



**HAL**  
open science

# The Steklov-Poincaré technique for data completion: preconditioning and filtering

Renaud Ferrier, Mohamed Larbi Kadri, Pierre Gosselet

► **To cite this version:**

Renaud Ferrier, Mohamed Larbi Kadri, Pierre Gosselet. The Steklov-Poincaré technique for data completion: preconditioning and filtering. *International Journal for Numerical Methods in Engineering*, 2018, 116 (4), pp.270-286. 10.1002/nme.5924 . hal-01632526

**HAL Id: hal-01632526**

**<https://hal.science/hal-01632526>**

Submitted on 10 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License

# The Steklov-Poincaré technique for data completion: preconditioning and filtering

Renaud Ferrier<sup>1</sup>, Mohamed L. Kadri<sup>2</sup> and Pierre Gosselet<sup>1</sup>

<sup>1</sup> LMT, ENS Paris-Saclay/CNRS, 61 av. du président Wilson, 94235 Cachan, France

<sup>2</sup> LAMSIN, ENIT, BP 37, 1002 Tunis, Tunisie

October 25, 2017

## Abstract

This article presents a study of primal and dual Steklov-Poincaré approaches for the identification of unknown boundary conditions of elliptic problems. After giving elementary properties of the discretized operators, we investigate the numerical solution with Krylov solvers. Different preconditioning and acceleration strategies are evaluated. We show that costless filtering of the solution is possible by post-processing Ritz elements. Assessments are provided on a 3D mechanical problem.

**Keywords:** boundary data completion, inverse problem, KMF and Steklov-Poincaré algorithms

## 1 Introduction

Inverse problems play an important role in mechanical engineering since they are associated to many non-destructive control techniques. We focus on the problem of missing boundary values in static linear elasticity: we suppose that the geometry and constituents of a structure are perfectly known, whereas a part of the boundary conditions is missing. To compensate this lack of knowledge, another part of the boundary is over-specified in the sense that both Dirichlet's and Neumann's are known. The challenge is thus to propagate this extra information till the unknown boundary.

Mathematically speaking this data completion problem is a Cauchy problem on an elliptic partial differential equation. Under compatibility assumptions on the data, there exists a unique solution to this problem. Anyhow, the problem is ill-posed in the sense of Hadamard, meaning that the solution is very sensitive to variations in the inputs (which is a practical problem because inputs are issued from measurements). For further discussion about the ill-posedness of this problem, the reader may refer to [8, 4].

Because of this instability regularization is more than often necessary to find a practical solution. For instance, in [25], the boundary element method is combined to Tykhonov regularization. Another possibility is to weaken the verification of the boundary conditions and set the problem in terms of the minimization of an error-like functional, see [22] for model updating and [1] for the Cauchy problem. In [9] a minimization coupled with an iterative regularization method was proposed. Also, it is possible to set the problem in a stochastic framework, which enables to recover well-posedness at the very expensive price of identifying a probability density function [18].

In this paper, we study a specific class of method, the Kozlov-Maz'ya-Formin (KMF) and Steklov-Poincaré (SP) methods [21, 3, 19]. These methods consist in formulating the Cauchy problem as the equality of the solutions of two direct problems involving respectively the Dirichlet and the Neumann data (without weakening). This equality results in a compact problem which can be solved either by truncated SVD techniques [15] or more simply by an iterative solver with adapted stopping criterion.

The aim of this paper is to confront some variants of the standard Krylov solvers to the Cauchy problem. In particular, preconditioning, acceleration and filtering techniques are investigated in order to accelerate the resolution and make it more accurate.

The paper is organized as follows: in section 2 the discrete linear systems associated with the Cauchy problem are presented. In section 3, the variants of the linear solver are presented and assessed on an academic example. Finally, in section 4, a numerical 3D example is addressed in order to illustrate the stability of the proposed method.

## 2 Systems associated with the data completion problem

### 2.1 Notations and hypothesis

Let  $\Omega \subset \mathbb{R}^d$  ( $d = 2$  or  $3$ ) be a domain where we wish to solve a linear elasticity problem under the hypothesis of small deformation. Thanks to the linearity of the problem, we consider the variation of the system around an equilibrated configuration, which enables us to remove the external loads. Let  $u$  be the unknown displacement field,  $\varepsilon$  the symmetric gradient operator,  $\mathbb{H}$  the Hooke's tensor,  $\sigma$  the Cauchy stress tensor and  $n$  the outward normal vector.

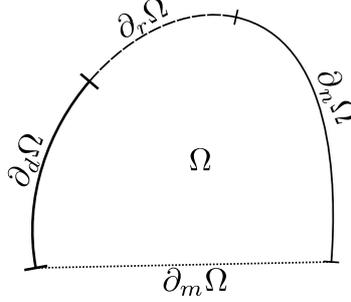


Figure 1: Notations for the boundary data completion problem.

We consider Figure 1. The boundary is split into four parts:  $\partial_d\Omega$  is submitted to Dirichlet's conditions,  $\partial_n\Omega$  is submitted to Neumann's conditions,  $\partial_m\Omega$  is the part of the boundary where data is missing – neither the displacement  $u_m$  nor the traction  $f_m$  are known – and  $\partial_r\Omega$  is the part of the boundary with redundant information – both Dirichlet's and Neumann's conditions, respectively  $\hat{u}_r$  and  $\hat{f}_r$ , are known. The equations of the Cauchy problem are thus:

$$\left. \begin{array}{l} \sigma = \mathbb{H} : \varepsilon(u) \\ \operatorname{div}(\sigma) = 0 \end{array} \right\} \text{in } \Omega, \quad \left. \begin{array}{l} u = 0 \text{ on } \partial_d\Omega \\ \sigma \cdot n = 0 \text{ on } \partial_n\Omega \end{array} \right\}, \quad \left. \begin{array}{l} u = \hat{u}_r \\ \sigma \cdot n = \hat{f}_r \end{array} \right\} \text{on } \partial_r\Omega \quad (1)$$

We suppose that  $\operatorname{meas}(\partial_d\Omega) > 0$ , so that all the forward problems considered afterwards will be well-posed. Note that this hypothesis can easily be lifted by using the techniques of the FETI domain decomposition method for floating substructures [11], this is briefly described in section 2.3.3.

We use a classical Finite Element approximation of the problem, using Lagrange elements to discretize the displacement field, and we note  $\mathbf{u}$  the vector of nodal displacement. We assume Dirichlet conditions were eliminated so that the stiffness matrix  $\mathbf{K}$  is sparse symmetric definite positive. We note with subscript  $i$  the degrees of freedom inside the domain and on the Neumann's boundary; with subscript  $r$  the redundant degrees of freedom where both displacement degrees of freedom  $\hat{\mathbf{u}}_r$  and nodal reactions  $\hat{\mathbf{f}}_r$  are known, with subscript  $m$  the missing degrees of freedom where neither the displacement  $\mathbf{u}_m$  nor the nodal reaction  $\mathbf{f}_m$  are known. We use the counting measure notation for the number of degrees of freedom, e.g.  $|m|$  is the number of degrees of freedom of type  $m$ .

The discrete equivalent to (1) can then be written as:

$$(\hat{\mathbf{u}}_r, \hat{\mathbf{f}}_r) \text{ being known, find } (\mathbf{u}_i, \mathbf{u}_m, \mathbf{f}_m) \text{ s.t. } \begin{pmatrix} \mathbf{K}_{ii} & \mathbf{K}_{ir} & \mathbf{K}_{im} \\ \mathbf{K}_{ri} & \mathbf{K}_{rr} & \mathbf{K}_{rm} \\ \mathbf{K}_{mi} & \mathbf{K}_{mr} & \mathbf{K}_{mm} \end{pmatrix} \begin{pmatrix} \mathbf{u}_i \\ \hat{\mathbf{u}}_r \\ \mathbf{u}_m \end{pmatrix} = \begin{pmatrix} 0 \\ \hat{\mathbf{f}}_r \\ \mathbf{f}_m \end{pmatrix} \quad (2)$$

In order to simplify our study, we assume that there is no element bearing both redundant and missing degrees of freedom. Note that it is always possible to comply to this assumption by “forgetting” a redundant data (typically the Dirichlet part) which converts a degree of freedom of type  $r$  into a degree of freedom of type  $i$ . This assumption results in  $\mathbf{K}_{rm} = \mathbf{0}$ .

## 2.2 First condensation

As classically done in domain decomposition methods, we can eliminate the internal degrees of freedom and condense the problem on the missing and redundant degrees of freedom:

$$\begin{aligned} \mathbf{u}_i &= -\mathbf{K}_{ii}^{-1}(\mathbf{K}_{ir}\hat{\mathbf{u}}_r + \mathbf{K}_{im}\mathbf{u}_m) & \mathbf{S}_d &= \mathbf{K}_{mm} - \mathbf{K}_{mi}\mathbf{K}_{ii}^{-1}\mathbf{K}_{im} \\ \begin{pmatrix} \mathbf{Z}_d & \mathbf{C}_{rm} \\ \mathbf{C}_{rm}^T & \mathbf{S}_d \end{pmatrix} \begin{pmatrix} \hat{\mathbf{u}}_r \\ \mathbf{u}_m \end{pmatrix} &= \begin{pmatrix} \hat{\mathbf{f}}_r \\ \mathbf{f}_m \end{pmatrix} & \text{with} & \mathbf{Z}_d = \mathbf{K}_{rr} - \mathbf{K}_{ri}\mathbf{K}_{ii}^{-1}\mathbf{K}_{ir} \\ & & & \mathbf{C}_{rm} = -\mathbf{K}_{ri}\mathbf{K}_{ii}^{-1}\mathbf{K}_{im} \end{aligned} \quad (3)$$

$\mathbf{S}_d$  is the Schur complement of the structure of the Missing degrees of freedom considering Dirichlet boundary conditions of the Redundant part.  $\mathbf{Z}_d$  is the Schur complement of the structure on the Redundant part (assuming Dirichlet boundary conditions of the Missing part).  $\mathbf{C}_{rm}$  is the coupling between the Missing and the Redundant part realized through the Interior degrees of freedom; note that except in very pathological cases without practical interest (typically if  $|i| < |r|$  or  $|i| < |m|$ ), it is a full rank matrix.

