On Optimization and Convergence of Observation Channels and Quantizers in Stochastic Control

Serdar Yüksel and Tamás Linder

Abstract—This paper studies the optimization of observation channels and quantizers in partially observed stochastic control problems. Continuity properties of the optimal cost in channels and quantizers are explored. Sufficient conditions for sequential compactness under total variation and setwise convergence are presented. Conditions for existence of optimal channels and quantizers are established. It is shown that quantizers are extreme points of the space of channels in an appropriate sense.

I. Introduction

In stochastic control, one is often concerned with the following problem: Given a dynamical system, an observation channel (stochastic kernel), a cost function, and an action set, when does there exist an optimal policy, and what is an optimal control policy? The theory for such problems is advanced, and practically significant, spanning a wide variety of applications in engineering, economics, and natural sciences.

In this paper, we are interested in a dual problem with the following questions to be explored: Given a dynamical system, a cost function, an action set, and a set of observation channels, how can one optimize the observation channels? What is an optimal observation channel subject to constraints on the channel? What is the right convergence notion for continuity in such observation channels for optimization purposes?

We start with the probabilistic setup of the problem. Let $\mathbb{X} \subset \mathbb{R}^n$, be a Borel set in which elements of a controlled Markov process $\{X_t, t \in \mathbb{Z}_+\}$ live. Here and throughout the paper $\mathbb{Z}_+$ denotes the set of nonnegative integers and $\mathbb{N}$ denotes the set of positive integers. Let $\mathbb{Y} \subset \mathbb{R}^m$ be a Borel set, and let an observation channel Q be defined as a stochastic kernel (regular conditional probability) from $\mathbb{X}$ to $\mathbb{Y}$, such that $Q(\,\cdot\,|x)$ is a probability measure on the (Borel) σ-algebra $\mathcal{B}(\mathbb{Y})$ on $\mathbb{Y}$ for every $x \in \mathbb{X}$, and $Q(A|\,\cdot\,) : \mathbb{X} \to [0,1]$ is a Borel measurable function for every $A \in \mathcal{B}(\mathbb{Y})$. Let a decision maker (DM) be located at the output an observation channel Q, with inputs X_t and outputs Y_t. Let $\mathbb{U}$ be a Borel subset of some Euclidean space. An *admissible policy* Π is a sequence of control functions $\{\gamma_t, t \in \mathbb{Z}_+\}$ such that γ_t is measurable with respect to the σ-algebra generated by the information variables

$$I_t = \{Y_{[0,t]}, U_{[0,t-1]}\}, \quad t \in \mathbb{N}, \qquad I_0 = \{Y_0\}.$$

where $U_t = \gamma_t(I_t)$, $t \in \mathbb{Z}_+$, are the $\mathbb{U}$-valued control actions

and we used the notation

$$Y_{[0,t]} = \{Y_s, 0 \le s \le t-1\}, \quad U_{[0,t-1]} = \{U_s, 0 \le s \le t-1\}.$$

The joint distribution of the state, control, and observation processes is determined by the above and the following relationships:

$$\Pr\big((X_0, Y_0) \in B\big) = \int_B P(dx_0) Q(dy_0|x_0), \quad B \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}),$$

where P is the (prior) distribution of the initial state X_0, and

$$\Pr\bigg((X_t, Y_t) \in B \,\bigg|\, x_{[0,t-1]}, y_{[0,t-1]}, u_{[0,t-1]}\bigg)$$
$$= \int_B P(dx_t|x_{t-1}, u_{t-1}) Q(dy_t|x_t), \quad B \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}),$$

where $P(\cdot|x, u)$ is a stochastic kernel from $\mathbb{X} \times \mathbb{U}$ to $\mathbb{X}$.

One way of presenting the problem in a familiar setting is the following: Consider a dynamical system described by the discrete-time equations

$$X_{t+1} = f(X_t, U_t, W_t), \qquad Y_t = g(X_t, V_t)$$

for some measurable functions f, g, with $\{W_t\}$ being independent and identically distributed (i.i.d) system noise process and $\{V_t\}$ an i.i.d. disturbance process, which are independent of X_0 and each other. Here, the second equation represents the communication channel Q, as it describes the relation between the state and observation variables.

With the above setup, let the objective of the decision maker be the minimization of the cost

$$J(P, Q, \Pi) = E_P^{Q,\Pi}\bigg[\sum_{t=0}^{T-1} c(X_t, U_t) \bigg], \tag{1}$$

over the set of all admissible policies Π, where $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is a Borel measurable cost function and $E_P^{Q,\Pi}$ denotes the expectation with initial state probability measure given by P under policy Π and given channel Q. We adapt the convention that random variables are denoted by capital letters and lowercase letters denote their realizations. Also, given a probability measure μ the notation $Z \sim \mu$ means that Z is a random variable with distribution μ.

In the paper, we will regard quantizers as a particular class of channels. With this interpretation, we will be interested in the following problems:

PROBLEM P1. CONTINUITY ON THE SPACE OF CHANNELS (STOCHASTIC KERNELS) Suppose $\{Q_n, n \in \mathbb{N}\}$ is a sequence

of communication channels converging in some sense to a channel Q. When does

$$Q_n \to Q$$

imply

$$\inf_{\Pi \in \mathscr{P}} J(P, Q_n, \Pi) \to \inf_{\Pi \in \mathscr{P}} J(P, Q, \Pi)?$$

PROBLEM P2. CONCAVITY ON THE SPACE OF CHANNELS Is the optimization problem

$$J(P, Q) := \inf_{\Pi} E_P^{Q,\Pi} \left[\sum_{t=0}^{T-1} c(X_t, U_t) \right]$$

convex/concave on the space of channels?

