Quadratic Vector Equations and Multilinear Pagerank

Federico Poloni
Joint work with B. Meini

University of Pisa

Networks: from Matrix Functions to Quantum Physics
Oxford, August 2017

Pagerank extensions

Pagerank [Page '98]

Input: transition probabilities $P_{ij} = P[i \to j]$, 'personalization vector' v.
Output: 'importance' score x_i of each node

$$x = \alpha P x + (1-\alpha) v, \qquad \mathbf{1}^\top x = 1.$$

Multilinear Pagerank [Gleich-Lim-Yu '15]

Input: two-step transition probabilities $P_{ijk} = P[i \to j \to k]$, 'personalization vector' v.
Output: 'importance' score x_i of each node

$$x = \alpha \sum_{i,j=1}^{n} P_{ij:} x_i x_j + (1-\alpha) v, \qquad \mathbf{1}^\top x = 1.$$

(Not just a 2^{nd}-order Markov model: that would rank pairs of nodes, X_{ij}).

Back to simpler models

A toy problem

A cell splits into two identical ones with probability p, or dies without reproducing with probability $1 - p$. Starting from a single cell, what is the probability that the whole colony eventually becomes extinct?

Back to simpler models

A toy problem

A cell splits into two identical ones with probability p, or dies without reproducing with probability $1 - p$. Starting from a single cell, what is the probability that the whole colony eventually becomes extinct?

dies immediately — splits in two, and they both get extinct

$$x = (1 - p) + px^2$$

The extinction probability x is a solution of this equation – but which one? $x = 1$ or the other?

Looking at the whole story

$x^{(h)} = P$[extinction within h generations] satisfies

$$x^{(0)} = 0, \quad x^{(h+1)} = (1-p) + p(x^{(h)})^2.$$

Easy to show that the sequence $x^{(0)} \leq x^{(1)} \leq x^{(2)} \leq \dots$ converges to the smaller solution of $x = (1-p) + px^2$.

Answer to the riddle

The colony dies out with probability $\begin{cases} \frac{1-p}{p} & p > \frac{1}{2}, \\ 1 & p \leq \frac{1}{2}. \end{cases}$

What does this simple problem tell us about matrices and tensors?

A vector version

Now we have cells of n types:

$$\begin{aligned} P_{ijk} &= P[\text{cell of type } k \text{ splits into } i \text{ and } j], \\ v_k &= P[\text{cell of type } k \text{ dies without offspring}] = 1 - \sum_{i,j} P_{ijk}, \\ x_k &= P[\text{colony starting from one type-}k \text{ cell dies out}]. \end{aligned}$$

$$x = \sum_{i,j=1}^{n} P_{ij:} x_i x_j + v$$

[Kolmogorov 1940s, Etessami-Yannakakis '05, Bean-Kontoleon-Taylor '08]

Quadratic vector equations

$$x = \sum_{i,j=1}^{n} P_{ij:} x_i x_j + v \qquad (*)$$

Many properties are similar to those of the scalar version:

- $x = \mathbf{1}$ is a solution.
- The corresponding fixed-point equation converges monotonically (starting from $x^{(0)} = \mathbf{0}$).
- The extinction probabilities are given by its limit point x^*. Every other solution of (*) is entrywise larger than x^* (minimal solution).
- Many other fixed-point recurrences (e.g., Newton's method, Gauss-Seidel-like variants. . .) converge monotonically to x^* as well.

[Hautphenne-Latouche-Remiche '11, Etessami-Stewart-Yannakakis '12, P '13]

Numerical experiments

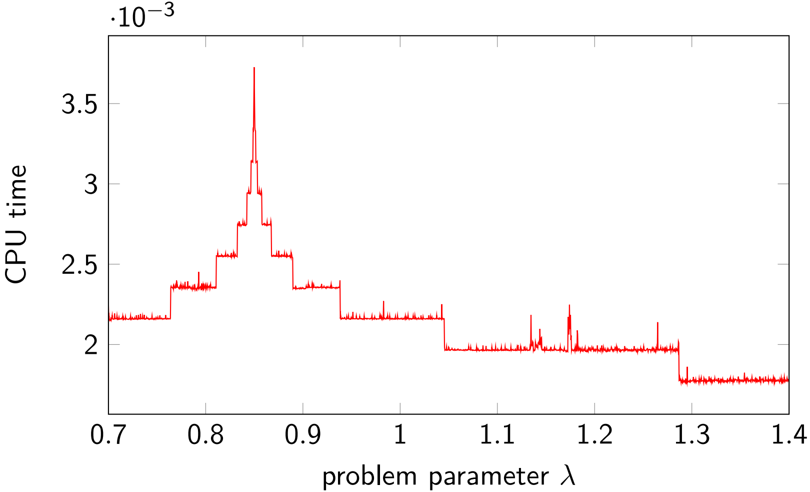

Figure: CPU time for Newton's method on a parameter-dependent problem [Bean-Kontoleon-Taylor '08, Ex. 1]; lower=better

Critical problems

Why the spike? Minimal solution $x^* \approx \mathbf{1}$ for that parameter value.

