Markov Chains and Web Ranking: a Multilevel Adaptive Aggregation Method

Hans De Sterck
Department of Applied Mathematics, University of Waterloo

Quoc Nguyen; Steve McCormick, John Ruge, Tom Manteuffel (Boulder, Colorado)
1. Simple Markov Chain Example

- 5 states
- each outgoing edge same probability (random walk on directed graph)
Simple Markov Chain Example

• start in one state with probability 1: what is the stationary probability vector after $\infty$ number of steps?

$$x_{i+1} = B x_i$$

• stationary probability:

$$B x = x \quad \|x\|_1 = 1$$

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

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

\[ B \mathbf{x} = \mathbf{x} \quad \|\mathbf{x}\|_1 = 1 \]

- \( B \) is column-stochastic

\[ b_{i,j} \geq 0, \quad \sum_i b_{i,j} = 1 \quad \forall j \]

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

\[ \exists! \, \mathbf{x} : B \mathbf{x} = \mathbf{x} \quad \text{and} \quad \|\mathbf{x}\|_1 = 1, \quad x_i > 0 \quad \forall i \]

(no probability sinks!)
Problem Statement

$B x = x \quad \|x\|_1 = 1$

- $B$ is column-stochastic
- $B$ is irreducible
- $B$ is a-periodic
  \[ \Rightarrow \]
  \[ x = \lim_{n \to \infty} B^n x_0 \quad \text{for any } x_0 \]

- largest eigenvalue of $B$: $\lambda_1 = 1$
3. Traditional, One-Level Iterative Methods

- **Power Method**
  \[ x_{i+1} = B x_i \]

- **Weighted Jacobi Method (WJAC)**
  \[ B x = x \]
  \[ (B - I) x = 0 \quad A x = 0 \]
  \[ A = L + D + U \]
  \[ x_{i+1} = N((1 - w) x_i + w D^{-1}(L + U) x_i) \]

**normalization:**
\[ N(x) = \frac{x}{\|x\|_1} \quad (x \neq 0) \]
Traditional, One-Level Iterative Methods

- **Power Method**
  - convergence factor: $|\lambda_2|$
  - convergence is very slow when $|\lambda_2| \approx 1$

- **WJAC**: similar
4. Aggregation

- form three coarse, aggregated states

\[ x_{c,I} = \sum_{i \in I} x_i \]

\[ x_c^T = [8/19 \ 10/19 \ 1/19] \]

\[ B_c x_c = x_c \]

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

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

(Simon and Ando, 1961)
Aggregation

\[ B_c x_c = x_c \]

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

\[ x_c = P^T x \]

\[ B_c = P^T B \text{diag}(x) P \text{diag}(P^T x)^{-1} \]

(Krieger, Horton, ...)

\[
P = \begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]
5. Two-Level Acceleration by Aggregation

**Algorithm:** Two-level acceleration by aggregation

choose initial guess \( x \)

repeat

\[ x \leftarrow N(\text{Relax}(A, x)) \text{ \( \nu \) times} \]

\[ A_c = P^T A \text{ diag}(x) P \text{ diag}(P^T x)^{-1} \]

\[ x_c \leftarrow \text{solve } A_c x_c = 0, \|x_c\|_1 = 1 \quad (\text{coarse-level solve}) \]

\[ x = N(\text{diag}(P \text{ diag}(P^T x)^{-1} x_c) x) \quad (\text{coarse-level correction}) \]

end

≈ Iterative Aggregation/Disaggregation Algorithm (IAD)

UBC CS May 2007
hdesterck@uwaterloo.ca
6. Relation with Multigrid for PDEs

\[-u_{xx} - u_{yy} = f(x, y)\]

- high-frequency error is removed by relaxation
- low-frequency-error needs to be removed by coarse-grid correction
multigrid hierarchy: V-cycle

- multigrid V-cycle:
  - relax (=smooth) on various grids
  - transfer error using restriction ($P^T$) and interpolation ($P$)
7. Multilevel Adaptive Aggregation (MAA) Method

**Algorithm:** Multilevel Adaptive Aggregation method (V-cycle)

\[ x = \text{MAA}_V(A, x, \nu_1, \nu_2) \]

\[
\begin{align*}
\text{begin} \\
\quad x &\leftarrow N(\text{Relax}(A, x)) \quad \nu_1 \text{ times} \\
\quad \text{build } P \text{ based on } x \text{ and } A \quad (P \text{ is rebuilt every level and cycle}) \\
\quad A_c &= P^T A \text{ diag}(x) P \text{ diag}(P^T x)^{-1} \\
\quad x_c &= \text{MAA}_V(A_c, N(P^T x, \nu_1, \nu_2)) \quad \text{(coarse-level solve)} \\
\quad x &= N(\text{diag}(P \text{ diag}(P^T x)^{-1} x_c) x) \quad \text{(coarse-level correction)} \\
\quad x &\leftarrow N(\text{Relax}(A, x)) \quad \nu_2 \text{ times} \\
\text{end}
\]

