ISIT'2013 1569719743

Dynamic Joint Source–Channel Coding with Feedback

Tara Javidi
Electrical & Computer Engineering
University of California, San Diego
Email: tara@ece.ucsd.edu

Andrea Goldsmith
Electrical Engineering
Stanford University
Email: andrea@wsl.stanford.edu

***Abstract*—This paper considers real time joint source–channel coding of a Markov source over a discrete memoryless channel with noiseless feedback. The encoder incurs a cost which is minimized along with a real-time end-to-end distortion. The problem is mapped to a partially observable Markov decision problem and the corresponding optimality equations, in the form of dynamic programming equations, are derived. As a consequence of the dynamic programming formulation, basic structural properties of the optimal encoding and decoding strategies are established. In addition, the problem formulation and solution obtained for dynamic joint source–channel coding with noiseless feedback is shown to encompass a much broader class of problems including that of information acquisition and real time tracking.**

I. Introduction

The coding theorem for discrete momoryless channels and ergodic sources asserts the possibility of reliable communication for all sources with information rate less than the Shannon capacity of the channel [1] (and no such communication is possible for any higher rate). However, this result holds only in the limit of an infinite number of channel uses. This fundamental observation, as noted by many [2], means little for practical systems that are constrained by delay requirements of the application layer. An alternative view seeks to find the optimum encoder and decoder for a system with one unit of time delay. Such a formulation is naturally motivated by control and tracking applications in which the source symbols represent time sensitive information.

In this paper, we consider joint source–channel coding of a Markov source over a discrete memoryless channel (DMC) with perfect output feedback. The encoding and transmission are assumed to incur a general cost which is to be minimized along with a real-time end-to-end distortion. The existence of the noiseless feedback allows the encoding, and subsequently the decoding, to be selected dynamically in order to (re)evaluate and trade-off the cost of encoding with the distortion loss. We formulate this problem as a partially observable Markov decision problem (POMDP) and derive a set of optimality equations in which the expected total cost of communication and distortion is minimized subject to causality of coding and encoding strategies.

While dynamic joint source–channel coding with real time distortion and noise-free feedback falls within the class of of real-time information transmission problems, it has a broad applicability well beyond this class. In particular, we show that our problem formulation and solution for dynamic joint source-channel coding is applicable to a large class of information acquisition problems such as real-time tracking.

The remainder of this paper is organized as follows. In Section II, we introduce the problem of dynamic joint source–channel coding with noiseless feedback and communication cost constraints. Posed as a partially observable Markov Decision problem, we derive the optimality equations in both finite horizon and infinite horizon setups. Furthermore, to illustrate the approach, we specialize the results for the case of joint source–channel coding over a binary input channel. In this special case, we also illustrate the significance of the optimality equation in guiding simple heuristic policy design. In Section III, we introduce the problem of dynamic tracking in a noisy environment and show that it is a special case of dynamic joint source–channel coding with noiseless feedback. Finally, we conclude the paper in Section IV.

Notations and Definitions: A random variable is denoted by an upper case letter (e.g. X) and its realization is denoted by a lower case letter (e.g. x). Similarly, a random vector and its realization are denoted by bold face symbols (e.g. $\mathbf{X}$ and $\mathbf{x}$). The i-th element of vector $\mathbf{v}$ is denoted by v_i or $v(i)$. For any set S, $|S|$ denotes the cardinality of S. Throughout the paper, for a given set E, we use notation $\mathcal{M}(E)$ to denote the set of all real-valued functions on E. The entropy function on a vector $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_M) \in [0,1]^M$ is defined as $H(\boldsymbol{\rho}) := \sum_{i=1}^{M} \rho_i \log(1/\rho_i)$. Also, the *Kullback–Leibler (KL) divergence* between two probability distributions P_Z and P'_Z is defined as $D(P_Z||P'_Z) := \sum_{z\in\mathcal{Z}} P_Z(z) \log \frac{P_Z(z)}{P'_Z(z)}$ with the convention $0 \log \frac{a}{0} = 0$ and $b \log \frac{b}{0} = \infty$ for $a, b \in [0,1]$ with $b \neq 0$. Finally, we call a matrix A sub-rectangular if $A_{ij}, A_{i'j'} > 0$ implies that $A_{i'j}, A_{ij'} > 0$.

