Challenges in achieving scalable and robust linear solvers

L. Grigori

Alpines

Inria Paris and LJLL, Sorbonne University

with H. Al Daas, MPI Magdebourg

P. Jolivet, CNRS IRIT

P. H. Tournier, LJLL, CNRS, Alpines, Sorbonne University

September 18, 2019

European Research Council

Established by the European Commission

informatics mathematics

SORBONNE UNIVERSITÉ

CRÉATEURS DE FUTURS DEPUIS 1257

Plan

Challenge in getting scalable and robust solvers

On large scale computers, Krylov solvers reach less than 2% of the peak performance.

- Typically, each iteration of a Krylov solver requires
  - Sparse matrix vector product
    $\rightarrow$ point-to-point communication
  - Dot products for orthogonalization
    $\rightarrow$ global communication

- When solving complex linear systems arising, e.g. from large discretized systems of PDEs with strongly heterogeneous coefficients
  most of the highly parallel preconditioners lack robustness
  - wrt jumps in coefficients / partitioning into irregular subdomains, e.g. one level DDM methods (Additive Schwarz, RAS)
  - A few small eigenvalues hinder the convergence of iterative methods

Can we have both scalable and robust methods ?

Difficult ... but crucial ...

since complex and large scale applications very often challenge existing methods

Focus on increasing scalability by reducing coummunication/increasing arithmetic intensity while preserving robustness/dealing with small eigenvalues.

- Robust preconditioners that guarantee the condition number of preconditioned matrix
  - Robust multilevel Additive Schwarz, using Geneo framework
- Enlarged Krylov methods
  - reduce communication,
  - increase arithmetic intensity - compute sparse matrix-set of vectors product.

Can we have both scalable and robust methods ?

Difficult ... but crucial ...
since complex and large scale applications very often challenge existing methods

Focus on increasing scalability by reducing coummunication/increasing arithmetic intensity while preserving robustness/dealing with small eigenvalues.

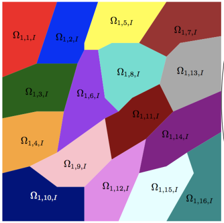
- Robust preconditioners that guarantee the condition number of preconditioned matrix
  - Robust multilevel Additive Schwarz, using Geneo framework
- Enlarged Krylov methods
  - reduce communication,
  - increase arithmetic intensity - compute sparse matrix-set of vectors product.

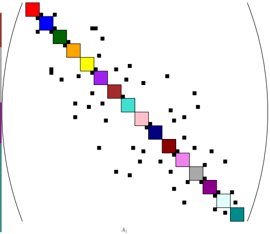
Notations

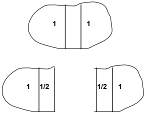
Solve $M^{-1}Ax = M^{-1}b$, where $A \in \mathbb{R}^{n \times n}$ is SPD

Notations:

- DOFs partitioned into $\{\Omega_{1j}\}_{j=1}^{N_1}$ overlapping domains of dimensions $n_{11}, n_{12}, \ldots n_{1,N_1}$
- $R_{1j} \in \mathbb{R}^{n_{1j} \times n}$ restriction operator, $R_{1j} = I_n(\Omega_{1j}, :)$
- $A_{1j} \in \mathbb{R}^{n_{1j} \times n_{1j}}$: restriction of A to domain j, $A_{1j} = R_{1j} A R_{1j}^T$
- $\{D_{1j}\}_{j=1}^{N_1}$: algebraic partition of unity , $I_n = \sum_{j=1}^{N_1} R_{1j}^T D_{1j} R_{1j}$







Additive and Restrictive Additive Schwarz methods

- Original idea from Schwarz algorithm at the continuous level (Schwarz 1870)
- Symmetric formulation, Additive Schwarz (1989) defined as

$$M_{AS}^{-1} := \sum_{j=1}^{N_1} R_{1j}^T A_{1j}^{-1} R_{1j}$$

- Restricted Additive Schwarz (Cai & Sarkis 1999) defined as

$$M_{RAS}^{-1} := \sum_{j=1}^{N_1} R_{1j}^T D_{1j} A_{1j}^{-1} R_{1j}$$

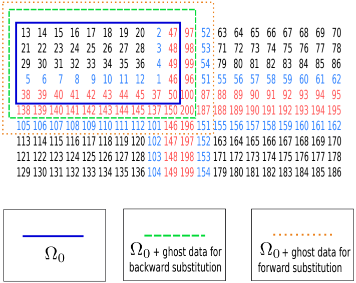
- In practice, RAS more efficient than AS

Relation between IC0 and RAS

- Consider an Alternating Min Max layers ordering for IC0
- Duplicate data on domain j, include all DOFs at distance 2 plus a constant number of other DOFs.

- With $L_j L_j^T$ the IC0 factor of domain j, IC0 preconditioner is

$$M_{IC0}^{-1} := \sum_{j=1}^{N_1} R_{1j}^T D_{1j} (L_j L_j^T)^{-1} R_{1j}$$



- For structured 2D grids, RAS with IC0 in subdomains and overlap 2 similar to IC0 (modulo a constant number of extra DOFs per subdomain)
- with S. Moufawad and S. Cayrols (proofs in their Phds thesis)

Upper bound for the eigenvalues of $M_{AS,1}^{-1}A$

Let k_{1c} be number of distinct colours to colour the subdomains of A. The following holds (e.g. Chan and Mathew 1994)

$$\lambda_{max}(M_{AS,1}^{-1}A) \leq k_{1c}$$

$\rightarrow$ Two level preconditioners are required

Two level preconditioners

Given a coarse subspace S_1, $S_1 = \text{span}\,(V_1)$, $V_1 \in \mathbb{R}^{n \times n_2}$, the coarse grid $A_2 = V_1^T A V_1$.
the two level AS preconditioner is,

$$M_{AS,2}^{-1} := V_1 \left(A_2\right)^{-1} V_1^T + \sum_{j=1}^{N_1} R_{1j}^T \left(A_{1j}\right)^{-1} R_{1j}$$

Let k_{1c} be minimum number of distinct colors so that $\{span\{R_{1j}^T\}\}_{1 \leq i \leq N_1}$ of the same color are mutually A-orthogonal. The following holds (e.g. Chan and Mathew 1994)

$$\lambda_{max}(M_{AS,2}^{-1}) \leq k_{1c} + 1$$

How to compute the coarse subspace $S_1 = \text{span}\,(V_1)$

- Nicolaides 87 (CG): kernel of the operator (constant vectors)

$$V_1 \quad := \quad \left(R_{1j}^T D_{1j} R_{1j} 1\right)_{j=1:N_1}$$

- Other early references: [Morgan 92] (GMRES), [Chapman, Saad 92], [Kharchenko, Yeremin 92], [Burrage, Ehrel, and Pohl, 93]
- Estimations of eigenvectors corresponding to smallest eigenvalues / knowledge from the physics
- Geneo [Spillane et al., 2014]: through solving local Gen EVPs, bounds smallest eigenvalue for standard FE and bilinear forms, SPD input matrix

subd	dofs	AS	AS-ZEM (V_1)	GenEO (V_1)
4	1452	79	54 (24)	16 (46)
8	29040	177	87 (48)	16 (102)
16	58080	378	145 (96)	16 (214)