PROBLEM P3: EXISTENCE OF OPTIMAL CHANNELS Let $\mathcal{Q}$ be a set of communication channels. When do there exist minimizing and maximizing channels for the problems

$$\inf_{Q \in \mathcal{Q}} \inf_{\Pi \in \mathscr{P}} E_P^{Q,\Pi} \left[\sum_{t=0}^{T-1} c(X_t, U_t) \right]$$

and

$$\sup_{Q \in \mathcal{Q}} \inf_{\Pi \in \mathscr{P}} E_P^{Q,\Pi} \left[\sum_{t=0}^{T-1} c(X_t, U_t) \right].$$

The answers may help solve problems in application areas such as:

- For a partially observed stochastic control problem, sometimes we have control over the observation channels by encoding/quantization. When does there exist an optimal quantizer for such a setup? (Optimal quantization)
- Given an uncertainty set for the observation channels, and tools available for learning the channel, can one identify a worst element/best element? (Robust control)
- When we do not know the channel, but have statistical tools and empirical measurements to learn, typically under mild technical conditions, the empirical distributions converge to the actual distribution, in some sense. Does this imply that we could design the optimal control policies based on empirical estimates, and does the optimal cost converge to the correct limit as the number of measurements grows? (Consistency of empirical controllers)

In the following, we will address problems P1-P3. We will consider the case with $T = 1$. The multi-stage case is addressed in [26] and [27].

A. Relevant literature

The problems stated are related to three main areas of research: Robust control, optimal quantizer design and design of experiments.

References [6], [7], [8] have considered both optimal control and estimation and the related problem of optimal control design when the channel is unknown. Similarly, there are connections with robust detection, when the source distribution to be detected belongs to some set. Recently, [25] considered continuity and other functional properties of minimum mean square estimation problems under Gaussian channels.

We will observe that the optimal cost is concave in the space of observation channels. This is related to statistics in the context of comparison of experiments as discussed by Blackwell [5].

A related area is on the theory of optimal quantization: References [1], [13] are related as these papers study the effects of uncertainties in the input distribution and consider robustness in the quantizer design. References [17] and [19] study the consistency of optimal quantizers based on empirical data for an unknown source. In the context of decentralized detection, [21] studied certain topological properties and the existence of optimal quantizers.

The rest of the paper is organized as follows. In the next section, we introduce three relevant topologies on the space of communication channels. The continuity problem is considered in Section III. In Section IV, concavity properties of the optimization problems are discussed. We study the problem of existence of optimal channels in Section V, followed by applications in quantization in Section VI. Concluding remarks and discussions in Section VIII end the paper.

II. SOME TOPOLOGIES ON THE SPACE OF COMMUNICATION CHANNELS

One question that we wish address is the choice of an appropriate notion of convergence for a sequence of observation channels. Toward this end, we first review three notions of convergence for probability measures.

Let $\mathcal{P}(\mathbb{R}^N)$ denote the family of all probability measure on $(\mathbb{X}, \mathcal{B}(\mathbb{R}^N))$ for some $N \in \mathbb{N}$. Let $\{\mu_n, n \in \mathbb{N}\}$ be a sequence in $\mathcal{P}(\mathbb{R}^N)$. Recall that $\{\mu_n\}$ is said to converge to $\mu \in \mathcal{P}(\mathbb{R}^N)$ *weakly* if $\int_{\mathbb{R}^N} c(x)\mu_n(dx) \to \int_{\mathbb{R}^N} c(x)\mu(dx)$ for every continuous and bounded $c : \mathbb{R}^N \to \mathbb{R}$. On the other hand, $\{\mu_n\}$ is said to converge to $\mu \in \mathcal{P}(\mathbb{R}^N)$ *setwise* if $\int_{\mathbb{R}^N} c(x)\mu_n(dx) \to \int_{\mathbb{R}^N} c(x)\mu(dx)$ for every measurable and bounded $c : \mathbb{R}^N \to \mathbb{R}$. Setwise convergence can also be defined through pointwise convergence on Borel subsets of $\mathbb{R}^N$ (see, e.g., [16]), that is $\mu_n(A) \to \mu(A)$, for all $A \in \mathcal{B}(\mathbb{R}^N)$. since the space of simple functions are dense in the space of bounded and measurable functions under the supremum norm.

For two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$, the *total variation* metric is given by

$$\|\mu - \nu\|_{TV} := 2 \sup_{B \in \mathcal{B}(\mathbb{R}^N)} |\mu(B) - \nu(B)|$$

$$= \sup_{f : \|f\|_\infty \leq 1} \left| \int f(x)\mu(dx) - \int f(x)\nu(dx) \right|, \quad (2)$$

where the infimum is over all measurable real f such that $\|f\|_\infty = \sup_{x \in \mathbb{R}^N} |f(x)| \leq 1$. A sequence $\{\mu_n\}$ is said to converge to $\mu \in \mathcal{P}(\mathbb{R}^N)$ in total variation if $\|\mu_n - \mu\|_{TV} \to 0$.

Setwise convergence is equivalent to pointwise convergence on Borel sets whereas total variation requires uniform convergence on Borel sets. Thus these three convergence notions are in increasing order of strength: convergence in total variation implies setwise convergence, which in turn implies weak convergence.

A. Convergence of information (observation) channels

Here $\mathbb{X} = \mathbb{R}^n$ and $\mathbb{Y} = \mathbb{R}^m$, and $\mathcal{Q}$ denotes the set of all observation channels (stochastic kernels) with input space $\mathbb{X}$ and output space $\mathbb{Y}$. For $P \in \mathcal{P}(\mathbb{X})$ and $Q \in \mathcal{Q}$ we let PQ denote the joint distribution induced on $(\mathbb{X} \times \mathbb{Y}, \mathcal{B}(\mathbb{X} \times \mathbb{Y}))$ by channel Q with input distribution P:

$$PQ(A) = \int_A Q(dy|x)P(dx), \quad A \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}).$$

Definition 2.1 (Convergence of Channels):

 (i) A sequence of channels $\{Q_n\}$ converges to a channel Q *weakly at input P* if $PQ_n \to PQ$ weakly.
 (ii) A sequence of channels $\{Q_n\}$ converges to a channel Q *setwise at input P* if $PQ_n \to PQ$ setwise, i.e., if $PQ_n(A) \to PQ(A)$ for all Borel sets $A \subset \mathbb{X} \times \mathbb{Y}$.
 (iii) A sequence of channels $\{Q_n\}$ converges to a channel Q in *total variation at input P* if $PQ_n \to PQ$ in total variation, i.e., if $\|PQ_n - PQ\|_{TV} \to 0$.

If we introduce the equivalence relation $Q \equiv Q'$ if and only if $PQ = PQ'$, $Q, Q' \in \mathcal{Q}$, then the convergence notions in Definition 2.1 only induce the corresponding topologies (resp. metrics) on the resulting equivalence classes in $\mathcal{Q}$, instead of $\mathcal{Q}$. Since in most of the development the input distribution P is fixed, there should be no confusion when (somewhat incorrectly) we talk about the induced topologies (resp. metrics) on $\mathcal{Q}$.

