Preconditioners for Krylov subspace methods

L. Grigori

EPFL and PSI

December 10, 2024

Plan

Preconditioned Krylov subspace methods with domain decomposition

- One level Additive Schwarz (AS) methods
- Two level preconditioners

Plan

Preconditioned Krylov subspace methods with domain decomposition

- One level Additive Schwarz (AS) methods
- Two level preconditioners

Preconditioned Krylov subspace methods

- Solve by using iterative methods

$$Ax = b.$$

- Convergence depends on $\kappa(A)$ and the eigenvalue distribution (for SPD matrices).
- To accelerate convergence, solve

$$M^{-1}Ax = M^{-1}b,$$

  where
  - M approximates well the inverse of A and/or
  - improves $\kappa(A)$, the condition number of A.
- Ideally, we would like to bound $\kappa(A)$, independently of the size of the matrix A.

Domain decomposition methods: notations

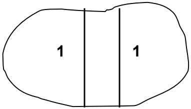
Solve $M^{-1}Ax = M^{-1}b$, where $A \in \mathbb{R}^{n \times n}$ is SPD

For $\mathcal{N} = \{1, \ldots, n\}$, let $\mathcal{N}_i \subset \mathcal{N}$ for $i = 1 \ldots N$ be the subset of DOF of subdomain i, referred to as Ω_i, possibly with overlap. We define:

- The restriction operator $R_i \in \mathbb{R}^{n_i \times n}$, $R_i = I_n(\mathcal{N}_i, :)$.
- The prolongation operator, $R_i^T \in \mathbb{R}^{n \times n_i}$

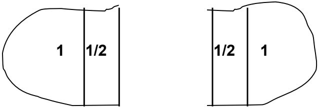
- The matrix associated to domain i, $A_i \in \mathbb{R}^{n_i \times n_i}$,

$$A_i = R_i A R_i^T$$



- The algebraic partition of unity $(D_i)_{1 \leq i \leq N}$,

$$I_n = \sum_{i=1}^{N} R_i^T D_i R_i$$



Additive and Restrictive Additive Schwarz methods

- Original idea from Schwarz algorithm at the continuous level (Schwarz 1870)
- Restricted Additive Schwarz (Cai & Sarkis 1999) defined as

$$M_{RAS}^{-1} := \sum_{i=1}^{N} R_i^T D_i A_i^{-1} R_i$$

- Symmetric formulation, Additive Schwarz (1989) defined as

$$M_{AS}^{-1} := \sum_{i=1}^{N} R_i^T A_i^{-1} R_i$$

- In practice, RAS more efficient than AS

Two level preconditioners

Given a coarse subspace $V_0 \in \mathbb{R}^{n \times n_0}$ and Z its basis, $V_0 = \text{span } Z$, let $R_0 = Z^T$, the coarse grid $R_0 A R_0^T$.
The two level AS preconditioner is

$$M_{AS,2}^{-1} := R_0^T \left(R_0 A R_0^T \right)^{-1} R_0 + \sum_{i=1}^{N} R_i^T \left(A_i \right)^{-1} R_i$$

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

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

Convergence theory

Results from e.g. [Chan and Mathew, 1994, Dolean et al., 2015].

$$M_{AS,2}^{-1}A := \sum_{i=0}^{N} R_i^T \left(A_i\right)^{-1} R_i A = \sum_{i=0}^{N} P_i, \text{ where } P_i = R_i^T \left(A_i\right)^{-1} R_i A$$

P_i are orthogonal projection matrices in the A inner product since

$$\begin{aligned} P_i P_i &= R_i^T \left(A_i\right)^{-1} R_i A R_i^T \left(A_i\right)^{-1} R_i A = R_i^T \left(A_i\right)^{-1} R_i A = P_i \\ A P_i &= A R_i^T \left(A_i\right)^{-1} R_i A = P_i^T A \end{aligned}$$

Recall that $a(u, v) = v^T A u$ and $||P_i|| \leq 1$.

$$\begin{aligned} \lambda_{max}(M_{AS,2}^{-1}A) &= \sup_{u \in \mathbb{R}^n} \frac{a(M_{AS,2}^{-1}Au, u)}{a(u, u)} \\ &= \sup_{u \in \mathbb{R}^n} \sum_{i=0}^{N} \frac{a(P_i u, u)}{||u||_a^2} = \sup_{u \in \mathbb{R}^n} \sum_{i=0}^{N} \frac{a(P_i u, P_i u)}{||u||_a^2} \\ &\leq \sum_{i=0}^{N} \sup_{u \in \mathbb{R}^n} \frac{a(P_i u, P_i u)}{||u||_a^2} \leq N + 1 \end{aligned}$$

Convergence theory (contd)

If we define a-orthogonal projectors

$$\tilde{P}_i = \sum_{j \in \Theta_i} P_j, \text{ for } i = 1, \dots k_c$$

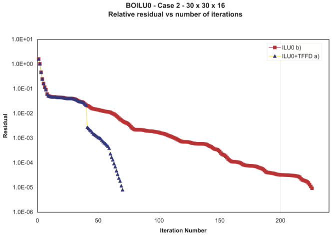
where Θ_i is a set of indices with the same color (that is the indices denoting disjoint subdomains). We can apply the same reasoning and obtain

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

The need for two level preconditioners

- When solving complex linear systems arising, e.g. from large discretized systems of PDEs with strongly heterogeneous coefficients (high contrast, multiscale).