II. Dynamic Joint Source-Channel Coding with Feedback

Consider the problem of dynamic joint source-channel coding (JSCC) over a discrete memoryless channel with noiseless feedback as depicted in Fig. 1. During a decision horizon interval T, the transmitter wishes to communicate a Markov source symbol θ_t to the receiver, where θ_t belongs to the source message set $\Omega = \{1, 2, \ldots, M\}$ with a transition probability matrix Q. The DMC is described by finite input and output sets $\mathcal{X}$ and $\mathcal{Y}$, and the collection of conditional probabilities $P(Y|X)$. Given an encoding function $e_t \colon \Omega \to \mathcal{X}$, agreed upon by the transmitter and the receiver to be used at a given time t, the transmitter produces channel input X_t, as a function of the source θ_t:

$$X_t = e_t(\theta_t). \tag{1}$$

We further assume that the encoding function e incurs an encoding cost of $c(e)$ to capture the complexity of encoding

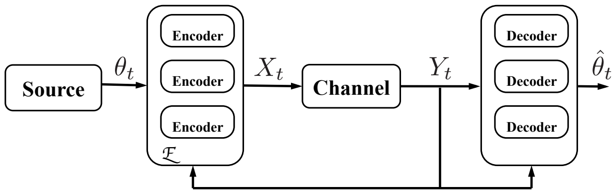

Fig. 1. Dynamic joint source channel coding over a discrete memoryless channel with perfect output feedback. The encoding and decoding functions are selected in real time based on the channel output so far.

and the communication cost associated with a channel use. We also allow for a zero cost encoding $\star$, where the choice of $\star$ encoding at time t indicates the decision to remain silent for this transmission in order to save the communication cost.

Given some decoding function $d_t \colon \mathcal{Y}^t \to \Omega$ to be utilized after observing the t channel outputs $Y_1, \ldots, Y_t$, the receiver estimates the message θ_t as

$$\hat{\theta}_t = d_t(Y_{1:t}), \tag{2}$$

and incurs a distortion loss of $\delta(\theta, \hat{\theta})$ where $\delta : \Omega \times \Omega \to \mathbb{R}^+$ is a known and given distortion function.

The problem of dynamic joint source channel coding with feedback considers the case where the encoding function at any given time t can be selected, in real time, from a fixed and known class of encoding functions $\mathcal{E} \subset \Omega \times \mathcal{X} \cup \{\star\}$ as a function of the past channel outputs $Y_1, \ldots, Y_{t-1}$ which, thanks to the noiseless causal feedback, are also available to the transmitter. To account for the real time (sequential) selection of the encoding function, it is only natural that the decoding function is also selected in a real time (sequential) manner. Let E_{t+1} and D_t be the random elements, measurable with respect to $\sigma(Y_{1:t})$, denoting the encoding and decoding functions at time t, $t = 1, \ldots, T$. The objective is to find functions E_{t+1} and D_t to minimize the expected total distortion plus encoding cost. In other words, given a Bayesian prior on the initial state $\theta_1 \sim \boldsymbol{\rho}$, the objective is to design encoding and decoding functions E_t and D_t, $t = 1, \ldots, T$, which minimize the weighted total distortion loss and encoding cost

$$J_T(\boldsymbol{\rho}) := \mathbb{E}\left[\sum_{t=1}^{T} \delta(\theta_t, \hat{\theta}_t) + \lambda \sum_{t=1}^{T} c(E_t)\right]. \tag{3}$$

A slightly less stringent objective is that of minimizing the expected (long-term) average cost criterion:

$$J(\boldsymbol{\rho}) := \lim_{t\to\infty} \frac{1}{T} \mathbb{E}\left[\sum_{t=1}^{T} \delta(\theta_t, \hat{\theta}_t) + \lambda \sum_{t=1}^{T} c(E_t)\right]. \tag{4}$$

Remark 1. The total and average cost criteria (3) and (4), for a proper choice of multiplier λ, is equivalent to the Lagrangian relaxation of the problem of minimizing distortion subject to a constraint on the expected encoding cost.

