Zero-Delay Coding over a Noisy Channel

I. Zero-Delay Lossy Coding: Noisy Channel Case

A. Optimal Quantizers

We will assume throughout that the source $\{X_t\}_{t\geq 0}$ is a discrete-time Markov process with probability matrix P, which is irreducible and aperiodic (and thus admits a unique invariant measure). After encoding, the (compressed) information is sent over a discrete memoryless *noisy* channel with input alphabet $\mathcal{M} := \{1, \dots, M\}$ and output alphabet $\mathcal{M}' := \{1, \dots, M'\}$. We assume the encoder has access to feedback from previous channel outputs.

Thus, the encoder is defined by an encoder policy $\{\gamma_t^e\}_{t\geq 0}$, where $\gamma_t^e : (\mathcal{M}')^t \times \mathbb{X}^{t+1} \to \mathcal{M}$. That is, the encoder can use all past channel outputs and all past and current source inputs to generate the current encoder output. This can be viewed as the encoder policy selecting a quantizer $Q_t : \mathbb{X} \to \mathcal{M}$ using past information, then quantizing X_t as $q_t = Q_t(X_t)$ [1]. Then q_t is sent over the channel, and the channel output q_t' is determined by the transition probability $T(q_t'|q_t)$. The decoder generates the reconstruction $\hat{X}_t$ without delay, using decoder policy $\{\gamma_t^d\}_{t\geq 0}$, where $\gamma_t^d : (\mathcal{M}')^{t+1} \to \hat{\mathbb{X}}$. Thus we have $\hat{X}_t = \gamma_t^d(q'_{[0,t]})$. We now present some noisy-channel analogs to the original noiseless case.

In a similar fashion to the noiseless case, one can restrict to optimal encoders and assume the optimal corresponding decoder is used. We then consider the discounted cost problem; for some $\beta \in (0,1)$, we wish to minimize:

$$J_\beta(\mu, \gamma) := \lim_{T\to\infty} \mathbf{E}_\mu^\gamma \left[\frac{1}{T} \sum_{t=0}^{T-1} \beta^t d(X_t, \hat{X}_t) \right]$$

For finite horizon problems, we have the following results on the structure of optimal zero-delay codes, the first from Witsenhausen and the second from Walrand and Varaiya (which were later generalized further in the literature, see e.g. [2]–[4]):

Theorem I.1. *[5] For the problem of coding a Markov source over a finite time horizon T, any zero delay encoder policy $\gamma = \{\gamma_t\}$ can be replaced, without loss in distortion performance, by a policy $\hat{\gamma} = \{\hat{\gamma}_t\}$ which only uses $q'_{[0,t-1]}$ and X_t to generate q_t, i.e., such that $q_t = \hat{\gamma}_t(q'_{[0,t-1]}, X_t)$ for all $t = 1, \dots, T-1$.*

We define the following conditional probability measure on $\mathbb{X}$. Let $\mathcal{P}(\mathbb{X})$ be the space of probability measures on $\mathbb{X}$, and define $\pi_t \in \mathcal{P}(\mathbb{X})$ as:

$$\pi_t(A) := \Pr(X_t \in A | q'_{[0,t-1]})$$

Theorem I.2. *[6] For the problem of coding a Markov source over a finite time horizon T, any zero delay encoder policy $\gamma = \{\gamma_t\}$ can be replaced, without loss in distortion performance, by a policy $\hat{\gamma} = \{\hat{\gamma}_t\}$ which only uses π_t and X_t to generate q_t, i.e., such that $q_t = \hat{\gamma}_t(\pi_t, X_t)$ for all $t = 1, \dots, T-1$. Alternatively, at time t such a policy uses π_t to select a quantizer $Q_t = \hat{\gamma}_t(\pi_t)$ (where $Q_t : \mathbb{X} \to \mathcal{M}$), and then q_t is obtained by $q_t = Q_t(X_t)$.*

Encoders with the above structure are called *Walrand-Varaiya type* policies in [1], [7], or alternatively, Markov policies (because they endow Markov properties onto the process $\{\pi_t\}$, which we will see momentarily). We also have an infinite-horizon analog of the above theorem, which was proven in [7]:

Theorem I.3. *[7, Theorem 7] For the problem of coding an irreducible and aperiodic Markov source, for any initial distribution μ, there exists a stationary Walrand-Varaiya type policy γ^* that solves the infinite-horizon discounted cost problem, i.e. one that satisfies:*

$$J_\beta(\mu, \gamma^*) = \inf_{\gamma\in\Gamma} J_\beta(\mu, \gamma)$$

where Γ is the set of all admissible encoder policies and $J_\beta(\mu, \gamma)$ is defined in (I-A)*.*

Under Walrand-Varaiya type policies, it was shown in [1] that $\{\pi_t\}$ is a controlled Markov process with control $\{Q_t\}$. More specifically, we have the following result:

Theorem I.4. *[1] Under a Walrand-Varaiya type policy, the update equation for π_t is given by*

$$\pi_{t+1}(x_{t+1}) = \frac{\sum_{q_t} T(q_t'|q_t) \sum_{x_t \in Q_t^{-1}(q_t)} P(x_{t+1}|x_t)\pi_t(x_t)}{\sum_{q_t} T(q_t'|q_t)\pi_t(Q_t^{-1}(q_t))} \tag{1}$$

Therefore π_{t+1} is conditionally independent of $(\pi_{[0,t-1]}, Q_{[0,t-1]})$ given π_t and Q_t, and hence $\{\pi_t\}$ is a controlled Markov process with control $\{Q_t\}$.

And we define the following cost function for this Markov decision process (MDP) in terms of π_t and Q_t (this is the average distortion if the optimal decoder is used for a given Q_t).

$$c(\pi_t, Q_t) := \sum_{i=1}^{M} \min_{\hat{x}\in\hat{\mathbb{X}}} \sum_{x\in Q_t^{-1}(i)} \pi_t(x) d(x, \hat{x}) \tag{2}$$

Note that by this definition of $c(\pi_t, Q_t)$ and our assumption that we are using an optimal decoder for a given encoder of

the optimal Walrand-Varaiya type, we have:

$$\mathbf{E}_{\mu}^{\gamma}\left[\frac{1}{T}\sum_{t=0}^{T-1}c(\pi_t,Q_t)\right]=\mathbf{E}_{\mu}^{\gamma}\left[\frac{1}{T}\sum_{t=0}^{T-1}d(X_t,\hat{X}_t)\right]$$

B. A Topology on Quantizers

This is defined exactly as in the noiseless case (i.e. $Q_n \to Q$ weakly iff $PQ_n \to PQ$ weakly). Under this topology, the results in [1] can be extended to show the following property of the controlled Markov chain $\{\pi_t\}$.

Lemma I.5. *[1, Lemma 11]. The transition kernel $P(d\pi_{t+1}|\pi_t, Q_t)$ is weakly continuous in (π_t, Q_t). That is,*

$$\int_{\mathcal{P}(\mathbb{X})\times\mathcal{Q}} f(\pi^{'})P(d\pi^{'}|\pi,Q)$$

is continuous on $\mathcal{P}(\mathbb{X})\times\mathcal{Q}$ for all continuous bounded f.

II. Q-learning and Quantized Q-learning

This section applies to a general MDP and is not specific to the noisy/noiseless case, so it has been omitted. We again note that the only non-trivial assumption from [8, Theorem 3.2] that we must prove is the following.

Assumption II.1. *Under a random exploration policy γ, the process $\{\pi_t\}_{t\geq 0}$ admits a unique invariant measure.*

Where a memoryless exploration policy γ is one in which

$$\Pr(\gamma(\cdot)=u_i)=p_i \quad \forall i=1,\ldots,|\mathbb{U}|$$

where $p_i > 0\ \forall i$ and $\sum_i p_i = 1$.

III. Unique Ergodicity Under a Memoryless Exploration Policy

Recall our setup from Section I. In particular, we have a controlled Markov process $\{\pi_t\}$, with control $\{Q_t\}$, where $Q_t : \mathbb{X} \to \mathcal{M}$. Here $\mathbb{X}$ is our source alphabet, $\mathcal{M}$ is our message set, we have $Q_t(X_t) = q_t$, and q_t' is obtained through $T(q_t'|q_t)$.

