

Deterministic list codes for state-constrained arbitrarily varying channels

Anand D. Sarwate
Michael Gastpar



Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2007-6

<http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-6.html>

January 8, 2007

Copyright © 2007, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Acknowledgement

The work of A.D. Sarwate and M. Gastpar was supported in part by the National Science Foundation under award CCF-0347298.

Deterministic list codes for state-constrained¹ arbitrarily varying channels

Anand D. Sarwate *Student Member, IEEE*, and Michael Gastpar *Member, IEEE*

Abstract

The capacity for the discrete memoryless arbitrarily varying channel (AVC) with cost constraints on the jammer is studied using deterministic list codes under both the maximal and average probability of error criteria. For a cost function $l(\cdot)$ on the state set and constraint Λ on the jammer, the achievable rates are upper bounded by the random coding capacity $C_r(\Lambda)$. For maximal error, the rate $R = C_r(\Lambda) - \epsilon$ is achievable using list codes with list size $O(\epsilon^{-1})$. For average error, an integer $L_{\text{sym}}(\Lambda)$, called the *symmetrizability*, is defined. It is shown that any rate below $C_r(\Lambda)$ is achievable under average error using list codes of list size $L > L_{\text{sym}}$. An example is given for a class of discrete additive AVCs.

I. INTRODUCTION

The arbitrarily varying channel (AVC) is a model for communication subject to time-varying interference. This interference is modeled by a channel state parameter which can vary arbitrarily across time, and coding schemes for these channels are required to give a guarantee on the probability of error for all channel state sequences. In this sense the AVC can be thought of as adversarial model in which the channel state is controlled by a *jammer* who wishes to foil the communication between the encoder and decoder.

Because reliable communication over AVCs is for the worst case, the coding problem for deterministic codes with the maximal probability of error criterion is quite difficult. The best bounds on the capacity for deterministic coding under maximal error C_d are due to Csiszár and Körner [1]. By relaxing the error criterion or expanding the coding strategies, results become easier to prove and in many cases coincide. The first results on AVCs were proved for the case where the encoder and decoder share a source of common randomness that they can use to perform *randomized coding* [2]. The largest rate for which the probability of decoding error can be driven to 0 for any message and any state sequence is called the random coding capacity under maximal error, and is denoted by C_r . By requiring deterministic coding but defining the error probability to be the averaged over all messages, Ahlswede [3] proved the following dichotomy: the deterministic coding capacity for average error \bar{C}_d is either 0 or $\bar{C}_d = C_r$. A condition for \bar{C}_d to be positive, called *symmetrizability*, was shown to be necessary by Ericson [4] and sufficient by Csiszár and Narayan [5].

Another way of relaxing the coding problem is to consider *list coding* [6], in which the decoder is allowed to output a list of codewords of size no more than L and an error is counted only if the transmitted codeword is not in the list. For deterministic coding, list coding capacities C_L and \bar{C}_L for AVCs have been investigated for both maximal and average error. For maximal error, Ahlswede [7], [8] found that for any rate $R < C_r$ there exists a constant list size $L(R)$ such that R is achievable with list codes of

Manuscript received December XX, 2006; revised XXXXXXXXXXXXXXXX.

The authors are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley CA 94720-1770 USA.

The work of A.D. Sarwate and M. Gastpar was supported in part by the National Science Foundation under award CCF-0347298.

list size $L(R)$. The average error case was solved independently by Blinovsky, Narayan, and Pinsker [9], [10] and Hughes [11]. They proposed an extended notion of symmetrizability and showed that for non-symmetrizable channels there is a constant list size L so that all rates $R < C_r$ are achievable with list codes of list size L .

Another modification to the AVC scenario is to impose a cost $l(\cdot)$ on the possible states and to constrain the jammer to be arbitrary subject to a total budget Λn . For example, the inputs and outputs may be binary and the jammer can flip a fraction Λ of the bits. The random coding capacity for maximal error $C_r(\Lambda)$ and the deterministic coding capacity for average error $\bar{C}_d(\Lambda)$ were found by Csiszár and Narayan [5], [12]. The deterministic coding capacity for maximal error is intimately connected to the problem of error correcting codes, as the bit-flipping example shows.

The purpose of this paper is to extend the results on list decoding to the case of constrained AVCs. Our primary results are to show that some of the results in [8], [10], [11] also hold for the case of cost-constrained AVCs:

- For maximal error, for any rate $R = C_r(\Lambda) - \epsilon$ there exists list codes of size $L(\epsilon) = O(\epsilon^{-1})$ that achieve the rate R .
- For average error there exists a number $L_{\text{sym}}(\Lambda)$, called the *symmetrizability* of the channel, such that the list coding capacity is C_r for lists larger than $L_{\text{sym}}(\Lambda)$.

The proofs of these results parallel those for the unconstrained case. We follow the same arguments as Ahlswede [8] for the maximal error case and Hughes [11] for the average error case.

In the next section we will describe the channel model in more detail. In Section III we prove the maximal error result and in Section IV the average error result. We will close with some examples and their implications for other problems of interest.

II. CHANNEL MODELS AND MAIN RESULTS

An arbitrarily varying channel (AVC) is a collection of $\mathcal{W} = \{W(\cdot|\cdot, s) : s \in \mathcal{S}\}$ of channels from an input alphabet \mathcal{X} to an output alphabet \mathcal{Y} parameterized by a state $s \in \mathcal{S}$. Here we will assume the sets \mathcal{X} , \mathcal{Y} and \mathcal{S} are finite. If $\mathbf{x} = (x_1, x_2, \dots, x_n)$, $\mathbf{y} = (y_1, y_2, \dots, y_n)$ and $\mathbf{s} = (s_1, s_2, \dots, s_n)$ are length n vectors, the probability of \mathbf{y} given \mathbf{x} and \mathbf{s} is given by:

$$W(\mathbf{y}|\mathbf{x}, \mathbf{s}) = \prod_{i=1}^n W(y_i|x_i, s_i). \quad (1)$$

We think of the state as being controlled by a malicious adversary, called the *jammer*, whose objective is to maximize the probability of decoder error and thereby minimize the capacity.

We are interested in the case where there is a cost function $l : \mathcal{S} \rightarrow \mathbb{R}^+$ on the jammer. We will assume $\max_s l(s) \leq l_{\text{max}} < \infty$. The cost of an n -tuple is

$$l(\mathbf{s}) = \sum_{k=1}^n l(s_k) \quad (2)$$

The state obeys a state constraint Λ if

$$l(\mathbf{s}) \leq n\Lambda \quad a.s. \quad (3)$$

Note that if $\Lambda = l_{\text{max}}$ then the state constraint is inoperative and we return to the unconstrained case.

As an example, take the AVC given by $y = x \oplus s$ with $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, 1\}$ and $l(s) = s$. This is similar to a binary symmetric channel (BSC) with flip probability Λ . In a BSC the channel will flip close to Λn bits *with high probability* and the decoder error must be small for *most error patterns*. For

this AVC, however, the channel is constrained to flip *no more than* Λn bits but the decoder error must be small for *any error pattern* of weight Λn .

An (n, N, L) *deterministic list code* C for the AVC is a pair of maps (ψ, ϕ) where the encoding function is $\psi : \{1, 2, \dots, N\} \rightarrow \mathcal{X}^n$ and the decoding function is $\phi : \mathcal{Y}^n \rightarrow \{1, 2, \dots, N\}^L$. The *rate* of the code is $R = \log(N/L)$. The *codebook* is the set of vectors $\{\mathbf{x}_i : 1 \leq i \leq N\}$, where $\mathbf{x}_i = \psi(i)$. The decoding region for message i is $D_i = \{\mathbf{y} : i \in \phi(\mathbf{y})\}$. We will often specify a code by the pairs $\{(\mathbf{x}_i, D_i) : i = 1, 2, \dots, N\}$, with the encoder and decoder implicitly defined.

We have two notions of error probability, *maximal* and *average*. The maximal error is given by:

$$\varepsilon(C, \cdot) = \max_{\mathbf{s}: l(\mathbf{s}) \leq \Lambda} \max_i \mathbb{P}(i \notin \phi(\mathbf{y}) | \mathbf{x}_i, \mathbf{s}) = \max_{\mathbf{s}: l(\mathbf{s}) \leq \Lambda} \max_i W(D_i^c | X^n = \mathbf{x}_i, \mathbf{s}) . \quad (4)$$

The average probability of error is given by

$$\bar{\varepsilon}(C, \cdot) = \max_{\mathbf{s}: l(\mathbf{s}) \leq \Lambda} \frac{1}{N} \sum_{i=1}^N W(D_i^c | \mathbf{x}_i, \mathbf{s}) . \quad (5)$$

A rate R is called *achievable* under maximal (average) error with list size L if there exists of sequence of (n, N, L) codes of rate greater than R whose maximal (average) error converges to 0 as $n \rightarrow \infty$. The supremum of all achievable rates is the list- L capacity. We will denote the list- L capacity for maximal error by $C_L(\Lambda)$ and for average error by $\bar{C}_L(\Lambda)$.

In some cases the capacity $\bar{C}_L(\Lambda) = 0$. A necessary and sufficient condition for this is if the AVC is L -symmetrizable. We say an AVC is m -symmetrizable if there is channel $U : \mathcal{X}^m \rightarrow \mathcal{S}$ such that

$$V(y|x, x_1, \dots, x_m) = \sum_s W(y|x, s) U(s|x_1, x_2, \dots, x_m) \quad (6)$$

is a symmetric function in (x, x_1, \dots, x_m) . For a distribution $P \in \mathcal{P}(\mathcal{X})$ let us define the quantity

$$\Lambda_m(P) = \min_{U: \mathcal{X}^m \rightarrow \mathcal{S}} \sum_{x^m} \sum_s P(x^m) U(s|x^m) l(s) . \quad (7)$$

We define the *symmetrizability* $L_{\text{sym}}(P, \Lambda)$ of the channel to be the largest integer such that

$$\Lambda_{L_{\text{sym}}(P, \Lambda)}(P) \leq \Lambda \quad (8)$$

$$\Lambda_{L_{\text{sym}}(P, \Lambda)+1}(P) > \Lambda \quad (9)$$

Intuitively, the jammer can “simulate $L_{\text{sym}}(P, \Lambda)$ codewords of type P ” within its cost constraint Λ , but simulating $L_{\text{sym}}(P, \Lambda) + 1$ codewords will make it exceed the constraint. When it is clear from context we will suppress the dependence on Λ and write $L_{\text{sym}}(P)$ for $L_{\text{sym}}(P, \Lambda)$.

A. Main Results

Our main results on list coding are analogous to those in [8], [10], [11] for unconstrained discrete AVCs. The primary contribution of this work is an extension of these results to the constrained AVC setting. The goal is to show that list coding using finite lists is sufficient to achieve the *random coding capacity* of the AVC:

$$C_r(\Lambda) = \max_{P(x)} \min_{Q(s): \mathbb{E}_Q[l(s)] \leq \Lambda} I(X \wedge Y) . \quad (10)$$

Here the minimization is over all distributions $Q(s)$ on \mathcal{S} whose expected cost no more than Λ , and the maximization is over all distributions on \mathcal{X} .

For maximal error, we prove that a rate $R = C_r - \delta$ is achievable using list codes of list size $L > \delta^{-1} \log |\mathcal{Y}| + 1$. Here the size of the list grows as we approach capacity, but does not scale with the blocklength. To prove this result, we first construct a code with nearly $2^{nH(X)}$ codewords that is list-decodable with lists of size $2^{n \max_Q H(X|Y)}$. We then subsample this list code to obtain a code with finite list size and rate $C_r - \delta$. It is not clear that this list size is necessary – matching converse bounds are not shown.