A. Related Work

In the interest of brevity and because of the inherent difference in our cost and real time constraints, we refrain from discussing the large body of work on joint source–channel coding in the classical context where the reconstruction of the source symbols is allowed with a (large) delay.

The problem of dynamic joint source channel coding with feedback and communication/distortion cost generalizes the problem of variable length coding over discrete memoryless channels with feedback tackled by Burnashev [3], by considering a dynamic source and also by accounting for a cumulative and generalized notion of distortion. In other words, when the source transition matrix is an identity matrix, the encoding cost is fixed and non-zero for all functions but $\star$, and the expected terminal distortion is the weighted probability of error, the problem above matches that considered by Burnashev. Note that due to the static nature of the source in [3], where Q is the identity matrix, it is only natural to expect a random time after which the receiver's knowledge about the message is within an acceptable reliability constraint and, hence, after which the transmitter becomes silent. Mathematically, this means that $\lim_{T\to\infty} J_T(\boldsymbol{\rho})$ exists and is finite and for which an asymptotically tight set of lower and upper bounds are known [3].

The problem of dynamic JSCC with feedback also generalizes the problem of causal encoding and decoding studied by Walrand and Varayia [4] where the source alphabet Ω was taken to be the same as the channel input alphabet $\mathcal{X}$ and in which case no coding was shown to be optimal for certain channels with strong symmetric properties.

The model above is also closely related to the problem of sequential real time source coding of Borkar et al. [5], where a sequential quantization of a Markov source has been considered. More specifically, the objective in [5] is to jointly optimize the entropy rate of the quantized process (in order to obtain a better compression rate) as well as a suitable distortion measure. Similar to our formulation, the real time (sequential) choice of quantizer there affects the symbol-by-symbol real time distortion as well as the encoding cost. However, this cost is measured in terms of the efficiency/rate of the overall compression achieved. In other words, instead of an explicit cost function $c(\cdot)$ associated with each symbol produced by the quantizer, the authors rely on the entropy rate of the quantized process as a proxy for the source encoding costs.

We end this subsection by noting that the problem of joint source-channel coding with feedback has interesting connections to the problems of active hypothesis testing [6] as well as quickest detection problems, even though exploration of these connections remains beyond the scope of this paper.

B. Analysis

Given a Bayesian prior ρ on the initial distribution of θ_1, the above problem is a partially observable Markov decision problem (POMDP) where the state transition matrix is Q and the observations are noisy. This problem is equivalent to an MDP whose information state (sufficient statistics) at time t is the belief vector $\boldsymbol{\rho}(t) = [\rho_1(t), \ldots, \rho_M(t)]$ where

$$\rho_i(t) = \text{Prob}(\theta_t = i \,|Y_{1:t}\,),$$

and the information state space is defined as $\mathbb{P}(\Omega) = \{\boldsymbol{\rho} \in [0,1]^M : \sum_{i=1}^{M} \rho_i = 1\}$.

Let $\Phi^e(\boldsymbol{\rho}, y)$ denote the posterior belief state, given current belief (posterior), the encoding function e, and channel output

y. This means that for all $i \in \Omega$, posterior belief

$$\Phi^e(\boldsymbol{\rho}(t), y)(i) := \frac{\rho_i^+(t) P(Y = y | X = e(i)))}{\sum_{j=1}^M \rho_j^+(t) P(Y = y | X = e(j)))},$$

where

$$\rho_i^+(t) := \mathbb{P}(\theta_{t+1} = i | Y^t) = \sum_{j \in \Omega} \rho_j(t) Q_{ji}$$

is the *pre-transmission one step predictor* of the source symbol.

