} \gamma_{\text{out}}^K \mathbb{E}^K, \\ \widehat{(\gamma_\times^T \mathbf{H})}_{|F} &= -\frac{1}{Z_T + Z_K} \gamma_{\text{out}}^T \mathbb{E}^T + \frac{1}{Z_T + Z_K} \gamma_{\text{out}}^T \mathbb{E}^K. \end{aligned}$$



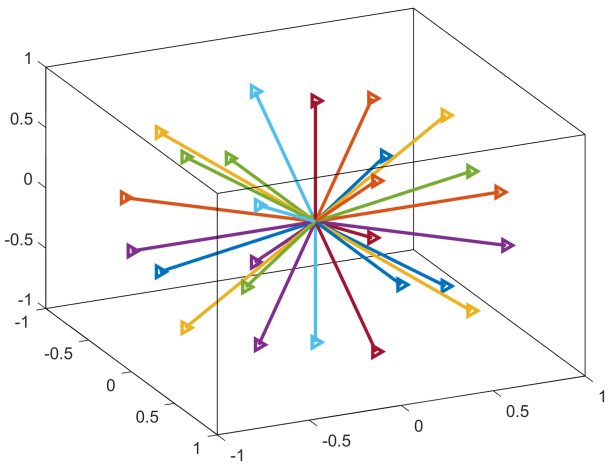
Boundary numerical traces

$$\begin{aligned} \widehat{(\gamma_t \mathbf{E})}_{|F} &= \frac{Z_{\partial\Omega}}{Z_T + Z_{\partial\Omega}} \gamma_{\text{out}}^T \mathbb{E}^T + \frac{Z_T}{Z_T + Z_{\partial\Omega}} \mathbf{g}^T, \\ \widehat{(\gamma_\times^T \mathbf{H})}_{|F} &= -\frac{1}{Z_{\partial\Omega} + Z_T} \gamma_{\text{out}}^T \mathbb{E}^T + \frac{1}{Z_{\partial\Omega} + Z_T} \mathbf{g}^T. \end{aligned}$$

Plane wave Galerkin space

An electromagnetic plane wave $\mathbf{E} = \overrightarrow{p}\, e^{ik\,\overrightarrow{d}\cdot\overrightarrow{x}}$ is parameterized by:

- a direction of propagation $\overrightarrow{d} = (d_x, d_y, d_z)$,
- an electric polarisation $\overrightarrow{p} = (p_x, p_y, p_z)$ orthogonal to $\overrightarrow{d}$.



26 plane wave directions and
52 polarisations.

- In each element T, the approximated solution $\mathbb{E}^T := (\mathbf{E}^T, \mathbf{H}^T)$ is

$$\begin{cases} \mathbf{E}^K = \displaystyle\sum_{\mathrm{j}=1}^{N} u_{\mathrm{j}} \underbrace{\overrightarrow{p_{\mathrm{j}}}\, e^{ik\overrightarrow{d_{\mathrm{j}}}\cdot\overrightarrow{x}}}_{\mathbf{w}^{\mathrm{j}}}, \\ \mathbf{H}^K = \displaystyle\sum_{\mathrm{j}=1}^{N} u_{\mathrm{j}} \left(\overrightarrow{p_{\mathrm{j}}} \times \overrightarrow{d_{\mathrm{j}}}\right) e^{ik\overrightarrow{d_{\mathrm{j}}}\cdot\overrightarrow{x}}. \end{cases}$$

- **What is the optimal number N of plane waves ?**

  N small leads to a bad approximation.

  N large leads to rounding errors.

Some references about the optimal choice of basis functions

What is the optimal number of plane waves ?

- A. Moiola. Trefftz-discontinuous Galerkin methods for time-harmonic wave problems. Diss. ETH Zurich. 2011.
- A. Moiola, R. Hiptmair, I. Perugia. Vekua theory for the Helmholtz operator. Zeit. für ang. Math. und Phys. 2011.

Strategies to minimize the effect of numerical pollution

- T. Luostari, T. Huttunen, P. Monk. Improvements for the ultra weak variational formulation, Int. J. Numer. Meth. Engng 94. 2013.
- S. Congreve, J. Gedicke, I. Perugia. Numerical investigation of the conditioning for plane wave discontinuous Galerkin methods, Vol. 126 of Lecture Notes in Comp. Sc. and Eng., Springer. 2019.
- H. Barucq, A. Bendali, J. Diaz, S. Tordeux. Local strategies for improving the conditioning of the plane-wave Ultra-Weak Variational Formulation. J.C.P. 2021.

Some alternatives exist (evanescent modes or Quasi-Trefftz methods)

- E. Parolin, D. Huybrechs, A. Moiola. Stable approximation of Helmholtz solutions by evanescent plane waves. 2022.
- L.-M. Imbert-Gérard, and P. Monk. "Numerical simulation of wave propagation in inhomogeneous media using generalized plane waves." ESAIM: Mathematical Modelling and Numerical Analysis 51.4 2017.
- H. S. Fure, S. Pernet, M. Sirdey, S. Tordeux. A discontinuous Galerkin Trefftz type method for solving the two dimensional Maxwell equations. SN Partial Differ. Equ. Appl. 2020.

The Trefftz matrix for a structured mesh

$$\mathbf{A}[\mathbb{E}] \;=\; \mathbf{F},$$

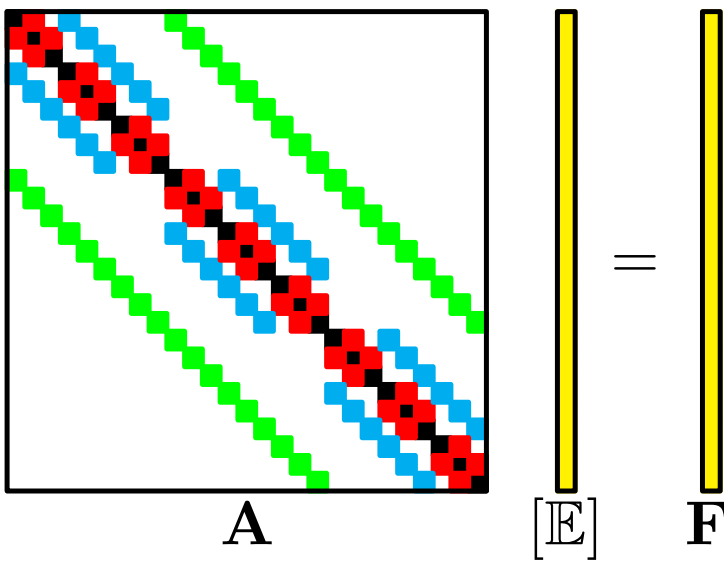
where

$$\mathbf{A}_{i,j} := \mathbf{a}(\mathbf{w}^j, \mathbf{w}^i)$$

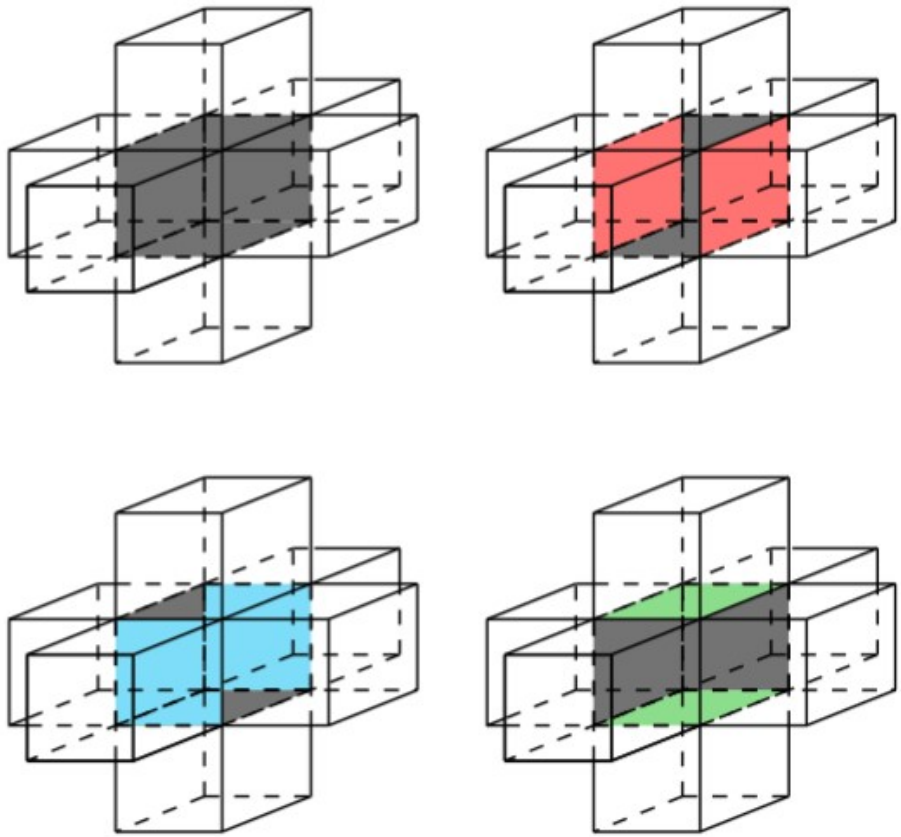
$$\mathbf{F}_i := \mathbf{l}(\mathbf{w}^i) \quad \text{for} \;\; i, j = 1, N_{\text{dof}}.$$

$$N_{\text{dof}} := N \times N_{\text{elem}}$$

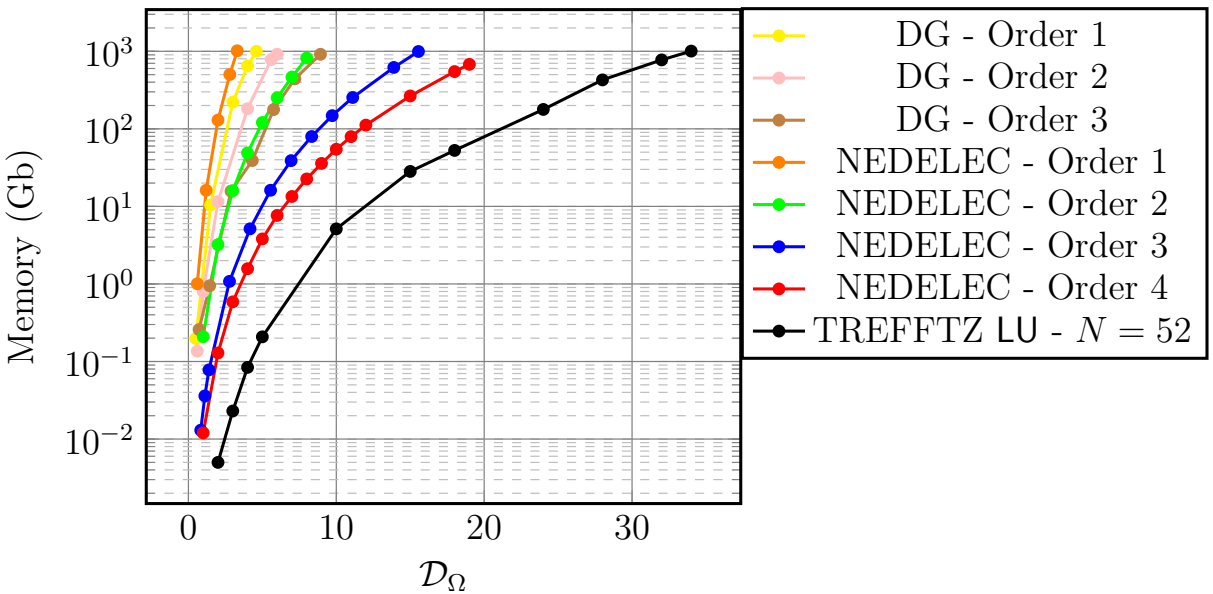
$$size(\mathbf{A}) = N^2 \times 7 \times N_{\text{elem}}$$



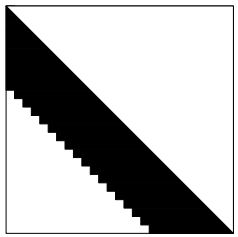
The matrix $\mathbf{A}$ for $N_{\text{elem}} = 27 = 3^3$ cubes.



