Rational Inattention in Controlled Markov Processes

Ehsan Shafieepoorfard, Maxim Raginsky and Sean P. Meyn

***Abstract*— The paper poses a general model for optimal control subject to information constraints, motivated in part by recent work on information-constrained decision-making by economic agents. In the average-cost optimal control framework, the general model introduced in this paper reduces to a variant of the linear-programming representation of the average-cost optimal control problem, subject to an additional mutual information constraint on the randomized stationary policy. The resulting optimization problem is convex and admits a decomposition based on the Bellman error, which is the object of study in approximate dynamic programming. The structural results presented in this paper can be used to obtain performance bounds, as well as algorithms for computation or approximation of optimal policies.**

I. INTRODUCTION

In typical applications of stochastic dynamic programming, the controller has access to limited information about the state of the system. Unlike much of the existing literature on problems with imperfect state information, in this paper it is assumed that the system designer has to decide not only about the control policy, but also about the observation channel based on which the control is derived. (A very partial list of references that touch upon similar themes includes [1]–[5].) There are various applications which fit in this framework, either in engineering or economics.

In economic decision-making, the amount of information required to make a truly optimal decision will typically exceed what an agent can handle. In his seminal work [6], Christopher Sims (who has shared the 2011 Nobel Memorial Prize in Economics with Thomas Sargent) adds an information-processing constraint to a specific kind of dynamic programming problem which is frequently used in macroeconomic models. The aim, as he expresses, is to support John Maynard Keynes' well-known view that real economic behavior is inconsistent with the idea of continuously optimizing agents interacting in continuously clearing markets. Sims uses the term "rational inattention" to describe the setting in which information-constrained agents strive to make the best use of whatever information they are able to handle.

E. Shafieepoorfard and M. Raginsky are with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA; {shafiee1,maxim}@illinois.edu.

S.P. Meyn is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA; meyn@ece.ufl.edu.

Research supported in part by the NSF under CAREER award no. CCF-1254041 and grant no. ECCS-1135598, and in part by the UIUC College of Engineering under Strategic Research Initiative on "Cognitive and Algorithmic Decision Making."

Sims considers a model in which a representative agent decides about his consumption over subsequent periods of time, while his computational ability to reckon his wealth – the state of the dynamic system – is limited. A special case is considered in which income in one period adds uncertainty of wealth in the next period. Other modeling assumptions reduce the model to an LQG control problem. As one justification for introducing the information constraint, Sims remarks [7] that "most people are only vaguely aware of their net worth, are little-influenced in their current behavior by the status of their retirement account, and can be induced to make large changes in savings behavior by minor 'informational' changes, like changes in default options on retirement plans." Quantitatively, the information constraint is stated in terms of an upper bound on the mutual information in the sense of Shannon [8] between the state of the system and the observation available to the agent.

Extensions of this work include Matejka and McKay [9], who recovered the well-known multinomial logit model for a situation where an individual must choose among discrete alternatives yielding different values not perfectly known ex-ante. In the sequel [10], they extend this model to study Nash equilibria in a Bertrand oligopoly setting where N firms produce the same commodity with different qualities not perfectly known ex-ante; it is shown that this information friction leads to increased prices for the commodity. On a similar theme, Peng [11] is motivated by the observation that "investors have limited time and attention to process information." In a factor model for asset returns, he investigates the dynamics of prices and mutual information when there is uncertainty in the decision making process.

Given the appeal and generality of these questions, there is ample motivation for the creation of a general theory for optimal control subject to information constraints. The contribution of the present paper is to initiate the development of such a theory, which would in turn enable us to address many problems in macroeconomics and engineering in a systematic fashion. We focus on the average-cost optimal control framework and show that the construction of an optimal information-constrained controller reduces to a variant of the linear-programming representation of the average-cost optimal control problem, subject to an additional mutual information constraint on the randomized stationary policy. The resulting optimization problem is convex and admits a decomposition in terms of the Bellman error, which is the object of study in approximate dynamic programming. The structure revealed in this paper can be used to obtain performance bounds, as well as algorithms for computation or approximation of optimal policies.

II. Preliminaries and notation

All spaces are assumed to be standard Borel (i.e., isomorphic to a Borel subset of a complete separable metric space), and will be equipped with their Borel σ-algebras. If X is such a space, then $\mathcal{B}(\mathsf{X})$ will denote its Borel σ-algebra, and $\mathcal{P}(\mathsf{X})$ will denote the space of all probability measures on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$. We will denote by $M(\mathsf{X})$ the space of all measurable functions $f : \mathsf{X} \to \mathbb{R}$ and by $C_b(\mathsf{X}) \subseteq M(\mathsf{X})$ the space of all bounded continuous functions. We will often use bilinear form notation for expectations: for any $f \in L^1(\mu)$,

$$\langle \mu, f \rangle \triangleq \int_{\mathsf{X}} f(x)\mu(\mathrm{d}x) = \mathbb{E}[f(X)],$$

where in the last expression it is understood that X is an X-valued random object with $\mathrm{Law}(X) = \mu$.

Given two spaces X and Y, a mapping $K(\cdot|\cdot) : \mathcal{B}(\mathsf{Y}) \times \mathsf{X} \to [0,1]$ is a *Markov* (or *stochastic*) *kernel* if $K(\cdot|x) \in \mathcal{P}(\mathsf{Y})$ for all $x \in \mathsf{X}$ and $x \mapsto K(B|x)$ is measurable for every $B \in \mathcal{B}(\mathsf{Y})$. The space of all such Markov kernels will be denoted by $\mathcal{M}(\mathsf{Y}|\mathsf{X})$. Markov kernels $K \in \mathcal{M}(\mathsf{Y}|\mathsf{X})$ act on measurable functions $f \in M(\mathsf{Y})$ from the left as

$$Kf(x) \triangleq \int_{\mathsf{Y}} f(y)K(\mathrm{d}y|x), \qquad \forall x \in \mathsf{X}$$

and on probability measures $\mu \in \mathcal{P}(\mathsf{X})$ from the right as

$$\mu K(B) \triangleq \int_{\mathsf{X}} K(B|x)\mu(\mathrm{d}x), \qquad \forall B \in \mathcal{B}(\mathsf{Y}).$$

Moreover, $Kf \in M(\mathsf{X})$ for any $f \in M(\mathsf{Y})$, and $\mu K \in \mathcal{P}(\mathsf{Y})$ for any $\mu \in \mathcal{P}(\mathsf{X})$. The *relative entropy* (or *information divergence*) between any two $\mu, \nu \in \mathcal{P}(\mathsf{X})$ [8] is defined as

$$D(\mu\|\nu) \triangleq \begin{cases} \left\langle \mu, \log \dfrac{\mathrm{d}\mu}{\mathrm{d}\nu} \right\rangle, & \text{if } \mu \prec \nu \\ +\infty, & \text{otherwise} \end{cases}$$

Given any probability measure $\mu \in \mathcal{P}(\mathsf{X})$ and any Markov kernel $K \in \mathcal{M}(\mathsf{Y}|\mathsf{X})$, we can define a probability measure $\mu \otimes K$ on the product space $(\mathsf{X} \times \mathsf{Y}, \mathcal{B}(\mathsf{X}) \otimes \mathcal{B}(\mathsf{Y}))$ via its action on the rectangles $A \times B$, $A \in \mathcal{B}(\mathsf{X}), B \in \mathcal{B}(\mathsf{Y})$:

$$(\mu \otimes K)(A \times B) \triangleq \int_A K(B|x)\mu(\mathrm{d}x).$$

Note that $\mu \otimes K(\mathsf{X} \times B) = \mu K(B)$ for all $B \in \mathcal{B}(\mathsf{X})$. The *Shannon mutual information* [8] in the pair (μ, K) is

$$I(\mu, K) \triangleq D(\mu \otimes K \| \mu \otimes \mu K), \tag{1}$$

where, for any $\mu \in \mathcal{P}(\mathsf{X})$ and $\nu \in \mathcal{P}(\mathsf{Y})$, $\mu \otimes \nu$ denotes the product measure defined via $(\mu \otimes \nu)(A \times B) \triangleq \mu(A)\nu(B)$ for all $A \in \mathcal{B}(\mathsf{X}), B \in \mathcal{B}(\mathsf{Y})$. The functional $I(\mu, K)$ is concave in μ and convex in K. If (X, Y) is a pair of random objects with $\mathrm{Law}(X, Y) = \Gamma = \mu \otimes K$, then we will also write $I(X; Y)$ or $I(\Gamma)$ for $I(\mu, K)$.