We recall that the Schur complements inherit several properties from the original matrix [30]. In particular, in the case we consider, they are symmetric definite positive. Moreover, we have the relation:

$$\|\mathbf{u}_m\|_{\mathbf{S}_d} = \|\mathbf{u}\|_{\mathbf{K}} \text{ where } \mathbf{u} = \begin{pmatrix} \mathbf{K}_{ii}^{-1}\mathbf{K}_{im}\mathbf{u}_m \\ \mathbf{0}_r \\ \mathbf{u}_m \end{pmatrix} \quad (4)$$

in words, the norm associated with  $\mathbf{S}_d$  is the energy norm of the field obtained by discrete harmonic lifting (assuming zero Dirichlet boundary condition on Redundant degrees of freedom). Note that for a regular problem (typically a domain with regular shape and good aspect ratio, enough Dirichlet boundary conditions, moderate heterogeneities) with not too fine discretization, the Schur complements can be considered to be well conditioned matrices, in particular compared to the other operators that will be encountered in the rest of the paper.

In the following, we consider various strategies to solve the equation (2) or equivalently (3).

*Remark 1.* An important class of methods relaxes the satisfaction of the measurements, which particularly makes sense in the presence of noisy data. This is typically what is done in the modified error in constitutive relation [23]. In the fading regularization method [10], the identified field is the limit of the sequence indexed by  $j$  (a well suited norm shall be used for each term in the functional to minimize):

$$\begin{aligned} &\text{minimize } \|\mathbf{u}_r^j - \hat{\mathbf{u}}_r\|^2 + \|\mathbf{f}_r^j - \hat{\mathbf{f}}_r\|^2 + \|\mathbf{u}_{m,r}^j - \mathbf{u}_{m,r}^{j-1}\|^2 + \|\mathbf{f}_{m,r}^j - \mathbf{f}_{m,r}^{j-1}\|^2 \\ &\text{under the constraint } \begin{pmatrix} \mathbf{Z}_d & \mathbf{C}_{rm} \\ \mathbf{C}_{rm}^T & \mathbf{S}_d \end{pmatrix} \begin{pmatrix} \mathbf{u}_r^j \\ \mathbf{u}_m^j \end{pmatrix} = \begin{pmatrix} \mathbf{f}_r^j \\ \mathbf{f}_m^j \end{pmatrix} \end{aligned} \quad (5)$$

In the following, we investigate strategies where the measures are not relaxed. Trying to enforce them all is highly unstable and thus we will use an iterative method in order to find a good approximation of the solution to (3) while controlling its regularity.

### Two-field approach

A first expression is worth mentioning. From (3), we directly obtain:

$$\begin{pmatrix} \mathbf{S}_d & -\mathbf{I} \\ \mathbf{C}_{rm} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u}_m \\ \mathbf{f}_m \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{rm}^T \hat{\mathbf{u}}_r \\ \hat{\mathbf{f}}_r - \mathbf{Z}_d \hat{\mathbf{u}}_r \end{pmatrix} \quad (6)$$

This is a rectangular formulation with  $m+r$  rows and  $2m$  columns. The problem is under-determined if  $|r| < |m|$ , that is to say when there is less measurements than data to identify; in that case the solution is not unique. The problem is overdetermined if  $|r| > |m|$ ; in that case the well-posedness condition is  $(\hat{\mathbf{f}}_r - \mathbf{Z}_d \hat{\mathbf{u}}_r) \in \text{range}(\mathbf{C}_{rm})$ , which means that measurements must be coherent for the solution to exist – or it should be sought in a minimal residual sense.

In the following we present two classical ways to obtain square systems.

## 2.3 Steklov-Poincaré's formulations

In this approach, we deduce from (3) two classical forward problems by alternatively “forgetting” one redundant information, which enables us to link  $\mathbf{u}_m$  and  $\mathbf{f}_m$ .

- If we assume Dirichlet condition on  $\partial_r\Omega$ , we have:

$$\mathbf{S}_d \mathbf{u}_m = \mathbf{f}_m - \mathbf{C}_{rm}^T \hat{\mathbf{u}}_r \quad (7)$$

and we note  $\mathbf{b}_d = -\mathbf{C}_{rm}^T \hat{\mathbf{u}}_r$ .

- If we assume Neumann condition on  $\partial_r\Omega$ , we have:

$$\mathbf{u}_m = \begin{pmatrix} \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{Z}_d & \mathbf{C}_{rm} \\ \mathbf{C}_{rm}^T & \mathbf{S}_d \end{pmatrix}^{-1} \begin{pmatrix} \hat{\mathbf{f}}_r \\ \mathbf{f}_m \end{pmatrix} \quad (8)$$

which can be written as, using classical block inverse formula:

$$\mathbf{u}_m = \mathbf{S}_n^{-1} (\mathbf{f}_m + \mathbf{b}_n) \quad \text{with} \quad \begin{aligned} \mathbf{S}_n &= \mathbf{S}_d - \mathbf{C}_{rm}^T \mathbf{Z}_d^{-1} \mathbf{C}_{rm} \\ \mathbf{b}_n &= -\mathbf{C}_{rm}^T \mathbf{Z}_d^{-1} \hat{\mathbf{f}}_r \end{aligned} \quad (9)$$

*Remark 2.*  $\mathbf{S}_n$  is the Schur complement of the structure on the Missing degrees of freedom considering Neumann boundary conditions of the Redundant part. Using the quotient formula we have equivalent expressions for  $\mathbf{S}_n$  and  $\mathbf{b}_n$ :

$$\begin{aligned} \mathbf{S}_n &= \mathbf{K}_{mm} - \begin{pmatrix} \mathbf{K}_{mi} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{K}_{ii} & \mathbf{K}_{ir} \\ \mathbf{K}_{ri} & \mathbf{K}_{rr} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{K}_{im} \\ 0 \end{pmatrix} \\ \mathbf{b}_n &= -\begin{pmatrix} \mathbf{K}_{mi} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{K}_{ii} & \mathbf{K}_{ir} \\ \mathbf{K}_{ri} & \mathbf{K}_{rr} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \hat{\mathbf{f}}_r \end{pmatrix} \end{aligned} \quad (10)$$

### 2.3.1 Primal approach

The primal Steklov-Poincaré formulation simply consists in eliminating the unknown traction  $\mathbf{f}_m$  from equations (7,9), resulting in the following system:

$$(\mathbf{S}_d - \mathbf{S}_n) \mathbf{u}_m = \mathbf{b}_d - \mathbf{b}_n \quad (11)$$

We also directly obtain the following expressions:

$$\begin{aligned} \mathbf{S}_d - \mathbf{S}_n &= \mathbf{C}_{rm}^T \mathbf{Z}_d^{-1} \mathbf{C}_{rm} \\ \mathbf{b}_d - \mathbf{b}_n &= \mathbf{C}_{rm}^T \mathbf{Z}_d^{-1} (\mathbf{Z}_d \hat{\mathbf{u}}_r - \hat{\mathbf{f}}_r) \end{aligned} \quad (12)$$

from which we deduce the following property:

**Proposition 1.**  $(\mathbf{S}_d - \mathbf{S}_n)$  is a symmetric positive matrix of size  $|m|$  and rank  $\min(|m|, |r|)$ .

We can recognize an equivalent minimization problem:

$$\mathbf{u}_m = \arg \min \|\mathbf{C}_{rm} \mathbf{u}_m - (\mathbf{Z}_d \hat{\mathbf{u}}_r - \hat{\mathbf{f}}_r)\|_{\mathbf{Z}_d^{-1}} \quad (13)$$

This equation enables us to understand the hidden choice made when using the primal Steklov-Poincaré approach: we recognize the second line of (6) taken in a least-energy sense. The first line can be used to post-process the nodal reactions.

In order to fully understand the system, one property inherited from the continuous formulation studied in [2] must be mentioned:

**Proposition 2.**  $(\mathbf{S}_d - \mathbf{S}_n)$  is the discrete analogue of a symmetric definite positive compact operator in the adapted displacement trace space of  $\partial_m\Omega$ .

This implies that  $(\mathbf{S}_d - \mathbf{S}_n)$ , even in the case where it is strictly definite, tends to possess very small positive eigenvalues.

From this property and equation (12), we can deduce that Matrix  $\mathbf{C}_{rm}$  possesses very small singular values. Mechanically speaking this matrix represents the nodal reaction on the clamped  $\partial_r\Omega$  resulting from a given displacement field on  $\partial_m\Omega$ ; small singular values correspond to unit norm displacement on  $\partial_m\Omega$  resulting in negligible effects on the faraway boundary  $\partial_r\Omega$ .

*Remark 3.* We expected the expressions in the right hand side of (12) to be less sensitive to cancellation error than their classical form on the left. In practice we never observed such a property, so that we preferred the classical form which makes it possible to use sparse solvers.

### 2.3.2 Dual approach

The Steklov-Poincaré dual formulation simply consists in eliminating the unknown displacement  $\mathbf{u}_m$  from equations (7,9), resulting in the following system:

$$(\mathbf{S}_n^{-1} - \mathbf{S}_d^{-1}) \mathbf{f}_m = \mathbf{S}_d^{-1} \mathbf{b}_d - \mathbf{S}_n^{-1} \mathbf{b}_n \quad (14)$$

We also directly obtain the following expressions using Sherman-Morrison formula on the expression of  $\mathbf{S}_n$  (9):

$$\begin{aligned} \mathbf{S}_n^{-1} - \mathbf{S}_d^{-1} &= \mathbf{S}_d^{-1} \mathbf{C}_{rm}^T \mathbf{Z}_n^{-1} \mathbf{C}_{rm} \mathbf{S}_d^{-1} \\ \mathbf{S}_d^{-1} \mathbf{b}_d - \mathbf{S}_n^{-1} \mathbf{b}_n &= \mathbf{S}_d^{-1} \mathbf{C}_{rm}^T (\hat{\mathbf{u}}_r - \mathbf{Z}_n^{-1} \hat{\mathbf{f}}_r) \end{aligned} \quad (15)$$

where  $\mathbf{Z}_n = \mathbf{Z}_d - \mathbf{C}_{rm} \mathbf{S}_d^{-1} \mathbf{C}_{rm}^T$  is the Schur complement of the structure on the boundary  $\partial_r\Omega$  assuming Neumann conditions on  $\partial_m\Omega$ .