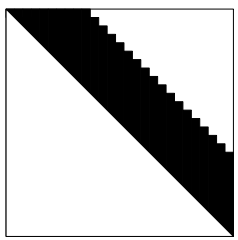
Memoy limit of the Trefftz direct solver



LU Factorisation



L



U

$\mathcal{D}_\Omega$	5λ	10λ	20λ	100λ	200λ
N_{elem}	25	1000	8000	1×10^6	8×10^6
N_{dof}	1300	52×10^3	416×10^6	52×10^6	416×10^6
LU Memory	312 Mb	5 Gb	80 Gb	50 Tb	800 Tb

Table: Memory cost of the LU factorisation with respect to the domain size $\mathcal{D}_\Omega$, for $N = 52$ and for $h = 1$.

Cessenat-Després decomposition

Singular regular decomposition

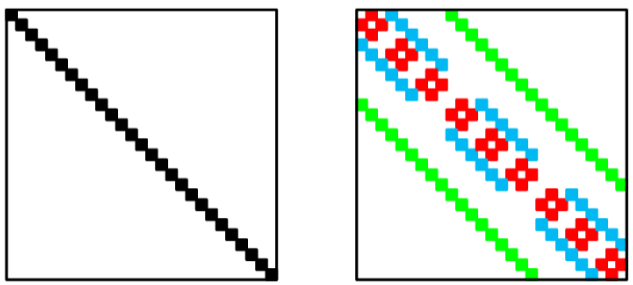
$$\mathbf{A} = \mathbf{M} + \mathbf{N},$$

with two conditions

1. $\mathbf{M}$ easily invertible,
2. $\rho(\mathbf{M}^{-1}\mathbf{N}) < 1$.

Picard fixed point algorithm: $\mathbf{M}[\mathbb{E}_{n+1}] = \mathbf{F} - \mathbf{N}[\mathbb{E}_n]$.

Cessenat-Després decomposition $=$ Jacobi decomposition[a]



The two conditions are theoretically satisfied.

[a] O. Cessenat and B. Despres, (1998). Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem. SIAM journal on numerical analysis, 35(1), 255-299.

Cessenat-Després decomposition

Singular regular decomposition

$$\mathbf{A} = \mathbf{M} + \mathbf{N},$$

with two conditions

1. $\mathbf{M}$ easily invertible,
2. $\rho(\mathbf{M}^{-1}\mathbf{N}) < 1$.

Picard solver based on Jacobi decomposition:

$$\mathbf{M}[\mathbb{E}^{\mathrm{jac}}_{n+1}] = -\mathbf{N}[\mathbb{E}^{\mathrm{jac}}_{n}] + \mathbf{F}.$$

Randomly divergent due to rounding errors.

$\mathcal{D}_\Omega$	10λ	30λ	50λ	80λ	100λ
$\rho(\mathbf{M}^{-1}\mathbf{N})$	$1-0.0097$	$1-0.0033$	$1-0.0024$	$1+0.00064$	$1-0.00043$

Table: The largest eigenvalue of $\mathbf{M}^{-1}\mathbf{N}$ computed with respect to the size $\mathcal{D}_\Omega$ of the domain.

Krylov methods

Coercivity: $\mathrm{Re}(\mathbf{a}(\mathbb{E}, \mathbb{E}')) > 0 \rightarrow$ Good framework for GMRES/Krylov Galerkin method.

$$[\mathbb{E}_{N_{\mathrm{kry}}}] = \mathrm{KG}(\mathbf{A}, \mathbf{F}, N_{\mathrm{kry}})$$

We approximate the solution $\mathbb{E}_{N_{\mathrm{kry}}} \in \mathbb{K}^{N_{\mathrm{kry}}}$ of

(Krylov Galerkin) Find $[\mathbb{E}_{N_{\mathrm{kry}}}] \in [\mathbb{K}^{N_{\mathrm{kry}}}]$, $[\mathbb{E}']^* \mathbf{A} [\mathbb{E}_{N_{\mathrm{kry}}}] = [\mathbb{E}']^* \mathbf{F}$, $\forall [\mathbb{E}'] \in [\mathbb{K}^{N_{\mathrm{kry}}}]$,

(GMRES) Find $[\mathbb{E}_{N_{\mathrm{kry}}}] \in [\mathbb{K}^{N_{\mathrm{kry}}}]$, $[\mathbb{E}']^* \mathbf{A}^* \mathbf{A} [\mathbb{E}_{N_{\mathrm{kry}}}] = [\mathbb{E}']^* \mathbf{A}^* \mathbf{F}$, $\forall [\mathbb{E}'] \in [\mathbb{K}^{N_{\mathrm{kry}}}]$,

where the Krylov space of dimension N_{kry} is

$$\left[\mathbb{K}^{N_{\mathrm{kry}}}\right] := \mathrm{span} \underbrace{\left[\begin{pmatrix} \vdots \\ \mathbf{F} \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ \mathbf{AF} \\ \vdots \end{pmatrix}, \ldots, \begin{pmatrix} \vdots \\ \mathbf{A}^{N_{\mathrm{kry}}-1}\mathbf{F} \\ \vdots \end{pmatrix} \right]}_{N_{\mathrm{kry}}} \Bigg\} N_{\mathrm{dof}},$$

$$\Longrightarrow size(Krylov) = N_{\mathrm{kry}} \times N_{\mathrm{dof}}, \text{ with } dim(\mathbf{F}) = N_{\mathrm{dof}}.$$

Krylov method issues

- **The convergence rate of the GMRES method is slow down by the smallest eigenvalues of $\mathbf{A}$.** GMRES Convergence Theorem[i]

$$\|\mathbb{E} - \mathbb{E}_n\| \leq C\,\varepsilon_n\,\|\mathbf{F}\|,$$

with

$$\varepsilon_n = \inf_{deg(p)\leq n-1} \sup_{\lambda\in\Lambda} \left|\frac{1}{\lambda} - p(\lambda)\right|.$$

Better estimates also exist [ii].

- **Need to reduce memory for large domains**

$$size(Krylov) + size(\mathbf{A}) = N_{\text{kry}} \times N \times N_{\text{elem}} + N^2 \times 7 \times N_{\text{elem}}$$

Our objectives

1. **Move away from zero** the eigenvalues.
2. **Reduce memory** cost.

[i] Saad, Youcef and Schultz, Martin H,GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems,SIAM,7,3,856–869,1986

[ii] J. Liesen, P. Tichý, (2012). The field of values bound on ideal GMRES.

Restart strategies for Krylov methods

GMRES solver converges theoretically in N_{dof} iterations, *ie* $N_{\text{kry}} = N_{\text{dof}}$.

- N_{elem} the number of elements
- $N = 52$ the number of basis functions
- $N_{\text{dof}} = N \times N_{\text{elem}}$ the number of degrees of freedom

$$\rightarrow \textbf{size(Krylov)} = N_{\text{kry}} \times N_{\text{dof}} = 52 \times 10^6 \times 52 \times 10^6 = \mathbf{43264}\ \textbf{Tb}.$$

Problems when N_{dof} becomes large, *ie* when N_{elem} is large

- High memory cost to store $\mathbb{K}^{N_{\text{kry}}}$.
- Accumulation of rounding errors at each iteration.

Idea: **Truncate the Krylov space** $\mathbb{K}^{N_{\text{kry}}}$ with $N_{\text{kry}}^{restart} < N_{\text{kry}}$

$$[\mathbb{E}]^{\text{k}} = [\mathbb{E}]^{\text{k}-1} + \text{GMRES}(\mathbf{A}, \mathbf{F} - \mathbf{A}[\mathbb{E}]^{\text{k}-1}), \quad [\mathbb{E}]^0 = 0.$$

Example with $N_{\text{kry}}^{restart} = 20$:

$$\textbf{size(Krylov)} = N_{\text{kry}}^{restart} \times N_{\text{dof}} = 20 \times 52 \times 10^6 = \mathbf{16}\ \textbf{Gb}.$$

$\rightarrow$ **restart strategy reduces the memory cost.**

Matrix free strategy

The **memory cost** of GoTEM3 **with matrix storage** is

$$\text{size(Krylov)} + \text{size(A)} = 20 \times N \times N_{\text{elem}} + N^2 \times 7 \times N_{\text{elem}}$$

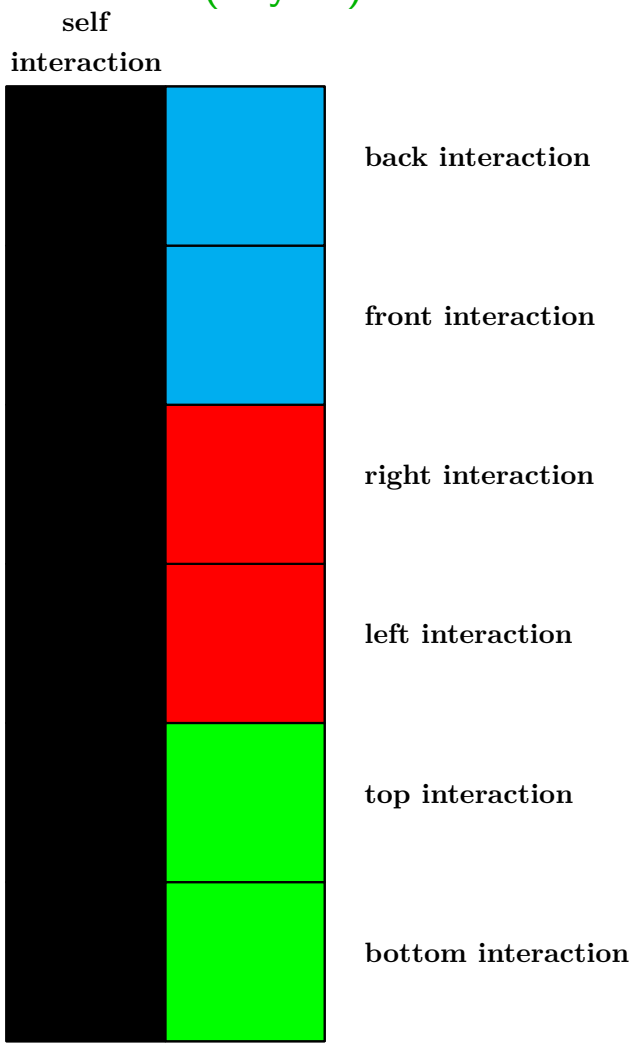
$$\frac{\text{size(A)}}{\text{size(Krylov)}} \simeq 7 \rightarrow \quad \textbf{when assembling :} \text{ the most costly structure is } \mathbf{A}.$$

Matrix free strategy

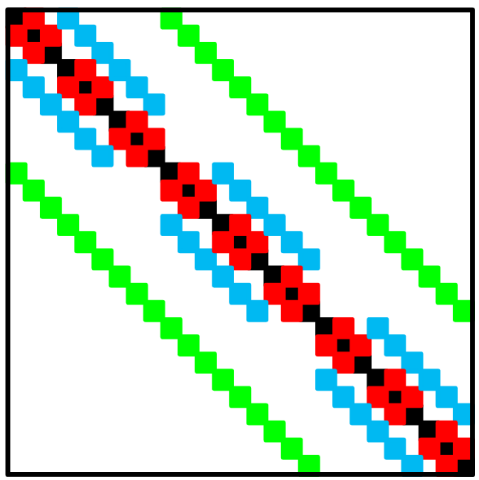
The **memory cost** of GoTEM3 **with matrix storage** is

$$\text{size(Krylov)} + \text{size(A)} = 20 \times N \times N_{\text{elem}} + N^2 \times 7 \times N_{\text{elem}}$$

$$\frac{\text{size(A)}}{\text{size(Krylov)}} \simeq 7 \rightarrow \quad \textbf{when assembling :} \text{ the most costly structure is } \mathbf{A}.$$



Structured mesh



$\mathbf{A}$

$U \mapsto \mathbf{A}\,U$ is an element by element product.

