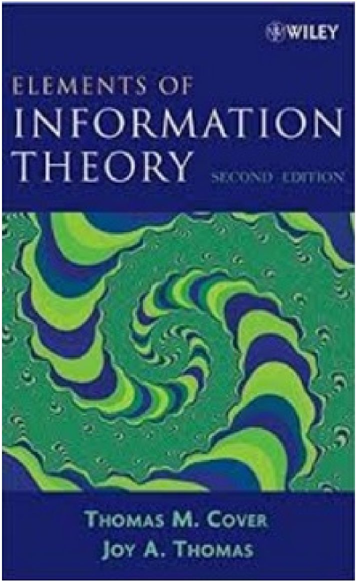
A relative entropy characterization of the growth rate of reward in risk-sensitive control

Venkat Anantharam
EECS Department, University of California, Berkeley

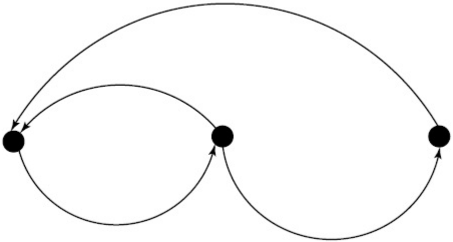
(joint work with Vivek Borkar , IIT Bombay)

August 26, 2016
2016 Conference on Applied Mathematics
The University of Hong Kong

To begin


WILEY
ELEMENTS OF
INFORMATION
THEORY
SECOND EDITION
THOMAS M. COVER
JOY A. THOMAS

Cover and Thomas, 2nd Edition, Problem 4.16

- Consider binary strings constrained to have at least one 0 and at most two 0s between any pair of 1s.
- What is the growth rate of the number of such sequences (assuming we start with a 1, for instance)?
- 

Cover and Thomas, 2nd Edition, Problem 4.16

- Let $X(n) = \begin{bmatrix} X_1(n) \\ X_2(n) \\ X_3(n) \end{bmatrix}$, where $X_i(n)$ is the number of paths of length n ending in state i.
- Then

$$X(n) = AX(n-1) = A^2X(n-2) = \ldots = A^{n-1}X(1) = A^n \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

where

$$A := \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

- Solution :
  $\log \rho$, where ρ is the Perron-Frobenius eigenvalue of A.

Perron-Frobenius eigenvalue

Every irreducible nonnegative square matrix A has an eigenvalue ρ, called its Perron-Frobenius eigenvalue such that:

- $\rho > 0$ (in particular ρ is real);
- ρ is at least as big as the absolute value of any eigenvalue of A;
- ρ admits left and right eigenvectors that are unique up to scaling and can be chosen to have strictly positive coordinates;
- $\log \rho$ is the “growth rate” of A^n.

Courant-Fischer formula

- Let $A \in \mathbb{R}^{d \times d}$ be a positive definite matrix.
- Its largest eigenvalue is given by

$$\rho = \max_{x \in \mathbb{R}^d, x \neq 0} \frac{x^T A x}{x^T x} .$$

Courant-Fischer formula

- Let $A \in \mathbb{R}^{d \times d}$ be a positive definite matrix.
- Its largest eigenvalue is given by

$$\rho = \max_{x \in \mathbb{R}^d, x \neq 0} \frac{x^T A x}{x^T x} \, .$$

- Is there an analogous characterization of the Perron-Frobenius eigenvalue of an irreducible nonnegative matrix?

Collatz-Wielandt formula

Let A be an irreducible nonnegative $d \times d$ matrix. Then its Perron-Frobenius eigenvalue ρ satisfies:

$$\rho = \sup_{x \,:\, x(i) > 0 \forall i} \; \min_{1 \le i \le d} \frac{\sum_{j=1}^{d} a(i,j) x(j)}{x(i)} ,$$

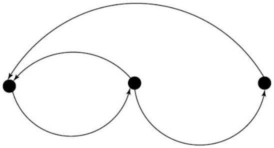
and

$$\rho = \inf_{x \,:\, x(i) > 0 \forall i} \; \max_{1 \le i \le d} \frac{\sum_{j=1}^{d} a(i,j) x(j)}{x(i)} .$$

But Problem 4.16 goes on a different tack.