Finally, given a triple of jointly distributed random objects (X, Y, Z) with $\Gamma = \mathrm{Law}(X, Y, Z)$, we will say that they form a *Markov chain* $X \to Y \to Z$ if there exist Markov kernels $K_1 \in \mathcal{M}(\mathsf{Y}|\mathsf{X})$ and $K_2 \in \mathcal{M}(\mathsf{Z}|\mathsf{Y})$ so that Γ can be disintegrated as $\Gamma = \mu \otimes K_1 \otimes K_2$ (in words, if X and Z are conditionally independent given Y). The mutual information satisfies the *data processing inequality*: if $X \to Y \to Z$ is a Markov chain, then

$$I(X; Z) \le I(X; Y). \tag{2}$$

In words, no additional processing can increase information.

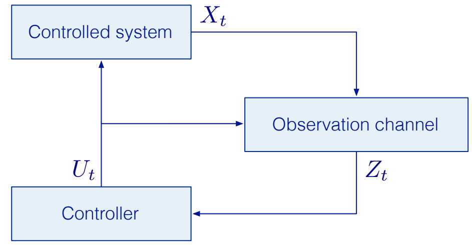

Fig. 1. System model.

III. Some general considerations

Consider a model in which the controller is constrained to observe the system being controlled through an information-limited channel. This is illustrated in the block diagram shown in Figure 1, consisting of:

- the (time-invariant) controlled system, specified by an initial condition $\mu \in \mathcal{P}(\mathsf{X})$ and a stochastic kernel $Q \in \mathcal{M}(\mathsf{X}|\mathsf{X} \times \mathsf{U})$, where X is the state space and U is the control (or action) space;
- the observation channel, specified by a sequence $\underline{W}$ of stochastic kernels $W_t \in \mathcal{M}(\mathsf{Z}|\mathsf{X}^t \times \mathsf{Z}^{t-1} \times \mathsf{U}^{t-1})$, $t = 1, 2, \ldots$, where Z is some observation space[1]; and
- the feedback controller, specified by a sequence $\underline{\Phi}$ of stochastic kernels $\Phi_t \in \mathcal{M}(\mathsf{U}|\mathsf{Z})$, $t = 1, 2, \ldots$.

The X-valued state process $\{X_t\}_{t=1}^\infty$, the Z-valued observation process $\{Z_t\}_{t=1}^\infty$, and the U-valued control process $\{U_t\}_{t=1}^\infty$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and have the causal ordering

$$X_1, Z_1, U_1, \ldots, X_t, Z_t, U_t, \ldots \tag{3}$$

where, $\mathbb{P}$-almost surely, $\mathbb{P}(X_1 \in A) = \mu(A)$ for all $A \in \mathcal{B}(\mathsf{X})$, and for all $t = 1, 2, \ldots$, $B \in \mathcal{B}(\mathsf{Z})$, $C \in \mathcal{B}(\mathsf{U})$, $D \in \mathcal{B}(\mathsf{X})$ we have

$$\mathbb{P}(Z_t \in B | X^t, Z^{t-1}, U^{t-1}) = W_t(B | X^t, Z^{t-1}, U^{t-1}) \tag{4a}$$

$$\mathbb{P}(U_t \in C | X^t, Z^t, U^{t-1}) = \Phi_t(C | Z_t) \tag{4b}$$

$$\mathbb{P}(X_{t+1} \in D | X^t, Z^t, U^t) = Q(D | X_t, U_t) \tag{4c}$$

This specification ensures that, for each t, the next state X_{t+1} is conditionally independent of X^{t-1}, Z^t, U^{t-1} given X_t, U_t (which is the usual case of a controlled Markov process), and that the control U_t is conditionally independent of X^t, Z^{t-1}, U^{t-1} given Z_t. In other words, at each time t the controller takes as input only the most recent observation Z_t, which amounts to the assumption that there is a *separation*

[1] This observation space can be very general — for example, the case when Z is a Cartesian product of countably many copies of $\mathsf{X} \times \mathsf{U}$ is allowed, so we can cover the case of arbitrarily long (or even perfect) controller memory.

structure between the observation channel and the controller. This assumption is common in the literature [1], [3], [4].

In the spirit of the rational inattention framework, we assume that the amount of information flow that can be maintained across the observation channel per time step is constrained, and wish to *design* a suitable channel $\underline{W}$ and a controller $\underline{\Phi}$ to minimize a given performance objective under the information constraint. For maximum flexibility, we grant the system designer the freedom to choose the observation space Z as well. In other words, the designer is allowed to choose an optimal *representation* for the data supplied to the controller.

As we will demonstrate shortly, the choice $\mathsf{Z} = \mathcal{P}(\mathsf{X})$ (i.e., letting Z be the space of beliefs on the state space) is information-theoretically optimal. For now, let us see how much insight we can gain in the case of a fixed Z. Then the problem of *optimal control with rational inattention* can be formulated as follows. Let $c : \mathsf{X} \times \mathsf{U} \to \mathbb{R}^+$ be a given measurable one-step state-action cost function. Given a fixed finite horizon T and some $R \geq 0$, we seek an observation channel $\underline{W}$ and a controller $\underline{\Phi}$ in order to

$$\text{minimize } \mathbb{E}\left[\sum_{t=1}^{T} c(X_t, U_t)\right] \tag{5a}$$

$$\text{subject to } I(X_t; Z_t) \leq R, \ t = 1, 2, \ldots, T \tag{5b}$$

Later we focus on a related average-cost performance criterion.

