Fast Multilevel Numerical Methods for Random Walks on Directed Graphs

Hans De Sterck

Department of Applied Mathematics

University of Waterloo

University of Waterloo

RSA 2009

14th International Conference on Random Structures and Algorithms

collaborators

- Killian Miller, Manda Winlaw

  Department of Applied Mathematics, University of Waterloo
- Steve McCormick, Tom Manteuffel, John Ruge, Geoff Sanders

  Department of Applied Mathematics, University of Colorado at Boulder
- our area of research is numerical linear algebra methods for PDEs, in particular so-called algebraic multigrid methods, and we have recently started to apply these techniques to numerical linear algebra methods for Markov chains

1. problem formulation and example

- develop efficient numerical method for calculating stationary distributions of Markov chains:
  - finite-state (n states)
  - irreducible
  - large
  - sparse
  - slowly mixing
- goal: O(n) method

$\Rightarrow$ approach: use iterative method with multilevel aggregation to distribute probability on all scales quickly

hdesterck@uwaterloo.ca

problem formulation

$$B\,\mathbf{x} = \mathbf{x} \qquad \|\mathbf{x}\|_1 = 1 \qquad x_i \geq 0\,\forall i$$

- *B* is column-stochastic

$$0 \leq b_{ij} \leq 1 \;\forall i,\, j \qquad \mathbf{1}^T B = \mathbf{1}^T$$

- *B* is irreducible (every state can be reached from every other state in the directed graph)

$$\Longrightarrow$$

$$\exists!\;\mathbf{x}: \quad B\,\mathbf{x} = \mathbf{x} \qquad \|\mathbf{x}\|_1 = 1 \qquad x_i > 0\;\forall i$$

- singular M-matrix formulation

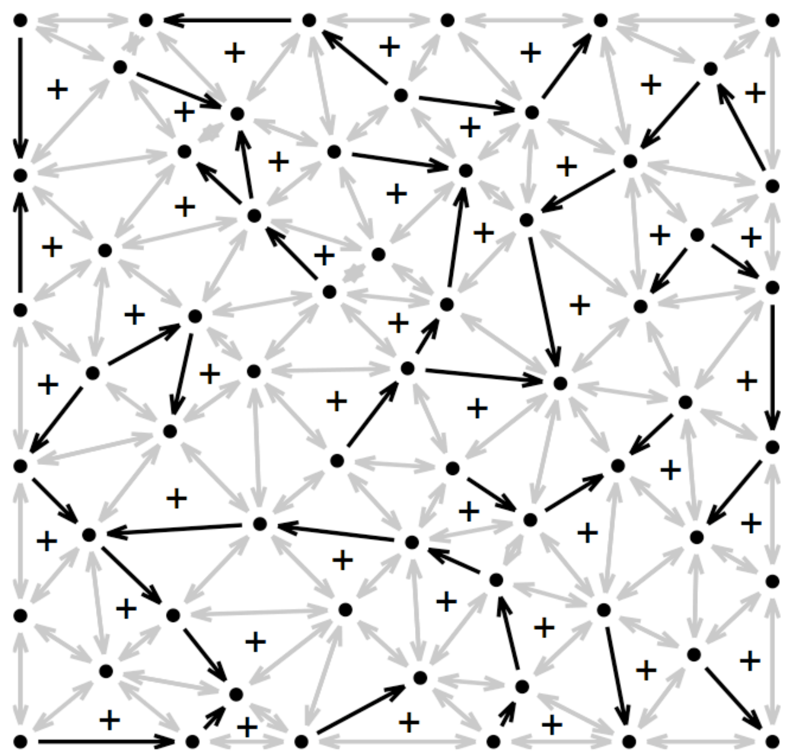
$$A = \begin{bmatrix} + & - & - & - & - \\ - & + & - & - & - \\ - & - & + & - & - \\ - & - & - & + & - \\ - & - & - & - & + \end{bmatrix}$$

$$A\,\mathbf{x} = 0 \qquad A = I - B$$

example

- example: random walk on directed planar graph
- choose n uniformly distributed random points in the unit square
- perform Delaunay triangulation on points
- choose a maximal subset of triangles that are not neighbours
- randomly delete one directed edge from each triangle in this subset

⇒ find stationary distribution of random walk



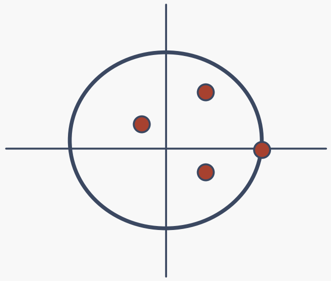
University of Waterloo

RSA 2009
hdesterck@uwaterloo.ca

2. power method convergence

- power method: $\mathbf{x}_{i+1} = B\,\mathbf{x}_i$
- largest eigenvalue of *B*: $\lambda_1 = 1$
- power method (nonperiodic *B*):
  - convergence rate: 1-$|\lambda_2|$
  - convergence is slow when 1-$|\lambda_2|$ $\rightarrow$ 0 for increasing n

    (we call this a slowly mixing Markov chain)
  - every power iteration is O(n) work