Entropy and Problem 4.16 of Cover and Thomas

- Consider all Markov chains compatible with the directed graph giving rise to A with Perron-Frobenius eigenvalue λ.



- Transition probability matrix $\begin{bmatrix} 0 & 1 & 0 \\ \alpha & 0 & 1-\alpha \\ 1 & 0 & 0 \end{bmatrix}$ for some $0 \leq \alpha \leq 1$.
- Maximize the entropy rate of this Markov chain over all α.
- Problem 4.16 asks you to verify that this equals $\log \rho$.

Entropy and relative entropy

- Entropy:

$$H(P) := -\sum_i P(i) \log P(i)\ .$$

- Properties: $H(P) \geq 0$, concave in P, maximized at the uniform distribution.

Entropy and relative entropy

- Entropy:

$$H(P) := -\sum_i P(i) \log P(i) \ .$$

- Properties: $H(P) \geq 0$, concave in P, maximized at the uniform distribution.
- Relative entropy:

$$D(Q \| P) = \sum_i Q(i) \log \frac{Q(i)}{P(i)} \ .$$

- Properties: $D(Q \| P) \geq 0$, jointly convex in (Q, P), equal to 0 iff $Q = P$.

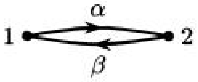
Entropy rate of a Markov chain

- Consider an irreducible finite state Markov chain with transition probabilities $p(j|i)$ and stationary distribution $\pi(\cdot)$.
- The entropy rate of the Markov chain is

$$\sum_{i,j} \pi(i) p(j|i) \log \frac{1}{p(j|i)} \ .$$

- Example:

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$



$$\text{Entropy rate} = \frac{\beta}{\alpha+\beta} h(\alpha) + \frac{\alpha}{\alpha+\beta} h(\beta) \ ,$$

where $h(p) := p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$.

Some notation

- Given A, an irreducible nonnegative $d \times d$ matrix, with Perron-Frobenius eigenvalue ρ, we will choose to write it as

$$a(i,j) = e^{r(i,j)} p(j|i), \text{ for all } i,j \ ,$$

  where $p(j|i)$ are transition probabilities.

- $\mathcal{P}_d$: probability distributions on $\{1, \dots, d\}$.

- $\mathcal{P}_{d \times d}$: probability distributions on $\{1, \dots, d\} \times \{1, \dots, d\}$.

Donsker-Varadhan characterization of the Perron-Frobenius eigenvalue

- A, irreducible nonnegative $d \times d$ with P-F eigenvalue ρ.
- Then

$$\log \rho = \sup_{\eta \in \tilde{\mathcal{G}}} \left[\sum_{i,j} \eta(i,j) r(i,j) - \sum_i \eta_0(i) \sum_j \eta_1(j|i) \log \frac{\eta_1(j|i)}{p(j|i)} \right],$$

where $\eta(i,j) = \eta_0(i)\eta_1(j|i)$ is a probability distribution, and $\tilde{\mathcal{G}}$ denotes the set of such probability distributions for which $\sum_i \eta(i,j) = \eta_0(j)$.

- Taking $p(j|i) = \frac{1}{\deg(i)}$ for all j such that $i \to j$ solves Problem 4.16.

Cumulant generating function and conjugate duality

Let $Q = (Q(i), 1 \leq i \leq d)$ be a probability distribution.
Let $\theta = (\theta(1), \ldots, \theta(d))^T$ be a real vector.
Then

$$\log(\sum_i Q(i) e^{\theta(i)}) = \sup_P \left(\sum_i \theta(i) P(i) - \sum_i P(i) \log \frac{P(i)}{Q(i)} \right) .$$

Cumulant generating function and conjugate duality

Let $Q = (Q(i), 1 \leq i \leq d)$ be a probability distribution.
Let $\theta = (\theta(1), \ldots, \theta(d))^T$ be a real vector.
Then

$$\log(\sum_i Q(i) e^{\theta(i)}) = \sup_P \left(\sum_i \theta(i) P(i) - \sum_i P(i) \log \frac{P(i)}{Q(i)} \right) .$$

There is an iceberg below the little tip of this formula:

- $\log(\sum_i Q(i) e^{\theta(i)})$ is $\log E[e^{\theta^T X}]$, where $P(X = e_i) = Q(i)$.