A. Key structural result

The optimization problem (5) seems formidable: for each time step $t = 1, \ldots, T$ we must design stochastic kernels $W_t(\mathrm{d}z_t | x^t, z^{t-1}, u^{t-1})$ and $\Phi_t(\mathrm{d}u_t | z_t)$ for the observation channel and the controller, and the complexity of the feasible set of W_t's grows with t. However, the fact that (a) both the controlled system and the controller are Markov, and (b) the cost function at each stage depends only on the current state-action pair, permits a drastic simplification — at each time t, we can limit our search to *memoryless* channels $W_t(\mathrm{d}z_t | x_t)$ without impacting either the average cost in (5a) or the information flow constraint in (5b):

Theorem 1 (Memoryless observation channels suffice)**.** *Under the specified information pattern* (4)*, for any controller specification* $\underline{\Phi}$ *and any channel specification* $\underline{W}$*, there exists another channel specification* $\underline{W}'$ *consisting of stochastic kernels* $W_t(\mathrm{d}z_t | x_t)$*,* $t = 1, 2, \ldots$*, such that*

$$\mathbb{E}\left[\sum_{t=1}^{T} c(X'_t, U'_t)\right] = \mathbb{E}\left[\sum_{t=1}^{T} c(X_t, U_t)\right]$$

and

$$I(X'_t; Z'_t) = I(X_t; Z_t), \qquad t = 1, 2, \ldots, T$$

where $\{(X_t, U_t, Z_t)\}$ *is the original process with* $(\mu, Q, \underline{W}, \underline{\Phi})$*, while* $\{X'_t, U'_t, Z'_t)\}$ *is the one with* $(\mu, Q, \underline{W}', \underline{\Phi})$*.*

Proof. To prove the theorem, we follow the approach used by Wistenhausen in [12]. We start with the following simple observation that can be regarded as an instance of the Shannon–Mori–Zwanzig Markov model [13]:

Principle of Irrelevant Information. Let $\Xi, \Theta, \Psi, \Upsilon$ be four random variables defined on a common probability space, such that Υ is conditionally independent of (Θ, Ξ) given Ψ. Then there exist four random variables $\Xi', \Theta', \Psi', \Upsilon'$ defined on the same spaces as the original tuple, such that $\Xi' \to \Theta' \to \Psi' \to \Upsilon'$ is a *Markov chain*, and moreover the bivariate marginals agree:

$$\begin{aligned}
\mathrm{Law}(\Xi, \Theta) &= \mathrm{Law}(\Xi', \Theta') \\
\mathrm{Law}(\Theta, \Psi) &= \mathrm{Law}(\Theta', \Psi') \\
\mathrm{Law}(\Psi, \Upsilon) &= \mathrm{Law}(\Psi', \Upsilon')
\end{aligned}$$

Using this principle, we can prove the following two lemmas (see Appendices for the proofs):

Lemma 1 (Two-Stage Lemma)**.** *Suppose* $T = 2$*. Then the kernel* $W_2(\mathrm{d}z_2 | x^2, z_1, u_1)$ *can be replaced by another kernel* $W'_2(\mathrm{d}z_2 | x_2)$*, such that the resulting variables* (X'_t, Z'_t, U'_t)*,* $t = 1, 2$*, satisfy*

$$\mathbb{E}[c(X'_1, U'_1) + c(X'_2, U'_2)] = \mathbb{E}[c(X_1, U_1) + c(X_2, U_2)]$$

and $I(X'_t; Z'_t) = I(X_t; Z_t)$*,* $t = 1, 2$*.*

Lemma 2 (Three-Stage Lemma)**.** *Suppose* $T = 3$*, and* Z_3 *is conditionally independent of* (X_i, Z_i, U_i)*,* $i = 1, 2$*, given* X_3*. Then the kernel* $W_2(\mathrm{d}z_2 | x^2, z_1, u_1)$ *can be replaced by another kernel* $W'_2(\mathrm{d}z_2 | x_2)$*, such that the resulting variables* (X'_i, Z'_i, U'_i)*,* $i = 1, 2, 3$*, satisfy*

$$\mathbb{E}\left[\sum_{t=1}^{3} c(X'_t, U'_t)\right] = \mathbb{E}\left[\sum_{t=1}^{3} c(X_t, U_t)\right]$$

and $I(X'_t; Z'_t) = I(X_t; Z_t)$ *for* $t = 1, 2, 3$*.*

Armed with these two lemmas, we can now prove the theorem by backward induction and grouping of variables. Fix any T. By the Two-Stage-Lemma, we may assume that W_T is memoryless, i.e., Z_T is conditionally independent of $X^{T-1}, Z^{T-1}, U^{T-1}$ given X_T. Now we apply the Three-Stage Lemma to

$$\Big| \underbrace{X^{T-3}, Z^{T-3}, U^{T-3}, X_{T-2}}_{\substack{\text{Stage 1} \\ \text{state}}}, \underbrace{Z_{T-2}}_{\substack{\text{Stage 1} \\ \text{observation}}}, \underbrace{U_{T-2}}_{\substack{\text{Stage 1} \\ \text{control}}} \Big|$$
$$\Big| \underbrace{X_{T-1}}_{\substack{\text{Stage 2} \\ \text{state}}}, \underbrace{Z_{T-1}}_{\substack{\text{Stage 2} \\ \text{observation}}}, \underbrace{U_{T-1}}_{\substack{\text{Stage 2} \\ \text{control}}} \Big| \underbrace{X_T}_{\substack{\text{Stage 3} \\ \text{state}}}, \underbrace{Z_T}_{\substack{\text{Stage 3} \\ \text{observation}}}, \underbrace{U_T}_{\substack{\text{Stage 3} \\ \text{control}}} \Big| \tag{6}$$

to replace $W_{T-1}(\mathrm{d}z_{T-1} | x^{T-1}, z^{T-2}, u^{T-2})$ with $W'_{T-1}(\mathrm{d}z_{T-1} | x_{T-1})$ without affecting the expected cost or the mutual information between the state and the observation at time $T-1$. We proceed inductively by merging the second and the third stages in (6), splitting the first stage in (6) into two, and then applying the Three-Stage Lemma to replace the original observation kernel W_{T-2} with a memoryless one. □

B. Long-term average cost criterion

Despite the simplification afforded by Theorem 1, the optimization problem (5) is still difficult even when the observation space Z is fixed, and the only general way of solving it is via infinite-dimensional dynamic programming.

Our goal is to gain theoretical insight into the structure of optimal control policies in the rational inattention framework. To that end, we make several simplifications:

1) We replace the finite-horizon cost criterion in (5a) with the one based on the long-term average cost.
2) We consider only stationary (time-invariant) observation channels and controllers.
3) Instead of separately optimizing over the observation space, the observation channel, and the controller, we collapse these decision variables into the choice of a Markov randomized stationary (MRS) control law $\Phi \in \mathcal{M}(\mathsf{U}|\mathsf{X})$ satisfying the information constraint.

Of these, Item 3) requires some explanation; in essence, it is justified by the simple fact that the mutual information $I(X;U)$ between two random variables (X,U) can be expressed as

$$I(X;U) = \inf\Big\{I(X;Z) : X \to Z \to U\Big\},$$

where the infimum is over all standard Borel spaces Z and all Z-valued random objects Z jointly distributed with (X,U), such that $X \to Z \to U$ forms a Markov chain. Indeed, for any such triple we have $I(X;U) \le I(X;Z)$ by the data processing inequality; the other direction follows by considering the "degenerate" Markov chain $X \to U \to U$. Consequently, for any choice of $\mu \in \mathcal{P}(\mathsf{X})$, Z, $W \in \mathcal{M}(\mathsf{Z}|\mathsf{X})$ and $\Phi \in \mathcal{M}(\mathsf{U}|\mathsf{Z})$, the kernel $\Psi = \Phi \circ W \in \mathcal{M}(\mathsf{U}|\mathsf{X})$ will satisfy $I(\mu,\Psi) \le I(\mu,W)$ (recall notation from (1)). Conversely, we can always factor any given $\Psi \in \mathcal{M}(\mathsf{U}|\mathsf{X})$ through some standard Borel space Z as $\Psi = \Phi \circ W$ with $\Phi \in \mathcal{M}(\mathsf{U}|\mathsf{Z})$ and $W \in \mathcal{M}(\mathsf{Z}|\mathsf{X})$, such that $I(\mu,W) = I(\mu,\Psi)$.

