

Provable Convergence rate for Asynchronous methods via Randomized Gauss-Seidel

Daniel B. Szyld

Temple University, Philadelphia

ICERM Workshop on
Asynchronous Methods for Numerical Linear Algebra
5 May 2026

Joint work with Andreas Frommer (Wuppertal)

Background. Classical iterative methods.

We want to solve the $n \times n$ linear system

$$Ax = b$$

$$\sum_{j=1}^n a_{ij}x_j = b_i, \quad i = 1, \dots, n$$

Background. Classical iterative methods.

We want to solve the $n \times n$ linear system

$$Ax = b$$

$$\sum_{j=1}^n a_{ij}x_j = b_i, \quad i = 1, \dots, n$$

A relaxation

$$x_i = \sum_{j=1, j \neq i}^n -\frac{a_{ij}}{a_{ii}} x_j + \frac{b_i}{a_{ii}}$$

Background. Classical iterative methods.

We want to solve the $n \times n$ linear system

$$Ax = b$$

$$\sum_{j=1}^n a_{ij}x_j = b_i, \quad i = 1, \dots, n$$

A relaxation

$$x_i = \sum_{j=1, j \neq i}^n -\frac{a_{ij}}{a_{ii}} x_j + \frac{b_i}{a_{ii}}$$

or: Solve for x_i in

$$a_{ii}x_i = b_i - \sum_{j=1, j \neq i}^n a_{ij} x_j$$

- ▶ Fixed point iteration
- ▶ Jacobi Method follows naturally. n relaxations at a time
- ▶ Gauss-Seidel uses most recent information in the rhs

More background: The block case

Consider now A permuted and partitioned into $q \times q$ blocks:

$$\mathbf{S}^T \mathbf{A} \mathbf{S} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \vdots & & \ddots & \vdots \\ A_{q1} & A_{q2} & \cdots & A_{qq} \end{bmatrix},$$

where $\mathbf{S} = [S_1, S_2, \dots, S_q]$ is the permutation matrix.

Thus $A_{ij} = S_i^T A S_j$.

We permute partition \mathbf{x} and \mathbf{b} accordingly: $\mathbf{x}_i = S_i^T \mathbf{x}$, $\mathbf{b}_i = S_i^T \mathbf{b}$.

Thus, one block equation looks like:

Solve for \mathbf{x}_i in

$$A_{ii} \mathbf{x}_i = \mathbf{b}_i - \sum_{j=1, j \neq i}^q A_{ij} \mathbf{x}_j$$

Block Algorithms

Algorithm 1 Classical block Jacobi iteration for a fixed partition.

for $m = 1, 2, \dots$ **and** $i = 1, \dots, q$ **do**
 Solve $A_{ii}\mathbf{x}_i^{m+1} = \mathbf{b}_i - \sum_{j=1, j \neq i}^q A_{ij}\mathbf{x}_j^m$
end for

We are interested in Asynchronous Iterations.

[What are asynchronous iterations?](#)

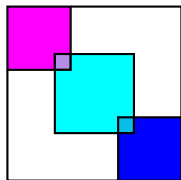
No iteration counts, just block relaxations

Algorithm 2 Asynchronous block Jacobi (or Block Gauss-Seidel) iteration for a fixed partition. Processor i solves local problem i

for $i = 1, \dots, q$ in parallel **do**
 read $\mathbf{x}_j, j \neq i$ from the other processors
 Solve $A_{ii}\mathbf{x}_i = \mathbf{b}_i - \sum_{j=1, j \neq i}^q A_{ij}\mathbf{x}_j$
 write \mathbf{x}_i
end for

Better method, use overlapping variables

- ▶ Instead of a partition of $\{1, 2, \dots, n\} = \cup_{i=1, \dots, q} \Omega_i$ with $\Omega_i \cap \Omega_j = \emptyset, i \neq j$, use overlapping sets Ω_i .
- ▶ Having overlapping variables is the foundation of the success of Schwarz methods (Domain Decomposition methods)¹.
- ▶ Additive/multiplicative Schwarz can be interpreted as Block Jacobi/Gauss-Seidel **with overlap**.
- ▶ Restricted Additive Schwarz (RAS): compute with overlap, communicate without overlap²

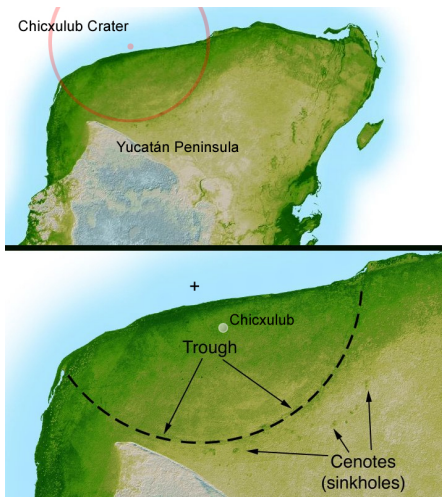


¹see ddm.org

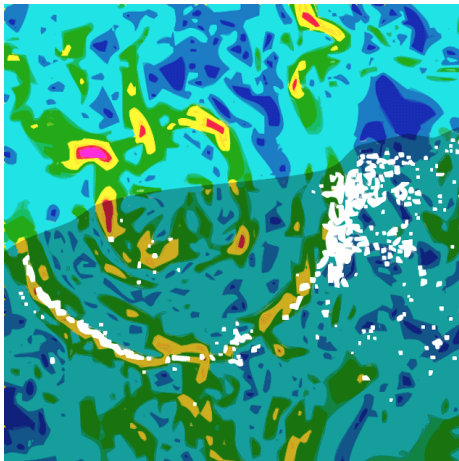
²[Cai, Sarkis, 1999], [Frommer, S, 2001]

A numerical experiment. Asynchronous Schwarz

Chicxulub Crater, created by a collision of an asteroid approx. 66 million years ago: Cretaceous-Paleogen boundary: extinction of dinosaurs, approx. diameter 180km (pictures NASA, 2010)



Artistic rendering of the gravity anomaly map (NASA, 2010). Different colors represent different gravity measurements, except the white dots, which are sinkholes called cenotes.