Matrix free strategy

The **memory cost ~~without~~ with** matrix free strategy is

$$\text{size(Krylov)} + \cancel{\text{size(A)}} = 20 \times N \times N_{\text{elem}} + \cancel{N^2 \times 7 \times N_{\text{elem}}}$$

$$\text{size(Krylov)} + \text{size(interactions)} = 20 \times N \times N_{\text{elem}} + N^2 \times 7$$

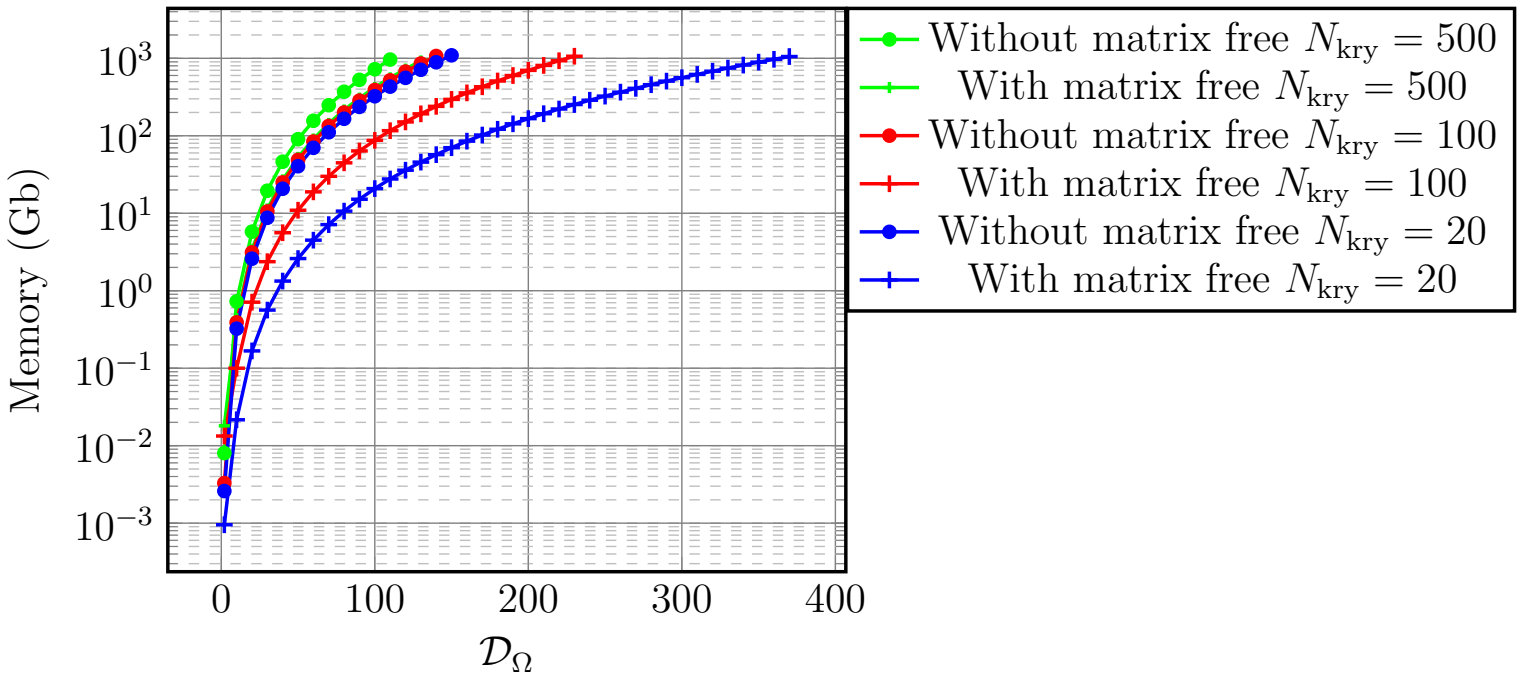
with matrix free: only the interactions between elements are stored

- $N_{\text{kry}} = 20$ the dimension of the Krylov space
- N_{elem} the number of basis functions
- $N = 52$ the number plane waves per element
- $N_{\text{dof}} = N \times N_{\text{elem}}$ the number of degrees of freedom

$\mathcal{D}_\Omega$	5λ	10λ	20λ	100λ	200λ
N_{elem}	25	1000	8000	1×10^6	8×10^6
N_{dof}	1300	52×10^3	416×10^3	52×10^6	416×10^6
Without matrix free	7.9 Mb	0.31 Gb	2.5 Gb	0.31 Tb	2.5 Tb
With matrix free	0.71 Kb	16 Mb	0.13 Gb	16 Gb	0.13 Tb

Table: Comparison with or without matrix free strategy.

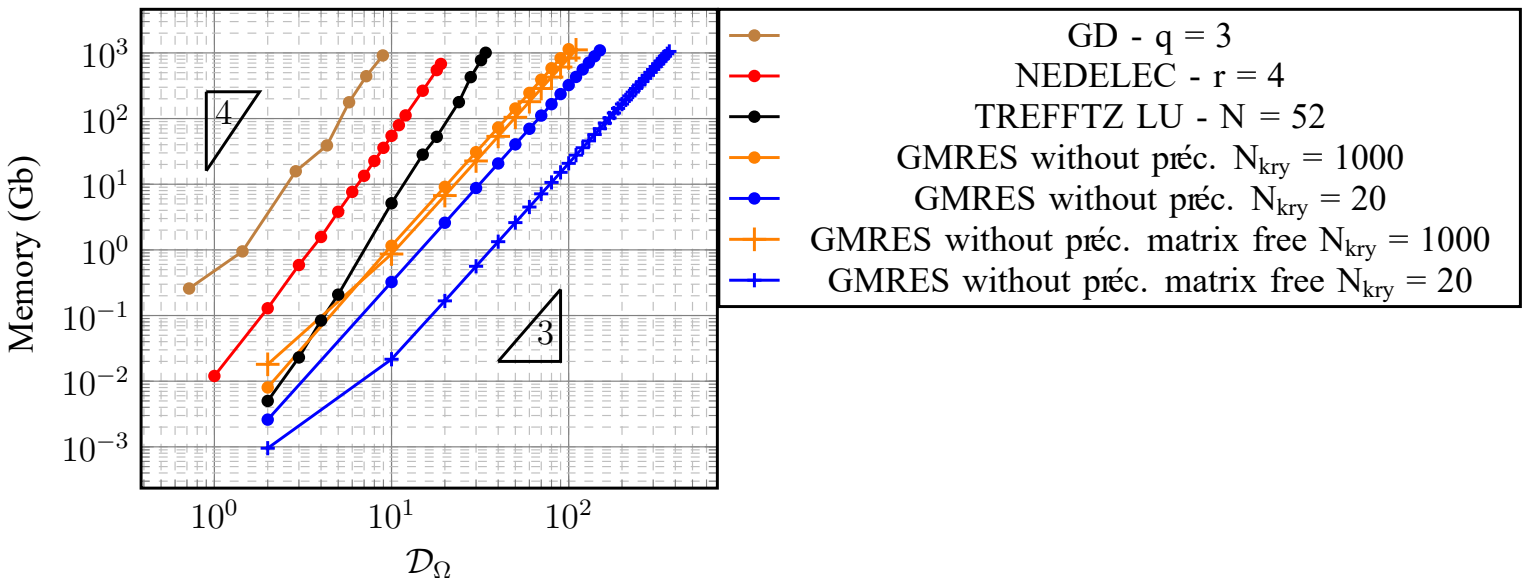
Reduction of the memory cost with the matrix free strategy



Comparison of the GMRES method memory costs with or without matrix free startegy, for different N_{kry}.

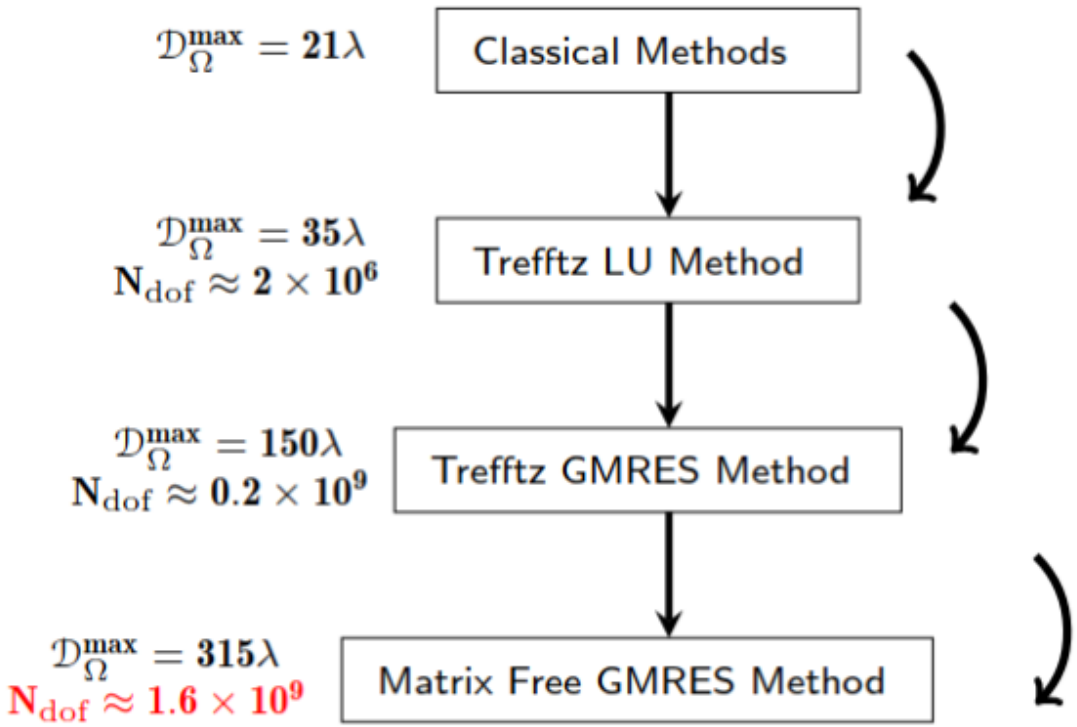
Restart $\searrow$ $\Longrightarrow$ Matrix free effect $\nearrow$

Comparison of memory costs



GoTEM3 : $\mathcal{D}_\Omega^{\max} = 315\lambda \iff N_{\text{elem}} = 30 \times 10^6 \iff N_{\text{dof}} = 1.6 \times 10^9$

Larger computational scenes



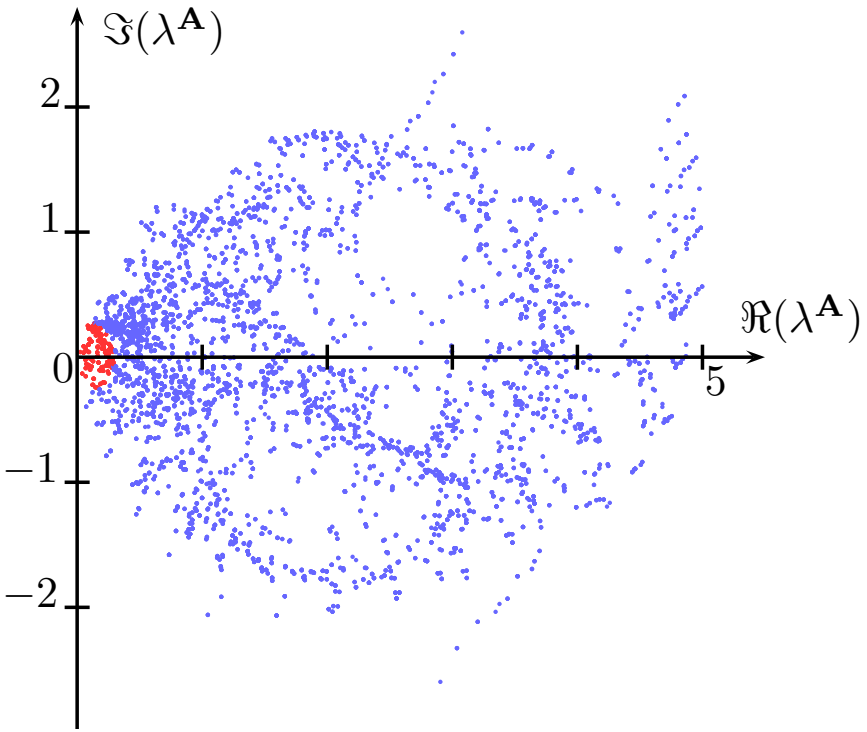
$\mathcal{D}_{\Omega}^{\max} = \mathbf{315\lambda}$ for $1\mathrm{Tb}$
$\approx 1.6 \times 10^9$ degrees of freedom

Cessenat-Després preconditioning

The convergence of the GMRES method depends on the spectrum of $\mathbf{A}$.

Objective: **move away from** 0 eigenvalues and avoid rounding errors.