In view of the above, we focus on the following problem: find an MRS control law $\Phi \in \mathcal{M}(\mathsf{U}|\mathsf{X})$ to

$$\text{minimize } \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} c(X_t,U_t)\right] \tag{7a}$$
$$\text{subject to } \limsup_{t\to\infty} I(X_t;U_t) \le R,\ t = 1,2,\dots \tag{7b}$$

IV. One-stage problem: solution via rate-distortion theory

Before we analyze the infinite-horizon problem (7), let us show that the one-stage case can be solved completely using *rate-distortion theory* [14] (a branch of information theory that deals with optimal compression of data subject to information constraints). To the best of our knowledge, this solution is originally due to Stratonovich [15] (see also [16, Sec. 9.7]). Because our subsequent development builds on these ideas, we briefly describe them here. Moreover, this analysis will provide additional justification for eliminating the decision variables Z and $W \in \mathcal{M}(\mathsf{Z}|\mathsf{X})$ in favor of an information-constrained controller $\Phi \in \mathcal{M}(\mathsf{U}|\mathsf{X})$ directly connected to the system being controlled.

When $T = 1$, the problem in (5) becomes

$$\text{minimize } \mathbb{E}\left[c(X,U)\right] \tag{8a}$$
$$\text{subject to } I(X;Z) \le R \tag{8b}$$

for a given law $\mu \in \mathcal{P}(\mathsf{X})$ of X, where the minimization is over all observation channels $W \in \mathcal{M}(\mathsf{Z}|\mathsf{X})$ and controllers $\Phi \in \mathcal{M}(\mathsf{U}|\mathsf{Z})$. If we denote the optimum value attained in (8) by $V(R,\mathsf{Z})$, then the quantity of interest is

$$V(R) \triangleq \inf_{\mathsf{Z}} V(R,\mathsf{Z}), \tag{9}$$

where the infimum is over all standard Borel observation spaces. In other words, we seek an observation space Z^*, an observation channel $W^* \in \mathcal{M}(\mathsf{Z}^*|\mathsf{X})$, and a controller $\Phi^* \in \mathcal{M}(\mathsf{U}|\mathsf{Z}^*)$, such that $I(X;Z^*) \le R$ and the resulting expected cost $\mathbb{E}[c(X,U^*)]$ is minimized.

In order to properly frame the main result of [15], we need to make some preliminary observations. If we fix Z and an observation channel $W \in \mathcal{M}(\mathsf{Z}|\mathsf{X})$, then the optimal choice of the controller $\Phi^*_W \in \mathcal{M}(\mathsf{U}|\mathsf{Z})$ is to let $\Phi^*_W(\mathrm{d}u|z)$ be supported on the set of all $u^* \in \mathsf{U}$, such that

$$\mathbb{E}[c(X,u^*)|Z=z] = \min_{u\in\mathsf{U}} \mathbb{E}[c(X,u)|Z=z].$$

In fact, under suitable assumptions on c, there exists a deterministic *measurable selector* $\varphi^*_W : \mathsf{Z} \to \mathsf{U}$, so that

$$\mathbb{E}[c(X,\varphi^*_W(z))|Z=z] = \min_{u\in\mathsf{U}} \mathbb{E}[c(X,u)|Z=z].$$

With this, we would then use the deterministic controller $\Phi^*_W(\mathrm{d}u|z) = \delta_{\varphi^*_W(z)}(\mathrm{d}u)$, where δ_u denotes the Dirac measure concentrated at $u \in \mathsf{U}$. Thus,

$$\begin{aligned} V(R,\mathsf{Z}) &= \inf_{\substack{W\in\mathcal{M}(\mathsf{Z}|\mathsf{X})\\ I(X;Z)\le R}} \mathbb{E}[c(X,\varphi^*_W(Z))] \\ &= \inf_{\substack{W\in\mathcal{M}(\mathsf{Z}|\mathsf{X})\\ I(X;Z)\le R}} \mathbb{E}\left[\min_{u\in\mathsf{U}} \mathbb{E}[c(X,u)|Z]\right]. \end{aligned} \tag{10}$$

We also need to introduce some notions from rate-distortion theory [14]. For a given $\mu \in \mathcal{P}(\mathsf{X})$ and a given $R \ge 0$, consider the set

$$\mathcal{I}_\mu(R) \triangleq \Big\{K \in \mathcal{M}(\mathsf{U}|\mathsf{X}) : I(\mu,K) \le R\Big\}.$$

The set $\mathcal{I}_\mu(R)$ is nonempty for every $R \ge 0$. To see this, note that any kernel $K_\diamond \in \mathcal{M}(\mathsf{U}|\mathsf{X})$ for which the function $x \mapsto K_\diamond(B|x)$ is constant (μ-a.e. for any $B \in \mathcal{B}(\mathsf{U})$) satisfies $I(\mu,K_\diamond) = 0$.

The *Shannon distortion-rate function* (DRF) of μ is defined as

$$D_\mu(R) \triangleq \inf_{K\in\mathcal{I}_\mu(R)} \langle \mu\otimes K, c\rangle. \tag{11}$$

We use the more cumbersome notation $D_\mu(R;c)$ when we need to specify the dependence of the DRF on the underlying cost function c. Starting with the easily proved variational expression

$$I(\mu,K) = \inf_{\nu\in\mathcal{P}(\mathsf{U})} D(\mu\otimes K \| \mu\otimes\nu)$$

(where the infimum is achieved uniquely by $\nu = \mu K$), we can introduce the Lagrangian relaxation

$$\mathcal{L}_\mu(K,\nu,s) \triangleq sD(\mu \otimes K \| \mu \otimes \nu) + \langle \mu \otimes K, c \rangle$$

for $s \geq 0$ and $\nu \in \mathcal{P}(\mathsf{U})$, and establish the following key results [14], [17]:

Proposition 1. *The DRF $D_\mu(R)$ is convex and nondecreasing in R. If $D_\mu(R) < \infty$, then a Markov kernel $K^* \in \mathcal{M}(\mathsf{U}|\mathsf{X})$ attains the infimum in* (11) *if and only if $I(\mu, K^*) = R$ and the Radon–Nikodym derivative of $\mu \otimes K^*$ w.r.t. $\mu \otimes \mu K^*$ takes the form*

$$\frac{\mathrm{d}(\mu \otimes K^*)}{\mathrm{d}(\mu \otimes \mu K^*)}(x,y) = \alpha(x) e^{-\frac{1}{s}c(x,u)} \tag{12}$$

where $\alpha : \mathsf{X} \to \mathbb{R}^+$ and $s \geq 0$ are such that

$$\int_{\mathsf{X}} \alpha(x) e^{-\frac{1}{s}c(x,u)} \mu(\mathrm{d}x) \leq 1, \qquad \forall u \in \mathsf{U} \tag{13}$$

and $-s$ is the slope of a line tangent to the graph of $D_\mu(R)$ at R:

$$D_\mu(R') + sR' \geq D_\mu(R) + sR, \qquad \forall R' \geq 0. \tag{14}$$

Proposition 2. *The DRF $D_\mu(R)$ can be expressed as*

$$D_\mu(R) = \sup_{s \geq 0} \inf_{\nu \in \mathcal{P}(\mathsf{U})} s \left[\left\langle \mu, \log \frac{1}{\int_{\mathsf{U}} e^{-\frac{1}{s}c(x,u)} \nu(\mathrm{d}u)} \right\rangle - R \right]$$

We are now in a position to state and prove the main result of this section:

Theorem 2 (Stratonovich)**.** *For any $R \geq 0$ such that $D_\mu(R) < \infty$, we have $V(R) = D_\mu(R)$, and the infimum over Z in* (9) *is attained at $\mathsf{Z}^* = \mathcal{P}(\mathsf{X})$.*