(V_1): size of the coarse space
AS-ZEM Nicolaides with rigid body motions, 6 per subdomain
Results for 3D elasticity problem provided by F. Nataf

How to compute the coarse subspace $S_1 = \text{span}\,(V_1)$

- Nicolaides 87 (CG): kernel of the operator (constant vectors)

$$V_1 \quad := \quad \left(R_{1j}^T D_{1j} R_{1j} 1\right)_{j=1:N_1}$$

- Other early references: [Morgan 92] (GMRES), [Chapman, Saad 92], [Kharchenko, Yeremin 92], [Burrage, Ehrel, and Pohl, 93]
- Estimations of eigenvectors corresponding to smallest eigenvalues / knowledge from the physics
- Geneo [Spillane et al., 2014]: through solving local Gen EVPs, bounds smallest eigenvalue for standard FE and bilinear forms, SPD input matrix

subd	dofs	AS	AS-ZEM (V_1)	GenEO (V_1)
4	1452	79	54 (24)	16 (46)
8	29040	177	87 (48)	16 (102)
16	58080	378	145 (96)	16 (214)

(V_1): size of the coarse space
AS-ZEM Nicolaides with rigid body motions, 6 per subdomain
Results for 3D elasticity problem provided by F. Nataf

Plan

Fictitious space lemma (Nepomnyaschikh 1991)

Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n_B\times n_B}$ be two SPD matrices. Suppose there exists

$$\begin{array}{rccc} \mathscr{R} : & \mathbb{R}^{n_B} & \rightarrow & \mathbb{R}^{n} \\ & v_{n_B} & \mapsto & \mathscr{R} v_{n_B}, \end{array}$$

such that the following holds

1. The operator $\mathscr{R}$ is surjective
2. There exists $c_u > 0$ such that

$$(\mathscr{R} v_{n_B})^T A (\mathscr{R} v_{n_B}) \leq c_u \; v_{n_B}^\top B v_{n_B}, \quad \forall v_{n_B} \in \mathbb{R}^{n_B}$$

3. Stable decomposition: there exists $c_l > 0$ such that $\forall v \in \mathbb{R}^n, \exists v_{n_B} \in \mathbb{R}^{n_B}$ with $v = \mathscr{R} v_{n_B}$ and

$$c_l \; v_{n_B}^\top B v_{n_B} \leq (\mathscr{R} v_{n_B})^\top A (\mathscr{R} v_{n_B}) = v^\top A v$$

Then, the spectrum of the operator $\mathscr{R} B^{-1} \mathscr{R}^T A$ is in the segment $[c_l, c_u]$.

Geneo two level DDM preconditioner

Consider the generalized eigenvalue problem for each domain j, for given τ:

$$\begin{gathered} Find\ (u_{1jk},\ \lambda_{1jk})\ \in \mathbb{R}^{n_{i,1}} \times \mathbb{R}, \lambda_{1jk} \leq 1/\tau \\ such\ that\ \tilde{A}_{1j}^{Neu} u_{1jk} = \lambda_{1jk} D_{1j} A_{1j} D_{1j} u_{1jk} \end{gathered}$$

where $\tilde{A}_{1j}^{Neu}$ is the Neumann matrix of domain i, V_1 basis of S_1,

$$\mathcal{S}_1 := \bigoplus_{j=1}^{N_1} D_{1j} R_{1j}^\top Z_{1j}, \quad Z_{1j} = \text{span}\left\{ u_{1jk} \mid \lambda_{1jk} < 1/\tau \right\}$$

$$M_{AS,2_{Geneo}}^{-1} := V_1 \left(V_1^T A V_1 \right)^{-1} V_1^T + \sum_{j=1}^{N_1} R_{1j}^T A_{1j}^{-1} R_{1j}$$

Theorem (Spillane, Dolean, Hauret, Nataf, Pechstein, Scheichl'14)

With two technical assumptions fulfilled by standard FE and bilinear forms

$$\kappa \left(M_{AS,2_{Geneo}}^{-1} A \right) \leq (k_{1c} + 1) \left(2 + (2k_{1c} + 1) k_1 \tau \right)$$

where k_{1c} = number of distinct colours to colour the graph of A,
k_1 = max number of domains that share a common vertex.

Local SPSD splitting of A wrt a subdomain

with H. Al Daas [Daas and Grigori, 2019]

- Challenge: can we find an algebraic stable decomposition ?
- We call $\{\tilde{A}_{1j}\}_{j=1}^{N_1}$, $\tilde{A}_{1j} \in \mathbb{R}^{n \times n}$ a splitting of A into local SPSD matrices if the following conditions are satisfied:

$$\begin{aligned} R_{1j,\Delta} \tilde{A}_{1j} &= 0, \\ u^\top \sum_{j=1}^{N_1} \tilde{A}_{1j} u &\leqslant k u^\top A u, \forall u \in \mathbb{R}^n, \end{aligned} \tag{1}$$

  where $R_{1j,\Delta}$ is the restriction operator to complimentary unknowns, $\Delta_{1j} = \Omega \setminus \Omega_{1j}$
- First condition means that $\tilde{A}_{1j}$ is local to subdomain Ω_{1j}, i.e., there is a permutation matrix $\mathcal{P}_j$, $\tilde{A}^j_{I,\Gamma} \in \mathbb{R}^{n_{1j} \times n_{1j}}$

$$\mathcal{P}_j \tilde{A}_{1j} \mathcal{P}_j^\top = \begin{pmatrix} \tilde{A}^j_{I,\Gamma} & 0 \\ 0 & 0 \end{pmatrix}.$$

- If these two conditions are satisfied, the construction of the coarse space can be obtained through the theory of Geneo.
- In Geneo, $k = k_1$, max number of domains that share a vertex.