Two very close solutions: double = trouble.

Critical problems

Why the spike? Minimal solution $x^* \approx \mathbf{1}$ for that parameter value.

Two very close solutions: double = trouble.

Idea

Is it possible to deflate the known solution $x = \mathbf{1}$?

Critical problems

Why the spike? Minimal solution $x^* \approx \mathbf{1}$ for that parameter value.

Two very close solutions: double = trouble.

Idea

Is it possible to deflate the known solution $x = \mathbf{1}$?

Yes! (somehow) Let $y = \mathbf{1} - x$, survival probability. Eqn becomes

$$y = H_y y, \qquad (H_y)_{ik} = \sum_j P_{ijk} + P_{jik} - P_{jik} y_j$$

y = Perron eigenvector of the nonnegative matrix H_y.

Perron iteration [Meini-P '11], [Bini-Meini-P '12]

- $y^{(k)}$ = Perron (maximal) eigenvector of $H_{y^{(k-1)}}$.
- Normalize $y^{(k)}$ (s.t. $x = \mathbf{1} - \alpha y^{(k)}$ approximate solution of the eqn).
- Iterate.

Numerical experiments

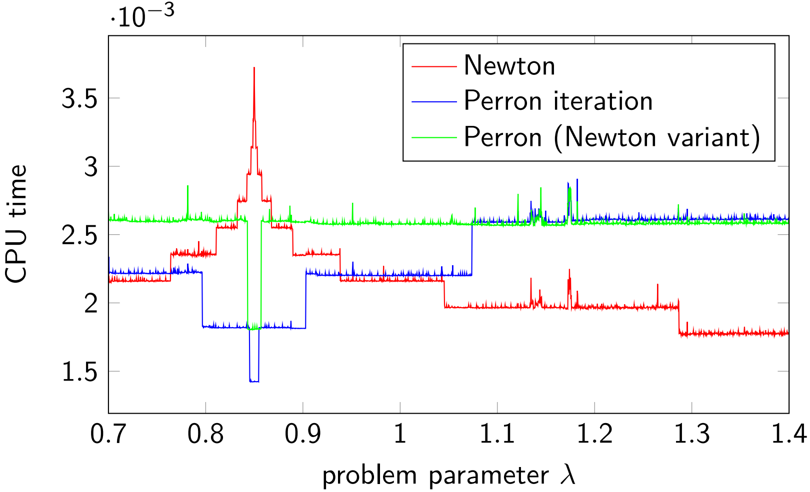

Figure: CPU times on a parameter-dependent problem [Bean-Kontoleon-Taylor '08, Ex. 1]; lower=better

Back to Multilinear Pagerank

Multilinear Pagerank [Gleich-Lim-Yu '15]

Input: two-step transition probabilities $P_{ijk} = P[i \to j \to k]$, 'personalization vector' v.

Output: 'importance' score x_i of each node

$$x = \alpha \sum_{i,j=1}^{n} P_{ij:} x_i x_j + (1-\alpha) v, \qquad \mathbf{1}^\top x = 1.$$

(Not just a 2[nd]-order Markov model: that would rank pairs of nodes, X_{ij}).

Very similar equation. Main differences:

- $\mathbf{1}$ no longer a solution;
- We seek a stochastic solution, which is not necessarily minimal.

Structure of the solutions

$$g(x) = \alpha \sum_{i,j=1}^{n} P_{ij:} x_i x_j + (1-\alpha) v$$

has predictable behaviour on the 'mass' of x: if $\mathbf{1}^\top x = w$, then $\mathbf{1}^\top g(x) = \alpha w^2 + (1-\alpha)$.

Consequence: every fixed-point x has $\mathbf{1}^\top x = 1$ or $\mathbf{1}^\top x = \frac{1-\alpha}{\alpha}$.

Theorem

Consider the iteration

$$x^{(k)} = g(x^{(k-1)}), \qquad x^{(0)} = \mathbf{0}.$$

- If $\alpha \leq \frac{1}{2}$, $x^{(k)} \to x^*$, the unique minimal solution with $\mathbf{1}^\top x^* = 1$.
- If $\alpha > \frac{1}{2}$, $x^{(k)} \to x^*$, the unique minimal solution with $\mathbf{1}^\top x^* = \frac{1-\alpha}{\alpha}$. There may be several solutions $x \geq x^*$ with $\mathbf{1}^\top x = 1$.