(Krieger, Horton 1994, but no satisfactory way to build P)
8. Choosing Aggregates Based on Strength

- error equation: \( A \text{diag}(x_i) e_i = 0 \)
- use strength of connection in \( A \text{diag}(x_i) \)
- define row-based strength (determine all states that strongly influence a row’s state)
- state that has largest value in \( x_i \) is seed point for new aggregate, and all unassigned states influenced by it join its aggregate
- repeat

(similar to Brandt, McCormick and Ruge, 1983)
9. Web Matrix Regularization

- web adjacency matrix $G$, then $B = N(G)$
- needs to be made irreducible, a-periodic
- need to add in extra links, with ‘coupling factor’ $\alpha$
Web Matrix Regularization

- PageRank (used by Google):

\[ B_{PR} = (1 - \alpha) N(G + e d^T) + \alpha N(e e^T) \]

\( \alpha = 0.15 \)
Web Matrix Regularization

- BackLink (to root page):

\[ B_{BL} = N((1 - \alpha - \epsilon) N(G') + \alpha e^{(1)} e^T + \epsilon I) \]
Web Matrix Regularization

- **BackButton** (add reverse of each link):

\[
B_{BB} = N((1 - \alpha - \epsilon) N(G) + \alpha N(G^T) + \epsilon I)
\]
Web Matrix Regularization

- second eigenvalue for PageRank: $\leq 1 - \alpha$
  \(\alpha = 0.15\) and \(\alpha = 0.01\)
Web Matrix Regularization

- second eigenvalue for BackLink ($\alpha=0.15$ and $\alpha=0.01$):
Web Matrix Regularization

- second eigenvalue for BackButton ($\alpha=0.15$ and $\alpha=0.01$):
Web Matrix Regularization

- second eigenvalue for BackButton:

\[ 1 - |\lambda_2| \approx O(1/n) \]
9. Performance of MAA

- total efficiency factor of MAA relative to WJAC:

\[ f_{MAA-WJAC}^{(tot)} = \frac{\log(r_{MAA}/t_{MAA})}{\log(r_{WJAC}/t_{WJAC})} \]

- \( f^{(tot)} = 2 \): MAA 2 times more efficient than WJAC

- \( f^{(tot)} = 1/2 \): MAA 2 times less efficient than WJAC
## Performance of MAA

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\gamma_{MAA}$</th>
<th>$\gamma_{WJAC}$</th>
<th>$\gamma_{WJAC}$</th>
<th>$f^{(tot)}_{MAA-WJAC}$</th>
<th>$f^{(as)}_{MAA-WJAC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PageRank, $\alpha = 0.15$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>0.355124</td>
<td>10</td>
<td>1.67</td>
<td>0.815142</td>
<td>1/2.74</td>
</tr>
<tr>
<td>4000</td>
<td>0.335889</td>
<td>9</td>
<td>1.67</td>
<td>0.805653</td>
<td>1/2.52</td>
</tr>
<tr>
<td>8000</td>
<td>0.387411</td>
<td>9</td>
<td>1.65</td>
<td>0.821903</td>
<td>1/2.79</td>
</tr>
<tr>
<td>16000</td>
<td>0.554686</td>
<td>12</td>
<td>1.78</td>
<td>0.836429</td>
<td>1/4.07</td>
</tr>
<tr>
<td>32000</td>
<td>0.502008</td>
<td>11</td>
<td>1.83</td>
<td>0.833367</td>
<td>1/3.94</td>
</tr>
<tr>
<td>64000</td>
<td>0.508482</td>
<td>11</td>
<td>1.75</td>
<td>0.829696</td>
<td>1/3.86</td>
</tr>
<tr>
<td>128000</td>
<td>0.532518</td>
<td>12</td>
<td>1.75</td>
<td>0.829419</td>
<td>1/4.31</td>
</tr>
<tr>
<td>PageRank, $\alpha = 0.01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>0.321062</td>
<td>10</td>
<td>1.77</td>
<td>0.956362</td>
<td>3.42</td>
</tr>
<tr>
<td>4000</td>
<td>0.658754</td>
<td>20</td>
<td>1.75</td>
<td>0.980665</td>
<td>2.16</td>
</tr>
<tr>
<td>8000</td>
<td>0.758825</td>
<td>22</td>
<td>1.65</td>
<td>0.976889</td>
<td>1.88</td>
</tr>
<tr>
<td>16000</td>
<td>0.815774</td>
<td>27</td>
<td>1.77</td>
<td>0.979592</td>
<td>1.45</td>
</tr>
<tr>
<td>32000</td>
<td>0.797182</td>
<td>29</td>
<td>1.82</td>
<td>0.979881</td>
<td>1.35</td>
</tr>
<tr>
<td>64000</td>
<td>0.786973</td>
<td>33</td>
<td>1.79</td>
<td>0.980040</td>
<td>1.19</td>
</tr>
<tr>
<td>128000</td>
<td>0.854340</td>
<td>38</td>
<td>1.72</td>
<td>0.980502</td>
<td>1.05</td>
</tr>
</tbody>
</table>
## Performance of MAA

