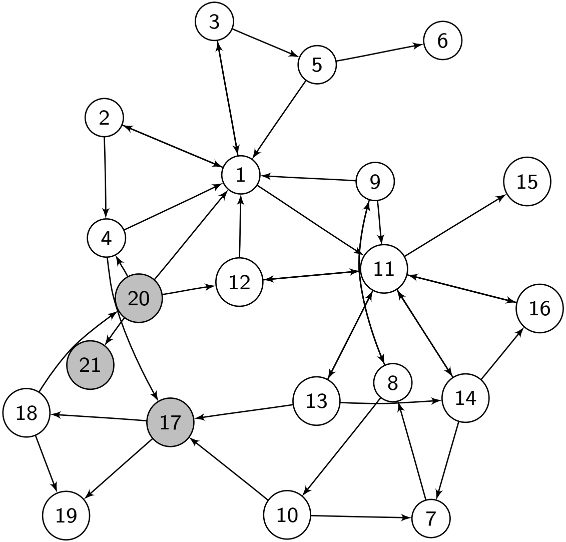
Optimization of the HOTS score of a website's pages

Olivier Fercoq and Stéphane Gaubert

INRIA Saclay and CMAP Ecole Polytechnique

June 21st, 2012

Toy example with 21 pages



Nodes = web pages

Arcs = hyperlinks

(21) : controlled page

(1) : non controlled page

Context

A webmaster controls a given number of pages:

- May add links
- Must respect the content
- Wishes to maximize:
  - Sum of PageRank values of the site
  - HITS authority score of the home page
  - Sum of HOTS score values of the site

PageRank: Brin and Page, 1998
HITS: Kleinberg, 1998
HOTS: Tomlin, 2003

Tomlin's HOTS algorithm: irreducible case

A: adjacency matrix of the web graph (irreducible)
ρ: web traffic

$$\max_{\rho \geq 0} - \sum_{i,j \in [n]} \rho_{i,j} (\log(\frac{\rho_{i,j}}{A_{i,j}}) - 1)$$

$$\sum\nolimits_{j \in [n]} \rho_{i,j} = \sum\nolimits_{j \in [n]} \rho_{j,i} \ , \ \forall i \in [n] \qquad (p_i)$$

$$\sum\nolimits_{i,j \in [n]} \rho_{ij} = 1 \qquad (\mu)$$

Optimal ρ while PageRank gives a specific ρ
(uniform probability is arbitrary)

Dual problem: irreducible case

Minimize: $$\theta(p,\mu) := \sum_{i,j\in[n]} A_{ij} e^{p_i - p_j + \mu} - \mu \ .$$

θ is convex and differentiable.

$$\frac{\partial\theta}{\partial\mu}(p,\mu) = 0 \Rightarrow \mu = -\log\Big(\sum_{i,j\in[n]} A_{ij} e^{p_i - p_j}\Big)$$

$$\frac{\partial\theta}{\partial p}(p,\mu) = 0 \Rightarrow p \text{ is a fixed point of } f, \text{ where}$$

$$f_i(x) = \frac{1}{2}\log(A^T e^x)_i - \frac{1}{2}\log(Ae^{-x})_i \qquad g_i(d) = \left(\frac{(A^T d)_i}{(Ad^{-1})_i}\right)^{1/2}$$

e^{p_i} is interpreted as the temperature of page i

The matrix balancing problem

Given a $n \times n$ matrix A, find a diagonal positive matrix D such that $X = DAD^{-1}$ verifies

$$\sum_{j \in [n]} X_{i,j} = \sum_{k \in [n]} X_{k,i} \qquad \forall i \in [n]$$

Proposition (Eaves, Hoffman, Rothblum, H Schneider, 1985)

There exists $v \in \mathbb{R}^n$ such that $f(v) = v$ and $\sum_{i \in [n]} v_i = 0$ if and only if A has a diagonal similarity scaling if and only if A is completely reducible.

