Science & Technology Facilities Council

A Multi-preconditioned GMRES Algorithm

Tyrone Rees

Rutherford Appleton Laboratory

SIAM Applied Linear Algebra : June 2012

Joint work with:
Chen Greif(UBC), Daniel Szyld (Temple University), Scott MacLachlan (Tufts)

The specific problem we are interested in:

Consider solving

$$\mathcal{A}\mathbf{x} = \mathbf{b}$$

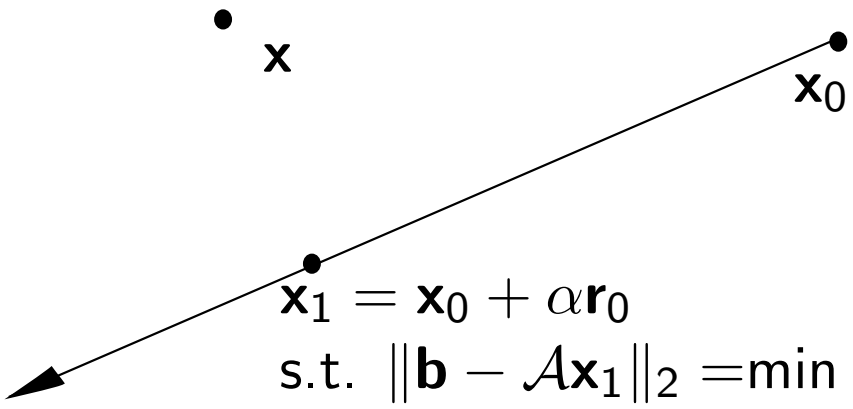
where $\mathcal{A}$ is a large, sparse matrix, using a Krylov subspace method.

Suppose we have two (or more!) possible preconditioners,

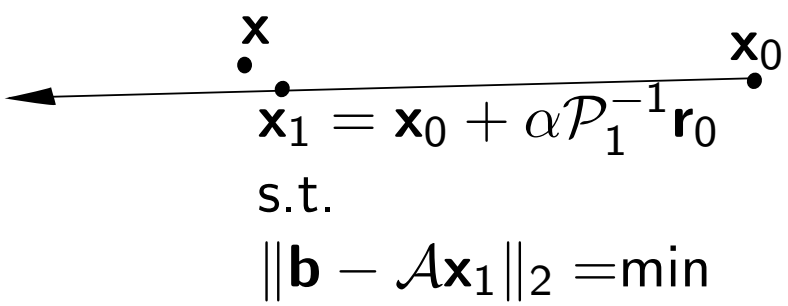
$$\mathcal{P}_1 \qquad \text{and} \qquad \mathcal{P}_2,$$

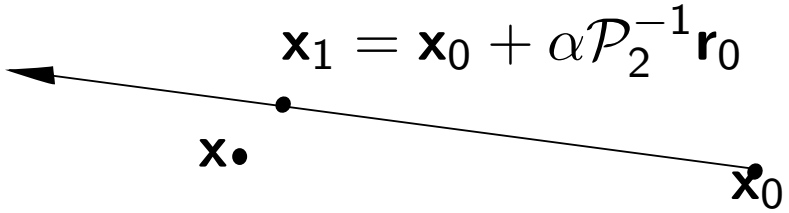
Can we (optimally) combine information from more than one preconditioner?

Diagrams


$\mathbf{x}$
$\mathbf{x}_0$
$\mathbf{x}_1 = \mathbf{x}_0 + \alpha \mathbf{r}_0$
s.t. $\|\mathbf{b} - \mathcal{A}\mathbf{x}_1\|_2 =$min

Diagrams


$\mathbf{x}$
$\mathbf{x}_0$
$\mathbf{x}_1 = \mathbf{x}_0 + \alpha \mathcal{P}_1^{-1} \mathbf{r}_0$
s.t.
$\|\mathbf{b} - \mathcal{A}\mathbf{x}_1\|_2 =$min


$\mathbf{x}_1 = \mathbf{x}_0 + \alpha \mathcal{P}_2^{-1} \mathbf{r}_0$
$\mathbf{x}$
$\mathbf{x}_0$

Multi-preconditioning

Two relevant methods

Multi-preconditioned conjugate gradients (MPCG): Bridson & Greif (2006)

- Combines multiple preconditioners automatically in a (locally) optimal way
- Requires $\mathcal{A}$ and $\{\mathcal{P}_i\}$ to be symmetric positive definite
- We lose the short-term-recurrence of PCG

Flexible GMRES (FGMRES): Saad (1993)

- Allows variable preconditioners – e.g. $\mathcal{P}_1$ on odd iterations, $\mathcal{P}_2$ on even iterations
- Uses all preconditioners, but nontrivial subspace is being constructed and optimality properties are not fully understood

Full Multi-preconditioned Arnoldi

Pick $\mathbf{x}_0$, let $V_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$.
Let $Z_1 = [\mathcal{P}_1^{-1}V_1 \ \cdots \ \mathcal{P}_t^{-1}V_1] \in \mathbb{R}^{n\times t}$
`for` $i = 1 \ldots$ `max_its`
    $Q = AZ_i$
    `for` $j = 1 \ldots i$
        $H_{j,i} = V_j^T W$
        $W = W - V_j H_{j,i}$
    `end`
    $W = V_{i+1}H_{i+1,i}$ (skinny QR factorization)
    $Z_{i+1} = [\mathcal{P}_1^{-1}V_{i+1} \ \cdots \ \mathcal{P}_t^{-1}V_{i+1}]$
`end`