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\gamma_{MAA}$</th>
<th>$it_{MAA}$</th>
<th>$c_{\text{grid,MAA}}$</th>
<th>$\gamma_{WJAC}$</th>
<th>$f_{MAA-WJAC}^{(\text{tot})}$</th>
<th>$f_{MAA-WJAC}^{(\text{as})}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BackLink, $\alpha = 0.15$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>0.331226</td>
<td>11</td>
<td>1.67</td>
<td>0.839540</td>
<td>1/3.11</td>
<td>1/3.04</td>
</tr>
<tr>
<td>4000</td>
<td>0.344225</td>
<td>11</td>
<td>1.75</td>
<td>0.851397</td>
<td>1/3.30</td>
<td>1/3.18</td>
</tr>
<tr>
<td>8000</td>
<td>0.361255</td>
<td>11</td>
<td>1.69</td>
<td>0.858532</td>
<td>1/3.04</td>
<td>1/3.24</td>
</tr>
<tr>
<td>16000</td>
<td>0.358282</td>
<td>11</td>
<td>2.03</td>
<td>0.866344</td>
<td>1/3.75</td>
<td>1/4.11</td>
</tr>
<tr>
<td>32000</td>
<td>0.369351</td>
<td>11</td>
<td>2.26</td>
<td>0.868116</td>
<td>1/3.99</td>
<td>1/4.39</td>
</tr>
<tr>
<td>64000</td>
<td>0.368789</td>
<td>11</td>
<td>1.88</td>
<td>0.868889</td>
<td>1/3.30</td>
<td>1/3.53</td>
</tr>
<tr>
<td>128000</td>
<td>0.369744</td>
<td>11</td>
<td>1.78</td>
<td>0.871525</td>
<td>1/3.07</td>
<td>1/3.22</td>
</tr>
<tr>
<td>BackLink, $\alpha = 0.01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>0.452383</td>
<td>16</td>
<td>1.89</td>
<td>0.952865</td>
<td>2.01</td>
<td>1/1.21</td>
</tr>
<tr>
<td>4000</td>
<td>0.778003</td>
<td>28</td>
<td>1.76</td>
<td>0.953782</td>
<td>1.41</td>
<td>1/4.23</td>
</tr>
<tr>
<td>8000</td>
<td>0.749847</td>
<td>20</td>
<td>1.72</td>
<td>0.970096</td>
<td>2.23</td>
<td>1/2.23</td>
</tr>
<tr>
<td>16000</td>
<td>0.745776</td>
<td>23</td>
<td>1.96</td>
<td>0.976919</td>
<td>1.87</td>
<td>1/2.11</td>
</tr>
<tr>
<td>32000</td>
<td>0.855323</td>
<td>28</td>
<td>1.93</td>
<td>0.981223</td>
<td>1.66</td>
<td>1/3.04</td>
</tr>
<tr>
<td>64000</td>
<td>0.868049</td>
<td>32</td>
<td>1.96</td>
<td>0.983076</td>
<td>1.45</td>
<td>1/3.15</td>
</tr>
<tr>
<td>128000</td>
<td>0.837747</td>
<td>31</td>
<td>1.83</td>
<td>0.985161</td>
<td>1.65</td>
<td>1/2.09</td>
</tr>
</tbody>
</table>
## Performance of MAA