If in addition A is irreducible, then v is unique.

v minimizes $\theta_0(p) = \theta(p, 0)$

Algorithms for matrix balancing

- Convex optimization algorithms
- Coordinate descent: (M Schneider and Zenios, 1989) select a coordinate i and set

$$d_i \leftarrow \left(\frac{(A^T d)_i}{(Ad^{-1})_i} \right)^{\frac{1}{2}}$$

- DomEig (Johnson, Pitkin, Stanford, 2000)

$$\min \left\{ \lambda_{\max}(A + \mathrm{diag}(v)) \mid \sum_{i \in [n]} v_i = 0 \right\} = \min_{p \in \mathbb{R}^n} \theta_0(p)$$

Fixed point approach for matrix balancing

- Let f and g be the functions defined by

$$f_i(x) = \frac{1}{2}\log(A^T e^x)_i - \frac{1}{2}\log(Ae^{-x})_i$$

$$g_i(d) = \left(\frac{(A^T d)_i}{(Ad^{-1})_i}\right)^{1/2}$$

- The solution of the matrix balancing problem verifies

$$x = f(x) \text{ or } d = g(d)$$

- Ideal HOTS algorithm (irreducible case):

$$x_{k+1} = f(x_k)$$

- Does it converge ? (Knight, 2008: not proved)

Convergence of the fixed point scheme

The fixed point operator

$$f(x) = \frac{1}{2}(\log(A^T e^x) - \log(Ae^{-x}))$$

is monotone, additively homogeneous.

Theorem

If A is irreducible and $A + A^T$ is primitive, then $\forall x \in \mathbb{R}^n$,

$$\limsup_{k \to \infty} \left(\| f^k(x) - v \| \right)^{\frac{1}{k}} \leq |\lambda_2(P)|$$

$$P = \frac{1}{2} \left(\mathrm{diag}(A^T e^v)^{-1} A^T \mathrm{diag}(e^v) + \mathrm{diag}(Ae^{-v})^{-1} A \mathrm{diag}(e^{-v}) \right)$$

Proof

Nonlinear Perron-Frobenius theory (Nussbaum and followers)

Comparison of algorithms

	$A = \begin{bmatrix} \epsilon & 1 \\ 2 & 0 \end{bmatrix}$	CMAP 1,500 p	NZ Uni 413,639 p
$\lvert\lambda_2(P)\rvert$	0.9993	0.8739	0.9774
Matlab's fminunc	0.015 s	948 s	Out of memory
DomEig	0.5 s	$>$ 600 s	$>$ 600 s
Coordinate descent	0.001 s	0.03 s	6.06 s
Fixed point (HOTS)	0.004 s	0.02 s	7.52 s

A is nearly imprimitive.

CMAP website and surroundings

NZ Uni dataset: New Zealand Universities

Irreducibility by adding small positive values to all the entries

Tomlin's HOTS: handling reduciblity

Network flow model with constraints on the modified network

$$A' = \begin{bmatrix} A & 1 \\ 1^T & 0 \end{bmatrix}$$

$$\max_{\rho \geq 0} \; -\sum_{i,j \in [n+1]} \rho_{i,j} \left(\log\left(\frac{\rho_{i,j}}{A'_{i,j}}\right) - 1 \right)$$

$$\sum_{j \in [n+1]} \rho_{i,j} = \sum_{j \in [n+1]} \rho_{j,i} \;, \; \forall i \in [n+1] \qquad (p_i)$$

$$\sum_{i,j \in [n+1]} \rho_{ij} = 1 \qquad (\mu)$$

$$\sum_{j \in [n]} \rho_{n+1,j} = 1 - \alpha \qquad (a)$$

$$1 - \alpha = \sum_{i \in [n]} \rho_{i,n+1} \qquad (b)$$

Dual function

$$\theta(p,\mu,a,b) = \sum_{i,j\in[n]} A_{ij} e^{p_i - p_j + \mu} + \sum_{i\in[n]} e^{-b-p_{n+1}+p_i+\mu}$$
$$+ \sum_{j\in[n]} e^{a+p_{n+1}-p_j+\mu} - (1-\alpha)a - \mu + (1-\alpha)b$$
$$\mu(p) = \log\left(\frac{2\alpha-1}{\sum_{i,j\in[n]} A_{i,j} e^{p_i-p_j}}\right)$$
$$a(p) = \log\left(\frac{1-\alpha}{2\alpha-1} \frac{\sum_{i,j\in[n]} A_{i,j} e^{p_i-p_j}}{\sum_{j\in[n]} e^{p_{n+1}-p_j}}\right)$$
$$b(p) = \log\left(\frac{1-\alpha}{2\alpha-1} \frac{\sum_{i,j\in[n]} A_{i,j} e^{p_i-p_j}}{\sum_{i\in[n]} e^{p_i-p_{n+1}}}\right)$$

We denote $\lambda(p) := (\mu(p), a(p), b(p))$

Iterative scheme

Define f^λ to be the fixed point operator associated to the matrix balancing problem on the matrix

$$A = \begin{bmatrix} e^\mu A & e^{a+\mu} 1 \\ e^{-b+\mu} 1^T & 0 \end{bmatrix}$$