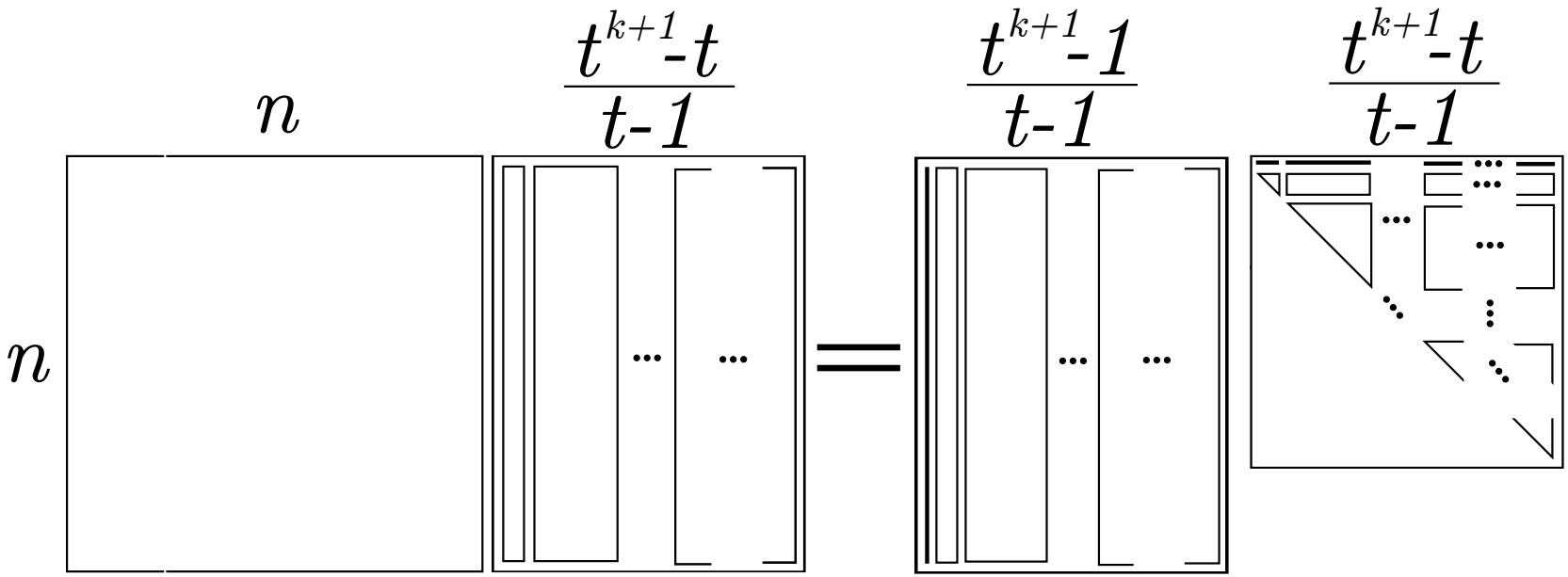
Full Multi-preconditioned Arnoldi

Pick $\mathbf{x}_0$, let $V_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$.
Let $Z_1 = [\mathcal{P}_1^{-1}V_1 \ \cdots \ \mathcal{P}_t^{-1}V_1] \in \mathbb{R}^{n\times t}$
`for` $i = 1 \ldots$ `max_its`
    $Q = AZ_i$
    `for` $j = 1 \ldots i$
        $H_{j,i} = V_j^T W$
        $W = W - V_j H_{j,i}$
    `end`
    $W = V_{i+1}H_{i+1,i}$ (skinny QR factorization)
    $Z_{i+1} = [\mathcal{P}_1^{-1}V_{i+1} \ \cdots \ \mathcal{P}_t^{-1}V_{i+1}]$
`end`

$$\mathcal{A}[Z_1 \ \cdots \ Z_k] = [V_1 \ldots V_{k+1}]\widetilde{H}_k$$

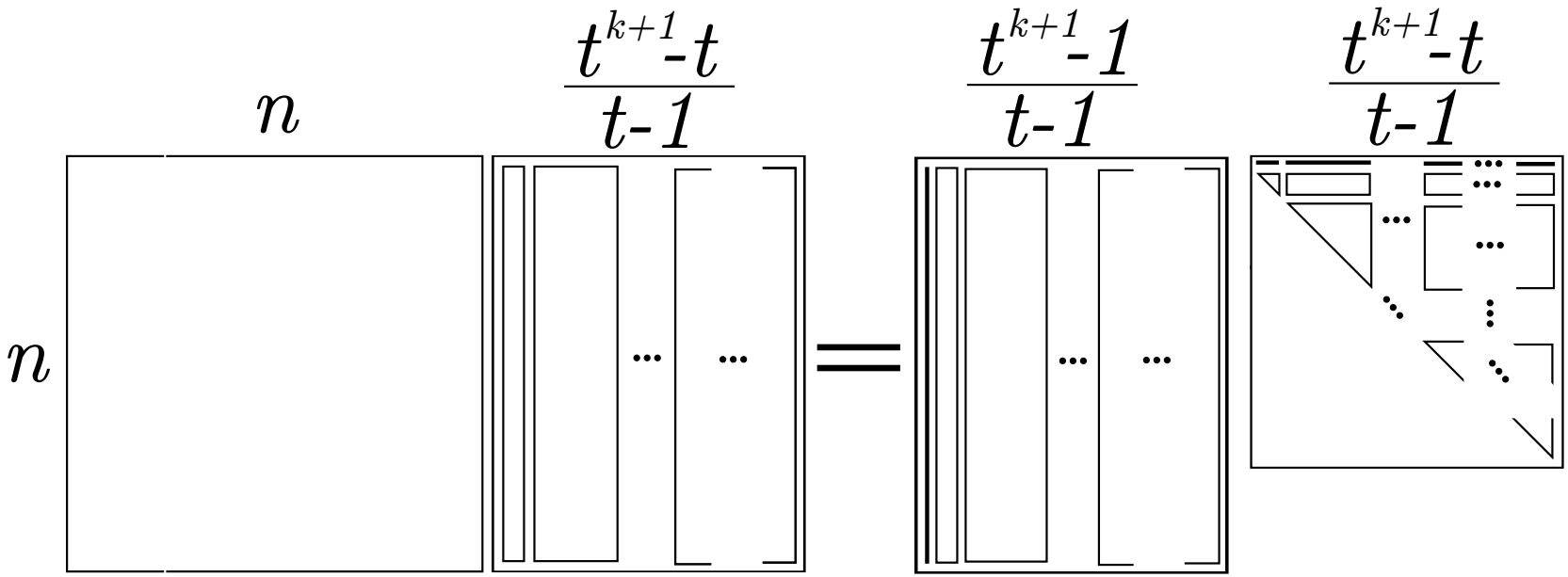
Full Multipreconditioned Arnoldi

$$\mathcal{A}[Z_1 \ \cdots \ Z_k] = [V_1 \ldots V_{k+1}]\widetilde{H}_k$$



Full Multipreconditioned Arnoldi

$$\mathcal{A}[Z_1 \ \cdots \ Z_k] = [V_1 \ldots V_{k+1}]\widetilde{H}_k$$



Impractical!