Cumulant generating function and conjugate duality

Let $Q = (Q(i), 1 \leq i \leq d)$ be a probability distribution.
Let $\theta = (\theta(1), \ldots, \theta(d))^T$ be a real vector.
Then

$$\log(\sum_i Q(i) e^{\theta(i)}) = \sup_P \left(\sum_i \theta(i) P(i) - \sum_i P(i) \log \frac{P(i)}{Q(i)} \right) .$$

There is an iceberg below the little tip of this formula:

- $\log(\sum_i Q(i) e^{\theta(i)})$ is $\log E[e^{\theta^T X}]$, where $P(X = e_i) = Q(i)$.
- Given a convex function $f(z)$ for $z \in \mathbb{R}^d$,

  $$\hat{f}(\theta) := \sup_z \left(\theta^T z - f(z) \right)$$

  is convex, and

  $$f(z) = \sup_\theta \left(z^T \theta - \hat{f}(\theta) \right) .$$

Minimax theorem

Let $f(x, y)$ be a function on $\mathcal{X} \times \mathcal{Y}$, where:

- $\mathcal{X}$ is a compact convex subset of some Euclidean space.
- $\mathcal{Y}$ is a convex subset of some Euclidean space.
- f is concave in x for each fixed y.
- f is convex in y for each fixed x.

Then

$$\sup_x \inf_y f(x, y) = \inf_y \sup_x f(x, y) \ .$$

Donsker-Varadhan from Collatz-Wielandt (1)

- $$\begin{aligned} \rho &= \inf_{x\,:\,x(i)>0\forall i} \max_{1\le i\le d} \frac{\sum_{j=1}^{d} a(i,j)x(j)}{x(i)}, \\ &= \inf_{x\,:\,x(i)>0\forall i} \sup_{\gamma\in\mathcal{P}_d} \sum_{i=1}^{d} \gamma(i) \frac{\sum_{j=1}^{d} e^{r(i,j)} p(j|i)x(j)}{x(i)} \\ &= \inf_{x\,:\,x(i)>0\forall i} \sup_{\gamma\in\mathcal{P}_d} \sum_{i=1}^{d}\sum_{j=1}^{d} \gamma(i)p(j|i)e^{r(i,j)+\log x(j)-\log x(i)} \end{aligned}$$

- So

$$\log\rho = \inf_{u\in\mathbb{R}^d} \sup_{\gamma\in\mathcal{P}_d} \log\Big(\sum_{i=1}^{d}\sum_{j=1}^{d}\gamma(i)p(j|i)e^{r(i,j)+u(j)-u(i)}\Big).$$

Donsker-Varadhan from Collatz-Wielandt (2)

$$
\begin{aligned}
\log \rho &= \inf_{u \in \mathbb{R}^d} \sup_{\gamma \in \mathcal{P}_d} \log \Big(\sum_{i=1}^{d} \sum_{j=1}^{d} \gamma(i) p(j|i) e^{r(i,j)+u(j)-u(i)} \Big) . \\
&= \inf_{u \in \mathbb{R}^d} \sup_{\gamma \in \mathcal{P}_d} \sup_{\eta \in \mathcal{P}_{d \times d}} \Bigg[\sum_{i,j} \eta(i,j)(r(i,j) + u(j) - u(i)) \\
&\qquad - \sum_{i,j} \eta(i,j) \log \frac{\eta(i,j)}{\gamma(i) p(j|i)} \Bigg] \\
&= \sup_{\gamma \in \mathcal{P}_d} \sup_{\eta \in \mathcal{P}_{d \times d}} \inf_{u \in \mathbb{R}^d} \Bigg[\sum_{i,j} \eta(i,j)(r(i,j) + u(j) - u(i)) \\
&\qquad - \sum_{i} \eta_0(i) \log \frac{\eta_0(i)}{\gamma(i)} - \sum_{i} \eta_0(i) \sum_{j} \eta_1(j|i) \log \frac{\eta_1(j|i)}{p(j|i)} \Bigg]
\end{aligned}
$$