For average error, we show that any rate below C_r is achievable using lists of size larger than $L_{\text{sym}}(P^*)$ where P^* is the maximizing input distribution in (10). Hughes [11] proved that for the unconstrained AVC there is a single L_{sym} that acts as a threshold for list decoding, so $\bar{C}_L = 0$ for $L \leq L_{\text{sym}}$ and $\bar{C}_L = C_r$ for $L > L_{\text{sym}}$. For constrained jammers we do not have such a dichotomy for list coding. Since each input distribution P has a corresponding symmetrizability $L_{\text{sym}}(P)$, we may have the case that $\Lambda_{L_{\text{sym}}(P^*)}(P) > \Lambda$, in which case we can achieve

$$I(P) = \min_{Q(s): \mathbb{E}_Q[l(s)] \leq \Lambda} I(X \wedge Y) , \quad (11)$$

with lists of size smaller than $L_{\text{sym}}(P^*)$. We will define $L_{\text{sym}}(\Lambda) = L_{\text{sym}}(P^*, \Lambda)$ to be the *symmetrizability of the channel*. Csiszár and Narayan observed in [5] that the deterministic coding capacity for constrained AVCs may be positive but smaller than the corresponding random code capacity. The analogous statement for list codes is that the capacity for list codes with list size smaller than $L_{\text{sym}}(\Lambda)$ may be positive but smaller than the random coding capacity.

The proof closely follows that of Hughes with some modifications to deal with the state constraint. We note that the proof by Blinovsky, Narayan, and Pinsker [10] relies on the elimination technique [3], which does not hold for cost-constrained AVCs [5].

B. Notation conventions

The proofs of AVC results tend to generate a surfeit of notation. Let $[N] = \{1, 2, \dots, N\}$ and $\Sigma_L = \{J \subset [N] : |J| = L\}$. Let $\Sigma_L(-\{i\}) = \{J \in \Sigma_L : i \notin J\}$. We will use boldface symbols to denote tuples, so $\mathbf{z} = (z_1, z_2, \dots, z_n)$. We will denote sets by calligraphic letters, such as \mathcal{Z} , and the set of all distributions on a set \mathcal{Z} by $\mathcal{P}(\mathcal{Z})$. For $P \in \mathcal{P}(\mathcal{Z})$ we write $H(P)$ for the entropy under distribution P . Given a sequence $\mathbf{z} \in \mathcal{Z}^n$, let $N(\zeta|\mathbf{z}) = |\{i : z_i = \zeta\}|$, the number of times ζ appears in \mathbf{z} . We denote the type of \mathbf{z} by

$$T_{\mathbf{z}} = \frac{1}{n} (N(\zeta_1|\mathbf{z}), N(\zeta_2|\mathbf{z}), \dots, N(\zeta_{|\mathcal{Z}|}|\mathbf{z})) . \quad (12)$$

The set of all sequences of a fixed type P will be denoted by

$$\mathcal{T}(P) = \{\mathbf{z} \in \mathcal{Z}^n : T_{\mathbf{z}} = P\} . \quad (13)$$

The set of all types of sequences of length n will be denoted by $\mathcal{P}_n(\mathcal{Z}) \subset \mathcal{P}(\mathcal{Z})$. We denote the maximum variation between two distributions P and Q by

$$d_m(P, Q) = \max_{\zeta \in \mathcal{Z}} |P(\zeta) - Q(\zeta)| . \quad (14)$$

For a distribution $P \in \mathcal{P}(\mathcal{Z})$, the set

$$T_P^\epsilon = \{\mathbf{z} \in \mathcal{Z}^n : d_m(P, T_{\mathbf{z}}) \leq \epsilon\} \quad (15)$$

is the δ -(strongly) typical set.

Let $V(Y|Z)$ be a channel. Then we denote the (V, ϵ) -shell of \mathbf{z} by

$$T_V^\epsilon(\mathbf{z}) = \{\mathbf{y} \in \mathcal{Y}^n : d_m(T_{\mathbf{y}\mathbf{z}}, WT_{\mathbf{z}}) < \epsilon\} . \quad (16)$$

We have

$$V^n(\{\mathbf{y} : T_{\mathbf{y}\mathbf{z}} = P_{YZ}\}|\mathbf{x}) \leq \exp(-nD(P_{XZ} \parallel V \times P_Z)) . \quad (17)$$

$$V^n(\{\mathbf{y} : T_{\mathbf{x}\mathbf{y}\mathbf{z}} = P_{XYZ}\}|\mathbf{x}) \leq \exp(-nI(Y \wedge X|Z)) . \quad (18)$$

We will denote the set of all \mathbf{s} that satisfy the cost constraint by $\mathcal{S}^n(\Lambda)$:

$$\mathcal{S}^n(\Lambda) = \{\mathbf{s} : l(\mathbf{s}) \leq \Lambda\} . \quad (19)$$

We will also define the set of all distributions on \mathcal{S} satisfying the constraint.

$$\mathcal{P}(\mathcal{S}, \Lambda) = \left\{ Q \in \mathcal{P}(\mathcal{S}) : \sum_s Q(s)l(s) \leq \Lambda \right\} . \quad (20)$$

For an AVC $\mathcal{W} = \{W(Y|X, S) : s \in \mathcal{S}\}$ with state constraint Λ we can define a set of channels $\bar{\mathcal{W}}(\Lambda)$ by

$$\bar{\mathcal{W}}(\Lambda) = \left\{ V(Y|X) : V(y|x) = \sum_s W(y|x, s)Q(s), \quad Q(s) \in \mathcal{P}(\mathcal{S}, \Lambda) \right\} . \quad (21)$$

We will suppress the explicit dependence on Λ .

III. LIST DECODING FOR MAXIMAL ERROR

The arbitrarily varying channel with deterministic codes and maximal error is directly related to the design of error correcting codes. In the bit-flipping example given in the previous section, a list code of list size 1 under maximal error must correct every error pattern of weight Λn for every codeword. It is sometimes easier to prove the existence of list codes with small but constant list size that perform well under maximal error. For unconstrained AVCs, the list coding capacity was investigated by Ahlswede [8], [13] using hypergraph coloring arguments [14], [15]. In this section we generalize his result to the case of a constrained jammer but without using the hypergraph terminology.

Theorem 1 (List decoding for maximal error): Let $\mathcal{W} = \{W(\cdot|\cdot, s) : s \in \mathcal{S}\}$ be an arbitrarily varying channel with constraint function $l(s)$ and state constraint Λ . Fix a rate $R < C_r(\Lambda)$. Then R is achievable under maximal error with deterministic list codes of list size

$$L = O\left(\frac{1}{C_r(\Lambda) - R}\right) . \quad (22)$$

In other words,

$$C_d(L, \Lambda) \geq C_r(\Lambda) - O(L^{-1}) . \quad (23)$$

The result is proved in two steps – first we show that list codes of exponential list size exist, and then we construct a code of finite list size by sampling codewords from the larger list code. We first need some results on types and jointly typical sets.

A. Preliminaries

We need to define some more sets. First we define an AVC-version of a (V, ϵ) shell:

$$T_{\bar{W}, \mathbf{s}}^\epsilon(\mathbf{x}) = \left\{ \mathbf{y} \in \mathcal{Y}^n : \max_{a,b} \left| \frac{1}{n} N(a, b | \mathbf{x}, \mathbf{y}) - \frac{1}{n} \sum_{k: x_k = a} W(y|a, s_k) \right| < \epsilon n \right\} \quad (24)$$

We can take the union over all \mathbf{s} satisfying the state constraint Λ to get the set of all \mathbf{y} sequences that could have been generated from \mathbf{x} and some permissible state \mathbf{s} :

$$T_{\bar{W}}^\epsilon(\mathbf{x}) = \bigcup_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} T_{\bar{W}, \mathbf{s}}^\epsilon(\mathbf{x}) \quad (25)$$

The content of part 4 of Lemma 1 is to relate the $\mathbf{y} \in T_{\bar{W}}^\epsilon(\mathbf{x})$ with the $\mathbf{y} \in T_V^\delta(\mathbf{x})$ for some $V \in \bar{\mathcal{W}}$.

A channel $V(Y|X)$ and distribution $P \in \mathcal{P}(X)$ induces a distribution $P'(Y) \in \mathcal{P}(Y)$ and a "reverse channel" $V'(X|Y)$ related by

$$V(Y|X)P(X) = V'(X|Y)P'(Y) \quad (26)$$

This then defines a $T_{V'}^\epsilon(\mathbf{y})$, the (V', ϵ) -shell for V' for a fixed \mathbf{y} . We define the set of all channels in $\bar{\mathcal{W}}$ that are "consistent" with $T_{\mathbf{y}}$ and P by

$$\mathcal{W}^*(\delta) = \{V \in \bar{\mathcal{W}} : d_m(P', T_{\mathbf{y}}) < \delta\}. \quad (27)$$

Lemma 1: For $P \in \mathcal{P}(X)$ with $\min_a P(a) > 0$ and $0 < \epsilon < \min_a P(a)$ and n sufficiently large the following statements all hold:

- 1) We can bound the size of the typical set by

$$\exp(n(H(P) - o(\epsilon))) \leq |T_P^\epsilon| \leq \exp(n(H(P) + o(\epsilon))) \quad (28)$$

- 2) For $\mathbf{x} \in \mathcal{X}^n$ and \mathbf{s} such that $l(\mathbf{s}) \leq \Lambda$ we have for some $E(\epsilon) > 0$:

$$\mathbb{P}(T_{\bar{W}, \mathbf{s}}^\epsilon(\mathbf{x}) | \mathbf{x}, \mathbf{s}) \geq 1 - \exp(-nE(\epsilon)) \quad (29)$$

- 3) For $P \in \mathcal{P}(X)$, channel $V \in \bar{\mathcal{W}}$, $\mathbf{y} \in \mathcal{Y}^n$, and $\epsilon > 0$

$$|T_V^\epsilon(\mathbf{y})| \leq \exp\left(n \left(\sum_y H(V'(x|y)) T_{\mathbf{y}}(y) + o(\epsilon) \right)\right) \quad (30)$$

$$|T_{\bar{W}^*}^\epsilon(\mathbf{y})| \leq \exp\left(n \left(\max_{V \in \bar{\mathcal{W}}} \sum_y H(V'(x|y)) P'(y) + o(\epsilon) \right)\right) \quad (31)$$

- 4) There exist constants c_1, c_2, c_3 , and c_4 so that for sufficiently small $\epsilon > 0$, $\mathbf{x} \in T_P^\epsilon$, and $\mathbf{y} \in T_{\bar{W}}^\epsilon(\mathbf{x})$ we have

$$\mathbf{x} \in \bigcup_{V \in \mathcal{W}^*(c_1 \epsilon)} T_{V'}^{c_2 \epsilon}(\mathbf{y}) \quad (32)$$

$$\mathbf{y} \in \bigcup_{V \in \mathcal{W}^*(c_3 \epsilon)} T_V^{c_4 \epsilon}(\mathbf{x}) \quad (33)$$

Proof: We take up the different items in turn.