B. Classes of assumptions

Throughout the paper the following classes of assumptions will be adopted for the cost function c and the (Borel) set $\mathbb{U} \subset \mathbb{R}^k$ in different contexts:

ASSUMPTIONS.

A1. The function $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is non-negative, bounded, and continuous on $\mathbb{X} \times \mathbb{U}$.

A2. The function function $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is non-negative, measurable, and bounded.

A3. The function $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is non-negative, measurable, bounded, and continuous on $\mathbb{U}$ for every $x \in \mathbb{X}$.

A4. $\mathbb{U}$ is a compact set.

A5. $\mathbb{U}$ is a convex set.

III. PROBLEM P1: CONTINUITY OF THE OPTIMAL COST IN CHANNELS

In this section, we consider continuity properties under total variation, setwise convergence and weak convergence. We consider the single-stage case, and thus investigate the continuity of the functional

$$
\begin{aligned}
J(P,Q) &= \inf_{\Pi} E_P^{Q,\Pi}\big[c(X_0, U_0)\big] \\
&= \inf_{\gamma \in \mathcal{G}} \int_{\mathbb{X} \times \mathbb{Y}} c(x, \gamma(y))Q(dy|x)P(dx)
\end{aligned}
$$

in the channel Q, where $\mathcal{G}$ is the collection of all Borel measurable functions mapping $\mathbb{Y}$ into $\mathbb{U}$. Note that by our previous notation, $\Pi = \gamma$ is an admissible first-stage control policy. As

before, in this section $\mathcal{Q}$ denotes the set of all channels with input space $\mathbb{X}$ and output space $\mathbb{Y}$.

Let us first look for conditions under which an optimal control policy exists; i.e, when the infimum in $\inf_\gamma E_P^{Q,\gamma}[c(X,U)]$ is a minimum.

Theorem 3.1: Suppose assumptions A3 and A4 hold. Then, there exists an optimal control policy for any channel Q.

A. Weak convergence

1) Absence of continuity under weak convergence: We state the following. The proofs are available in [27].

Theorem 3.2: $J(P,Q)$ is not continuous on $\mathcal{Q}$ under weak convergence.

2) Upper semi-continuity under weak convergence:

Theorem 3.3: Suppose assumptions A1 and A5 hold. If $\{Q_n\}$ is a sequence of channels converging weakly at input P to a channel Q, then

$$\limsup_{n \to \infty} J(P, Q_n) \le J(P, Q),$$

that is, $J(P,Q)$ is upper semi-continuous on $\mathcal{Q}$ under weak convergence.

B. Continuity properties under setwise convergence

1) Upper semi-continuity under setwise convergence:

Theorem 3.4: Under assumption A2 the optimal cost

$$J(P,Q) := \inf_\gamma E_P^{Q,\gamma}[c(X,U)]$$

is sequentially upper semi-continuous on the set of communication channels $\mathcal{Q}$ under setwise convergence.

2) Absence of continuity under setwise convergence:

Theorem 3.5: The optimal cost

$$J(P,Q) := \inf_\gamma E_P^{Q,\gamma}[c(X,U)]$$

is not sequentially continuous on the set of communication channels $\mathcal{Q}$ under setwise convergence, even for continuous cost functions and compact $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{U}$.

C. Continuity under total variation

Theorem 3.6: Under assumption A2 the optimal cost $J(P,Q)$ is continuous on the set of communication channels $\mathcal{Q}$ under under the topology of total variation.

IV. PROBLEM P2: CONCAVITY ON THE SPACE OF CHANNELS

The following is a result with important consequences in decentralized stochastic control problems as is elaborated on in [26] and [27].

Theorem 4.1: The expression

$$J(P,Q) = \inf_\gamma E_P^{Q,\gamma}[c(X,U)]$$

is a concave function of Q.

Proof of Theorem 4.1. For $\alpha \in [0,1]$ and $Q', Q'' \in \mathcal{Q}$, let $Q = \alpha Q' + (1-\alpha)Q'' \in \mathcal{Q}$, i.e.,

$$Q(A|x) = \alpha Q'(A|x) + (1-\alpha)Q''(A|x)$$

for all $A \in \mathcal{B}(\mathbb{Y})$ and $x \in \mathbb{X}$. Noting that $PQ = \alpha PQ' + (1 - \alpha)PQ''$, we have

$$
\begin{aligned}
& J(P,Q) = J(P, \alpha Q' + (1 - \alpha)Q'') \\
& = \inf_{\gamma} E_P^{Q,\gamma}[c(X,U)] = \inf_{\gamma \in \mathcal{G}} \int c(x, \gamma(y)) PQ^*(dx, dy) \\
& = \inf_{\gamma \in \mathcal{G}} \left(\alpha \int c(x, \gamma(y)) PQ'(dx, dy) \right. \\
& \qquad \left. + (1 - \alpha) \int c(x, \gamma(y)) PQ''(dx, dy) \right) \\
& \geq \inf_{\gamma \in \mathcal{G}} \left(\alpha \int c(x, \gamma(y)) PQ'(dx, dy) \right) \\
& \qquad + \inf_{\gamma \in \mathcal{G}} \left((1 - \alpha) \int c(x, \gamma(y)) PQ''(dx, dy) \right) \\
& = \alpha J(P, Q') + (1 - \alpha) J(P, Q'') \qquad\qquad (3)
\end{aligned}
$$

proving that $J(P, Q)$ is concave in Q. ∎

V. PROBLEM P3: EXISTENCE OF OPTIMAL CHANNELS

Here we study characterizations of compactness (or sequential compactness) which will be useful in obtaining existence results.

The discussion on weak convergence showed us that weak convergence does not induce a strong enough topology, i.e., under which useful continuity properties can be obtained. In the following, we will obtain conditions for sequential compactness for the other two convergence notions, that is, for setwise convergence and total variation.

We first discuss setwise convergence. A set of probability measures $\mathcal{M}$ on some measurable space is said to be *setwise sequentially precompact* if every sequence in $\mathcal{M}$ has a subsequence converging setwise to a probability measure (not necessarily in $\mathcal{M}$). For two finite measures ν and μ defined on the same measurable space we write $\nu \leq \mu$ if $\nu(A) \leq \mu(A)$ for all measurable A.