University of Waterloo

hdesterck@uwaterloo.ca

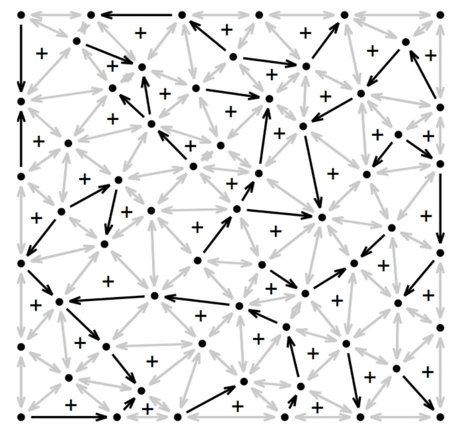
numerical results: one-level iteration for random graph problem

- start from random intial guess $\mathbf{x}_0$
- let $A = D - (L + U)$
- iterate on $\mathbf{x}_{i+1} = (I + w\, D^{-1}\, A)\, \mathbf{x}_i$

  with $w = 0.7$

  until

$$\frac{\|A\,\mathbf{x}_i\|_1}{\|A\,\mathbf{x}_0\|_1} < 10^{-8}$$

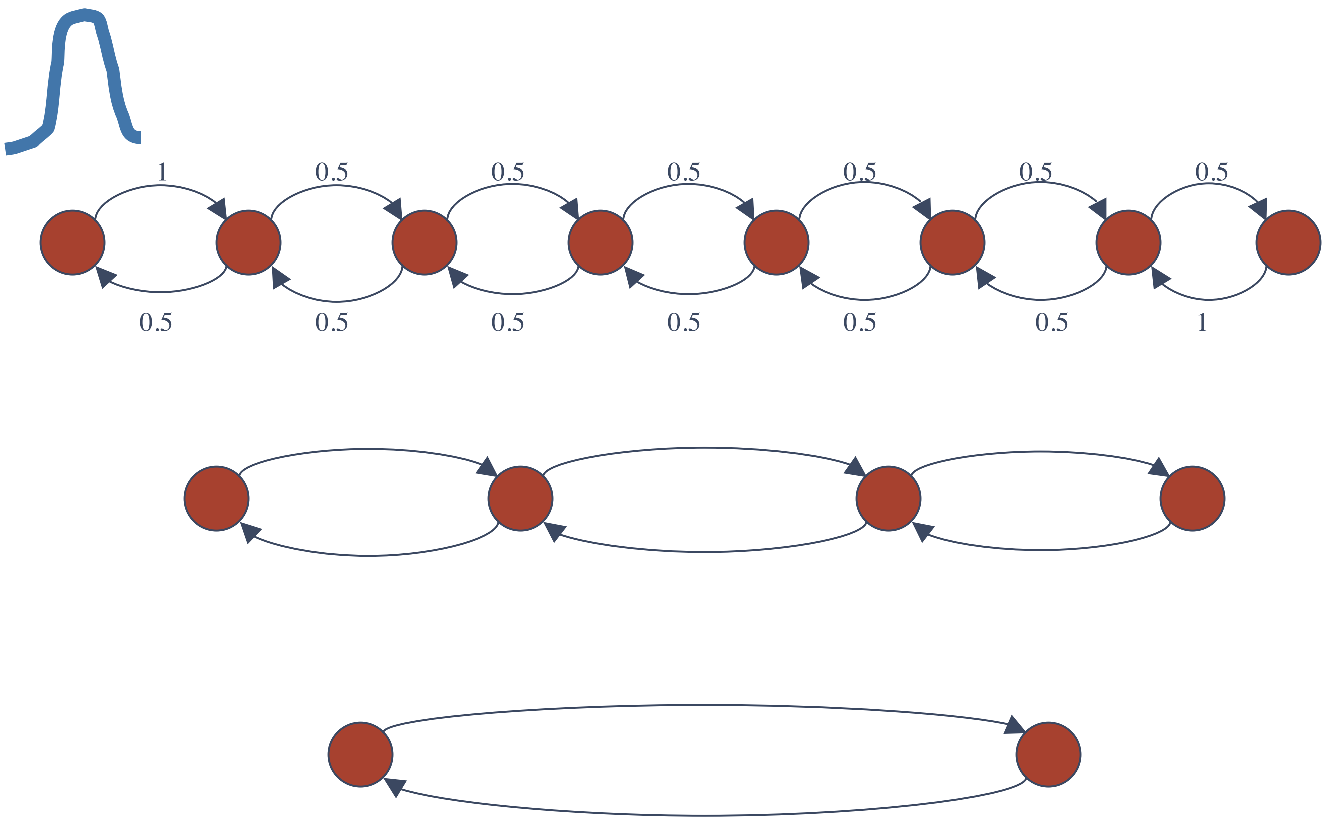


n	it
128	322
256	494
512	1010
1024	1768

University of Waterloo

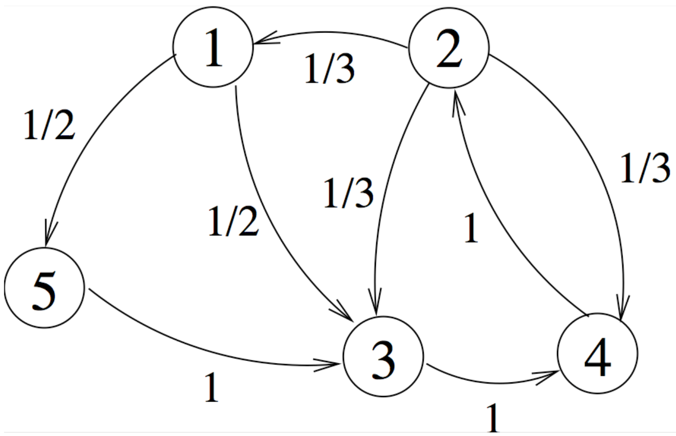
hdesterck@uwaterloo.ca

why/when is power method slow?
why multilevel methods?