Construction of the coarse space for 2nd level

Consider the generalized eigenvalue problem for each domain j, for given τ:

$$\begin{aligned} &F\mathit{ind}\ (u_{1jk},\ \lambda_{1jk})\ \in \mathbb{R}^{n_{1j}} \times \mathbb{R}, \lambda_{1jk} \le 1/\tau \\ &\mathit{such\ that}\ R_{1j}\tilde{A}_{1j}R_{1j}^T u_{1jk} = \lambda_{1jk} D_{1j} A_{1j} D_{1j} u_{1jk} \end{aligned}$$

where $\tilde{A}_{1j}$ is a local SPSD splitting of A suitably permuted, $S_1 = \text{span}\,(V_1)$,

$$\mathcal{S}_{1,ALSP} := \bigoplus_{j=1}^{N_1} D_{1j} R_{1j}^\top Z_{1j}, \quad Z_{1j} = \text{span}\,\{u_{1jk} \mid \lambda_{1jk} < 1/\tau\} \tag{2}$$

$$M_{AS,2_{ALSP}}^{-1} := V_1 \left(V_1^T A V_1\right)^{-1} V_1^T + \sum_{j=1}^{N_1} R_{1j}^T A_{1j}^{-1} R_{1j} \tag{3}$$

$$\kappa\left(M_{AS,2_{ALSP}}^{-1} A\right) \le (k_c + 1)\,(2 + (2k_c + 1)k\tau)$$

where k_{1c} is the number of distinct colors required to color the graph of A, $k \le N_1$, where N_1 is the number of subdomains

Local SPSD splitting of A wrt a subdomain

- For each domain j, we impose the condition

$$u^\top \tilde{A}_{1j} u \quad \leqslant u^\top A u, \forall u \in \mathbb{R}^n,$$

  there exists a decomposition $A = \tilde{A}_{1j} + C$, where $\tilde{A}_{1j}$ and C are SPSD
- $\tilde{A}_{1j}$ is local to subdomain Ω_{1j}
- Consider domain 1, where B_{11} corresponds to interior DOFs, B_{22} the DOFs at the interface of 1 with all other domains, and B_{33} the remaining DOFs:

$$A = \begin{pmatrix} B_{11} & B_{12} & \\ B_{21} & B_{22} & B_{23} \\ & B_{32} & B_{33} \end{pmatrix}$$

- We note $S(B_{22})$ the Schur complement with respect to B_{22},

$$S(B_{22}) = B_{22} - B_{21} B_{11}^{-1} B_{12} - B_{23} B_{33}^{-1} B_{32}.$$

Characterization of algebraic local SPSD splittings

Algebraic local SPSD splitting lemma

Let $A \in \mathbb{R}^{n\times n}$, an SPD matrix, and $\tilde{A}_{11} \in \mathbb{R}^{n\times n}$ be partitioned as follows

$$A = \begin{pmatrix} B_{11} & B_{12} & \\ B_{21} & B_{22} & B_{23} \\ & B_{32} & B_{33} \end{pmatrix}, \quad \tilde{A}_{11} = \begin{pmatrix} B_{11} & B_{12} & \\ B_{21} & \tilde{B}_{22} & \\ & & 0 \end{pmatrix}$$

where $B_{ii} \in \mathbb{R}^{m_i \times m_i}$ is non trivial matrix for $i \in \{1,2,3\}$. If $\tilde{B}_{22} \in \mathbb{R}^{m_2\times m_2}$ is a symmetric matrix verifying the following inequalities

$$u^T B_{21} B_{11}^{-1} B_{12} u \leq u^T \tilde{B}_{22} u \leq u^T \left(B_{22} - B_{23} B_{33}^{-1} B_{32} \right) u, \quad \forall u \in \mathbb{R}^{m_2},$$

then $A - \tilde{A}_{11}$ is SPSD, that is the following inequality holds

$$0 \leq u^T \tilde{A}_{11} u \leq u^T A u, \quad \forall u \in \mathbb{R}^n.$$

Remember: $\quad S(B_{22}) = B_{22} - B_{21} B_{11}^{-1} B_{12} - B_{23} B_{33}^{-1} B_{32}.$

Properties of the algebraic splitting

Algebraic local SPSD splitting $\tilde{A}_1$ satisfies

$$u^T B_{21} B_{11}^{-1} B_{12} u \leq u^T \tilde{B}_{22} u \leq u^T \left(B_{22} - B_{23} B_{33}^{-1} B_{32} \right) u, \quad \forall u \in \mathbb{R}^{m_2} \qquad (4)$$

1. The set of matrices $\tilde{A}_{11}$ that verify the condition (4) is not empty
2. There exist matrices $\tilde{B}_{22}$ that verify the condition (4) with strict inequalities, e.g.

   $$\tilde{B}_{22} := \frac{1}{2} S(B_{22}) + B_{21} B_{11}^{-1} B_{12},$$

   where $S(B_{22}) = B_{22} - B_{21} B_{11}^{-1} B_{12} - B_{23} B_{33}^{-1} B_{32}$. Then we have,

   $$\tilde{B}_{22} - B_{21} B_{11}^{-1} B_{12} = \tilde{B}_{22} - \left(B_{22} - B_{23} B_{33}^{-1} B_{32} \right) = \frac{1}{2} S(B_{22}),$$

   which is an SPD matrix. Hence, the strict inequalities in (4) follow.
3. The left and right inequalities are optimal

Dimension of coarse subspace

Let $\tilde{A}_j^1, \tilde{A}_j^2$ be two SPSD splittings of A associated to domain j and S_{ALSP}^1, S_{ALSP}^2 associated coarse subspaces, eq (2). If

$u^T \tilde{A}_j^1 u \leq u^T \tilde{A}_j^2 u, \ \ \forall u \in \mathbb{R}^n, \ \ j = 1, \dots N_1,$ then $dim(S_{ALSP}^1) \geq dim(S_{ALSP}^2)$

- $dim(S_{1,ALSP})$ associated to $\tilde{A}_{1j}$, $j = 1, \dots, N_1$ constructed similarly to $\tilde{A}_{11}$ below, is minimal,

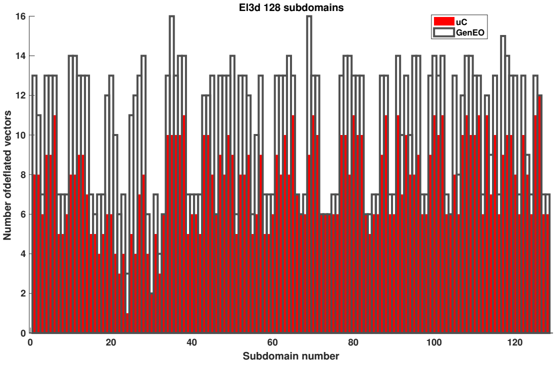
$$\tilde{A}_{11} = \begin{pmatrix} B_{11} & B_{12} & \\ B_{21} & B_{22} - B_{23} B_{33}^{-1} B_{32} & \\ & & 0 \end{pmatrix}$$

- $dim(S_{1,ALSP})$ associated to $\tilde{A}_{1j}$, $j = 1, \dots, N_1$ constructed similarly to $\tilde{A}_{1j}$ below, is maximal (dimension of the overlap at least for each domain j),

$$\tilde{A}_{11} = \begin{pmatrix} B_{11} & B_{12} & \\ B_{21} & B_{21} B_{11}^{-1} B_{12} & \\ & & 0 \end{pmatrix}$$

Comparison with Geneo

- Number of deflated vectors per subdomain in GenEO (black) versus minimal number of deflated vectors per subdomain by ALSP
- 3D elasticity problem, 128 domains



Multilevel methods

Several different multilevel methods exist, only a selection presented here. Often based on hierarchical meshing, both in multigrid and DDM.