The following sufficient condition follows from Dunford and Schwartz ([11], p. 305-306) (see also Hernandez-Lerma and Lasserre ([16], p. 7)):

Lemma 5.1 (Dunford-Schwartz): Let μ be a finite Borel measure on a locally compact metric space $\mathbb{T}$. Assume a set of Borel probability measures $\Psi \subset \mathcal{P}(\mathbb{T})$ satisfies

$$ P \leq \mu, \quad \text{for all } P \in \Psi. $$

Then Ψ is setwise sequentially precompact.

As before, $PQ \in \mathcal{P}(\mathbb{X} \times \mathbb{Y})$ denotes the joint probability measure induced by input P and channel Q, where $\mathbb{X} = \mathbb{R}^n$ and $\mathbb{Y} = \mathbb{R}^m$. A simple consequence of the preceding majorization criterion is the following.

Lemma 5.2: Let ν be a finite measure on $\mathcal{B}(\mathbb{X} \times \mathbb{Y})$ and let P be a probability measure on $\mathcal{B}(\mathbb{X})$. Suppose $\mathcal{Q}$ is a set of channels such that

$$ PQ \leq \nu, \quad \text{for all } Q \in \mathcal{Q}. $$

Then $\mathcal{Q}$ is setwise sequentially precompact at input P in the sense that any sequence in $\mathcal{Q}$ has a subsequence $\{Q_n\}$ such that $Q_n \to Q$ setwise at input P for some channel Q.

For a probability density function p on $\mathbb{R}^N$ we let P_p denote the induced probability measure: $P_p(A) = \int_A p(x) \, dx$, $A \in \mathcal{B}(\mathbb{R}^N)$. The next lemma gives a sufficient condition for precompactness under total variation.

Lemma 5.3: Let μ be a finite Borel measure on $\mathbb{R}^N$ and let $\mathcal{F}$ be an equicontinuous and uniformly bounded family of probability density functions. Define $\Psi \subset \mathcal{P}(\mathbb{R}^N)$ by

$$ \Psi = \{P_p : P_p \leq \mu, p \in \mathcal{F}\}. $$

Then Ψ is precompact under total variation.

Lemma 5.4: Let $\mathcal{Q}$ be a set of channels such that $\{PQ : Q \in \mathcal{Q}\}$ is a precompact set of probability measures under total variation. Then $\mathcal{Q}$ is precompact under total variation at input P.

The following theorem, when combined with the preceding results, gives sufficient conditions for the existence of best and worst channels when the given family of channels $\mathcal{Q}$ is closed under the appropriate convergence notion.

Theorem 5.1: Recall problem P2.

(i) There exist a worst channel in $\mathcal{Q}$, that is, a solution for the maximization problem

$$ \sup_{Q \in \mathcal{Q}} J(P, Q) = \sup_{Q \in \mathcal{Q}} \inf_{\gamma} E_P^{Q,\gamma} E[c(X, U)] $$

when the set $\mathcal{Q}$ is weakly sequentially compact and assumptions A1, A4, and A5 hold.

(ii) There exist a worst channel in $\mathcal{Q}$ when the set $\mathcal{Q}$ is setwise sequentially compact and assumption A2 holds.

(iii) There exist best and worst channels in $\mathcal{Q}$, that is, solutions for the minimization problem $\inf_{Q \in \mathcal{Q}} J(P, Q)$ and the maximization problem $\sup_{Q \in \mathcal{Q}} J(P, Q)$ when the set $\mathcal{Q}$ is compact under total variation and assumption A2 holds.

VI. QUANTIZERS AS A CLASS OF CHANNELS

Here we consider the problem of convergence and optimization of quantizers. We start with the definition of a quantizer.

Definition 6.1: An M-cell *vector quantizer*, q, is a (Borel) measurable mapping from $\mathbb{X} = \mathbb{R}^n$ to the finite set $\{1, 2, \ldots, M\}$, characterized by a measurable partition $\{B_1, B_2, \ldots, B_M\}$ such that $B_i = \{x : q(x) = i\}$ for $i = 1, \ldots, M$. The B_i are called the cells (or bins) of q.

REMARKS.

- For later convenience we allow for the possibility that some of the cells of the quantizer are empty.
- Traditionally, in source coding theory, a quantizer is a mapping $q : \mathbb{R}^n \to \mathbb{R}$ with a finite range. Thus q is defined by a partition and a reconstruction value in $\mathbb{R}^n$ for each cell in the partition. That is, for given cells $\{B_1, \ldots, B_M\}$ and reconstruction values $\{c_1, \ldots, c_M\} \subset \mathbb{R}^n$, we have $q(x) = c_i$ if and only if $x \in B_i$. In our definition, we do not include the reconstruction values.

A quantizer q with cells $\{B_1, \ldots, B_M\}$, however, can also be characterized as a stochastic kernel Q from $\mathbb{X}$ to $\{1, \ldots, M\}$) defined by

$$ Q(i|x) = 1_{\{x \in B_i\}}, \quad i = 1, \ldots, M $$

so that $q(x) = \sum_{i=1}^{M} Q(i|x)$. We denote by $\mathcal{Q}_D(M)$ the space of all M-cell quantizers represented in the channel form. In addition, we let $\mathcal{Q}(M)$ denote the set of (Borel) stochastic kernels from $\mathbb{X}$ to $\{1, \ldots, M\}$, i.e., $Q \in \mathcal{Q}(M)$ if and only if $Q(\cdot|x)$ is probability distribution on $\{1, \ldots, M\}$ for all $x \in \mathbb{X}$, and $Q(i|\cdot)$ is Borel measurable for all $i = 1, \ldots, M$. Note that $\mathcal{Q}_D(M) \subset \mathcal{Q}(M)$, and by our definition $\mathcal{Q}_D(M-1) \subset \mathcal{Q}_D(M)$ for all $M \geq 2$. We note that elements of $\mathcal{Q}(M)$ are sometimes referred to in the literature as random quantizers.

Lemma 6.1: The set of quantizers $\mathcal{Q}_D(M)$ is setwise sequentially precompact at any input P.