RSA 2009
hdesterck@uwaterloo.ca

3. multilevel aggregation



$$B = \begin{bmatrix} 0 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1/2 & 1/3 & 0 & 0 & 1 \\ 0 & 1/3 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B\,\mathbf{x} = \mathbf{x} \qquad \|\mathbf{x}\|_1 = 1$$

$$\mathbf{x}^T = [2/19\ 6/19\ 4/19\ 6/19\ 1/19]$$

RSA 2009
hdesterck@uwaterloo.ca

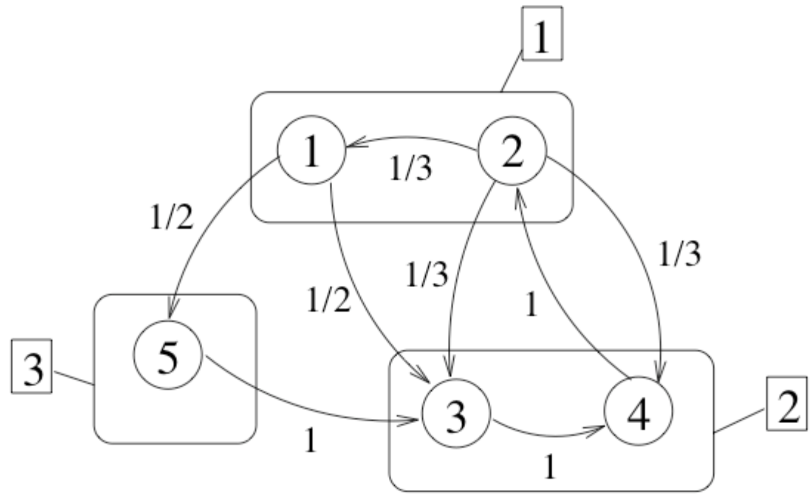
aggregation

- form three coarse, aggregated states

$$B_c\,\mathbf{x}_c = \mathbf{x}_c$$

$$b_{c,IJ} = \frac{\sum_{j\in J} x_j \left(\sum_{i\in I} b_{ij}\right)}{\sum_{j\in J} x_j}$$

$$B_c = \begin{bmatrix} 1/4 & 3/5 & 0 \\ 5/8 & 2/5 & 1 \\ 1/8 & 0 & 0 \end{bmatrix}$$



$$x_{c,I} = \textstyle\sum_{i\in I} x_i$$

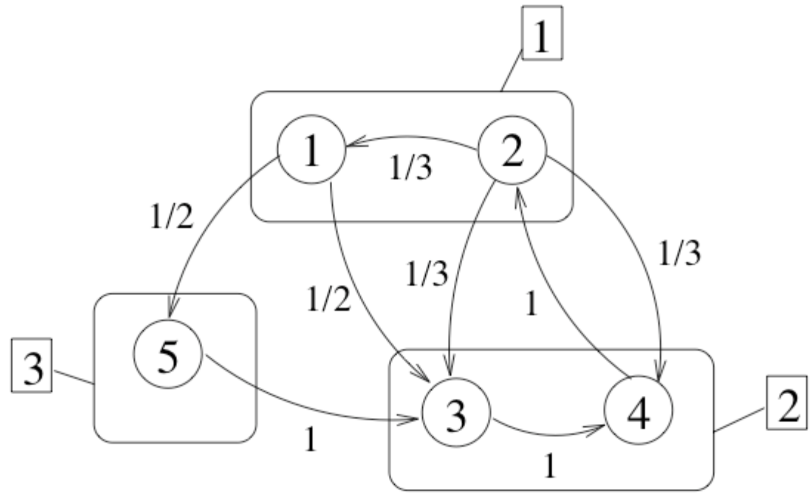
$$\mathbf{x}_c^T = [8/19\;\; 10/19\;\; 1/19]$$

(Simon and Ando, 1961)

matrix form of aggregation

$$B_c \mathbf{x}_c = \mathbf{x}_c$$

$$b_{c,IJ} = \frac{\sum_{j \in J} x_j \left(\sum_{i \in I} b_{ij} \right)}{\sum_{j \in J} x_j}$$



$$B_c = Q^T \, B \, \mathrm{diag}(\mathbf{x}) \, Q \, \mathrm{diag}(Q^T \, \mathbf{x})^{-1}$$

