

# A variational formula for risk-sensitive control

Vivek S. Borkar,  
IIT Bombay

May 9, 2018, IMA, Minneapolis

PART I: Risk-sensitive reward for Markov chains (*joint work with V. Anantharam, Uni. of California, Berkeley*)

PART II: Risk-sensitive cost for reflected diffusions (*joint work with Ari Arapostathis, Uni. of Texas at Austin, and K. Suresh Kumar, IIT Bombay*)

PART III: Linear/dynamic programming approaches for degenerate finite risk-sensitive reward problems

# **PART I**

**Courant-Fisher** formula for principle eigenvalue of a positive definite matrix  $A \in \mathcal{R}^{d \times d}$ :

$$\lambda = \max_{0 \neq x \in \mathcal{R}^d} \frac{x^T A x}{x^T x}.$$

Consider an irreducible nonnegative  $Q \in \mathcal{R}^{d \times d}$ . Then the **Perron-Frobenius** theorem guarantees a positive principle eigenvalue with an associated positive eigenvector.

Is there a counterpart of the **Courant-Fisher** formula?

**YES !!** The **Collatz-Wielandt** formula for the principal eigenvalue of an irreducible nonnegative matrix

$$Q = [[q(i, j)]] \in \mathcal{R}^{d \times d} :$$

$$\begin{aligned} \lambda &= \sup_{x=[x_1, \dots, x_d]^T, x_i > 0 \forall i} \min_{i: x_i > 0} \left( \frac{(Qx)_i}{x_i} \right) \\ &= \inf_{x=[x_1, \dots, x_d]^T, x_i > 0 \forall i} \max_{i: x_i > 0} \left( \frac{(Qx)_i}{x_i} \right). \end{aligned}$$

An alternative characterization: write

$$Q = DP, \text{ where}$$

$$D = \text{diag}(d_1, \dots, d_d),$$

$$P = [[p(j|i)]] \text{ stochastic.}$$

Also define

$\mathcal{G}_0 := \{ (\pi, \tilde{P}) : \pi \text{ is a stationary probability for the stochastic matrix } \tilde{P} = [[\tilde{p}(j|i)]] \}$ .

Then the following representation holds:

$$\log \lambda = \sup_{(\pi, \tilde{P}) \in \mathcal{G}_0} \left( \sum_i \pi(i) [d(i) - D(\tilde{p}(\cdot|i) || p(\cdot|i))] \right).$$

This is the **Donsker-Varadhan** formula for the principal eigenvalue of a nonnegative matrix.

(cf. the book by **Dembo-Zeitouni**)

Infinite dimensional generalization of Perron-Frobenius theorem is given by the **Krein-Rutman** theorem: Let

1.  $B$  be a Banach space with a 'positive cone'  $K$  such that  $K - K$  is dense in  $B$ ,

2.  $T : B \mapsto B$  a compact positive linear operator which is strongly positive.

Then a principal eigenvalue (unique, positive) / eigenvector (positive) exist.

Our interest is in the following *nonlinear* scenario arising in *Risk-Sensitive Control*: Consider

- a controlled Markov chain  $\{X_n\}$  on a compact metric state space  $S$ ,
- an associated control process  $\{Z_n\}$  in a compact metric control space  $U$ ,
- a *per stage reward function*  $r : S \times U \times S \mapsto \mathcal{R}$  such that  $r \in C(S \times U \times S)$ ,



- a controlled transition kernel  $p(dy|x, u)$  with **full support**<sup>\*</sup>, such that

$$P(X_{n+1} \in A | X_m, Z_m, m \leq n) = p(A | X_n, Z_n)$$

and, the maps

$$(x, u) \mapsto \int f(y)p(dy|x, u), \quad f \in C(S), \quad \|f\| \leq 1,$$

are equicontinuous.

<sup>\*</sup>This can be relaxed via an approximation argument.

The *control problem* is to maximize the asymptotic growth rate of the exponential reward:

$$\lambda := \sup_{x \in S} \sup \liminf_{N \uparrow \infty} \frac{1}{N} \log E \left[ e^{\sum_{m=0}^{N-1} r(X_m, Z_m, X_{m+1})} \mid X_0 = x \right].$$

The second supremum is over all admissible (i.e., non-anticipative) controls. We allow *relaxed* (i.e., probability measure valued) controls.

Define

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

$$T^{(n)} := T \circ T \circ \dots \circ T \text{ (} n \text{ times)}, \quad T^{(0)} := Id.$$

$T^{(n)} : C(S) \mapsto C(S)$ ,  $n \geq 0$ , is the **Nisio** semigroup, satisfying:

1. strictly increasing:  $f > g \implies T^{(n)}f > T^{(n)}g$ ,
2. strongly positive:  $f \geq 0$ ,  $f \neq \theta \implies T^{(n)}f \in \text{int}(C^+(S))$ ,
3. positively 1-homogeneous:  $c > 0 \implies T^{(n)}(cf) = cT^{(n)}f$ ,
4. compact.