Multi-preconditioned Arnoldi

Pick $\mathbf{x}_0$, let $V_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$.
Let $Z_1 = [\mathcal{P}_1^{-1}V_1 \ \cdots \ \mathcal{P}_t^{-1}V_1] \in \mathbb{R}^{n\times t}$
`for` $i = 1\ldots$
    $Q = AZ_i$
    `for` $j = 1\ldots i$
        $H_{j,i} = V_j^T W$
        $W = W - V_j H_{j,i}$
    `end`
    $W = V_{i+1}H_{i+1,i}$ (skinny QR factorization)
    $Z_{i+1} = [\mathcal{P}_1^{-1}V_{i+1}^{(1)} \ \cdots \ \mathcal{P}_t^{-1}V_{i+1}^{(t)}]$
`end`

Multi-preconditioned Arnoldi

Pick $\mathbf{x}_0$, let $V_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$.
Let $Z_1 = [\mathcal{P}_1^{-1}V_1 \ \cdots \ \mathcal{P}_t^{-1}V_1] \in \mathbb{R}^{n\times t}$
`for` $i = 1\ldots$
    $Q = AZ_i$
    `for` $j = 1\ldots i$
        $H_{j,i} = V_j^T W$
        $W = W - V_j H_{j,i}$
    `end`
    $W = V_{i+1}H_{i+1,i}$ (skinny QR factorization)
    $Z_{i+1} = [\mathcal{P}_1^{-1}V_{i+1}^{(1)} \ \cdots \ \mathcal{P}_t^{-1}V_{i+1}^{(t)}]$
`end`

$$\mathcal{A}[Z_1 \ \cdots \ Z_k] = [V_1 \ldots V_{k+1}]\widetilde{H}_k$$

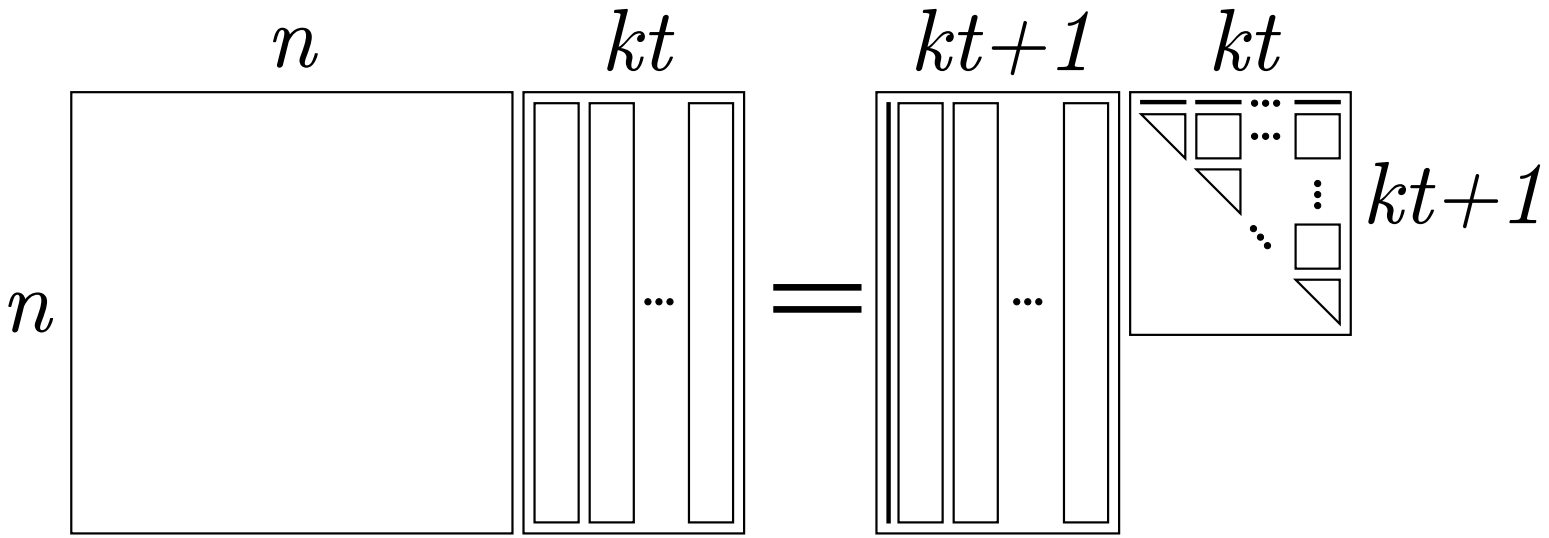
Various possibilities for generating the search directions

Could replace $Z_{i+1} = [\mathcal{P}_1^{-1} V_{i+1}^{(1)} \ \cdots \ \mathcal{P}_t^{-1} V_{i+1}^{(t)}]$ by taking a mix of all columns, say: $Z_{i+1} = [\mathcal{P}_1^{-1} V_{i+1} \mathbf{1} \ \cdots \ \mathcal{P}_t^{-1} V_{i+1} \mathbf{1}]$, where $\mathbf{1}$ is a vector of all ones.

Practical evidence shows "mixing" is typically more effective; no analytical observations to support this.

Multi-preconditioned Arnoldi

$$\mathcal{A}[Z_1 \ \cdots \ Z_k] = [V_1 \ldots V_{k+1}]\widetilde{H}_k$$

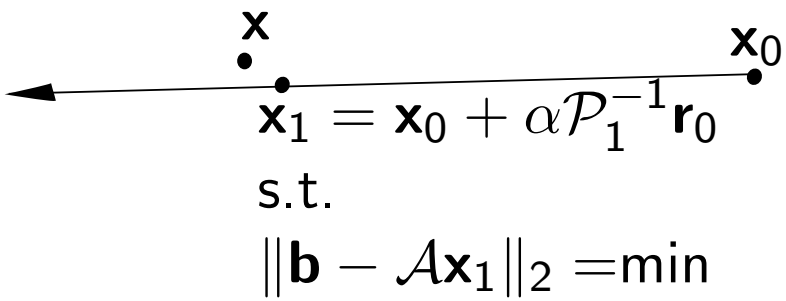


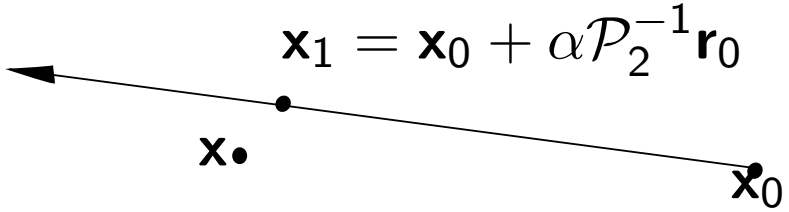
Multi-preconditioned GMRES

Find vector $\mathbf{y}_k$ s.t. $\mathbf{x}_k = \mathbf{x}_0 + Z_k \mathbf{y}_k$
i.e., find $\mathbf{y}_k$ which minimizes

