State constraints and list decoding for the AVC

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

Abstract

List decoding for arbitrarily varying channels (AVCs) under state constraints is investigated. It is shown that rates within ϵ of the randomized coding capacity of AVCs with input-dependent state can be achieved under maximal error with list decoding using lists of size $O(1/\epsilon)$. Under average error an achievable rate region and converse bound are given for lists of size L. These bounds are based on two different notions of symmetrizability and do not coincide in general. An example is given that shows that for list size L the capacity may be positive but strictly smaller than the randomized coding capacity. This behavior is different than the situation without state constraints.

I. Introduction

The arbitrarily varying channel (AVC) is a model for communication subject to time-varying interference [**?**]. The time variation is captured by a channel state parameter and coding schemes for these channels are required to give a guarantee on the probability of error for all channel state sequences. The AVC is thought of as an adversarial model in which the channel state is controlled by a *jammer* who wishes to foil the communication between the encoder and decoder.

This short paper addresses the problem of list-decoding in an AVC when the state sequence is constrained. The constraint comes by imposing a per-letter cost $l(\cdot)$ on the state sequence and requiring the cost of the state sequence chosen by the jammer for n channel uses to be less than a total budget Λn. The randomized and deterministic coding capacity for this AVC variant was found by Csiszár and Narayan [**?**], [**?**]. In particular, they showed that the deterministic coding capacity under average error $\bar{C}_d(\Lambda)$ may be positive but strictly smaller than the randomized coding capacity $C_r(\Lambda)$. This is a qualitatively

Manuscript received September 2009; revised XXXXXXXXXXXXXX.

A.D. Sarwate is with the Information Theory and Applications Center at the University of California, San Diego, La Jolla CA 92093-0447. M. Gastpar is 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.

different situation from AVCs without constraints [**?**], where $\bar{C}_d$ is either 0 or equal to C_r. They also showed that *symmetrizability* as defined by Ericson [**?**] is sufficient for $\bar{C}_d(\Lambda)$ to be positive [?].

In list-decoding, the decoder is allowed to output a list of L messages and an error is declared only if the list does not contain the transmitted message. For AVCs without constraints, list-decoding capacities have been investigated under both maximal and average error. For maximal error, Ahlswede [**?**], [**?**] found a quantity C_{dep} such a rate $C_{\mathrm{dep}} - \epsilon$ is achievable with lists of size $O(1/\epsilon)$. We extend this result to the situation with cost constraints and define a quantity $C_{\mathrm{dep}}(\Lambda)$ such that a rate $C_{\mathrm{dep}}(\Lambda) - \epsilon$ is achievable under list-decoding with list size $O(1/\epsilon)$. This result on maximal error can be used to find the randomized coding capacity of AVCs where the state can depend on the transmitted codeword as well as rateless code constructions [**?**].

The average error list-L capacity $\bar{C}_L$ without constraints was found independently by Blinovsky, Narayan, and Pinsker [**?**], [**?**] and Hughes [**?**]. These authors defined the symmetrizability $\hat{L}_{\mathrm{sym}}$ of an AVC and showed that there is a constant list size $\hat{L}_{\mathrm{sym}}$ so that for $L \le \hat{L}_{\mathrm{sym}}$ the list-L capacity is 0 and for $L > \hat{L}_{\mathrm{sym}}$ the list-L capacity is equal to the randomized coding capacity C_r. We show that under state constraints the behavior is qualitatively different. The ability of the jammer to symmetrize the channel depends on the input distribution P and the cost constraint Λ. We define two kinds of symmetrizability for list-decoding under state constraints. We show that for list size L the coding strategy of Hughes [?] can be used with input distributions P such that L is larger than the *weak symmetrizability* $\tilde{L}_{\mathrm{sym}}(P, \Lambda)$. We also prove a new converse for input distributions P such that L is smaller than the *strong symmetrizability* $L_{\mathrm{sym}}(P, \Lambda)$.

In general, $L_{\mathrm{sym}}(P, \Lambda) < \tilde{L}_{\mathrm{sym}}(P, \Lambda)$, which gives a gap between our achievable region and converse. Closing this gap seems non-trivial; we conjecture that the converse can be tightened. However, our results do imply a significant difference between the constrained and unconstrained setting. Without constraints, the list-L capacity $\bar{C}_L$ is either 0 or equal to the randomized coding capacity C_r. We show via a simple example that under cost constraints (analogous to [?]) the list-L capacity $\bar{C}_L(\Lambda)$ may be positive but strictly smaller than the randomized coding capacity $C_r(\Lambda)$.