This leads to an abstract **Collatz-Wielandt** formula:

**Theorem** There exist  $\rho > 0, \psi \in \text{int}(C^+(S))$  such that  $T\psi = \rho\psi$  and

$$\begin{aligned}\rho &= \inf_{f \in \text{int}(C^+(S))} \sup_{\mathcal{M}^+(S)} \frac{\int T f d\mu}{\int f d\mu} \\ &= \sup_{f \in \text{int}(C^+(S))} \inf_{\mathcal{M}^+(S)} \frac{\int T f d\mu}{\int f d\mu}.\end{aligned}$$

Furthermore,  $\log \rho$  is the optimal reward for the risk-sensitive control problem.

The proof uses **Nussbaum-Ogiwara** formulation of a nonlinear Krein-Rutman theorem.

## A variational formula:

Let  $\mathcal{G} :=$  the set of probability measures

$$\eta(dx, du, dy) \in \mathcal{P}(S \times U \times S)$$

which disintegrate as

$$\eta(dx, du, dy) = \eta_0(dx)\eta_1(du|x)\eta_2(dy|x, u),$$

such that  $\eta_0$  is invariant under the transition kernel

$$\int_U \eta_2(dy|x, u)\eta_1(du|x).$$

Let  $D(\cdot \| \cdot)$  denote the Kullback-Leibler divergence or relative entropy.

**Theorem** Under above hypotheses,

$$\log \rho = \sup_{\eta \in \mathcal{G}} \left( \int \int \int \eta_0(dx) \eta_1(du|x) \left[ \int r(x, u, y) \eta_2(dy|x, u) - D(\eta_2(dy|x, u) \| p(dy|x, u)) \right] \right).$$

This generalizes the **Donsker-Varadhan** formula.

Hypotheses can be relaxed to:

1.  $\text{Range}(r) = [-\infty, \infty)$  with  $e^r \in C(S \times U \times S)$ ,
2.  $p(dy|x, u)$  need not have full support.

(extension via an approximation argument)

- We have an equivalent concave maximization problem, as opposed to a ‘team’ problem one would obtain from the usual ‘log transformation’.
- If  $\rho(\varphi)$  denotes the asymptotic growth rate for a randomized Markov control  $\varphi$ , then  $\rho = \max \rho(\varphi)$  (sufficiency of randomized Markov controls).
- Related to entropy-penalized control (Bierkens-Kapper, Guan-Raginsky-Willett, Todorov)

# Applications

1. Growth rate of the number of directed paths in a graph  
(requires  $-\infty$  as a possible reward).
2. Portfolio optimization in the framework of Bielecki, Hernandez-Hernández and Pliska.
3. Problem of minimizing the exit rate from a domain.



# **PART II**

Reflected diffusion:

$$\begin{aligned}dX(t) &= b(X(t), v(t))dt + \sigma(X(t))dW(t) - \gamma(t)d\xi(t), \\d\xi(t) &= I\{X(t) \in \partial Q\}d\xi(t),\end{aligned}$$

for  $t \geq 0$ . Here:

1.  $Q$  open bounded with  $C^3$  boundary  $\partial Q$ .
2.  $b$  continuous,  $b(\cdot, u)$  Lipschitz uniformly in  $u$ .
3.  $\sigma$  is  $C^{1,\beta_0}$  and uniformly non-degenerate.

4.  $\gamma_i(x) = \sigma(x)\sigma(x)^T \eta(x)$  where  $\eta(x)$  is the unit outward normal.

5.  $W(\cdot)$  a BM,  $v(\cdot)$  a non-anticipative control  $\in$  a compact action space.

OBJECTIVE: Minimize

$$\lim_{t \uparrow \infty} \frac{1}{t} \log E \left[ e^{\int_0^t r(X(s), v(s)) ds} \right],$$

where  $r$  is continuous.

**Nisio semigroup:** For  $t \geq 0$ ,

$$S_t f(x) := \inf_{v(\cdot)} E_x \left[ e^{\int_0^t r(X(s), v(s)) ds} \right].$$

Then  $S_t : C(\bar{Q}) \mapsto C(\bar{Q})$  is a semigroup of strongly continuous bounded Lipschitz, monotone operators with infinitesimal generator  $\mathcal{G}$  defines by

$$\mathcal{G}f(x) = \frac{1}{2} \text{tr} \left( \sigma(x) \sigma^T(x) \nabla^2 f(x) \right) + \min_v [\langle b(x, v), \nabla f(x) \rangle + r(x, v) f(x)].$$

Also, supperadditive, positively 1-homogeneous, strongly positive, completely continuous.