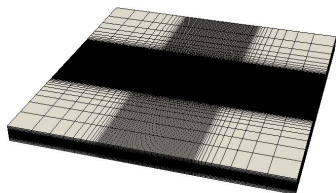


A cenote in Yucatan



Our experiments

We want to compute the gravitational potential Φ on a parallelepiped geometric domain of dimensions $250\text{km} \times 250\text{km} \times 15\text{km}$. Finite element mesh



Equation to solve: $\Delta\Phi = -4\pi G\delta\rho$

- ▶ $G = 6.672 \times 10^{-11} \text{m}^3 \text{kg}^{-1} \text{s}^{-2}$ gravitational constant
- ▶ $\delta\rho$ anomaly density distribution computed from data acquisition on a salt dome (produced by the impact)

Three discretizations of box

- ▶ case I has 2 491 632 DOF (256 subdomains)
- ▶ case II has 19 933 056 DOF (512 subdomains)
- ▶ case III has 146 707 292 DOF (1024 subdomains)
- ▶ 1068 processors - 17,088 cores (half 1.6 Ghz with 2x2MB of cache, half 2.93 Ghz with 2x4MB of cache)
- ▶ In each subdomain solve linear system directly³

³F. Magoulès, DBS, C. Venet, Asynchronous Optimized Schwarz Methods with and without Overlap. *Numerische Mathematik*, **137** (2017) 199–227.

Three discretizations of box

- ▶ case I has 2 491 632 DOF (256 subdomains)
- ▶ case II has 19 933 056 DOF (512 subdomains)
- ▶ case III has 146 707 292 DOF (1024 subdomains)
- ▶ 1068 processors - 17,088 cores (half 1.6 Ghz with 2x2MB of cache, half 2.93 Ghz with 2x4MB of cache)
- ▶ In each subdomain solve linear system directly³

OOO – case	iter	time	upt min	upt max	time
I (256)	1722	43	1030	1917	40
II (512)	3379	777	2257	4438	591
III (1024)	8331	3888	5251	13274	863

³F. Magoulès, DBS, C. Venet, Asynchronous Optimized Schwarz Methods with and without Overlap. *Numerische Mathematik*, **137** (2017) 199–227.

Three discretizations of box

- ▶ case I has 2 491 632 DOF (256 subdomains)
- ▶ case II has 19 933 056 DOF (512 subdomains)
- ▶ case III has 146 707 292 DOF (1024 subdomains)
- ▶ 1068 processors - 17,088 cores (half 1.6 Ghz with 2x2MB of cache, half 2.93 Ghz with 2x4MB of cache)
- ▶ In each subdomain solve linear system directly³

OOO – case	iter	time	upt min	upt max	time
I (256)	1722	43	1030	1917	40
II (512)	3379	777	2257	4438	591
III (1024)	8331	3888	5251	13274	863

- ▶ Asynchronous is 22% of the time

³F. Magoulès, DBS, C. Venet, Asynchronous Optimized Schwarz Methods with and without Overlap. *Numerische Mathematik*, **137** (2017) 199–227.

Three discretizations of box

- ▶ case I has 2 491 632 DOF (256 subdomains)
- ▶ case II has 19 933 056 DOF (512 subdomains)
- ▶ case III has 146 707 292 DOF (1024 subdomains)
- ▶ 1068 processors - 17,088 cores (half 1.6 Ghz with 2x2MB of cache, half 2.93 Ghz with 2x4MB of cache)
- ▶ In each subdomain solve linear system directly³

OOO – case	iter	time	upt min	upt max	time
I (256)	1722	43	1030	1917	40
II (512)	3379	777	2257	4438	591
III (1024)	8331	3888	5251	13274	863

- ▶ Asynchronous is 22% of the time
- ▶ More experiments including coarse grid correction in [C. Glusa, E. Boman, E. Chow, S. Rajamanickam, and DBS, SISC **42** (2020) C384–C409]

³F. Magoulès, DBS, C. Venet, Asynchronous Optimized Schwarz Methods with and without Overlap. *Numerische Mathematik*, **137** (2017) 199–227.

Asynchronous parallel methods for fixed point problems

Long history mostly from the 1980's and 1990's

Very selected references:

Papers: [Chazan, Miranker, 1969], [Robert, 1976], [Baudet, 1978], [El Tarazi, 1982], [Bertsekas, 1983], [Üresin, Dubois, 1989]

Books: [Bertsekas, Tsitsiklis, 1989], [Bahi, Contassot-Vivier, Couturier, 2008]

Surveys: [Frommer, S., 2000] [Spiteri, 2020]

Asynchronous parallel methods for fixed point problems

Long history mostly from the 1980's and 1990's

Very selected references:

Papers: [Chazan, Miranker, 1969], [Robert, 1976], [Baudet, 1978], [El Tarazi, 1982], [Bertsekas, 1983], [Üresin, Dubois, 1989]

Books: [Bertsekas, Tsitsiklis, 1989], [Bahi, Contassot-Vivier, Couturier, 2008]

Surveys: [Frommer, S., 2000] [Spiteri, 2020]

Key result: For a fixed point iteration $u(n+1) = \mathcal{T}u(n)$, if $\|\mathcal{T}\| < 1$, and some other (natural) conditions, asynchronous iteration converges to the unique fixed point.

All theorems show asynchronous method converge, but results do not give a convergence rate.

Asynchronous parallel methods for fixed point problems

Long history mostly from the 1980's and 1990's

Very selected references:

Papers: [Chazan, Miranker, 1969], [Robert, 1976], [Baudet, 1978], [El Tarazi, 1982], [Bertsekas, 1983], [Üresin, Dubois, 1989]

Books: [Bertsekas, Tsitsiklis, 1989], [Bahi, Contassot-Vivier, Couturier, 2008]

Surveys: [Frommer, S., 2000] [Spiteri, 2020]

Key result: For a fixed point iteration $u(n+1) = \mathcal{T}u(n)$, if $\|\mathcal{T}\| < 1$, and some other (natural) conditions, asynchronous iteration converges to the unique fixed point.

All theorems show asynchronous method converge, but results do not give a convergence rate.