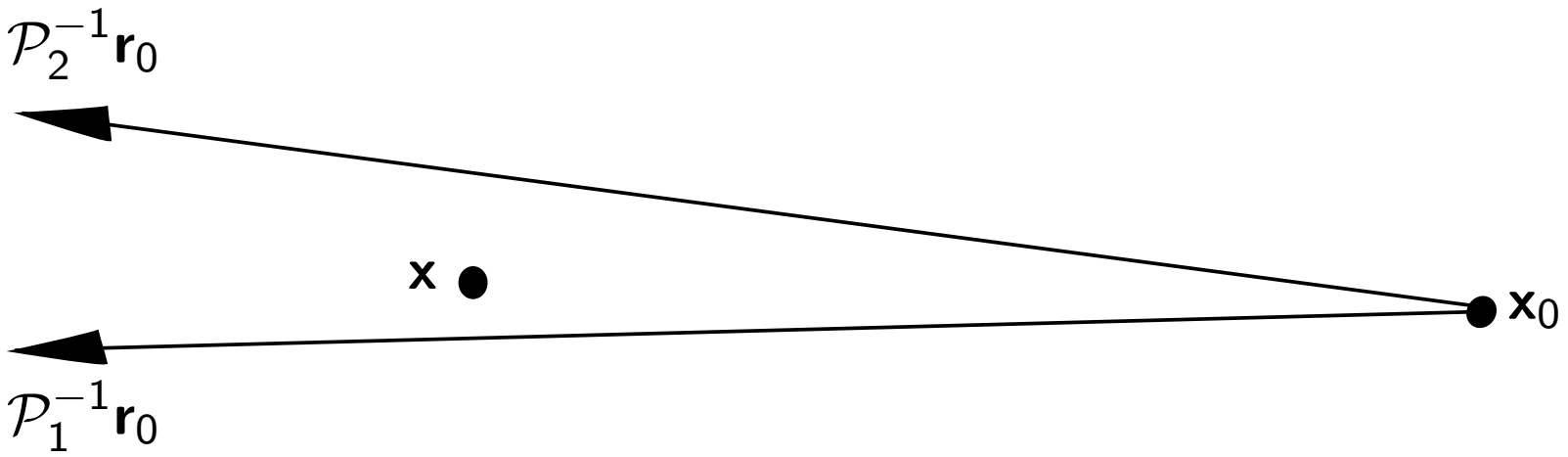
$$\begin{aligned}
\|\mathbf{b} - \mathcal{A}\mathbf{x}_k\|_2 &= \|\mathbf{b} - \mathcal{A}(\mathbf{x}_0 + [Z_1 \cdots Z_k]\mathbf{y}_k)\|_2 \\
&= \|\mathbf{r}_0 - \mathcal{A}[Z_1 \cdots Z_k]\mathbf{y}_k\|_2 \\
&= \|\mathbf{r}_0 - [V_1 \cdots V_{k+1}]\widetilde{H}_k \mathbf{y}_k\|_2 \\
&= \|V_1\|\mathbf{r}_0\|_2 - [V_1 \cdots V_{k+1}]\widetilde{H}_k \mathbf{y}_k\|_2 \\
&= \|[V_1 \cdots V_{k+1}](\|\mathbf{r}_0\|_2 \mathbf{e}_1 - \widetilde{H}_k \mathbf{y}_k)\|_2 \\
&= \|\|\mathbf{r}_0\|_2 \mathbf{e}_1 - \widetilde{H}_k \mathbf{y}_k\|_2
\end{aligned}$$

Diagrams

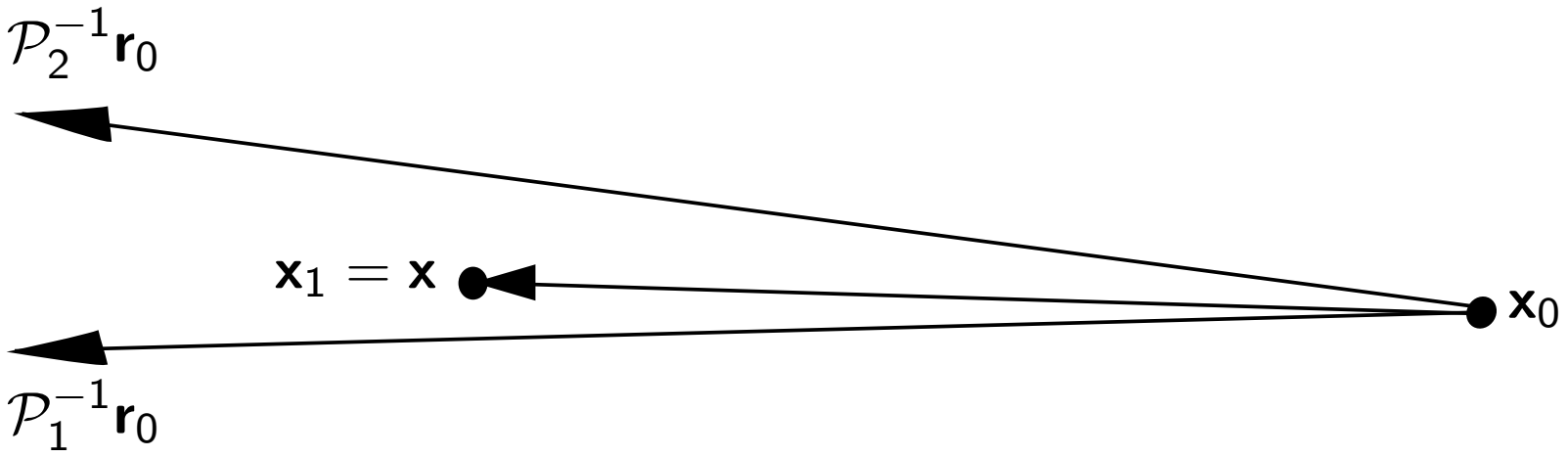

$\mathbf{x}$
$\mathbf{x}_0$
$\mathbf{x}_1 = \mathbf{x}_0 + \alpha \mathcal{P}_1^{-1} \mathbf{r}_0$
s.t.
$\|\mathbf{b} - \mathcal{A}\mathbf{x}_1\|_2 = \min$