Furthermore, for an encoding function e, define the corresponding Markov operator $\mathbb{T}^e : \mathcal{M}(\mathbb{P}(\Omega)) \to \mathcal{M}(\mathbb{P}(\Omega))$ as

$$(\mathbb{T}^e g)(\boldsymbol{\rho}) := \mathbb{E}\left[g\left(\Phi^e(\boldsymbol{\rho} Q, Y)\right)\right].$$

In addition, for declaration θ' and $\boldsymbol{\rho} \in \mathbb{P}(\Omega)$, let $\Delta(\boldsymbol{\rho}, \theta') = \sum_i \rho_i(t) \delta(i, \theta')$ denote the expected distortion.

Fact 1 (Propositions 3.1 in [7])**.** *Define recursively the functions:*

$$V_T(\boldsymbol{\rho}) = \min_{\theta'} \Delta(\boldsymbol{\rho}, \theta') \tag{5}$$

$$V_t(\boldsymbol{\rho}) = \min_{\theta', e} \Delta(\boldsymbol{\rho}, \theta') + c(e) + (\mathbb{T}^e V_{t+1})(\boldsymbol{\rho}). \tag{6}$$

Then $V_1(\boldsymbol{\rho}_1)$, known as the optimal value function at $t = 1$, is equal to the minimum cost $J_T(\boldsymbol{\rho}_1)$ in (3) with the prior belief $\boldsymbol{\rho}_1$.

Important consequences of the above set of dynamic programming equations are as follows.

1) The minimizers of (5) and (6) constitute deterministic optimal Markov policy (selecting declarations and inspection regions as deterministic functions of the sufficient statistic $\boldsymbol{\rho}$).
2) The optimal declaration is the Bayes' risk minimizer which, at any given belief vector $\boldsymbol{\rho}$, minimizes the expected current distortion. When the distortion is measured in terms of error probability (when $\delta(\theta, \theta') = \mathbf{1}_{\theta = \theta'}$), the optimal declaration is nothing but the maximum aposterior decoder, i.e. $\hat{\theta}^* = \arg\max_i \rho_i(t)$.
3) If $\min_{\theta'} \Delta(\boldsymbol{\rho} Q, \theta') \leq c(e)$ for all $e \subset \mathcal{E}$, $e \neq \star$, then it is optimal for the transmitter to skip the last transmission. The result can be generalized for skipping the last τ transmissions.

Proposition 1. *Assume that the source Q is aperiodic and irreducible. In addition, assume that $P(\cdot | X = x)$ is absolutely continuous with respect to $P(\cdot | X = x')$, for all $x, x' \in \mathcal{X}$. Then there exists scalar V^* and bounded function $W \in \mathcal{M}(\mathbb{P}\Omega)$ such that*

$$V^* + W(\boldsymbol{\rho}) = \min_{\theta', e} \Delta(\boldsymbol{\rho}, \theta') + c(e) + (\mathbb{T}^e W)(\boldsymbol{\rho}), \tag{7}$$

and V^* is equal to the minimum cost $J(\boldsymbol{\rho}_1)$ in (4) (which becomes independent of initial prior).

Proof: Equation (7) is referred to as the average cost optimality equation (ACOE) and has a dynamic programming interpretation and flavor. However, unlike the case of finite horizon (and infinite horizon with discounted cost criterion), the problem of minimizing the average expected cost criterion is not guaranteed to coincide with the solution to the ACOE. Here, the main challenge in the proof arises due to the uncountable nature of the state space $\mathbb{P}(\Omega)$ and a lack of guarantees for the ergodicity of the information state process $\rho(t)$, even when Q is ergodic (see [8] for a detailed discussion).

However, when Q is an aperiodic and irreducible Markov chain and when $P(\cdot|x)$ and $P(\cdot|x')$ are absolutely continuous with respect to each other for all $x, x' \in \mathcal{X}$, one can obtain optimality equation (7). More specifically, for a given $y \in \mathcal{Y}$ and encoding function $e \subset \mathcal{E}$, define the $M \times M$ substochastic matrix $P^{(y,e)}$ to be such that

$$P^{(y,e)}(i,j) := Q_{i,j} P(Y = y | X = e(j))$$

and for a finite sequence of encoding functions $\underline{e} = (e_1, e_2, \ldots, e_n)$ and the corresponding sequence of channel outputs $\underline{y} = (y_1, y_2, \ldots, y_n)$, define the product matrix

$$P^{(\underline{y}, \underline{e})} := P^{(y_1, e_1)} P^{(y_2, e_2)} \ldots P^{(y_n, e_n)}.$$