Proof (Sketch). One direction, $V(R) \geq D_\mu(R)$, is relatively straightforward. Fix Z, and suppose that $W^* \in \mathcal{M}(\mathsf{Z}|\mathsf{X})$ and $\Phi^* \in \mathcal{M}(\mathsf{U}|\mathsf{Z})$ attain the optimal value $V(R, \mathsf{Z})$. Consider the Markov kernel $K \in \mathcal{M}(\mathsf{U}|\mathsf{X})$ obtained by composing Φ^* and W^*: $K = \Phi^* \circ W^*$. The joint action of W^* and Φ^* can be described by a Markov chain $X \to Z^* \to U^*$, where $\mathrm{Law}(X) = \mu$, $\mathrm{Law}(Z^*|X = x) = W^*(\cdot|x)$, and $\mathrm{Law}(U^*|X = x, Z^* = z) = \mathrm{Law}(U|Z^* = z) = \Phi^*(\cdot|z)$. Moreover, $\mathrm{Law}(X, U^*) = \mu \otimes K$. Since $I(X; Z^*) \leq R$, we have $K \in \mathcal{I}_\mu(R)$ by the data processing inequality (2). Consequently, $D_\mu(R) \leq \langle \mu \otimes K, c \rangle = V(R, \mathsf{Z})$. Taking the infimum over Z, we get $D_\mu(R) \leq V(R)$.

To prove the other direction, let $\mathsf{Z} = \mathcal{P}(\mathsf{X})$[2] and consider the optimal kernel $K^* \in \mathcal{M}(\mathsf{U}|\mathsf{X})$ that achieves the infimum in (11). Using (12) and the Bayes' rule, we can compute the posterior distribution (belief state)

$$\check{K}^*(\mathrm{d}x|u) = \frac{e^{-\frac{1}{s}c(x,u)}\mu(\mathrm{d}x)}{\int_{\mathsf{X}} e^{-\frac{1}{s}c(x,u)}\mu(\mathrm{d}x)}.$$

Using the minimal sufficiency property of the belief state [18], [19] and the fact that K^* attains the DRF, it can be shown that the kernel $W \in \mathcal{M}(\mathsf{Z}|\mathsf{X})$ given by the composition of K^* and the deterministic mapping $u \mapsto \check{K}^*(\cdot|u)$ is feasible for the problem (8) with $\mathsf{Z} = \mathcal{P}(\mathsf{X})$. Moreover, if we choose the controller $\Phi \in \mathcal{M}(\mathsf{U}|\mathsf{Z})$ in such a way that $\Phi(\mathrm{d}u|z)$ is supported on the set $\{u \in \mathsf{U} : \check{K}^*(\cdot|u) = z\}$, then the resulting cost will not exceed $D_\mu(R)$. [In fact, with the choice $\mathsf{Z} = \mathcal{P}(\mathsf{X})$, we can write $D_\mu(R)$ in the form (10).] Again, taking the infimum over Z, we get $V(R) \leq D_\mu(R)$. □

[2]Note that because X is standard Borel, the space $\mathsf{Z} = \mathcal{P}(\mathsf{X})$ is a complete separable metric space w.r.t. any metric compatible with weak convergence of probability measures, and so Z is standard Borel as well.

V. Average-cost optimal control with rational inattention

We now turn to the analysis of the infinite-horizon control problem (7) with an information constraint. In multi-stage control problems, such as this one, the control law has a *dual effect* [20]: it affects both the cost at the current stage and the uncertainty about the state at future stages. The presence of the mutual information constraint (7b) enhances this dual effect, since it prevents the controller from ever learning "too much" about the state. This, in turn, limits the controller's future ability to keep the average cost low. These considerations suggest that, in order to bring rate-distortion theory to bear on the problem (7), we cannot use the one-stage cost c as the distortion function. Instead, we must modify it to account for the effect of the control action on future costs. As we will see, this modification implies a certain stochastic relaxation of the Average Cost Optimality Equation (ACOE) that characterizes optimal performance achievable by any MRS control law in the absence of information constraints.

A. Reduction to single-stage optimization

To construct this modification in a principled manner, we first reduce the dynamic optimization problem (7) to a suitable static (single-stage) problem. Once this has been carried out, we will be able to take advantage of the results of Section IV. The reduction is based on the so-called *convex-analytic* approach to controlled Markov processes [21]–[25], which we briefly summarize here. Given a Markov control problem with initial state distribution μ and transition kernel Q, the long-run expected average cost of an MRS control law $\Phi \in \mathcal{M}(\mathsf{U}|\mathsf{X})$ is given by

$$J_\mu(\Phi) = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\sum_{t=1}^{T} c(X_t, U_t) \right]. \tag{15}$$

We wish to find an MRS control law Φ^* that would minimize $J_\mu(\Phi)$ simultaneously for all μ. Any MRS control law Φ induces a transition kernel Q_Φ on the state space X:

$$Q_\Phi(A|x) \triangleq \int_{\mathsf{U}} Q(A|x,u)\Phi(\mathrm{d}u|x), \quad \forall A \in \mathcal{B}(\mathsf{X}).$$

We say that Φ is *stable* if:

1) There exists a probability measure $\pi_\Phi \in \mathcal{P}(\mathsf{X})$ which is invariant under Q_Φ, i.e., $\pi_\Phi = \pi_\Phi Q_\Phi$.
2) The average cost $J_{\pi_\Phi}(\Phi)$ is finite, and moreover

$$J_{\pi_\Phi}(\Phi) = \langle \Gamma_\Phi, c \rangle = \int_{\mathsf{X} \times \mathsf{U}} c(x,u)\Gamma_\Phi(\mathrm{d}x, \mathrm{d}u),$$

where $\Gamma_\Phi \triangleq \pi_\Phi \otimes \Phi$.

Let $\mathcal{K} \subset \mathcal{M}(\mathsf{U}|\mathsf{X})$ denote the space of all such stable laws.

Theorem 3. *Suppose that the following assumptions are satisfied:*

- *(**A.1**) The one-stage cost function c is nonnegative, lower semicontinuous, and coercive, i.e., there exist two sequences of compact sets $\mathsf{X}_n \uparrow \mathsf{X}$ and $\mathsf{U}_n \uparrow \mathsf{U}$ such that*
$$\lim_{n\to\infty} \inf_{x\in\mathsf{X}_n^c, u\in\mathsf{U}_n^c} c(x,u) = +\infty.$$
- *(**A.2**) The transition kernel Q is continuous, i.e., $Qf \in C_b(\mathsf{X}\times\mathsf{U})$ for any $f \in C_b(\mathsf{X})$.*

Let $J^* \triangleq \inf_\Phi \inf_\mu J_\mu(\Phi)$. Then there exists a control law $\Phi^* \in \mathcal{K}$, such that

$$J_{\pi_{\Phi^*}}(\Phi^*) = J^* = \inf_{\Phi\in\mathcal{K}} \langle \Gamma_\Phi, c\rangle. \tag{16}$$

B. Bellman error minimization via marginal decomposition

For our purposes, it is convenient to decompose the infimum over Φ in (16) by first fixing the marginal $\pi \in \mathcal{P}(\mathsf{X})$. Consider the set of all stable control laws that leave π invariant,

$$\mathcal{K}_\pi \triangleq \left\{\Phi \in \mathcal{K} : \pi = \pi_\Phi\right\}$$