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$x_{c,I} = \textstyle\sum_{i \in I} x_i$$

$$\mathbf{x}_c = Q^T \, \mathbf{x}$$

(Krieger, Horton, ... 1990s)

two-level aggregation method

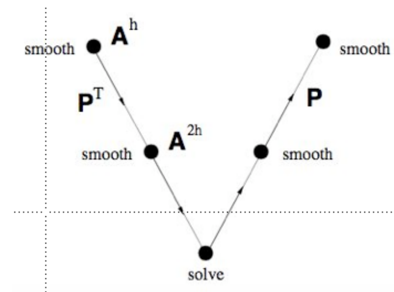
repeat

fine-level relaxation: $\mathbf{x}^* = B\,\mathbf{x}_i$
build Q
build $B_c = Q^T B \,\mathrm{diag}(\mathbf{x}^*)\, Q\, (\mathrm{diag}(Q^T\,\mathbf{x}^*))^{-1}$
coarse-level solve: $B_c\,\mathbf{x}_c = \mathbf{x}_c$
fine-level update: $\mathbf{x}_{i+1} = \mathrm{diag}(\mathbf{x}^*)\, Q\, (\mathrm{diag}(Q^T\,\mathbf{x}^*))^{-1}\mathbf{x}_c$

(note: there is a convergence proof for this two-level method, Marek and Mayer 1998, 2003)

multilevel aggregation method

Algorithm: Multilevel Adaptive Aggregation method (V-cycle)



$\mathbf{x} = \mathsf{AM\_V}(A, \mathbf{x}, \nu_1, \nu_2)$

begin

$\mathbf{x} \leftarrow \mathsf{Relax}(A, \mathbf{x}) \quad \nu_1$ times

build Q based on $\mathbf{x}$ and A (Q is rebuilt every level and cycle)

$R = Q^T$ and $P = \mathrm{diag}(\mathbf{x})\, Q$

$A_c = R\, A\, P$

$\mathbf{x}_c = \mathsf{AM\_V}(A_c\,\mathrm{diag}(P^T\,\mathbf{1})^{-1}, P^T\,\mathbf{1}, \nu_1, \nu_2)$ (coarse-level solve)

$\mathbf{x} = P\,(\mathrm{diag}(P^T\,\mathbf{1}))^{-1}\mathbf{x}_c$ (coarse-level correction)

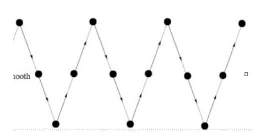
$\mathbf{x} \leftarrow \mathsf{Relax}(A, \mathbf{x}) \quad \nu_2$ times

end

(Krieger, Horton 1994)

University of Waterloo

(note: O(n) work per cycle:
n + n/2 + n/4 + n/8 + ... < 2 n)



hdesterck@uwaterloo.ca

aggregation strategy

- fine-level relaxation should efficiently distribute probability within aggregates (smooth out local, high-frequency errors)

- coarse-level update will efficiently distribute probability between aggregates (smooth out global, low-frequency errors)

- base aggregates on 'strong connections' in $A \, \mathrm{diag}(\mathbf{x}_i)$

University of Waterloo

hdesterck@uwaterloo.ca

aggregation strategy

scaled problem matrix:

$$\hat{A} = A\,\mathrm{diag}(\mathbf{x}_i)$$

strong connection: coefficient is large in either of rows *i* or *j*

$$-\hat{a}_{ij} \geq \theta \max_{k \neq i}\{-\hat{a}_{ik}\} \quad \text{or} \quad -\hat{a}_{ji} \geq \theta \max_{k \neq j}\{-\hat{a}_{jk}\}$$

($\theta \in (0,1)$, $\theta=0.25$)

University of Waterloo

hdesterck@uwaterloo.ca

‘neighbourhood’ aggregation strategy

Algorithm 2: neighborhood-based aggregation, $\{Q_J\}_{J=1}^m \longleftarrow$ **Neighbourhood-hoodAgg** $(A\,\mathrm{diag}(\mathbf{x}),\ \theta)$

For all points i, build strong neighbourhoods $\mathcal{N}_i$ based on $A\,\mathrm{diag}(\mathbf{x})$ and θ.
Set $\mathcal{R} \leftarrow \{1, ..., n\}$ and $J \leftarrow 0$.

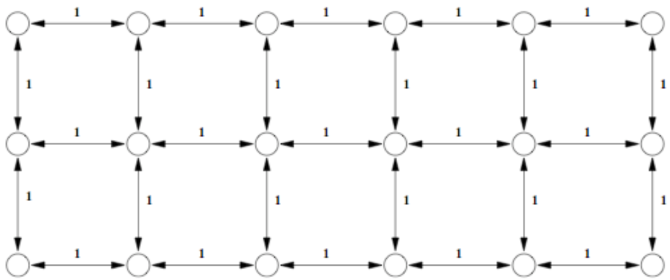
```
/* 1st pass:  assign entire neighborhoods to aggregates */
```