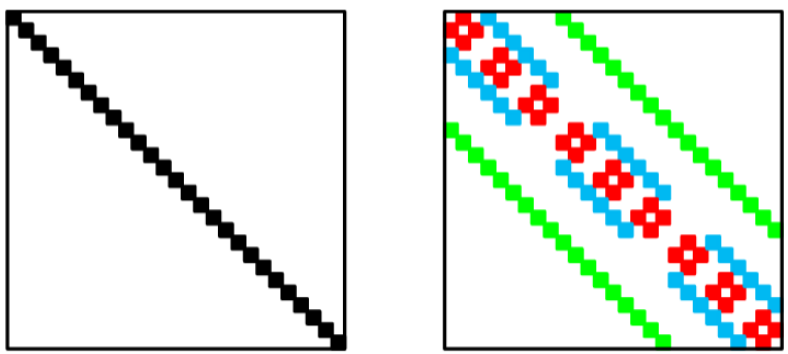
Idea: use **Preconditioned GMRES** which solves $\mathbf{A}^{\#}\mathbf{A}[\mathbb{E}] = \mathbf{A}^{\#}\mathbf{F}$.



Real and imaginary parts of the eigenvalues $\lambda^{\mathbf{A}}$ of the matrix $\mathbf{A}$, for a domain with size $\mathcal{D}_\Omega = 6\lambda$.

$$\mathbf{A} = \mathbf{M} + \mathbf{N}$$

Cessenat-Després decomposition [a]

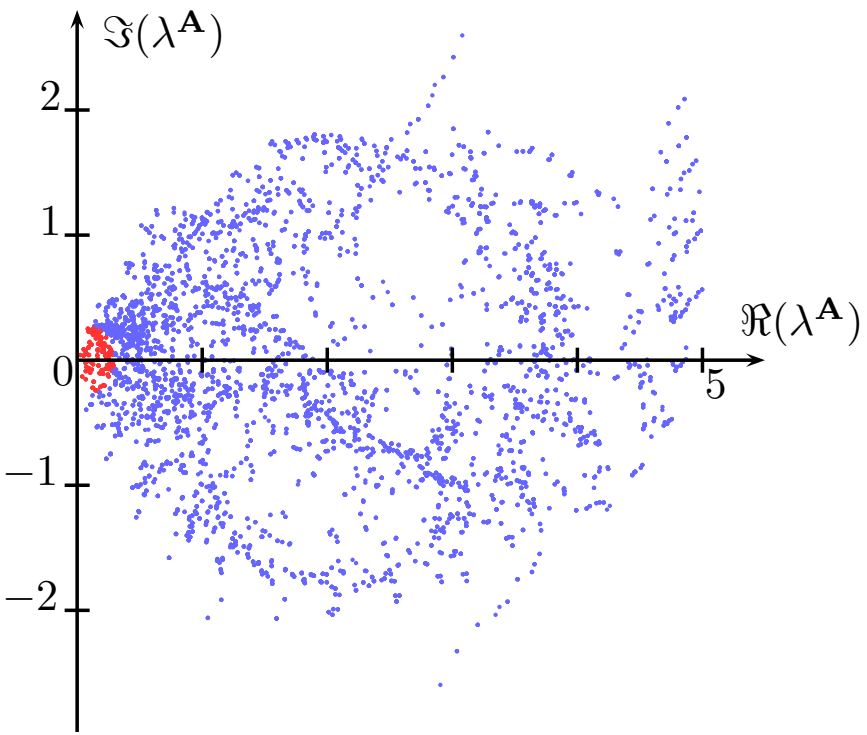


[a] O. Cessenat and B. Despres, (1998). Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem. SIAM journal on numerical analysis, 35(1), 255-299.

Cessenat-Després preconditioning

The convergence of the GMRES method depends on the spectrum of $\mathbf{A}$.

Objective: **move away from** 0 eigenvalues and avoid rounding errors.



Real and imaginary parts of the eigenvalues $\lambda^{\mathbf{A}}$ of the matrix $\mathbf{A}$, for a domain with size $\mathcal{D}_\Omega = 6\lambda$.

Idea: use a **preconditioned GMRES** which solves $\mathbf{A}^{\#}\mathbf{A}[\mathbb{E}] = \mathbf{A}^{\#}\mathbf{F}$.

Cessenat-Després iteration

$$\mathbf{M}[\mathbb{E}]^{n+1} = -\mathbf{N}[\mathbb{E}]^{n} + \mathbf{F}$$

Define $\mathbf{A}^{\#} = \mathbf{M}^{-1}$ with $\mathbf{M}$ block diagonal.

Symmetric preconditioning

$$\widetilde{\mathbf{A}} = \mathbf{M}^{-\frac{1}{2}}\mathbf{A}\mathbf{M}^{-\frac{1}{2}} = \mathbf{Id} + \mathbf{M}^{-\frac{1}{2}}\mathbf{N}\mathbf{M}^{-\frac{1}{2}}$$

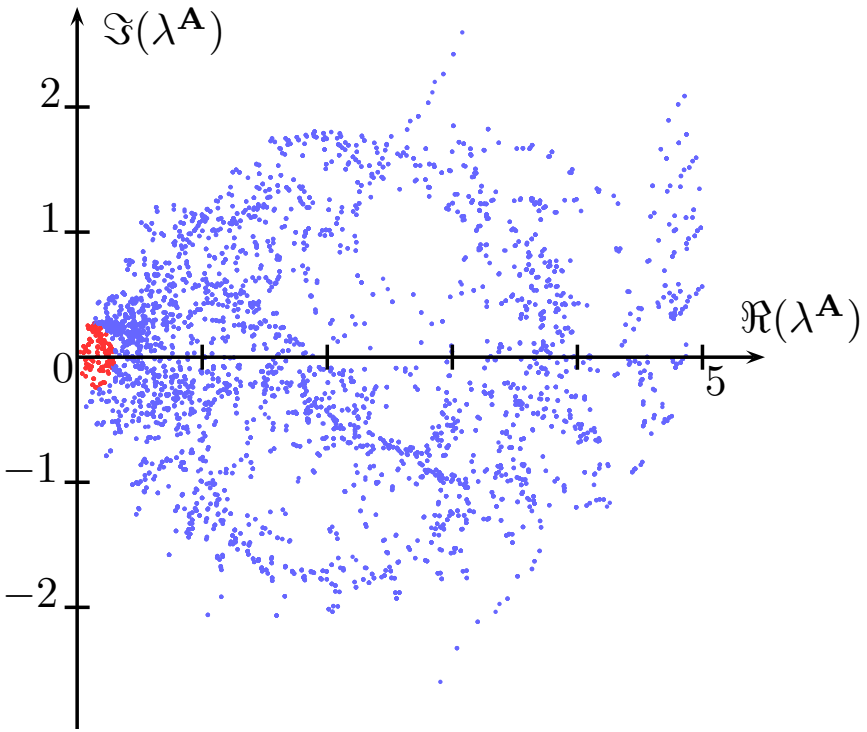
Computation of $\mathbf{M}^{-\frac{1}{2}}$ is trivial.

Cessenat-Després preconditioning

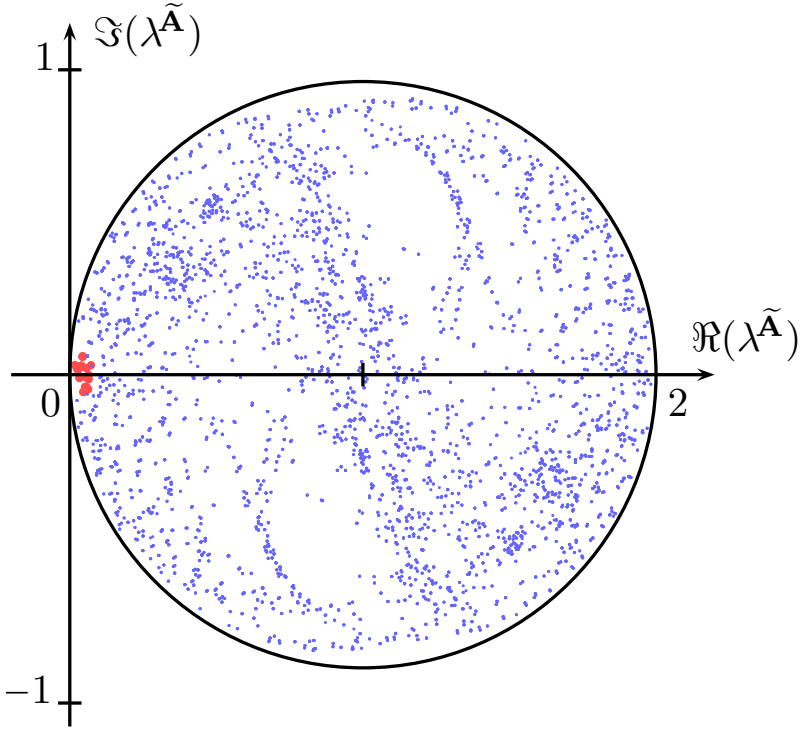
The convergence of the GMRES method depends on the spectrum of $\mathbf{A}$.

Objective: **move away from** 0 eigenvalues and avoid rounding errors.

$$\tilde{\mathbf{A}} = \mathbf{Id} + \mathbf{M}^{-\frac{1}{2}}\mathbf{N}\mathbf{M}^{-\frac{1}{2}}$$
is contractant



Real and imaginary parts of the eigenvalues $\lambda^{\mathbf{A}}$ of the matrix $\mathbf{A}$, for a domain with size $\mathcal{D}_\Omega = 6\lambda$.



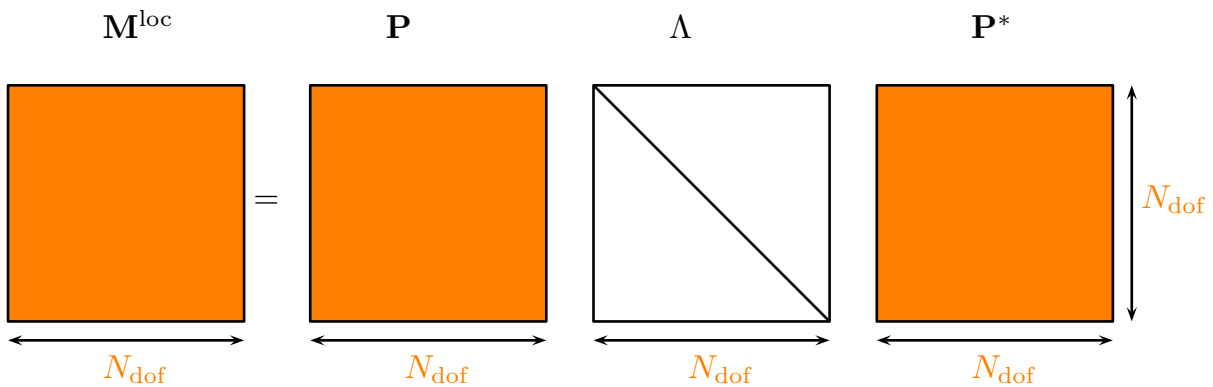
Real and imaginary parts of the eigenvalues $\lambda^{\widetilde{\mathbf{A}}}$ of the matrix $\widetilde{\mathbf{A}}$, for a domain with size $\mathcal{D}_\Omega = 6\lambda$.

Basis reduction strategy

$\mathbf{M}^{\mathrm{loc}}$ is a **symmetric positive definite** matrix

$$\mathbf{M}^{\mathrm{loc}}_{\ell,k} = (\gamma_{\mathrm{out}}\mathbf{w}_k, \gamma_{\mathrm{out}}\mathbf{w}_\ell)$$

for $\ell, k = 1, N$.



$\mathbf{P}$ is the eigenvector matrix.
Λ is the eigenvalue matrix.