$\mathbf{x}_1 = \mathbf{x}_0 + \alpha \mathcal{P}_2^{-1} \mathbf{r}_0$
$\mathbf{x}$
$\mathbf{x}_0$

Diagrams


$\mathcal{P}_2^{-1}\mathbf{r}_0$
$\mathbf{x}$
$\mathbf{x}_0$
$\mathcal{P}_1^{-1}\mathbf{r}_0$

Diagrams


$\mathcal{P}_2^{-1}\mathbf{r}_0$
$\mathbf{x}_1 = \mathbf{x}$
$\mathbf{x}_0$
$\mathcal{P}_1^{-1}\mathbf{r}_0$

Comparison of costs

Costs at the kth iteration:

	Mat.-vec. prods	inner products	pre. solves
MPGMRES	t^k	$\frac{t^k-1}{t-1} + \frac{t^{2(k-1)}+3t^{k-1}}{2}$	t^k
tMPGMRES	t	$(k-\frac{3}{2})t^2 + \frac{5}{2}t$	t
GMRES	1	k+1	1

Comparison of costs

Costs at the kth iteration:

	Mat.-vec. prods	inner products	pre. solves
MPGMRES	t^k	$\frac{t^k-1}{t-1} + \frac{t^{2(k-1)}+3t^{k-1}}{2}$	t^k
tMPGMRES	t	$(k-\frac{3}{2})t^2 + \frac{5}{2}t$	t
GMRES	1	k+1	1

Easily parallelized

Characterizing the search space

GMRES preconditioned with a right-preconditioner $\mathcal{P}$ finds the vector that minimizes the 2-norm of the residual over all vectors of the form

$$\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + \mathcal{P}^{-1}\mathbf{y}_k,$$

where $\mathbf{y}_k$ is a member of the Krylov subspace

$$\mathcal{K}_k(\mathcal{A}\mathcal{P}^{-1}, \mathbf{r}^{(0)}) = \mathrm{span}(\mathbf{r}^{(0)}, \mathcal{A}\mathcal{P}^{-1}\mathbf{r}^{(0)}, \ldots, (\mathcal{A}\mathcal{P}^{-1})^{k-1}\mathbf{r}^{(0)}).$$

Characterizing the search space

GMRES preconditioned with a right-preconditioner $\mathcal{P}$ finds the vector that minimizes the 2-norm of the residual over all vectors of the form

$$\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + \mathcal{P}^{-1}\mathbf{y}_k,$$

where $\mathbf{y}_k$ is a member of the Krylov subspace

$$\mathcal{K}_k(\mathcal{A}\mathcal{P}^{-1}, \mathbf{r}^{(0)}) = \mathrm{span}(\mathbf{r}^{(0)}, \mathcal{A}\mathcal{P}^{-1}\mathbf{r}^{(0)}, \ldots, (\mathcal{A}\mathcal{P}^{-1})^{k-1}\mathbf{r}^{(0)}).$$

The extension for MPGMRES: the first two iterates satisfy

$$\mathbf{x}^{(1)} - \mathbf{x}^{(0)} \in \mathrm{span}\{\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_2^{-1}\mathbf{r}^{(0)}\}$$
$$\mathbf{x}^{(2)} - \mathbf{x}^{(0)} \in \mathrm{span}\{\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_2^{-1}\mathbf{r}^{(0)}, \mathcal{P}_1^{-1}\mathcal{A}\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_1^{-1}\mathcal{A}\mathcal{P}_2^{-1}\mathbf{r}^{(0)}, \mathcal{P}_2^{-1}\mathcal{A}\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_2^{-1}\mathcal{A}\mathcal{P}_2^{-1}\mathbf{r}^{(0)}\},$$

and the rest follow the same pattern.

Breakdowns

All breakdowns in standard GMRES are 'lucky'.

This is not the case with MPGMRES....

e.g. if $\mathcal{P}_1 = \mathcal{P}_2$, the matrix $Z_1 = [\mathcal{P}_1^{-1}\mathbf{r}^{(0)}\ \mathcal{P}_2^{-1}\mathbf{r}^{(0)}]$ will be of rank one

Breakdowns

All breakdowns in standard GMRES are 'lucky'.

This is not the case with MPGMRES....

e.g. if $\mathcal{P}_1 = \mathcal{P}_2$, the matrix $Z_1 = [\mathcal{P}_1^{-1}\mathbf{r}^{(0)}\ \mathcal{P}_2^{-1}\mathbf{r}^{(0)}]$ will be of rank one

...but not a problem in general – we can monitor the subdiagonal entries of the upper Hessenberg matrix:

- if 0 on subdiagonal, and not converged – must be a linearly dependent vector: discard and all is fine
- if 0 on the subdiagonal, and residual small enough – lucky breakdown!
- if no zero on subdiagonal, no problem!