This set may very well be empty for some π. Regardless, assuming that the conditions of Theorem 3 are satisfied, we can write

$$J^* = \inf_{\Phi\in\mathcal{K}} \langle \Gamma_\Phi, c\rangle = \inf_{\pi\in\mathcal{P}(\mathsf{X})} \inf_{\Phi\in\mathcal{K}_\pi} \langle \pi\otimes\Phi, c\rangle.$$

We are now in a position to introduce the information constraint. Let $\mathcal{K}_\pi(R) \triangleq \mathcal{K}_\pi \cap \mathcal{I}_\pi(R)$. Then we define the optimal steady-state value of the information-constrained average-cost control problem as

$$J^*(R) \triangleq \inf_{\pi\in\mathcal{P}(\mathsf{X})} \inf_{\Phi\in\mathcal{K}_\pi(R)} \langle \pi\otimes\Phi, c\rangle. \tag{17}$$

As a first step to understanding solutions to (17), we consider each candidate invariant distribution $\pi \in \mathcal{P}(\mathsf{X})$ separately:

Proposition 3. *For any $\pi \in \mathcal{P}(\mathsf{X})$, let*

$$J^*_\pi(R) \triangleq \inf_{\Phi\in\mathcal{K}_\pi(R)} \langle \pi\otimes\Phi, c\rangle.$$

Then

$$J^*_\pi(R) = \inf_{\Phi\in\mathcal{I}_\pi(R)} \sup_{h\in C_b(\mathsf{X})} \langle \pi\otimes\Phi, c + Qh - h\rangle \tag{18}$$

Remark 1. *The function $h \in C_b(\mathsf{X})$ plays the role of a Lagrange multiplier associated with the constraint $\Phi \in \mathcal{K}_\pi$, which is what can be expected from the theory of average-cost optimal control [25, Ch. 9].*

On setting $\eta = \langle \pi\otimes\Phi, c\rangle$, the function $c + Qh - h - \eta$ is the Bellman error associated with h that is central to approximate dynamic programming.

Remark 2. *Both in* (18) *and elsewhere, we can extend the supremum over h to all $h \in L^1(\pi)$ without affecting the value of $J^*_\pi(R)$.*

Proof. Define the function

$$\iota_\pi(\Phi) \triangleq \begin{cases} 0, & \Phi \in K_\pi \\ +\infty, & \Phi \notin K_\pi \end{cases}$$

Then we can write

$$J^*_\pi(R) = \inf_{\Phi\in\mathcal{I}_\pi(R)} \left[\left\langle \pi\otimes\Phi, c\right\rangle + \iota_\pi(\Phi)\right]. \tag{19}$$

Moreover,

$$\iota_\pi(\Phi) = \sup_{h\in C_b(\mathsf{X})} \left[\left\langle \pi Q_\Phi, h\right\rangle - \left\langle \pi, h\right\rangle\right] \tag{20}$$

Indeed, if $\Phi \in K_\pi$, then the right-hand side of (20) is zero. On the other hand, suppose that $\Phi \notin K_\pi$. Since X is standard Borel, any two probability measures $\mu, \nu \in \mathcal{P}(\mathsf{X})$ are equal if and only if $\langle\mu, h\rangle = \langle\nu, h\rangle$ for all $h \in C_b(\mathsf{X})$. Consequently, there exists some $h_0 \in C_b(\mathsf{X})$ such that $\langle\pi, h_0\rangle \neq \langle\pi Q_\Phi, h_0\rangle$. There is no loss of generality if we assume that $\langle\pi Q_\Phi, h_0\rangle - \langle\pi, h_0\rangle > 0$. Then by considering functions $h_0^n = nh_0$ for all $n = 1, 2, \dots$ and taking the limit as $n \to \infty$, we can make the right-hand side of (20) grow without bound. Thus, we have proved (20). Substituting it into (19), we get (18). □

To analyze the optimization problem (17), let us fix some π and consider the *dual value*

$$J_{*,\pi}(R) \triangleq \sup_{h\in C_b(\mathsf{X})} \inf_{\Phi\in\mathcal{I}_\pi(R)} \langle \pi\otimes\Phi, c + Qh - h\rangle \tag{21}$$

Clearly, $J^*_\pi(R) \geq J_{*,\pi}(R)$ for all π and R. Moreover:

Proposition 4. *Suppose that assumption* (**A.1**) *above is satisfied, and that $J^*_\pi(R) < \infty$. Then the primal value $J^*_\pi(R)$ and the dual value $J_{*,\pi}(R)$ are equal.*

Proof. Let $\mathcal{P}^0_{\pi,c}(R) \subset \mathcal{P}(\mathsf{X}\times\mathsf{U})$ be the closure, in the weak topology, of the set of all probability measures $\Gamma \in \mathcal{P}(\mathsf{X}\times\mathsf{U})$, such that $\Gamma(A\times\mathsf{U}) = \pi(A)$, $I(\Gamma) \leq R$, and $\langle\Gamma, c\rangle < \infty$. Since $J^*_\pi(R) < \infty$ by hypothesis, we can write

$$J^*_\pi(R) = \inf_{\Gamma\in\mathcal{P}^0_{\pi,c}(R)} \sup_{h\in C_b(\mathsf{X})} \langle \Gamma, c + Qh - h\rangle. \tag{22}$$

Because c is coercive and nonnegative, the set $\{\Gamma \in \mathcal{P}(\mathsf{X}\times\mathsf{U}) : \langle\Gamma, c\rangle < \infty\}$ is tight [26, Proposition 1.4.15], so its closure is weakly sequentially compact by Prohorov's theorem. Moreover, because the function $\Gamma \mapsto I(\Gamma)$ is weakly lower semicontinuous [8], the set $\{\Gamma : I(\Gamma) \leq R\}$ is closed. Therefore, the set $\mathcal{P}^0_{\pi,c}(R)$ is closed and tight, hence weakly sequentially compact. Moreover, the sets $\mathcal{P}^0_{\pi,c}(R)$ and $C_b(\mathsf{X})$ are both convex, and the objective function on the right-hand side of (22) is affine in Γ and linear in h. Therefore, by Sion's minimax theorem [27] we may interchange the supremum and the infimum to conclude that $J^*_\pi(R) = J_{*,\pi}(R)$. □

We are now in a position to relate the optimal value $J^*_\pi(R)$ to a suitable rate-distortion problem.

Theorem 4. *Consider a probability measure $\pi \in \mathcal{P}(\mathsf{X})$ such that $J^*_\pi(R) < \infty$, and the supremum over $h \in C_b(\mathsf{X})$ in* (21) *is attained by some h^*. Then there exists an MRS control law $\Phi^* \in \mathcal{M}(\mathsf{U}|\mathsf{X})$, such that $I(\pi, \Phi^*) = R$,*

$$J_\pi^*(R) = \langle \pi \otimes \Phi^*, c + Qh^* - h^* \rangle,$$

and the Radon–Nikodym derivative of $\pi \otimes \Phi^*$ w.r.t. $\pi \otimes \pi\Phi^*$ takes the form

$$\frac{\mathrm{d}(\pi \otimes \Phi^*)}{\mathrm{d}(\pi \otimes \pi\Phi^*)}(x,u) = \frac{e^{-\frac{1}{s}d(x,u)}}{\int_{\mathsf{U}} e^{-\frac{1}{s}d(x,u)}\pi\Phi^*(\mathrm{d}u)}, \tag{23}$$

where $d(x,u) \triangleq c(x,u) + Qh^*(x,u)$, and $s \ge 0$ satisfies

$$D_\pi(R'; c + Qh^*) + sR' \ge D_\pi(R; c + Qh^*) + sR \tag{24}$$

for all R'.

