Rocking in Two by Two: From Collatz-Wielandt to Donsker-Varadhan (Extended abstract)

V. Anantharam  
Department of Electrical Engineering  
and Computer Sciences  
University of California, Berkeley, CA 94720.  
Email: ananth@berkeley.edu

V. S. Borkar  
Department of Electrical Engineering  
Indian Institute of Technology, Powai, Mumbai 400076.  
Email: borkar.vs@gmail.com

Abstract—We derive a variational formula for the optimal growth rate of reward in the infinite horizon risk-sensitive control problem for discrete time Markov decision processes with compact state and action spaces, extending a formula of Donsker and Varadhan for the Perron-Frobenius eigenvalue of a positive operator. This can be viewed as an abstract version of the Collatz-Wielandt formula for the Perron-Frobenius eigenvalue of a non-negative matrix. This leads to a concave maximization formulation of the problem of determining the optimal growth rate of risk-sensitive reward.

I. INTRODUCTION

The celebrated Courant-Fisher formula gives a variational characterization of the principal eigenvalue of a positive definite matrix. Somewhat less known is the Collatz-Wielandt formula for the Perron-Frobenius eigenvalue \( \lambda \) of a non-negative irreducible matrix \( Q = [q_{ij}] \in \mathbb{R}^{d \times d} \):

\[
\lambda = \lim_{n \to \infty} \frac{1}{n} \log \left( \max_{x \in \mathbb{R}^d, x > 0} \frac{e^{Q x}}{x} \right) = \lim_{n \to \infty} \frac{1}{n} \log \left( \min_{x \in \mathbb{R}^d, x > 0} \frac{e^{-Q x}}{x} \right).
\]

This has an intimate relationship with the asymptotic growth rate for mean multiplicative reward arising in the infinite time horizon risk-sensitive control in Markov decision theory setting. The optimal reward multiplier per step then turns out to be the Perron-Frobenius eigenvalue of a positive 1-homogeneous nonlinear operator. This statement, in fact, is equivalent to the associated multiplicative dynamic programming equation. The existence of this and the corresponding eigenfunction is ensured by the nonlinear Krein-Rutman theorem of [10, Theorem 3.1.1 and Proposition 3.1.5] under suitable conditions. An abstract version of the aforementioned Collatz-Wielandt formula characterizes this eigenvalue. We build on this to provide a variational formula for the optimal growth rate of reward in the spirit of the Donsker-Varadhan formula for Perron-Frobenius eigenvalue of a nonnegative matrix [5, section 3.1.2], [6], [9].

The details of the results presented here along with further ramifications and examples appear in [1]. They can be viewed as a discrete time counterpart of the results of [2].

II. NOTATION

Our notation is as follows: For a compact metric space \( \mathcal{X} \), \( \mathcal{M}(\mathcal{X}) \) and \( \mathcal{P}(\mathcal{X}) \) will denote respectively the space of finite (signed) Borel measures on \( \mathcal{X} \) and the space of probability measures on \( \mathcal{X} \), both with the topology of weak convergence. \( \mathcal{C}(\mathcal{X}) \) will denote the Banach space of continuous maps \( \mathcal{X} \to \mathbb{R} \) with the supremum norm, denoted by \( \| \cdot \| \). Thus \( \mathcal{M}(\mathcal{X}) \) is the dual Banach space of \( \mathcal{C}(\mathcal{X}) \), with the weak* topology. Let \( \mathcal{S} \) be a prescribed compact metric space called the state space and \( \mathcal{U} \) another compact metric space, called the action space. We shall consider an \( S \)-valued controlled Markov process \( (X_n, n \geq 0) \) controlled by a \( U \)-valued control process \( (Z_n, n \geq 0) \) defined as follows. Consider a complete probability space \( (\Omega, \mathcal{F}, \mathcal{P}) \) where \( \Omega := (\mathcal{S} \times \mathcal{U})^\mathbb{N} \) and \( \mathcal{F} \) is its product Borel \( \sigma \)-field. For \( \omega = [(\omega_0, \omega'_0), (\omega_1, \omega'_1), (\omega_2, \omega'_2), \ldots] \in \Omega \) with \( \omega_i \in \mathcal{S} \) and \( \omega'_i \in \mathcal{U} \) for all \( i \), define ‘canonical’ random variables \( X_i = \omega_i, Z_i = \omega'_i, i \geq 0 \). The probability measure \( \mathcal{P} \) on \( (\Omega, \mathcal{F}) \) is then the law of \( ((X_n, Z_n), n \geq 0) \) defined as follows. The law of \( X_0 \) is prescribed and the law of \( ((X_n, Z_n), n \geq 0) \) is constructed inductively. For this purpose, define two increasing families of sub-\( \sigma \)-fields of \( \mathcal{F} : \mathcal{F}_n := \sigma(X_m, m \leq n; Z_m, m < n) \) and \( \mathcal{F}_\infty := \sigma(X_m, m \leq n; Z_m, m \leq n) \) for \( n \geq 0 \). First define the conditional law of \( Z_0 \) given \( \mathcal{F}_{0} \) as \( \phi_0(du|X_0) \), where