  - Flow in porous media
  - Elasticity problems
  - CMB data analysys (no PDE)



- Most of the existing preconditioners lack robustness
  - wrt jumps in coefficients / partitioning into irregular subdomains, e.g. one level DDM methods (block Jacobi, RAS), incomplete LU
  - A few small eigenvalues hinder the convergence of iterative methods

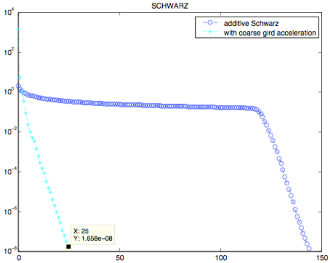
How to compute the coarse subspace $V_0 = \text{span } Z$

- Nicolaides 87 (CG): kernel of the operator (constant vectors) for a Poisson like problem works well

$$Z \quad := \quad \left(R_i^T D_i R_i 1\right)_{i=1:N}$$

Z defined as in (Nicolaides 1987):

$$Z = \begin{pmatrix} 1_{\Omega_1} & & & \\ & 1_{\Omega_2} & & \\ & & \ddots & \\ & & & 1_{\Omega_N} \end{pmatrix}$$



Courtesy of F. Nataf

How to compute the coarse subspace $V_0 = \text{span}\, Z$

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

$$Z \quad := \quad \left(R_i^T D_i R_i 1\right)_{i=1:N}$$

- Z is formed by estimations of eigenvectors corresponding to smallest eigenvalues / knowledge from the physics
- Krylov methods can be used to obtain those estimations

Example of two level preconditoner

Consider the matrix [Tang et al., 2009]:

$$P := I - AZE^{-1}Z^T, \quad E := Z^TAZ$$

where

- Z is the deflation subspace matrix of full rank
- E is the coarse grid correction, a small dense invertible matrix
- P is called deflation matrix, $PAZ = 0$

Example of preconditioner

$$P_{2lvl}^{-1} = M^{-1}P + ZE^{-1}Z^T,$$

where M is the first level preconditioner (eg based on block Jacobi).

- $P_{2lvl}^{-1}AZ = Z$
- The small eigenvalues are shifted to 1.
- P_{2lvl} is not SPD, even when A is, better choices available, but more expensive.

Example of two level preconditoner

Consider the matrix [Tang et al., 2009]:

$$P := I - AZE^{-1}Z^T, \quad E := Z^TAZ$$

where

- Z is the deflation subspace matrix of full rank
- E is the coarse grid correction, a small dense invertible matrix
- P is called deflation matrix, $PAZ = 0$

Example of preconditioner

$$P_{2lvl}^{-1} = M^{-1}P + ZE^{-1}Z^T,$$

where M is the first level preconditioner (eg based on block Jacobi).

- $P_{2lvl}^{-1}AZ = Z$
- The small eigenvalues are shifted to 1.
- P_{2lvl} is not SPD, even when A is, better choices available, but more expensive.

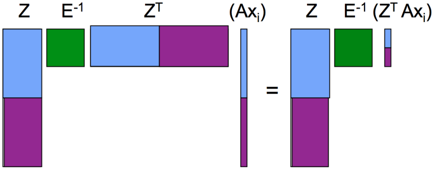
Two level preconditioners (contd)

Computing the preconditioner requires

- Deflation subspace Z, which can be formed by
  - Eigenvectors corresponding to smallest eigenvalues - from previous linear systems solved with different right hand sides, etc.
  - Using knowledge from the physics, partition of the unity, etc.
- Computing AZ and $E = Z^T AZ$.

Applying the preconditioner at each iteration requires

- Computing $y = ZE^{-1}Z^T(Ax_i) = ZE^{-1}Z^T v$
  $\Rightarrow$ involves collective communication when computing $Z^T v$, and solving a linear system with E.



Two level preconditioners (contd)

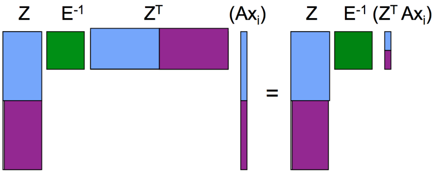
Computing the preconditioner requires

- Deflation subspace Z, which can be formed by
  - Eigenvectors corresponding to smallest eigenvalues - from previous linear systems solved with different right hand sides, etc.
  - Using knowledge from the physics, partition of the unity, etc.
- Computing AZ and $E = Z^T AZ$.

Applying the preconditioner at each iteration requires

- Computing $y = ZE^{-1}Z^T(Ax_i) = ZE^{-1}Z^T v$

  $\Rightarrow$ involves collective communication when computing $Z^T v$,
  and solving a linear system with E.



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.

Chan, T. F. and Mathew, T. P. (1994).
Domain decomposition algorithms.
Acta Numerica, 3:61–143.

Dolean, V., Jolivet, P., and Nataf, F. (2015).
An introduction to domain decomposition methods.
Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
Algorithms, theory, and parallel implementation.

Grigori, L., Moufawad, S., and Nataf, F. (2016).
Enlarged Krylov Subspace Conjugate Gradient Methods for Reducing Communicatio n.
SIAM Journal on Scientific Computing, 37(2):744–773.
Also as INRIA TR 8266.

Grigori, L. and Tissot, O. (2019).
Scalable linear solvers based on enlarged krylov subspaces with dynamic reduct ion of search directions.
SIAM Journal on Scientific Computing, 34(1):206–239.

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.

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

References (2)

Saad, Y. (2003).
Iterative Methods for Sparse Linear Systems.
Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2nd edition.

Tang, J. M., Nabben, R., Vuik, C., and Erlangga, Y. A. (2009).
Comparison of two-level preconditioners derived from deflation, domain decomposition and multigrid methods.
J. Sci. Comput., 39:340–370.