From the absolute continuity of $P(\cdot|x)$ and $P(\cdot|x')$ with respect to each other and the finiteness of $\mathcal{Y}$, we have that for all sequences $\underline{e} = (e_1, e_2, \ldots, e_n)$ and $\underline{y} = (y_1, y_2, \ldots, y_n)$

$$P_{ij}^{(\underline{y}, \underline{e})} \geq Q_{ij}^n . \prod_{i=1}^n \min_x P(Y = y_i | X = x).$$

On the other hand, from the aperiodicity and irreducibility of Q, there exists m for which $Q^n, \forall n \geq m$, is subrectangular. Appealing to Lemma B.1. and Theorems 1-3 of [8], we have the assertion of the proposition.

■

Similarly, the above dynamic programming result (and the corresponding optimality equation) has the following simple consequences:

1) The minimizer of the right hand side of (7) constitutes a deterministic *stationary* optimal Markov policy (selecting declarations and inspection regions is a deterministic and *time invariant* function of the sufficient statistic $\boldsymbol{\rho}$):
   - The optimal declaration is the Bayes' risk minimizer which, at any given belief vector $\boldsymbol{\rho}$, minimizes the expected current distortion. When the distortion is measured in terms of error probability (when $\delta(\theta, \theta') = \mathbf{1}_{\theta = \theta'}$), the optimal declaration is nothing but the maximum aposterior decoder, i.e. $\hat{\theta}^* = \arg\max_i \rho_i(t)$.
   - The optimal encoding function is given as :

$$\arg\min_e \{c(e) + (\mathbb{T}^e W)(\boldsymbol{\rho}) - W(\boldsymbol{\rho})\} \tag{8}$$

2) An important consequence of Proposition 1 is that the V^* can be approximated by considering the finite horizon problem of (5)-(6) as T gets large [8]. An alternative technique relies on the vanishing discount technique with an infinite horizon discounted cost criterion, which enables computational solutions via value/policy iteration.

Note that characterization of the optimal policy has the following intuitive interpretation associated with the following notion of information utility.

Definition 1. Associated with a functional $h : \mathbb{P}(\Theta) \to \mathbb{R}_+$, the *information utility* of action a at information state $\boldsymbol{\rho}$ is defined as $\mathcal{IU}(a, \boldsymbol{\rho}, h) := h(\boldsymbol{\rho}) - (\mathbb{T}^a h)(\boldsymbol{\rho})$.

Together with (8), this results in the following optimal deterministic choice of encoding function:

$$\arg\max_{e} \{\mathcal{IU}(e, \boldsymbol{\rho}, W) - c(e)\}$$

maximizing the information utility minus cost at any given belief $\boldsymbol{\rho}$.

C. Example: Joint Source–Channel Coding with Feedback over a Binary Input Channel

We now consider a special, and perhaps more conventional, case of the problem where the encoding cost is zero in case of silence, i.e. when encoding function $\star$ is selected, and otherwise incurs a unit cost: $c(e) = 1_{X \neq \star}, e \in \mathcal{E}$. Furthermore, in this example, we assume that the channel input is binary $\mathcal{X} = \{0, 1\}$.

The transmitter produces channel inputs X_t for $t = 1, \ldots, T$, which it can compute as a function of the source θ_t and (thanks to the noiseless causal feedback) also of the past channel outputs $Y_1, \ldots, Y_{t-1}$:

$$X_t = e_t(\theta_t, Y_{1:t-1}), \quad t = 1, \ldots, T, \tag{9}$$

for some encoding function $e_t \colon \Omega \times \mathcal{Y}^{t-1} \to \{0, 1\} \cup \{\star\}$, where again the choice of $\star$ symbol indicates the transmitter's decision to remain silent. In order to avoid ambiguity regarding the timing of the transmitter's silent intervals on the receiver's side, the transmitter's choice of silence symbol is assumed to be only a function of past channel outputs (which is known to both transmitter and receiver) and not the messages. In other words, we restrict our encoding functions such that

$$X_t = e_t(\theta_{1:t}, Y_{1:t-1}) = \star \quad \Rightarrow e_t(\theta'_{1:t}, Y_{1:t-1}) = \star. \tag{10}$$