Our goal: Provide a convergence rate for these methods

First step: Randomized (point) Gauss-Seidel

First, sequential Gauss-Seidel:

We had: Solve for x_i in

$$a_{ii}x_i = b_i - \sum_{j=1, j \neq i}^n a_{ij} x_j$$

$$A = B - D, H = B^{-1}C, c = B^{-1}b$$

Algorithm 3 Sequential relaxation for $Ax = b$, with initial approximation x^0 (“Gauss-Seidel” if $H = D^{-1}B$)

for $k = 1, 2, \dots$ until a convergence criterion is satisfied **do**

$i = k - n \lfloor (k-1)/n \rfloor$ $\{i \in \{1, \dots, n\} \text{ with } i \equiv k \pmod{n}\}$

$x_i^{k+1} = \sum_{j=1}^n h_{ij}x_j^k + c_i, \quad x_\ell^{k+1} = x_\ell^k \text{ for } \ell \neq i$

end for

n relaxations = one classical iteration

Randomized Gauss-Seidel

If the index i to select which row to relax is chosen randomly, with probability p_i , $0 < p_i < 1$, $\sum_{i=1}^n p_i = 1$, we call this algorithm **randomized Gauss-Seidel**

Algorithm 4 Randomized iteration for $Ax = b$ with initial approximation x^0 (“randomized Gauss-Seidel” if $H = D^{-1}B$)

for $k = 1, 2, \dots$ **do**

 choose index i with probability p_i

$$x_i^{k+1} = \sum_{j=1}^n h_{ij}x_j^k + c_i, \quad x_\ell^{k+1} = x_\ell^k \text{ for } \ell \neq i$$

end for

Analyzed by [Leventthal and Lewis, 2010] and [Griebel and Oswald, 2012] for A Hermitian positive definite.

Gauss-Southwell

Gauss had another iteration scheme in mind, popularized by Southwell [see a nice historical recount by Y. Saad, 2020]. Instead of going cyclically $i = 1, 2, \dots, n$, choose to relax the component of maximum residual. That is, with $r^k = b - Ax^k$, choose i so that

$$|r_i^k| = \max_{j=1}^n |r_j^k|.$$

Gauss-Southwell

Gauss had another iteration scheme in mind, popularized by Southwell [see a nice historical recount by Y. Saad, 2020]. Instead of going cyclically $i = 1, 2, \dots, n$, choose to relax the component of maximum residual. That is, with $r^k = b - Ax^k$, choose i so that

$$|r_i^k| = \max_{j=1}^n |r_j^k|.$$

It is a greedy algorithm. Other options include:
Fix weights $\beta_i > 0$ and choose i such that

$$\beta_i |r_i^k| = \max_{j=1}^n \beta_j |r_j^k|.$$

Gauss-Southwell Algorithm

Algorithm 5 Greedy relaxation for $Ax = b$, with initial approximation x^0 (“Gauss-Southwell” if $H = D^{-1}B$)

fix a greedy pick rule

for $k = 1, 2, \dots$ **do**

 determine an index i according to the greedy pick rule

$$x_i^{k+1} = \sum_{j=1}^n h_{ij} x_j^k + c_i, \quad x_\ell^{k+1} = x_\ell^k \text{ for } \ell \neq i$$

end for

Converges in many fewer relaxations (iterations), but the max in the greedy rule makes each relaxation more expensive.

A non-Hermitian

We consider general splittings $A = M - N$ and iteration matrix $H = M^{-1}N$.

Def. For $u \in \mathbb{R}^n$ with $u_i > 0, \forall i$, the weighted ℓ_1 -norm on \mathbb{C}^n is

$$\|x\|_{u,1} = \sum_{j=1}^n u_j |x_j|$$

The induced operator norm for $A \in \mathbb{C}^{n \times n}$ is the weighted column sum-norm

$$\|A\|_{u,1} = \max_{j=1}^n \frac{1}{u_j} \sum_{i=1}^n u_i |a_{ij}|$$

“Preconditioned” residual is

$$\hat{r}^k = c - (I - H)x^k = M^{-1}(b - Ax^k) = M^{-1}A(x^* - x^k)$$

Our results in the point case

We prove⁴ that **in expected value sense** and using a **weighted ℓ_1 -norm**, the bound on the convergence factor for the preconditioner residual for Randomized Gauss-Seidel is the same as for Gauss-Southwell, but without the extra cost per relaxation.

⁴A. Frommer and DBS, On the convergence of randomized and greedy relaxation schemes for solving nonsingular linear systems of equations, *Numer. Algo.* **92** (2023) 639–664.

Numerical Examples

Convection-diffusion equation for a concentration

$$c = c(x, y, t) : [0, 1] \times [0, 1] \times [0, T] \rightarrow \mathbb{C}$$

$$\begin{aligned} \frac{\partial}{\partial t} c &= -\frac{\partial}{\partial x} \alpha \frac{\partial}{\partial x} c - \frac{\partial}{\partial y} \beta \frac{\partial}{\partial y} c + \frac{\partial}{\partial x} \nu c + \frac{\partial}{\partial y} \mu c, \\ c(x, y, t) &= 0 \text{ if } (x, y) \in \partial([0, 1] \times [0, 1]), \\ c(x, y, 0) &= c_0(x, y) \text{ for } (x, y) \in [0, 1] \times [0, 1]. \end{aligned}$$

$\alpha = \alpha(x, y) > 0$, $\beta = \beta(x, y) > 0$, and the velocity field $(\nu, \mu) = (\nu(x, y), \mu(x, y))$. Discretize in space using standard finite differences with N interior equispaced grid points in each direction.

Numerical Examples (cont.)

Obtain the semi-discretized system

$$\frac{\partial}{\partial t} c = Bc, \quad c(x, y, 0) = c_0(x, y) \text{ for all grid points } x, y,$$

where now $c = c(t)$ is a two-dimensional array, each component corresponding to one grid point. Using the implicit Euler rule as a symplectic integrator means that at a given time t and a stepsize τ we have to solve

$$\underbrace{\left(I + \frac{\tau}{2} B\right)}_{=: A} c(t+1) = \frac{\tau}{2} Bc(t)$$

for $c(t+1)$.