Tomlin's HOTS algorithm:

$$p_{k+1} = f^{\lambda(p_k)}(p_k) = F(p_k)$$

- F is homogeneous but not monotone
- Expansive in Thomson's metric ($d(x, y) = \max_i x_i - y_i - \min_j x_j - y_j$)

Elements of the proof

Theorem (Lyapounov function)

The ideal HOTS operator verifies $\theta_0(f(p)) \leq \theta_0(p)$

Theorem (local contraction in projective space)

Denote $F(p) = f^{\lambda(p)}(p)$ *and* p^* *such that* $F(p^*) = p^*$*.*
Then all the eigenvalues of $\nabla F(p^*)$ *belong to* $(-1, 1]$
and the eigenvalue 1 is simple.

Convergence of HOTS algorithm

Theorem

If there exists a primal feasible point with the same pattern as A, then the HOTS algorithm converges to the HOTS vector (unique up to an additive constant) with a linear rate of convergence equal to $|\lambda_2(\nabla F)|$.

Proof

Use the Lyapunov function
Prove that all limit points $(\bar{p}, \bar{\lambda})$ minimize $\theta(p, \lambda)$
Conclude thanks to the local contraction in projective space

Optimization of link-based rankings

Effort concentrated on PageRank

- Avrachenkov and Litvak, 2006, one page
- Matthieu and Viennot, 2006, unconstrained problem
- de Kerchove, Ninove, van Dooren, 2008
- Ishii and Tempo, 2010
- Csáji, Jungers and Blondel, 2010
- F., Akian, Bouhtou, Gaubert, 2011

Perron vector optimization, HITS and HOTS optimization

- Fercoq, 2011 (arXiv:1111.2234)

Tomlin's HOTS optimization

Obligatory links $\mathcal{O}$, optional links $\mathcal{F}$, prohibited links $\mathcal{I}$
$N(p) := \log(\sum_{i\in[n]} e^{p_i})$, J is the set of hyperlinks selected
The HOTS optimization problem is:

$$\max_{J\subseteq\mathcal{F}, p\in\mathbb{R}^n} \{ U(p) \ ; \ f^{\lambda(p)}(A(J), p) = p \ , \ N(p) = 0 \ , \}$$

Relaxed HOTS optimization problem:

$$\max_{A\in\mathbb{R}^{n\times n}, p\in\mathbb{R}^n} U(p)$$

$$f^{\lambda(p)}(A, p) = p \ , \quad N(p) = 0$$

$$A_{i,j} = 1 \ , \ \forall(i,j) \in \mathcal{O}$$

$$A_{i,j} = 0 \ , \ \forall(i,j) \in \mathcal{I}$$

$$0 \leq A_{i,j} \leq 1 \ , \ \forall(i,j) \in \mathcal{F}$$

Matrix of partial derivatives

Denote $F(A,p) = f^{\lambda(p)}(A,p)$.

Proposition

The derivative of $U \circ p$ is given by $g_{i,j} = \sum_l w_l \frac{\partial F^l}{\partial A_{i,j}}$ where

$$w = (-\nabla U^T + (\nabla U^T e)\nabla N^T)(\nabla_p F - I)^{\#}$$

Moreover, the matrix $(g_{i,j})_{i,j}$ has rank at most 3.

Proposition

Let $z = -\nabla U^T + (\nabla U^T e)\nabla N^T$ and v s.t. $v^T \nabla_p F = v^T$. The fixed point scheme defined by

$$\forall k \in \mathbb{N}, \quad w_{k+1} = (z + w_k \nabla_p F)(I - \frac{1}{v^T e} e v^T)$$

converges in geometric speed to w.

Polak's Master algorithm model

- For any weighted adjacency matrix A, $J(A) = U(p(A))$
  $J_n(A) = U(p_{k_n})$ where k_n is the first nonnegative integer k such that $\|p_{k+1} - p_k\| \leq \Delta(n)$
  $\mathcal{B}_n(A)$ an approximate gradient iteration
- Let $\omega \in (0,1)$, $\sigma' \in (0,1)$, $n_{-1} \in \mathbb{N}$ and $A_0 \in \mathcal{C}$,
  For $i \in \mathbb{N}$, compute A_{i+1} and the smallest $n_i \in \mathbb{N}$ s.t.