II. Definitions and main results

We will use calligraphic type for sets and $[M] = \{1, 2, \ldots, M\}$ for integers M. For sets $\mathcal{X}$ and $\mathcal{Y}$, the set $\mathcal{P}(\mathcal{X})$ is the set of probability distributions on $\mathcal{X}$, $\mathcal{P}_n(\mathcal{X})$ is the set of all distributions of composition n, and $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ is the set of all conditional distributions on $\mathcal{Y}$ conditioned on $\mathcal{X}$. For random variables (X, Y) with joint distribution P_{XY} we will write P_X and P_Y for the marginal distributions and $P_{X|Y}$

for the conditional distribution of X given Y. For a distribution $\bar{P} \in \mathcal{P}(\mathcal{X}^m)$ we will denote by P_i the i-th marginal of $\bar{P}$. Let $d_{\max}(P, Q)$ be the maximum deviation (ℓ_∞ distance) between two probability distributions P and Q.

A. Channel model and codes

An 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 $s \in \mathcal{S}$, where all alphabets are finite. If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ and $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ are length n vectors, the probability of $\mathbf{y}$ given $\mathbf{x}$ and $\mathbf{s}$ is given by:

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

We are interested in the case where there is a bounded cost function $l : \mathcal{S} \to \mathbb{R}^+$ on the jammer. The cost of an n-tuple is

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

The state obeys a state constraint Λ if

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

An (n, N, L) *deterministic list code* C for the AVC is a pair of maps (ψ, ϕ) where the encoding function is $\psi : \{1, 2, \ldots, N\} \to \mathcal{X}^n$ and the decoding function is $\phi : \mathcal{Y}^n \to \{1, 2, \ldots, N\}^L$. 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 $\{(\mathbf{x}_i, D_i) : i = 1, 2, \ldots, N\}$, with the encoder and decoder implicitly defined.

The *maximal* and *average* error probabilities ε_L and $\bar{\varepsilon}_L$ are given by

$$\varepsilon_L = \max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \max_i \left(1 - W(D_i|X^n = \mathbf{x}_i, \mathbf{s})\right) \tag{4}$$

$$\bar{\varepsilon}_L = \max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \frac{1}{N} \sum_{i=1}^{N} \left(1 - W(D_i|\mathbf{x}_i, \mathbf{s})\right) \ . \tag{5}$$

A rate R is called achievable under maximal (average) list-decoding with list size L if for any $\epsilon > 0$ there exists a sequence of (n, N, L) list codes rate at least $R - \epsilon$ whose maximal (average) error converges to 0. The list-L capacity is the supremum of achievable rates. We denote the list-L capacities under maximal and average error by $C_L(\Lambda)$ and $\bar{C}_L(\Lambda)$, respectively.

B. Symmetrizability and information quantities

We call a channel $V(y|x_1, x_2, \ldots, x_m)$ from $\mathcal{X}^m$ to $\mathcal{Y}$ *symmetric* if for any permutation π on $[m]$,

$$V(y|x_1, x_2, \ldots, x_m) = V(y|x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(m)}) \quad \forall (x_1, x_2, \ldots, x_m, y) \ . \tag{6}$$

A channel $U(s|x_1, x_2, \ldots, x_m)$ *symmetrizes* an AVC $\mathcal{W}$ if

$$V(y|x, x_1, \ldots, x_m) = \sum_{s \in \mathcal{S}} W(y|x, s) U(s|x_1, x_2, \ldots, x_m) \tag{7}$$

is a symmetric channel. We denote by $\mathcal{U}_{\mathrm{sym}}(m)$ the set of channels which symmetrize $\mathcal{W}$:

$$\mathcal{U}_{\mathrm{sym}}(m) = \{U(s|x^m) : V(y|x, x_1, \ldots, x_m) \text{ is symmetric}\} \ . \tag{8}$$

Note that $\mathcal{U}_{\mathrm{sym}}$ is a convex subset of channels $U(s|x_1, \ldots, x_m)$ defined by equality constraints from (6).

For a distribution $P \in \mathcal{P}(\mathcal{X})$ we define the *strong symmetrizing cost* $\lambda_m(P)$ to be the smallest expected cost of a channel $U(s|x^m)$ that symmetrizes the AVC $\mathcal{W}$ whose input $\bar{P}(x^m)$ may be correlated but has marginals equal to P:

$$\lambda_m(P) = \min_{U \in \mathcal{U}_{\mathrm{sym}}(m)} \max_{\bar{P} \in \mathcal{P}(\mathcal{X}^m) : P_i = P} \sum_{x^m} \sum_{s} \bar{P}(x^m) U(s|x^m) l(s) \ . \tag{9}$$

We call an AVC *strongly* m*-symmetrizable* under the constraint Λ if $\lambda_m(P) \leq \Lambda$. We define the *strong symmetrizability* $L_{\mathrm{sym}}(P, \Lambda)$ of the channel under input P to be the largest integer m such that $\lambda_m(P) < \Lambda$. That is,

$$L_{\mathrm{sym}}(P, \Lambda) = \max \{m : \lambda_m(P) < \Lambda\} \ . \tag{10}$$

