Optimization of Perron eigenvectors and applications:

from web ranking to chronotherapeutics

Olivier Fercoq

September 17th, 2012







Outline

Optimization of PageRank

Optimization of other scores

Convergence of HOTS algorithm

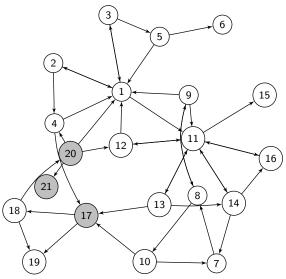
Chronotherapeutics

Web ranking

A webmaster controls a given number of pages:

- May add links
- Must respect the content (the goal of a site is to provide information or service)
- Wishes to maximize:
 - Income (number of clicks on ads, number of sales)
 - Visibility (Sum of PageRank values of the site, PageRank of home page in Google)

Toy example with 21 pages



Nodes = web pages Arcs = hyperlinks

21 : controlled page

1 : non controlled page

Definition of PageRank [Brin and Page, 1998]

- Random web surfer moves from page i to page j with probability $\frac{1}{D_i}$ (D_i = degree of page i)
- $\pi = \text{invariant measure of the Markov chain}$

$$\pi_i = \sum_{j:j\to i} \frac{\pi_j}{D_j}$$

- An important page is a page linked to by important pages
- Markov chain model may be reducible

Definition of PageRank [Brin and Page, 1998]

- Random web surfer moves from page i to page j with probability $\frac{1}{D_i}$ (D_i = degree of page i)
- $\pi = \text{invariant measure of the Markov chain}$

$$\pi_i = \alpha \sum_{j:j\to i} \frac{\pi_j}{D_j} + (1-\alpha)z_i$$

- An important page is a page linked to by important pages
- Markov chain model may be reducible
 - \rightarrow with probability 1α , surfer gets bored and resets: new research from page i with probability z_i
- Transition matrix: $P_{i,j} > 0, \forall i, j$ (usually $\alpha = 0.85$)
- PageRank is the unique invariant measure π of P

The PageRank optimization problem

- Well studied subject: Avratchenkov and Litvak, 2006 Mathieu and Viennot 2006
 De Kerchove, Ninove and Van Dooren 2008
 Csáji, Jungers and Blondel 2010...
- Obligatory links \mathcal{O} , facultative links \mathcal{F} , prohibited links \mathcal{I} (Strategy set proposed by Ishii and Tempo, 2010)
- Utility $\varphi(\pi, P) = \sum_{i} r_{i,j} \pi_i P_{i,j}$
- $r_{i,j}$ is viewed as reward by click on $i \rightarrow j$
- [Fercoq, Akian, Bouhtou, Gaubert, to appear in IEEE TAC]

Reduction to ergodic control

Proposition

 $\mathcal{P}_i = \mathsf{set}$ of admissible transition probabilities from Page i

The PageRank Optimization problem is equivalent to the ergodic control problem with process X_t :

$$\max_{(\nu_t)_{t\geq 0}} \liminf_{T\to +\infty} \frac{1}{T} \mathbb{E}(\sum\nolimits_{t=0}^{T-1} r_{X_t,X_{t+1}})$$

$$\nu_t \in \mathcal{P}_{X_t}, \forall t \geq 0$$

 $\mathbb{P}(X_{t+1} = j | X_t = i, \nu_t = p) = p_j, \forall i, j \in [n], \forall p \in \mathcal{P}_i, \forall t \geq 0$ where ν_t is a function of the history $(X_0, \nu_0, \dots, X_{t-1}, \nu_{t-1}, X_t)$

Exponential size of the action sets

- At each page i, an action corresponds equivalently to
 - select $J \subseteq \mathcal{F}_i$
 - select $\nu \in \mathcal{P}_i$, a uniform measure on J
- 2ⁿ hyperlink configurations by controlled page
- Classical Markov Decision Process techniques fail
- Csáji, Jungers and Blondel, 2010: graph rewriting to optimize the rank of a single page
- Our solution: action sets have a concise description