- 1) This can be found in [16, Lemma 2.3].
- 2) Fix sequences \mathbf{x} and \mathbf{s} and let $\{Y_i : i \in [n]\}$ be independent random variables with distribution $\{W(\cdot|x_i, s_i)\}$. Let $g_{(a,b)}(Y_1, \dots, Y_n) = N(a, b|\mathbf{x}, Y_1^n)$. Then

$$\mathbb{E}[g_{(a,b)}(Y_1, \dots, Y_n)] = \sum_{k:x_k=a} W(\cdot|a, s_k)$$

If $\{\tilde{Y}_i\}$ are independent copies of $\{Y_i : i = 1, \dots, n\}$, then we have

$$\left| g_{(a,b)}(Y_1, \dots, Y_i, \dots, Y_n) - g_{(a,b)}(Y_1, \dots, \tilde{Y}_i, \dots, Y_n) \right| \leq 1$$

almost surely, so by standard concentration inequalities [17, Corollary 2.4.14], for any $\epsilon > 0$

$$\mathbb{P} \left(|g_{(a,b)}(Y_1^n) - \mathbb{E}[g_{(a,b)}(Y_1^n)]| \geq \epsilon \right) \leq \exp \left(-nD \left(\frac{1+\epsilon}{2} \parallel \frac{1}{2} \right) \right)$$

Taking a union bound over all $(a, b) \in \mathcal{X} \times \mathcal{Y}$ shows that there exists an $E(\epsilon) > 0$ so that for n sufficiently large

$$\mathbb{P} \left(T_{W, \mathbf{s}}^\epsilon(\mathbf{x})^c | \mathbf{x}, \mathbf{s} \right) \leq \exp(-nE(\epsilon)) .$$

- 3) Equation (30) is in [16, Lemma 2.13]. For a fixed channel $V \in \bar{\mathcal{W}}(\Lambda) \cap \mathcal{P}_n(\mathcal{Y}|\mathcal{X})$, we have (30). Since $|\mathcal{P}_n(\mathcal{Y}|\mathcal{X})| < (n+1)^{|\mathcal{X}||\mathcal{Y}|}$, taking a union bound over (30) gives (31).
- 4) Fix $\epsilon > 0$. Since $\mathbf{y} \in T_{\mathcal{W}}^\epsilon(\mathbf{x})$, from (25) we know there exists an \mathbf{s} such that $\mathbf{y} \in T_{W, \mathbf{s}}^\epsilon(\mathbf{x})$. We can define a channel

$$V(b|a) = \frac{1}{N(a|\mathbf{x})} \sum_{k:x_k=a} W(y|a, s_k) ,$$

so that

$$\max_{a,b} \left| N(a, b|\mathbf{x}, \mathbf{y}) - \frac{nP(a)}{N(a|\mathbf{x})} \sum_{k:x_k=a} W(y|a, s_k) \right| = \max_{a,b} |N(a, b|\mathbf{x}, \mathbf{y}) - P(a)V(b|a)| . \quad (34)$$

Since $\mathbf{x} \in T_P^\epsilon$ we know

$$\left| 1 - \frac{nP(a)}{N(a|\mathbf{x})} \right| \leq \frac{n\epsilon}{N(a|\mathbf{x})}$$

Thus

$$\max_{a,b} \left| N(a, b|\mathbf{x}, \mathbf{y}) - \sum_{k:x_k=a} W(y|a, s_k) \right| \leq \epsilon + \frac{n\epsilon}{N(a|\mathbf{x})} \leq c_3\epsilon$$

for large n as long as $\min P(a) > 0$. Thus

$$\mathbf{y} \in T_{W, \mathbf{s}}^\epsilon(\mathbf{x}) \subset T_V^{c_3\epsilon}(\mathbf{x}) .$$

We must now show that $V \in \mathcal{W}^*(c_1\epsilon)$ for some c_1 . If we marginalize (34) over a we obtain for $\mathbf{y} \in T_V^{c_3\epsilon}(\mathbf{x})$.

$$\max_b |N(b|\mathbf{y}) - P'(b)| < c_3\epsilon|\mathcal{X}|n . \quad (35)$$

So for $c_1 > c_3|\mathcal{X}|$ we have $d_m(T_{\mathbf{y}}, P') < c_1\epsilon$, so by (27) we have proved (33). To show (32), let $V \in \bar{\mathcal{W}}$ and suppose $\mathbf{y} \in T_V^\epsilon(\mathbf{x})$. Then we have

$$d_m(T_{\mathbf{x}, \mathbf{y}}, PV) < \epsilon$$

Marginalizing over the distribution on X gives:

$$d_m(T_{\mathbf{y}}, P') < |\mathcal{X}|\epsilon$$

Therefore we have

$$d_m(T_{\mathbf{x}, \mathbf{y}}, T_{\mathbf{y}}V') < (|\mathcal{X}| + 1)\epsilon$$

Let $c_2 = (|\mathcal{X}| + 1)$. Thus $\mathbf{x} \in T_{V'}^{c_2\epsilon}(\mathbf{y})$ and $V \in \mathcal{W}^*(c_1\epsilon)$, proving (33). ■

B. List codes with large lists

Lemma 2: Let $(\mathcal{W}, l(\cdot), \Lambda)$ be a constrained AVC. For any $\epsilon > 0$ there is an n sufficiently large there is an (n, N, L, δ) list code with

$$N \geq \exp(n(H(P(x)) - \epsilon)) \quad (36)$$

$$L \leq \exp\left(n\left(\max_{V \in \mathcal{W}(\Lambda)} H(V'(x|y)|P'(y)) + \epsilon\right)\right) \quad (37)$$

$$\delta \leq \exp(-nE(\epsilon)) \quad (38)$$

Proof: Let the codewords of the code be $\mathbf{x} \in T_P^\epsilon$. For each channel output \mathbf{y} let the decoder output the list $T_{\mathcal{W}^*(c_1\epsilon)}^{c_3\epsilon}(\mathbf{y})$.

Equation (36) follows from Lemma 1 part 1. The bound (37) on the list size follows from Lemma 1 part 3. To bound the error in (38) note that with probability $\exp(-nE(\epsilon))$ we have $\mathbf{y} \in T_{\mathcal{W}}^\epsilon(\mathbf{x})$ (by Lemma 1 part 2) and hence by Lemma 1 part 4 we have $\mathbf{x} \in T_{V'}^{c_1\epsilon}(\mathbf{y})$ for some $V' \in \mathcal{W}^*(c_2\epsilon)$. ■

C. List reduction

We can now prove a generalization of Ahlswede's result [8] to the constrained AVC. Given a small gap ϵ from capacity, we subsample the previous code with exponential list sizes to obtain a code with finite list size $O(\epsilon^{-1})$ that can achieve rates ϵ away from capacity.

Lemma 3: Let $(\mathcal{W}, l(\cdot), \Lambda)$ be a constrained AVC whose randomized coding capacity is $C_r(\Lambda)$. For any $\epsilon' > 0$ there exists a list code of rate $R > C_r(\Lambda) - \epsilon'$ and list size

$$L' < \left\lceil \frac{\log |\mathcal{Y}|}{C_r(\Lambda) - R} \right\rceil + 1. \quad (39)$$

Proof: By Lemma 2 there exists N, L , and δ satisfying (36) – (38) for any ϵ so that there exists an (n, N, l, δ) list code $\mathcal{C}_L = \{(\mathbf{u}_i, D_i) : i \in [N]\}$. Note that $N/L = \exp(n(C_r - 2\epsilon))$. We will subsample this codebook to find our code of constant list size.

Let $N' = \exp(nR)$ and $\mathcal{C}_{L'} = \{(\mathbf{x}_j, D_j) : j \in [N']\}$ be a collection of N' codewords selected uniformly from \mathcal{C}_L . We will prove that there exists a constant L' so that no $\mathbf{y} \in \mathcal{Y}^n$ is in more than L' decoding sets D_j with high probability. Fix $\mathbf{y} \in \mathcal{Y}^n$ and note that from the definition of \mathcal{C}_L we have

$$\mathbb{E}(\mathbf{1}(\mathbf{y} \in D_j)) = \mathbb{P}(\mathbf{y} \in D_j) \leq \frac{L}{N} \quad (40)$$

For a fixed \mathbf{y} , the chance that more than L' decoding regions contain \mathbf{y} out of N' choices can be bounded above using Sanov's Theorem [18, Theorem 12.4.1]. That is, for any $\lambda > 0$ we can choose n large enough so that

$$\mathbb{P}\left(\frac{1}{N'} \sum_{j=1}^{N'} \mathbf{1}(\mathbf{y} \in D_j) > \frac{L'}{N'}\right) \leq \exp\left(-N' \left(D\left(\frac{L'}{N'} \parallel \frac{L}{N}\right) - \lambda\right)\right) \quad (41)$$

$$= \exp\left(-N' \left(\frac{L'}{N'} \log \frac{L'/N'}{L/N} + \left(1 - \frac{L'}{N'}\right) \log \frac{1 - L'/N'}{1 - L/N} - \lambda\right)\right) \quad (42)$$

To deal with the second term we use the inequality $-(1-a)\log(1-a) \leq 2a$ (for small a) on the term $(1 - L'/N')\log(1 - L'/N')$ and discard the small positive term $-(1 - L'/N')\log(1 - L/N)$. Choosing $\lambda = L'/N'$ we can have

$$\mathbb{P}\left(\frac{1}{N'} \sum_{j=1}^{N'} \mathbf{1}(\mathbf{y} \in D_j) > \frac{L'}{N'}\right) \leq \exp\left(-L' \log L' + L' \log 2^{n(C_r - R - 2\epsilon)} + 3L'\right) \quad (43)$$

$$= \exp\left(-n \left(L'(C_r - R - 2\epsilon) - \frac{L'}{n}(\log L' - 3)\right)\right) \quad (44)$$

Now we take a union bound over all \mathbf{y} to get

$$\mathbb{P}\left(\frac{1}{N'} \sum_{j=1}^{N'} \mathbf{1}(\mathbf{y} \in D_j) > \frac{L'}{N'} \quad \forall \mathbf{y}\right) \leq \exp\left(-n \left(L'(C_r - R - 2\epsilon) - \log |\mathcal{Y}| - \frac{L'}{n}(\log L' - 3)\right)\right) \quad (45)$$

So for n sufficiently large, if we choose ϵ sufficiently small we can have

$$L' < \left\lceil \frac{\log |\mathcal{Y}|}{C_r - R} \right\rceil + 1. \quad (46)$$

■

Theorem 1 now follows from the preceding Lemma.

IV. LIST DECODING FOR AVERAGE ERROR

In the case where we simultaneously allow list codes measure performance by the average error over the codebook, we can also achieve all rates below the randomized coding capacity with finite list sizes. We need to first relate the symmetrizability $L_{\text{sym}}(P)$ to the rate $I(P)$ defined in (11). The following theorem shows that if $I(P)$ is positive, then $L_{\text{sym}}(P)$ is finite.

Theorem 2 (Finite symmetrizability): If $C_r(\Lambda) = 0$ then $L_{\text{sym}}(P) = \infty$ for all P . If $C_r(\Lambda) > 0$ then

$$L_{\text{sym}}(P) \leq \frac{\log(\min(|\mathcal{Y}|, |\mathcal{S}|))}{I(P)}. \quad (47)$$

Proof: Suppose $C_r(\Lambda) = 0$. Then for all P we know $I(P) = 0$. Without loss of generality, we may take $P(x) > 0$ for all $x \in \mathcal{X}$. For such P there exists distribution $Q(s) \in \mathcal{P}(\mathcal{S}, \Lambda)$ so that

$$\sum_s W(y|x, s)Q(s) = P_Y(y) .$$

That is, the input and output are independent under $Q(s)$. If we define $U(s|x^L) = Q(s)$ it is clear that

$$V(x, x^L) = \sum_s W(y|x, s)U(s|x^L)$$

is symmetric in (x, x_1, \dots, x_L) and that

$$\Lambda_L(P) = \sum_{s, x^L} P(x^L)U(s|x^L)l(s) = \sum_s Q(s)l(s) \leq \Lambda$$

Since this holds for all L , the channel is L -symmetrizable for all L and thus $L_{\text{sym}}(P) = \infty$.