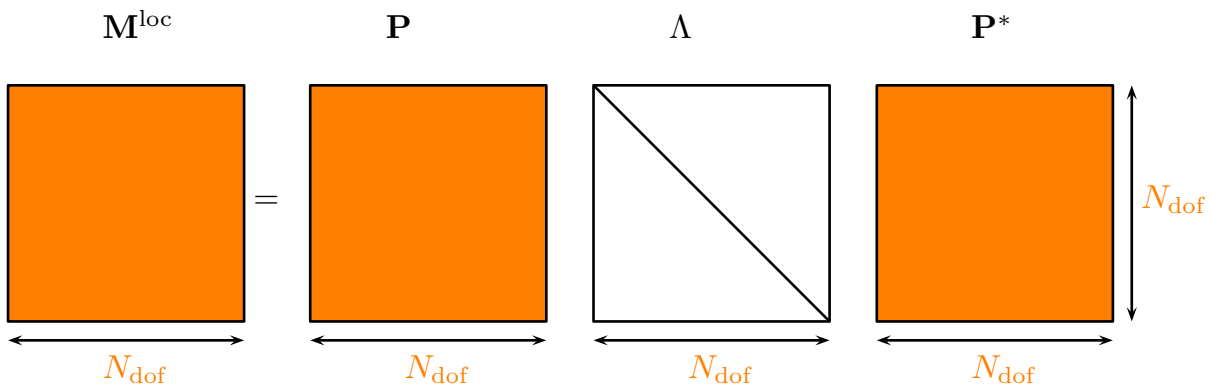
$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_{\mathrm{dof}}} \quad \text{with } \lambda_{\mathrm{i}} = \Lambda_{\mathrm{i,i}}.$$

Basis reduction strategy

$\mathbf{M}^{\text{loc}}$ is a **symmetric positive definite** matrix

$$\mathbf{M}^{\text{loc}}_{\ell,k} = (\gamma_{\text{out}}\mathbf{w}_k, \gamma_{\text{out}}\mathbf{w}_\ell)$$

for $\ell, k = 1, N$.

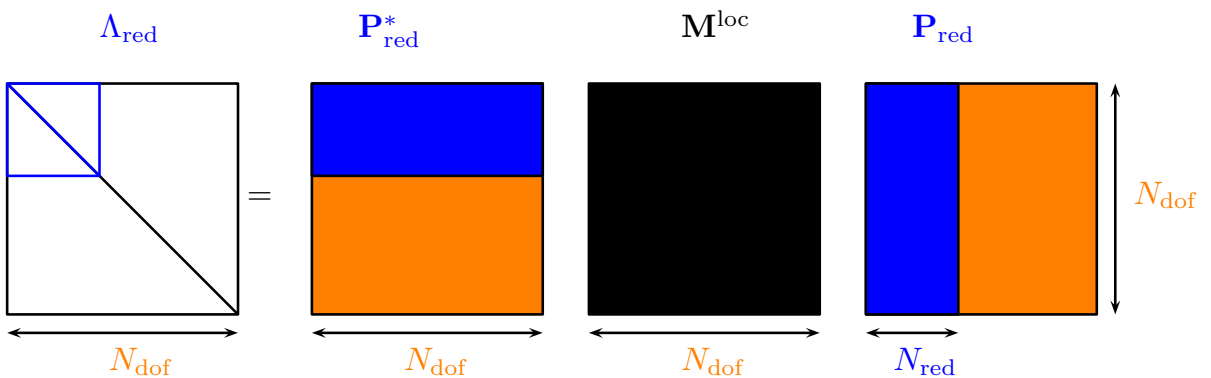


$\mathbf{P}$ is the eigenvector matrix.
Λ is the eigenvalue matrix.

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_{\text{dof}}} \quad \text{with } \lambda_{\text{i}} = \Lambda_{\text{i,i}}.$$

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{N_{\text{red}}} \\ \vdots \\ \lambda_{N_{\text{dof}}-1} \\ \lambda_{N_{\text{dof}}} \end{pmatrix} \rightarrow \textbf{Filtering} \rightarrow \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{N_{\text{red}}} \end{pmatrix}$$

$$\frac{\lambda_{\text{i}}}{\lambda_1} < \varepsilon$$



Normalized Reduced Basis

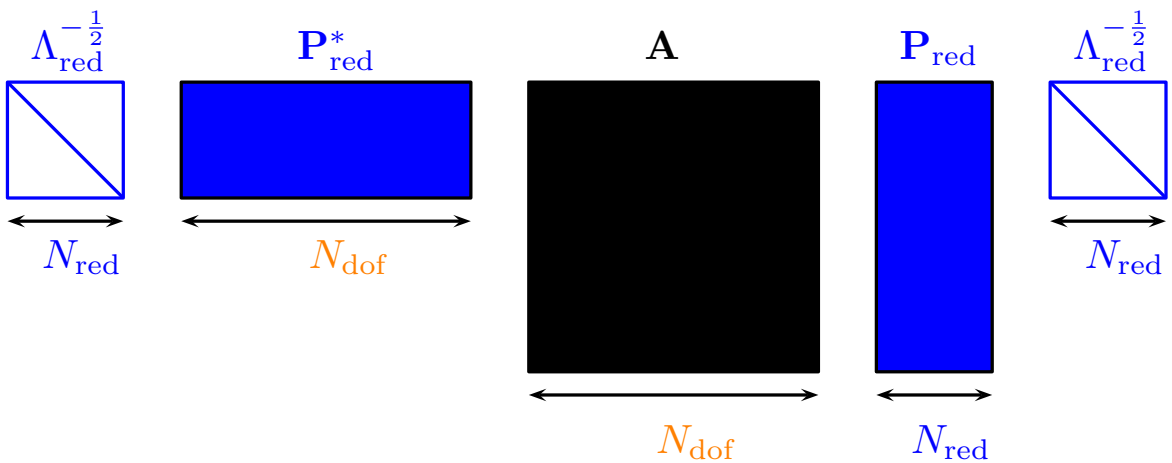
The unknown of the problem is represented as

$$\mathbb{E} = \sum_{i=1}^{N} [\mathbb{E}]_i \mathbf{w}^i = \sum_{i=1}^{N} [\mathbf{y}]_i \widetilde{\mathbf{w}}^i \simeq \sum_{i=1}^{N_{red}} [\mathbf{y}]_i \widetilde{\mathbf{w}}^i, \quad \mathbf{Y}_{red} = [\mathbf{y}]_{i=1,N_{red}}.$$

This approximation **does not erase information** since $\frac{\lambda_i}{\lambda_1} < \varepsilon$, with ε well chosen.

The reduced basis can be normalized and the new matrix is

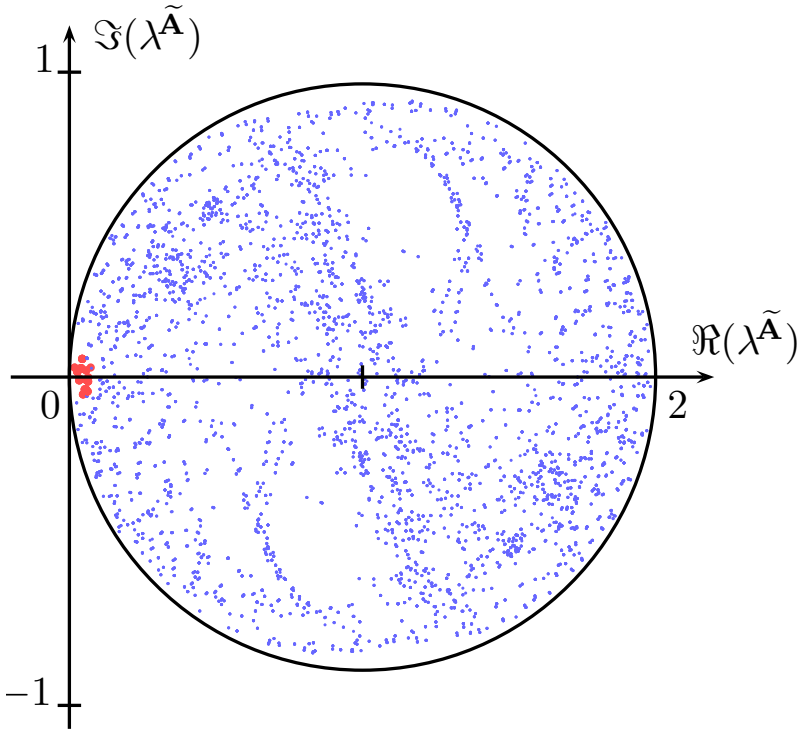
$$\mathbf{A}_{red} = \mathbf{I}_{red} + \mathbf{N}_{red} =$$



$[\mathbb{E}^{n+1}] = -\mathbf{N}_{red}[\mathbb{E}^n] + \mathbf{F}_{red}$, with $\mathbf{N}_{red} := \Lambda_{red}^{-\frac{1}{2}} \mathbf{P}_{red}^{*} \mathbf{N} \mathbf{P}_{red} \Lambda_{red}^{-\frac{1}{2}}$, and $[\mathbb{E}^0] = 0$.

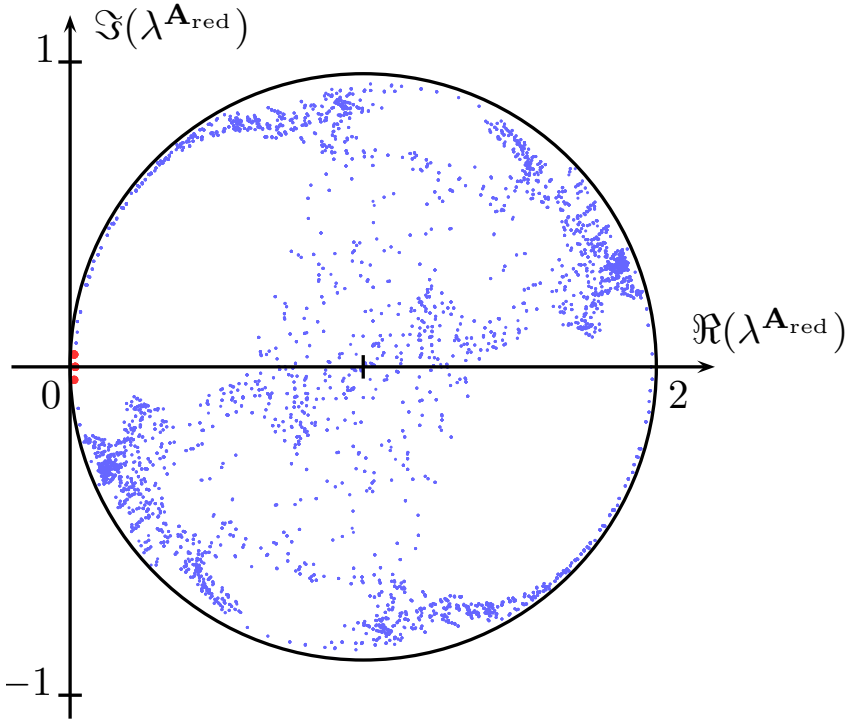
Improvement of the conditioning with basis reduction

Non reduced preconditioned matrix $\widetilde{\mathbf{A}}$



Real and imaginary parts of the eignevalues $\lambda^{\widetilde{\mathbf{A}}}$ of $\widetilde{\mathbf{A}}$, for $\mathcal{D}_\Omega = 6\lambda$.

Reduced preconditioned matrix $\mathbf{A}_{\mathrm{red}}$



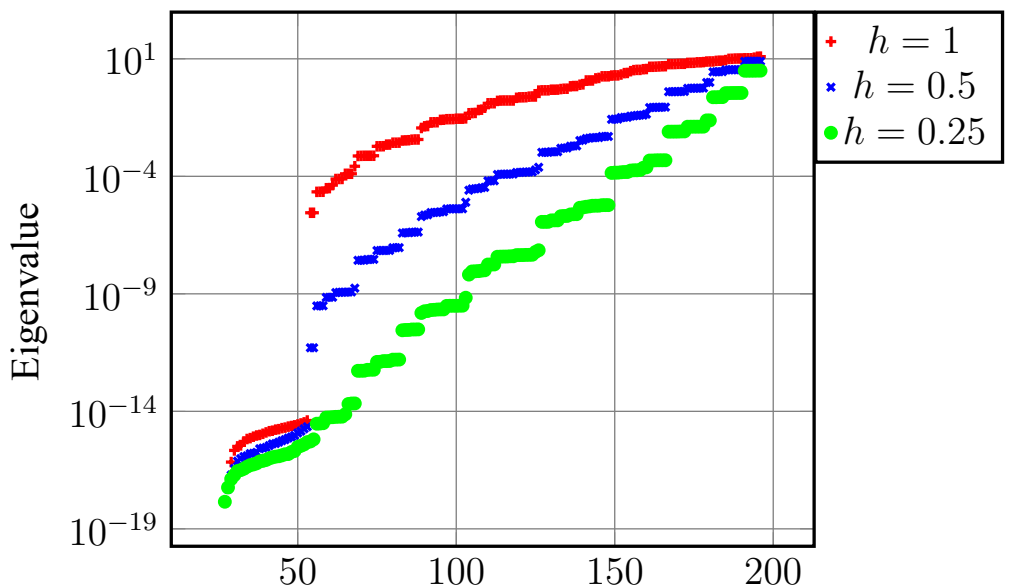
Real and imaginary parts if the eigenvalues $\lambda^{\mathbf{A}_{\mathrm{red}}}$ of $\mathbf{A}_{\mathrm{red}}$, for $\mathcal{D}_\Omega = 6\lambda$ and $\varepsilon = 10^{-5}$.