Admissible transition probabilities

Theorem

Optimization of PageRank

The convex hull of the set of admissible transition probabilities is either a simplex or a polyhedron defined by:

$$\begin{array}{ll} \forall j \in \mathcal{I}_i \ , & x_j = (1 - \alpha)z_j \\ \forall j \in \mathcal{O}_i \setminus \{j_0\} \ , & x_j = x_{j_0} \\ \forall j \in \mathcal{F}_i \ , & (1 - \alpha)z_j \leq x_j \leq x_{j_0} \\ & \text{and} & \sum_{j \in [n]} x_j = 1 \end{array}$$



- Implicitly defined actions: vertices of the polytope
- Concise description ⇒ polynomial time separation oracle
 ⇒ well-described polyhedron
 [Groetschel, Lovász, Schrijver, 1988]

Well-described Markov Decision Processes

Define

A well-described MDP is a finite MDP where the action sets are defined implicitly as the vertices of well-described polyhedra and the transitions and costs are linear

Theorem

The infinite horizon average cost problem on well-described MDP is solvable in polynomial time, even if there are exponentially many actions

Corollary

The PageRank optimization problem with local constraints is solvable in polynomial time

Optimization of PageRank

Resolution by Dynamic Programming

The ergodic dynamic programming equation

$$w_i + \psi = \max_{\nu \in \mathcal{P}_i} \nu(r_{i,\cdot} + w), \quad \forall i \in [n]$$

$$\text{tion } (w, v) \in \mathbb{R}^n \times \mathbb{R}. \text{ The constant } v \text{ is } v \text{ in } v \text{ in$$

has a solution $(w, \psi) \in \mathbb{R}^n \times \mathbb{R}$. The constant ψ is unique and is the value of the ergodic control problem

• To get an optimal strategy, select $\forall i$ a maximizing $\nu \in \mathcal{P}_i$

Resolution by Dynamic Programming

The ergodic dynamic programming equation

$$w_i + \psi = \max_{\nu \in \mathcal{P}_i} \nu(r_{i,\cdot} + w), \quad \forall i \in [n]$$
 (1)

has a solution $(w, \psi) \in \mathbb{R}^n \times \mathbb{R}$. The constant ψ is unique and is the value of the ergodic control problem

- ullet To get an optimal strategy, select orall i a maximizing $u \in \mathcal{P}_i$
- The unique solution of the discounted equation $w_i = \max_{\nu \ : \ \alpha\nu + (1-\alpha)z \in \mathcal{P}_i} \alpha\nu(r_{i,\cdot} + w) + (1-\alpha)zr_{i,\cdot}, \forall i \in [n] \ \ (2)$ is solution of (1) with $\psi = (1-\alpha)zw$
- The fixed point scheme for (2) has contracting factor α independent of the dimension: complexity of optimization

$$O\left(\frac{\log(\epsilon)}{\log(\alpha)}\sum_{i\in[n]}|\mathcal{O}_i|+|\mathcal{F}_i|\log(|\mathcal{F}_i|)\right)$$

Existence of a master page

Theorem

Assume $\forall i, j, r_{i,j} = r'_i$ and let $v = (I_n - \alpha S)^{-1}r'$ be the mean reward before teleportation

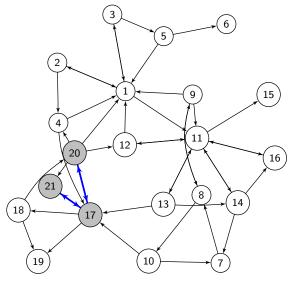
P=lpha S+(1-lpha)ez is an optimal link strategy if and only if

$$\begin{cases} v_j > \frac{v_i - r_i}{\alpha} \Rightarrow & \textit{facultative link } (i, j) \textit{ is activated} \\ v_j < \frac{v_i - r_i}{\alpha} \Rightarrow & \textit{facultative link } (i, j) \textit{ is desactivated} \end{cases}$$

v gives a total order of preference for page pointing

De Kerchove, Ninove and Van Dooren, 2008: similar result when one only requires that there exists a path from every page of the site to the rest of the web

Web graph optimized for PageRank



21 : controlled page

1: non controlled page

__ added links

PageRank sum: $0.0997 \rightarrow 0.1694$

Clique is not optimal

Conclusion for PageRank optimization

- Polynomial time solvability of the PageRank optimization problem
- Introduction of well-described Markov Decision Processes
- Very fast optimization algorithm based on value iteration
 For a problem with 413,639 pages and 2,319,174
 facultative links, solution in 81s with Intel Xeon 2.98Ghz
- Qualitative results on the optimal strategies
- Application to spam detection by minimizing the PageRank of known spam pages and propagating "spamicity"