We wish to show that if we choose the Q_t randomly, $\{\pi_t\}$ admits a unique invariant measure. In order to apply the same POMDP methods we used in the noiseless case, we will note that the conditional probability of q_t' given x_t is determined by T and the distribution of Q_t. We denote the resulting measurement kernel by $S(q_t'|x_t) = \sum_{q\in\mathcal{M}} T(q_t'|q)R(\{Q \in \mathcal{S} : Q(x_t) = q\})$, where (as in the noiseless case) we define R as a distribution on the set of quantizers S according to our random exploration policy (that is, the distribution of Q_t is described entirely by R, which is positive everywhere).

As we did in the noiseless case, we have unique ergodicity of the process $\{\pi_t\}_{t\geq 0}$ if we can show that the *filter* $\pi_t^*(A) := \Pr(X_t \in A|q'_{[0,t]})$ is stable (in total variation in expectation). To this end, we impose the same assumption on our set of quantizers $\mathcal{S}$ as in the noiseless case:

Assumption III.1. *For all $x \in \mathbb{X}$ and for all $q \in \mathcal{M}$ we have $\{Q \in \mathcal{S} : Q(x) = q\} \neq \emptyset$.*

Then, as in the noiseless case, we have that $R(\{Q \in \mathcal{S} : Q(x) = q\})$ is positive . But then this implies that $S(q'|x)$ is positive everywhere (if not, we must have that $T(q'|q) = 0$ for all q, so q' is not a valid channel output).

Then, filter stability follows from [9, Corollary 5.5], and so $\{\pi_t\}_{t\geq 0}$ admits a unique invariant measure. Therefore all the assumptions for quantized Q-learning to converge are met. The slightly modified algorithms are below.

IV. Algorithms

A. Quantizing π_t

Since the state space $\mathbb{X}$ is finite, say with $|\mathbb{X}| = m$, then $\mathcal{P}(\mathbb{X})$ is a simplex in $\mathbb{R}^m$. For a given belief π_t and n, we wish to find the nearest (in terms of Eucilidean distance) $\hat{\pi}_t = [\frac{k_1}{n}, \ldots, \frac{k_m}{n}]$, where $k_i \in \mathbb{Z}$. Then we can use the algorithm in e.g. [10], [11] to quantize π_t as follows.

Algorithm 1: Predictor Quantization

Require: $n \geq 1, \pi_t = (p_1, \ldots, p_m)$
for $i = 1$ **to** m **do**
  $k_i' = \lfloor np_i + \frac{1}{2} \rfloor$
end for
$n' = \sum_i k_i'$
if $n = n'$ **then**
  return $(\frac{k_1'}{n}, \ldots, \frac{k_m'}{n})$
end if
for $i = 1$ **to** m **do**
  $\delta_i = k_i' - np_i$
end for
Sort δ_i s.t. $\delta_{i_1} \leq \ldots \leq \delta_{i_m}$
$\Delta = n' - n$
if $\Delta > 0$ **then**

$$k_{i_j} = \begin{cases} k'_{i_j} & j = 1, \ldots, m - \Delta \\ k'_{i_j} - 1 & j = m - \Delta + 1, \ldots, m \end{cases}$$

else

$$k_{i_j} = \begin{cases} k'_{i_j} + 1 & j = 1, \ldots, |\Delta| \\ k'_{i_j} & j = |\Delta| + 1, \ldots, m \end{cases}$$

end if
return $(\frac{k_1'}{n}, \ldots, \frac{k_m'}{n})$

We have the following lemma regarding the radius of these quantization bins under the above algorithm.

Lemma IV.1. *[10, Proposition 2] The maximum radius of the quantization regions for π_t under the L_∞ norm is given by*

$$b_\infty = \frac{1}{n}(1 - \frac{1}{m})$$

Also note that the number of bins for π_t when using **Algorithm 1** is related to n by the following relation: # bins $= \binom{n+m-1}{m-1}$ [10].