We define the *weak symmetrizing cost* $\tilde{\lambda}_m(P)$ to be the smallest expected cost of a channel $U(s|x^m)$ that symmetrizes the AVC $\mathcal{W}$ with independent inputs:

$$\tilde{\lambda}_m(P) = \min_{U \in \mathcal{U}_{\mathrm{sym}}(m)} \sum_{x^m} \sum_{s} P^m(x^m) U(s|x^m) l(s) \ , \tag{11}$$

where P^m is the product distribution $P \times P \times \cdots \times P$. We call an AVC *weakly* m*-symmetrizable* if $\tilde{\lambda}_m(P) \leq \Lambda$. Similarly, the *weak symmetrizability* $\tilde{L}_{\mathrm{sym}}(P, \Lambda)$ is the largest integer m such that $\tilde{\lambda}_m(P) < \Lambda$. That is,

$$\tilde{L}_{\mathrm{sym}}(P, \Lambda) = \max \left\{ m : \tilde{\lambda}_m(P) < \Lambda \right\} \ . \tag{12}$$

For a fixed input distribution $P(x)$ on $\mathcal{X}$ and channel $V(y|x)$, we will use the notation $I(P, V)$ to denote the mutual information between the input and output of the channel:

$$I(P, V) = \sum_{x,y} V(y|x) P(x) \log \frac{V(y|x) P(x)}{P(x) \sum_{x'} V(y|x') P(x')} \ . \tag{13}$$

We define the following two information sets:

$$\mathcal{Q}(\Lambda) = \left\{ Q \in \mathcal{P}(\mathcal{S}) : \sum_{s} l(s) Q(s) \le \Lambda \right\} \tag{14}$$

$$\mathcal{U}(P, \Lambda) = \left\{ U \in \mathcal{P}(\mathcal{S}|\mathcal{X}) : \sum_{s,x} U(s|x) P(x) l(s) \le \Lambda \right\} \ . \tag{15}$$

These in turn can be used to define two information quantities:

$$C_{\mathrm{std}}(\Lambda) = \max_{P \in \mathcal{P}(\mathcal{X})} \min_{Q \in \mathcal{Q}(\Lambda)} I\left(P, \sum_{s} W(y|x,s) Q(s)\right) \tag{16}$$

$$C_{\mathrm{dep}}(\Lambda) = \max_{P \in \mathcal{P}(\mathcal{X})} \min_{U \in \mathcal{U}(P,\Lambda)} I\left(P, \sum_{s} W(y|x,s) U(s|x)\right) \ . \tag{17}$$

C. Main results

Our first result extends the strategy of Ahlswede to the case of constrained AVCs under maximal error.

Theorem 1 (List decoding for maximal error): Let $\mathcal{W}$ be an arbitrarily varying channel with state cost function $l(s)$ and cost constraint Λ. Then for any $\epsilon > 0$ the rate

$$R = C_{\mathrm{dep}}(\Lambda) - \epsilon \tag{18}$$

is achievable under maximal error using list decoding with list size

$$L = O\left(\frac{1}{\epsilon}\right) \ . \tag{19}$$

Furthermore, the capacity $C_L(\Lambda)$ under maximal error using list decoding with list size L is bounded:

$$C_{\mathrm{dep}}(\Lambda) - O(L^{-1}) \le C_L(\Lambda) \le C_{\mathrm{dep}}(\Lambda) \ . \tag{20}$$

The proof is given in Appendix I. This result can be used together with a message authentication strategy [**?**] to show that $C_{\mathrm{dep}}(\Lambda)$ is the randomized coding capacity of AVCs with input-dependent state [?].

For average error we can show an achievable rate region and converse bound which in general do not coincide. Proofs of Theorems 2 and 3 are given in Appendix II. In both cases the results constrain the set of input distributions in $\mathcal{P}(\mathcal{X})$. The intuition for the converse is that for any codebook with codewords of type P, the jammer can choose a symmetrizing channel $U \in \mathcal{U}_{\mathrm{sym}}(L)$ such that the expected cost under any joint distribution with marginals equal to P is within the cost constraint. Operationally, the jammer chooses L codewords from the codebook and uses them as inputs to U to generate a state sequence $\mathbf{s}$ which satisfies the cost constraints.

Theorem 2 (Converse for average error): Let $\mathcal{W}$ be an arbitrarily varying channel with state cost function $l(\cdot)$ and cost constraint Λ. Then we have the following upper bound on $\bar{C}_L(\Lambda)$:

$$\bar{C}_L(\Lambda) \leq \max_{P \in \mathcal{P}(\mathcal{X}) : L_{\mathrm{sym}}(P,\Lambda) < L} \ \min_{Q \in \mathcal{Q}(\Lambda)} I\left(P, \sum_s W(y|x,s) Q(s) \right) \ . \tag{21}$$

For achievability we extend the coding strategy of Hughes [?] in a manner analogous to [?] to show an achievable rate for input distributions P such that $L > \tilde{L}_{\mathrm{sym}}(P,\Lambda)$.