The dual formula satisfy equivalent properties to the primal approach.

**Proposition 3.**  $(\mathbf{S}_n^{-1} - \mathbf{S}_d^{-1})$  is a symmetric positive matrix of size  $|m|$  and rank  $\min(|m|, |r|)$ .

We can also recognize an equivalent minimization problem:

$$\mathbf{f}_m = \arg \min \|\mathbf{C}_{rm} \mathbf{S}_d^{-1} \mathbf{f}_m - (\hat{\mathbf{f}}_r - \mathbf{Z}_n \hat{\mathbf{u}}_r)\|_{\mathbf{Z}_n^{-1}} \quad (16)$$

This equation enables us to see that the dual Steklov-Poincaré can also be derived from (6) by taking the second line in a least-energy sense (with a measure of the energy slightly different from the primal approach) and using the first line in order to replace displacement by efforts.

### 2.3.3 Dual approach in the absence of Dirichlet conditions

In the absence of Dirichlet conditions ( $\partial_d\Omega = \emptyset$ ), the operator  $\mathbf{S}_n$  is not invertible. This is the equivalent situation to floating subdomains in domain decomposition methods. In that case, we let  $\mathbf{S}_n^\dagger$  be a pseudo-inverse of  $\mathbf{S}_n$  and  $\mathbf{R}$  be a basis of the rigid body motions of the whole structure.  $\mathbf{R}_m$  is the restriction of  $\mathbf{R}$  on  $\partial_m\Omega$ ; we have  $\mathbf{S}_n \mathbf{R}_m = \mathbf{0}$ .

Formula (9) expressing the solution to a full Neumann problem must be adapted:

$$\begin{aligned} \mathbf{u}_m &= \mathbf{S}_n^\dagger (\mathbf{f}_m + \mathbf{b}_n) + \mathbf{R}_m \boldsymbol{\alpha} \\ \mathbf{0} &= \mathbf{R}_m^T (\mathbf{f}_m + \mathbf{b}_n) \end{aligned} \quad (17)$$

in the first line, the displacement is defined up to a rigid body motion characterized by the amplitude  $\boldsymbol{\alpha}$ ; the second line corresponds to the well-posedness condition of the Neumann problem.

This leads to the following constrained system:

$$\begin{pmatrix} (\mathbf{S}_n^\dagger - \mathbf{S}_d^{-1}) & \mathbf{R}_m \\ \mathbf{R}_m^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{f}_m \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_d^{-1} \mathbf{b}_d - \mathbf{S}_n^\dagger \mathbf{b}_n \\ -\mathbf{R}_m^T \mathbf{b}_n \end{pmatrix} \quad (18)$$

which is typically solved using a convenient initialization/projection algorithm [12].

### 3 Solving the Steklov-Poincaré systems

In this section, we consider iterative algorithms to solve the Steklov-Poincaré formulations. Indeed using iterative solvers with adapted stopping criteria is a common way not to let small singular values perturb the solution, in particular we will make use of the L-curve [14] in order to stop the iterations. The systems being symmetric definite positive, we use the Conjugate Gradient (CG) algorithm, as was proposed in [5]. We propose and compare various preconditioning strategies and other acceleration techniques and we propose a low cost regularization of the solution based on Ritz eigencomponents. Note that we also tried algorithms based on the minimization of the residual (and harmonic Ritz eigencomponents) without significant difference.

This section briefly recalls the algorithms used, then investigates preconditioning, acceleration and filtering strategies. The methods are compared on a simple example in the last subsection.

#### 3.1 Quick reminder

##### 3.1.1 Preconditioned Conjugate gradient (PCG)

Conjugate Gradient is a classical Krylov solver for symmetric positive definite systems. We use the following notations: we solve  $\mathbf{Ax} = \mathbf{b}$ , where  $\mathbf{A}$  is a symmetric positive definite matrix,  $\mathbf{x}_i$  is the approximation at iteration  $i > 0$ ,  $\mathbf{x}_0$  is the chosen initialization, the residual is  $\mathbf{r}_i = \mathbf{b} - \mathbf{Ax}_i$ . We use a symmetric positive definite preconditioner  $\mathbf{M}$  (in practice, only the inverse of  $\mathbf{M}$  is used), so that we in fact solve the equivalent symmetric system  $\mathbf{L}^{-1}\mathbf{AL}^{-T}(\mathbf{L}^T\mathbf{x}) = \mathbf{L}^{-1}\mathbf{b}$  where  $\mathbf{M} = \mathbf{LL}^T$ .

We note  $\mathbf{z}_i = \mathbf{M}^{-1}\mathbf{r}_i$  the preconditioned residual and we define the Krylov subspace at the iteration  $i > 0$ :

$$\mathcal{K}_i(\mathbf{M}^{-1}\mathbf{A}, \mathbf{z}_0) = \text{span}(\mathbf{z}_0, \mathbf{M}^{-1}\mathbf{Az}_0, \dots, (\mathbf{M}^{-1}\mathbf{A})^{i-1}\mathbf{z}_0) \quad (19)$$

The Conjugate Gradient satisfies the following search principle expressed either in terms of orthogonality or in terms of optimality:

$$\mathbf{x}_i \in \mathbf{x}_0 + \mathcal{K}_i(\mathbf{M}^{-1}\mathbf{A}, \mathbf{z}_0), \quad \mathbf{x}_i = \arg \min \|\mathbf{x} - \mathbf{x}_i\|_{\mathbf{A}} \Leftrightarrow \mathbf{r}_i \perp \mathcal{K}_i(\mathbf{M}^{-1}\mathbf{A}, \mathbf{z}_0) \quad (20)$$

In our case  $\mathbf{A}$  possesses very small eigenvalues resulting in the incapacity to correctly evaluate the error committed in the associated directions.

Algorithm 1 recalls the Conjugate Gradient algorithm. The comments aligned on the right recall how to compute Ritz eigencomponents (see below). Note that the algorithm is written in a form suitable for multiple right-hand sides (aka Block CG), that is to say  $\mathbf{b}$  and  $\mathbf{x}_0$  may have  $N \geq 1$  columns. In this case  $\delta_i$  and  $\gamma_i$  are  $N \times N$  symmetric non-negative matrices,  $\delta_i^{-1}$  is the matrix inverse (or pseudo-inverse for more generality) and  $\gamma_i^{-1/2}$  is the matrix square root. Ample details can be found in the dedicated literature [27]. The formulas for the block computation of Ritz components are adapted from [29].

##### 3.1.2 Discussion on the under-determined case

In the case  $|m| > |r|$ , the matrix  $\mathbf{C}_{rm}$  has a right-kernel and thus  $\mathbf{A}$  is rank-deficient. Let  $\mathbf{R}$  be a basis of the kernel of  $\mathbf{A}$ . Let  $\mathbf{x}$  be a solution to  $\mathbf{Ax} = \mathbf{b}$ , then for all vector  $\alpha$ ,  $(\mathbf{x} + \mathbf{R}\alpha)$  is also a solution.

The conjugate gradient will search the solution in the Krylov subspace. Note that, even if  $\mathbf{A}$  is rank-deficient, we still have  $\mathbf{z}_0 \in \text{Range}(\mathbf{M}^{-1}\mathbf{A})$  and thus  $\mathcal{K}_i(\mathbf{M}^{-1}\mathbf{A}, \mathbf{z}_0) \subset \text{Range}(\mathbf{M}^{-1}\mathbf{A})$ . The conjugate gradient thus converges to the intersection between  $(\mathbf{x} + \mathbf{R}\alpha)$  and  $\text{Range}(\mathbf{M}^{-1}\mathbf{A})$ , which we note  $\bar{\mathbf{x}}$ . Simple computation leads to:

$$\bar{\mathbf{x}} = (\mathbf{I} - \mathbf{R}(\mathbf{R}^T\mathbf{M}\mathbf{R})^{-1}\mathbf{R}^T\mathbf{M})\mathbf{x} \quad (21)$$

$\bar{\mathbf{x}}$  is the projection of  $\mathbf{x}$  on  $\text{Range}(\mathbf{M}^{-1}\mathbf{A})$  parallel to  $\text{Range}\mathbf{R}$ . It can also be characterized as:

$$\bar{\mathbf{x}} = \arg \min_{\mathbf{y} \in \text{Range}(\mathbf{M}^{-1}\mathbf{A})} \|\mathbf{x} - \mathbf{y}\|_{\mathbf{A}} \Rightarrow \bar{\mathbf{x}} = \mathbf{M}^{-1}\mathbf{A}(\mathbf{A}\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A})^\dagger \mathbf{A}\mathbf{M}^{-1}\mathbf{Ax} \quad (22)$$

In the case of synthetic experiments where data are numerically generated from a given  $\mathbf{x}$ . If there are less measurements than missing data, the true error during the iterations must then be evaluated with respect to  $\bar{\mathbf{x}}$  instead of  $\mathbf{x}$ .

In the practical case, once a good approximation of  $\bar{\mathbf{x}}$  was found, one still has the opportunity to improve the solution by adding a term of the form  $\mathbf{R}\alpha$  with  $\alpha$  chosen out of mechanical considerations.

*Remark 4.* The over-determined case is naturally dealt with by the Steklov-Poincaré formulation (because of the equivalent minimization problem). It has no interaction with the solver.

### 3.1.3 Ritz values and vectors

We assume the solver converged in  $m$  iterations. We note with capital letters the concatenation of vectors stored during the resolution: e.g.  $\mathbf{P}_m = (\mathbf{p}_0, \dots, \mathbf{p}_{m-1})$ .

Ritz vectors  $\mathbf{V}_m$  form a specific basis of the Krylov subspace, which can be computed at no extra cost, and which satisfies the following relations:

$$\mathbf{V}_m^T \mathbf{A} \mathbf{V}_m = \text{diag}(\theta_i) \quad , \quad \mathbf{V}_m^T \mathbf{M} \mathbf{V}_m = \mathbf{I} \quad (23)$$

$(\theta_i)$  are the Ritz values, they are approximations of the generalized eigenvalues of the pair formed by the operator  $\mathbf{A}$  and the preconditioner  $\mathbf{M}$ . We assume they are stored in decreasing order. The following remark briefly recall how they can be processed at nearly no cost.