Numerical Examples A non-HPD

We use non-vanishing and non-constant velocity field describing a re-circulating flow,

$$(\nu(x, y), \mu(x, y)) = \sigma(4x(x-1)(1-2y), -4y(y-1)(1-2x)).$$

Two cases:

$\sigma = 1$ (weak convection) and $\sigma = 400$ (strong convection).

The diffusion coefficients are $\alpha = \beta = 1$.

Numerical Examples A non-HPD

We use non-vanishing and non-constant velocity field describing a re-circulating flow,

$$(\nu(x, y), \mu(x, y)) = \sigma(4x(x-1)(1-2y), -4y(y-1)(1-2x)).$$

Two cases:

$\sigma = 1$ (weak convection) and $\sigma = 400$ (strong convection).

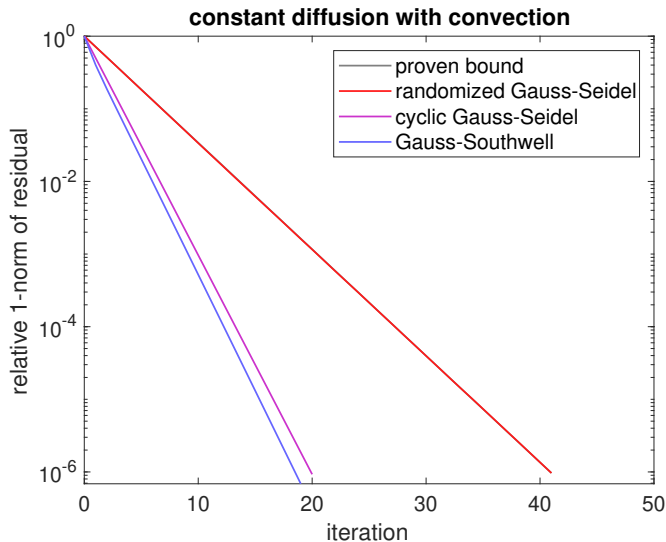
The diffusion coefficients are $\alpha = \beta = 1$.

Here, A is an H -matrix, and we use the following greedy rule for Gauss-Southwell giving “optimal” convergence factor

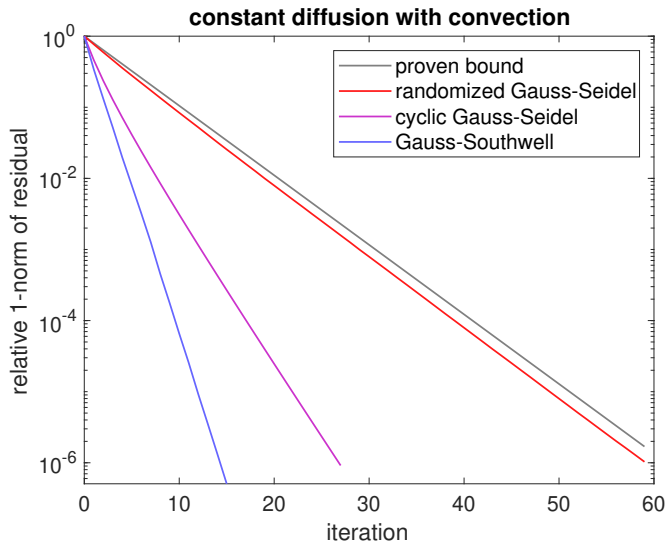
$$(1 - \rho_i) \frac{u_j}{|a_{ij}|} |r_i^k| = \max_{j=1}^n (1 - \rho_j) \frac{u_j}{|a_{jj}|} |r_j^k|$$

with $u > 0$ such that $u^T A > 0$.

Numerical example with weak convection ($\sigma = 1$)



Numerical example with strong convection ($\sigma = 400$)



Second step: The block case

But we want to consider overlapping variables, i.e., Multiplicative Schwarz

To that end, let the index set $\Omega = \{1, \dots, n\}$ and q non-empty subsets $\Omega_i \subset \Omega$, $i = 1, \dots, q$. These correspond to DoF in domain i . Let $n_i = |\Omega_i|$, the cardinality of Ω_i . We want $\cup_{i=1}^q \Omega_i = \Omega$, and thus $\sum_{i=1}^q n_i \geq n$.

Second step: The block case

But we want to consider overlapping variables, i.e., Multiplicative Schwarz

To that end, let the index set $\Omega = \{1, \dots, n\}$ and q non-empty subsets $\Omega_i \subset \Omega$, $i = 1, \dots, q$. These correspond to DoF in domain i . Let $n_i = |\Omega_i|$, the cardinality of Ω_i . We want $\cup_{i=1}^q \Omega_i = \Omega$, and thus $\sum_{i=1}^q n_i \geq n$.

We also define Ω_{-i} , the complement of Ω_i , i.e., $\Omega_{-i} := \Omega \setminus \Omega_i$, so that $|\Omega_{-i}| = n - n_i$.

Second step: The block case

But we want to consider overlapping variables, i.e., Multiplicative Schwarz

To that end, let the index set $\Omega = \{1, \dots, n\}$ and q non-empty subsets $\Omega_i \subset \Omega$, $i = 1, \dots, q$. These correspond to DoF in domain i . Let $n_i = |\Omega_i|$, the cardinality of Ω_i . We want $\cup_{i=1}^q \Omega_i = \Omega$, and thus $\sum_{i=1}^q n_i \geq n$.

We also define Ω_{-i} , the complement of Ω_i , i.e., $\Omega_{-i} := \Omega \setminus \Omega_i$, so that $|\Omega_{-i}| = n - n_i$.

Then, the prolongation matrices $S_i \in \mathbb{R}^{n \times n_i}$ (sometimes called sketching matrices), whose columns are columns of the $n \times n$ identity matrix corresponding to the indices in Ω_i , so that for example if $\Omega_i = \{1, 3, 8\}$, then $S_i = [\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_8]$, where \mathbf{e}_j is the j th column of I . Similarly, we have $S_{-i} \in \mathbb{R}^{n \times (n - n_i)}$.