Proof: Proof follows from Lemma 5.2 and the interpretation above regarding a quantizer as a channel. In particular, a majorizing finite measure is obtained by the relation: $PQ(B \times \{i\}) \leq P(B)$ for every $B \in \mathcal{B}(\mathbb{R}^m)$ and $i = 1, \ldots, M$. ∎

The following simple lemma provides a useful formula.

Lemma 6.2: A sequence $\{Q_n\}$ in $\mathcal{Q}(M)$ converges to a Q in $\mathcal{Q}(M)$ setwise at input P if and only if

$$\int_A Q_n(i|x)P(dx) \to \int_A Q(i|x)P(dx) \quad \forall A \in \mathcal{B}(\mathbb{X}), \forall i$$

Proof: The lemma follows by noticing that for any $Q \in \mathcal{Q}(M)$ and measurable $B \subset \mathbb{X} \times \{1, \ldots, M\}$,

$$PQ(B) = \int_B Q(dy|x)P(dx) = \sum_{i=1}^{M} \int_{B_i} Q(i|x)P(dx)$$

where $B_i = \{x \in \mathcal{X} : (x, i) \in B\}$. ∎

The following counterexample shows that the space of quantizers $\mathcal{Q}_D(M)$ is not closed under setwise convergence:

Let $\mathbb{X} = [0,1]$ and P the uniform distribution on $[0,1]$. Let $L_{nk} = \left[\frac{2k-2}{2n}, \frac{2k-1}{2n}\right)$ and let $B_{n,1} = \bigcup_{i=1}^{n} L_{nk}$ and $B_{n,2} = [0,1] \setminus B_{n,1}$. Define $\{Q_n\}$ as the sequence of 2-cell quantizers given by

$$Q_n(1|x) = 1_{\{x \in B_{n,1}\}}, \quad Q_n(2|x) = 1_{\{x \in B_{n,2}\}}.$$

The proof of Riemann-Lebesgue Lemma can be used to show that for all $A \in \mathcal{B}([0,1])$,

$$\lim_{n\to\infty} \int_A Q_n(dy|x)P(dx) = \lim_{n\to\infty} \int_0^1 \frac{1}{2} f_n(t)\, dt = \frac{1}{2}P(A),$$

and thus, by Lemma 6.2, Q_n converges setwise to Q given by $Q(1|x) = Q(2|x) = \frac{1}{2}$ for all $x \in [0,1]$. However, Q is not a (deterministic) quantizer. ∎

Definition 6.2: The class of *finitely randomized quantizers* $\mathcal{Q}_{FR}(M)$ is the convex hull of $\mathcal{Q}_D(M)$, i.e., $Q \in \mathcal{Q}_{FR}(M)$ if and only if there exist $k \in \mathbb{N}$, $Q_1, \ldots, Q_k \in \mathcal{Q}_D(M)$, and $\alpha_1, \ldots, \alpha_k \in [0,1]$ with $\sum_{i=1}^{k} \alpha_i = 1$, such that

$$Q(i|x) = \sum_{j=1}^{k} \alpha_j Q_j(i|x), \quad \text{for all } i = 1, \ldots, M \text{ and } x \in \mathbb{X}.$$

The next result shows that $\mathcal{Q}_R(M)$ is the (sequential) closure of the convex hull of $\mathcal{Q}_D(M)$.

Theorem 6.1: For any $Q \in \mathcal{Q}(M)$ there exists a sequence $\{\hat{Q}_n\}$ of finitely randomized quantizers in $\mathcal{Q}_{FR}(M)$ which converges to Q setwise at any input P.

Proof: We will prove the existence of a sequence $\{\hat{Q}_n\}$ in $\mathcal{Q}_{FR}(M)$ such that $\hat{Q}_n(\cdot|x) \to Q(\cdot|x)$ setwise for all $x \in \mathbb{X}$.

Let $\mathcal{P}_M = \{z \in \mathbb{R}^M : z_1 + \cdots + z_M = 1, z_i \geq 0, i = 1, \ldots, M\}$ denote the probability simplex in $\mathbb{R}^M$ and note that each $Q \in \mathcal{Q}(\mathcal{M})$ is uniquely represented by the function $Q^v : \mathbb{X} \to \mathcal{P}_M$ defined by

$$Q^v(x) = (Q(1|x), Q(2|x), \ldots, Q(M|x)).$$

For a positive integer n let $\mathcal{P}_{M,n}$ be the collection of probability vectors in $\mathcal{P}_M$ with rational components having common denominator n, i.e.,

$$\mathcal{P}_{M,n} = \left\{z \in \mathcal{P}_M : z_i \in \{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}, i = 1, \ldots, M\right\}.$$

Clearly, any $z \in \mathcal{P}_M$ can be approximated within error $1/n$ in the l_∞ sense by a member of $\mathcal{P}_{M,n}$, i.e.,

$$\max_{z \in \mathcal{P}_M} \min_{z' \in \mathcal{P}_{M,n}} \|z - z'\|_\infty = \max_{z \in \mathcal{P}_M} \min_{z' \in \mathcal{P}_{M,n}} \max_{i=1,\ldots,M} |z_i - z_i'| \leq \frac{1}{n}.$$

Breaking ties in a predetermined manner, we can make the selection of z' for a given z unique, and thus define a Borel measurable mapping $q_n : \mathcal{P}_M \to \mathcal{P}_{M,n}$ such that $z' = q_n(z)$ approximates z in the above sense. Given $Q \in \mathcal{Q}(\mathcal{M})$, use this mapping to define $Q_n \in \mathcal{Q}(\mathcal{M})$ through the relation

$$Q_n^v(x) = q_n(Q^v(x)).$$

(The measurability of $Q(i|x)$ in x follows from the measurability of the mapping q_n.) Let $\{z^{(1)}, \ldots, z^{(L(n))}\}$ be an enumeration of those elements of $\mathcal{P}_{M,n}$ for which the sets

$$S_j = \{x : Q_n^v(x) = z^{(j)}\}, \quad j = 1, \ldots, L(n)$$

are not empty (clearly, $L(n) \leq (n+1)^M$). Note that the S_i form an $\mathcal{A}$-measurable partition of $\mathbb{X}$ and we have