Uniqueness (or not) of stochastic solutions also in [Gleich-Lim-Yu '15].

Numerical experiments

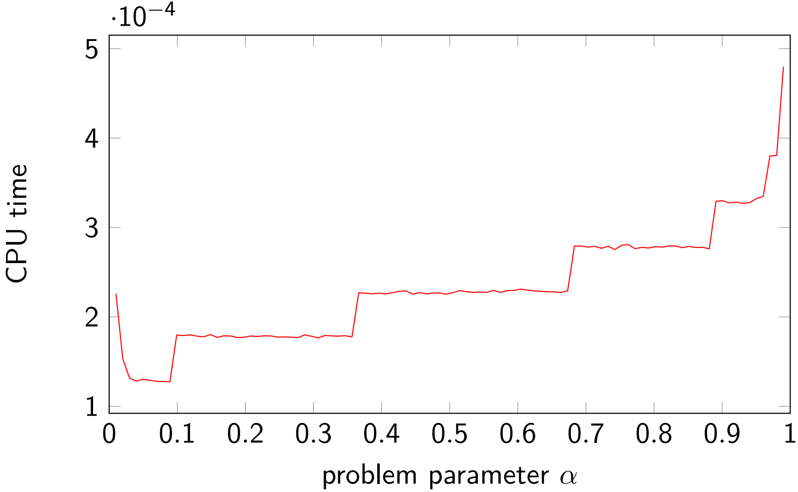

Figure: CPU time for Newton's method on a parameter-dependent multilinear pagerank problem [Gleich-Lim-Yu '5, Ex. R6_5]; lower=better.

Large and small α

- The good When $\alpha \leq \frac{1}{2}$, there is a unique stochastic solution, and many fixed point iterations (e.g. Newton's method, Gauss-Seidel-like variants...) converge to it monotonically.
- The bad When $\alpha > \frac{1}{2}$, there may be multiple stochastic solutions, and convergence may be problematic — even if we enforce $\mathbf{1}^\top x^{(k)} = 1$.

Large and small α

- The good When $\alpha \leq \frac{1}{2}$, there is a unique stochastic solution, and many fixed point iterations (e.g. Newton's method, Gauss-Seidel-like variants...) converge to it monotonically.
- The bad When $\alpha > \frac{1}{2}$, there may be multiple stochastic solutions, and convergence may be problematic — even if we enforce $\mathbf{1}^\top x^{(k)} = 1$.
- The ugly Perron-based iterations can be used for the bad case $\alpha > \frac{1}{2}$:
  - Compute minimal (sub-stochastic) solution x^*;
  - Change variable $y = x - x^*$;
  - Interpret resulting equation as $y = H_y y$;
  - Fixed-point iteration: $y^{(k)} =$ Perron eigenvector of $H_{y^{(k-1)}}$ (or variants)

(Another algorithm involving Perron vectors is in [Benson-Gleich-Lim '17].)

Our goal

[Gleich-Lim-Yu '15] contains 29 small-size benchmark problems ($n \in \{3, 4, 6\}$), some of them with difficult convergence.

The best methods there (`newton` and `innout`) can solve 23 and 26 of them, respectively.

Goal: develop a numerical method that can reliably solve all of them (in a reasonable number of iterations).

Example (`R6_3`, [Gleich-Lim-Yu '15])

$$\frac{1}{4}\left[\begin{array}{cccccc|cccccc|cccccc|cccccc|cccccc|cccccc}
0&0&0&0&0&0&0&0&0&0&0&2&1&0&0&0&0&0&0&0&4&0&4&0&0&2&4&0&0&0&0&0&0&0&0&4\\
0&0&0&0&0&0&0&0&4&0&0&0&0&4&0&0&0&0&0&0&0&0&0&0&0&2&0&0&2&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&4&0&0&0&2&0&4&2&0&0&2&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&4&0&4&0&0&0&0&0&0&0&2&0&4&0&0\\
0&0&0&0&0&0&0&4&0&0&0&2&0&0&0&0&0&0&0&4&0&0&0&0&0&0&0&4&0&0&0&0&4&0&2&0\\
4&4&4&4&4&4&4&0&0&4&4&0&1&0&4&4&4&4&4&0&0&0&0&0&0&0&0&0&0&4&0&0&0&0&0&0
\end{array}\right]$$

Improvements

Improvement 1: use Newton's method on $y - \mathrm{pvec}(H_y) = 0$, where pvec is the map that computes the Perron vector of a matrix.