Donsker-Varadhan from Collatz-Wielandt (3)

$$
\begin{aligned}
\log \rho &= \sup_{\eta \in \mathcal{P}_{d\times d}} \inf_{u \in \mathbb{R}^d} \left[\sum_{i,j} \eta(i,j)\big(r(i,j) + u(j) - u(i)\big) \right. \\
&\qquad \left. - \sum_i \eta_0(i) \sum_j \eta_1(j|i) \log \frac{\eta_1(j|i)}{p(j|i)} \right] \\
&= \sup_{\eta \in \tilde{\mathcal{G}}} \left[\sum_{i,j} \eta(i,j) r(i,j) - \sum_i \eta_0(i) \sum_j \eta_1(j|i) \log \frac{\eta_1(j|i)}{p(j|i)} \right] .
\end{aligned}
$$

Average reward Markov decision problem

- Let $\mathcal{S} := \{1, \ldots, d\}$ and let U be a finite set.
- $[p(j|i,u)]$: transition probabilities from $\mathcal{S}$ to $\mathcal{S}$ for $u \in U$.
- Assume irreducibility for convenience.
- $r(i,u,j)$: one-step reward for transition from i to j under u.
- Aim:

$$\sup_{\mathcal{A}} \liminf_{N\to\infty} \frac{1}{N} \sum_{m=0}^{N-1} r(X_m, Z_m, X_{m+1}) \ ,$$

  where $\mathcal{A}$ is the set of causal randomized control strategies.
- Call this growth rate λ.

Ergodic characterization of the optimal reward

- Write probability distributions $\eta(i,u,j)$ as

$$\eta(i,u,j) = \eta_0(i)\eta_1(u|i)\eta_2(j|i,u) \ .$$

- Let $\mathcal{G}$ denote the set of η satisfying

$$\sum_{i,u} \eta(i,u,j) = \eta_0(j) \ , \quad \text{for all } j \ .$$

- Then

$$\lambda = \sup_{\eta \in \mathcal{G}} \sum_{i,u,j} \eta(i,u,j) r(i,u,j) \ .$$

- This is based on linear programming duality, starting from the average cost dynamic programming equation:

$$\lambda + h(i) = \max_{u \in U} \sum_{j} p(j|i,u) \left(r(i,u,j) + h(j) \right) \ .$$

Risk-sensitivity (1)

- Consider a random reward R, whose distribution depends on some choices.

- One can incorporate sensitivity to risk by posing the problem of maximizing $E[R] - \frac{1}{2}\theta \mathrm{Var}(R)$.

- $$\begin{array}{ll} \theta > 0 \Leftrightarrow & \text{Risk-averse} \\ \theta < 0 \Leftrightarrow & \text{Risk-seeking} \end{array}$$

- In a framework with Markovian dynamics, it is easier to work with a criterion more aligned to large deviations theory than the variance.

Risk-sensitivity (2)

- Write

$$E[e^{-\theta R}] = e^{-\theta E[R]} E[e^{-\theta(R-E[R])}] \simeq e^{-\theta E[R]} \left(1 + \frac{\theta^2}{2}\mathrm{Var}(R)\right) .$$

- Hence

$$\begin{aligned} -\frac{1}{\theta}\log E[e^{-\theta R}] &\simeq E[R] - \frac{1}{\theta}\log(1 + \frac{\theta^2}{2}\mathrm{Var}(R)) \\ &\simeq E[R] - \frac{\theta}{2}\mathrm{Var}(R) . \end{aligned}$$

- $$\begin{array}{cccc} \text{Risk-averse} & \Leftrightarrow & \theta > 0 \Longrightarrow & \text{Minimize } E[e^{-\theta R}] \\ \text{Risk-seeking} & \Leftrightarrow & \theta < 0 \Longrightarrow & \text{Maximize } E[e^{-\theta R}] . \end{array}$$

- The risk-seeking case corresponds to portfolio growth rate maximization.