*Remark 5* (Computation of Ritz elements). If during the CG iterations, we compute the normalized vector  $\hat{\mathbf{z}}_i = (-1)^i \mathbf{z}_i / \sqrt{\mathbf{z}_i^T \mathbf{r}_i}$ , then  $\hat{\mathbf{Z}}_i$  is a basis of  $\mathcal{K}_{i+1}$  which satisfies:

$$\hat{\mathbf{Z}}_i^T \mathbf{M} \hat{\mathbf{Z}}_i = \mathbf{I} \quad \text{and} \quad \hat{\mathbf{Z}}_i^T \mathbf{A} \hat{\mathbf{Z}}_i =: \mathbf{T}$$

where  $\mathbf{T}$  is a tridiagonal symmetric matrix whose eigenvalues are the Ritz values. The coefficients of  $\mathbf{T}$  can be recovered from the coefficients computations of the CG [29], even in the case of a block solver as given in Algorithm 1. If we note  $\mathbf{U}$  its orthonormal eigenvectors, then  $\mathbf{V}_i = \hat{\mathbf{Z}}_i \mathbf{U}$  are the Ritz vectors. We note  $\mathbf{u}$  the first row of Matrix  $\mathbf{U}$ , which corresponds to the decomposition of  $\mathbf{z}_0$  on the basis of Ritz vectors :  $\mathbf{z}_0 = \|\mathbf{r}_0\|_{\mathbf{M}^{-1}} \mathbf{V} \mathbf{u}^T = \|\mathbf{r}_0\|_{\mathbf{M}^{-1}} \sum_i^m \mathbf{v}_i u_i$  or equivalently  $u_i = \mathbf{v}_i^T \mathbf{r}_0 / \|\mathbf{r}_0\|_{\mathbf{M}^{-1}}$ .

---

#### Algorithm 1 Block Conjugate Gradient with Ritz computation

---

```

 $\mathbf{x}_0$  given
 $\mathbf{r}_0 = \mathbf{b} - \mathbf{A} \mathbf{x}_0$ 
 $\mathbf{z}_0 = \mathbf{M}^{-1} \mathbf{r}_0$ ,  $\mathbf{p}_0 = \mathbf{z}_0$   $\phi_{-1} = 0$ 
for  $i = 0, 1, \dots, m$  (convergence) do
   $\mathbf{q}_i = \mathbf{A} \mathbf{p}_i$ 
   $\delta_i = (\mathbf{p}_i^T \mathbf{q}_i)$ ,  $\gamma_i = (\mathbf{z}_i^T \mathbf{r}_i)$ ,  $\alpha_i = \delta_i^{-1} \gamma_i$   $\hat{\mathbf{z}}_i = (-1)^i \mathbf{z}_i \delta_i^{-1/2}$ 
   $\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{p}_i \alpha_i$ 
   $\mathbf{r}_{i+1} = \mathbf{r}_i - \mathbf{q}_i \alpha_i$ 
   $\mathbf{z}_{i+1} = \mathbf{M}^{-1} \mathbf{r}_{i+1}$ 
   $\phi_i = (\mathbf{q}_i^T \mathbf{z}_{i+1})$ ,  $\beta_i = \delta_i^{-1} \phi_i$   $\mathbf{T}_{i+1,i+1} = \gamma_i^{-1/2} (\delta_i + \phi_{i-1}^T \phi_{i-1}) \gamma_i^{-1/2}$ 
   $\mathbf{p}_{i+1} = \mathbf{z}_{i+1} - \mathbf{p}_i \beta_i$   $(i > 0) \mathbf{T}_{i+1,i} = \mathbf{T}_{i,i+1} = \gamma_{i-1}^{-1/2} \delta_{i-1} \gamma_{i-1}^{-1} \gamma_i^{1/2}$ 
end for
 $(\theta_i)_{i < m} =$  Eigenvalues of  $\mathbf{T}$   $\mathbf{T}$  is symmetric bloc tridiagonal
 $\mathbf{U} =$  Eigenvectors of  $\mathbf{T}$ ,
Return: Solution  $\mathbf{x}_{m+1}$ , Ritz values  $(\theta_i)$ , Ritz vectors  $\mathbf{V}_m = \hat{\mathbf{Z}}_m \mathbf{U}$ 

```

---

### 3.1.4 Stopping criterion based on the L-curve

Thanks to its embedded minimization principle, PCG is a robust solver (even if some extra care may be required – like full reorthogonalization – to resist numerical errors). In our case, the quantity  $\|\mathbf{x} - \mathbf{x}_i\|_{\mathbf{A}}$  is decreased at each iteration but because of the small eigenvalues of  $\mathbf{A}$  this is often paid by an explosion of  $\|\mathbf{x}_i\|_2$  and in fact no convergence of  $\|\mathbf{x} - \mathbf{x}_i\|_2$ . The L-curve, explained for example in [14] aims at stopping iterations before the solution explodes by finding the bottom-left “corner” of a curve of the type  $(\log(\|\mathbf{r}_i\|), \log(\|\mathbf{x}_i\|))$  where  $\|\mathbf{r}_i\|$  shall be a computable measure of the error and  $\|\mathbf{x}_i\|$  a norm of the solution. The corner realizes a compromise, but its detection (or even its existence) is not evident.

## 3.2 Illustration of the raw methods

We consider the test-case presented on Figure 2a. The structure is a rectangle of dimensions  $5\text{mm} \times 10\text{mm}$ . We consider plane stress and linear elastic isotropic behavior. The Poisson ratio is 0.3; in the homogeneous case, the Young modulus is  $E = 70000$  MPa; in the heterogeneous case, each element is granted a random Young modulus accordingly to an uniform probability law in  $[7000; 70000]$  MPa. The reference solution is built with

Neumann conditions of the form:  $f_{m_x} = f_0(1 + y/y_0)$ ,  $\hat{f}_{r_x} = f_0(1 - y/y_0)$  and  $\hat{f}_{r_y} = f_{m_y} = 0$ . With  $f_0 = 1$  N.mm<sup>-1</sup> and  $y_0 = 1$  mm. We use a regular mesh with continuous piecewise linear triangle finite element, there is a total of 288 degrees of freedom, with  $|m| = 38$ ,  $|r| = 40$ . This choice implies that the problem possesses only positive eigenvalue but they may be very small.

The solution to the forward problem serves as a reference for the inverse computations. When noise is injected, it takes the form of a Gaussian white noise:  $\hat{\mathbf{u}}_r^n = (\mathbf{I} + N_l \mathbf{G}) \hat{\mathbf{u}}_r$ , with  $\mathbf{u}_r^n$  the noisy data,  $N_l$  the level of noise,  $\mathbf{I}$  the identity matrix and  $\mathbf{G}$  a diagonal matrix of uncorrelated Gaussian random variables with mean 0 and covariance 1.

The reference for the homogeneous case is represented on Figure 2b; the evolution of the x-component of the displacement on the right side  $u_{m_x}^{ref}$  is plotted on Figure 2c.

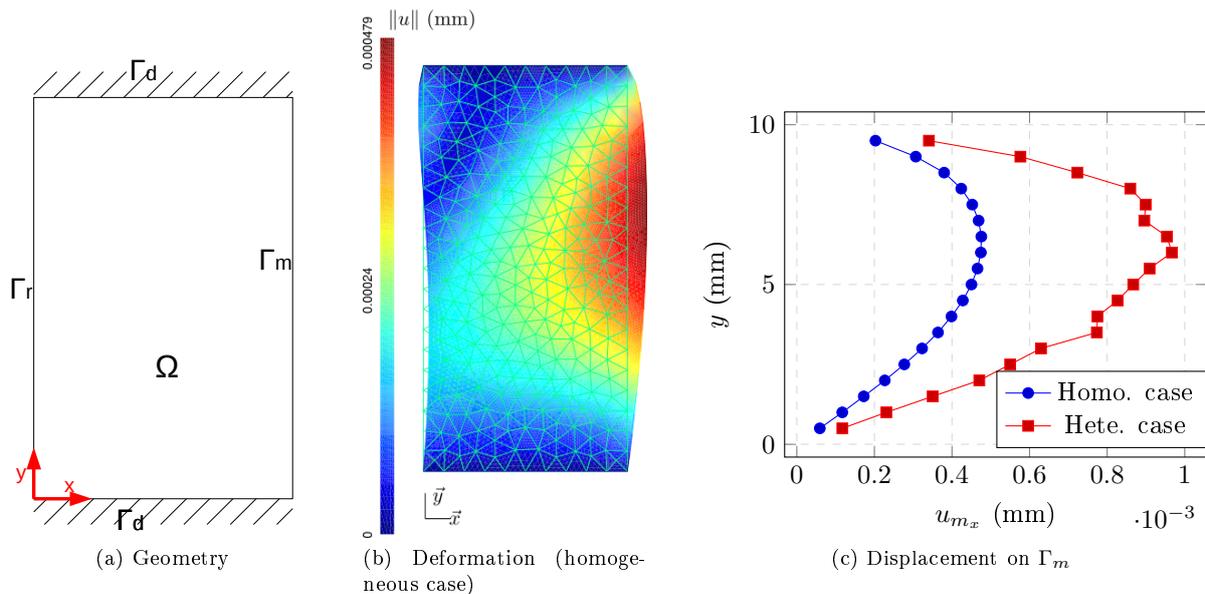


Figure 2: Reference problem and solution

In order to evaluate the quality of the solution  $\mathbf{u}$ , we use the following error  $e_m = (u_m^{ref} - u_m) / \max |u_m^{ref}|$ . In order to illustrate the methods, we choose to plot L-curves with realistic measurements: abscissa is the Euclidean norm of the residual  $\|\mathbf{r}\|_2$ , ordinate is the Euclidean norm of the unknown vector  $\|\mathbf{x}_m\|_2$  (displacement in primal nodal traction in dual). Both quantities are normalized by their initial value.