\[
\phi_0(du|x_0) : \mathcal{S} \to \mathcal{P}(\mathcal{U})
\]

is a prescribed kernel, i.e. \( \phi_0(du|x) \) is a probability distribution in \( \mathcal{P}(\mathcal{U}) \) for all \( x \) and \( \phi_0(A|x) \) is Borel measurable in \( x \) for all Borel subsets \( A \subset \mathcal{U} \). Let \( P^0 \) denote the law of \( ((X_0, Z_0), (X_1, Z_1), \ldots, (X_n, Z_n)) \), defined as a probability measure on \( (\Omega, \mathcal{F}_n) \), starting with \( n = 0 \). Define the law of \( X_{n+1} \) given \( \mathcal{F}_n \) as \( p(dy|x_n, Z_n) \) where

\[
p(dy|x, u) : \mathcal{S} \times \mathcal{U} \to \mathcal{P}(\mathcal{S})
\]

is a prescribed kernel, i.e. \( p(dy|x, u) \) is a probability distribution in \( \mathcal{P}(\mathcal{S}) \) for all \( (x, u) \in \mathcal{S} \times \mathcal{U} \) and \( p(A|x, u) \) is Borel measurable in \( (x, u) \) for all Borel subsets \( A \subset \mathcal{S} \). Define the conditional law of \( Z_{n+1} \) given \( \mathcal{F}_{n+1} \) as

\[
\phi_{n+1}(du|(X_0, Z_0), \ldots, (X_n, Z_n), X_{n+1})
\]
where
\[ \phi_{n+1}(du(x_0, u_0), \ldots, x_n, u_n, x_{n+1}) : (S \times U)^n \times S \mapsto P(U) \]
is a prescribed kernel for each \( n \). These together define \( P_{n+1} \).
By the Ionescu-Tulcea theorem (p. 101, [11]), we define a unique \( P \) on \((\Omega, F)\). By construction, for all Borel \( A \subset S \),
\[ P(X_{n+1} \in A | F_n) = P(X_{n+1} \in A | X_n, Z_n) = p(A | X_n, Z_n). \]
The \((Z_n, n \geq 0)\) constructed above will be referred to as admissible controls. We shall also consider two special classes of admissible controls: stationary Markov controls of the form
\[ Z_n = v(X_n) \quad \forall n, \]
for some measurable \( v : S \mapsto U \), and randomized stationary Markov controls satisfying
\[ P(Z_n \in A | F_n) = P(Z_n \in A | X_n) = \varphi(A | X_n) \quad \forall n, \]
for all Borel \( A \subset U \) and some kernel \( \varphi(du(x) : S \mapsto P(U) \).
By a standard abuse of terminology, we identify these with the maps \( v(\cdot) \), \( \varphi(\cdot | \cdot) \) resp. The sets thereof will be denoted by \( SM \) and \( RM \) respectively. We view \( SM \) as a subset of \( RM \) by identifying \( v(\cdot) \) with \( \delta_{v(\cdot)} \), the Dirac measure at \( v(\cdot) \).

The infinite horizon risk-sensitive reward we seek to characterize is
\[ \lambda := \sup_{\varphi \in SM} \lim_{N \to \infty} \frac{1}{N} \times \int \log E \left[ e^{\sum_{m=0}^{N-1} r(X_m, Z_m, X_{m+1}) | X_0 = x} \right], \quad (2) \]
where the second supremum is over all admissible controls. Here \( r(x, u, y) \) is the ‘per stage reward’ on transitioning from \( x \) to \( y \) under action \( u \). It should be noted that we will allow \( e^{r(x, u, y)} = 0 \) for some \( (x, u, y) \), so \( r(x, u, y) \) should be thought of as being allowed to the extended real value \(-\infty\). Throughout the paper, we make the following assumptions about \( r(x, u, y) \) and \( p(dy|x, u) \).