Risk-sensitive control problem

- Let $\mathcal{S} := \{1, \ldots, d\}$ and let U be a finite set.
- $[p(j|i, u)]$: transition probabilities from $\mathcal{S}$ to $\mathcal{S}$ for $u \in U$.
- Assume irreducibility for convenience.
- $r(i, u, j)$: one-step reward for transition from i to j under u.
- Aim:

$$\max_{i} \sup_{\mathcal{A}} \liminf_{N \to \infty} \frac{1}{N} \log E\left[e^{\sum_{m=0}^{N-1} r(X_m, Z_m, X_{m+1})} \middle| X_0 = i \right],$$

  where $\mathcal{A}$ is the set of causal randomized control strategies.
- Call this growth rate λ.

Statement of the problem

Formal problem statement

- Let $\mathcal{S}$ and U be compact metric spaces.

- Let $p(dy|x,u) \; : \; \mathcal{S} \times U \mapsto \mathcal{P}(\mathcal{S})$ be a prescribed kernel. Here $\mathcal{P}(\mathcal{S})$ is the set of probability distributions on $\mathcal{S}$ with the topology of weak convergence.

- Let $r(x,u,y) \; : \; \mathcal{S} \times U \times \mathcal{S} \to [-\infty, \infty)$. This is the per-stage reward function.

- Causal control strategies are defined in terms of kernels $\phi_0(du|x_0)$ and

$$\phi_{n+1}(du|(x_0,u_0),\ldots,(x_n,u_n),x_{n+1}) \; , \quad n \geq 0 \; .$$

- Aim:

$$\sup_x \sup_{\mathcal{A}} \liminf_{N\to\infty} \frac{1}{N} \log E\left[e^{\sum_{m=0}^{N-1} r(X_m, Z_m, X_{m+1})} \middle| X_0 = x \right] ,$$

where $\mathcal{A}$ is the set of causal randomized control strategies.

- Call this growth rate λ.

Technical assumptions

- **(A0)**: $e^{r(x,u,y)} \in C(\mathcal{S} \times U \times \mathcal{S})$.
- **(A1)**: The maps $(x, u) \to \int f(y)p(dy|x, u)$, $f \in C(\mathcal{S})$ with $\|f\| \leq 1$, are equicontinuous.

This case where (**A0**) and (**A1**) hold is developed by a limiting argument starting with the case with the stronger assumptions:

- **(A0+)**: Condition **(A0)** holds and we also have $e^{r(x,u,y)} > 0$ for all (x, u, y).
- **(A1+)**: Condition **(A1)** holds and we also have $p(dy|x, u)$ having full support for all (x, u).

The first main result (1)

- Define the operator $T \ : \ C(\mathcal{S}) \to C(\mathcal{S})$ by

$$Tf(x) := \sup_{\phi \in \mathcal{P}(U)} \int \int p(dy|x,u)\phi(du)e^{r(x,u,y)}f(y) \ .$$

- Let $C^+(\mathcal{S}) := \{f \in C(\mathcal{S}) \ : \ f(x) > 0 \ \forall x\}$ denote the cone of nonnegative functions in $C(\mathcal{S})$.
- Theorem: Under assumptions **(A0+)** and **(A1+)** there exists a unique $\rho > 0$ and $\psi \in \mathrm{int}(C^+(\mathcal{S}))$ such that

$$\rho\psi(x) = \sup_{\phi \in \mathcal{P}(U)} \int \int p(dy|x,u)\phi(du)e^{r(x,u,y)}\psi(y) \ .$$

- Thus ρ may be considered the Perron-Frobenius eigenvalue of T. Note that T is a nonlinear operator.