Figures 3a and 3b, 4a, 4b compare L-curves and distribution of error (at the corner iteration) for the basic primal and dual approaches in either a noiseless heterogeneous case or a homogeneous situation with 10% noise. It is worthy to notice that in the noiseless case (figure 3b), both approach have well identified corners in the L-curves and the dual method gives way better results than the primal method. This agrees with [19]. On the other hand, as the noise level increases (figure 4b), the corners are much harder to detect and the superiority of the dual method seems to vanish (as will be visible in table 1 in that case the primal method generates a very small error even if the residual is rather large).

### 3.3 Preconditioners

Because of the compactness of the continuous system, the problem we solve is intrinsically not well posed, with a lower part of the spectrum populated by many small eigenvalues. The interest of preconditioning is not as clear as for well-posed problems. For instance in domain decomposition methods applied to forward (well posed) problems, preconditioning enables to mitigate the effects of the discretization, of the decomposition and of the bad properties of the system (large heterogeneities), leading to reasonable condition numbers on large class of problems.

In the case of inverse problems, preconditioning can not cure the lower part of the spectrum which drives the convergence. Thus solvers can be used “raw” (non-preconditioned). One may find, in [26] and the associated references, the illustration of raw solvers successfully applied to compact systems. In the literature, Steklov-Poincaré methods were also most often tried without preconditioner, with certain success [5, 19, 3].

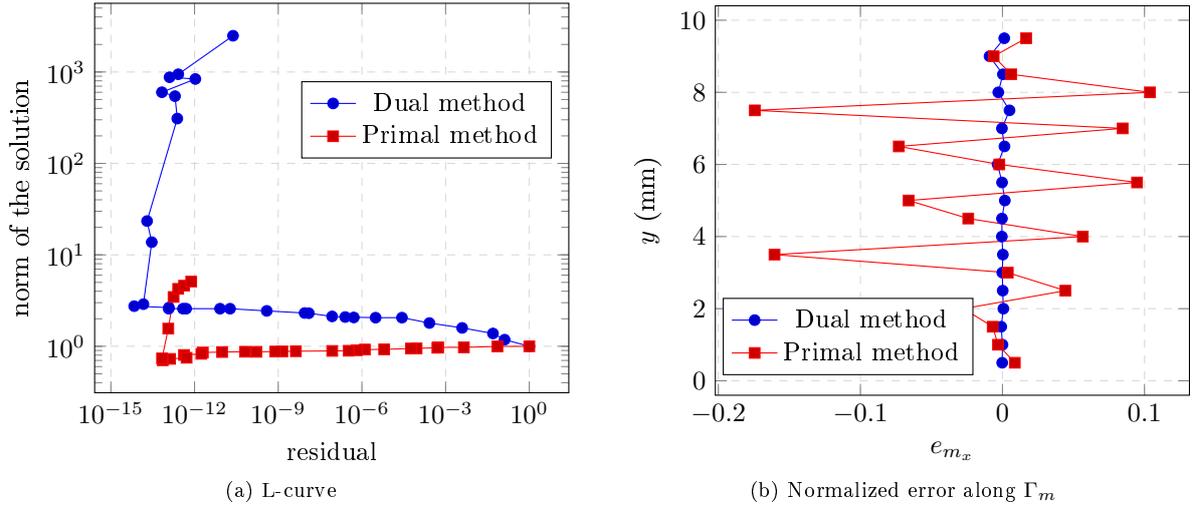


Figure 3: Primal and dual SP methods in the heterogeneous case without noise

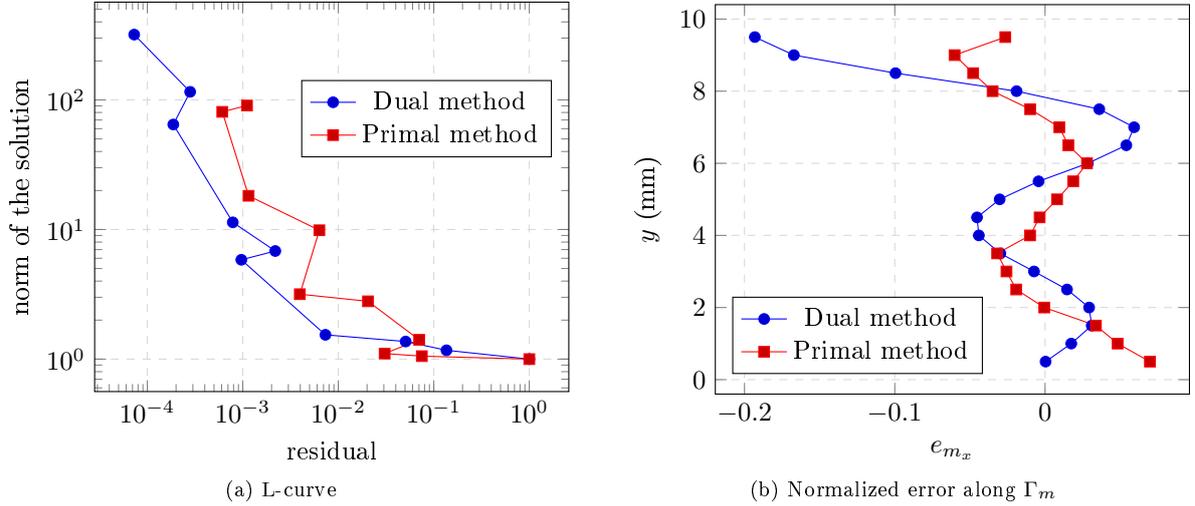


Figure 4: Primal and dual SP methods in the homogeneous case with 10% noise

In the following, we investigate preconditioning strategies. We focus on the primal approach even though one could easily design the dual counterparts. Yet it appeared from our experiments that for these problems primal and dual approaches strongly differ and in fact, for now, we have no good preconditioning strategies for the dual approach.

Table 1 gives a synthetic view of the performance of the methods we discuss in the following (for the noisy homogeneous and noiseless heterogeneous cases respectively). For the different methods, we compare the index of the “corner iteration”, and for that iteration, the true error in Euclidean norm, the normalized residual (in Euclidean norm) and the Euclidean norm of the displacement.

### 3.3.1 KMF preconditioner

The KMF preconditioner consists in preconditioning the primal formulation by  $\mathbf{S}_d^{-1}$ . The resulting system clearly has its spectrum inside the interval  $]0, 1[$  which makes it prone to stationary iterations. Stationary iterations are slow to converge but they have the advantage to be more stable than Krylov iterations in particular in the absence of complex eigenvalues [17].

The preconditioner being symmetric definite positive, its use can be interpreted as changing the orthogonality properties in the Krylov solver. Moreover, as said in (4), the norm associated with  $\mathbf{S}_d$  is equivalent to the energy norm. These properties make the KMF preconditioner a regularization in itself since it tends to promote most

Method	Corner iteration	$\ \mathbf{e}_m\ _2$	$\ \mathbf{r}_m\ _2/\ \mathbf{r}_{m,0}\ _2$	$\ \mathbf{u}_m\ _2$
Noisy homogeneous case				
Dual	4	$1.91 \cdot 10^{-4}$	$1.09 \cdot 10^{-4}$	0.00162
Primal	3	$8.88 \cdot 10^{-5}$	$1.20 \cdot 10^{-3}$	0.00160
Primal KMF	4	$1.69 \cdot 10^{-4}$	$1.10 \cdot 10^{-3}$	0.00161
Primal Sym.	4	$2.10 \cdot 10^{-4}$	$4.97 \cdot 10^{-4}$	0.00161
Primal MultiPrec	3	$4.98 \cdot 10^{-4}$	$1.24 \cdot 10^{-4}$	0.00166
Noiseless heterogeneous case				
Dual	20	$3.01 \cdot 10^{-5}$	$2.92 \cdot 10^{-16}$	0.00317
Primal	24	$4.52 \cdot 10^{-4}$	$4.22 \cdot 10^{-15}$	0.00314
Primal KMF	22	$6.07 \cdot 10^{-5}$	$3.17 \cdot 10^{-15}$	0.00317
Primal Sym.	20	$3.81 \cdot 10^{-5}$	$8.33 \cdot 10^{-16}$	0.00318
Primal MultiPrec	no corner	$9.40 \cdot 10^{-5}$	$7.28 \cdot 10^{-7}$	0.00317

Table 1: Synthetic view of the methods performance.

energetic modes. This role is particularly important in the cases where the Schur complement significantly differs from identity. This is typically the case of heterogeneous structures where energy often is not evenly distributed where the true error with KMF preconditioner is 10 times smaller than for the raw primal approach.

### 3.3.2 Symmetric “à la dual” preconditioner

Since in general the dual method has a better behavior compared to the primal approach, we propose to build a version of the primal approach as close to the dual approach as possible.

By the simple change of variable  $\tilde{\mathbf{f}}_m = \mathbf{S}_d \mathbf{u}_m$ , and pre-multiplying (12) by  $\mathbf{S}_d^{-1}$ , we obtain:

$$\mathbf{S}_d^{-1} \mathbf{C}_{rm}^T \mathbf{Z}_n^{-1} \mathbf{C}_{rm} \mathbf{S}_d^{-1} \tilde{\mathbf{f}}_m = \mathbf{S}_d^{-1} \mathbf{C}_{rm}^T (\hat{\mathbf{u}}_r - \mathbf{Z}_n^{-1} \hat{\mathbf{f}}_r) \quad (24)$$

compared to the dual formulation (15),  $\mathbf{Z}_d$  was replaced by  $\mathbf{Z}_n$ .