Ranking algorithms

- Perron vector (Kendall and Babigton Smith 1939)
- HITS (Kleinberg 1998)
- SALSA (Lempel and Moran 2000)
- HOTS (Tomlin 2003)
- CenterRank (Blondel et al 2004)
- Matrix scaling (Smith 2005)

Definition of HITS [Kleinberg, 1998]

- Given a query, we build a subgraph of the web graph, focused on the query
- Hub and authority scores

$$ho \; \mathsf{hub}_j = \sum_{i:j o i} \mathsf{aut}_i \qquad \qquad
ho \; \mathsf{aut}_i = \sum_{j:j o i} \mathsf{hub}_j$$

- A: adjacency matrix of the subgraph: $A^T A$ aut = ρ^2 aut
- Uniqueness guaranteed with $(A^TA + \xi ee^T)$ aut $= \rho^2$ aut
- PageRank, SALSA and CenterRank also rank web pages according to a Perron vector

Kleinberg's HITS optimization

Obligatory links \mathcal{O} , facultative links \mathcal{F} , prohibited links \mathcal{I} J is the set of hyperlinks selected, $N(x) = \sum_{i \in [n]} x_i^2$ The HITS optimization problem is:

$$\max_{J\subseteq\mathcal{F},\mathbf{x}\in\mathbb{R}_+^n}\{\varphi(\mathbf{x})\ ;\ (A(J)^TA(J)+\xi ee^T)\mathbf{x}=\rho^2\mathbf{x}\ ,\ N(\mathbf{x})=1\}$$

Relaxed HITS optimization problem: weighted adjacency matrices → differentiable optimization

Perron vector Optimization Problem

We study the following Perron vector Optimization Problem:

$$\max \ J(M) = \varphi(u(M))$$
$$M \in h(C)$$

- $h: \mathbb{R}^m \to \mathbb{R}^{n \times n}_+$ and $\varphi: \mathbb{R}^n_+ \to \mathbb{R}$ are differentiable
- $u: \mathbb{R}^{n \times n}_{+} \to \mathbb{R}^{n}_{+}$ is the function that to an irreducible matrix associates its normalized Perron vector
- C convex, h(C) a set of nonnegative irreducible matrices

Chosen approach

- Nonconvex optimization problem
- Special case with h and φ linear and C a polytope is already NP-hard
- We abandon global optimality: first order method
- But efficient and scalable computation

Derivative for Perron vector optimization

Proposition (Derivative has rank one)

Denote $R = (M - \rho I_n)^{\#}$ the reduced resolvent at ρ and N the normalization function of u

Let
$$w^T = (-\nabla \varphi^T + (\nabla \varphi \cdot u)\nabla N^T)R$$
, then

$$g_{ij} = \frac{\partial J}{\partial M_{ij}} = \nabla \varphi^{\mathsf{T}} \frac{\partial u}{\partial M_{ij}} = w_i u_j$$

Proof

Consequence of [Deutch and Neumann, 1985]

$$\frac{\partial u}{\partial M_{ii}}(M) = -Re_i u_j + (\nabla N^T Re_i u_j)u$$

Iterative scheme for the derivative

Proposition

Let M be a primitive matrix with Perron vectors u and v. We denote $\tilde{M} = \frac{1}{a}M$, P = uv, $z = \frac{1}{a}(-\nabla \varphi^T + (\nabla \varphi \cdot u)\nabla N^T)$ $w_{k+1} = (z + w_k \tilde{M})(I_n - P) \rightarrow w$ Then

with a geometric rate of convergence $\frac{|\lambda_2|}{\alpha}$

- Direct application of Deutsch and Neumann formula: $O(mn^3)$
- Inversion of the linear system defining w: $O(n^3)$
- Iterative algorithm: O(m+n) operations per iteration