Suppose now that $C_r(\Lambda) > 0$. Let P be an input distribution for which $I(P) > 0$. Suppose that under P the channel is L -symmetrizable. Therefore there is a channel $U(S|X^L)$ that symmetrizes W . Let X_1, X_2, \dots, X_L be independent with distribution P and let (S, X^L) have distribution $U(S|X^L)P(X_1) \cdots P(X_L)$. Then $(X, X^L) \rightarrow (X, S) \rightarrow Y$ is a Markov chain, so by the Data Processing inequality we have

$$\begin{aligned} I(XS \wedge Y) &\geq I(XX^L \wedge Y) \\ &\geq I(X \wedge Y) + \sum_{j=1}^L I(X_j \wedge Y) \\ &= (L+1)I(X \wedge Y) \\ I(S \wedge Y|X) &\geq L \cdot I(X \wedge Y) \\ L &\leq \frac{I(S \wedge Y|X)}{I(X \wedge Y)} \\ &\leq \frac{\log(\min(|\mathcal{Y}|, |\mathcal{S}|))}{I(P)} \end{aligned}$$

This same bound holds for $L_{\text{sym}}(P)$. ■

Since we do not have a matching lower bound on $L_{\text{sym}}(P)$, it is not immediately clear that the P^* maximizing $I(P)$ will have the smallest symmetrizability. This implies that rates below $C_r(\Lambda)$ may be achievable using list codes with lists smaller than $L_{\text{sym}}(P^*)$. An example of this is given in Section V. A similar issue arises in the non-list coding case, which can be thought of as list coding with list size 1. There, the capacity may be positive but strictly less than the random coding capacity. However, for lists larger than $L_{\text{sym}}(P^*)$ the capacity of the constrained AVC for deterministic coding and average error is equal to the random coding capacity, as shown in the following theorem.

Theorem 3 (List decoding for average error): Let $\mathcal{W} = \{W(\cdot|\cdot, s) : s \in \mathcal{S}\}$ be an arbitrarily varying channel with constraint function $l(s)$ and state constraint Λ . Let $L_{\text{sym}} = L_{\text{sym}}(P^*)$ be the symmetrizability of the AVC for the input distribution P^* that maximizes $I(P)$. Then the deterministic-coding capacity of \mathcal{W} for list size $M > L_{\text{sym}}$ and average error is given by

$$\bar{C}_M(\Lambda) = C_r(\Lambda) \tag{48}$$

The proof of this theorem parallels that of Hughes [11], whose proof is in turn based on the original proof by Csiszár and Narayan [5]. The converse follows from the fact that the jammer can simulate up to L_{sym} codewords via its state input, so there is a constant probability that the decoder will not be able to construct a list containing the correct codeword whose size is smaller than L_{sym} .

The decoding rule we use is an extension of the Hughes rule to the case with constraints. To show that C_r is achievable for L_{sym} , we use the fact that a random codebook with fixed type P enjoys certain properties (Lemma 5). We then show that nonsymmetrizability for $M > L_{\text{sym}}$ implies a certain “separation” of probability distributions (Lemma 6), which we can use to show that the decoding rule will be unambiguous (Lemma 7). The codebook properties plus this unambiguous decoding allows us to show that $I(P)$ is achievable with fixed input type P (Lemma 8). Maximizing over P gives the result.

A. The converse

We have the following extension of [5, Lemma 1] and [11, Lemma 4] for these channels. The result is that error will be large for list codes with list size smaller than the symmetrizability of the channel.

Lemma 4: For an AVC $\{\mathcal{W}, l(\cdot), \Lambda\}$ with symmetrizability $L_{\text{sym}}(P)$, every list- L code with blocklength n and codewords of fixed type P with $N \geq 2$ codewords satisfies

$$\max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \bar{\varepsilon}(\mathbf{s}) \geq \left(1 - \frac{L}{K+1}\right) \left(\frac{N-K}{N}\right) - \frac{1}{n} \cdot \frac{l_{\max}^2}{(\Lambda - \Lambda_K(P))^2} \quad (49)$$

where $K = \min(N-1, L_{\text{sym}}(P))$.

Proof: Fix a codebook with codewords $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ and list- L decoder ϕ . Since the channel is K -symmetrizable we know there is a channel $U : \mathcal{X}^K \rightarrow \mathcal{S}$ such that

$$V(y|x_0^K) = \sum_{s \in \mathcal{S}} W(y|x_0, s) U(s|x_1^K) \quad (50)$$

is symmetric in (x_0, \dots, x_K) .

For any $J \subset \{1, 2, \dots, N\}$ with $|J| = K$ let \mathbf{S}_J be a random variable distributed like

$$Q^n(\mathbf{s}_J) = \prod_{k=1}^n U(s_{J,k} | \{x_{j,k} : j \in J\}) \quad (51)$$

That is, the k -th element of \mathbf{S}_J is formed by passing the k -th elements of the codewords $\{\mathbf{x}_j : j \in J\}$ through the channel U . Then we have

$$\mathbb{E}[W^n(\mathbf{y}|\mathbf{x}_i, \mathbf{S}_J)] = \sum_{\mathbf{s}} W^n(\mathbf{y}|\mathbf{x}_i, \mathbf{s}) U^n(\mathbf{s}|\{\mathbf{x}_j : j \in J\}) \quad (52)$$

We will use the symmetry of this function to obtain our bound.

Pick now a set $G \subset \{1, 2, \dots, N\}$ with $|G| = K+1$. We have for $i \in G$ that

$$\begin{aligned} \sum_{i \in G} \mathbb{E}[\varepsilon(i, \mathbf{S}_{G-\{i\}})] &= \sum_{i \in G} \left(1 - \sum_{\mathbf{y}: i \in \phi(\mathbf{y})} \mathbb{E}[W^n(\mathbf{y}|\mathbf{x}_i, \mathbf{S}_{G-\{i\}})]\right) \\ &= K+1 - \sum_{i \in G} \sum_{\mathbf{y}: i \in \phi(\mathbf{y})} \mathbb{E}[W^n(\mathbf{y}|\mathbf{x}_{i_0}, \mathbf{S}_{G-\{i_0\}})] \\ &\geq K+1 - L \sum_{\mathbf{y} \in \mathcal{Y}^n} \mathbb{E}[W^n(\mathbf{y}|\mathbf{x}_{i_0}, \mathbf{S}_{G-\{i_0\}})] \\ &= K+1 - L. \end{aligned} \quad (53)$$

Suppose the jammer chooses the following strategy : it randomly chooses a subset J uniformly from all subsets of size K , generates an \mathbf{S}_J and sends that. The expected error for this strategy is

$$\begin{aligned}
\binom{N}{K}^{-1} \sum_{J \in \Sigma_K} \mathbb{E} [\bar{\varepsilon} \mathbf{S}_J] &= \binom{N}{K}^{-1} \sum_J \frac{1}{N} \sum_{i=1}^N \mathbb{E} [\varepsilon(i, \mathbf{S}_J)] \\
&\geq \frac{1}{N} \binom{N}{K}^{-1} \sum_G \sum_{i \in G} \mathbb{E} [\varepsilon(i, \mathbf{S}_{G-\{i\}})] \\
&\geq \frac{\binom{N}{K+1} (K+1-L)}{N \cdot \binom{N}{K}} \\
&= \left(1 - \frac{L}{K+1}\right) \left(\frac{N-K}{N}\right). \tag{54}
\end{aligned}$$

We now turn to the state constraint. Note that

$$\mathbb{E} [\bar{\varepsilon}(\mathbf{S}_J)] \leq \max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \bar{\varepsilon}(\mathbf{s}) + \mathbb{P}(l(\mathbf{S}_J) > \Lambda) \tag{55}$$

So

$$\max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \bar{\varepsilon}(\mathbf{s}) \geq \mathbb{E} [\bar{\varepsilon}(\mathbf{S}_J)] - \mathbb{P}(l(\mathbf{S}_J) > \Lambda) \tag{56}$$

Suppose $\Lambda_K(P) < \Lambda$ and U attains the minimum in the definition of $\Lambda_K(P)$. Then by expanding out the expectation we can see that

$$\mathbb{E} [l(\mathbf{S}_J)] = \Lambda_K(P). \tag{57}$$

Furthermore

$$\text{Var}(l(\mathbf{S}_J)) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(l(s_{J,k})) \leq \frac{l_{\max}^2}{n} \tag{58}$$

Therefore Chebyshev's bound gives us

$$\begin{aligned}
\mathbb{P}(l(\mathbf{S}_J) \leq \Lambda) &= \mathbb{P}(l(\mathbf{S}_J) - \mathbb{E}[l(\mathbf{S}_J)] > \Lambda - \Lambda_K(P)) \\
&\geq \frac{l_{\max}^2}{n} \cdot \frac{1}{(\Lambda - \Lambda_K(P))^2}. \tag{59}
\end{aligned}$$

From (54) and (58)

$$\max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \bar{\varepsilon}(\mathbf{s}) \geq \left(1 - \frac{L}{K+1}\right) \left(\frac{N-K}{N}\right) - \frac{1}{n} \cdot \frac{l_{\max}^2}{(\Lambda - \Lambda_K(P))^2}, \tag{60}$$

as desired. ■

B. Decoding rule

In order to describe the decoding rule we will use, we must define the set

$$\mathcal{G}_\eta(\Lambda) = \{P_{XSY} : D(P_{XSY} \| P_X \times P_S \times W) \leq \eta, \mathbb{E}[l(s)] \leq \Lambda\} \tag{61}$$

We can think of $\mathcal{G}_\eta(\Lambda)$ as those joint types which are close to those generated from the AVC via independent inputs of type P_X and P_S .

Definition 1 (Decoding rule): Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ be a given codebook and suppose \mathbf{y} was received. Let $\mathcal{L}(\mathbf{y})$ denote the list decoded from \mathbf{y} . Then put $i \in \mathcal{L}(\mathbf{y})$ if and only if

- 1) there exists an $\mathbf{s} \in \mathcal{S}^n$ such that $T_{\mathbf{x}_i \mathbf{s} \mathbf{y}} \in \mathcal{G}_\eta(\Lambda)$.
- 2) for every choice of L other distinct codewords $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_L}$ such that $T_{\mathbf{x}_{i_j} \mathbf{s}_j \mathbf{y}} \in \mathcal{C}_\eta$ for some $\mathbf{s}_j \in \mathcal{S}^n$ and all $j = 1, 2, \dots, L$ we have

$$I(YX \wedge X^L | S) \leq \eta \quad (62)$$

where $P_{YX^L S}$ is the joint type of $(\mathbf{y}, \mathbf{x}_i, \mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_L}, \mathbf{s})$.

This is the decoding rule used by Hughes in [11] modified in the natural way suggested by Csiszár and Narayan [5].

C. Codebook generation

We use Lemma 1 from [11].

Lemma 5 (Codebook existence): For any $L \geq 1$, $\epsilon > 0$, $n \geq n_0(\epsilon, L)$, $N \geq L \exp(n\epsilon)$, and type P , there exist codewords $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$, each of type P , such that for every $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{s} \in \mathcal{S}^n$, and joint type $P_{X^L S}$ we have the following for $k = 1, 2, \dots, L$:

- 1) If $I(X \wedge S) \geq \epsilon$ then

$$\frac{1}{N} |\{i : (\mathbf{x}_i, \mathbf{s}) \in P_{XS}\}| \leq \exp(-n\epsilon/2) . \quad (63)$$

- 2) If $I(X \wedge X_k S) \geq |R - I(X_k \wedge S)|^+ + \epsilon$ then

$$\frac{1}{N} |\{i : (\mathbf{x}_i, \mathbf{x}_j, \mathbf{s}) \in P_{X X_k S} \text{ for some } j \neq i\}| \leq \exp(-n\epsilon/2) . \quad (64)$$

- 3) Also, for any \mathbf{x}

$$|\{j : (\mathbf{x}, \mathbf{x}_j, \mathbf{s}) \in P_{X X_k S}\}| \leq \exp(n(|R - I(X_k \wedge X S)|^+ + \epsilon)) . \quad (65)$$

- 4) Moreover, if $R < \min_k I(X_k \wedge S)$, then $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ can be selected to further satisfy

$$|\{J \in \Sigma_L : (\mathbf{x}_i, \mathbf{x}_J, \mathbf{s}) \in P_{X X^L S}\}| \leq \exp(n\epsilon) . \quad (66)$$

- 5) If $R < \min_k I(X_k \wedge S)$ and $I(X \wedge X^L S) \geq \epsilon$ then

$$\frac{1}{N} |\{i : (\mathbf{x}_i, \mathbf{x}_J, \mathbf{s}) \in P_{X X^L S} \text{ for some } J \in \Sigma_L(-\{i\})\}| \leq \exp(-n\epsilon/2) . \quad (67)$$

The proof of this lemma is contained in [11] and requires some large deviations results [5], [19] that have proved useful in many other coding problems for AVCs under average error [11], [20]–[22]. We note that these properties do not depend on the presence of the state constraint.