- Multispace and multilevel BDDC [Mandel, Sousedik, Dohrmann'08]
- Algebraic multilevel additive Schwarz method [Borzi, De Simone, Di Serafino'13]
- Multilevel Schwarz domain decomposition solver for elasticity problems [Kong and Cai'16]
- Multilevel balancing domain decomposition at extreme scales [Badia, Martin, Principe'16]
- Three level method based on applying recursively the two-level Generalized Dryja-Smith-Widlund preconditioner [Heinlein, Klawonn, Rheinbach, Rover, 18]

Extension of the theory to a multilevel method

with H. Al Daas, P. Jolivet, P. H. Tournier [Daas et al., 2019]
Based on a hierarchy of robust coarse spaces S_i defined for $i = 2 : L$.

- Given coarse space S_1, $S_1 = \text{span}\,(V_1)$, coarse grid matrix $A_2 = V_1^\top A V_1$, preconditioner for A is:

$$M_{MAS}^{-1} = M_{A_1,MAS}^{-1} = V_1 A_2^{-1} V_1^T + \sum_{j=1}^{N_1} R_{1j}^\top A_{1j}^{-1} R_{1j}, \tag{5}$$

- For level $i = 2 : L$, define preconditioner M_i^{-1} for A_i based on AS and additive coarse grid correction,

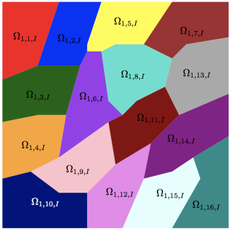
$$M_{A_i,MAS}^{-1} = V_i A_{i+1}^{-1} V_i^T + \sum_{j=1}^{N_i} R_{ij}^\top A_{ij}^{-1} R_{ij}, \tag{6}$$

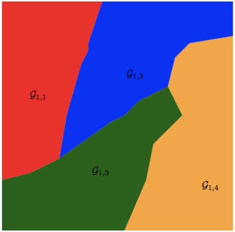
- Coarse space S_i chosen such that condition number of $M_{i,MAS}^{-1} A_i$ is bounded.
- $S_i = \text{span}\,(V_i)$, coarse grid matrix $A_{i+1} = V_i^T A_i V_i$

Aggregation of subdomains into superdomains

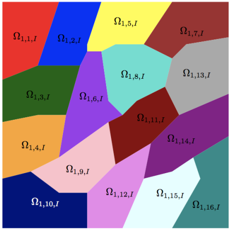
For each level $i = 1 : L$ and $j = 1 : N_i$,

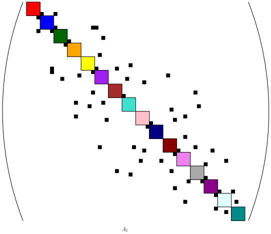
- $\Omega_i = [1 : n_i]$ is the set of unknowns at level i, partitioned into N_i overlapping subdomains, $\Omega_{i,j}$
- $R_{i,j} \in \mathbb{R}^{n_{i,j} \times n_i}$ restriction operator, $R_{i,j} = I_{n_i}(\Omega_{i,j}, :)$
- $R_{i,j,\Delta}$ restriction operator to complimentary unknowns, $\Delta_{i,j} = \Omega_i \setminus \Omega_{i,j}$, $R_{i,j} = I_{n_i}(\Omega_{i,j}, :)$
- Define superdomains $\mathcal{G}_{i,k}$, $k = 1 : N_{i+1}$, as the concatenation of d neighboring domains, $\bigcup_{k=1}^{N_{i+1}} \mathcal{G}_{i,k} = \{1, \ldots, N_i\}$.





Multilevel Additive Schwarz M_{MAS}





for level $i = 1$ and each domain $j = 1 : N_1$ in parallel ($A = A_1$) **do**

$A_{1j} = R_{1j} A_1 R_{1j}^T$ *(local matrix associated to domain j)*

$\tilde{A}_{1j}^{Neu}$ is Neumann matrix of domain j (local SPSD splitting)

Solve Gen EVP, set $Z_{1j} = \text{span}\left\{ u_{1jk} \mid \lambda_{1jk} < \frac{1}{\tau} \right\}$

Find $(u_{1jk}, \lambda_{1jk}) \in \mathbb{R}^{n_{1j}} \times \mathbb{R}$

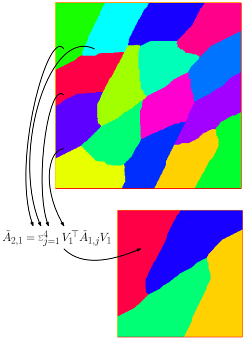
$\tilde{A}_{1j}^{Neu} u_{1jk} = \lambda_{1jk} D_{1j} A_{1j} D_{1j} u_{1jk}$.

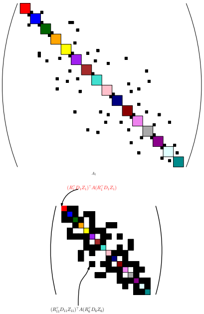
Let $\mathcal{S}_1 = \bigoplus_{j=1}^{N_1} D_{1j} R_{1j}^\top Z_{1j}$, $S_1 = \text{span}\,(V_1)$, $A_2 = V_1^T A_1 V_1$, $A_2 \in \mathbb{R}^{n_2 \times n_2}$

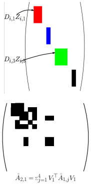
end for

Preconditioner defined as: $M_{A_1,MAS}^{-1} = V_1 A_2^{-1} V_1^T + \sum_{j=1}^{N_1} R_{1j}^\top A_{1j}^{-1} R_{1j}$