Coupled power and gradient iterations

- Approximation of value and gradient M is a weighted adjacency matrix, $u_{k+1} = \frac{Mu_k}{N(Mu_k)}$ $J_n(M) = \varphi(u_{k_n})$ and $g_n(M) = w_{k_n} u_{k_n}^T$ where k_n is the first nonnegative integer k such that $||u_{k+1} - u_k|| < \Delta(n)$
- Gradient step rule $\mathcal{B}_n(M)$ is an interrupted Armijo line search that may fail
- Polak's Master algorithm model for precision control Let $\omega \in (0,1)$, $\sigma' \in (0,1)$, $n_{-1} \in \mathbb{N}$ and $M_0 \in \mathcal{C}$, For $i \in \mathbb{N}$, compute M_{i+1} and the smallest $n_i \geq n_{i-1}$ s.t.

$$egin{aligned} M_{i+1} &\in \mathcal{B}_{n_i}(M_i) \ J_{n_i}(M_{i+1}) - J_{n_i}(M_i) \leq -\sigma'(\Delta(n_i))^\omega \end{aligned}$$

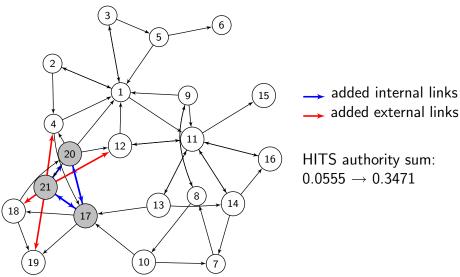
Convergence

Theorem

Let $(M_i)_{i>0}$ be a sequence constructed by the coupled power and gradient iterations for the resolution of the perron vector optimization problem

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

Web graph optimized for HITS



Tomlin's ideal HOTS algorithm

Network flow model of web traffic

A is the adjacency matrix of the web graph (first irreducible) Web surfers minimize the entropy of the traffic

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

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

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

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

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

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

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

 p_i is called the "temperature" of page i

Tomlin's HOTS: handling reduciblity

Network flow model with constraints on the modified network

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

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

$$\sum_{i,j\in[n+1]}\rho_{i,j}=1\tag{\mu}$$

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

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

Dual function: general case

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

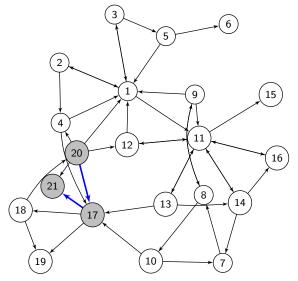
Effective HOTS:

$$d_i = F_i(d) = \left(\frac{(A^T d)_i + e^{a(d)}}{(Ad^{-1})_i + e^{-b(d)}}\right)^{1/2}$$

Proposition (HOTS optimization)

Assuming the HOTS algorithm $d^{k+1} = F(d^k)$ converges, the coupled power and gradient iterations converges to a stationary point of the HOTS optimization problem

Web graph optimized for HOTS



added links

HOTS score sum: $0.142 \rightarrow 0.169$

Page 21 has no outlink

Convergence of the ideal HOTS algorithm

Ideal HOTS vector defined by

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

Proposition

If A is irreducible and $A + A^T$ is primitive, then the ideal HOTS algorithm $d^{k+1} = f(d^k)$ converges to the ideal HOTS vector (unique up to a multiplicative constant)

Proof of convergence

```
Proof
f_i(d) = \left(\frac{(A^T d)_i}{(Ad^{-1})_i}\right)^{1/2}
f is monotone: d_1 > d_2 \Rightarrow f(d_1) > f(d_2)
f is homogeneous: \lambda \in \mathbb{R}_+ \Rightarrow f(\lambda d) = \lambda f(d)
A irreducible \Rightarrow G(f) is strongly connected \Rightarrow f has an
eigenvector (f(d^*) = \mu d^*) [Gaubert, Gunawardena 2004]
A irreducible \Rightarrow \frac{\partial f}{\partial d}(d^*) irreducible \Rightarrow eigenvector is unique
[Nussbaum, 1988]
A + A^T primitive \Rightarrow \frac{\partial f}{\partial d}(d^*) primitive
⇒ power algorithm converges [Nussbaum, 1988]
```

Convergence of the effective HOTS algorithm

Theorem

If there exists a feasible point with the same pattern as A, then the HOTS algorithm converges to the HOTS vector d^* s.t. $d_i^* = F_i(d^*) = \left(\frac{(A^T d^*)_i + e^{a(d^*)}}{(A(d^*)^{-1})_i + e^{-b(d^*)}}\right)^{1/2}$ (unique up to a multiplicative constant) with a linear rate of convergence equal to $|\lambda_2(\nabla F(d^*))|$.