An expression for the Jacobian can be found using eigenvector derivatives.

Theorem [Meini-P '17]

The Jacobian of the map $w = \mathrm{pvec}(H_y)$ is

$$J = \alpha(w\mathbf{1}^\top - (I - H_y + w\mathbf{1}^\top)^{-1} \sum_j P_{:j:} w_j)$$

(similar to [Bini-Meini-P '11] for the extinction probability problem.)

Improvements

Improvement 2 Use homotopy continuation techniques: first solve the problem for an ‘easy’ α, then increase its value slowly.

Theorem [Meini-P ’17]

Let x_α be the solution vector for a certain value of $\alpha \in (0,1)$. Then,

$$x_{\alpha+h} = x + \left(I - \alpha(\sum_j P_{:j:} + P_{j::}) \right)^{-1} \left(v - \sum P_{ij:}(x_\alpha)_i (x_\alpha)_j \right) h + O(h^2).$$

Step-size heuristic: estimate the neglected second-order term $\frac{\mathrm{d}x_\alpha}{\mathrm{d}\alpha^2}$, and use it to choose the next step size.

Numerical results

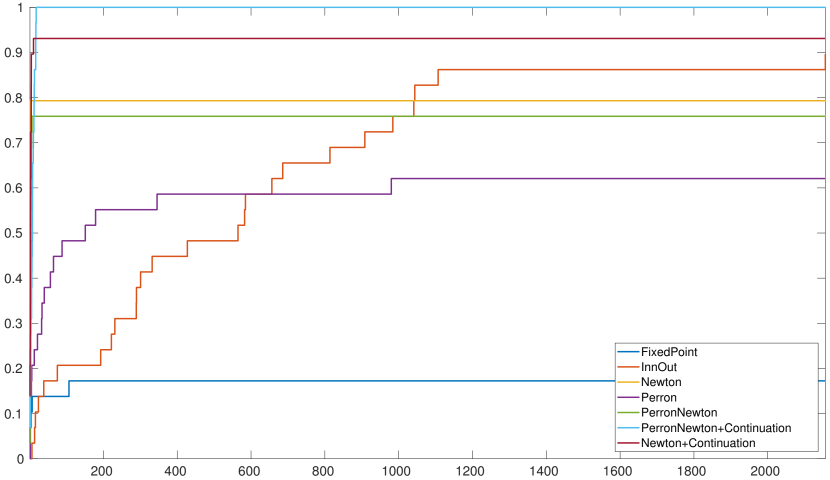

Figure: Performance profile for the 29 examples with $\alpha = 0.99$

Numerical results

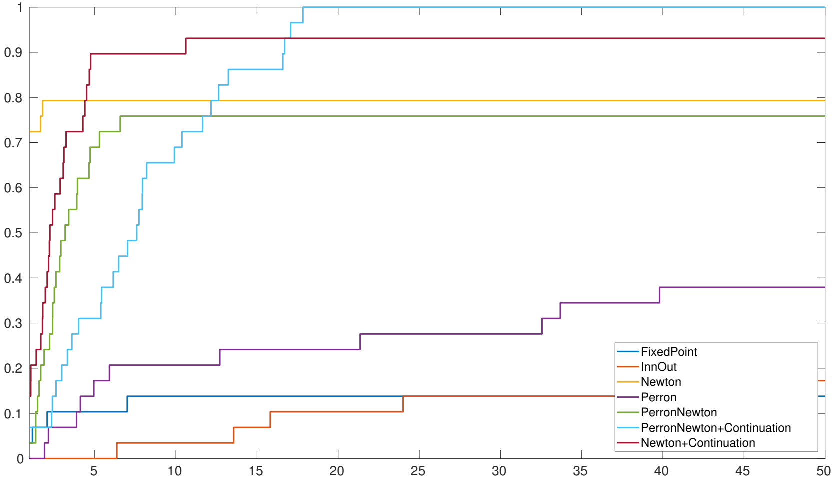

Figure: Zoomed performance profile for the 29 examples with $\alpha = 0.99$

Conclusions

- More understanding for multilinear pagerank problems — from analogy with population models.
- New numerical strategies: Perron-based methods, continuation.
- Can handle all benchmark problems successfully.
- TO-DO: test at real-world scale.

Conclusions

- More understanding for multilinear pagerank problems — from analogy with population models.
- New numerical strategies: Perron-based methods, continuation.
- Can handle all benchmark problems successfully.
- TO-DO: test at real-world scale.

Thanks for your attention!