B. Quantized Q-learning

Using the above algorithm to quantize π_t, we have the following algorithm for quantized Q-learning.

Algorithm 2: Quantized Q-learning

Require: source alphabet $\mathbb{X}$, transition kernel $P(x_{t+1}|x_t)$, initial distribution π_0, quantization parameter n, quantizer set $\mathcal{Q}$, exploration policy γ, time horizon T

Initialize Q-table of size $\binom{n+m-1}{m-1} \times |\mathcal{Q}|$
Initialize x_0 according to π_0
Quantize π_0 using **Algorithm 1**, call this $\hat{\pi}_0$
Select quantizer Q_0 according to γ
$q_0 = Q_0(x_0)$
for $t = 0$ **to** $T - 1$ **do**
    Compute $c(\pi_t, Q_t)$ (see (2))
    Receive x_{t+1} according to $P(x_{t+1}|x_t)$
    Receive π_{t+1} according to update equation (see (1))
    Quantize π_{t+1} using **Algorithm 1**, call this $\hat{\pi}_{t+1}$
    Update Q-table
    Select quantizer Q_{t+1} according to γ
    $q_{t+1} = Q_{t+1}(x_{t+1})$
    Receive q'_t according to $T(q'_t|q_t)$
end for
return $\gamma^*(\pi) = \operatorname{argmin}_{Q \in \mathcal{Q}}(\text{Q-table}(\pi, Q))$

Theorem IV.2. *Under Assumption III.1 and as $n \to \infty$, the above algorithm gives a near-optimal policy for the zero-delay coding problem.*

Proof. As in the noiseless case, the proof follows from the fact that $\{\pi_t\}$ admits a unique invariant measure and from [8, Theorem 3.2]. □

V. Examples

Currently filling out this section with simulations. Unfortunately, they take a while longer than the noiseless case (presumably due to the extra complexity of computing π_t). Preliminary results look similar to the noiseless case, but the benefits of finer quantization are generally less obvious.

References

[1] T. Linder and S. Yüksel, "On optimal zero-delay coding of vector markov sources," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5975–5991, Oct. 2014.

[2] A. Mahajan and D. Teneketzis, "Optimal design of sequential real-time communication systems," *IEEE Transactions on Information Theory*, vol. 55, pp. 5317–5338, Nov. 2009.

[3] D. Teneketzis, "On the structure of optimal real-time encoders and decoders in noisy communication," *IEEE Transactions on Information Theory*, vol. 52, pp. 4017–4035, Sep. 2006.

[4] S. Yüksel, "On optimal causal coding of partially observed Markov sources in single and multi-terminal settings," *IEEE Transactions on Information Theory*, vol. 59, pp. 424–437, Jan. 2013.

[5] H. Witsenhausen, "On the structure of real-time source coders," *Bell System Technical Journal*, vol. 58, no. 6, pp. 1437–1451, 1979.

[6] J. Walrand and P. Varaiya, "Optimal causal coding-decoding problems," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 6, pp. 814–820, Nov. 1983.

[7] R. Wood, T. Linder, and S. Yüksel, "Optimal zero delay coding of markov sources: Stationary and finite memory codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5968–5980, Sep. 2017.

[8] A. Kara, N. Saldi, and S. Yüksel, "Q-learning for MDPs with general spaces: Convergence and near optimality via quantization under weak continuity," 2021. DOI: 10.48550/ARXIV.2111.06781.

[9] R. van Handel, "The stability of conditional markov processes and markov chains in random environments," *Ann. Probab.*, vol. 37, no. 5, pp. 1876–1925, Sep. 2009.

[10] Y. Reznik, "An algorithm for quantization of dicrete probability distributions," *DCC 2011*, pp. 333–342, Mar. 2011.

[11] N. Saldi, S. Yüksel, and T. Linder, "Asymptotic optimality of finite model approximations for partially observed markov decision processes with discounted cost," *IEEE Transactions on Automatic Control*, vol. 65, no. 1, pp. 130–142, 2020.