Examples

Demonstration of the relative richness of the search space

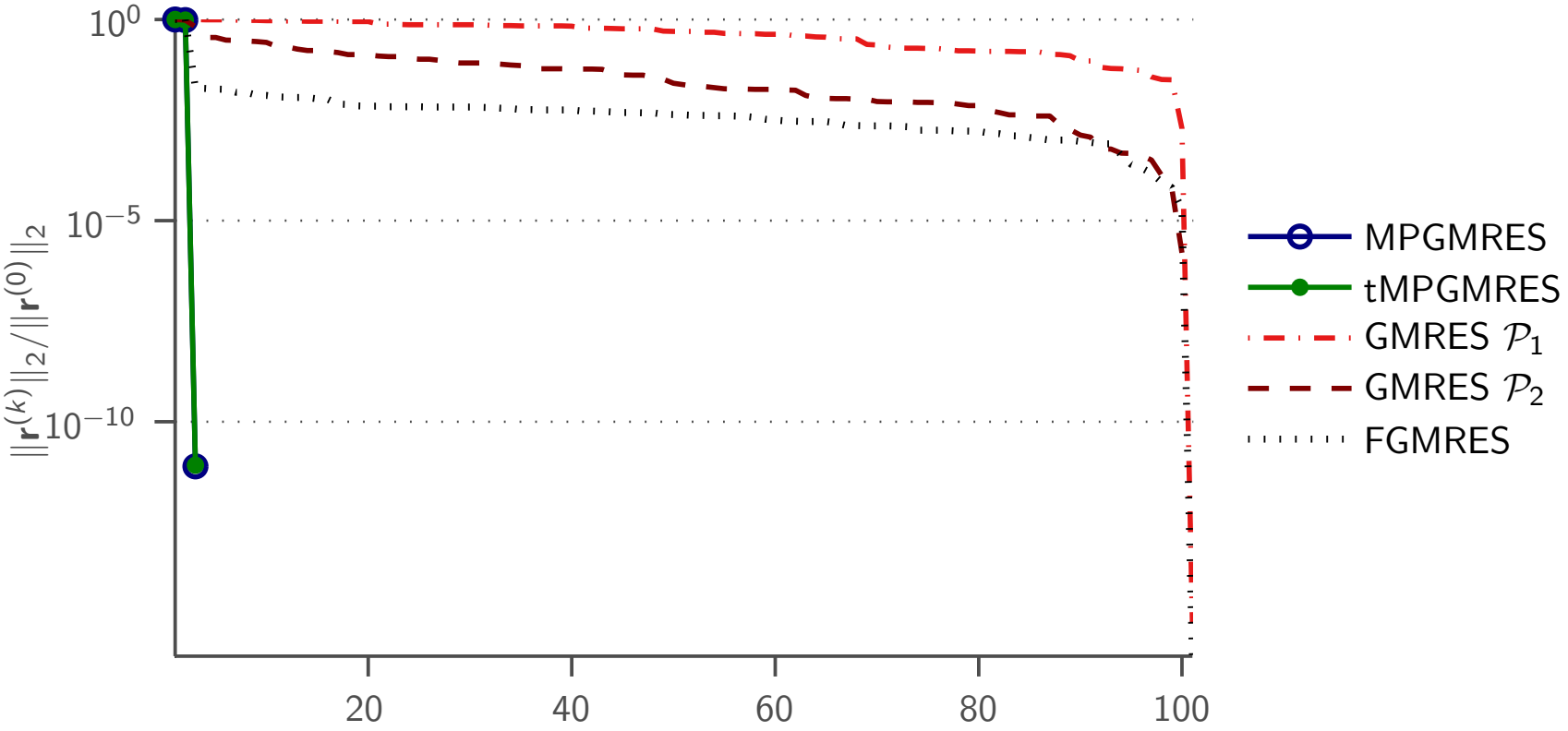
Given $\mathcal{P}_1$ and $\mathcal{P}_2$, take $\mathbf{x}^{(0)} = \mathbf{0}$. Then

$$\mathcal{P}_1^{-1}\mathcal{A}\mathcal{P}_2^{-1}\mathbf{b}$$

lies in the search space after two iterations.

Therefore if $\mathbf{b}$ is an eigenvector of $\mathcal{A}\mathcal{P}_1^{-1}\mathcal{A}\mathcal{P}_2^{-1}$, MPGMRES will converge after two iterations.

Two iterations



Domain decomposition

Consider the advection-diffusion equation on $\Omega = [0,1]^2$:

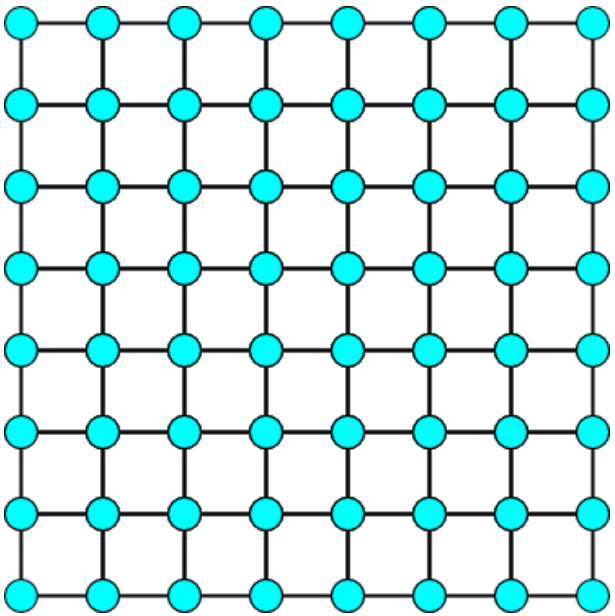
$$\begin{aligned} -\nabla^2 u + \boldsymbol{\omega} \cdot \nabla u &= f \qquad \text{in } \Omega \\ u &= 0 \qquad \text{on } \partial\Omega. \end{aligned}$$

Upon discretization by finite differences we get the matrix equation

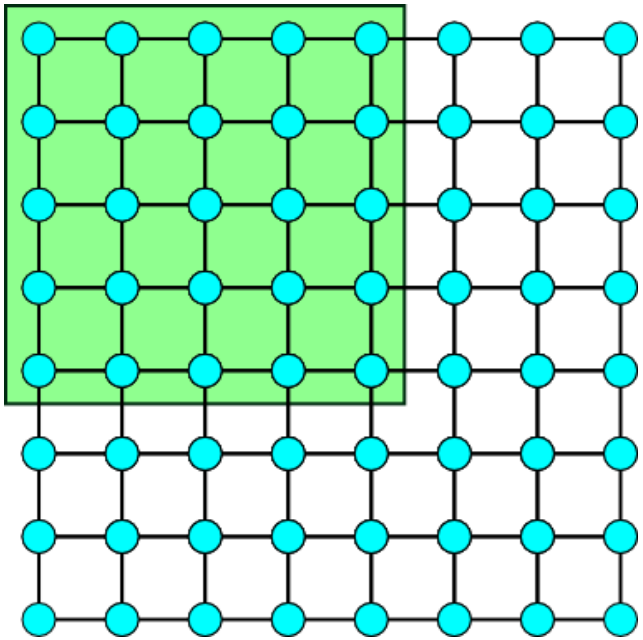
$$A\mathbf{u} = \mathbf{b},$$

where A is a real positive, but nonsymmetric, matrix.