Second step: The block case

But we want to consider overlapping variables, i.e., Multiplicative Schwarz

To that end, let the index set $\Omega = \{1, \dots, n\}$ and q non-empty subsets $\Omega_i \subset \Omega$, $i = 1, \dots, q$. These correspond to DoF in domain i . Let $n_i = |\Omega_i|$, the cardinality of Ω_i . We want $\cup_{i=1}^q \Omega_i = \Omega$, and thus $\sum_{i=1}^q n_i \geq n$.

We also define Ω_{-i} , the complement of Ω_i , i.e., $\Omega_{-i} := \Omega \setminus \Omega_i$, so that $|\Omega_{-i}| = n - n_i$.

Then, the prolongation matrices $S_i \in \mathbb{R}^{n \times n_i}$ (sometimes called sketching matrices), whose columns are columns of the $n \times n$ identity matrix corresponding to the indices in Ω_i , so that for example if $\Omega_i = \{1, 3, 8\}$, then $S_i = [\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_8]$, where \mathbf{e}_j is the j th column of I . Similarly, we have $S_{-i} \in \mathbb{R}^{n \times (n - n_i)}$. Furthermore,

let $E_i = S_i S_i^T \in \mathbb{R}^{n \times n}$, and $E_{-i} = S_{-i} S_{-i}^T \in \mathbb{R}^{n \times n}$, and in fact $E_i + E_{-i} = I$.

Block Gauss-Seidel now with overlap

Algorithm 4 Block Gauss-Seidel, repeated

for $k = 1, 2, \dots$ until a convergence criterion is satisfied **do**
 $i = k - q \lfloor (k - q) / q \rfloor \quad \{i \in \{1, \dots, 1\} \text{ with } i \equiv k \pmod{q}\}$
 Solve $A_{ii} \mathbf{x}_i^{k+1} = \mathbf{b}_i - \sum_{j=1, j \neq i}^q A_{ij} \mathbf{x}_j^k$
 The other components are unchanged
end for

Block Gauss-Seidel now with overlap

Algorithm 4 Block Gauss-Seidel, repeated

for $k = 1, 2, \dots$ until a convergence criterion is satisfied **do**
 $i = k - q \lfloor (k - q) / q \rfloor$ $\{i \in \{1, \dots, 1\}$ with $i \equiv k \pmod{q}\}$
 Solve $A_{ii} \mathbf{x}_i^{k+1} = \mathbf{b}_i - \sum_{j=1, j \neq i}^q A_{ij} \mathbf{x}_j^k$
 The other components are unchanged
end for

Algorithm 5 Classical Block Gauss-Seidel iteration for a fixed covering (or partition), or Algebraic Multiplicative Schwarz

for $k = 1, 2, \dots$ until a convergence criterion is satisfied **do**
 $i = k - q \lfloor (k - q) / q \rfloor$ $\{i \in \{1, \dots, 1\}$ with $i \equiv k \pmod{q}\}$
 Solve $A_{ii} \mathbf{x}_i^{k+1} = \mathbf{b}_i - S_i^T A S_{-i} \mathbf{x}_{-i}^k$
 The other components are unchanged, i.e.,
 $\mathbf{x}^{k+1} = E_{-i} \mathbf{x}^k + S_i \mathbf{x}_i^{k+1}$
end for

Randomized Block Gauss-Seidel and Block Gauss-Southwell

Algorithm 6 Randomized Block G-S or Randomized Schwarz

for $k = 1, 2, \dots$ until a convergence criterion is satisfied **do**

 choose index i with probability p_i ($i \in \{1, \dots, q\}$)

 Solve $A_{ii}\mathbf{x}_i^{k+1} = \mathbf{b}_i - S_i^T A S_{-i} \mathbf{x}_{-i}^k$

 Other components are unchanged, i.e., $\mathbf{x}^{k+1} = E_{-i} \mathbf{x}^k + S_i \mathbf{x}_i^{k+1}$

end for

Randomized Block Gauss-Seidel and Block Gauss-Southwell

Algorithm 8 Randomized Block G-S or Randomized Schwarz

for $k = 1, 2, \dots$ until a convergence criterion is satisfied **do**

choose index i with probability p_i ($i \in \{1, \dots, q\}$)

Solve $A_{ii}\mathbf{x}_i^{k+1} = \mathbf{b}_i - S_i^T A S_{-i} \mathbf{x}_{-i}^k$

Other components are unchanged, i.e., $\mathbf{x}^{k+1} = E_{-i} \mathbf{x}^k + S_i \mathbf{x}_i^{k+1}$

end for

Algorithm 9 Greedy Block Gauss-Seidel, or Greedy Schwarz

for $k = 1, 2, \dots$ until a convergence criterion is satisfied **do**

choose index i that satisfies the greedy pick rule

Solve $A_{ii}\mathbf{x}_i^{k+1} = \mathbf{b}_i - S_i^T A S_{-i} \mathbf{x}_{-i}^k$

Other components are unchanged, i.e., $\mathbf{x}^{k+1} = E_{-i} \mathbf{x}^k + S_i \mathbf{x}_i^{k+1}$

end for

Block Convergence Results HPD Case, Assumption

Assumption. There exist positive (small) constants α_i , $i = 1, \dots, q$, so that for each $\mathbf{x} \in V = \mathbb{R}^n$, there exist vectors $\mathbf{x}_i \in V_i$ such that $\mathbf{x} = \sum_{i=1}^q \mathbf{x}_i$ and

$$\sum_{i=1}^q \frac{1}{\alpha_i} \|S_i \mathbf{x}_i\|_A^2 \leq \|\mathbf{x}\|_A^2. \quad (1)$$

Block Convergence Results HPD Case

Theorem. (i) In randomized Block Gauss-Seidel, the expected values for the squares of the norms of the errors $\mathbf{x}^k - \mathbf{x}^*$ satisfy

$$\mathbb{E} \left(\|\mathbf{x}^k - \mathbf{x}^*\|_A^2 \right) \leq (1 - \alpha^{\text{rGS}})^k \|\mathbf{x}^0 - \mathbf{x}^*\|_A^2,$$

with

$$\alpha^{\text{rGS}} = \min_{i=1}^q \frac{p_i}{\alpha_j}.$$

The factor $1 - \alpha^{\text{rGS}}$ becomes smallest if we take $p_i = \alpha_i / \sum_{j=1}^q \alpha_j$, for all i , in which case

$$\alpha^{\text{rGS}} = \alpha_{\text{opt}} = \frac{1}{\sum_{j=1}^q \alpha_j}.$$

Block Convergence Results HPD Case (cont.)