Since  $\mathbf{Z}_d \geq \mathbf{Z}_n$  (for the ordering of SPD matrices), and because the bad conditioning of the system is mostly caused by the smallest eigenvalues, we can expect the dual approach to have a slightly better numerical behavior. However, using Neumann formula, we can check that the two approaches are first order equivalent:

$$\begin{aligned} \mathbf{S}_n^{-1} - \mathbf{S}_d^{-1} &= (\mathbf{S}_d - (\mathbf{S}_d - \mathbf{S}_n))^{-1} - \mathbf{S}_d^{-1} = \mathbf{S}_d^{-1} (\mathbf{I} - (\mathbf{S}_d - \mathbf{S}_n) \mathbf{S}_d^{-1})^{-1} - \mathbf{S}_d^{-1} \\ &= \mathbf{S}_d^{-1} (\mathbf{I} + (\mathbf{S}_d - \mathbf{S}_n) \mathbf{S}_d^{-1} + o(\mathbf{S}_d - \mathbf{S}_n)) - \mathbf{S}_d^{-1} = \mathbf{S}_d^{-1} (\mathbf{S}_d - \mathbf{S}_n) \mathbf{S}_d^{-1} + o(\mathbf{S}_d - \mathbf{S}_n) \end{aligned} \quad (25)$$

As expected, the experiments show that this method is almost equivalent to the dual approach.

### 3.3.3 Multipreconditioning

In [28], one can find a physical interpretation of the classical preconditioners for FETI and BDD domain decomposition method. This interpretation leads us to considering a preconditioner made out of weighted local inverses of the operators:

$$(\mathbf{S}^d - \mathbf{S}^n)^{-1} \simeq \frac{1}{2} (\mathbf{S}_n^{-1} - \mathbf{S}_d^{-1}) \frac{1}{2} \quad (26)$$

The factor 1/2 is the counterpart of the scaling operator of FETI and BDD[20] in the case of Schur complements computed on the same domain. In the following, we show evidence that this preconditioner is not far from optimal but in a sense which is not adapted to the problem we wish to solve.

Our investigation is based on the use of MultiPreconditioned Conjugate Gradient (MPCG) [7, 13, 6]. Starting from a family of potential preconditioners, in our case  $(\mathbf{S}_d^{-1}, \mathbf{S}_n^{-1})$ , multipreconditioned Krylov solvers let the solver find the optimal linear combination at each iteration for the underlying minimization problem (20). Schematically, at the iteration  $i$ , the preconditioner takes the form  $(\alpha_i^n \mathbf{S}_n^{-1} + \alpha_i^d \mathbf{S}_d^{-1})$  where  $\alpha_i^n$  and  $\alpha_i^d$  are computed thanks to optimality conditions.

*Remark 6* (MPCG algorithm.). In order to apply multipreconditioning to Conjugate Gradient, one simply needs to consider  $\mathbf{z}, \mathbf{p}, \mathbf{q}$  as two-column matrices and to replace the preconditioning step by  $\mathbf{z} = [\mathbf{S}_n^{-1} \mathbf{r}, \mathbf{S}_d^{-1} \mathbf{r}]$  (concatenation of two vectors); then  $\mathbf{p}^T \mathbf{q}$  is a  $2 \times 2$  matrix and  $\alpha$  is a  $2 \times 1$  vector. Note that full reorthogonalization must be used because multipreconditioning breaks the short recurrence of CG.

Figure 5 compares the iterations of CG and MPCG in terms of error (measured by  $\|\mathbf{r}\|_2$ ) and norm of the error (measured by  $\|\mathbf{e}\|_2$ ) for the heterogeneous noiseless case. We observe that the initial convergence is faster with the multipreconditioner, but a breakdown quickly occurs caused by the bad properties of the matrix  $\mathbf{p}^T \mathbf{q} = \mathbf{p}^T \mathbf{A} \mathbf{p}$ .

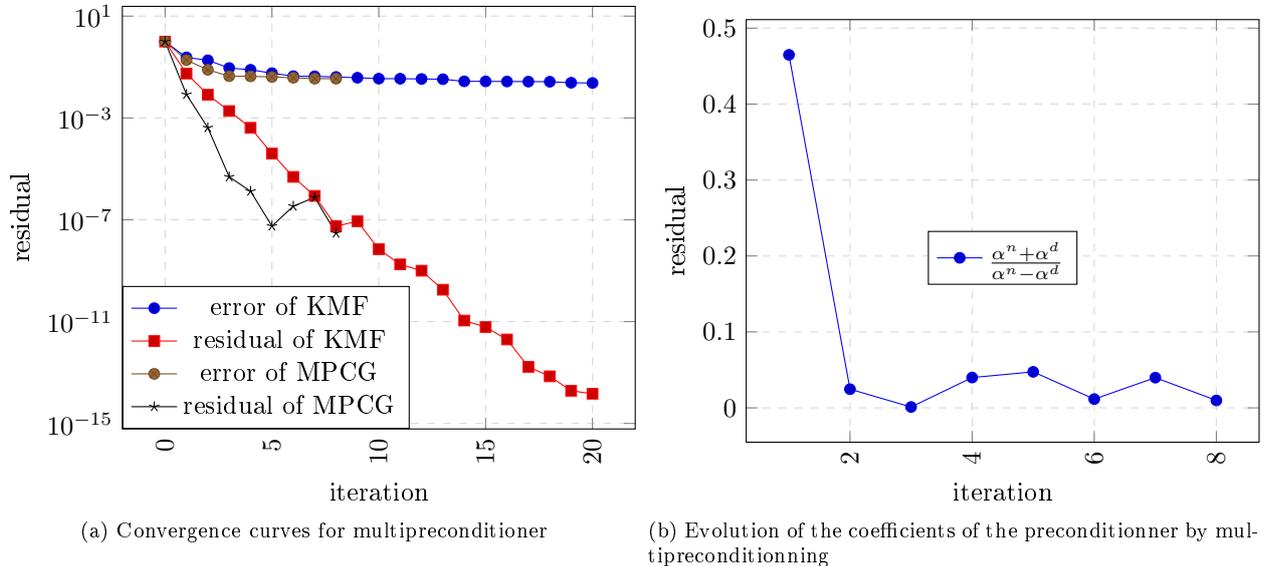


Figure 5: Performance of multipreconditioned conjugate gradient in the heterogeneous noiseless case.

After little variations in the first iterations, it seems that the intuitive combination, corresponding to (26),  $\alpha^n = -\alpha^d$  is quasi-optimal with respect to the solver minimization property. This can be viewed on figure 5b. Anyhow selecting  $\alpha^n = -\alpha^d$  means preconditioning a compact operator by another compact operator which is clearly a bad idea from the stability point of view and leads to early breakdowns. Moreover, the post-filtering which will be presented in Section 3.5 is not as straightforward as for the classical KMF preconditioner.

### 3.4 Acceleration by multiple right-hand sides

Block solvers form a simple and efficient family of methods to accelerate the iterations [27]: instead of solving for a vector right hand side, we artificially add other columns. In our case where the right-hand side is the difference of two vectors, we can solve the following system:

$$\begin{aligned} \text{in primal:} \quad & (\mathbf{S}_d - \mathbf{S}_n) [\mathbf{u}_m, \tilde{\mathbf{u}}_m] = [\mathbf{b}_d - \mathbf{b}_n, \mathbf{b}_d + \mathbf{b}_n] \\ \text{in dual:} \quad & (\mathbf{S}_n^{-1} - \mathbf{S}_d^{-1}) [\mathbf{f}_m, \tilde{\mathbf{f}}_m] = [\mathbf{S}_d^{-1} \mathbf{b}_d - \mathbf{S}_n^{-1} \mathbf{b}_n, \mathbf{S}_d^{-1} \mathbf{b}_d + \mathbf{S}_n^{-1} \mathbf{b}_n] \end{aligned} \quad (27)$$

It is also possible, in particular in case of one of the two vectors  $[\mathbf{b}_d, \mathbf{b}_n]$  being null, to use random right-hand side.

Block Krylov solvers involve block-operations whose implementation is highly optimized. They generate larger Krylov subspaces. In our case, they usually converge in two times less iterations than regular CG. Moreover, we can combine Block CG with any preconditioner, short recurrence still applies, and even Ritz post-analysis is possible (see the next subsection).

### 3.5 A posteriori Ritz filtering

The solution obtained by conjugate gradient can be improved thanks to Ritz vectors post-analysis. We assume the solver converged in  $m$  iterations.

Once the Ritz elements were determined, a truncation can be applied in order to keep only the  $m' \leq m$  largest eigenvalues, the projected solution (written with subscript  $R$ ) can be computed:

$$\mathbf{x}_{R,m'} = \mathbf{x}_0 + \|\mathbf{r}_0\|_{\mathbf{M}^{-1}} \sum_{i=0}^{m'} \frac{u_i}{\theta_i} \mathbf{v}_i \quad \text{where} \quad u_i = \frac{\mathbf{v}_i^T \mathbf{r}_0}{\|\mathbf{r}_0\|_{\mathbf{M}^{-1}}} \quad (28)$$

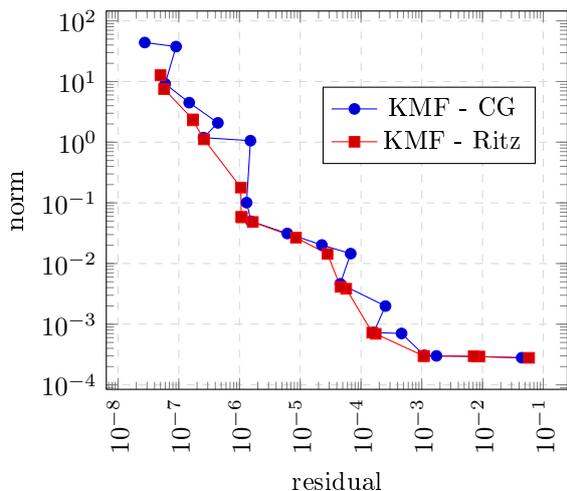
We have:

$$\begin{aligned}\|\mathbf{x} - \mathbf{x}_{R,i}\|_{\mathbf{A}}^2 &= \|\mathbf{x} - \mathbf{x}_0\|_{\mathbf{A}}^2 - \|\mathbf{r}_0\|_{\mathbf{M}^{-1}}^2 \sum_i^{m'} \frac{u_i^2}{\theta_i} \\ \|\mathbf{x}_{R,i} - \mathbf{x}_0\|_{\mathbf{M}}^2 &= \|\mathbf{r}_0\|_{\mathbf{M}^{-1}}^2 \sum_i^{m'} \frac{u_i^2}{\theta_i^2}\end{aligned}\tag{29}$$