D. Nonsymmetrizability and separation

We need to prove that using lists longer than the symmetrizability of the AVC will lead to a sufficient “distance” property for us to use in our decoding rule for the AVC. The proof is given in Appendix II.

Lemma 6: Let $\beta > 0$, $\{\mathcal{W}, l(\cdot), \Lambda\}$ be an AVC, $P \in \mathcal{P}(\mathcal{X})$ with $I(P) > 0$ and $\min_x P(x) \geq \beta$, and $M = L_{\text{sym}}(P) + 1$. Then there exists a $\zeta > 0$ such that every collection of distributions $\{U_i \in \mathcal{P}(\mathcal{X}^M \times \mathcal{S}) : i = 1, 2, \dots, M\}$ satisfy

$$\max_{j \neq i} \sum_{y, x^{M+1}} \left| \sum_s W(y|x_i, s) U_i(x_{-\{i\}}^{M+1}, s) P(x_i) - \sum_s W(y|x_j, s) U_j(x_{-\{j\}}^{M+1}, s) P(x_j) \right| \geq \zeta \quad (68)$$

Furthermore, for any AVC and $\alpha > 0$ there exists a $\zeta > 0$ such that (68) holds for any collection of U_i 's for which a P can be found with

$$\sum_{x^{M+1}, s} P(x_i) U_i(x_{-\{i\}}^M, s) l(s) \leq \Lambda_M(P) - \alpha \quad (69)$$

for all $i = 1, 2, \dots, M + 1$.

E. Non ambiguity of decoding

We must ensure that our decoding rule will not output a list of size larger than $L_{\text{sym}}(P) + 1$. The next lemma shows that for sufficiently small η there are no random variables that can force the decoding rule to output a list that is too large.

Lemma 7: Let $M = L_{\text{sym}}(P) + 1$ for the AVC $\{\mathcal{W}, l(\cdot), \Lambda\}$. If $\beta > 0$ then no tuple of rv's (Y, X^{M+1}, S^{M+1}) can satisfy

$$\min_x P(x) \geq \beta \quad (70)$$

$$P_{X_i} = P \quad (71)$$

$$P_{Y X_i S_i} \in \mathcal{G}_\eta(\Lambda) \quad (72)$$

$$I\left(Y X_i \wedge X_{-\{i\}}^{M+1} \middle| S_i\right) \leq \eta \quad 1 \leq i \leq M + 1 \quad (73)$$

Proof: Assume, to the contrary, that there does exist a tuple of random variables (Y, X^{M+1}, S^{M+1}) satisfying (70)–(73). This will lead to a bound on a certain KL-divergence which, via Pinsker's inequality, becomes a bound on total variational distance that contradicts the conclusion of Lemma 6.

Let $W_i = W(\cdot, S_i)$. For every i we then have the following divergence bound:

$$\begin{aligned} D\left(P_{Y X^{M+1} S_i} \middle\| W_i \times P_{X_i} \times P_{X_{-\{i\}}^{M+1} S_i}\right) \\ &= D\left(P_{Y X_i S_i} \middle\| W_i \times P_{X_i} \times P_{S_i}\right) + D\left(P_{X_{-\{i\}}^{M+1} | Y X_i S_i} \middle\| P_{X_{-\{i\}}^{M+1} | S_i} \middle| P_{Y X_i S_i}\right) \\ &= D\left(P_{Y X_i S_i} \middle\| W_i \times P_{X_i} \times P_{S_i}\right) + I\left(Y X_i \wedge X_{-\{i\}}^{M+1} \middle| S_i\right) \\ &\leq 2\eta, \end{aligned}$$

where the last line follows from (72), (61) and (73).

Projecting the distributions onto $\mathcal{Y} \times \mathcal{X}^{M+1}$ cannot increase the divergence, so if we define

$$W_i P_{X_{-\{i\}}^{M+1}}(y | x_{-\{i\}}^{M+1} | x_i) = \sum_s W(y | x_i, s) P_{X_{-\{i\}}^{M+1} S_i}(x_{-\{i\}}^{M+1}, s), \quad (74)$$

we get

$$D\left(P_{Y X^{M+1}} \middle\| W_i P_{X_{-\{i\}}^{M+1}} \times P_{X_i}\right) < 2\eta$$

To use Lemma 6 we must turn this divergence bound into a bound on a total variational distance. We can use Pinsker's inequality [16, p. 58, Problem 17] to show that the KL-divergence is an upper bound on the variational distance:

$$\sum_{y, x^{L+1}} \left| P_{Y X^{M+1}}(y, x^{L+1}) - W_i P_{X_{-\{i\}}^{M+1}}(y | x_{-\{i\}}^{M+1} | x_i) P_{X_i}(x_i) \right| < \sqrt{(2 \ln 2) \eta} \quad \forall i \in [M + 1] \quad (75)$$

Since the bound holds for all i , we know that the second terms must be close to each other, and by using (74)

$$\begin{aligned}
& \max_{i \neq j} \sum_{y, x^{L+1}} \left| W_i P_{X_{-\{j\}}^{M+1}}(y|x_{-\{j\}}^{M+1}|x_j) P_{X_j}(x_j) - W_i P_{X_{-\{i\}}^{M+1}}(y|x_{-\{i\}}^{M+1}|x_i) P_{X_i}(x_i) \right| \\
&= \max_{i \neq j} \sum_{y, x^{L+1}} \left| \sum_s W(y|x_j, s) P_{X_{-\{j\}}^{M+1}}(x_{-\{j\}}^{M+1}, s|x_j) P_{X_j}(x_j) - \sum_s W(y|x_i, s) P_{X_{-\{i\}}^{M+1}}(x_{-\{i\}}^{M+1}, s|x_i) P_{X_i}(x_i) \right| \\
&< 2\sqrt{(2 \ln 2)\eta}
\end{aligned} \tag{76}$$

By choosing η small enough we violate the conclusion of Lemma 6, which is the desired contradiction. \blacksquare

F. Achievability of rates

The statement of the following lemma is identical to Lemma 3 of Hughes, and a proof is included for completeness in Appendix I.

Lemma 8: For any $\delta > 0$, $\beta > 0$ and $P \in \mathcal{P}(\mathcal{X})$ with $I(P) > 0$ and $\min_x P(x) \geq \beta$, for $M = L_{\text{sym}}(P) + 1$ there exists a list- M code of constant type P such that

$$R = \frac{1}{n} \log \left(\frac{N}{M} \right) > I(P) - \delta, \tag{77}$$

$$\max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \bar{\epsilon}(\mathbf{s}) < \exp(-n\gamma), \tag{78}$$

for all $n \geq n_2(\beta, \delta, \mathcal{W})$ and $\gamma(\beta, \delta, \mathcal{W})$.

G. Putting it all together

We now provide the proof of Theorem 3.

Proof: Let $P \in \mathcal{P}(\mathcal{X})$ with $\min P(x) > \beta$ for some $\beta > 0$ and $I(P) > 0$. Suppose $M \leq L_{\text{sym}}(P)$. For any rate $R > 0$, we know that for n large enough $L_{\text{sym}} < N - 1$. Then from Lemma 4 the error is lower bounded by

$$\max_{\mathbf{s} \in \mathcal{S}^n} \bar{\epsilon}(\mathbf{s}) \geq \left(1 - \frac{M}{L_{\text{sym}}(P) + 1} \right) \left(\frac{N - L_{\text{sym}}(P)}{N} \right) - \frac{1}{n} \cdot \frac{l_{\max}^2}{(\Lambda - \Lambda_{L_{\text{sym}}(P)}(P))^2} \tag{79}$$

So for large n the first term is strictly positive and not decreasing with n , which establishes that no positive rate is achievable if $M \leq L_{\text{sym}}(P)$.

Suppose now that $M = L_{\text{sym}}(P) + 1$. Then Lemma 8 shows that codebooks of type P can achieve rates arbitrarily close to $I(P)$. Maximizing over P and choosing $M = L_{\text{sym}}(P^*) + 1$, we see that for any $\delta > 0$, rates

$$R > \max_{P \in \mathcal{P}(\mathcal{X})} \min_{Q \in \mathcal{P}(\mathcal{S}, \Lambda)} I(X \wedge Y) - \delta = C_r(\Lambda) - \delta \tag{80}$$

are achievable. Therefore $C_L(\Lambda) = C_r(\Lambda)$ for $L > L_{\text{sym}}(P^*)$. \blacksquare

V. EXAMPLES

We now turn to an example of an additive cost-constrained AVC. Let $\mathcal{X} = \{-1, 1\}$ and let $\mathcal{S} = \{-\sigma, -\sigma + 1, \dots, \sigma\}$ for some integer σ . The output Y of this channel is given by

$$Y = X + S, \quad (81)$$

that is, the real addition of the input and state. This is similar in spirit to the example given by Hughes [11], but is more closely related to [23], which analyzes a game between power constrained noise and an encoder with binary inputs.

We will consider two kinds of cost constraints on the jammer for this AVC. The first is an L_1 constraint:

$$l_1(s) = |s|. \quad (82)$$

The second is an L_2 constraint:

$$l_2(s) = s^2. \quad (83)$$

For each of these constraints we will compute capacities for different values of Λ .

The first question to settle is that of the randomized coding capacity $C_r(\Lambda)$ for this channel, given by (10). By symmetry, we may assume that the input distribution P^* is uniform on the set $\{-1, 1\}$. Since max-min is min-max:

$$C_r(\Lambda) = \min_{Q(s) \in \mathcal{P}(\mathcal{S}, \Lambda)} I(X \wedge X + S) \quad (84)$$

$$= \min_{Q(s) \in \mathcal{P}(\mathcal{S}, \Lambda)} H(X + S) - H(S) \quad (85)$$

This minimization can be carried out numerically by convex optimization methods. Some details are given in Appendix III. Capacities for the L_1 case are shown in Figure 1 and for L_2 are shown in Figure 2. As the cost constraint Λ is increased, the randomized coding capacity decreases, and for smaller alphabet sizes the constraint becomes inactive.

For coding under average probability of error, we would also like to know the symmetrizability of these channels. For each candidate list size L , we must determine if there exists a channel $U : \mathcal{X}^L \rightarrow \mathcal{S}$ satisfying (6). The channel $U(S|X^L)$ itself must be symmetric, so it is only a function of the type of X^L . We can rewrite the averaged channel as

$$\sum_s W(y|x, s)U(s|t) \quad (86)$$

where $t \in [0, 1, \dots, L]$ counts the number of 1's in X^L . The condition that V be symmetric can be rewritten as:

$$\sum_s W(y| -1, s)U(s|t) - \sum_s W(y| +1, s)U(s|t-1) \quad \forall y, t \quad (87)$$

where U must satisfy

$$\sum_s \sum_{t=0}^L \binom{L}{t} 2^{-L} l(s) U(s|t) \leq \Lambda. \quad (88)$$

To obtain a convex objective function, note that (87) holds if and only if

$$f(U) = \sum_y \sum_{t=1}^L \left(\sum_s W(y| -1, s)U(s|t) - \sum_s W(y| +1, s)U(s|t-1) \right)^2 = 0 \quad (89)$$

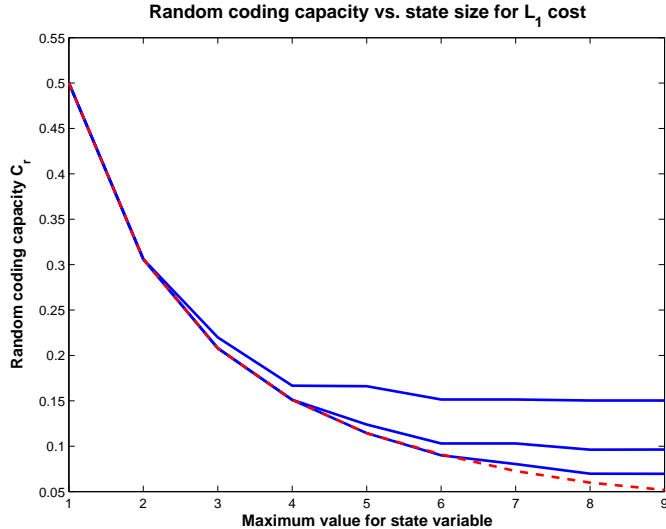


Fig. 1. Random coding capacity $C_r(\Lambda)$ versus σ for $\Lambda = 1.5, 2,$ and 2.5 with loss function $l_1(s)$. The dashed line is the randomized coding capacity for the unconstrained case.