Proof. Using Proposition 4 and the definition (21) of the dual value $J_{*,\pi}(R)$, we can express $J_\pi^*(R)$ as a pointwise supremum of a family of DRF's:

$$J_\pi^*(R) = \sup_{h \in C_b(\mathsf{X})} \left[D_\pi(R; c + Qh) - \langle \pi, h \rangle \right]. \tag{25}$$

Since $J_\pi^*(R) < \infty$, we can apply Proposition 1 separately for each $h \in C_b(\mathsf{X})$. In particular, taking any $h^* \in C_b(\mathsf{X})$ that achieves the supremum in (25) and using (12) with

$$\alpha(x) = \frac{1}{\int_{\mathsf{U}} e^{-\frac{1}{s}d(x,u)}\pi\Phi^*(\mathrm{d}u)},$$

we get (23). In the same way, (24) follows from (14). □

Theorem 5. *The optimal value $J_\pi^*(R)$ admits the following variational representation:*

$$J_\pi^*(R) = \sup_{s \ge 0} \sup_{h \in C_b(\mathsf{X})} \inf_{\nu \in \mathcal{P}(\mathsf{U})} \Bigg\{ - \langle \pi, h \rangle + s \left[\left\langle \pi, \log \frac{1}{\int_{\mathsf{U}} e^{-\frac{1}{s}[c(x,u)+Qh(x,u)]}\nu(\mathrm{d}u)} \right\rangle - R \right] \Bigg\}$$

Proof. Immediate from (25) and Proposition 2. □

To complete the computation of the optimal steady-state value $J^*(R)$ defined in (17), we need to consider all candidate invariant distributions $\pi \in \mathcal{P}(\mathsf{X})$ for which $\mathcal{K}_\pi(R)$ is nonempty, and then choose among them any π that attains the smallest value of $J_\pi^*(R)$ (assuming this value is finite). The corresponding control law will then be characterized by Theorem 4. Of course, even if we can guarantee the existence of such an optimal π under suitable assumptions on the cost c and the transition law Q, it will be exceedingly difficult to solve for it explicitly. On the other hand, if $J_\pi^*(R) < \infty$ for *some* π, then Theorem 4 ensures that there exists a *suboptimal* control law satisfying the information constraint.

C. Some implications

We close this section by examining some implications of the results proved above. Using Theorem 5, we see that $J_\pi^*(R)$ is equal to the value of the following optimization problem:

$$\begin{aligned} &\text{maximize } \lambda \\ &\quad \text{s.t. } s\left\langle \pi, \log\frac{1}{\int_{\mathsf{U}} e^{-\frac{1}{s}[c(x,u)+Qh(x,u)]}\nu(\mathrm{d}u)} - \frac{h}{s} \right\rangle \\ &\qquad\qquad \ge \lambda + sR, \quad \forall \nu \in \mathcal{P}(\mathsf{U}) \\ &\qquad \lambda \ge 0,\ s \ge 0,\ h \in C_b(\mathsf{X}) \end{aligned}$$

Let us examine what happens as we relax the information constraint, i.e., let $R \to \infty$. From the fact that the DRF is convex and nondecreasing in R, and from the characterization (24) of $-s$ as the slope of a tangent to the graph of the DRF at R, this is equivalent to letting s approach 0 (with the convention that $sR \to 0$ even as $R \to \infty$). Let us recall *Laplace's principle* [28]: for any $\nu \in \mathcal{P}(\mathsf{U})$ and any measurable function $F : \mathsf{U} \to \mathbb{R}$ such that $e^{-F} \in L^1(\nu)$, we have

$$-\lim_{s \downarrow 0} s \log \int_{\mathsf{U}} e^{-\frac{1}{s}F(u)}\nu(\mathrm{d}u) = \operatorname*{ess\,inf}_{u \in \mathsf{U}} F(u),$$

where the essential infimum is defined w.r.t. ν. In view of this, the limit of $J_\pi^*(R)$ as $R \to \infty$ is the value of

$$\begin{aligned} &\text{maximize } \lambda \\ &\quad \text{s.t. } \left\langle \pi, \inf_{u \in \mathsf{U}} [c(x,u) + Qh(x,u)] - h \right\rangle \ge \lambda \\ &\qquad \lambda \ge 0,\ h \in C_b(\mathsf{X}) \end{aligned}$$

Performing now the minimization over $\pi \in \mathcal{P}(\mathsf{X})$ as well, we see that the limit of $J^*(R)$ as $R \to \infty$ is given by the value of the following problem:

$$\begin{aligned} &\text{maximize } \lambda \\ &\quad \text{s.t. } \inf_{u \in \mathsf{U}} [c(x,u) + Qh(x,u)] - h \ge \lambda \\ &\qquad \lambda \ge 0,\ h \in C_b(\mathsf{X}) \end{aligned}$$

In particular, under Assumptions (**A.1**) and (**A.2**) of Theorem 3, there exist a function h^∞ and a number $\lambda^\infty \ge 0$ that solve the Average Cost Optimality Equation (ACOE)

$$h^\infty(x) + \lambda^\infty = \inf_{u \in \mathsf{U}} [c(x,u) + Qh^\infty(x,u)],$$

and $J^*(R) \to \lambda^\infty$ as $R \to \infty$. When a nontrivial information constraint is present, i.e., when $R \in (0, \infty)$, we may consider instead a *stochastic relaxation* of the ACOE, given by

$$\begin{aligned} &\langle \pi, h \rangle + \lambda + sR \\ &= s \inf_{\nu \in \mathcal{P}(\mathsf{U})} \left\langle \pi, \log \frac{1}{\int_{\mathsf{U}} e^{-\frac{1}{s}[c(x,u)+Qh(x,u)]}\nu(\mathrm{d}u)} \right\rangle, \end{aligned}$$

which must be solved for h, s, and λ. If a solution triple (h, s, λ) exists, and if $\nu^* \in \mathcal{P}(\mathsf{U})$ attains the infimum on the right-hand side, then a control law $\Phi \in \mathcal{M}(\mathsf{U}|\mathsf{X})$ can be defined through

$$\Phi(\mathrm{d}u|x) = \frac{e^{-\frac{1}{s}[c(x,u)+Qh(x,u)]}\nu^*(\mathrm{d}u)}{\int_{\mathsf{U}} e^{-\frac{1}{s}[c(x,u)+Qh(x,u)]}\nu^*(\mathrm{d}u)}.$$

Moreover, if $\pi = \pi_\Phi$, then by Theorems 4 and 5, $J_\pi^*(R) = \lambda$, and (h, Φ) are an optimizing pair for $J_\pi^*(R)$.

VI. CONCLUSIONS

The main contributions of this paper are to pose a general *rational inattention model* for optimal control with information constraints, and to reveal structure for the associated optimal control equations. Future work will include the construction of reinforcement learning algorithms that make use of this structure to obtain approximately optimal policies based on online measurements of the system to be controlled. This is particularly natural in applications to finance or economics.