Multilevel Additive Schwarz M_{MAS}







for level $i = 2$ to $\log_d N_i$ **do**
  for each domain $j = 1 : N_i$ in parallel **do**
    $\tilde{A}_{ij} = \sum_{k=(j-1)d+1}^{jd} V_{i-1}^T \tilde{A}_{i-1,k} V_{i-1}$ (local SPSD splitting)
    $A_{ij} = R_{ij} A_i R_{ij}^T$ (local matrix associated to domain j)
    Solve Gen EVP, $Z_{ij} = \text{span}\left\{u_{ijk} \mid \lambda_{ijk} < \frac{1}{\tau}\right\}$
      Find $(u_{ijk}, \lambda_{ijk}) \in \mathbb{R}^{n_{ij}} \times \mathbb{R}$
      $R_{ij} \tilde{A}_{ij} R_{ij}^\top u_{ijk} = \lambda_{ijk} D_{ij} A_{ij} D_{ij} u_{ijk}$
    Let $\mathcal{S}_i = \bigoplus_{j=1}^{N_i} D_{ij} R_{ij}^\top Z_{ij}$, $S_i = \text{span}(V_i)$, $A_{i+1} = V_i^T A_i V_i$, $A_{i+1} \in \mathbb{R}^{n_{i+1} \times n_{i+1}}$
end for end for

$$M_{A_i,MAS}^{-1} = V_i A_{i+1}^{-1} V_i^T + \sum_{j=1}^{N_i} R_{ij}^\top A_{ij}^{-1} R_{ij}$$

Definition of local SPSD matrices at each level

For each level $i+1 = 2 : L$, $N_{i+1} = N_i/d$, coarse grid matrix A_{i+1}

- Let $\tilde{A}_{i,j}$, $j = 1, \ldots, N_i$ be local SPSD splittings of A_i, that is

$$u^\top \sum_{j=1}^{N_i} \tilde{A}_{i,j} u \le k_i u^\top A_i u \quad \forall u \in \mathbb{R}^{n_i},$$

- Let $\mathcal{G}_{i,1}, \ldots, \mathcal{G}_{i,N_{i+1}}$ be a set of superdomains at level i associated with the partitioning at level $i+1$.
- The matrices $\tilde{A}_{i+1,j}$, $j = 1, \ldots, N_{i+1}$, defined as:

$$\tilde{A}_{i+1,j} = \sum_{k \in \mathcal{G}_{i,j}} V_i^\top \tilde{A}_{i,k} V_i \tag{7}$$

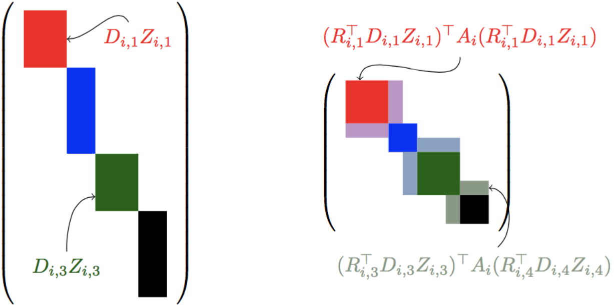
  are local SPSD splittings of A_{i+1}, that is

$$R_{i+1,j,\Delta} \tilde{A}_{i+1,j} \;=\; 0, \text{ that is } \mathcal{P}_j \tilde{A}_j \mathcal{P}_j^\top = \begin{pmatrix} \tilde{A}^j_{I,\Gamma} & 0 \\ 0 & 0 \end{pmatrix}$$

$$u^\top \sum_{j=1}^{N_{i+1}} \tilde{A}_{i+1,j} u \;\le\; k_{i+1} u^\top A_{i+1} u \quad \forall u \in \mathbb{R}^{n_i},$$

Construction of coarse space

- Correspondence between the columns of V_i (on the left) and the rows and columns of A_{i+1} (on the right).
- No overlapping in V_i is possible through a nonoverlapping partition of unity.



Robustness and efficiency of multilevel AS

Theorem (Al Daas, LG, Jolivet, Tournier)

Given the multilevel preconditioner defined at each level $i = 1 : \log_d N_1$ as

$$M^{-1}_{A_i,MAS} = V_i A^{-1}_{i+1} V_i^T + \sum_{j=1}^{N_i} R_{ij}^\top A_{ij}^{-1} R_{ij}$$

then $M^{-1}_{MAS} = M^{-1}_{A_1,MAS}$ and,

$$\kappa(M^{-1}_{A_i,MAS} A_i) \leq (k_{ic}+1)\left(2 + (2k_{ic}+1)k_i\tau\right),$$

where k_{ic} = number of distinct colours to colour the graph of A_i,
k_i = max number of domains that share a common vertex at level i.

- $k_i \leq k_1$ for $i = 2 : L$
- If Neumann matrices are used at the first level, k_1 is bounded by the maximum number of subdomains at level 1 that share an unknown.
- k_{ic} is the minimum number of colors required to colour the graph of A_i
- Constants independent of N_1, number of subdomains at level 1

Robustness and efficiency of multilevel AS

Theorem (Al Daas, LG, Jolivet, Tournier)

Given the multilevel preconditioner defined at each level $i = 1 : \log_d N_1$ as

$$M^{-1}_{A_i,MAS} = V_i A^{-1}_{i+1} V_i^T + \sum_{j=1}^{N_i} R_{ij}^\top A_{ij}^{-1} R_{ij}$$

then $M^{-1}_{MAS} = M^{-1}_{A_1,MAS}$ and,

$$\kappa(M^{-1}_{A_i,MAS} A_i) \leq (k_{ic} + 1)\,(2 + (2k_{ic} + 1)k_i\tau)\,,$$

where k_{ic} = number of distinct colours to colour the graph of A_i,
k_i = max number of domains that share a common vertex at level i.

- $k_i \leq k_1$ for $i = 2 : L$
- If Neumann matrices are used at the first level, k_1 is bounded by the maximum number of subdomains at level 1 that share an unknown.
- k_{ic} is the minimum number of colors required to colour the graph of A_i
- Constants independent of N_1, number of subdomains at level 1

Efficiency of multilevel AS

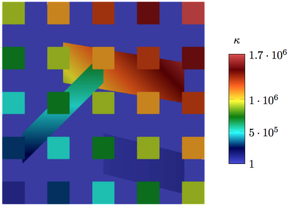
Communication efficiency

- Construction of M_{MAS}^{-1} preconditioner requires $O(\log_d N_1)$ messages.
- Application of M_{MAS}^{-1} preconditioner requires $O((\log_2 N_1)^{\log_d N_1})$ messages per iteration.

Difusion and linear elasticity test cases

- Heterogeneous difusion problem, FreeFem++ using $\mathbb{P}_2$ FE in 3D and $\mathbb{P}_4$ in 2D.

$$\begin{aligned} -\nabla \cdot (\kappa(x)\nabla u) &= f \text{ in } \Omega \\ u &= 0 \text{ on } \partial\Omega_D \\ \frac{\partial u}{\partial n} &= 0 \text{ on } \partial\Omega_N \end{aligned}$$

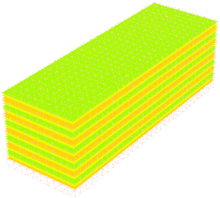


where

$$\kappa(x) = \begin{cases} 10^5([9y]), \textit{if } [9x] \equiv [9y] \equiv 0(mod 2), \\ 1, \quad \textit{otherwise}. \end{cases}$$

- Heterogeneous linear elasticity problem, FreeFem++ using $\mathbb{P}_2$ FE in 3D and $\mathbb{P}_3$ in 2D.

$$\begin{aligned} \operatorname{div}(\sigma(u)) + f &= 0 && \text{on } \Omega, \\ u &= u_D && \text{on } \partial\Omega_D, \\ \sigma(u) \cdot n &= g && \text{on } \partial\Omega_N, \end{aligned}$$



- Young's modulus E and Poisson's ratio ν, $(E_1, \nu_1) = (2 \cdot 10^{11}, 0.25)$, and $(E_2, \nu_2) = (10^7, 0.45)$.