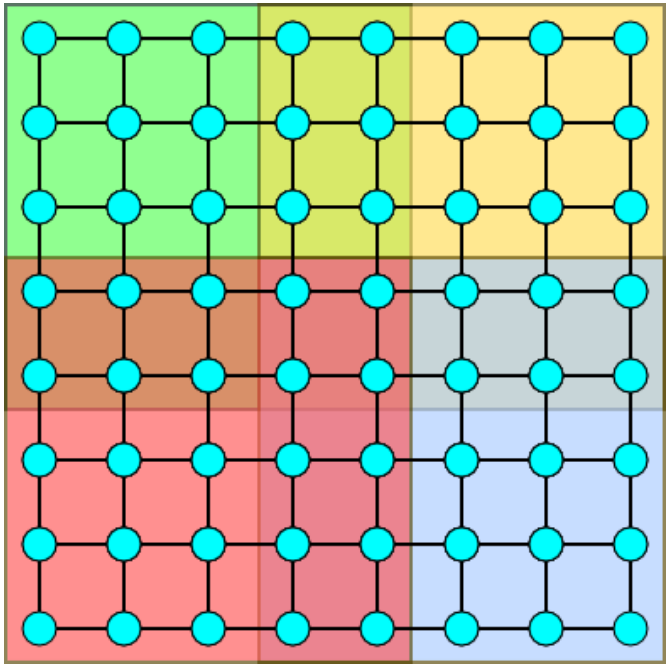
Domain decomposition (cont.)



Domain decomposition (cont.)



Domain decomposition (cont.)



$$A^{(i)} = R_{i,\delta} A R_{i,\delta}^T, i = 1, ..., 4,$$

where $R_{i,\delta}$ is a restriction matrix, and δ denotes the number of nodes overlapping ($\delta = 1$ here).

Additive Schwarz preconditioner

The additive Schwarz preconditioner has its inverse defined as

$$M^{-1} = \sum_{i=1}^{t} R_{i,\delta}^T (A^{(i)})^{-1} R_{i,\delta}.$$

Additive Schwarz preconditioner

The additive Schwarz preconditioner has its inverse defined as

$$M^{-1} = \sum_{i=1}^{t} R_{i,\delta}^T (A^{(i)})^{-1} R_{i,\delta}.$$

Well suited to a multi-preconditioned approach: take each solve on a subdomain as a preconditioner.

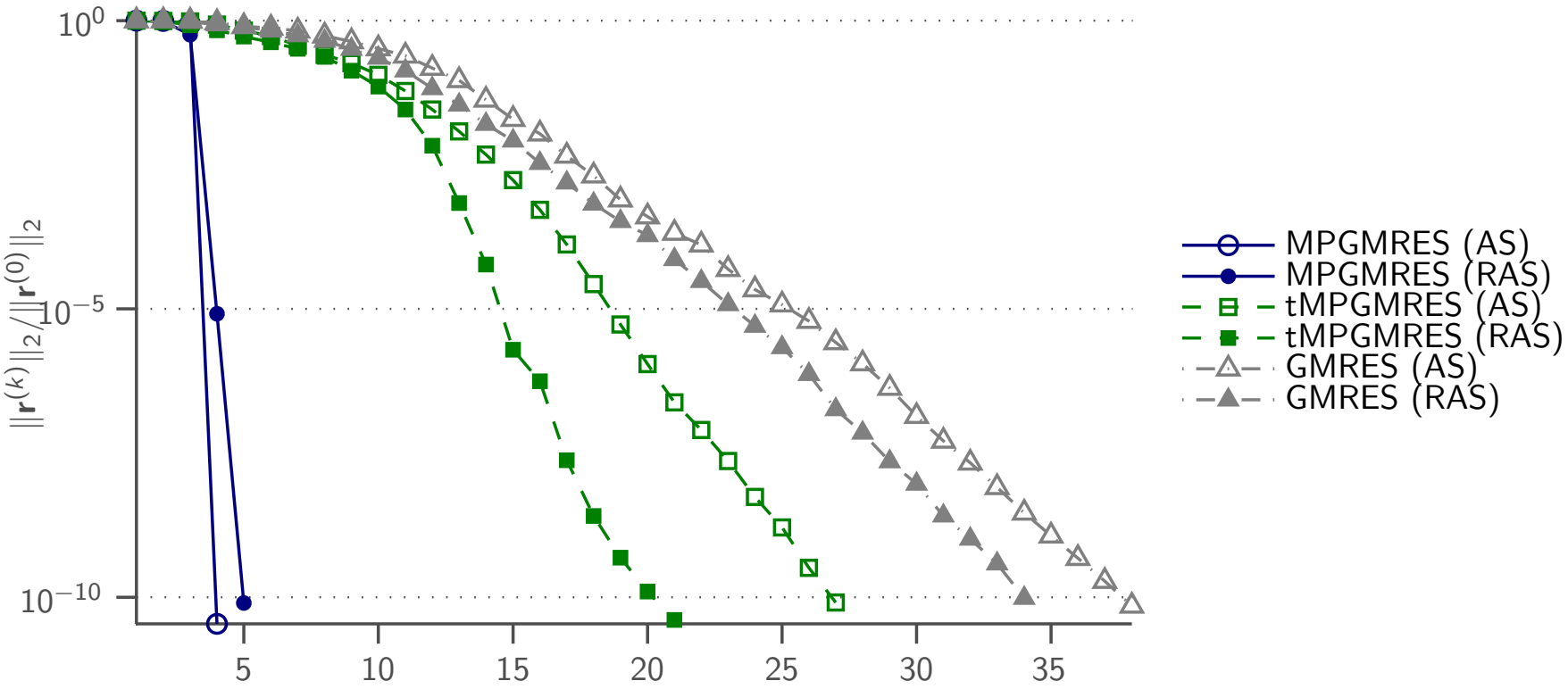
Restricted Additive Schwarz

The restricted Additive Schwarz preconditioner is defined as

$$M^{-1} = \sum_{i=1}^{t} R_{i,0}^T (R_{i,\delta} A R_{i,\delta}^T)^{-1} R_{i,\delta}.$$

This has the effect of removing the overlap in the preconditioner, hence improving convergence.