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\gamma_{MAA}$</th>
<th>$it_{MAA}$</th>
<th>$c_{grid,MAA}$</th>
<th>$\gamma_{WJAC}$</th>
<th>$f^{(tot)}_{MAA-WJAC}$</th>
<th>$f^{(as)}_{MAA-WJAC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BackButton, $\alpha = 0.15$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>0.746000</td>
<td>35</td>
<td>1.74</td>
<td>0.981331</td>
<td>2.36</td>
<td>1/1.41</td>
</tr>
<tr>
<td>4000</td>
<td>0.800454</td>
<td>39</td>
<td>1.64</td>
<td>0.982828</td>
<td>2.70</td>
<td>1/1.36</td>
</tr>
<tr>
<td>8000</td>
<td>0.786758</td>
<td>40</td>
<td>1.53</td>
<td>0.992129</td>
<td>3.15</td>
<td>1.17</td>
</tr>
<tr>
<td>16000</td>
<td>0.851671</td>
<td>50</td>
<td>1.62</td>
<td>0.992330</td>
<td>3.00</td>
<td>1/1.38</td>
</tr>
<tr>
<td>32000</td>
<td>0.988423</td>
<td>214</td>
<td>1.64</td>
<td>0.998366</td>
<td>4.92</td>
<td>1/2.88</td>
</tr>
<tr>
<td>64000</td>
<td>0.973611</td>
<td>185</td>
<td>1.59</td>
<td>0.999013</td>
<td>9.95</td>
<td>1.40</td>
</tr>
<tr>
<td>128000</td>
<td>0.943160</td>
<td>116</td>
<td>1.55</td>
<td>0.999693</td>
<td>34.64</td>
<td>9.90</td>
</tr>
<tr>
<td>BackButton, $\alpha = 0.01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>0.658032</td>
<td>23</td>
<td>1.68</td>
<td>0.999563</td>
<td>106.02</td>
<td>46.05</td>
</tr>
<tr>
<td>4000</td>
<td>0.794123</td>
<td>29</td>
<td>1.71</td>
<td>0.999345</td>
<td>73.02</td>
<td>19.78</td>
</tr>
<tr>
<td>8000</td>
<td>0.841182</td>
<td>39</td>
<td>1.70</td>
<td>0.997624</td>
<td>23.49</td>
<td>2.64</td>
</tr>
<tr>
<td>16000</td>
<td>0.835592</td>
<td>44</td>
<td>1.78</td>
<td>0.998696</td>
<td>19.72</td>
<td>4.42</td>
</tr>
<tr>
<td>32000</td>
<td>0.845457</td>
<td>56</td>
<td>1.83</td>
<td>0.999114</td>
<td>39.58</td>
<td>8.22</td>
</tr>
<tr>
<td>64000</td>
<td>0.959561</td>
<td>81</td>
<td>1.75</td>
<td>0.999660</td>
<td>75.05</td>
<td>5.74</td>
</tr>
<tr>
<td>128000</td>
<td>0.921870</td>
<td>42</td>
<td>1.70</td>
<td>0.999963</td>
<td>816.62</td>
<td>103.79</td>
</tr>
</tbody>
</table>
10. Web Matrix Regularizations as a Function of Coupling Factor $\alpha$

- PageRank
Web Matrix Regularizations as a Function of Coupling Factor $\alpha$

- BackLink

![Graph of BackLink](image)
Web Matrix Regularizations as a Function of Coupling Factor $\alpha$

- BackLink
Web Matrix Regularizations as a Function of Coupling Factor $\alpha$

- BackButton
11. Conclusions

- PageRank regularization of webgraph with $\alpha=0.15$ seems to be a pretty good model for web ranking
Conclusions

• MAA cannot beat Power Method speed for PageRank reasons:
  ▪ $\lambda_2=0.85$, independent of $n$
  ▪ there are global connections on all scales
  ▪ Power Method scales perfectly (also in parallel)
  ▪ $0.85^{10}=0.2$  $0.85^{20}=0.04$

still:
  ▪ MAA provides information about how web pages are clustered, at all levels
Conclusions

• MAA can dramatically outperform Power Method for Markov Chains for which $|\lambda_2(n)| \rightarrow 1$ for large $n$

reason:
  ▪ multilevel nature of MAA allows to bridge scales (Markov Chain has only local connections in this case)
Comparison: Strength-based Aggregation

• major differences with previously considered multilevel methods for Markov chains (Horton, Krieger,...):
  ▪ we use AMG strength of connection based on scaled problem matrix $A \text{diag}(x_i)$
  ▪ our aggregates based on columns of strength matrix (~ AMG coarsening)

• IAD methods
  ▪ normally only 2-level
  ▪ aggregates fixed, based on previously known, regular structure of Markov chain
Two-Level Acceleration by Aggregation

- multiplicative correction: error equation

\[ x = \text{diag}(x_i) e_i \]

\[ A \text{diag}(x_i) e_i = 0 \]

\[ P^T A \text{diag}(x_i) P e_c = 0 \]

\[ x_c = \text{diag}(P^T x_i) e_c \]

\[ A_c x_c = 0 \]

\[ P^T A \text{diag}(x) P \text{diag}(P^T x)^{-1} x_c = 0 \]