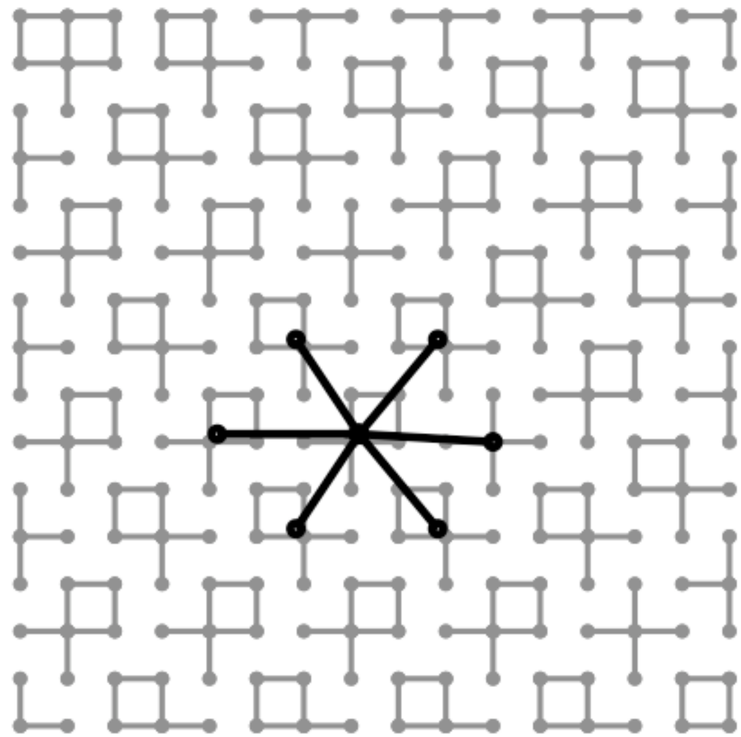
for $i \in \{1, ..., n\}$ **do**
    if $(\mathcal{R} \cap \mathcal{N}_i) = \mathcal{N}_i$ **then**
        $J \leftarrow J + 1$.
        $Q_J \leftarrow \mathcal{N}_i,\ \hat{Q}_J \leftarrow \mathcal{N}_i$.
        $\mathcal{R} \leftarrow \mathcal{R} \setminus \mathcal{N}_i$.
    end
end
$m \leftarrow J$.

```
/* 2nd pass:  put remaining points in aggregates they are most
   connected to */
```

while $\mathcal{R} \neq \emptyset$ **do**
    Pick $i \in \mathcal{R}$ and set $J \leftarrow \mathrm{argmax}_{K=1,...,m}\ \mathrm{card}\,(\mathcal{N}_i \cap Q_K)$.
    Set $\hat{Q}_J \leftarrow Q_J \cup \{i\}$ and $\mathcal{R} \leftarrow \mathcal{R} \setminus \{i\}$.
end
for $J \in \{1, ..., m\}$ **do** $Q_J \leftarrow \hat{Q}_J$.

aggregation: periodic 2D lattice



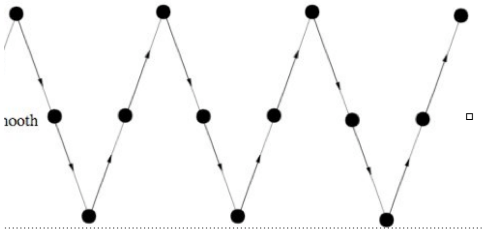


$$B_c = Q^T \, B \, \mathrm{diag}(\mathbf{x}^*) \, Q \, (\mathrm{diag}(Q^T \, \mathbf{x}^*))^{-1}$$

University of Waterloo

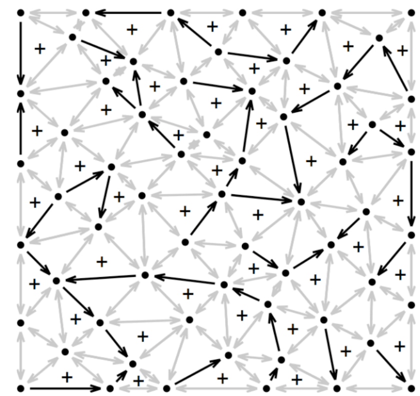
hdesterck@uwaterloo.ca

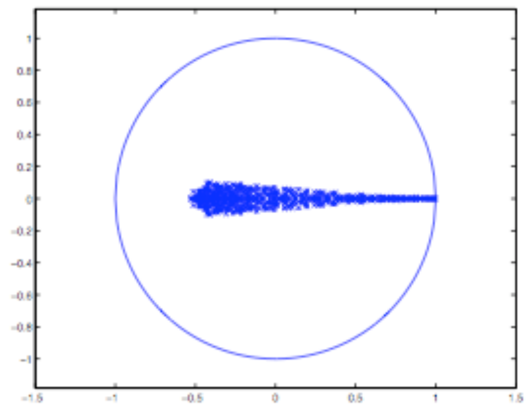
numerical results: aggregation multigrid for random walk problem



	1-level	aggregation		
n	iterations	iterations	C_{op}	levels
128	322	95	1.12	3
256	494	107	1.13	3
512	1010	156	1.14	3
1024	1768	220	1.15	4
2048		352	1.15	4

$$C_{op} = \frac{\sum_{l=0} \mathrm{nonzeros}(A_l)}{nonzeros(A_0)}$$



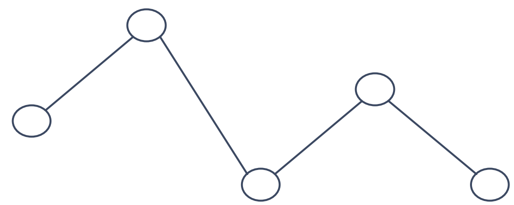


does not work as well as we would like!