The first main result (2)

Let $\mathcal{M}^+(S)$ denote the set of positive measure on $\mathcal{S}$. We have the following characterizations of the Perron-Frobenius eigenvalue.

- $$\rho = \inf_{f \in \text{int}(C^+(\mathcal{S}))} \sup_{\mu \in \mathcal{M}^+(S)} \frac{\int Tf(x)\mu(dx)}{\int f(x)\mu(dx)} .$$
- $$\rho = \sup_{f \in \text{int}(C^+(\mathcal{S}))} \inf_{\mu \in \mathcal{M}^+(S)} \frac{\int Tf(x)\mu(dx)}{\int f(x)\mu(dx)} .$$
- These formulae can be viewed as a version of the Collatz-Wielandt formula for the Perron-Frobenius eigenvalue of the nonlinear operator T.
- Finally, we have $\lambda = \log \rho$.

The second main result

- Theorem: Under assumptions **(A0)** and **(A1)** we have

$$\lambda = \sup_{\eta \in \mathcal{G}} \Bigg(\int \int \int \eta(dx, du, dy) r(x, u, y) \\ - \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) \| p(dy|x, u)) \Bigg) ,$$

  where $\tilde{\eta}(dx, du) := \eta_0(dx)\eta_1(du|x)$.

- This is a generalization of the Donsker-Varadhan formula to characterize the growth rate of reward in risk-sensitive control.

Structure of the proof

Structure of the proof

- The Collatz-Wielandt formula for the Perron-Frobenius eigenvalue ρ of the nonlinear operator T comes from an application of the nonlinear Krein-Rutman theorem of Ogiwara.

Structure of the proof

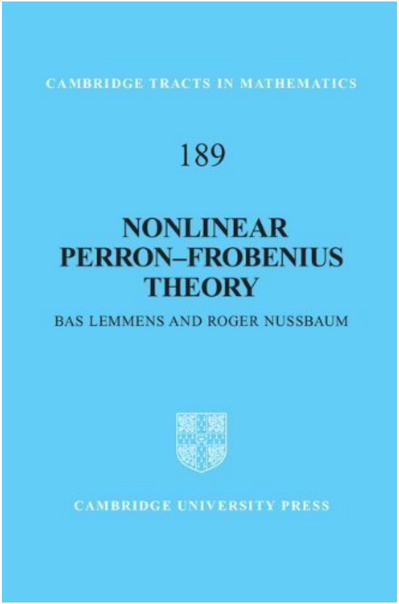
- The Collatz-Wielandt formula for the Perron-Frobenius eigenvalue ρ of the nonlinear operator T comes from an application of the nonlinear Krein-Rutman theorem of Ogiwara.
- The identification of $\log \rho$ with λ comes from observing that iterates of T form the Bellman-Nisio semigroup , so that the eigenvalue problem for T expresses the abstract dynamic programming principle.

Structure of the proof

- The Collatz-Wielandt formula for the Perron-Frobenius eigenvalue ρ of the nonlinear operator T comes from an application of the nonlinear Krein-Rutman theorem of Ogiwara.
- The identification of $\log \rho$ with λ comes from observing that iterates of T form the Bellman-Nisio semigroup , so that the eigenvalue problem for T expresses the abstract dynamic programming principle.
- The generalized Donsker-Varadhan formula under the assumptions **(A0+)** and **(A1+)** comes from a calculation analogous to the one giving the usual Donsker-Varadhan formula from the usual Collatz-Wielandt formula.

Structure of the proof

- The Collatz-Wielandt formula for the Perron-Frobenius eigenvalue ρ of the nonlinear operator T comes from an application of the nonlinear Krein-Rutman theorem of Ogiwara.
- The identification of $\log \rho$ with λ comes from observing that iterates of T form the Bellman-Nisio semigroup , so that the eigenvalue problem for T expresses the abstract dynamic programming principle.
- The generalized Donsker-Varadhan formula under the assumptions **(A0+)** and **(A1+)** comes from a calculation analogous to the one giving the usual Donsker-Varadhan formula from the usual Collatz-Wielandt formula.
- The generalized Donsker-Varadhan formula under the assumptions **(A0)** and **(A1)** comes from taking the limit in a perturbation argument.