Let  $C_{\gamma,+}^2(\bar{Q}) := \{f : \bar{Q} \mapsto [0, \infty) : f \text{ is } C^2 \text{ with } \langle \nabla f(x), \gamma(x) \rangle = 0 \text{ for } x \in \partial Q\}$ .

Nonlinear **Krein-Rutman** theorem  $\implies$  There exists unique pair  $(\rho, \varphi) \in \mathcal{R} \times C_{\gamma,+}^2(\bar{Q})$  satisfying  $\|\varphi\|_{0,\bar{Q}} = 1$  such that

$$S_t \varphi = e^{\rho t} \varphi.$$

This solves

$$\mathcal{G}\varphi(x) = \rho\varphi(x), \quad x \in Q, \quad \langle \nabla\varphi(x), \gamma(x) \rangle = 0, \quad x \in \partial Q.$$

Abstract **Collatz-Wielandt** formula  $\implies$

$$\begin{aligned}\rho &= \inf_{f \in C_{\gamma,+}^2(\bar{Q}), f > 0} \sup_{\nu \in \mathcal{P}(\bar{Q})} \int \frac{\mathcal{G}f}{f} d\nu \\ &= \sup_{f \in C_{\gamma,+}^2(\bar{Q}), f > 0} \inf_{\nu \in \mathcal{P}(\bar{Q})} \int \frac{\mathcal{G}f}{f} d\nu\end{aligned}$$

In uncontrolled case, the first formula is the convex dual of the **Donsker-Varadhan** formula for principle eigenvalue of  $\mathcal{G}$ :

$$\rho = \sup_{\nu \in \mathcal{P}(\bar{Q})} \left( \int_{\bar{Q}} r(x) \nu(dx) - I(\nu) \right)$$

where

$$I(\nu) := \inf_{f \in C_{\gamma,+}^2(\bar{Q}), f > 0} \int_{\bar{Q}} \left( \frac{\mathcal{G}f}{f} \right) d\nu.$$

# **PART III**

## Notation:

- $\{X_n, n \geq 0\}$  : controlled Markov chain with finite state space  $S := \{1, 2, \dots, s\}$  and finite action space  $U$ .
- $\{Z_n\}$  :  $U$ -valued control process.
- $P_u = [[p(j|i, u)]]_{i,j \in S}$  : controlled transition matrix.
- $r(\cdot, \cdot, \cdot) \in C(S \times U \times S)$  : 'per stage' reward.



# Risk-sensitive reward

**Aim:** Maximize the ‘asymptotic growth rate’

$$\lambda^* := \sup_i \sup_{\{Z_n\}} \liminf_{N \uparrow \infty} \frac{1}{N} \log E_i \left[ e^{\sum_{m=0}^{N-1} r(X_m, Z_m, X_{m+1})} \right].$$

Here:

- $E_i[\dots]$  := expectation w.r.t.  $X_0 = i$ , and,
- the inner supremum is over all admissible controls.

Consider a controlled Markov chain  $\{Y_n\}$  on  $S$  with state-dependent action space at state  $i$  given by:

$$\tilde{U}_i := \cup_{u \in U} (\{u\} \times V_{i,u}),$$

where

$$V_{i,u} := \{q(\cdot|i, u) : q(\cdot|i, u) \geq 0, \sum_j q(j|i, u) = 1\}.$$

This is isomorphic to  $\mathcal{P}(S)$ . Let

$$K := \cup_{i \in S} (\{i\} \times \tilde{U}_i).$$

The (controlled) transition probabilities of  $\{Y_n\}$  are

$$\tilde{p}(j|i, (u, q(\cdot|i, u))) := q(j|i, u).$$

Define per stage reward  $\tilde{r} : K \times S \mapsto \mathcal{R}$  by:

$$\tilde{r}(i, (u, q(\cdot|i, u)), j) := r(i, u, j) - D(q(\cdot|i, u) || p(\cdot|i, u)).$$

Let  $\{(Z_n, Q_n), n \geq 0\} :=$  the  $\tilde{U}_{Y_n}$ -valued control process.

Consider the problem:

Maximize the long run average reward

$$\liminf_{N \uparrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} E_x [\tilde{r}(Y_n, (Z_n, Q_n), Y_{n+1})].$$

Define the corresponding ergodic occupation measure  $\gamma \in \mathcal{P}(K \times S)$  by

$$\gamma(i, (u, dq), j) := \gamma_1(i) \gamma_2(u, dq|i) \gamma_3(j|i, (u, q)),$$

where  $\gamma_1$  is an invariant probability distribution (not necessarily unique) under the transition kernel

$$\check{\gamma}(j|i) = \sum_u \int_{V_{i,u}} \gamma_2(u, dq|i) \gamma_3(j|i, (u, q)).$$

Let  $\mathcal{E} :=$  the set of such  $\gamma$ .