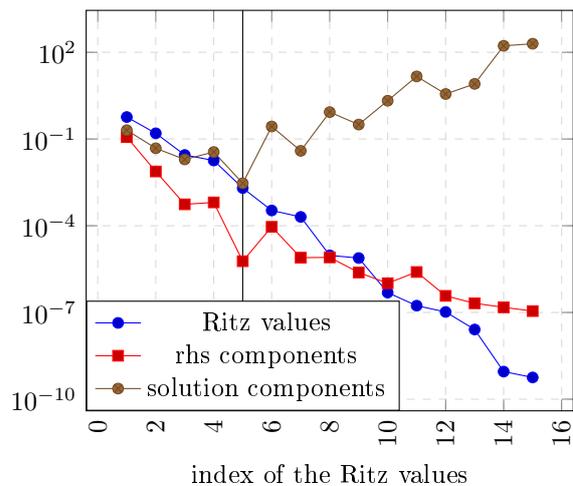
we see that the error (in the  $\mathbf{A}$  norm) is decreasing with  $m'$  whereas the norm of the solution (in the  $\mathbf{M}$  norm) is increasing. As soon as  $\theta_i < 1$ , the growth of the norm is faster than the decrease of the error. These properties ensure the possibility to define a ‘‘corner’’ without ambiguity once a precision/norm compromise was chosen on the L-curve indexed by the number of Ritz components (in the  $(\|\mathbf{e}_i\|_{\mathbf{A}}, \|\mathbf{x}_i - \mathbf{x}_0\|_{\mathbf{M}})$  frame). Unfortunately the monotonicity properties are not insured in a more practical frame, like  $(\|\mathbf{r}\|_2, \|\mathbf{x}\|_2)$ , where the Ritz L-curve could be compared to the classical CG L-curve, as on Figure 6a. Anyhow, we observe that the Ritz-L-curve is below and on the left side of the classical CG-L-curve, leading to a better accuracy-norm compromise.

The explosion of the norm is controlled as long as the  $(u_i)$  coefficients decrease faster than the eigenvalues, which is the classical Picard condition and which offers a quantitative alternative to the L-curve in order to stop the iterations.

On Figure 6b, an analogous to the Picard plot [16] is done: the Ritz values are plotted on the same graph as the coefficients of the projection of the right hand side on the Ritz basis, and the projection of the solution coefficient on the Ritz basis. Similarly to [24], one can propose as an alternative to the L-curve, to stop the reconstruction when solution components start to increase. Indeed, the increase of the components can be attributed to the effects of the noise on the data. On the presented graph, for example, the discrete Picard condition suggest to keep only the 5 first terms of the solution in the Ritz basis.



(a) L-curves for the KMF approach, obtained directly from CG or after Ritz filtering



(b) Picard plot: decomposition of the right-hand side and of the solution in the Ritz' basis

Figure 6: Filtering of KMF iterations by Ritz analysis, 10% noise homogeneous case.

Note that the Ritz post-filtering has of course strong similarities with the truncated SVD [15]. More precisely since  $\mathbf{A}$  is symmetric positive and  $\mathbf{M}$  is symmetric positive definite, the Ritz values approximate the generalized singular values of  $\mathbf{A}$  for the norm associated to  $\mathbf{M}$ .

Physically speaking, the Ritz filtering makes particular sense when used with the KMF preconditioner since in that case the Ritz modes are unit-energy modes on which the solution is decomposed.

## 4 Application on a 3D test-case

In this section, we propose to investigate a 3D test case by the non-preconditioned dual Steklov-Poincaré method and the KMF-primal approach. The results are filtered by the a posteriori Ritz analysis.

The geometry of the direct problem is described on the figure 7a: a cuboid is clamped on its four lateral faces. It is subjected to Neumann boundary conditions on the upper and on the lower faces. On the upper face,  $\Gamma_m$ , the support of the Neumann conditions are two ellipses; the lower face  $\Gamma_r$  is not loaded. The goal of the inverse problem is to recover the displacement on the upper face  $\Gamma_m$  from redundant data on the lower boundary  $\Gamma_r$ . In total, there are 1669 degrees of freedom of which 447 are with missing data and 444 are with redundant data.

We use the following heuristic as a stopping criterion: we consider the curve of the “solution components” in the Ritz basis, we filter out the small scale undulation by removing the local minima, then we use a polynomial interpolation and stop the selection when the interpolation curve is minimal (that is to say before it begins to increase). The reference solution and the results of the identification are illustrated on the figures 7 and 9.

First, we solve this problem with noiseless data on the redundant boundary. The Picard graph 8a suggests to project the problem on all the computed modes as the coefficients of the solution decrease on average.

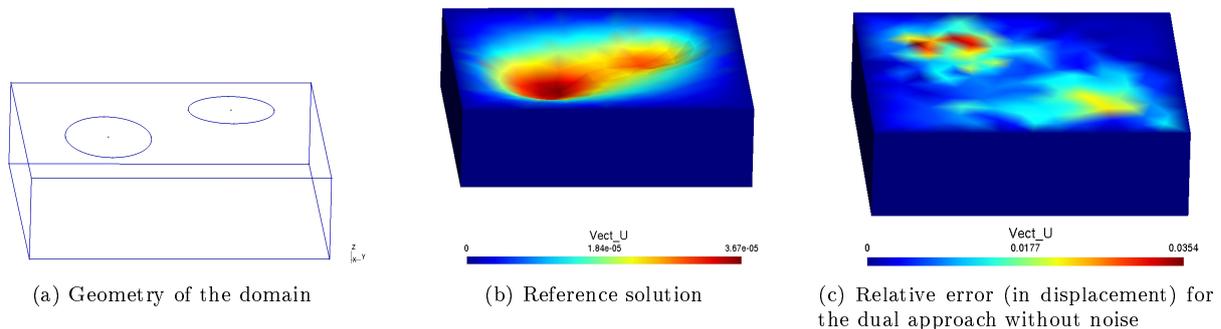


Figure 7: Illustration of the 3D problem

In the case when multiplicative Gaussian white noise is applied, the Picard plots of the figures 8b and 8c suggest to stop the iterations earlier as the amplitude of the noise increases.

On the table 2, the optimal number of modes and the error resulting from the process are presented. As the system is underdetermined, the error is measured with respect to the suitable  $\bar{\mathbf{u}}$ , projection of the reference field. We observe that the primal-KMF approach tends to be stopped after less iterations than the dual approach, but the approximation obtained is not as precise.

Method	KMF	KMF	KMF	Dual	Dual	Dual
Noise level	0%	1%	10%	0%	1%	10%
Nb of Ritz modes	50	30	15	50	35	20
error on $\mathbf{u}_m$	0.015586	0.088206	0.13869	0.0097854	0.066349	0.11712

Table 2: Error and number of Ritz modes for the primal-KMF and the dual methods

## 5 Conclusion

In this work, the primal and dual Steklov-Poincaré methods for solving the Cauchy problem have been investigated. First, we have studied preconditioning strategy for the primal approach: the KMF preconditioner which tends to promote most energetic modes, the symmetric preconditioner which results in a system very similar to the dual approach, and the one inspired by domain decomposition which leads to fast decrease of the error but which is highly unstable. Second we propose to use a block solver to accelerate the iterations. Last we showed that a post-filtering based on the Ritz elements was easy to set up and performed well since it lead to better L-curves and made it possible to derive a stopping criterion based on the Picard condition.

The block KMF-primal and dual algorithms were tested on a 3D identification case in order to illustrate their behavior on a larger problem.

The non-necessity of preconditioning the dual approach and its better behavior than the primal approach are not fully elucidated yet. These issues will be the subject of future work as well as the estimation of the influence of the finite element discretization on the quality of the identification.

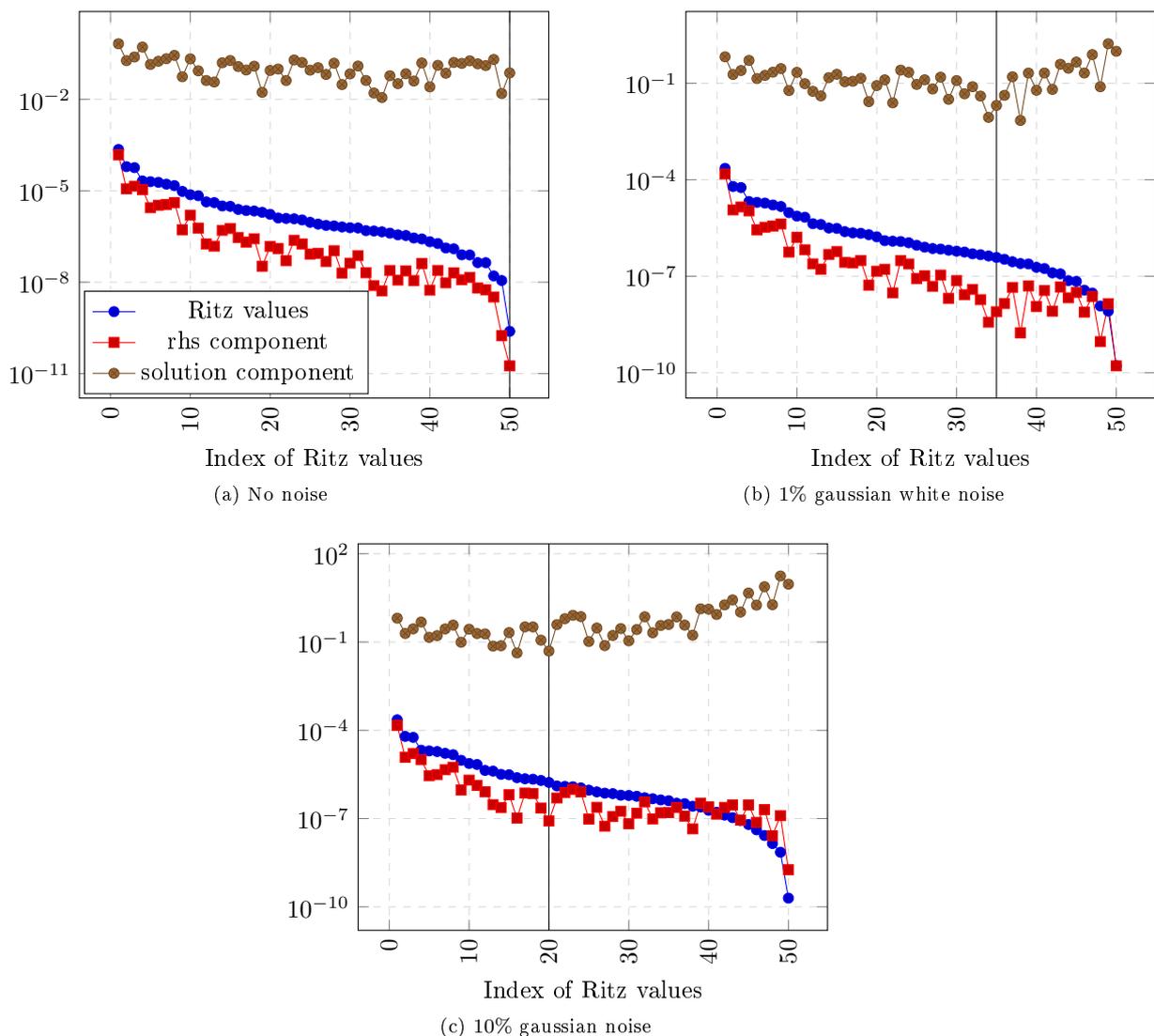


Figure 8: Picard plots for the 3D problem (with dual block solver) – the vertical line highlights the last selected component.