$$u := (z^{(1)}, z^{(2)}, \ldots, z^{L(n)}) \in (\mathcal{P}_M)^{L(n)}$$

and

$$Q_n^v(x) = z^{(j)} \quad \text{if } x \in S_j.$$

Viewed as a subset of $\mathbb{R}^{M \cdot L(n)}$, the set $(\mathcal{P}_M)^{L(n)}$ is compact and convex and therefore by the Krein-Milman theorem (see, e.g., [2]) it is the closure of the convex hull of its extreme points. The set of extreme points of $(\mathcal{P}_M)^{L(n)}$ is $(\mathcal{E}_M)^{L(n)}$, where $\mathcal{E}_M = \{e_1, \ldots, e_M\}$ is the standard basis for $\mathbb{R}^M$. In particular, we can find $u_1, \ldots, u_N \in (\mathcal{E}_M)^{L(n)}$ and $(\alpha_1, \ldots, \alpha_N) \in \mathcal{P}_N$ such that $\|u - \sum_{k=1}^{N} \alpha_k u_k\| \leq \frac{1}{n}$ ($\|\cdot\|$ denotes the standard Euclidean norm in any dimension). Since $u_k = (u_{k,1}, \ldots, u_{k,L(n)})$, where $u_{k,j} \in \mathcal{E}_M$ for all k and j, we can define the deterministic quantizers $Q_{n,k} \in \mathcal{Q}_D(\mathcal{M})$, $k = 1, \ldots, N$, by setting

$$Q_{n,k}^v(x) = u_{k,j} \quad \text{if } x \in S_j.$$

Putting things together, we obtain that

$$\left\| Q_n^v(x) - \sum_{k=1}^{N} \alpha_k Q_{n,k}^v(x) \right\| \leq \frac{1}{n} \quad \text{for all } x \in \mathbb{X}. \quad (4)$$

Define $\hat{Q}_n \in \mathcal{Q}(M)$ by

$$\hat{Q}_n(i|x) = \sum_{k=1}^{N} \alpha_k Q_{n,k}(i|x).$$

Combining (4) with $\|Q^v(x) - Q_n^v(x)\|_\infty \leq \frac{1}{n}$, we obtain

$$|Q(i|x) - \hat{Q}_n(i|x)| \leq \frac{2}{n} \quad \text{for all } x \in \mathbb{X} \text{ and } i = 1, \ldots, M$$

which implies that $\hat{Q}_n(\cdot|x) \to Q(\cdot|x)$ setwise for all $x \in \mathbb{X}$. Since each $\hat{Q}_n$ is a convex combination of deterministic quantizers in $\mathcal{Q}_D(M)$, the proof is complete. ∎

The preceding theorem has important consequences in that it tells us that the space of deterministic quantizers is a "basis" for the space of communication channels between $\mathbb{X}$ and $\{1, \ldots, M\}$ in an appropriate sense.

In the following we show that an optimal channel can be replaced with an optimal quantizer without any loss in performance.

Proposition 6.1: For any $Q \in \mathcal{Q}(M)$ there is a $Q' \in \mathcal{Q}_D(M)$ with $J(P, Q') \leq J(P, Q)$. If there exists an optimal channel in $\mathcal{Q}(M)$ for problem P3, then there is a quantizer in $\mathcal{Q}_D(M)$ that is optimal.

Proof: Only the first statement needs to be proved. We follow an argument common in the source coding literature (see, e.g., the Appendix of [24]).

For a policy $\gamma : \{1, \ldots, M\} \to \mathbb{U} = \mathbb{X}$ (with finite cost) define for all i,

$$B_i = \left\{ x : c(x, \gamma(i)) \leq c(x, \gamma(j)), \quad j = 1, \ldots, M \right\}.$$

Letting $B_1 = \bar{B}_1$ and $B_i = \bar{B}_i \setminus \bigcup_{j=1}^{i-1} B_j$, $i = 2, \ldots, M$, we obtain a partition $\{B_i, \ldots, B_M\}$ and a corresponding quantizer $Q' \in \mathcal{Q}_D(M)$. It is easy to see that $E_P^{Q', \gamma}[c(X, U)] \leq E_P^{Q, \gamma}[c(X, U)]$ for any $Q \in \mathcal{Q}(M)$. ∎

The following shows that setwise convergence of quantizers implies convergence under total variation.

Theorem 6.2: Let $\{Q_n\}$ be a sequence of quantizers in $\mathcal{Q}_D(M)\}$ which converges to a quantizer $Q \in \mathcal{Q}_D(M)$ setwise at P. Then, the convergence is also under total variation at P.

Proof: Let $B_1^n, \ldots, B_M^n$ be the cells of Q_n. Since $Q_n \to Q$ setwise at input P, we have $PQ_n(B \times \{i\}) \to PQ(B \times \{i\})$ for any $B \in \mathcal{B}(\mathbb{X})$. Since $PQ_n(B \times \{i\}) = \int_B 1_{\{x \in B_i^n\}} P(dx)$, we obtain

$$P(B \cap B_i^n) \to P(B \cap B_i), \quad \text{for all } i = 1, \ldots, M.$$

If $B_1, \ldots, B_M$ are the cells of Q, the above implies $P(B_j \cap B_i^n) \to P(B_j \cap B_i)$ for all $i, j \in \{1, \ldots, M\}$. Since both $\{B_i^n\}$ and $\{B_n\}$ are partitions of $\mathbb{X}$, we obtain

$$P(B_i^n \triangle B_i) \to 0 \quad \text{for all } i = 1, \ldots, M,$$

where $B_i^n \triangle B = (B_i^n \setminus B) \cup (B \setminus B_i^n)$. Then we have

$$\|PQ_n - PQ\|_{TV}$$
$$= \sup_{f: \|f\|_\infty \leq 1} \left| \sum_{i=1}^{M} \int_{\mathbb{X}} f(x, i) \left(1_{\{x \in B_i^n\}} - 1_{\{x \in B_i^n\}} \right) P(dx) \right|$$
$$\leq \sup_{f: \|f\|_\infty \leq 1} \sum_{i=1}^{M} \int_{B_i^n \triangle B_i} |f(x, i)| P(dx)$$
$$\leq \sum_{i=1}^{M} P(B_i^n \triangle B_i) \to 0 \tag{5}$$

and convergence in total variation follows. ∎

We next consider quantizers with convex codecells and a input distribution that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$ [15]. Assume $Q \in \mathcal{Q}_D(M)$ with cells $B_1, \ldots, B_M$, each of which is a convex subset of $\mathbb{R}^n$. By the separating hyperplane theorem, there exist pairs of complementary closed half spaces $\{(H_{i,j}, H_{j,i}) : 1 \leq i, j \leq M, i \neq j\}$ such that for all $i = 1, \ldots, M$,

$$B_i \subset \bigcap_{j \neq i} H_{i,j}.$$

Each $\bar{B}_i := \bigcap_{j \neq i} H_{i,j}$ is a closed convex polytope and by the absolute continuity of P one has $P(\bar{B}_i \setminus B_i) = 0$ for all $i = 1, \ldots, M$. One can thus obtain a $(P\text{–a.s})$ representation of Q by the $M(M-1)/2$ hyperplanes $h_{i,j} = H_{i,j} \cap H_{j,i}$.