Proof

F is homogeneous but not monotone: special proof required The ideal HOTS operator verifies

 $\theta(\log(f(d)), 0) \leq \theta(\log(d), 0)$: Lyapounov function

All the eigenvalues of $\nabla F(d^*)$ belong to (-1,1] and eigenvalue 1 is simple: local contraction in projective space

Drawbacks of HOTS

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

Relaxation of Perron ranking $\rho d = A^T d$ (a good page is a page pointed to by good pages) and anti-Perron ranking $\rho^{-1}d_i = \frac{1}{(Ad^{-1})_i}$ (a bad page is a page that points to bad pages) But anti-Perron penalizes pointing even to good pages

 Convergence rate may deteriorate when the size of the web graph grows:

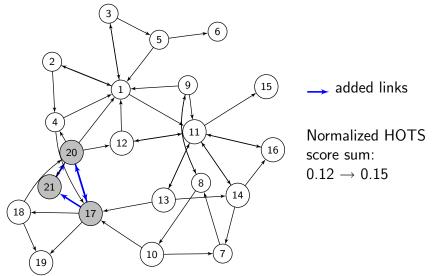
	CMAP 1,500 p	NZ Uni 413 kp	uk2002 18 Mp
$ \lambda_2(\nabla F) $	0.946	0.995	0.9994

Normalized HOTS

- Normalization of the adjacency matrix $M_{i,j} = \frac{A_{i,j}}{\sum_k A_{i,k}}$
- Relative entropy function
- Dangling nodes dealt with an additional fictitious node
- No more penalty for pointing to good pages: a bad page is a page that, in the mean, points to bad pages
- Better experimental convergence rate:

	CMAP 1,500 p	NZ Uni 413 kp	uk2002 18 Mp
$ \lambda_2(\nabla F) $	0.906	0.988	0.960

Web graph optimized for Normalized HOTS



Comparison of web ranking algorithms

Algorithm	convergence	external	harder than PR
	rate	links	to manipulate
Perron	good	++	
PageRank	0.85	-	reference
HITS	good	+++	
SALSA	good	+	
HOTS	bad		+ (?)
Normalized HOTS	acceptable (?)	=	+ (?)

Performances of web ranking algorithms

- + good characteristic
- = average
- bad characteristic
- (?) experimental likelihood only

From web ranking to chronotherapeutics

- For the web, scalability issues are very important
- So we developed a scalable optimization algorithm for Perron vector optimization
- Other situation where scalability is important:
 nonnegative matrices arise from monotone discretizations of age-structured Partial Differential Equations
- Optimization of chemotherapy infusion schedules [Basdevant, Clairambault, Lévi, 2006]
 Minimize the number of cancer cells while keeping the number of healthy cells above a toxicity threshold

Cancer chronotherapeutics

- Circadian clocks control cell proliferation
- Different behaviours (healthy or cancer cells)
- Chronotherapy [Lévi, 2002]
 Drug infusion schedules that depend on time
- [Billy, Clairambault, Fercoq, Gaubert, Lepoutre, Ouillon, Saito, Mathematics and Computers in Simulation 2011]

The cell cycle

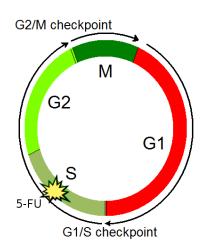
Cell proliferation: 4 main phases

5-FluoroUracil (5-FU): DNA damage in phase S

Proliferation stopped at the G2/M checkpoint

Figure: The cell cycle

G1: 1st growth phase, S: DNA synthesis G2: 2nd growth phase, M: mitosis Green and red correspond to the color of the nucleus with FUCCI reporters