Eigenvalues of $\mathbf{M}^{\mathrm{loc}}$

$\mathbf{N}_{\mathrm{red}}$ depends on the size of each element T and the threshold value ε.

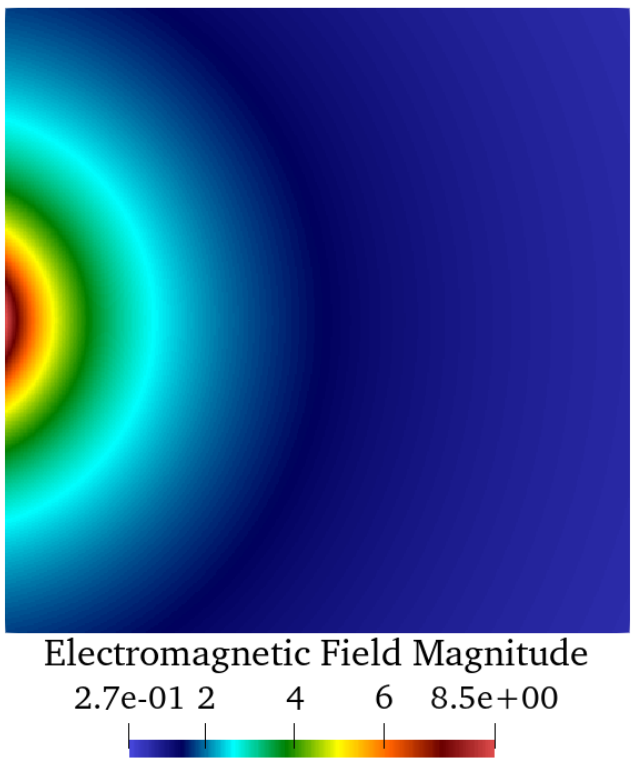
h \ ε	10^{-16}	10^{-15}	10^{-13}	10^{-11}	10^{-9}	10^{-7}	10^{-5}	10^{-4}	10^{-3}	10^{-2}
0.25	180	175	154	126	96	70	48	36	30	16
0.5	196	196	190	186	174	132	96	84	70	48
1	196	196	196	196	196	190	180	174	148	114

Table: Values of N_{red} with respect to ε and h, when performing basis reduction for $N = 196$ plane waves per element.



The quality of the solution depends on ε

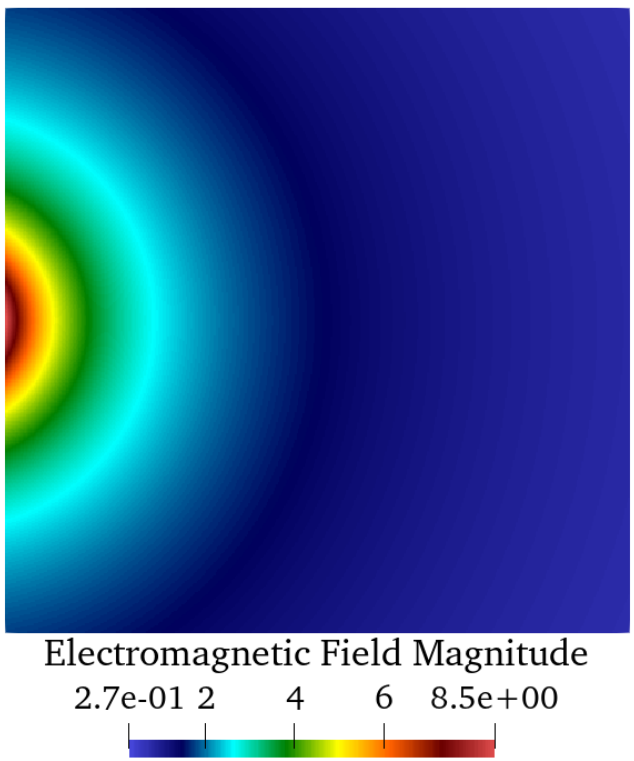
Field generated by a dipolar source located on the left of the domain for different values of ε with $N = 196$, $\mathcal{D}_\Omega = 5\lambda$, $h = 0.25\lambda$.



Electromagnetic field magnitude for $\varepsilon = 10^{-16}$, *ie* $N_{\text{red}} = 180$.

The quality of the solution depends on ε

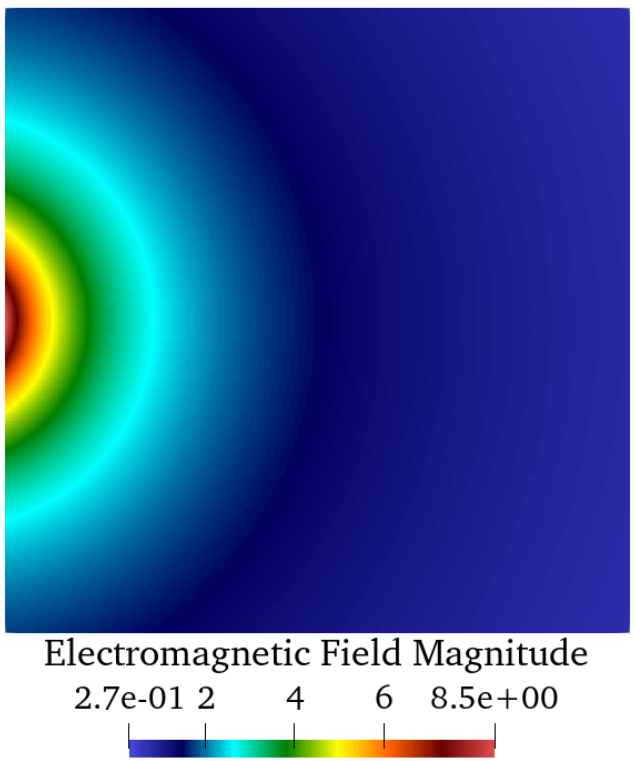
Field generated by a dipolar source located on the left of the domain for different values of ε with $N = 196$, $\mathcal{D}_\Omega = 5\lambda$, $h = 0.25\lambda$.



Electromagnetic field magnitude for $\varepsilon = 10^{-9}$, *ie* $N_{\text{red}} = 96$.

The quality of the solution depends on ε

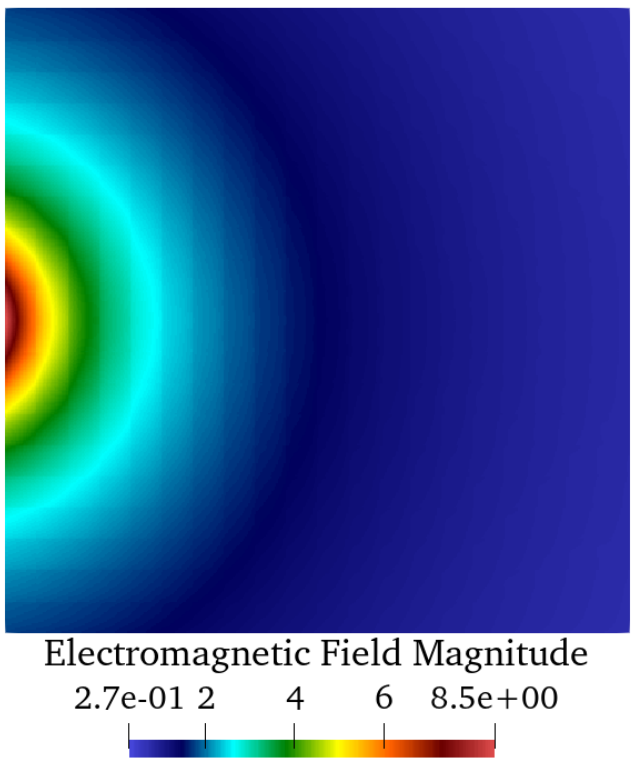
Field generated by a dipolar source located on the left of the domain for different values of ε with $N = 196$, $\mathcal{D}_\Omega = 5\lambda$, $h = 0.25\lambda$.



Electromagnetic field magnitude for $\varepsilon = 10^{-5}$, *ie* $N_{\text{red}} = 48$.

The quality of the solution depends on ε

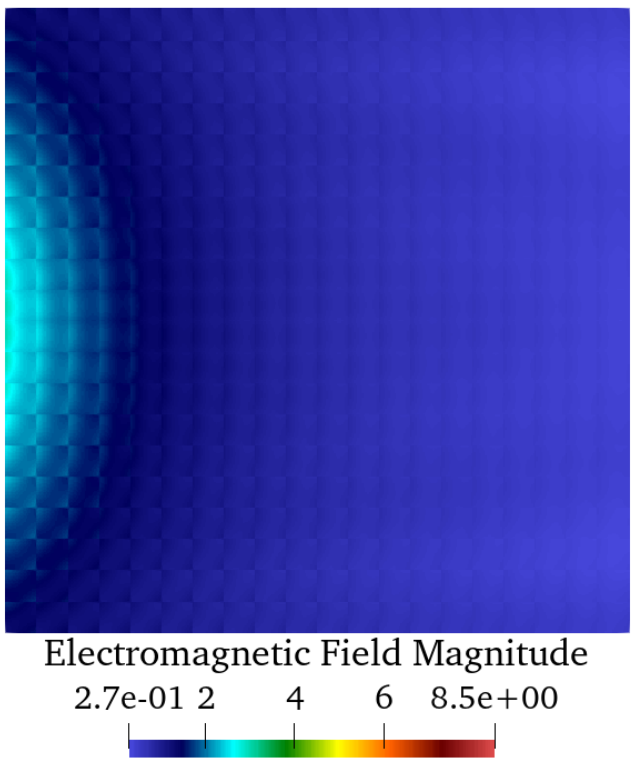
Field generated by a dipolar source located on the left of the domain for different values of ε with $N = 196$, $\mathcal{D}_\Omega = 5\lambda$, $h = 0.25\lambda$.



Electromagnetic field magnitude for $\varepsilon = 10^{-4}$, *ie* $N_{\text{red}} = 36$.

The quality of the solution depends on ε

Field generated by a dipolar source located on the left of the domain for different values of ε with $N = 196$, $\mathcal{D}_\Omega = 5\lambda$, $h = 0.25\lambda$.



Electromagnetic field magnitude for $\varepsilon = 10^{-2}$, *ie* $N_{\text{red}} = 16$.

The quality of the solution depends on ε

Field generated by a dipolar source located at the left of the domain for different values of ε with $N = 196$, $\mathcal{D}_\Omega = 5\lambda$, $h = 0.25\lambda$.

$$e^\varepsilon := \frac{\|\mathbb{E}^\varepsilon_{\text{red}} - \mathbb{E}^{\text{ex}}\|_\infty}{\|\mathbb{E}^{\text{ex}}\|_\infty}, \quad \text{with } \|\mathbb{E}\|_\infty := \max_{i=1,27} \mathbb{E}(x_i).$$

ε	10^{-16}	10^{-9}	10^{-5}	10^{-4}	10^{-2}
N_{red}	180	96	48	36	16
$N_{\text{dof,red}}$	1440000	768000	384000	288000	128000
Time (s)	4121	1085	254.9	141.8	25.61
Memory (Gb)	0.15	0.08	0.04	0.03	0.01
e^ε	6.35×10^{-7}	4.83×10^{-5}	6.33×10^{-3}	2.39×10^{-2}	0.59

Table: Results for a GMRES residual $= 10^{-12}$ ($N_{\text{kry}} = 100$).