To determine if the channel is L -symmetrizable, we minimize the function $f(U)$. If $\min f(U) = 0$ then the channel is symmetrizable, and if $\min f(U) > 0$ it is not. If we replace the square in (89) with an absolute value function, then we obtain a function similar to that in Lemma 6.

The plot in Figure 3 shows a plot of $\max I(P)$ against Λ for $l_2(\cdot)$ and $\sigma = 8$ under with list codes of fixed list sizes. As Λ increases, the capacity-achieving input distribution with $P(X = 1) = 1/2$ becomes L -symmetrizable for small L . However, suboptimal input distributions are not L -symmetrizable, and list codes of size L can still achieve some rates below C_r . In Figure 4 we show $\arg\max I(P)$ for the distributions P that are not L -symmetrizable. This shows the main difference between the constrained and unconstrained AVC problems – for some values of L and Λ the capacity may be positive but strictly less than the random coding capacity.

The extensive analysis in [23] found that the worst case power-constrained *noise* for binary modulation had support only on integer points. In this example, we are interested in the interplay between the list size, achievable rates, and cost constraint. In order to compute the random coding capacity we need to find the worst-case noise distribution, but this capacity is not in general realizable with deterministic codes. List coding relaxes the coding problem and, as we have shown here, approaches the performance of randomized coding for small list sizes.

VI. CONCLUSIONS

In this paper we have extended the earlier results [8], [10], [11] on list coding for arbitrarily varying channels to the case in which the state is subject to a cost constraint. The proof structures parallel those for the unconstrained case. Although somewhat technical, these results on list decoding for cost-constrained AVCs are useful for constructing codes for adversarial communication models and demonstrate some of the differences between unconstrained AVCs and constrained AVCs.

For maximal error, we first construct list codes of exponential list size and subsample them to show that rates approaching $C_r(\Lambda)$ can be achieved with increasing list sizes. For average error, the symmetrizability of the channel under for an input distribution P gives a criterion under which list decoding can be

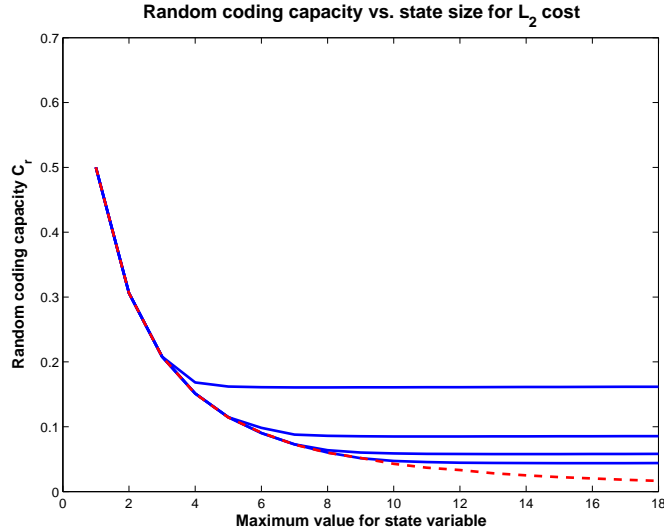


Fig. 2. Random coding capacity $C_r(\Lambda)$ versus σ for $\Lambda = 4, 8, 12, 16$ with loss function $l_2(s)$. The dashed line is the randomized coding capacity for the unconstrained case.

successful. For the P^* that is capacity-achieving for randomized coding, we can show that $\bar{C}_L(\Lambda) = C_r(\Lambda)$ for $L < L_{\text{sym}}(P^*)$.

For maximal error we have shown that lists of size $O(\epsilon^{-1})$ are sufficient to achieve rates $R = C_r(\Lambda) - \epsilon$. This means that if we can approach the randomized coding capacity with larger and larger lists, guaranteeing that the error will be small on every message for every state sequence. For deterministic codes, this implies that the state sequence can depend on the *codeword* as well as the message. A related channel model was proposed recently by Agarwal, Sahai, and Mitter [24] in the context of rate distortion theory, and randomized coding proved to be a key component of their strategy. In a subsequent paper we will use a list code to construct randomized codes with small key size for AVC models in which the codeword is known to the jammer.

For average error it may be the case that $L_{\text{sym}}(P) < L_{\text{sym}}(P^*)$ for some P with $I(P) > 0$. In this case the rate $I(P)$ is achievable with lists smaller than $L_{\text{sym}}(P^*)$. However, if achieving the randomized coding capacity is our goal, we require lists larger than $L_{\text{sym}}(P^*)$. As for non-list codes, Ahlswede's dichotomy theorem does not hold for constrained AVCs, and list coding shows one qualitative difference between constrained and unconstrained AVCs – the set of achievable rates increases with the list size until it reaches the randomized coding capacity in the constrained case, whereas for unconstrained AVCs there is an abrupt jump from 0 to C_r .

If we restrict our attention to linear codes, a connection between the notion of symmetrizability for list codes and generalized Hamming weights [25], [26] has been shown by Guruswami [27] for the case of list decoding from erasures. The r -th generalized Hamming weight $d_r(\mathcal{C})$ of a code \mathcal{C} is the minimum weight for the basis of an r -dimensional subcode of \mathcal{C} . For erasure channels, a list code \mathcal{C} can correct Λn errors with a list of size L if and only if $d_r(\mathcal{C}) > \Lambda n$ for $r = 1 + \lceil \log n \rceil$. The converse argument is similar to that for the average-error AVC – the error pattern can simulate r codewords if $d_r(\mathcal{C}) < \Lambda n$. It would be interesting to see how strong this connection is for more general AVCs.

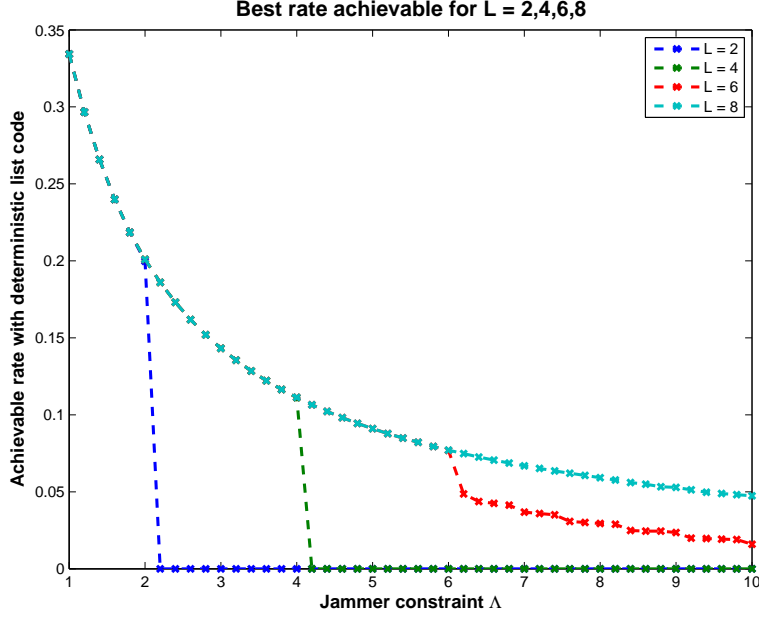


Fig. 3. The largest value of $I(P)$ achievable versus Λ for $\theta = 2$, $\sigma = 8$, and for different list sizes.

APPENDIX I PROOF OF LEMMA 8

Proof: Choose N to so that

$$I(P) - \delta < R < I(P) - \delta/2 . \quad (90)$$

Let ϵ and η be parameters to be chosen later in the proof. Choose N codewords according to Lemma 5 using this ϵ . The decoding rule will be given by Definition 1. Lemma 7 proves that for small η the decoding rule is unambiguous and that the list size needed is at most $M = L_{\text{sym}}(P) + 1$.

We must now bound the average probability of error for a fixed \mathbf{s} :

$$\varepsilon(\mathbf{s}) = \frac{1}{N} \sum_{i=1}^N \varepsilon(i, \mathbf{s}) . \quad (91)$$

We must bound this error using the properties of the codebook guaranteed by Lemma 5.

Suppose that \mathbf{x}_i was transmitted, the state sequence was \mathbf{s} , and \mathbf{y} was received, $i \notin \mathcal{L}(\mathbf{y})$. From the definition of the decoder in Definition 1 we know that \mathbf{y} must violate either the first or second of the two conditions in the decoding rule. Correspondingly, we can define:

$$A_i(\mathbf{s}) = \{\mathbf{y} : T_{\mathbf{x}_i \mathbf{s} \mathbf{y}} \in \mathcal{G}_\eta(\Lambda)\} \quad (92)$$

$$B_i(\mathbf{s}) = A_i(\mathbf{s}) \cap \bigcup_{P_{YXX^L S} \in \mathcal{H}_\eta(i, \mathbf{s})} E(P_{YXX^L S}) \quad (93)$$

$$\mathcal{H}_\eta(i, \mathbf{s}) = \{P_{YXX^L S} : \exists J \in \Sigma_L(-\{i\}), P_{YXX^L S} = T_{\mathbf{y} \mathbf{x}_i \mathbf{x}_J \mathbf{s}}, I(XY \wedge X^L | S) > \eta\} \quad (94)$$

$$E(P_{YXX^L S}) = \{\mathbf{y} : \exists (k, \mathbf{s}_k), k \in J, \mathbf{y} \in A_k(\mathbf{s}_k)\} \text{ for } P_{YXX^L S} \in \mathcal{H}_\eta(i, \mathbf{s}) \quad (95)$$

$$F(\mathbf{s}) = \{i : I(X \wedge S) < \epsilon, P_{X S} = T_{\mathbf{x}, \mathbf{s}}\} \quad (96)$$

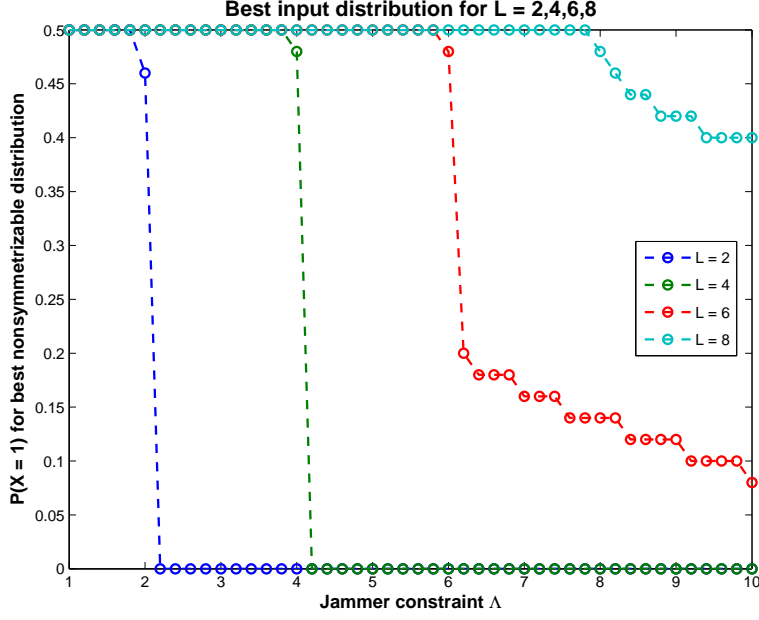


Fig. 4. The value of $P(X = 1)$ which is non-symmetrizable and achieves the highest rate versus Λ for $\theta = 2$, $\sigma = 8$, and different list sizes.