Difusion 2D and 3D with $2,048$ domains

- Number of outer iterations: 32
- Dimensions of A_2: $n_2 = 25 \times 2,048 = 51,200$
  A_3: $n_3 = 20 \times N_2$

Difusion 2D, $441x10^6$ **unknowns**

	2-level Geneo		3-level Geneo		
N2	CS	Solve	CS	Solve	Inner it.
4	2.4	11.9	6.5	27.4	14
16	1.8	11.3	3.6	15.4	15
64	1.9	12.1	3.0	16.7	14
256	2.4	18.4	2.8	13.9	13

- Number of outer iterations: 19
- Dimensions of A_2: $n_2 = 25 \times 2,048 = 51,200$
  A_3: $n_3 = 20 \times N_2$

Difusion 3D, $784x10^6$ **unknowns**

	2-level Geneo		3-level Geneo		
N2	CS	Solve	CS	Solve	Inner it.
4	7.0	20.9	16.9	43.6	17
16	5.0	19.8	7.7	26.7	17
64	5.1	20.1	5.8	32.7	15
256	5.2	24.1	5.3	22.6	14

2-level Geneo, CS: time needed to assemble and factor A_2 on N_2 procs, once V_1 was computed

3-level Geneo, CS: time to assemble local subdomain matrices $\{A_{2,j}\}_{j=1:N_2}$, level 2 local SPSD matrices, solve GenEVP concurrently, assemble and factor A_3 on one proc.

Elasticity 2D and 3D with $2,048$ domains

- Number of outer iterations: 73
- Dimensions of A_2: $n_2 = 50 \times 2,048 = 1.02 \cdot 10^5$
  A_3: $n_3 = 20 \times N_2$

Elasticity 2D, $441 x 10^6$ **unknowns**

	2-level Geneo		3-level Geneo		
N2	CS	Solve	CS	Solve	Inner it.
4	4.8	52.7	22.5	179.3	31
16	3.9	50.3	9.3	124.9	57
64	4.0	53.1	7.2	71.5	34
256	4.8	63.2	6.8	71.2	44

- Number of outer iterations: 45
- Dimensions of A_2: $n_2 = 25 \times 2,048 = 51,200$
  A_3: $n_3 = 20 \times N_2$

Elasticity 3D, $784 x 10^6$ **unknowns**

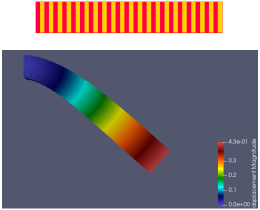
	2-level Geneo		3-level Geneo		
N2	CS	Solve	CS	Solve	Inner it.
4	28.5	46.9	78.9	296.7	23
16	17.3	35.4	24.5	124.5	23
64	15.0	33.2	15.4	62.2	21
256	13.6	40.7	10.6	50.7	23

2-level Geneo, CS: time needed to assemble and factor A_2 on N_2 procs, once V_1 was computed

3-level Geneo, CS: time to assemble local subdomain matrices $\{A_{2,j}\}_{j=1:N_2}$, level 2 local SPSD matrices, solve GenEVP concurrently, assemble and factor A_3 on one proc.

Parallel performance for linear elasticity

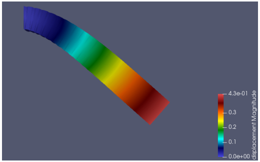
- Machine: IRENE (Genci), Intel Skylake 8168, 2,7 GHz, 24 cores each
- Stopping criterion: 10^{-5}
- Young's modulus E and Poisson's ratio ν take two values, $(E_1, \nu_1) = (2 \cdot 10^{11}, 0.35)$, and $(E_2, \nu_2) = (10^7, 0.45)$



Linear elasticity, $121x10^6$ unknowns, PETSc versus GenEO HPDDM

# P	PETSc GAMG			HPDDM				
	PCSetUp	KSPSolve	Total	Deflation subspace	Domain factor	Coarse matrix	Solve	Total
1,024	39.9	85.7	125.7	185.8	26.8	3.0	62.0	277.7
2,048	42.1	21.2	63.3	76.1	8.5	4.2	28.5	117.3
4,096	95.1	182.8	277.9	32.0	3.6	8.5	18.1	62.4

Parallel performance for linear elasticity (contd)



- Machine: IRENE (Genci), Intel Skylake 8168, 2,7 GHz, 24 cores each
- Stopping criterion: 10^{-5} (10^{-2} for 2nd level)
- Young's modulus E and Poisson's ratio ν take two values, $(E_1, \nu_1) = (2 \cdot 10^{11}, 0.35)$, and $(E_2, \nu_2) = (10^7, 0.45)$

Linear elasticity, $616 \cdot 10^6$ **unknowns, GenEO versus GenEO multilevel**

# P	Deflation subspace	Domain factor	Coarse matrix	Solve	Total	# iter
			GenEO			
8192	113.3	11.9	27.5	52.0	152.8	53
			GenEO multilevel			
8192	113.3	11.9	13.2	52.0	138.5	53

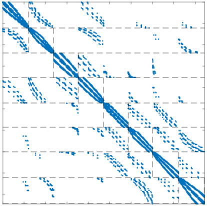
.
A_2 of dimension $328 \cdot 10^3 \times 328 \cdot 10^3$, 40 vectors per subdomain, 10^{-2} tolerance.
A_3 of dimension 5120×5120, 128 procs

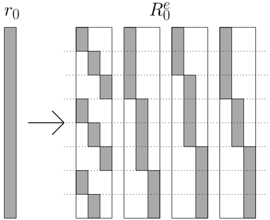
MUMPS for factoring subdomains, Arpack, Pardiso for coarse grids.