The larger is ε $\rightarrow$ the less rounding error $\rightarrow$ improvement of the conditionning
Too much reduction $\rightarrow$ bad quality of the approximation

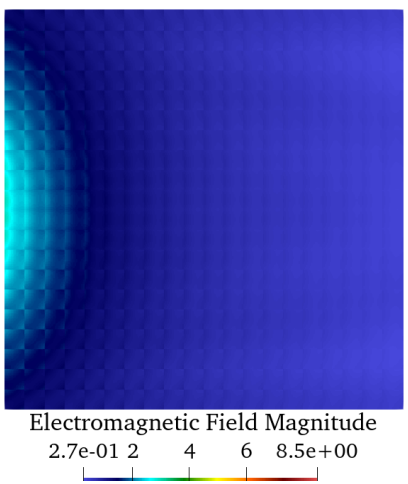
The quality of the solution depends on ε

Improvement on the computation time : 94%
Improvement on the memory cost : 75%

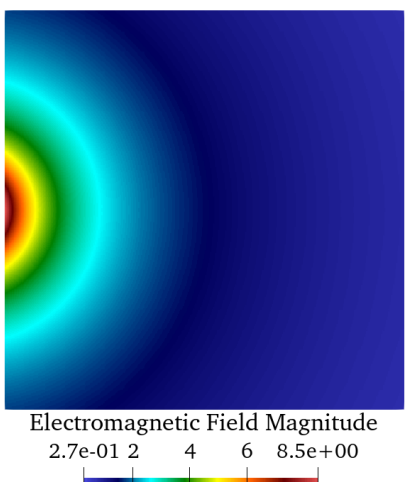
ε	10^{-16}	10^{-9}	10^{-5}	10^{-4}	10^{-2}
N_{red}	180	96	48	36	16
$\#\text{dof}_{\text{red}}$	1440000	768000	384000	288000	128000
Time (s)	4121	1085	254.9	141.8	25.61
Memory (Gb)	0.15	0.08	0.04	0.03	0.01
e^{ε}	6.35×10^{-7}	4.83×10^{-5}	6.33×10^{-3}	2.39×10^{-2}	0.59

GMRES residual $=10^{-12}$ but the approximation is too poor when ε is not correctly chosen.

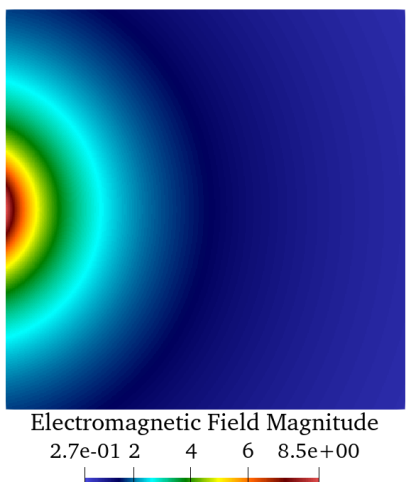
Too much reduction
$N_{\text{red}} = 16$, $\varepsilon = 10^{-2}$



Optimal trade-off
$N_{\text{red}} = 48$, $\varepsilon = 10^{-5}$



Reference solution
$N_{\text{red}} = 180$, $\varepsilon = 10^{-16}$

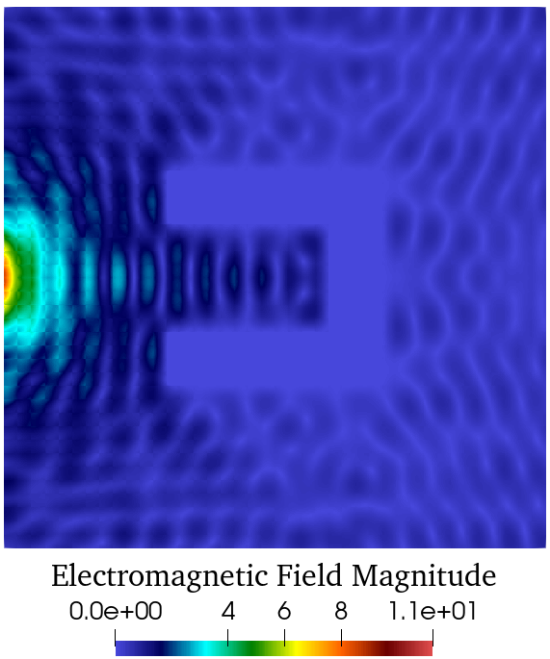


Application to a case with a scatterer

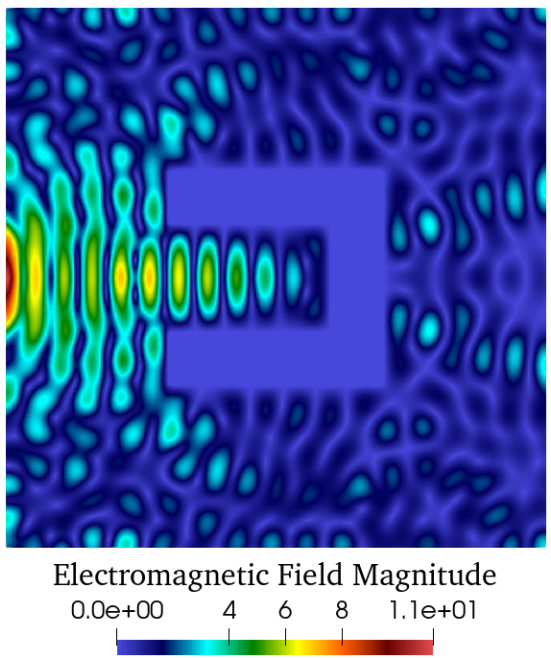
ε	10^{-7}	10^{-5}	10^{-4}	10^{-3}	10^{-2}
Trefftz error	N/A	3.4×10^{-2}	9.8×10^{-2}	0.15	0.74
Memory cost (Gb)	0.56	0.38	0.28	0.24	0.13
Time (hours)	55	27	11.3	7.23	0.58

GMRES residual $= 10^{-8}$ but the approximation is too poor when ε is not correctly chosen.

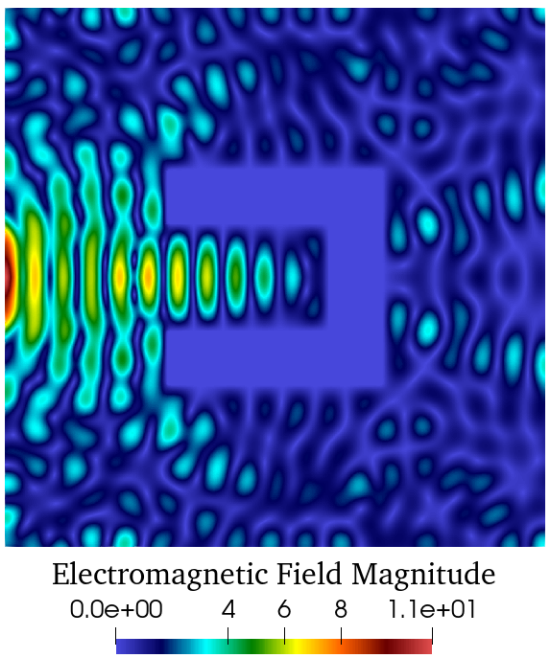
Too much reduction
$N_{\text{red}} = 16$, $\varepsilon = 10^{-2}$



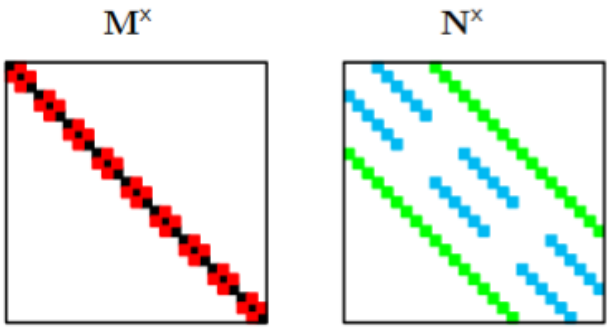
Optimal trade-off
$N_{\text{red}} = 36$, $\varepsilon = 10^{-4}$



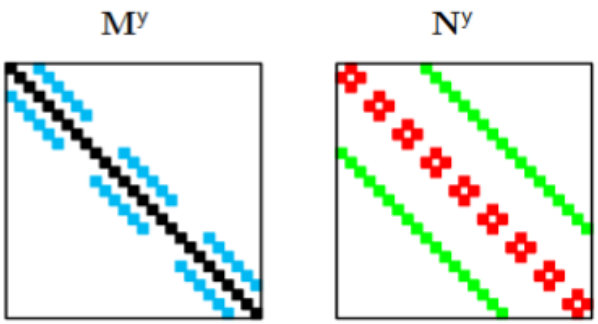
Reference solution
$N_{\text{red}} = 70$, $\varepsilon = 10^{-7}$



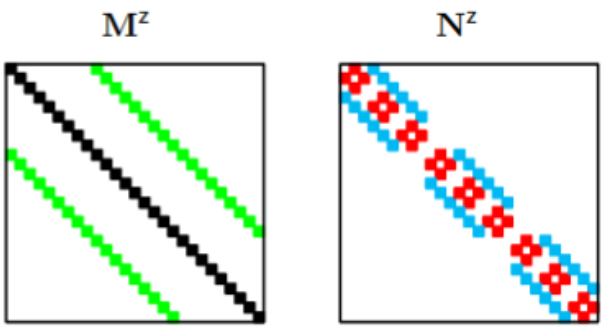
Global preconditioner



Structures of $\mathbf{M}^{x}$ and $\mathbf{N}^{x}$.



Structures of $\mathbf{M}^{y}$ and $\mathbf{N}^{y}$.



Structures of $\mathbf{M}^{z}$ and $\mathbf{N}^{z}$.

The new decomposition

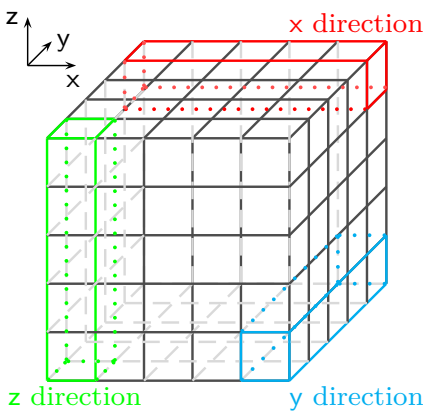
$$\mathbf{A} = \mathbf{M}^{x/y/z} + \mathbf{N}^{x/y/z},$$

can be used as preconditioners

1. $\mathbf{M}^{x/y/z}$ easily invertible,
2. ~~$\rho\left(\left(\mathbf{M}^{x/y/z}\right)^{-1}\mathbf{N}^{x/y/z}\right) < 1$~~,

and leads to preconditioners:

$$[\mathbf{y}] = \left(\mathbf{M}^{x/y/z}\right)^{-1}\mathbf{F}.$$



1D subdomains.

How to build a global preconditioner ?