After observing the t channel outputs $Y_1, \ldots, Y_t$, the receiver estimates the message θ_t as

$$\hat{\theta}_t = d_t(Y_{1:t}), \tag{11}$$

for some decoding function $d \colon \mathcal{Y}^t \to \Omega$.

The objective is to find encoding and decoding functions to minimize the expected total distortion plus transmission cost, i.e. to design encoding and decoding functions e_t and d_t, $t = 1, \ldots, T$, which minimize:

$$\lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} \delta(\theta_t, \hat{\theta}_t) + \lambda \sum_{t=1}^{T} \mathbf{1}_{\{X_t \neq \star\}}\right]$$

Remark 2. When $\delta(\theta_t, \hat{\theta}_t) = 1_{\hat{\theta}_t \neq \theta_t}$, the problem, via a Lagrangian relaxation, can be shown to be equivalent to minimizing Pe $:= \frac{1}{T}\sum P(\hat{\theta}_t \neq \theta_t)$ subject to a average number of channel uses.

Now let us specialize the more general result of Section II-B in the case of JSCC over binary input channels with feedback: At time t, the transmitter sequentially partitions the source symbols Ω into those that are mapped to input 0, S_0 and those mapped to input 1, referred to as $S_1 = \Omega - S_0$. Note that this partition at any given time t only depends on the current source symbol θ_t and the current belief state $\boldsymbol{\rho}(t-1)$.

Notice that for all $i \in \Omega$ and deterministic rule e_t, upon observing sample y_t, the belief state evolves as

$$\rho_i(t+1) = \begin{cases} \frac{\rho_i^+(t) f_1(Y)}{f_1(Y)\rho_{S_1}^+(t) + f_0(Y)\rho_{S_0}^+(t)} & \text{if } \ i \in S_1 \\ \frac{\rho_i^+(t) f_0(Y)}{f_1(Y)\rho_{S_1}^+(t) + f_0(Y)\rho_{S_0}^+(t)} & \text{if } \ i \in S_0 \end{cases}.$$

where $\rho_i^+(t) := \mathbb{P}(\theta_{t+1} = i | Y^t) = \sum_{j\in\Omega} \rho_j(t) Q_{ji}$, and $\rho_S^+(t) = \sum_{i \in S} \rho_i^+(t),$, $S = S_0, S_1$. And the optimal encoding function, which as seen before maximizes the information utility (all encoding functions have unit cost), can be identified as $\arg\max_e \{\mathcal{IU}(e, \boldsymbol{\rho}, W)\}$.

Despite the finite horizon and vanishing discount approximation techniques mentioned in Section II-B, solving for V^* and W can be computationally cumbersome (a closed form often cannot be obtained analytically). Following the seminal view of [9] and in lieu of solving for these functions, a useful technique is to consider replacing function $W(\cdot)$ with an alternative measures of informativity to derive useful heuristics. In particular, consider the following approximation based on two candidate functionals:

- Entropy: $H(\boldsymbol{\rho}) = \sum_{i=1}^{M} \rho_i \log \frac{1}{\rho_i}$
- Average log–likelihood: $U(\boldsymbol{\rho}) = \sum_{i=1}^{M} \rho_i \log \frac{1-\rho_i}{\rho_i}$

and the corresponding heuristic policies that maximize the information utility associated with these functionals. In other words, consider $\arg\max_e \{\mathcal{IU}(e, \boldsymbol{\rho}, H)\}$ and $\arg\max_e \{\mathcal{IU}(e, \boldsymbol{\rho}, U)\}$:

1) The first scheme is such that if transmission is chosen over silence, then the encoding is done in a manner to maximize $\mathcal{IU}(e, \boldsymbol{\rho}, H)$. In other words, set Ω is partitioned into sets $S_0(t), S_1(t)$ such that the mutual information of the sample distributions at the pre-transmission one step predictor belief

$$\sum_{k=0,1} \sum_{i \in S_k(t)} \rho_i^+(t) D\left(P(Y|X=k) \,\|P(Y)\right),$$

is maximized where

$$P(Y) = \sum_{l=0,1} \pi_l(t) P(Y|X=l)$$

and

$$\pi_l(t) = \sum_{i \in S_l(t)} \rho_i^+(t), \ l = 0, 1.$$

2) The second coding scheme is such that upon a transmission, $\mathcal{IU}(e, \boldsymbol{\rho}, U)$ is maximized. Now this is equivalent to partitioning message set Ω into sets $S_0(t), S_1(t)$ such that the EJS divergence [10] of the conditional sample/output distributions at the pre-transmission one step predictor belief

$$\sum_{k=0,1} \sum_{i \in S_k(t)} \rho_i^+(t) D\left(P(Y|X=k) \,\|P(Y_{-i})\right),$$

is maximized where

$$P(Y_{-i}) := \sum_{l \in \mathcal{X}} \hat{\pi}_{l,-i}(t) P(Y|X=l),$$

and

$$\hat{\pi}_{l,-i}(t) = \frac{1}{1 - \rho_i^+(t)} \sum_{j \in S_l(t) - \{i\}} \rho_j^+(t), \ l = 0, 1.$$

Contrast these constructions with randomized schemes under which each message, independently from the others, is assigned to set $S_0(t)$ with probability π_0^*, where π_0^* is a given fixed parameter. In other words, under this scheme and at any time t, partition $(S, \Omega - S)$, $S \in 2^{\Omega}$, is selected with probability λ_S^* where $\lambda_S^* = (\pi_0^*)^{|S|}(1 - \pi_0^*)^{M-|S|}$. Note that this strategy does not take advantage of the feedback and, when applied in the JSC problem, coincides with a random coding scheme with input distribution $(\pi_0^*, 1-\pi_0^*)$.

III. From Dynamic JSCC to Dynamic Tracking

We shall now show the somewhat surprising result that the JSCC with the binary input channel example of Section II-C can be exactly mapped to a visual tracking problem. Specifically, consider the problem of visually tracking a unique target of interest in a noisy image. We assume a time slotted scenario. In particular, consider the case where the search area is broken into finite segments indexed by $\Omega = \{1, 2, \ldots, M\}$ as shown in Figure 2, and the target is moving between these segments in a fixed and known Markovian fashion with the transition matrix Q. Let $\theta_t \in \Omega, t = 1, 2, \ldots$ denote the

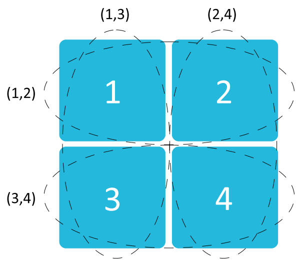

Fig. 2. Visually tracking of a moving target in a noisy image with 4 segments. Inspection regions consisting of neighboring segments are marked in dashed lines.

(segment) location of the target at time t and $\mathcal{Y}$ denote the observation space. Our tracker/decision-maker is interested in real time and sequential (generalized) tracking of the object, via visual inspection of selected segments. In other words, after observing t samples $Y_1, Y_2, \ldots, Y_t$, the tracker estimates the location of the target θ_t as

$$\hat{\theta}_t = d_t(Y_{1:t}), \tag{12}$$

for some declaration function $d\colon \mathcal{Y}^t \to \Omega$. We also assume that this declaration incurs a distortion loss of $\delta(\theta, \hat{\theta})$ where $\delta : \Omega \times \Omega \to \mathbb{R}^+$ is a known and given distortion function.