(A0): \( e^{r(x, u, y)} \in C(S \times U \times S) \). Furthermore, we assume for the time being that
\[ \int e^{r(x, u, y)} > 0 \quad \forall x, u, y. \]

Note that \( \int e^{r(x, u, y)} > 0 \quad \forall x, u, y \).

The first condition in (A1) is true, e.g., if \( S \) is a compact subset of \( \mathbb{R}^d \), \( U \) is a compact metric space, and \( p(dy|x, u) = \psi(y|x, u) dy \) with \( \psi(y | \cdot, \cdot) \in S \), equicontinuous. If in addition \( \psi(y | \cdot, \cdot) > 0 \), then \( \int p(dy|x, u) \) also holds. We shall drop the latter condition in our final result.

If \( p(dx) \) and \( q(dx) \) are finite nonnegative Borel measures on a compact metric space \( X \), we write \( D(p(dx) || q(dx)) \) for the relative entropy of \( p(dx) \) with respect to \( q(dx) \), defined by
\[ D(p(dx) || q(dx)) := \int p(dx) \log l(x) \quad \text{if we can write} \quad p(dx) = l(x)q(dx) \]
\[ := \infty \quad \text{otherwise.} \]

See e.g. [12] for some of the basic properties of relative entropy.

III. MAIN RESULTS

Define the operator \( T : C(S) \mapsto C(S) \) by
\[ Tf(x) := \sup_{\varphi \in P(U)} \int \int p(dy|x, u) \varphi(du)e^{r(x, u, y)} \psi(y), \quad (3) \]
with \( \rho \) given by
\[ \rho \psi(x) = \sup_{\varphi \in P(U)} \int \int p(dy|x, u) \varphi(du)e^{r(x, u, y)} \psi(y), \quad (3) \]
and \( \forall x \in S \),
\[ \log \rho = \sup_{n \to \infty} \frac{1}{n} \int \log E \left[ e^{\sum_{m=0}^{n-1} r(X_m, Z_m, X_{m+1}) | X_0 = x} \right], \]
where the supremum on the right is over all admissible controls and in fact is a maximum attained at some \( v^*(\cdot) \in SM \).

The formula (4) is an abstract version of the Collatz-Wielandt formula, see also [4]. An immediate consequence is that \( \log \rho \) equals the optimal growth rate of reward defined in (2). Our main result is Theorem 2 (ii) below, which follows by a limiting argument from part (i) thereof.

Theorem 2. (i) Under assumptions (A0), (A1), the optimal growth rate of reward \( \lambda \), as defined in (2), has the variational characterization
\[ \lambda = \log \rho = \sup_{\eta \in \mathcal{U}} \left( \int \int \eta(dx, du, dy)r(x, u, y) \right. \]
\[ - \int \int \eta(dx, du)D(\eta_2(dy|x, u) || p(dy|x, u)) \right), \]
where $\rho$ is defined as in Theorem 1.

(ii) The above holds even when $(\dagger), (\dagger\dagger)$ are dropped. \hfill $\square$

Note that we have reduced the problem of reward maximization with risk-sensitive criterion to a concave maximization problem. This is in contrast with a team problem that we would get if we follow the standard route through log transformation as in [7]-[8].

In [1], we discuss the following applications of the foregoing:

1) Consider a directed graph and a controlled Markov chain on this graph wherein at each node one is allowed to make any, possibly randomized, transition to one of its successors. The objective is to do this so as to maximize the exponential growth rate of the number of directed paths of length $n$ from the starting node as $n \uparrow \infty$. This has potential applications in constrained coding.

2) A standard portfolio optimization problem for maximization of exponential growth rate of wealth with risk-sensitive criterion has been analyzed in [3] by converting it to a cost minimization problem to which available machinery of risk-sensitive control can then be applied. Our framework allows us to consider the original reward maximization formulation and convert it to a concave maximization problem.

3) The problem of controlling the asymptotic exit rate from a domain is also amenable to this analysis.

ACKNOWLEDGMENT

Research of VA supported in part by the ARO MURI grant W911NF-08-1-0233, Tools for the Analysis and Design of Complex Multi-Scale Networks, the NSF grants CNS-0910702 and ECCS-1343398, and the NSF Science & Technology Center grant CCF-0939370, Science of Information. A part of this work was done while this author was visiting IIT Bombay. Research of VSB supported in part by a J. C. Bose Fellowship and grant 11IRCCSG014 from IIT Bombay. A part of this work was done while this author was visiting the University of California, Berkeley.

REFERENCES