Plan

Enlarged Krylov methods [LG, Moufawad, Nataf, 14]

- Partition the matrix into N domains
- Split the residual r_0 into t vectors corresponding to the N domains,





- Generate t new basis vectors, obtain an **enlarged** Krylov subspace

$$\mathcal{K}_{t,k}(A, r_0) = \text{span}\{R_0^e, AR_0^e, A^2 R_0^e, ..., A^{k-1} R_0^e\}$$

$$\mathcal{K}_k(A, r_0) \subset \mathcal{K}_{t,k}(A, r_0)$$

- Search for the solution of the system $Ax = b$ in $\mathcal{K}_{t,k}(A, r_0)$

Enlarged Krylov subspace methods based on CG

Defined by the subspace $\mathcal{K}_{t,k}$ and the following two conditions:

1. Subspace condition: $x_k \in x_0 + \mathcal{K}_{t,k}$
2. Orthogonality condition: $r_k \perp \mathcal{K}_{t,k}$

- At each iteration, the new approximate solution x_k is found by minimizing $\phi(x) = \frac{1}{2}(x^t Ax) - b^t x$ over $x_0 + \mathcal{K}_{t,k}$:

$$\phi(x_k) = min\{\phi(x), \forall x \in x_0 + \mathcal{K}_{t,k}(A, r_0)\}$$

- Can be seen as a particular case of a block Krylov method
  - $AX = S(b)$, such that $S(b)ones(t,1) = b$; $R_0^e = AX_0 - S(b)$
  - Orthogonality condition involves the block residual $R_k \perp \mathcal{K}_{t,k}$

Enlarged Krylov subspace methods based on CG

Defined by the subspace $\mathcal{K}_{t,k}$ and the following two conditions:

1. Subspace condition: $x_k \in x_0 + \mathcal{K}_{t,k}$
2. Orthogonality condition: $r_k \perp \mathcal{K}_{t,k}$

- At each iteration, the new approximate solution x_k is found by minimizing $\phi(x) = \frac{1}{2}(x^t A x) - b^t x$ over $x_0 + \mathcal{K}_{t,k}$:

$$\phi(x_k) = min\{\phi(x), \forall x \in x_0 + \mathcal{K}_{t,k}(A, r_0)\}$$

- Can be seen as a particular case of a block Krylov method
  - $AX = S(b)$, such that $S(b) ones(t,1) = b$; $R_0^e = AX_0 - S(b)$
  - Orthogonality condition involves the block residual $R_k \perp \mathcal{K}_{t,k}$

Related work

- Block Krylov methods [O'Leary, 1980]: solve systems with multiple rhs

$$AX = B,$$

by searching for an approximate solution $X_k \in X_0 + \mathcal{K}_k^{\square}(A, R_0)$,

$$\mathcal{K}_k^{\square}(A, R_0) = block - span\{R_0, AR_0, A^2R_0, ..., A^{k-1}R_0\}.$$

  - coopCG (Bhaya et al, 2012): solve one system by starting with t different initial guesses
  - BRRHS-CG [Nikishin and Yeremin, 1995]: use a block method with t-1 random right hand sides

- Multiple preconditioners
  - GMRES with multiple preconditioners [Greif, Rees, Szyld, 2011]
  - AMPFETI [Rixen, 97], [Gosselet et al, 2015]

- And to reduce communication: s-step methods, pipelined methods

Convergence analysis

Given

- A is an SPD matrix, x^* is the solution of $Ax = b$
- $||x^* - \overline{x}_k||_A$ is the k^{th} error of CG, $e_0 = x^* - x_0$
- $||x^* - x_k||_A$ is the k^{th} error of ECG

Result

CG

$$||x^* - \overline{x}_k||_A \leq 2||e_0||_A \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k$$

where $\kappa = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}$

ECG

$$||x^* - x_k||_A \leq 2||\hat{e}_0||_A \left(\frac{\sqrt{\kappa_t} - 1}{\sqrt{\kappa_t} + 1}\right)^k$$

where $\kappa_t = \frac{\lambda_{max}(A)}{\lambda_t(A)}$, $\hat{e}_0 \equiv E_0(\Phi_1^\top E_0)^{-1}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$, Φ_1 denotes the t eigenvectors associated to the smallest eigenvalues, and E_0 is the initial enlarged error.

From here on, results on enlarged CG obtained with O. Tissot [Grigori and Tissot, 2019].

Classical CG *vs.* Enlarged CG derived from Block CG

Algorithm 1 Classical CG

1: $p_1 = r_0(r_0^\top A r_0)^{-1/2}$
2: **while** $||r_{k-1}||_2 > \varepsilon||b||_2$ **do**
3: $\quad \alpha_k = p_k^\top r_{k-1}$
4: $\quad x_k = x_{k-1} + p_k \alpha_k$
5: $\quad r_k = r_{k-1} - A p_k \alpha_k$
6: $\quad z_{k+1} = r_k - p_k(p_k^\top A r_k)$
7: $\quad p_{k+1} = z_{k+1}(z_{k+1}^\top A z_{k+1})^{-1/2}$
8: **end while**

Cost per iteration

\# flops $= O(\frac{n}{P}) \leftarrow$ BLAS 1 & 2
\# words $= O(1)$
\# messages $= O(1)$ from SpMV + $O(logP)$ from dot prod + norm

Algorithm 2 ECG

1: $P_1 = R_0^e(R_0^{e\top} A R_0^e)^{-1/2}$
2: **while** $|| \sum_{i=1}^{\top} R_k^{(i)}||_2 < \varepsilon||b||_2$ **do**
3: $\quad \alpha_k = P_k^\top R_{k-1}$ $\qquad \triangleright$ $t \times t$ matrix
4: $\quad X_k = X_{k-1} + P_k \alpha_k$
5: $\quad R_k = R_{k-1} - A P_k \alpha_k$
6: $\quad Z_{k+1} = A P_k - P_k(P_k^\top A A P_k) - P_{k-1}(P_{k-1}^\top A A P_k)$
7: $\quad P_{k+1} = Z_{k+1}(Z_{k+1}^\top A Z_{k+1})^{-1/2}$
8: **end while**
9: $x = \sum_{i=1}^{\top} X_k^{(i)}$

Cost per iteration

\# flops $= O(\frac{nt^2}{P}) \leftarrow$ BLAS 3
\# words $= O(t^2) \leftarrow$ Fit in the buffer
\# messages $= O(1)$ from SpMV + $O(logP)$ from A-ortho

Construction of the search directions P_{k+1}

1 Construct Z_{k+1} s.t. $Z_{k+1}^\top AP_i = 0, \forall i \leq k$ by using:

1.a **Orthomin** as in block CG [OLeary, 1980] and original CG method [Hestenes and Stiefel, 1952]:

$$Z_{k+1} = R_k - P_k(P_k^\top AR_k)$$

1.b or **Orthodir** as in ECG [Grigori et al., 2016], based on Lanczos formula [Ashby et al., 1990]:

$$Z_{k+1} = AP_k - P_k(P_k^\top AAP_k) - P_{k-1}(P_{k-1}^\top AAP_k)$$

2 A-orthonormalize P_{k+1}, using e.g. A Cholesky QR:

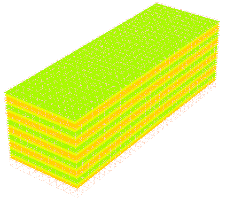
$$P_{k+1} = Z_{k+1}(Z_{k+1}^\top AZ_{k+1})^{-1/2}$$