An output $\mathbf{y} \in A_i^c(\mathbf{s})$ is atypical and fails the first criterion of the decoding rule. A joint distribution is in $\mathcal{H}_\eta(i, \mathbf{s})$ if it is the type of $(\mathbf{y}, \mathbf{x}, \mathbf{x}_J, \mathbf{s})$, where J is a list of L codewords not containing i , and if the mutual information condition in the second part of the decoding rule is violated. An output \mathbf{y} is in $E(T_{\mathbf{y}\mathbf{x}_i\mathbf{x}_J\mathbf{s}})$ if there is a $k \in J$ and a state sequence \mathbf{s}_k so that \mathbf{x}_k is an alternate candidate for being the transmitted codeword. Putting this together, the $\mathbf{y} \in B_i(\mathbf{s})$ fail the second decoding rule. Therefore we can write the error as:

$$\begin{aligned} \varepsilon(\mathbf{s}) &= \frac{1}{N} \left(\sum_{i \in F^c(\mathbf{s})} \varepsilon(i, \mathbf{s}) + \sum_{i \in F(\mathbf{s})} \varepsilon(i, \mathbf{s}) \right) \\ &\leq \frac{1}{N} \left(\sum_{i \in F^c(\mathbf{s})} \varepsilon(i, \mathbf{s}) + \sum_{i \in F(\mathbf{s})} W^n(A_i^c(\mathbf{s})|\mathbf{x}_i, \mathbf{s}) + \sum_{i \in F(\mathbf{s})} \sum_{P_{Y_{XX^L S}} \in \mathcal{H}_\eta} W^n(E(P_{Y_{XX^L S}}|\mathbf{x}_i, \mathbf{s})) \right) \end{aligned} \quad (97)$$

We must bound each of these three terms using the properties of our codebook guaranteed by Lemma 5 and some properties of types.

To bound the first term, note that for a joint type P_{XS} we can bound $|F^c(\mathbf{s})|$ using (1)

$$\begin{aligned} \frac{1}{N} \sum_{i \in F^c(\mathbf{s})} \varepsilon(i, \mathbf{s}) &\leq \frac{|F^c(\mathbf{s})|}{N} \\ &\leq \sum_{P_{XS} \in \mathcal{P}_n(\mathcal{X}, \mathcal{S})} \exp(-n\epsilon/2) \\ &\leq (n+1)^{|\mathcal{X}||\mathcal{S}|} \exp(-n\epsilon/2) \end{aligned} \quad (98)$$

To bound the second term, we use some facts about types. Note first that for $P_{XS} = T_{\mathbf{x}_i\mathbf{s}}$ and

$I(X \wedge S) = D(P_{XS} \parallel P_X \times P_S)$ we have:

$$D(P_{XSY} \parallel P_{XS} \times W) + I(X \wedge S) = D(P_{XSY} \parallel P_X \times P_S \times W) . \quad (99)$$

For $i \in F(\mathbf{s})$ we have $I(X \wedge S) < \epsilon$ and for $\mathbf{y} \in A_i^c(\mathbf{s})$ we have $D(P_{XSY} \parallel P_X \times P_S \times W) > \eta$, so

$$D(P_{XSY} \parallel P_{XS} \times W) > \eta - \epsilon . \quad (100)$$

Now, the second sum can be rewritten using (99) and (17):

$$\begin{aligned} \sum_{i \in F(\mathbf{s})} W^n(A_i^c(\mathbf{s})|\mathbf{x}_i, \mathbf{s}) &\leq \sum_{P_{XSY} \notin \mathcal{G}_\eta(\Lambda)} W^n(\{\mathbf{y} : P_{XSY} = T_{\mathbf{x}_i \mathbf{s} \mathbf{y}}\}|\mathbf{x}_i \mathbf{s}) \\ &\leq (n+1)^{|\mathcal{X}||\mathcal{S}||\mathcal{Y}|} \exp(-nD(P_{XSY} \parallel P_{XS} \times W)) \\ &\leq (n+1)^{|\mathcal{X}||\mathcal{S}||\mathcal{Y}|} \exp(-n(\eta - \epsilon)) . \end{aligned} \quad (101)$$

To bound the last term, let us fix $P_{YXX^L S} \in \mathcal{H}_\eta$. We will consider the cases where R is greater then or less than $\min_k I(X_k \wedge S)$, where $k \in [L]$.

- 1) Suppose $R < \min_k I(X_k \wedge S)$. In this case $|R - I(X_k \wedge S)|^+ = 0$. We consider two sub-cases. If $I(X \wedge X^L S) \geq \epsilon$ then part 5 of Lemma 5 says that for all k

$$\begin{aligned} \frac{1}{N} \sum_{i \in F(\mathbf{s})} W^n(E(P_{YXX^L S})|\mathbf{x}_i, \mathbf{s}) &\leq \frac{1}{N} |\{i : (\mathbf{x}_i, \mathbf{x}_J, \mathbf{s}) \in P_{YXX^L S} \text{ for some } J \in \Sigma_L(-\{i\})\}| \\ &\leq \exp(-n\epsilon/2) . \end{aligned} \quad (102)$$

If instead $I(X \wedge X^L S) < \epsilon$ we can bound

$$E(P_{YXX^L S}) \subset \bigcup_{J \in \Sigma_L(-\{i\}): T_{\mathbf{x}_i \mathbf{x}_J \mathbf{s}} = P_{YXX^L S}} \{\mathbf{y} : T_{\mathbf{y} \mathbf{x}_i \mathbf{x}_J \mathbf{s}} = P_{YXX^L S}\} . \quad (103)$$

Using a union bound, (18) and part 4 of Lemma 5 we obtain

$$\begin{aligned} \sum W^n(E(P_{YXX^L S})|\mathbf{x}_i, \mathbf{s}) &\leq \sum_{J \in \Sigma_L(-\{i\}): T_{\mathbf{x}_i \mathbf{x}_J \mathbf{s}} = P_{YXX^L S}} W^n(\{\mathbf{y} : T_{\mathbf{y} \mathbf{x}_i \mathbf{x}_J \mathbf{s}} = P_{YXX^L S}\}|\mathbf{x}_i, \mathbf{s}) \\ &\leq \sum_{J \in \Sigma_L(-\{i\}): T_{\mathbf{x}_i \mathbf{x}_J \mathbf{s}} = P_{YXX^L S}} \exp(-nI(Y \wedge X^L | XS)) \\ &\leq \exp(-n(I(Y \wedge X^L | XS) - \epsilon)) \end{aligned} \quad (104)$$

To bound the exponent, note that since $P_{YXX^L S} \in \mathcal{H}_\eta(i, \mathbf{s})$ and $I(X \wedge X^L S) < \epsilon$,

$$\begin{aligned} I(Y \wedge X^L | XS) &= I(YX \wedge X^L | S) - I(X \wedge X^L | S) \\ &> \eta - I(X \wedge X^L S) \\ &> \eta - \epsilon \end{aligned} \quad (105)$$

Then (104) and (105) give

$$\sum_{i \in F(\mathbf{s})} W^n(E(P_{YXX^L S})|\mathbf{x}_i, \mathbf{s}) \leq \exp(-n(\eta - 2\epsilon)) . \quad (106)$$

2) Suppose $R \geq \min_k I(X_k \wedge S)$, and pick k such that $R \geq I(X_k \wedge S)$. Then if $I(X \wedge X^L S) \geq |R - I(X_k \wedge S)|^+ \epsilon$ then part 2 of Lemma 5 says that for this k

$$\begin{aligned} \frac{1}{N} \sum_{i \in F(\mathbf{s})} W^n(E(P_{Y X X^L S}) | \mathbf{x}_i, \mathbf{s}) &\leq \frac{1}{N} |\{i : (\mathbf{x}_i, \mathbf{x}_j, \mathbf{s}) \in P_{X X_k S} \text{ for some } j \neq i\}| \\ &\leq \exp(-n\epsilon/2). \end{aligned} \quad (107)$$

Suppose now that $I(X \wedge X^L S) < |R - I(X_k \wedge S)|^+ \epsilon$. We may assume $P_{X_k} = P_X$, since X_k has the type of a codeword by the definition of $\mathcal{H}(i, \mathbf{s})$. Rewriting as before:

$$E(P_{Y X X^L S}) \subset \bigcup_{j \neq i: T_{\mathbf{x}_i \mathbf{x}_j \mathbf{s}} = P_{X X_k S}} \{\mathbf{y} : T_{\mathbf{y} \mathbf{x}_i \mathbf{x}_j \mathbf{s}} = P_{Y X X_k S}\}. \quad (108)$$

Using a union bound, (18) and part 3 of Lemma 5 we obtain

$$\begin{aligned} W^n(E(P_{Y X X^L S}) | \mathbf{x}_i, \mathbf{s}) &\leq \sum_{j \neq i: T_{\mathbf{x}_i \mathbf{x}_j \mathbf{s}} = P_{X X_k S}} W^n(\{\mathbf{y} : T_{\mathbf{y} \mathbf{x}_i \mathbf{x}_j \mathbf{s}} = P_{Y X X_k S}\} | \mathbf{x}_i, \mathbf{s}) \\ &\leq \sum_{j \neq i: T_{\mathbf{x}_i \mathbf{x}_j \mathbf{s}} = P_{X X_k S}} \exp(-n(I(Y \wedge X_k | X S))) \\ &\leq \exp(-n(I(Y \wedge X_k | X S) - |R - I(X_k \wedge X S)|^+ - \epsilon)) \end{aligned} \quad (109)$$

Because $R \geq \min_k I(X_k \wedge S)$, we have

$$\begin{aligned} R &> I(X \wedge X_k S) + I(X_k \wedge S) - \epsilon \\ &\geq I(X \wedge X_k | S) + I(X_k \wedge S) - \epsilon \\ &= I(X_k \wedge X S) - \epsilon \end{aligned} \quad (110)$$

Therefore

$$\begin{aligned} I(Y \wedge X_k | X S) - |R - I(X_k \wedge X S)|^+ - \epsilon &\geq I(Y \wedge X_k | X S) + I(X_k \wedge X S) - R - 2\epsilon \\ &\geq I(Y X S \wedge X_k) - R - 2\epsilon \\ &\geq I(X_k \wedge Y) - R - 2\epsilon. \end{aligned} \quad (111)$$

Since there $P_{Y X X^L S} \in \mathcal{H}_\eta(i, \mathbf{s})$, we must have $P_{Y X_k S_k} \in \mathcal{G}_\eta(\Lambda)$. This means

$$D(P_{Y X_k S_k} \parallel P_X \times P_{S_k} \times W) < \eta, \quad (112)$$

so projecting onto (Y, X_k) and using Pinsker's inequality [16, p. 58, Problem 17] we see the total variational distance between $P_{X_k Y}$ and $P_X W$ is less than η . This in turn means that we can for any $\delta > 0$ we can choose an η small enough so that $I(X_k \wedge Y) - I(P_X) < \delta/3$. This in turn means

$$I(X_k \wedge Y) - R \geq I(P_X) - R - \delta/3 \geq \delta/3 \quad (113)$$

From (109), (111), and (113)

$$\begin{aligned} W^n(E(P_{Y X X^L S}) | \mathbf{x}_i, \mathbf{s}) &\leq \exp(-n(I(X_k \wedge Y) - R - 2\epsilon)) \\ &\leq \exp(-n(\delta/3 - 3\epsilon)). \end{aligned} \quad (114)$$

Since the number of joint types in $\mathcal{H}_\eta(i, \mathbf{s})$ is at most polynomial in n and the bounds on the probability of $E(P_{Y X X^L S})$ in (98), (101), (102), (106), (107), (114) are exponentially decreasing in all cases, we have an exponential bound on the third term in the error sum.