Furthermore, the tracker is tasked with selecting a subset of the segments $S_t \subset \Omega$ for further visual inspection as a function of the past visual samples, $Y_1, Y_2, \ldots, Y_t$:

$$S_{t+1} = v_{t+1}(Y_{1:t}), \quad t = 1, \ldots, T-1 \tag{13}$$

for some visual inspection function $v_t : \mathcal{Y}^{t-1} \to 2^{\Omega}$. It is natural to assume that inspecting different sets of segments might require different levels of effort and resources. In general, we assume that inspecting subset S, incurs a known and fixed cost of $c(S)$. This cost might be associated with the complexity of inspecting set S, for example due to its size.

Upon the selection of a non-empty set S of segments to be inspected, the tracker obtains a noisy visual sample Y whose conditional distribution is known and given as f_1 if one of the segments contains the target and as f_0 otherwise, i.e.

$$Y_t \sim \begin{cases} f_1(\cdot) & \text{if } \theta_t \in S \\ f_0(\cdot) & \text{otherwise} \end{cases}.$$

The objective is to choose the inspection and declaration functions v_t and d_t in order to best track the target while accounting for the cumulative inspection cost:

$$J_T(\boldsymbol{\rho}) := \mathbb{E}\left[\sum_{t=1}^{T} \delta(\theta_t, \hat{\theta}_t) + \sum_{t=1}^{T} c(S_t)\right] \tag{14}$$

or the expected (long-term) average cost:

$$J(\boldsymbol{\rho}) := \lim_{t\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} \delta(\theta_t, \hat{\theta}_t) + \sum_{t=1}^{T} c(S_t)\right]. \tag{15}$$

Note that despite the apparent difference in the problem statements, the problem of visual tracking is nothing but a dynamic JSCC over a binary input channel where $f_x(\cdot)$ gives the output distribution given input $X = x$, and where the encoding function class $\mathcal{E}$ is nothing but all indicator functions associated with the possible inspection regions $S \subset \Omega$. In light of this equivalence, one can take advantage of the dynamic programming optimality equations given in Fact 1 and Proposition 1. Furthermore, this view allows us to utilize various coding heuristics/strategies in this seemingly different setup. In particular the schemes of Section II-C can provide heuristic tracking strategies.

IV. Discussion and Future Work

The problem of dynamic joint source–channel coding with noiseless feedback, where the encoding is selected dynamically to trade off a general encoding and transmission cost with that of an real-time distortion loss, is studied. Formulated as a partially observable Markov decision problem, optimality equations are derived and the structural properties of the optimal encoder and decoder have been established. Via an example of joint source–channel coding over a binary input channel, the main ideas have been illustrated and some heuristics have been proposed.

The problem of dynamic joint source–channel coding is shown to be applicable well beyond the realm of communications and information transmission. In particular, the problem arises in a variety of information acquisition problems with sequentially adaptive strategies such as tracking a moving target in a noisy background.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 2006.

[2] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Variable-Length Coding with Feedback in the Non-Asymptotic Regime," in *IEEE International Symposium on Information Theory (ISIT)*, 2010, pp. 231–235.

[3] M. V. Burnashev, "Data Transmission Over a Discrete Channel with Feedback Random Transmission Time," *Problemy Peredachi Informatsii*, vol. 12, no. 4, pp. 10–30, 1975.

[4] J. Walrand and P. Varaiya, "Optimal Causal Coding-Decoding Problems," *IEEE Transactions on Information Theory*, vol. 29, November 1983.

[5] S. K. Mitter V. S. Borkar and S. K. Tatikonda, "Optimal sequential vector quantization of markov sources," *SIAM Journal of Control and Optimization*, 2001.

[6] M. Naghshvar and T. Javidi, "Active Sequential Hypothesis Testing," 2012, available on arXiv:1203.4626.

[7] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control*, Athena Scientific, Belmont, Massachusetts, 1996.

[8] L. K. Platzman, "Optimal infinite-horizon undiscounted control of finite probabilistic systems," *SIAM Journal of Control And Optimization*, vol. 18, July 1980.

[9] M. H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill, Inc., 1970.

[10] M. Naghshvar and T. Javidi, "Extrinsic Jensen–Shnnon Divergence with Application in Active Hypothesis Testing," in *IEEE International Symposium on Information Theory (ISIT)*, 2012.