Also an ideal candidate for multi-preconditioning.


10^0
10^{-5}
10^{-10}
$\|\mathbf{r}^{(k)}\|_2/\|\mathbf{r}^{(0)}\|_2$
5
10
15
20
25
30
35
MPGMRES (AS)
MPGMRES (RAS)
tMPGMRES (AS)
tMPGMRES (RAS)
GMRES (AS)
GMRES (RAS)

Multigrid

Standard multigrid has two components:

$$\text{Smoother}: \quad \mathbf{x}^{(i+1)} \to \mathbf{x}^{(i)} + M\mathbf{r}^{(i)},$$

$$\text{Coarse}-\text{grid correction}: \quad \mathbf{x}^{(i+1)} \to \mathbf{x}^{(i)} + PA_C^{-1}R\mathbf{r}^{(i)}.$$

Standard (multiplicative) multigrid

$$\mathbf{x}^{(i+1/2)} = \mathbf{x}^{(i)} + \alpha_1 M\mathbf{r}^{(i)}$$
$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i+1/2)} + \alpha_2 PA_C^{-1} R\mathbf{r}^{(i+1/2)}$$

Taken together:

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} + \alpha_1 M\mathbf{r}^{(i)} + \alpha_2 PA_C^{-1} R\mathbf{r}^{(i)} - \alpha_1\alpha_2 PA_C^{-1} R\, AM\mathbf{r}^{(i)}$$

Standard (multiplicative) multigrid

$$\mathbf{x}^{(i+1/2)} = \mathbf{x}^{(i)} + \alpha_1 M\mathbf{r}^{(i)}$$
$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i+1/2)} + \alpha_2 PA_C^{-1}R\mathbf{r}^{(i+1/2)}$$

Taken together:

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} + \alpha_1 M\mathbf{r}^{(i)} + \alpha_2 PA_C^{-1}R\mathbf{r}^{(i)} - \alpha_1\alpha_2 PA_C^{-1}R\,AM\mathbf{r}^{(i)}$$

Fixed parameters at each step

Parallelizable (additive) multigrid

There have been attempts at a parallelizable MG routine – e.g. the BPX multigrid. These essentially take

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} + \alpha_1 M\mathbf{r}^{(i)} + \alpha_2 P A_C^{-1} R\mathbf{r}^{(i)}$$

for some choice of α_1, α_2.

Multi-preconditioned multigrid

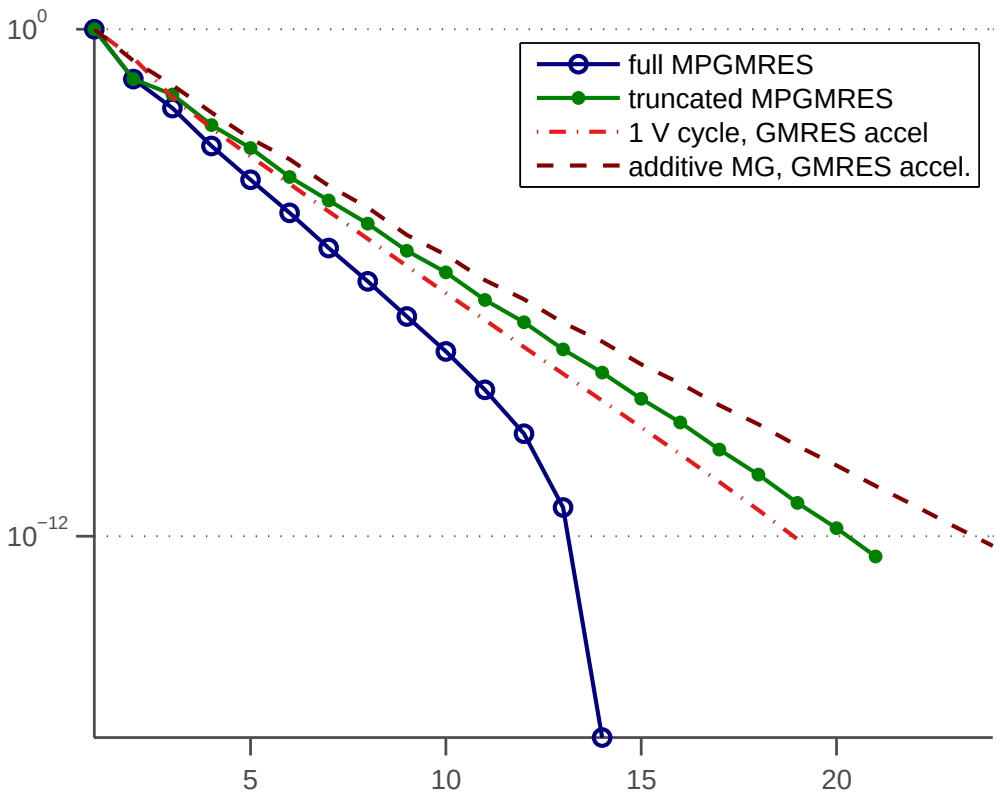
Lends itself nicely to MPGMRES:

$$\mathcal{P}_1 = M \qquad \text{(smoother)}$$
$$\mathcal{P}_2 = PA_C^{-1}R \qquad \text{(coase} - \text{grid correction)}$$

After two iterations, MPGMRES finds the vector which minimizes the residual over the space

$$\mathbf{x}^{(0)} + \mathrm{span}\{M\mathbf{r}^{(0)}, PA_C^{-1}R\mathbf{r}^{(0)}, MA\,M\mathbf{r}^{(0)}, MA\,PA_C^{-1}R\mathbf{r}^{(0)}, PA_C^{-1}R\,AM\mathbf{r}^{(0)}\}$$

Preliminary results



MPGMRES is an extension of the standard preconditioned GMRES which allows us to use more than one preconditioner.
The method:

- seems to work well when we have non-ideal preconditioners which complement each other
- can handle any number of candidate preconditioners
- can be parallelized, obtaining potential computational gains

Paper and MATLAB code available at

`www.cs.ubc.ca/~tyronere/`

Fortran 95 code (`HSL_MI28`) under development