Thus all three terms can be bounded by terms exponentially decreasing in n and the average error can be made as small as we like by choosing the block length to be large enough. These bounds are uniform in \mathbf{s} and hence hold for all \mathbf{s} for which the decoding rule is valid. Since the rule is valid if the list size is greater than L_{sym} , we are done. ■

APPENDIX II
PROOF OF LEMMA 6

Proof: Note that the outer sum in (68) is over all x^{M+1} . Define

$$V_i(x^{M+1}, s) = U_i(x_{-\{i\}}^{M+1}, s). \quad (115)$$

Let Π_{M+1} be the set of all permutations of $[M+1]$ and for $\pi \in \Pi_{M+1}$ let π_i be the image of i under π . Then

$$\begin{aligned} & \max_{j \neq i} \sum_{y, x^{M+1}} \left| \sum_s W(y|x_i, s) V_i(x^{M+1}, s) P(x_i) - \sum_s W(y|x_j, s) V_j(x^{M+1}, s) P(x_j) \right| \\ &= \max_{j \neq i} \sum_{y, x^{M+1}} \left| \sum_s W(y|x_i, s) V_{\pi_i}(x^{M+1}, s) P(x_i) - \sum_s W(y|x_j, s) V_{\pi_j}(x^{M+1}, s) P(x_j) \right| \end{aligned} \quad (116)$$

We can lower bound this by averaging over all $\pi \in \Pi_{M+1}$:

$$\max_{j \neq i} \sum_{y, x^{M+1}} \frac{1}{(M+1)!} \sum_{\pi \in \Pi_{M+1}} \left| \sum_s W(y|x_i, s) V_{\pi_i}(x^{M+1}, s) P(x_i) - \sum_s W(y|x_j, s) V_{\pi_j}(x^{M+1}, s) P(x_j) \right| \quad (117)$$

Now we use the convexity of $|\cdot|$ to pull the averaging inside to get a further lower bound:

$$F(\bar{V}, P) = \max_{j \neq i} \sum_{y, x^{M+1}} \left| \sum_s W(y|x_i, s) \bar{V}(x_{-\{i\}}^{M+1}, s) P(x_i) - \sum_s W(y|x_j, s) \bar{V}(x_{-\{j\}}^{M+1}, s) P(x_j) \right| \quad (118)$$

where we define

$$\begin{aligned} \bar{V}(x_{-\{i\}}^{M+1}, s) &= \frac{1}{(M+1)!} \sum_{\pi \in \Pi_{M+1}} V_{\pi_i}(x^{M+1}, s) \\ &= \frac{1}{(M+1)!} \sum_{l=1}^{M+1} \sum_{\pi \in \Pi_{M+1}: \pi_i=l} U_l(\pi(x^{M+1})_{-\{\pi_i\}}, s) \\ &= \frac{1}{(M+1)!} \sum_{l=1}^{M+1} \sum_{\sigma \in \Pi_M} U_l(\sigma(x_{-\{i\}}^{M+1}), s). \end{aligned}$$

Now note that \bar{V} is a symmetric function for all s . The function $F(\bar{V}, P)$ is continuous function on the compact set of symmetric distributions $\{\bar{V}\}$ and the set of distributions P with $\min_x P(x) \geq \beta$, so it has a minimum $\zeta = F(\bar{V}^*, P^*)$ for some (\bar{V}^*, P^*) .

We will prove that the $\zeta > 0$ by contradiction. Suppose $F(\bar{V}^*, P^*) = 0$. Then

$$\sum_s W(y|x_i, s) \bar{V}^*(x_{-\{i\}}^{M+1}, s) P^*(x_i) = \sum_s W(y|x_j, s) \bar{V}^*(x_{-\{j\}}^{M+1}, s) P^*(x_j)$$

So

$$\sum_y \sum_s W(y|x_i, s) \bar{V}^*(x_{-\{i\}}^{M+1}, s) P^*(x_i) = \sum_y \sum_s W(y|x_j, s) \bar{V}^*(x_{-\{j\}}^{M+1}, s) P^*(x_j)$$

$$\bar{V}^*(x_{-\{i\}}^{M+1}) P^*(x_i) = \bar{V}^*(x_{-\{j\}}^{M+1}) P^*(x_j) ,$$

which implies (see [11, Lemma A3]) that for all j :

$$\bar{V}^*(x_{-\{j\}}^{M+1}) P^*(x_j) = P^{*(M+1)}(x^{M+1}) .$$

Therefore

$$\sum_s W(y|x_1, s) \bar{V}^*(s|x^{M+1})$$

is symmetric in $(x_1, x_2, \dots, x_{M+1})$, which contradicts our assumption. Therefore $\zeta > 0$.

To bring in the state constraints, note that if (69) holds then

$$\sum_{x^{M+1}, s} P(x_i) \bar{V}(x_{-\{i\}}^M, s) l(s) \leq \Lambda_M(P) - \alpha \quad (119)$$

Again, because the AVC is not m -symmetrizable for $m > M$, the minimum of $F(\bar{V}, P)$ with the constraint (119) cannot be 0 or else $\Lambda_M(P) < \Lambda$. Thus we can still find a $\zeta > 0$ such that (68) holds. ■

APPENDIX III

OPTIMIZATION FOR THE ADDITIVE CHANNEL EXAMPLE

To find the random coding capacity we must minimize the quantity $I(X \wedge X + S)$. This can be stated as the following optimization problem. Let $I(Q) = I(X \wedge X + S)$ with $Q = Q(s)$ and $P(X = 1) = P(X = -1) = 1/2$. For a cost function $l(s) = |s|^\theta$, let $\Theta = (l(-A), l(-A + 1), \dots, l(A))^T$. The optimization is

$$\text{minimize} \quad I(Q) \quad (120)$$

$$\text{subject to} \quad \mathbf{1}^T Q = 1 \quad (121)$$

$$-Q(s) \leq 0 \quad \forall s \quad (122)$$

$$\Theta^T Q - \Lambda \leq 0 \quad (123)$$

Since the mutual information is convex in the distribution Q , this is a convex optimization problem in the vector Q . We can use the barrier method [28, Ch. 11] to solve this problem. By performing an optimization for for each value of σ and Λ we can create the plots shown in Figures 1 and 2.

In order to calculate the symmetrizability of the channel, we must solve the following program:

$$\text{minimize} \quad f(U) = \sum_y \sum_{t=1}^L \left(\sum_s W(y| -1, s) U(s|t) - \sum_s W(y| +1, s) U(s|t-1) \right)^2 \quad (124)$$

$$\text{subject to} \quad \sum_s U(s|t) = 1 \quad \forall t \quad (125)$$

$$-U(s|t) \leq 0 \quad \forall s, t \quad (126)$$

$$\sum_s \sum_{t=0}^L \binom{L}{t} 2^{-L} l(s) U(s|t) - \Lambda \leq 0 \quad (127)$$

Again, we can use the barrier method to optimize over the parameters σ and Λ to determine, for each list size, if the channel is L -symmetrizable.

REFERENCES

- [1] I. Csiszár and J. Körner, “On the Capacity of the Arbitrarily Varying Channel for Maximum Probability of Error,” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 57, pp. 87–101, 1981.
- [2] D. Blackwell, L. Breiman, and A. Thomasian, “The capacities of certain channel classes under random coding,” *Annals of Mathematical Statistics*, vol. 31, pp. 558–567, 1960.
- [3] R. Ahlswede, “Elimination of correlation in random codes for arbitrarily varying channels,” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 44, no. 2, pp. 159–175, 1978.
- [4] T. Ericson, “Exponential error bounds for random codes on the arbitrarily varying channel,” *IEEE Transactions on Information Theory*, vol. 31, no. 1, pp. 42–48, 1985.
- [5] I. Csiszár and P. Narayan, “The Capacity of the Arbitrarily Varying Channel Revisited : Positivity, Constraints,” *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 181–193, 1988.
- [6] P. Elias, “List decoding for noisy channels,” in *Wescon Convention Record, Part 2*. Institute of Radio Engineers (now IEEE), 1957, pp. 94–104.
- [7] R. Ahlswede, “Channel Capacities for List Codes,” *Journal of Applied Probability*, vol. 10, no. 4, pp. 824–836, 1973.
- [8] —, “The Maximal Error Capacity of Arbitrarily Varying Channels for Constant List Sizes,” *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1416–1417, 1993.
- [9] V. Blinovskiy and M. Pinsker, “Estimation of the size of the list when decoding over an arbitrarily varying channel,” in *Proceedings of 1st French-Israeli Workshop on Algebraic Coding*, ser. Lecture Notes in Computer Science, G. Cohen, S. Litsyn, A. Lobstein, and G. Zémor, Eds., no. 781. Berlin: Springer-Verlag, July 1993.
- [10] V. Blinovskiy, P. Narayan, and M. Pinsker, “Capacity of the arbitrarily varying channel under list decoding,” *Problems of Information Transmission*, vol. 31, no. 2, pp. 99–113, 1995.
- [11] B. Hughes, “The Smallest List for the Arbitrarily Varying Channel,” *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 803–815, 1997.
- [12] I. Csiszár and P. Narayan, “Arbitrarily Varying Channels with Constrained Inputs and States,” *IEEE Transactions on Information Theory*, vol. 34, no. 1, pp. 27–34, 1988.
- [13] R. Ahlswede, “Channels with Arbitrarily Varying Channel Probability Functions in the Presence of Noiseless Feedback,” *Zeitschrift für Wahrscheinlichkeit und verwandte Gebiete*, vol. 25, pp. 239–252, 1973.
- [14] —, “Coloring Hypergraphs : A New Approach fo Multi-user Source Coding – I,” *Journal of Combinatorics, Information, and System Sciences*, vol. 4, no. 1, pp. 76–115, 1979.
- [15] —, “Coloring Hypergraphs : A New Approach fo Multi-user Source Coding – II,” *Journal of Combinatorics, Information and System Sciences*, vol. 5, no. 3, pp. 220–268, 1980.
- [16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest: Akadémiai Kiadó, 1982.
- [17] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York: Springer, 1998.
- [18] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [19] R. Ahlswede, “A method of coding and an application to arbitrarily varying channels,” *Journal of Combinatorics, Information and System Sciences*, vol. 5, pp. 10–35, 1980.
- [20] I. Csiszár and P. Narayan, “Capacity of the Gaussian Arbitrarily Varying Channel,” *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 18–26, 1991.
- [21] J. Gubner, “On the Deterministic-Code Capacity of the Multiple-Access Arbitrarily Varying Channel,” *IEEE Transactions on Information Theory*, vol. 36, no. 2, pp. 262–275, 1990.
- [22] —, “State Constraints for the Multiple-Access Arbitrarily Varying Channel,” *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 27–31, 1991.
- [23] S. S. (Shitz) and S. Verdú, “Worst-Case Power-Constrained Noise Binary-Input Channels,” *IEEE Transactions on Information Theory*, vol. 38, no. 5, pp. 1494–1511, 1992.
- [24] M. Agarwal, A. Sahai, and S. Mitter, “Coding into a source: a direct inverse rate-distortion theorem,” in *45th Annual Allerton Conference on Communication, Control and Computation*, 2006.
- [25] V. K. Wei, “Generalized Hamming Weights for Linear Codes,” *IEEE Transactions on Information Theory*, vol. 37, no. 5, pp. 1412–1418, 1991.
- [26] M. A. Tsfasman and S. G. Vlăduț, “Geometric Approach to Higher Weights,” *IEEE Transactions on Information Theory*, vol. 41, no. 6, pp. 1564–1588, 1995.
- [27] V. Guruswami, “List Decoding From Erasures: Bounds and Code Constructions,” *IEEE Transactions on Information Theory*, vol. 49, no. 11, pp. 2826–2833, 2003.
- [28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.