The above average reward control problem is equivalent to the linear program:

**P0** Maximize

$$\sum_{i,j,u} \int \gamma(i, (u, dq), j) \tilde{r}(i, (u, q), j)$$

over  $\mathcal{E}$ .

(Recall that  $\mathcal{E}$  is specified by linear constraints.)

The maximum will be attained at an extreme point of  $\mathcal{E}$  corresponding to a stationary Markov policy.

This LP can be simplified as :

Maximize

$$\sum_{i,j} \int \gamma'(i, u, j) [r(i, u, j) - D(\tilde{q}(\cdot|i, u) \| p(\cdot|i, u))]$$

over

$$\tilde{\mathcal{E}} := \{\gamma' \in \mathcal{P}(S \times U \times S) : \gamma'(i, u, j) = \gamma_1(i) \varphi(u|i) q(j|i, u)\}$$

where  $\gamma_1(\cdot)$  is invariant under the transition kernel

$$\bar{\gamma}(j|i) := \sum_u \varphi(u|i) q(j|i, u).$$

The dual LP is :

Minimize  $\bar{\lambda}$  subject to

$$\bar{\lambda} \geq \lambda(i),$$

$$\lambda(i) + V(i) \geq \sum_j q(j|i, u) (\tilde{r}(i, (u, q(\cdot|i, u)), j) + V(j)),$$

$$\lambda(i) \geq \sum_j q(j|i, u) \lambda(j),$$

$$\forall i \in S, (u, q(\cdot|i, u)) \in \tilde{U}_i.$$

The proof goes through finite approximations.

Note that the LP has infinitely many constraints.

# Dynamic Programming

The equivalent *dynamic programming* formulation is:

$$\begin{aligned}\lambda^* &= \max_i \lambda(i), \\ \lambda(i) + V(i) &= \max_{u, q(\cdot|i, u)} \left( \sum_j q(j|i, u) (V(j) \right. \\ &\quad \left. + \tilde{r}(i, (u, q(\cdot|i, u), j))) \right), \quad (\dagger) \\ \lambda(i) &= \max_{(u, q(\cdot|i, u)) \in B_i} \sum_j q(j|i, u) \lambda(j), \\ &\quad \forall i \in S,\end{aligned}$$

where  $B_i$  is the Argmax in  $(\dagger)$ . Once again, the proof goes through finite approximations.



The maximization over  $q$  in (†) can be explicitly performed using the ‘Gibbs variational principle’ from statistical mechanics:

For fixed  $i, u$ , the maximum is attained at

$$q^*(j|i, u) := \frac{p(j|i, u)e^{r(i,u,j)+V(j)}}{\sum_k p(k|i, u)e^{r(i,u,k)+V(k)}} .$$

Substitute back, set

$$\Phi(i) := e^{V(i)}, \quad \Lambda(i) := e^{\lambda(i)}, \quad i \in S,$$

and exponentiate both sides of (†).

This leads to the multiplicative dynamic programming equations for infinite horizon risk-sensitive reward in the general degenerate case:

$$\Lambda(i)\Phi(i) = \max_u \sum_j p(j|i, u) \left( e^{r(i,u,j)} \Phi(j) \right), \quad (\dagger\dagger)$$

$$\Lambda(i) = \max_{u \in D_i} \sum_j \left( \frac{p(j|i, u) e^{r(i,u,j)} \Phi(j)}{\sum_k p(k|i, u) e^{r(i,u,k)} \Phi(k)} \right) \Lambda(j),$$

$i \in S,$

where  $D_i$  is the Argmax in  $(\dagger\dagger)$ . This is the analog of Howard-Kallenberg results for ergodic control.

Observe the occurrence of the '*twisted kernel*'.

## FUTURE PROBLEMS:

1. Extension to non-compact state spaces (cf. recent work of Cavazos-Cadena (Math. OR, 2018))
2. Degenerate case for diffusions

# References

1. V. Anantharam, V. S. Borkar, “A variational formula for risk-sensitive reward”, submitted.  
<http://arxiv.org/abs/1501.00676>
2. A. Arapostathis, V. S. Borkar, K. Suresh Kumar, “Risk-sensitive control and an abstract Collatz-Wielandt formula”, *Journal of Theoretical Probability*, May 2015.  
<http://arxiv.org/abs/1312.5834>

3. V. S. Borkar, “Linear and dynamic programming approaches to degenerate risk-sensitive reward processes’, *56th IEEE Conference on Decision and Control*, Dec. 12-15, 2017, Melbourne, Australia.

*THANK YOU !*