$$
\begin{aligned}
&n_i \geq n_{i-1} \\
&A_{i+1} = \mathcal{B}_{n_i}(A_i) \\
&J_{n_i}(A_{i+1}) - J_{n_i}(A_i) \leq -\sigma'(\Delta(n_i))^\omega
\end{aligned}
$$

Interrupted Armijo line search

Let $(\bar{M}_n)_{n\geq 0}$ be a sequence diverging to $+\infty$, $\sigma \in (0,1)$, $\alpha^0 > 0$, $\beta \in (0,1)$ and $\gamma > 0$.
Given $n \in \mathbb{N}$, $J_n = U(p_{k_n})$ and g_n is an approximate gradient.
Let m_n be the first $\mathbf{m} \in \mathbb{N}$ such that

$$J_n\left(\mathrm{P}_{\mathcal{C}}(A - \beta^{\mathbf{m}}\alpha^0 g_n(A))\right) - J_n(A) \leq -\sigma \frac{\|A - \mathrm{P}_{\mathcal{C}}(A - \beta^{\mathbf{m}}\alpha^0 g_n(A))\|_2^2}{\beta^{\mathbf{m}}\alpha^0}$$

If $m_n \leq \bar{M}_n$, then

$$\mathcal{B}_n(A) = \mathrm{P}_{\mathcal{C}}(x - \beta^{m_n}\alpha^0 g_n(A))$$

Otherwise $\mathcal{B}_n(A) = \emptyset$.

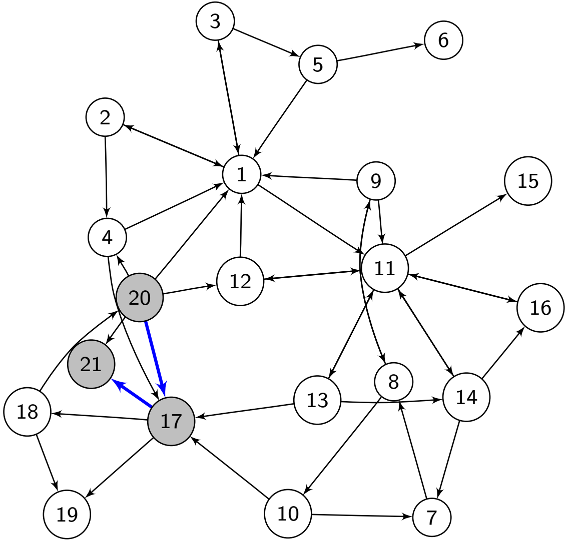
Convergence theorem for HOTS optimization

Theorem

Let $(A_i)_{i \geq 0}$ be a sequence constructed by the Master Algorithm Model for the resolution of the HOTS optimization problem such that $\mathcal{B}_n(A)$ is the Interrupted Armijo line search

Then every accumulation point of $(A_i)_{i \geq 0}$ is a stationary point of the optimization problem.

Web graph optimized for HOTS



added links

HOTS score sum:
$0.142 \rightarrow 0.169$

Numerical results

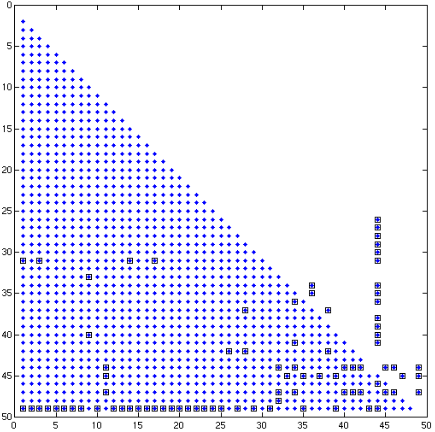
	CMAP	NZ Uni
Gradient (linear system)	0.82 s/it	out of memory
Gradient (iterative scheme)	0.12 s/it	62 s/it
Coupled iter. (master algo.)	0.04 s/it	2.9 s/it

Comparison of algorithms for HOTS optimization

CMAP website: 1,500 pages
New Zealand universities websites: (413,639 pages)

The iterative scheme formulation makes the problem scalable
The Master algorithm model gives a speedup from 3 to 30

Locally optimal solution



Local optimal solution of the HOTS optimization problem on CMAP website (1,500 pages)

Adjacency matrix restricted to the set of controlled pages

Pages sorted by w values

Squares: obligatory links

Conclusion

- General convergence proof: nonexpansive maps, nonlinear Perron-Frobenius theory
- Low rank property of derivatives
  Iterative scheme to compute it
  Scalable optimization algorithm
- Better understanding of HOTS algorithm:
  - quality of the convergence rate
  - spamming techniques / search engine optimizations