Let $\mathcal{Q}_C(M)$ denote the collection of M-cell quantizers with convex cells and consider a sequence $\{Q_n\}$ in $\mathcal{Q}_C(M)$. It can be shown (see the proof of Thm. 1 in [15]) that using an appropriate parametrization of the separating hyperplanes, a subsequence Q_{n_k} can be can be chosen which converges to a $Q \in \mathcal{Q}_C(M)$ in the sense that $P(B_i^{n_k} \triangle B_i) \to 0$ for all $i = 1, \ldots, M$, where the $B_i^{n_k}$ and the B_i are the cells of Q_{n_k} and Q, respectively. In view of (5), we obtain the following.

Theorem 6.3: The set $\mathcal{Q}_C(M)$ is compact under total variation at any input measure P that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$.

We can now state an existence result for optimal quantization (problem P1).

Theorem 6.4: Let P be absolutely continuous and suppose the goal is to find the best quantizer Q with M cells minimizing $J(P, Q) = \inf_\gamma E_P^{Q, \gamma}(X, U)$ under assumption A2, where Q is restricted to $\mathcal{Q}_C(M)$. Then an optimal quantizer exists.

Proof: Existence follows from Theorems 5.1 and 6.3. ∎

In the quantization literature finding an optimal quantizer means finding optimal codecells and corresponding reconstruction points. Our formulation does not require the existence of optimal reconstruction points (i.e., optimal policy γ). For cost functions of the form $c(x, u) = \|x - u\|^p$ for $x, u \in \mathbb{R}^n$ and some $p > 0$, the cells of "good" quantizers will be convex by Lloyd-Max conditions of optimality; see [15] for further results on convexity of bins for entropy constrained quantization problems. We note that [1] also considered such cost functions for existence results on optimal quantizers; Graf and Luschgy [12] considered more general norm-based cost functions.

The concavity property applies directly to the multi-stage case. That is, the optimization problem P1 is concave in the space of channels. The proof of this fact follows that of Theorem 4.1.

One further interesting problem regarding the multi-stage case is to consider adaptive observation channels. For example, one may aim to design optimal adaptive quantizers for a control problem. In this case, Markov Decision Process tools can be used for obtaining existence conditions for optimal channels and quantizers.

VII. DISCUSSION: EMPIRICAL CONSISTENCY OF OPTIMAL CONTROLLERS

One further issue to discuss is the connections of our results with *consistency* in learning the channel from empirical observations.

When one does not know the system dynamics, such as the observation channel, one typically attempts to learn the channel via test inputs or empirical observations. Let $\{(X_i, Y_i), i \in \mathbb{N}\}$ be an $\mathbb{X} \times \mathbb{Y}$-valued i.i.d sequence generated according to some distribution μ. Defining for every measurable $B \subset \mathbb{X} \times \mathbb{Y}$, and $n \in \mathbb{N}$, the empirical occupation measures

$$\mu_n(B) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{(X_i, Y_i) \in B\}},$$

one has $\mu_n(B) \to \mu(B)$ almost surely (a.s.) by the strong law of large numbers. However, it is generally not true that $\mu_n \to \mu$ setwise a.s. (e.g., μ_n never converges to μ setwise when either X_i or Y_i has a nonatomic distribution), in which case μ_n cannot converge to μ in total variation.

Again by the strong law, for any μ-integrable function f on $\mathbb{X} \times \mathbb{Y}$, one has, almost surely,

$$\lim_{n\to\infty} \int f(x,y)\mu_n(dx,dy) = \int f(x,y)\mu(dx,dy)$$

In particular, $\mu_n \to \mu$ weakly with probability one [10].

In the learning theoretic context, the convergence of the costs optimal for μ_n to the cost optimal for μ is called the consistency of empirical risk minimization (see [23] for an overview). In particular, if the cost function and the allowable control policies $\mathcal{F}$ are such that

$$\lim_{n\to\infty} \sup_{\gamma \in \mathcal{F}} \left| \int c(x,\gamma(y))\mu_n(dx,dy) - \int c(x,\gamma(y))\mu(dx,dy) \right|$$
$$= 0,$$

then we obtain consistency.

A class of measurable functions $\mathcal{E}$ is called a *Glivenko-Cantelli class* [9], if the integrals with respect to the empirical measures converge almost surely to the integrals with respect to the true measure uniformly over $\mathcal{E}$. Thus, if

$$\mathcal{G} = \{\gamma : c(x,\gamma(y)) \in \mathcal{E}\},$$

where $\mathcal{E}$ is a class of Glivenko-Cantelli family of functions, then we could establish consistency. One example of a Glivenko-Cantelli family of real functions on $\mathbb{R}^N$ is the family $\{f :$

$\|f\|_{BL} \le M\}$ for some $0 < M < \infty$, where $\| \cdot \|_{BL}$ denotes the bounded Lipschitz norm [9].

Thus, if we restrict the class of control policies, and given a cost function, we can obtain consistency and robustness to mismatch in the channel due to learning. The classification of the class of objective functions and policies which would lead to such a consistency result is a future research problem.

VIII. CONCLUDING REMARKS AND EXTENSIONS

This paper studied the structural and topological properties of some optimization problems in stochastic control in the space of observation channels. The main problem we considered is how to approach appropriate notions of convergence and distance while studying communication channels in the context of stochastic control problems.