Orthomin (Omin)	**Orthodir (Odir)**
→ Cheaper	→ More expensive
→ In practice breakdowns	→ In practice no breakdowns

Test cases

- 3 of 5 largest SPD matrices of Tim Davis' collection
- Heterogeneous linear elasticity problem discretized with FreeFem++ using $\mathbb{P}_1$ FE

$$
\begin{aligned}
\operatorname{div}(\sigma(u)) + f &= 0 && \text{on } \Omega, \\
u &= u_D && \text{on } \partial\Omega_D, \\
\sigma(u) \cdot n &= g && \text{on } \partial\Omega_N,
\end{aligned}
$$



- $u \in \mathbb{R}^d$ is the unknown displacement field, f is some body force.
- Young's modulus E and Poisson's ratio ν, $(E_1, \nu_1) = (2 \cdot 10^{11}, 0.25)$, and $(E_2, \nu_2) = (10^7, 0.45)$.

Name	Size	Nonzeros	Problem
Hook_1498	1,498,023	59,374,451	Structural problem
Flan_1565	1,564,794	117,406,044	Structural problem
Queen_4147	4,147,110	316,548,962	Structural problem
Ela_4	4,615,683	165,388,197	Linear elasticity

Enlarged CG: dynamic reduction of search directions

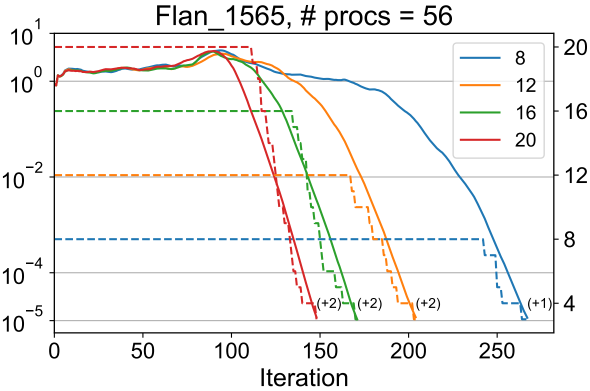


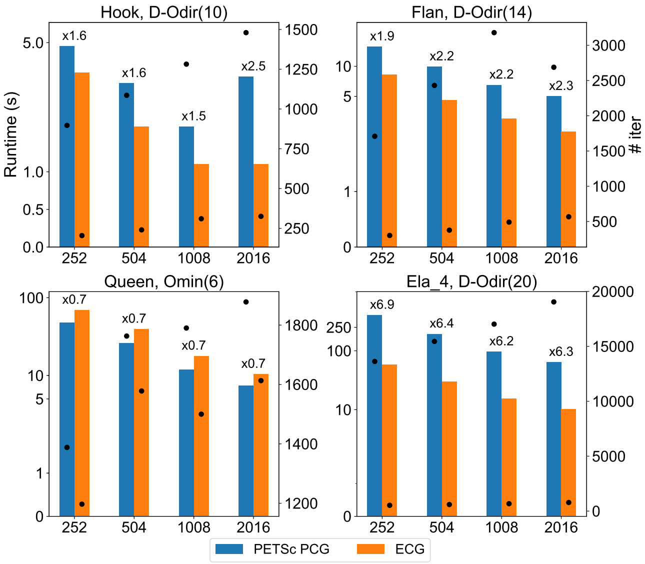
Figure : solid line: normalized residual (scale on the left), dashed line: number of search directions (scale on the right)

→ **Reduction occurs when the convergence has started**

Strong scalability

- Run on Kebnekaise, Umeå University (Sweden) cluster, 432 nodes with Broadwell processors (28 cores per node)
- Compiled with Intel Suite 18
- PETSc 3.7.6 (linked with the MKL)
- Pure MPI (no threading)
- Stopping criterion tolerance is set to 10^{-5} (PETSc default value)
- Block diagonal preconditioner, number blocks equals number of MPI processes
  - Cholesky factorization on the block with MKL-PARDISO solver

Strong scalability



Plan

Conclusions

Most of the methods discussed available in libraries:

- Multilevel Additive Schwarz
  - available in HPDDM as multilevel Geneo (P. Jolivet)
  - code for reproducing the results available at `https://github.com/prj-/aldaas2019multi`
- Krylov subspace methods:
  preAlps library `https://github.com/NLAFET/preAlps`:
  - Enlarged CG: Reverse Communication Interface
  - Enlarged GMRES will be available as well

Acknowledgements

- NLAFET H2020 european project, EMC^2 ERC Synergy project, ANR, Total

Prospects for the future

- Multilevel Additive Schwarz
  - from theory to practice, find an efficient local algebraic splitting that allows to solve the Gen. EVP locally on each processor

Collaborators: S. Cayrols, H. Al Daas, P. Jolivet, S. Moufawad, F. Nataf, P. H. Tournier, O. Tissot.

References (1)

Ashby, S. F., Manteuffel, T. A., and Saylor, P. E. (1990).
A taxonomy for conjugate gradient methods.
SIAM Journal on Numerical Analysis, 27(6):1542–1568.

Daas, H. A. and Grigori, L. (2019).
A class of efficient locally constructed preconditioners based on coarse spaces.
SIAM Journal on Matrix Analysis and Applications, 40:66–91.

Daas, H. A., Grigori, L., Jolivet, P., and Tournier, P. H. (2019).
A multilevel Schwarz preconditioner based on a hierarchy of robust coarse spaces.
Technical report, Inria, https://hal.archives-ouvertes.fr/hal-02151184.

Grigori, L., Moufawad, S., and Nataf, F. (2016).
Enlarged Krylov Subspace Conjugate Gradient Methods for Reducing Communication.
SIAM Journal on Scientific Computing, 37(2):744–773.

Grigori, L. and Tissot, O. (2019).
Scalable linear solvers based on enlarged Krylov subspaces with dynamic reduction of search directions.
SIAM Journal on Scientific Computing.

Hestenes, M. R. and Stiefel, E. (1952).
Methods of conjugate gradients for solving linear systems.
Journal of research of the National Bureau of Standards., 49:409–436.

OLeary, D. P. (1980).
The block conjugate gradient algorithm and related methods.
Linear Algebra and Its Applications, 29:293–322.

References (2)

Spillane, N., Dolean, V., Hauret, P., Nataf, F., Pechstein, C., and Scheichl, R. (2014).
Abstract robust coarse spaces for systems of pdes via gereralized eigenproblems in the overlap.
Numerische Mathematik, 126(4):741–770.