University of Waterloo

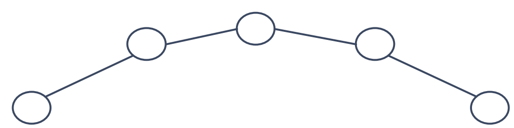
4. overlapping aggregates: we need 'smoothed aggregation'...

(Vanek, Mandel, and Brezina, Computing, 1996)

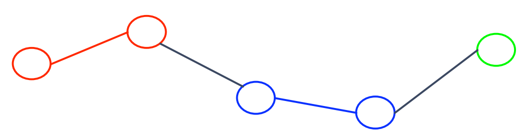


$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

after smoothing:

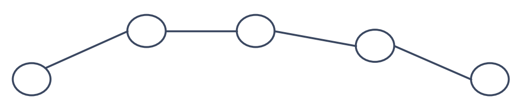


coarse grid correction with Q:



$$Q_s = \begin{bmatrix} \times & 0 & 0 \\ \times & \times & 0 \\ \times & \times & 0 \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}$$

coarse grid correction with Q_s:



University of Waterloo

hdesterck@uwaterloo.ca

smoothed aggregation

$$A_c = Q^T\, A\, \mathrm{diag}(\mathbf{x}_i)\, Q = R\, A\, P$$

- smooth the columns of *P* with weighted Jacobi:

$$P_s = (I + w\, D^{-1}\, A)\, \mathrm{diag}(\mathbf{x}_i)\, Q$$

$$w = 0.7$$

- smooth the rows of *R* with weighted Jacobi:

$$R_s = Q^T\, (I + w\, A\, D^{-1})$$

RSA 2009
hdesterck@uwaterloo.ca

smoothed aggregation: a problem with signs

- smoothed coarse level operator:

$$\begin{aligned} A_{cs} &= R_s\,(D-(L+U))\,P_s \\ &= R_s\,D\,P_s - R_s\,(L+U)\,P_s \end{aligned}$$

$$A = \begin{bmatrix} + & - & - & - & - \\ - & + & - & - & - \\ - & - & + & - & - \\ - & - & - & + & - \\ - & - & - & - & + \end{bmatrix}$$

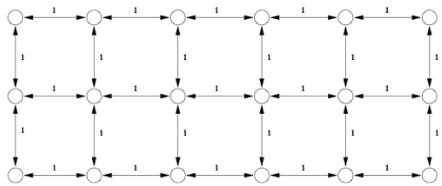
- problem: A_{cs} is not a singular M-matrix (signs wrong)
- solution:
  lumping approach
- well-posedness of this approach shown in *De Sterck et al., SIAM J. Sci. Comp., 2009*

$$S_{\{i,j\}} = \begin{array}{c c} & \begin{array}{ccccc} & i & & j & \end{array} \\ \begin{array}{c} \\ i \\ \\ j \\ \\ \end{array} & \begin{bmatrix} \ddots & \vdots & & \vdots & \\ \cdots & \beta_{\{i,j\}} & \cdots & -\beta_{\{i,j\}} & \cdots \\ & \vdots & & \vdots & \\ \cdots & -\beta_{\{i,j\}} & \cdots & \beta_{\{i,j\}} & \cdots \\ & \vdots & & \vdots & \end{bmatrix} \end{array}$$

University of Waterloo

hdesterck@uwaterloo.ca

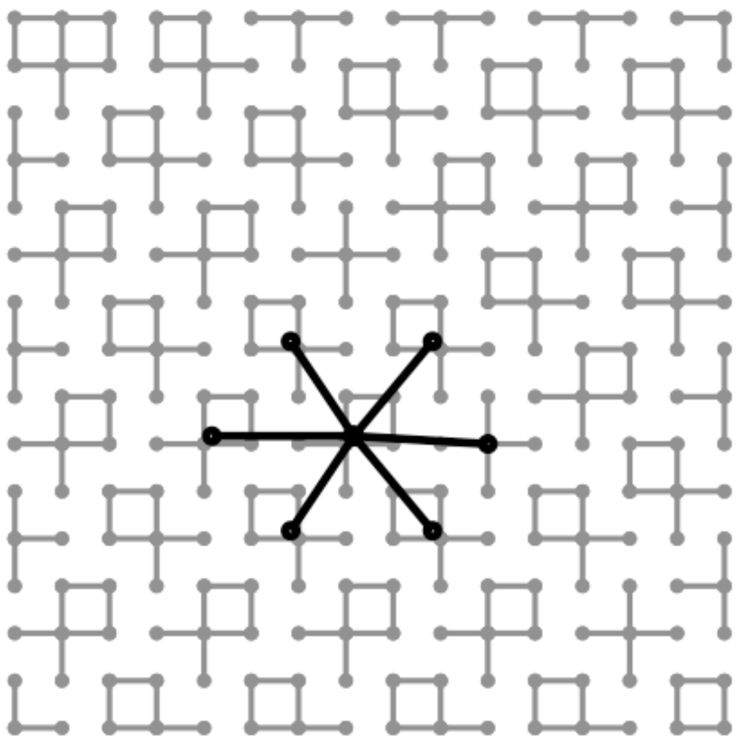
smoothed aggregation: periodic 2D lattice