(ii) In Gauss-Southwell with the greedy pick, then

$$\|\mathbf{x}^k - \mathbf{x}^*\|_A^2 \leq (1 - \alpha^{\text{GSW}})^k \|\mathbf{x}^0 - \mathbf{x}^*\|_A^2,$$

with

$$\alpha^{\text{GSW}} = \lambda_{\min}(A) \min_{i=1}^q \frac{\pi_i}{\lambda_{\max}(A_{ii})}, \quad \pi_i = \frac{1/\beta_i^2}{\sum_{j=1}^n 1/\beta_j^2}.$$

The factor $1 - \alpha^{\text{GSW}}$ becomes smallest if we take $\beta_i = 1/\sqrt{\lambda_{\max}(A_{ii})}$ for all i in the greedy pick rule, i.e., we choose i such that

$$\frac{\|\mathbf{r}_i^k\|_i^2}{\lambda_{\max}(A_{ii})} = \max_{j=1}^q \frac{\|\mathbf{r}_j^k\|_j^2}{\lambda_{\max}(A_{jj})}.$$

For A non-Hermitian, we need a new norm, and a new concept of block diagonally dominant

The usual approach for (strict) BDD is (by rows and cols., resp.)

$$\sum_{j \neq i, j=1}^q \|A_{ii}^{-1} A_{ij}\|_{ij} \leq (<) 1, \quad \sum_{i \neq j, j=1}^q \|A_{ij} A_{jj}^{-1}\|_{ij} \leq (<) 1$$

For A non-Hermitian, we need a new norm, and a new concept of block diagonally dominant

The usual approach for (strict) BDD is (by rows and cols., resp.)

$$\sum_{j \neq i, i=1}^q \|A_{ii}^{-1} A_{ij}\|_{ij} \leq (<) 1, \quad \sum_{i \neq j, j=1}^q \|A_{ij} A_{jj}^{-1}\|_{ij} \leq (<) 1$$

Instead, we define a Globally Strictly Block Diagonal Dominant matrix (GISBDD) A if for $i = 1, \dots, q$,

$$\|S_{-j}^T A S_j A_{jj}^{-1}\|_{(*, -j), (*, j)} : \rho_j \leq (<) < 1,$$

where $\|\cdot\|_{(*, -j), (*, j)}$ denotes the operator norm of the “block column” $S_{-j}^T A S_j A_{jj}^{-1}$ as a mapping from V_j to $V_{j'}$, i.e.,

$$\|S_{-j}^T A S_j A_{jj}^{-1} \mathbf{x}\|_{(*, -j)} \leq (<) < \rho_j \|\mathbf{x}_j\|_{(*, j)} \text{ for all } \mathbf{x}_j \in V_j,$$

and $\|\mathbf{x}_j\|_{(*, j)} = \|S_j \mathbf{x}_j\|_*$ for a norm in $V = \mathbb{R}^n$.

More about the norms

All we need is that

$$\begin{aligned}\|\mathbf{x}\|_* &= \|E_j \mathbf{x}\|_* + \|E_{-j} \mathbf{x}\|_* \text{ for all } \mathbf{x} \in \mathbb{R}^n, \\ \|\mathbf{x}\|_* &\leq \sum_{j=1}^q \|S_j^T \mathbf{x}\|_{(*,j)} \text{ for all } \mathbf{x} \in \mathbb{R}^n.\end{aligned}$$

In particular if A is an H-matrix (i.e., its companion matrix $\langle A \rangle$ is an M-matrix), then we can take

$$\|\mathbf{x}\|_* = \|\mathbf{x}\|_{\mathbf{u}} := \sum_{\ell=1}^n u_{\ell} \cdot |x_{\ell}| = \mathbf{u}^T |\mathbf{x}|,$$

with \mathbf{u} a positive vector such that $\mathbf{u}^T \langle A \rangle > 0$ (which always exists and e.g., one can take it as the Perron vector of A^T).

Block randomized convergence result

Theorem. (i) In randomized Block Gauss-Seidel, expected value for the norm of the residuals \mathbf{r}^k of the iterates \mathbf{x}^k at step k satisfies

$$\mathbb{E}(\|\mathbf{r}^k\|_*) \leq (1 - \alpha^r)^k \|\mathbf{r}^0\|_*,$$

with

$$\alpha^r = \min_{j=1}^q p_j / \gamma_j > 0 \quad \text{where} \quad \gamma_j = (1 - \rho_j)^{-1}.$$

Block randomized convergence result

Theorem. (i) In randomized Block Gauss-Seidel, expected value for the norm of the residuals \mathbf{r}^k of the iterates \mathbf{x}^k at step k satisfies

$$\mathbb{E}(\|\mathbf{r}^k\|_*) \leq (1 - \alpha^r)^k \|\mathbf{r}^0\|_*,$$

with

$$\alpha^r = \min_{j=1}^q \rho_j / \gamma_j > 0 \quad \text{where} \quad \gamma_j = (1 - \rho_j)^{-1}.$$

Moreover, α^r is maximized if one takes

$$\rho_j = \gamma_j / \sum_{i=1}^q \gamma_i, \quad j = 1, \dots, q;$$

its value then is $\alpha^r = \alpha_{\text{opt}} := 1 / \sum_{i=1}^q \gamma_i$.

Block randomized convergence result (cont.)

(ii) If in the Greedy Algorithm we take the greedy pick rule

$$\beta_j \|\mathbf{r}_j\|_{(*,j)} = \max_{i=1,\dots,q} \beta_i \|\mathbf{r}_{(*,i)}\|_{(*,i)}, \quad (2)$$

then putting

$$\alpha^g = \min_{j=1}^n \frac{\pi_j}{\gamma_j}, \text{ where } \pi_j = \frac{1/\beta_j}{\sum_{\ell=1}^n 1/\beta_\ell}, \gamma_j = (1 - \rho_j)^{-1}$$

we have

$$\|\mathbf{r}^k\|_* \leq (1 - \alpha^g)^k \|\mathbf{r}^0\|_*.$$

Moreover, α^g is maximized if we take

$$\beta_j = 1/\gamma_j = 1 - \rho_j, \quad j = 1, \dots, n;$$

its value is then identical to α_{opt} from part (i).