## References

- [1] Stéphane Andrieux, TN Baranger, and A Ben Abda. Solving Cauchy problems by minimizing an energy-like functional. *Inverse problems*, 22(1):115, 2006.
- [2] M Azaiez, A Ben Abda, and Jalel Ben Abdallah. Revisiting the Dirichlet-to-Neumann solver for data completion and application to some inverse problems. *Int. J. Appl. Math. Mech*, 1:106–121, 2005.
- [3] Mejdj Azaiez, Faker Ben Belgacem, and Henda El Fekih. On Cauchy’s problem: II. A variational Steklov–Poincaré theory. *Inverse Problems*, 22(6):1307–1336, 2006.
- [4] Faker Ben Belgacem. Why is the Cauchy problem severely ill-posed? *Inverse Problems*, 23(2):823–836, 2007.
- [5] Jalel Ben Abdallah. A conjugate gradient type method for the Steklov-Poincaré formulation of the cauchy-poisson problem. *Int. J. of Appl. Math. and Mech*, 9(1):27–40, 2007.
- [6] Christophe Bovet, Pierre Gosselet, and Nicole Spillane. Multipreconditioning for nonsymmetric problems: the case of orthomin and biCG. *Comptes rendus de l’académie des sciences (math.)*, 355(3):354–358, 2017.

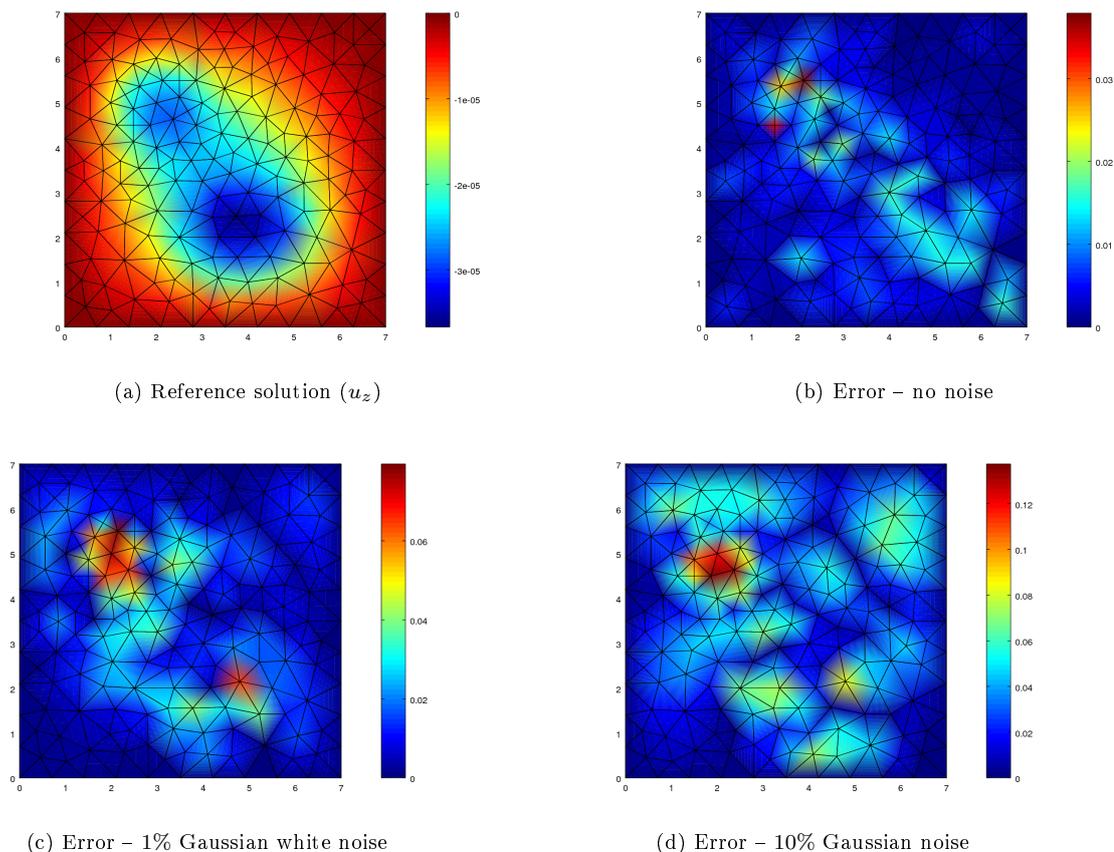


Figure 9: Solution and map of relative error ( $e_{m_z}$ ) – dual block approach – for various level of noise

- [7] Robert Bridson and Chen Greif. A multipreconditioned conjugate gradient algorithm. *SIAM Journal on Matrix Analysis and Applications*, 27(4):1056–1068, 2006.
- [8] JT Chen and KH Chen. Analytical study and numerical experiments for Laplace equation with overspecified boundary conditions. *Applied Mathematical Modelling*, 22(9):703–725, 1998.
- [9] A Cimetiere, F Delvare, M Jaoua, and F Pons. Solution of the Cauchy problem using iterated Tikhonov regularization. *Inverse Problems*, 17(3):553–570, 2001.
- [10] Alain Cimetiere, Franck Delvare, and Frédéric Pons. Une méthode inverse à régularisation évanescente. *Comptes Rendus de l’Académie des Sciences - Series IIB - Mechanics*, 328(9):639 – 644, 2000.
- [11] Charbel Farhat and François-Xavier Roux. A method of finite element tearing and interconnecting and its parallel solution algorithm. *International Journal for Numerical Methods in Engineering*, 32(6):1205–1227, 1991.
- [12] Charbel Farhat and François-Xavier Roux. Implicit parallel processing in structural mechanics. *Computational Mechanics Advances*, 2(1):1–124, 1994. North-Holland.
- [13] Pierre Gosselet, Daniel Rixen, François-Xavier Roux, and Nicole Spillane. Simultaneous FETI and block FETI: Robust domain decomposition with multiple search directions. *International Journal for Numerical Methods in Engineering*, 104(10):905–927, 2015.
- [14] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.*, 34:561–580, 1992.
- [15] Per Christian Hansen. The truncatedsvd as a method for regularization. *BIT Numerical Mathematics*, 27(4):534–553, 1987.

- [16] Per Christian Hansen. The discrete Picard condition for discrete ill-posed problems. *BIT Numerical Mathematics*, 30(4):658–672, 1990.
- [17] N. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, second edition, 2002.
- [18] Bangti Jin and Jun Zou. A bayesian inference approach to the ill-posed cauchy problem of steady-state heat conduction. *International journal for numerical methods in engineering*, 76(4):521–544, 2008.
- [19] Mohamed Larbi Kadri, Jalel Ben Abdallah, and Thouraya Nouri Baranger. Identification of internal cracks in a three-dimensional solid body via Steklov–Poincaré approaches. *Comptes Rendus Mécanique*, 339(10):674–681, 2011.
- [20] Axel Klawonn and Olof Widlund. Feti and Neumann-Neumann iterative substructuring methods: Connections and new results. *Communications on Pure and Applied Mathematics*, 54(1):57–90, 2001.
- [21] Vladimir Arkad’evich Kozlov, Vladimir Gilelevich Maz’ya, and AV Fomin. An iterative method for solving the Cauchy problem for elliptic equations. *Zhurnal Vychislitel’noi Matematiki i Matematicheskoi Fiziki*, 31(1):64–74, 1991.
- [22] P Ladeveze, M Reynier, and D Nedjar. Parametric correction of finite element models using modal tests. In *Inverse problems in engineering mechanics*, pages 91–100. Springer, 1993.
- [23] Pierre Ladevèze, Djamel Nedjar, and Marie Reynier. Updating of finite element models using vibration tests. *AIAA Journal*, 32(7):1485–1491, 1994.
- [24] G Landi, E Loli Piccolomini, and I Tomba. A stopping criterion for iterative regularization methods. *Applied Numerical Mathematics*, 106:53–68, 2016.
- [25] L Marin and D Lesnic. Regularized boundary element solution for an inverse boundary value problem in linear elasticity. *Communications in numerical methods in engineering*, 18(11):817–825, 2002.
- [26] Paolo Novati. Some properties of the arnoldi-based methods for linear ill-posed problems. *SIAM Journal on Numerical Analysis*, 55(3):1437–1455, 2017.
- [27] Dianne P O’Leary. The block conjugate gradient algorithm and related methods. *Linear algebra and its applications*, 29:293–322, 1980.
- [28] Daniel J. Rixen and Charbel Farhat. A simple and efficient extension of a class of substructure based preconditioners to heterogeneous structural mechanics problems. *International Journal for Numerical Methods in Engineering*, 44(4):489–516, 1999.
- [29] Yousef Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.
- [30] Fuzhen Zhang, editor. *The Schur Complement and Its Applications*, volume 4 of *Numerical Methods and Algorithms*. Springer-Verlag, US, 2005.