APPENDIX I
PROOF OF THE PRINCIPLE OF IRRELEVANT INFORMATION

Let $M(\mathrm{d}\upsilon|\psi)$ be the conditional distribution of Υ given Ψ, let $\Lambda(\mathrm{d}\psi|\theta,\xi)$ be the conditional distribution of Ψ given (θ,ξ), and disintegrate the joint distribution of Θ,Ξ,Ψ,Υ as

$$P(\mathrm{d}\theta,\mathrm{d}\xi,\mathrm{d}\psi,\mathrm{d}\upsilon) = P(\mathrm{d}\theta)P(\mathrm{d}\xi|\theta)\Lambda(\mathrm{d}\psi|\theta,\xi)M(\mathrm{d}\upsilon|\psi).$$

If we define $\Lambda'(\mathrm{d}\psi|\theta)$ by

$$\Lambda'(\cdot|\theta) = \int W(\cdot|\theta,\xi)P(\mathrm{d}\xi|\theta)$$

and let the tuple $(\Theta',\Xi',\Psi',\Upsilon')$ have the joint distribution

$$P'(\mathrm{d}\theta,\mathrm{d}\xi,\mathrm{d}\psi,\mathrm{d}\upsilon) = P(\mathrm{d}\theta)P(\mathrm{d}\xi|\theta)\Lambda'(\mathrm{d}\psi|\theta)M(\mathrm{d}\upsilon|\psi),$$

then it is easy to see that it has all of the desired properties.

APPENDIX II
PROOF OF THE TWO-STAGE LEMMA

Note that Z_1 only depends on X_1, and that only the second-stage expected cost is affected by the choice of W_2. We can therefore apply the Principle of Irrelevant Information to $\Theta = X_2$, $\Xi = (X_1, Z_1, U_1)$, $\Psi = Z_2$ and $\Upsilon = U_2$. Because both the expected cost $\mathbb{E}[c(X_t,U_t)]$ and the mutual information $I(X_t;Z_t)$ depend only on the corresponding bivariate marginals, the lemma is proved.

APPENDIX III
PROOF OF THE THREE-STAGE LEMMA

Again, Z_1 only depends on X_1, and only the second- and the third-stage expected costs are affected by the choice of W_2. By the law of iterated expectation, we have

$$\mathbb{E}[c(X_3,U_3)] = \mathbb{E}[\mathbb{E}[c(X_3,U_3)|X_2,U_2]] = \mathbb{E}[h(X_2,U_2)],$$

where $h(X_2,U_2) \triangleq \mathbb{E}[c(X_3,U_3)|X_2,U_2]$. Note that the functional form of h does not depend on the choice of W_2, since for any fixed realizations $X_2 = x_2$ and $U_2 = u_2$ we have, by hypothesis,

$$\begin{aligned} h(x_2,u_2) &= \int c(x_3,u_3)P(\mathrm{d}x_3,\mathrm{d}u_3|x_2,u_2) \\ &= \int c(x_3,u_3)Q(\mathrm{d}x_3|x_2,u_2)W_3(\mathrm{d}z_3|x_3)\Phi_3(\mathrm{d}u_3|\,\mathrm{d}z_3). \end{aligned}$$

Therefore, applying the Principle of Irrelevant Information to $\Theta = X_2$, $\Xi = (X_1,Z_1,U_1)$, $\Psi = Z_2$, and $\Upsilon = U_2$,

$$\begin{aligned} \mathbb{E}[c(X_2',U_2') + c(X_3',U_3')] &= \mathbb{E}[c(X_2',U_2') + h(X_2',U_2')] \\ &= \mathbb{E}[c(X_2,U_2) + h(X_2,U_2)] \\ &= \mathbb{E}[c(X_2,U_2) + c(X_3,U_3)], \end{aligned}$$

where the variables (X_t', Z_t', U_t') are obtained from the original ones by replacing $W_2(\mathrm{d}z_2|x^2,z_1,u_1)$ by $W_2'(\mathrm{d}z_2|x_2)$.

REFERENCES

[1] P. Varaiya and J. Walrand, "Causal coding and control for Markov chains," *Systems and Control Letters*, vol. 3, pp. 189–192, 1983.
[2] R. Bansal and T. Başar, "Simultaneous design of measurement and control strategies for stochastic systems with feedback," *Automatica*, vol. 25, no. 5, pp. 679–694, 1989.
[3] S. Tatikonda and S. Mitter, "Control over noisy channels," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1196–2001, July 2004.
[4] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1549–1561, September 2004.
[5] S. Yüksel and T. Linder, "Optimization and convergence of observation channels in stochastic control," *SIAM Journal on Control and Optimization*, vol. 50, no. 2, pp. 864–887, 2012.
[6] C. Sims, "Implications of rational inattention," *Journal of Monetary Economics*, vol. 50, no. 3, pp. 665–690, 2003.
[7] ——, "Rational inattention: Beyond the linear-quadratic case," *The American Economic Review*, vol. 96, no. 2, pp. 158–163, 2006.
[8] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. Holden-Day, 1964.
[9] F. Matejka and A. McKay, "Rational Inattention to discrete choices: A new foundation for the multinomial logit model," 2011, working paper.
[10] ——, "Simple market equilibria with rationally inattentive consumers," *The American Economic Review*, vol. 102, no. 3, pp. 24–29, 2012.
[11] L. Peng, "Learning with information capacity constraints," *Journal of Financial and Quantitative Analysis*, vol. 40, no. 2, pp. 307–329, 2005.
[12] H. S. Witsenhausen, "On the structure of real-time source coders," *Bell System Technical Journal*, vol. 58, no. 6, pp. 1437–1451, July 1979.
[13] S. P. Meyn and G. Mathew, "Shannon meets Bellman: Feature based Markovian models for detection and optimization," in *Proc. 47th IEEE CDC*, 2008, pp. 5558–5564.
[14] T. Berger, *Rate Distortion Theory, A Mathematical Basis for Data Compression*. Prentice Hall, 1971.
[15] R. L. Stratonovich, "On value of information," *Izvestiya of USSR Academy of Sciences, Technical Cybernetics*, vol. 5, pp. 3–12, 1965.
[16] ——, *Information Theory*. Moscow, USSR: Sovetskoe Radio, 1975.
[17] I. Csiszár, "On an extremum problem of information theory," *Studia Scientiarum Mathematicarum Hungarica*, vol. 9, pp. 57–71, 1974.
[18] P. R. Halmos and L. J. Savage, "Application of the Radon–Nikodym theorem to the theory of sufficient statistics," *Annals of Mathematical Statistics*, vol. 20, pp. 225–241, 1949.
[19] G. Picci, "Some connections between the theory of sufficient statistics and the identifiability problem," *SIAM Journal on Applied Mathematics*, vol. 33, no. 3, pp. 383–398, 1977.
[20] Y. Bar-Shalom and E. Tse, "Dual effect, certainty equivalence, and separation in stochastic control," *IEEE Transactions on Automatic Control*, vol. 19, pp. 494–500, 1974.
[21] A. Manne, "Linear programming and sequential decisions," *Management Science*, vol. 6, pp. 257–267, 1960.
[22] V. S. Borkar, "A convex analytic approach to Markov decision processes," *Probabillity Theory and Related Fields*, vol. 78, no. 583-602, 1988.
[23] O. Hernández-Lerma and J. B. Lasserre, "Linear programming and average optimality of Markov control processes on Borel spaces — unbounded costs," *SIAM Journal on Control and Optimization*, vol. 32, no. 2, pp. 480–500, March 1994.
[24] V. S. Borkar, "Convex analytic methods in Markov decision processes," in *Handbook of Markov Decision Processes*, E. Feinberg and A. Shwartz, Eds. Boston, MA: Kluwer, 2001.
[25] S. P. Meyn, *Control Techniques for Complex Networks*. Cambridge Univ. Press, 2008.
[26] O. Hernández-Lerma and J. B. Lasserre, *Markov Chains and Invariant Probabilities*. Birkhäuser, 2003.
[27] M. Sion, "On general minimax theorems," *Pacific Journal of Mathematics*, vol. 8, pp. 171–176, 1958.
[28] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. New York: Wiley, 1997.