In other words

We proved that in **expected value sense** and using a **weighted block one-norm**, the bound on the convergence factor for the residual for Randomized Gauss-Seidel (**with overlap**) is the same as for block Gauss-Southwell, but without the extra cost per block relaxation.

Experiments

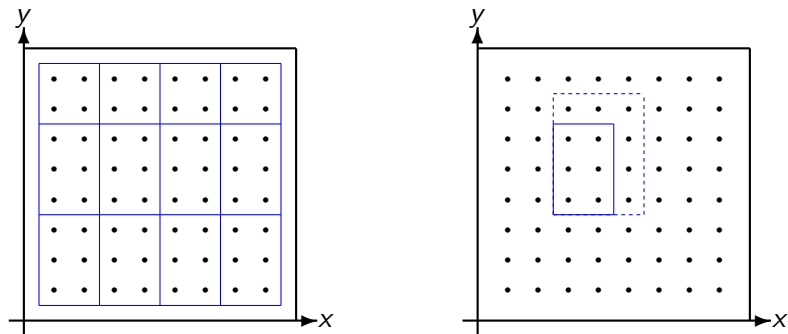
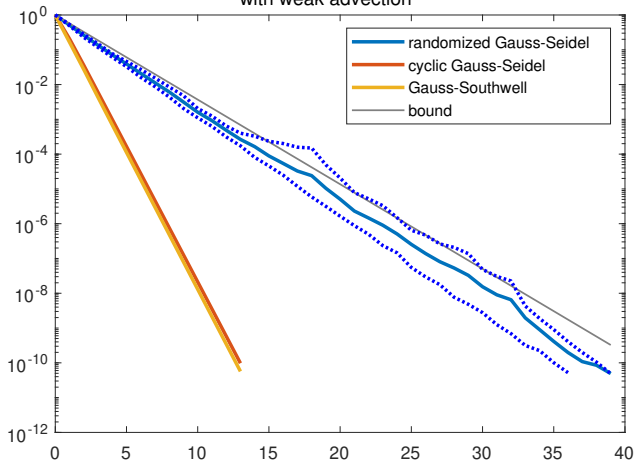


Figure: Example for a covering with 2-dimensional blocks, $N = 8$. Left: Partitioning with 12 blocks, $n_x = 2$, $n_y = 3$. Right: Extension of one 2×3 block of the partitioning in an overlapping covering with $ovl_x = 1$, $ovl_y = 1$

Numerical example with weak advection. No overlap

$N = 100, n_x = 4, n_y = 4, \text{ovl}_x = 0, \text{ovl}_y = 0$

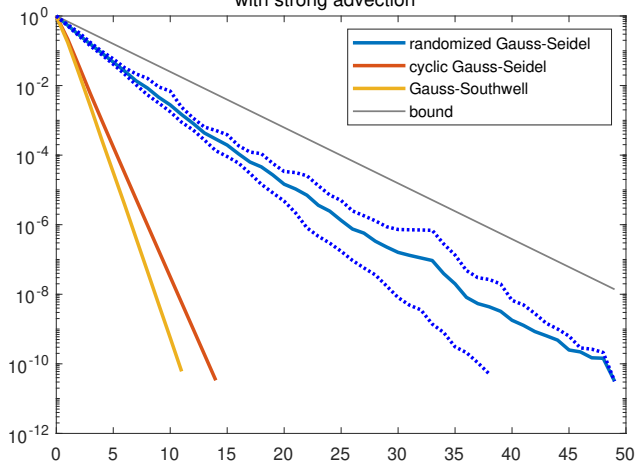
with weak advection



Numerical example with strong advection. No overlap

$N = 100, n_x = 4, n_y = 4, \text{ovl}_x = 0 \text{ ovl}_y = 0$

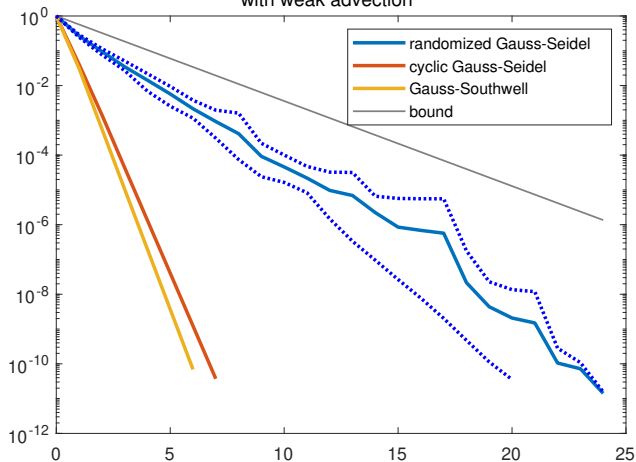
with strong advection



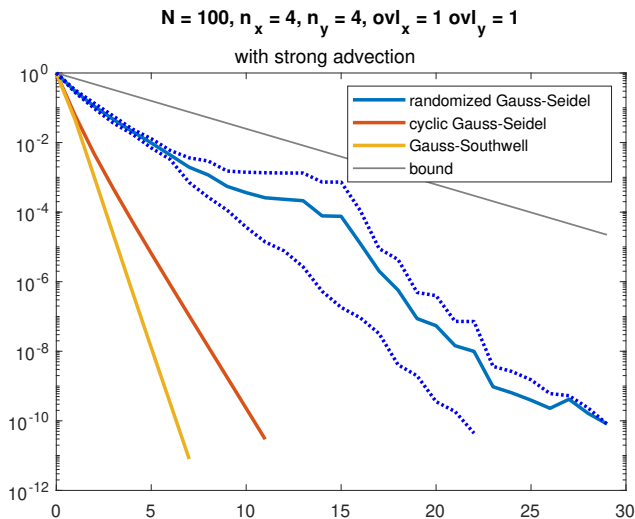
Numerical example with weak advection. With overlap

$N = 100, n_x = 4, n_y = 4, \text{ovl}_x = 1 \text{ ovl}_y = 1$

with weak advection



Numerical example with strong advection. With overlap



Observation

The improvement in convergence speed due to the overlap is largest for the randomized algorithm where the average number of iterations needed to reduce the residual by a factor of 10^{-10} decreases by 43% (from 49 to 28), while it decreases only by 21% (from 14 to 11) for the cyclic method (and by 36% for Gauss-Southwell).