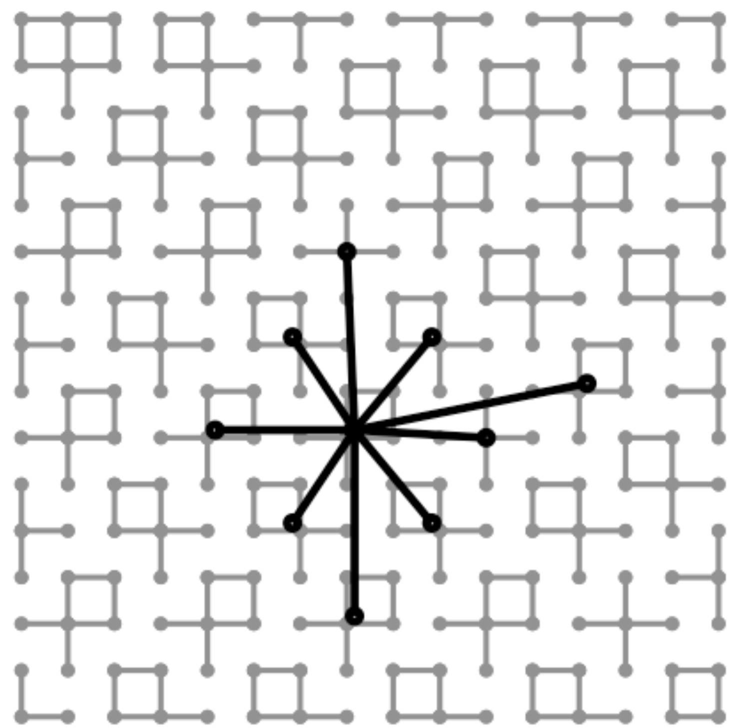


$$A_c = R_s \, A \, P_s$$

unsmoothed

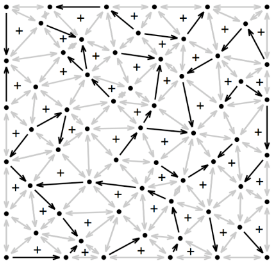
smoothed





hdesterck@uwaterloo.ca

numerical results: smoothed aggregation multigrid for random graph problem



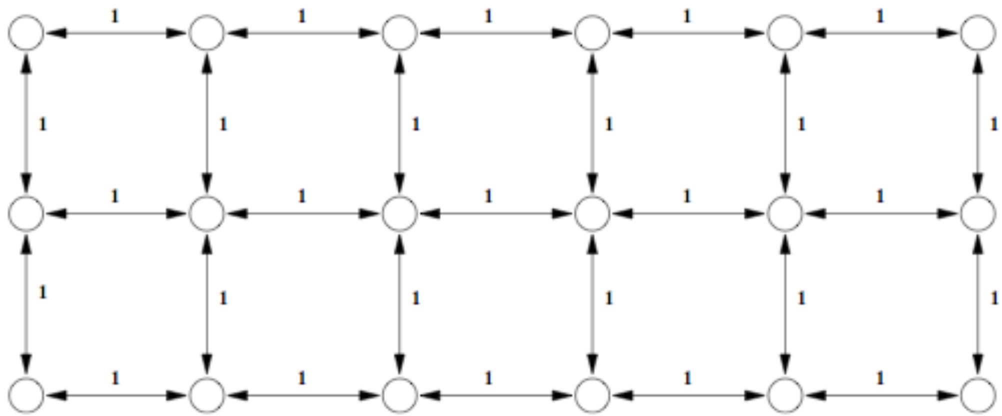
	1-level	aggregation			smoothed aggregation			
n	iterations	iterations	C_{op}	levels	iterations	C_{op}	levels	R_{lump}
128	322	95	1.12	3				
256	494	107	1.13	3				
512	1010	156	1.14	3	36	1.28	3	2.5e-4
1024	1768	220	1.15	4	39	1.31	4	1.2e-4
2048		352	1.15	4	33	1.31	4	6.0e-5
4096					46	1.35	4	2.3e-4
8192					35	1.37	4	2.0e-4
16384					51	1.36	5	9.4e-5
32768					43	1.38	5	1.6e-4

$$C_{op} = \frac{\sum_{l=0} \text{nonzeros}(A_l)}{nonzeros(A_0)}$$

University of Waterloo

RSA 2009
hdesterck@uwaterloo.ca

numerical results: smoothed aggregation multigrid for periodic 2D lattice problem



	1-level	aggregation			smoothed aggregation			
n	iterations	iterations	C_{op}	levels	iterations	C_{op}	levels	R_{lump}
64	197	47	1.23	3	16	1.26	3	0
256	760	96	1.26	3	17	1.34	3	0
1024	2411	242	1.25	4	17	1.32	4	0
4096		328	1.26	5	18	1.34	5	0
16384					18	1.33	5	0
32768					19	1.34	6	0

University of Waterloo

hdesterck@uwaterloo.ca

numerical results: smoothed aggregation multigrid for tandem queueing network problem