CAMBRIDGE TRACTS IN MATHEMATICS
189
NONLINEAR PERRON–FROBENIUS THEORY
BAS LEMMENS AND ROGER NUSSBAUM
CAMBRIDGE UNIVERSITY PRESS

Nonlinear Krein-Rutman theorem of Ogiwara Preliminaries

- Let B be a real Banach space and B^+ a closed convex cone in B with vertex at 0, satisfying $B^+ \cap (-B^+) = \{0\}$, and having nonempty interior.

- For $x, y \in B$, write $x \geq y$ if $x - y \in B^+$, $x > y$ if $x - y \in B^+ - \{0\}$, and $x \gg y$ if $x - y \in \text{int}(B^+)$.

- $T : B \mapsto B$, mapping B^+ into itself is called:
  - strongly positive if $x > y \implies Tx \gg Ty$;
  - positively homogeneous if $T(\alpha x) = \alpha Tx$ if $x \in B^+$ and $\alpha > 0$.

- Let $T^{(n)}$ denote the n-fold iteration of T.

Nonlinear Krein-Rutman theorem of Ogiwara

- Theorem (Ogiwara) : For a compact, strongly positive, positively homogeneous map T from an ordered Banach space (B, B^+) to itself, $\lim_{n\to\infty} \|T^{(n)}\|^{\frac{1}{n}}$ exists, and is strictly positive, is an eigenvalue of T, is the only positive eigenvalue of T, and admits an eigenvector in the interior of B^+ that is unique up to multiplication by a positive constant.

An application

- For each $u \in U$, a finite set, let G_u be a directed graph on $\mathcal{S} := \{1, \ldots, d\}$, with each vertex having positive outdegree for each u.

- We wish to maximize the growth rate of the number of paths, starting from 1 say, where we also get to choose which graph to use at each time (possibly randomized).

- Result :
  Among all stationary $\mathcal{S} \times U$-valued Markov chains (X_n, Z_n) such that if the transition from (i, u) to (j, v) has positive probability then $i \to j$ is in G_u, maximize $H(X_1 | X_0, U_0)$.

Another application (preliminaries)

- Let $\mathcal{S} := \{1, \ldots, d\}$ and let U be a finite set.
- $[p(j|i,u)]$: transition probabilities from $\mathcal{S}$ to $\mathcal{S}$ for $u \in U$.
- Let $\mathcal{S}_0 \subseteq \mathcal{S}$ and $\mathcal{S}_1 := \mathcal{S}_0^c$ be nonempty.
- Assume $[p(j|i,u)]$ is irreducible for each u.
- Assume $d(i,u) := \sum_{j \in \mathcal{S}_1} p(j|i,u) > 0$ for all $i \in \mathcal{S}_1$.
- Define

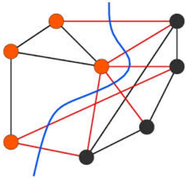
$$q(j|i,u) := \frac{p(j|i,u)}{d(i,u)} \text{ for } i \in \mathcal{S}_1.\ u \in U\ .$$

Another application (result)

- Aim:

$$\max_{i \in \mathcal{S}_1} \sup_{\mathcal{A}} \liminf_{N \to \infty} \frac{1}{N} \log P(\tau > N) \ .$$

where τ is the first hitting time of $\mathcal{S}_0$.



- Can be solved based on the observation that

$$P(\tau > N) = E[e^{\sum_{m=0}^{N-1} \log(d(X_m, Z_m))}] \ .$$

The most obvious open questions

- How does one remove the compactness assumptions on $\mathcal{S}$ and U?

- What about continuous time?

  (There is a version of the generalized Collatz-Wielandt formula for reflected controlled diffusions in a bounded domain, due to Araposthasis, Borkar, and Suresh Kumar: http://arxiv.org/abs/1312.5834)

The end