Asynchronous block iterations (with overlap) aka asynchronous Schwarz

First, synchronous

Algorithm 10 Classical block Jacobi iteration in parallel for a fixed partition. Processor i solves local problem i

for $k = 1, 2, \dots$ until a convergence criterion is satisfied **do**

for $i = 1, \dots, q$ (in parallel) **do**

 read $\mathbf{x}_j^k, j \neq i$ from the other processors

 Solve $A_{ii}\mathbf{x}_i^{k+1} = \mathbf{b}_i - S_i^T A S_{-i}\mathbf{x}_{-i}^k$

 write \mathbf{x}_i^{k+1}

end for

 Wait or Synchronize

end for

Asynchronous block iterations (with overlap) aka asynchronous multiplicative Schwarz

No iteration count, just block relaxations

Algorithm 11 Asynchronous block Jacobi iteration in parallel for a fixed partition. Processor i solves local problem i

```
for  $i = 1, \dots, q$  (in parallel) do  
  read  $\mathbf{x}_j, j \neq i$  from the other processors  
  Solve  $A_{ii}\mathbf{x}_i = \mathbf{b}_i - S_i^T A S_{-i}\mathbf{x}_{-i}$   
  write  $\mathbf{x}_i$   
end for
```

Randomized view of Asynchronous Iterations (2002)

At each time stamp k

$$x_i^k = \begin{cases} \mathcal{T}_i \left(x_1^{s_1(k)}, \dots, x_q^{s_q(k)} \right) & \text{with probability } p_i \\ x_i^{k-1} & \text{with probability } 1 - p_i \end{cases}$$

[Strikwerda, LAA, 2002] where he also had $s_i(k)$ as random variables

Of course $\sum_{i=1}^q p_i = 1$ Strikwerda proved that $\mathbb{E}(\|x^k - x^*\|) \rightarrow 0$

for $\mathcal{H} = H - I - D^{-1}A$, $\rho(\mathcal{H}) < 1$

and in fact $\mathbb{E}(\|x^k - x^*\|) = O(R^{-k})$ for some real number R

(radius of analicity of a matrix $M(z) = I - z[I - P + s(z)PH]$,

$P = \text{diag}(p_i)$, $s(z)$ related to randomized $s_i(k)$)

Randomized view of Asynchronous Iterations (2015)

[Avron, Druinsky, Gupta, *J ACM*, 2015] consider $Ax = b$, A SPD Essentially Asynchronous Randomized (point) Jacobi (\equiv Randomized Gauss-Seidel). Let $A = D - B$, $D = \text{diag}(A)$. $H = D^{-1}A$, $c = D^{-1}b$. Computational model here:

1. Bounded delays: $k - s_i(k) \leq d$.
2. Atomic write: only one component is updated at every time step.

Theorem. If $\|H\|_\infty$ small enough and the probabilities are uniform, then

$$\mathbb{E}(\|x^k - x^*\|_A^2) \leq \beta \alpha^{m/(d_0+d)} \|x^0 - x^*\|_A^2$$

where β, α functions of $\lambda_{\max}(H)$, d , n , $\|H\|_\infty$, and $\kappa(H)$, and

$$d_0 = \left\lceil \frac{\log(1/2)}{\log(1 - \lambda_{\max}/n)} \right\rceil$$

Randomized view of Asynchronous Iterations (2026)

Answer to the challenge: We can interpret Asynchronous Schwarz iterations, that is, Asynchronous Block Gauss-Seidel as randomized processes (in the same manner).

Our theorems show now provable linear convergence rate in the expected value sense for A HPD or for A “block diagonally dominant by columns” using the block one-norm (with overlap).

Contributions. In a nutshell

- ▶ We prove that **in expected value sense** and using a **weighted max-norm**, the convergence factor for Randomized Gauss-Seidel is the same as for Gauss-Southwell, but without the extra cost per relaxation.
- ▶ We show this result for nonsymmetric generalized diagonally dominant matrices and H -matrices. (The HPD case is known).
- ▶ Bound appears to be sharp.

Contributions. In a nutshell

- ▶ We prove that **in expected value sense** and using a **weighted max-norm**, the convergence factor for Randomized Gauss-Seidel is the same as for Gauss-Southwell, but without the extra cost per relaxation.
- ▶ We show this result for nonsymmetric generalized diagonally dominant matrices and H -matrices. (The HPD case is known).
- ▶ Bound appears to be sharp.
- ▶ In our experiments “cyclical” (or classical) Gauss-Seidel faster than Randomized Gauss-Seidel, both for Hermitian and non-Hermitian problems.
But our goal was more theoretical than to offer a new method,

Contributions. In a nutshell

- ▶ We prove that **in expected value sense** and using a **weighted max-norm**, the convergence factor for Randomized Gauss-Seidel is the same as for Gauss-Southwell, but without the extra cost per relaxation.
- ▶ We show this result for nonsymmetric generalized diagonally dominant matrices and H -matrices. (The HPD case is known).
- ▶ Bound appears to be sharp.
- ▶ In our experiments “cyclical” (or classical) Gauss-Seidel faster than Randomized Gauss-Seidel, both for Hermitian and non-Hermitian problems.
But our goal was more theoretical than to offer a new method, and to use this to be able to consider **block methods**, and analysis of asynchronous iterations.

For the asynchronous iterative methods

- ▶ We have then a provable convergence result for asynchronous iterations in the block case, for nonsymmetric problems, using a new concept of block diagonal dominance.

Published paper (for first step)

Andreas Frommer and Daniel B. Szyld,
On the convergence of randomized and greedy relaxation schemes
for solving nonsingular linear systems of equations,
Numerical Algorithms **92** (2023) 639–664.

It can be found at: <https://faculty.cst.temple.edu/~szyld>

Block case and Asynchronous convergence in preparation.