McKendrick age-structured population model

$$\frac{\partial n_i(t,x)}{\partial t} + \frac{\partial n_i(t,x)}{\partial x} + \left(d_i(t,x) + K_{i\to i+1}(t,x)\right)n_i(t,x) = 0$$

$$n_{i+1}(t,0) = \int_0^\infty K_{i\to i+1}(t,x)n_i(t,x)dx$$

$$n_1(t,0) = 2 \int_0^\infty K_{I\to 1}(t,x) n_I(t,x) dx$$

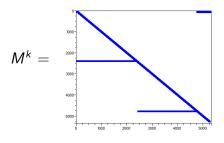
If d and K are T-periodic: Floquet eigenvalue λ

$$n_i(t,x) \sim C^0 N_i(t,x) e^{\lambda t}$$

 N_i is bounded and T-periodic

Block Leslie model

After a monotone discretization, we get an age-structured population with ages $1, \ldots n$



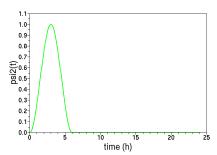
$$n^{k+1} = M^k n^k$$
, $\mathbf{M} = M^{N_T - 1} \dots M^1 M^0$, $n^T = \rho n^1 = \mathbf{M} n^1$

 $\frac{1}{T}\log(\rho)$: approximate growth rate of the population $\rho \geq R$: viability constraint

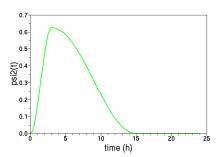
Modelling cancer cells

Two independent populations with the same model $K_{2\rightarrow 3}(x,t)=\kappa_{2\rightarrow 3}(x).\psi_2(t)$

Less synchronized proliferation gives an increased growth rate [Altinok, Lévi, Goldbeter, 2007]



Circadian control for healthy cells



Circadian control for cancer cells

Floquet eigenvalue optimization problem

Control:
$$K_{2\to 3}(x,t) = \kappa_{2\to 3}(x).\psi_2(t).(1-g(t))$$
 $g(t)=0$: no drug $g(t)=1$: transition blocking infusion 5-FU acts on phase S , thus on the G_2/M checkpoint only

$$\min_{g(\cdot)} \lambda_{C}(g)$$
 $\lambda_{H}(g) \geq \Lambda$
 $g \ 24\text{h-periodic}$

Optimal long term viable chemotherapy infusions

Discretized optimization problem

Discretized systems for cancer and healthy cells: nonnegative matrices $\mathbf{M}_{C}(x)$ and $\mathbf{M}_{H}(x)$

$$\min_{x \in [0,1]^{N_T}} \frac{1}{T} \log \left(\rho(\mathbf{M}_C(x)) \right)$$
$$\frac{1}{T} \log \left(\rho(\mathbf{M}_H(x)) \right) \ge \Lambda$$

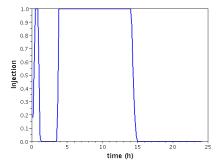
Proposition

The coupled power and gradient algorithm can be adapted to this case

Proof

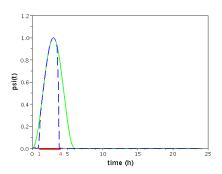
 $\frac{\partial
ho}{\partial \mathbf{M}_{i,i}} = u_i v_j$ and constraint dealt with a multiplier's method

Locally optimal strategy

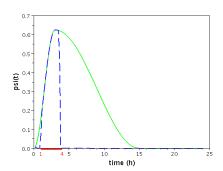


24h-periodic drug infusions g(t)

Locally optimal strategy



Action of drug infusion on transition rate for healthy cells



Action of drug infusion on transition rate for cancer cells

 G_2/M transition rate without drug $\psi_2(t)$ Drug induced transition rate $\psi_2(t)(1-g(t))$

Under the locally optimal strategy found, transitions G_2/M are restricted to lie between 1 a.m. and 4 a.m.

Conclusion chronotherapeutics

- We modelled a chemotherapy optimization problem with an age-structured proliferation model
- Optimization problem with periodic controls
- Optimization of chemotherapy shows the interest of chronotherapy
- Work in progress: Combination with a more realistic drug pharmacokinetics and pharmacodynamics model

Conclusion

- PageRank optimization: ergodic control very fast scalable algorithm for global optimum
- Other web rankings: small rank property of derivatives scalable optimization algorithm (but only local optimality)
- Convergence of HOTS
- Application of Perron vector optimization to chronotherapeutics
- Main open problem: determination of bounds for the Perron value optimization problem