FIG. 5.6. *Tandem queueing network.*

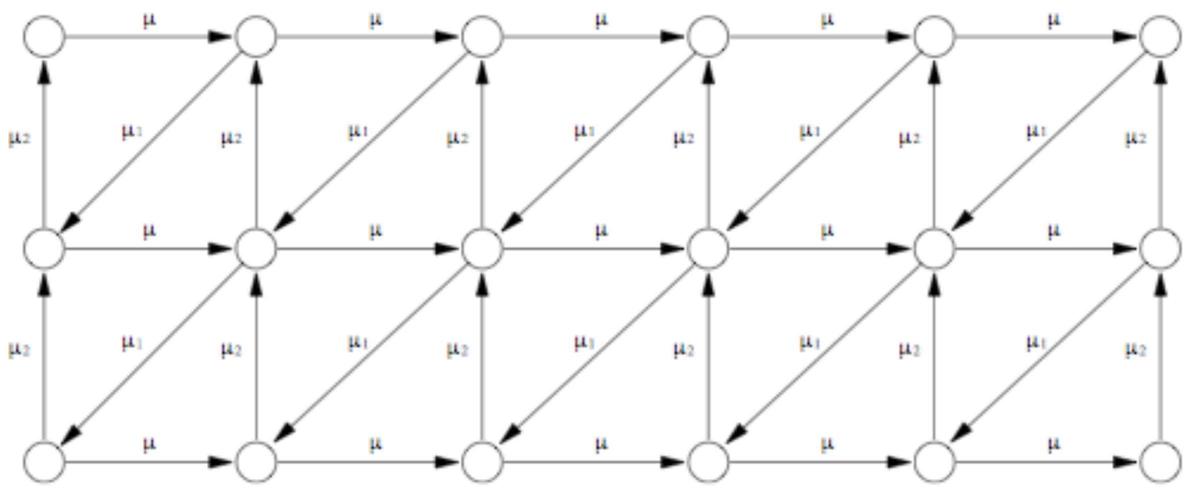
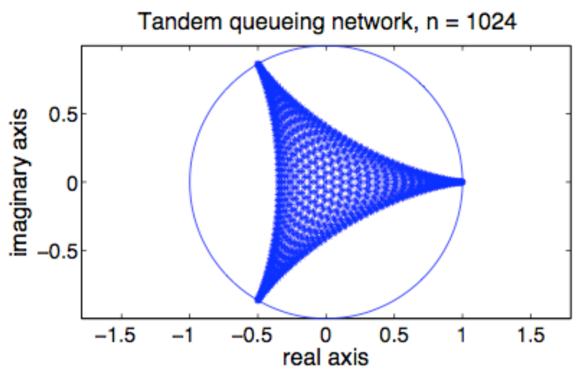


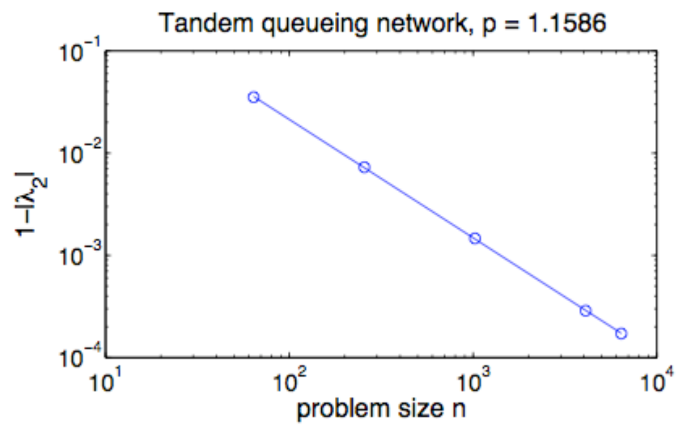
FIG. 5.7. *Graph for tandem queueing network.*

University of Waterloo

hdesterck@uwaterloo.ca

numerical results: smoothed aggregation for tandem queueing network problem





	1-level	aggregation			smoothed aggregation			
n	iterations	iterations	C_{op}	levels	iterations	C_{op}	levels	R_{lump}
256	967	103	1.23	3	17	1.25	3	1.7e-3
1024	4004	256	1.21	4	20	1.25	4	1.4e-3
4096		425	1.22	4	19	1.24	4	9.5e-4
16384					22	1.24	5	5.1e-4
65536					18	1.25	6	3.5e-4

University of Waterloo

hdesterck@uwaterloo.ca

6. discussion

- multilevel smoothed aggregation gets us close to O(n) algorithm for some slowly mixing Markov chains
- slowly mixing Markov chains are OK (their stationary distribution can be calculated efficiently)

- very little theory exists for these methods
  - convergence
  - optimal convergence (O(n))
- there is optimal convergence theory for SPD matrix discretizations of some elliptic PDEs (Brandt, Stueben, ...)

discussion

- we have several variants of these algorithms that also work well

- we are working on similar multilevel aggregation approach to speed up Markov Chain Monte Carlo methods for lattice spin systems (make groups of groups ... of spins and flip them together)

RSA 2009
hdesterck@uwaterloo.ca

7. questions

- any suggestions for further test problems for our algorithms? (large, sparse, irreducible, slowly mixing)
  - real-life problems
  - theoretical models that people care about
- any suggestions for 'pathological' chains that will 'break' our algorithm?
- which classes of Markov chains will this work well for, and which classes not? (how can these classes be characterized?)
- (optimal) convergence proof?

hdesterck@uwaterloo.ca

thanks!

University of Waterloo

hdesterck@uwaterloo.ca