By successively applying every preconditioner (x, y or z), we obtain a global preconditioner associating $[\mathbf{y}]$ to $[\mathbf{F}]$ with an iterative scheme

$$\begin{cases} \mathbf{M}^{\mathsf{x}}[\mathbb{E}_0] &= \mathbf{F}, \\ \mathbf{M}^{\mathsf{y}}[\mathbb{E}_1] &= \mathbf{F} - \mathbf{N}^{\mathsf{y}}[\mathbb{E}_0], \\ \mathbf{M}^{\mathsf{z}}[\mathbf{y}] &= \mathbf{F} - \mathbf{N}^{\mathsf{z}}[\mathbb{E}_1]. \end{cases}$$

The **reduced preconditioned GMRES problem** is

$$\mathbf{P}_{\mathrm{red}}^{\mathsf{xyz}} \mathbf{A}_{\mathrm{red}}[\mathbb{E}]_{\mathrm{red}} = \mathbf{P}_{\mathrm{red}}^{\mathsf{xyz}} \mathbf{F}_{\mathrm{red}},$$

where $\mathbf{P}_{\mathrm{red}}^{\mathsf{xyz}}$ is the reduced version of

$$\mathbf{P}^{\mathsf{xyz}} := \left(\mathbf{M}^{\mathsf{z}}\right)^{-1} - \left(\mathbf{M}^{\mathsf{z}}\right)^{-1} \mathbf{N}^{\mathsf{z}} \left(\mathbf{M}^{\mathsf{y}}\right)^{-1} + \left(\mathbf{M}^{\mathsf{z}}\right)^{-1} \mathbf{N}^{\mathsf{z}} \left(\mathbf{M}^{\mathsf{y}}\right)^{-1} \mathbf{N}^{\mathsf{y}} \left(\mathbf{M}^{\mathsf{x}}\right)^{-1}.$$

Better conditioning with successive decompositions

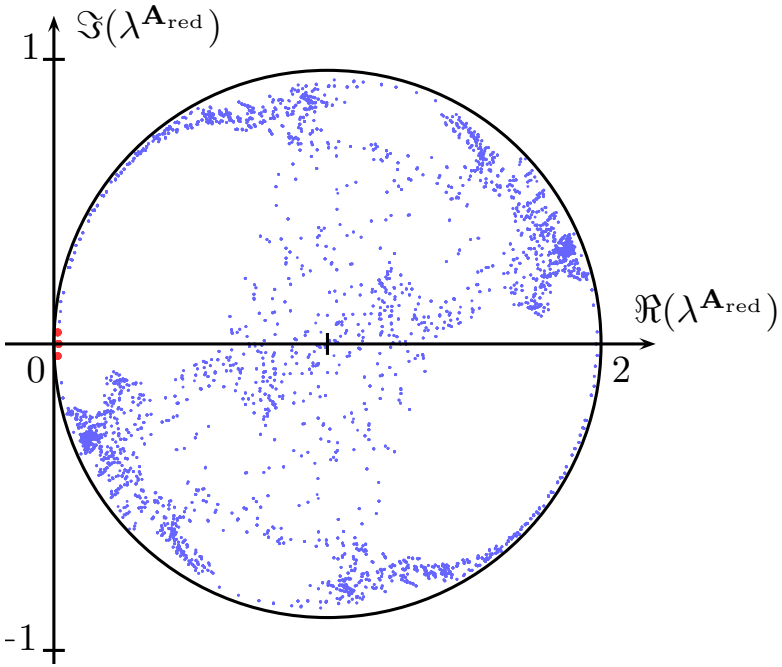


Figure: Real and imaginary parts of the eigenvalues $\lambda^{\mathbf{A}_{\mathrm{red}}}$ of $\mathbf{A}_{\mathrm{red}}$.

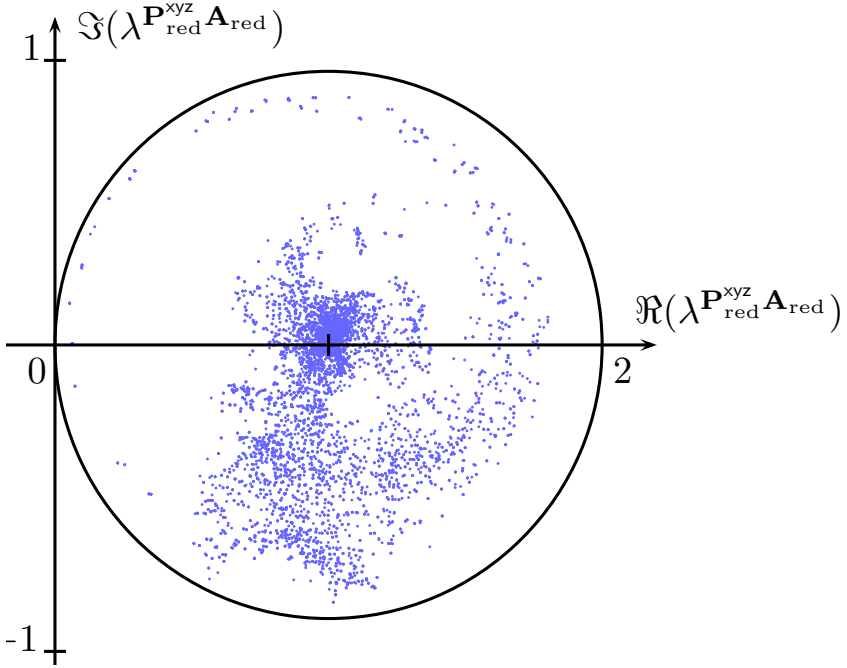


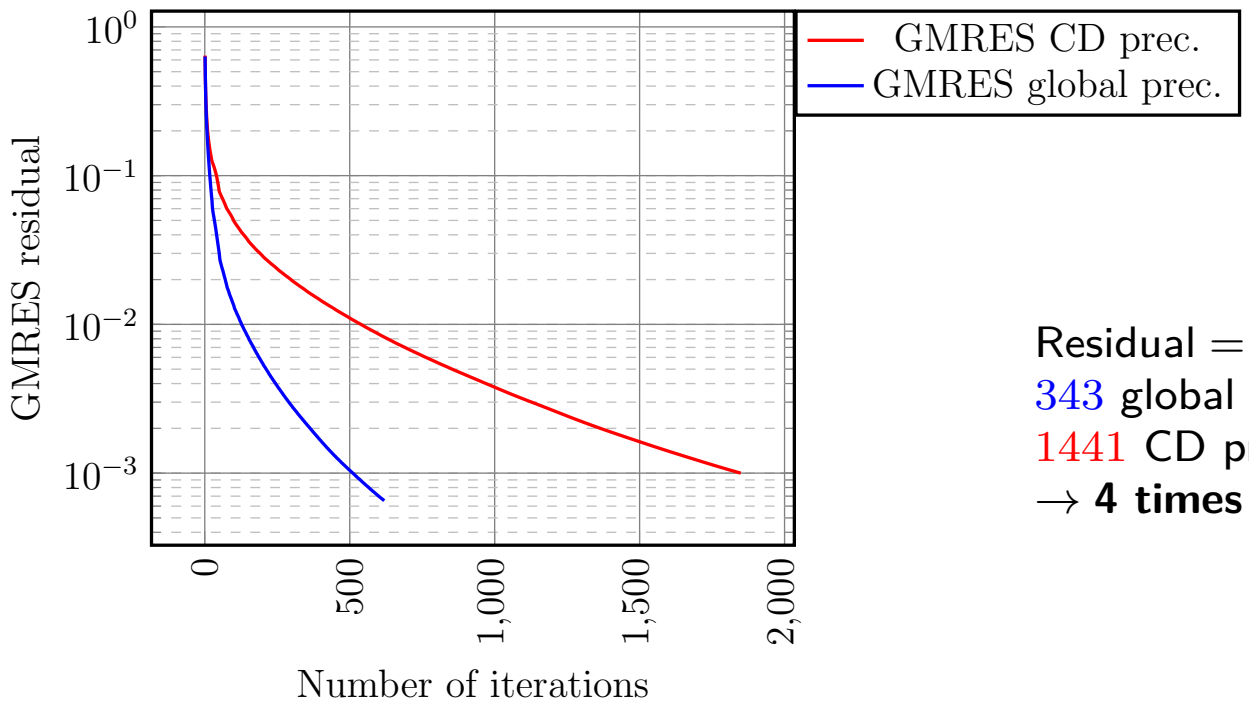
Figure: Real and imaginary parts of the eigenvalues $\lambda^{\mathbf{P}_{\mathrm{red}}^{\mathrm{xyz}}\mathbf{A}_{\mathrm{red}}}$ de $\mathbf{P}_{\mathrm{red}}^{\mathrm{xyz}}\mathbf{A}_{\mathrm{red}}$.

Convergence acceleration with the global preconditioner

Numerical case: electromagnetic dipole in a cup

$$\Omega = [0, 200] \times [0, 40] \times [0, 40],\ h = 1,\ R_{\partial\Omega} = 0.9,\ N = 196,\ N_{\text{red}} = 196,\ \varepsilon = 10^{-9},\ N_{\text{kry}} = 25.$$

$$\text{GMRES residual} = \frac{\|\mathbf{A}[\mathbb{E}] - \mathbf{F}\|_2}{\|\mathbf{F}\|_2}.$$



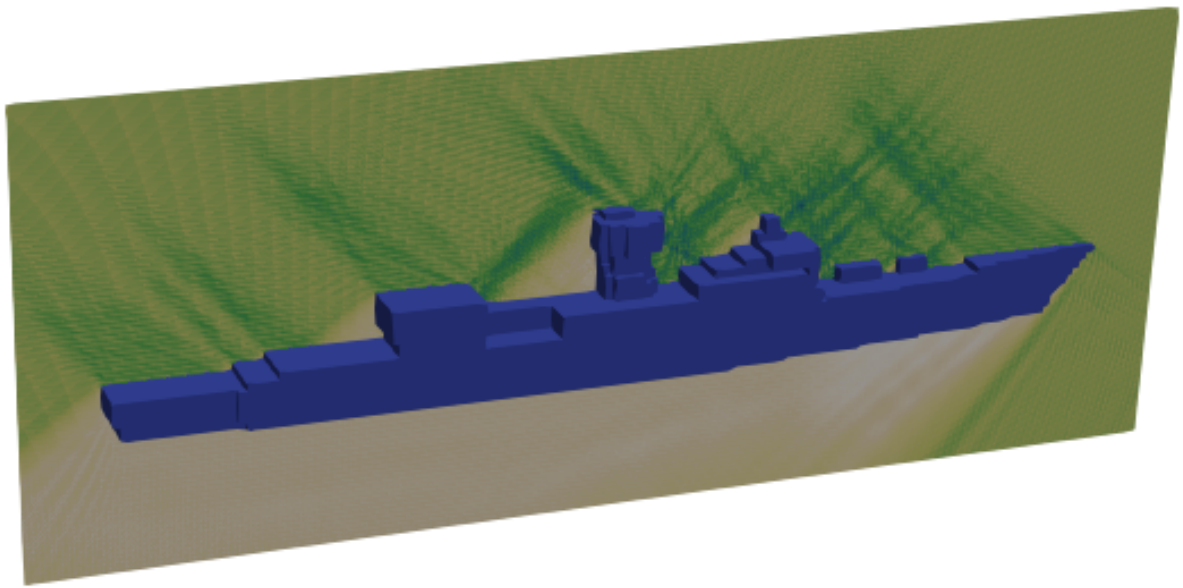
Residual $= 1.11 \times 10^{-2}$
343 global preconditioner
1441 CD preconditioner
$\rightarrow$ **4 times less iterations.**

Figure: Comparison of the convergence rate between the GMRES method preconditioned with the Cessenat-Després or the global preconditioners.

A numerical case

Boat size $= 24 \times 61 \times 154\ \lambda^3$.
$R_{\partial\Omega} = 0$.
Preconditioned GMRES with Cessenat-Després.

Interpretation : electromagnetic near field and shadow cone under the boat.



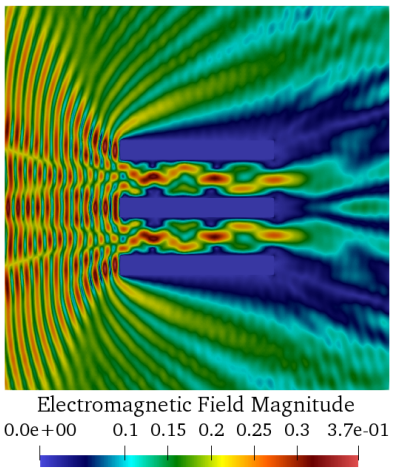
3D electromagnetic wave striking the boat from the top, for $\lambda = 1$ meter and $h = \lambda/4$.

#elem	N_{red}	N	$\#\text{dof}_{\text{red}}$	#dof
$> 14.4 \times 10^6$	46	52	$> 0.663 \times \mathbf{10^9}$	$> 0.75 \times \mathbf{10^9}$

GMRES iterations	Memory cost	Computation time	GMRES residual
800	389 Gb	28.6 hours	3.8×10^{-2}

Iterative Trefftz solver GoTEM3

Maximal size for 1Tb : $(315\lambda)^3$
$\approx 1.6 \times 10^9$ degrees of freedom



Thank you for your attention !

References, preprint :

- Fure, H.S., Pernet, S., Sirdey, M. et al. A discontinuous Galerkin Trefftz type method for solving the two dimensional Maxwell equations. SN Partial Differ. Equ. Appl. 1, 23 (2020).
- S. Pernet, N. Serdiuk, M. Sirdey, S. Tordeux. Discontinuous Galerkin Method based on Riemann fluxes for the time domain Maxwell System. [Research Report] RR-9431, INRIA Bordeaux - Sud-Ouest. 2021, pp.55. hal-03396721.
- S. Pernet, M. Sirdey, S. Tordeux. Ultra-Weak variational formulation for heterogeneous Maxwell problem in the context of high performance computing. 2022. hal-03642116.