It was observed that the optimization problem is concave in such channels. One implication of this observation is that in a decentralized control problem, if signaling is present, the original convex problem (which may be convex under a nested, partially nested or a stochastically nested information structure) loses its convexity.

The restriction to Euclidean state spaces is not essential and many (but not all) of the results can be extended to the case where $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{U}$ are Polish spaces. In particular, all the results in Sections III and IV carry through without change, except Theorem 3.3. The results of Section V hold for this more general setup (however, in Lemma 5.3 we need the additional condition that the space is σ-compact). Likewise, the most results in Section VI on quantization hold more generally (in fact, Theorem 6.1 holds for an arbitrary measurable space), but two of the main results, Theorems 6.3 and 6.4, do need the assumption that $\mathbb{X}$ is a finite-dimensional Euclidean space.

A. Sufficient conditions for continuity under setwise and weak convergence

It turns out that we need a uniform convergence principle for setwise convergence to be sufficient for continuity:

$$\lim_{n\to\infty} \sup_{\gamma \in \mathcal{G}} \left| \int \left(\int Q(dy|x)c(x,\gamma(y)) \right.\right.$$
$$\left.\left. - \int Q_n(dy|x)c(x,\gamma(y)) \right) P(dx) \right| = 0,$$

to be able to have continuity under setwise convergence. Thus, one important question of practical interest, is the following: What type of stochastic control problems, cost functions, and allowable policies lead to solutions which admit such a uniform convergence principle under setwise convergence? Some sufficient conditions for uniform setwise convergence are presented in [22].

B. Robustness in source distribution

The results in this paper may also be applicable to robustness to source distribution. We considered robustness in the observation channel. However, some of results also apply almost without change to robustness in the input distribution. Ornstein's $\bar{d}$-distance [18], [13] is also a meaningful distance

when the uncertainty is in the input distribution, but requires the cost function to be specific. Such problems for the specific case of minimum mean square estimation were considered in [25]. The important general problem merits further research.

REFERENCES

[1] E. A. Abaya and G. L. Wise, "Convergence of vector quantizers with applications to optimal quantization, *SIAM Journal on Applied Mathematics*, vol. 44, pp. 183–189, 1984.

[2] A. Barvinok, *A Course in Convexity*, vol. 54 of Graduate Studies in Mathematics. Providence, RI: American Mathematical Society, 2002.

[3] P. Billingsley, *Convergence of Probability Measures*. New York: Wiley 1968.

[4] J. von Neumann, "A certain zero-sum two-person game equivalent to the optimal assignment problem," *Contrib. to the Theory of Games*, vol. 2, pp. 5–12, Princeton University Press, Princeton, New Jersey, 1953.

[5] D. Blackwell, "The comparison of experiments," *in Proc. Second Berkeley Symposium on Mathematical Statistics and Probability,* pp. 93–102. Univ. of California Press, Berkeley, 1951.

[6] C.D. Charalambous and F. Rezaei, "Stochastic uncertain systems subject to relative entropy constraints: Induced norms and monotonicity properties of minimax games", *IEEE Transactions on Automatic Control*, vol. 52, no. 4, pp 647–663, May 2007.

[7] F. Rezaei, C.D. Charalambous and N. U. Ahmed, "Optimization of stochastic uncertain systems with variational norm constraints", in *Proc. IEEE Conference on Decision and Control*, pp. 2159–2163, New Orleans, LA, USA, Dec. 2007

[8] Y. Socratous, F. Rezaei and C.D. Charalambous, "Nonlinear estimation for a class of systems", *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1930-01938, Apr. 2009.

[9] R.M. Dudley, E. Gine, and J. Zinn, "Uniform and universal Glivenko-Cantelli classes", *Journal of Theoretical Probability*, vol. 4, pp. 485–510, 1991.

[10] R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, Cambridge, 2nd ed., 2002.

[11] N. Dunford and J. Schwartz, *Linear operators I*, Interscience, New York, 1958

[12] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*. Berlin, Heidelberg: Springer Verlag, 2000.

[13] R. M. Gray and L. D. Davisson, "Quantizer mismatch," *IEEE Transactions on Communications*, vol. 23, pp. 439–443, 1975.

[14] A. György and T. Linder, "Optimal entropy-constrained scalar quantization of a uniform source", *IEEE Transactions on Information Theory*, vol. 46, pp. 2704–2711, Nov. 2000.

[15] A. György and T. Linder, "Codecell convexity in optimal entropy-constrained vector quantization, *IEEE Transactions on Information Theory*, vol. 49, pp. 1821–1828, Jul. 2003.

[16] O. Hernandez-Lerma, J.B. Lasserre, *Markov Chains and Invariant Probabilities*, BirkhäuserVerlag, Basel, 2003.

[17] T. Linder, "On the training distortion of vector quantizers", *IEEE Transactions on Information Theory*, vol. 46, pp. 1617–1623, 2000.

[18] D. Ornstein, "An application of ergodic theory to probability theory, *The Annals of Probability*, vol. 1, pp. 43–58, 1973.

[19] D. Pollard, "Quantization and the method of k-means, *IEEE Transactions on Information Theory*, vol. 28, pp. 199–205, 1982.

[20] W. Rudin, *Real and Complex Analysis*, New York: McGraw-Hill, 3rd ed., 1987.

[21] J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers", *IEEE Transactions on Communications*, Vol. 41, pp. 550–558, Apr. 1993.

[22] F. Topsoe, "Uniformity in convergence of measures", *Z. Wahrschein-lichkeitsth*, vol. 39, pp. 1–30, 1977.

[23] V. N. Vapnik *The Nature of Statistical Learning Theory* Springer, New York, 2nd ed., 2000.

[24] H. S. Witsenhausen, "On the structure of real-time source coders," *Bell Syst. Tech. J.*, 58:1437-1451, July/August 1979.

[25] Y. Wu and S. Verdú, "Functional properties of MMSE", *Proc. IEEE International Symposium on Information Theory*, Austin, TX, June 13-18, 2010.

[26] S. Yüksel and T. Linder, "Optimization and Convergence of Observation Channels in Stochastic Control", submitted to *IEEE American Control Conference*, 2011.

[27] S. Yüksel and T. Linder, "Optimization and Convergence of Observation Channels in Stochastic Control", submitted to *SIAM Journal on Control and Optimization*, 2010 (available on arXiv).