Theorem 3 (Achievability for average error): Let $\mathcal{W}$ be an arbitrarily varying channel with state cost function $l(\cdot)$ and cost constraint Λ. Then we have the following lower bound on $\bar{C}_L(\Lambda)$:

$$\bar{C}_L(\Lambda) \geq \max_{P \in \mathcal{P}(\mathcal{X}) : \tilde{L}_{\mathrm{sym}}(P,\Lambda) < L} \ \min_{Q \in \mathcal{Q}(\Lambda)} I\left(P, \sum_s W(y|x,s) Q(s) \right) \ . \tag{22}$$

If P^* is the maximizing input distribution for $C_{\mathrm{std}}(\Lambda)$, then for list size $L > \tilde{L}_{\mathrm{sym}}(P^*, \Lambda)$ we have

$$\bar{C}_L(\Lambda) = C_{\mathrm{std}}(\Lambda) \ . \tag{23}$$

III. Example and discussion

We will now show via an example that the behavior of list-decoding under average error with state constraints is qualitatively different from that without constraints. In particular when the jammer must satisfy a constraint $\Lambda < \infty$, positive rates may be achievable with list sizes that are smaller than the unconstrained symmetrizability, and for a fixed list size the list-L capacity may be positive but strictly smaller than the randomized coding capacity. Let the input $\mathcal{X} = \{0,1\}$, state $\mathcal{S} = \{0, 1, \ldots, \sigma\}$ and the channel be defined by:

$$Y = X + S \ . \tag{24}$$

We will consider a quadratic cost function $l(s) = s^2$.

Without constraints, Hughes [?] has found that the randomized capacity is

$$C_r(\infty) = -\log \cos \frac{\pi}{\sigma + 3} \ . \tag{25}$$

He also showed that for unconstrained AVCs the list-L capacity obeys a strict threshold :

$$C_L(\infty) = \begin{cases} -\log \cos \frac{\pi}{\sigma+3} & L > \sigma \\ 0 & L \leq \sigma \end{cases} \tag{26}$$

We are interested in the case when there is a cost constraint Λ on the jammer. We must calculate the minimum mutual information for different input distributions:

$$I\left(P, \Lambda\right) = \min_{Q \in \mathcal{P}(\mathcal{S}) : \mathbb{E}_Q[l(s)] \leq \Lambda} I\left(X \ \wedge \ Y\right) \ . \tag{27}$$

The randomized-coding capacity under the cost constraint Λ is the max of $I\left(P, \Lambda\right)$ over P.

$$C_r(\Lambda) = \max_{P \in \mathcal{P}(\mathcal{X})} I\left(P, \Lambda\right) \ . \tag{28}$$

These calculations can be easily done numerically.

To calculate the symmetrizability constraints, note that the because the channel (24) is deterministic, the symmetry constraints imply that any channel $U \in \mathcal{U}_{\mathrm{sym}}$ must also be symmetric. Therefore $U(s|x_1, x_2, \ldots, x_L)$ is only a function of the type of $(x_1, x_2, \ldots, x_L)$. Let t denote this type. We now view $\mathcal{U}_{\mathrm{sym}}$ as containing channels $U(s|t)$. Note that for $y = 0$ we have

$$\sum_s W(0|0, s) U(s|t) = U(0|t) \ , \tag{29}$$

and by the symmetry constraint we have

$$U(0|t) = 0 \qquad t = 1, 2, \ldots, L \ . \tag{30}$$

Similarly, for $y = \sigma + 1$ we have

$$U(\sigma|t) = 0 \qquad t = 0, 1, \ldots, L-1 \ . \tag{31}$$

Finally, for $y = 1, 2, \ldots, \sigma$ we have

$$\sum_s W(y|0, s) U(s|t) = U(y|t) \tag{32}$$

$$= \sum_s W(y|1, s) U(s|t-1) \tag{33}$$

$$= U(y-1|t-1) \qquad y = 1, 2, \ldots, \sigma, \quad t = 1, 2, \ldots, L \tag{34}$$

The conditions (30), (31), and (34) characterize the linear symmetry constraints in $\mathcal{U}_{\mathrm{sym}}$.

Thus for each input distribution P we can find

$$f(P) = \min_{U \in \mathcal{U}_{\mathrm{sym}}} \sum_{s,t} l(s) U(s|t) \binom{L}{t} P(0)^{L-t} P(1)^t \ . \tag{35}$$

This is a simple linear program. To calculate the strong L-symmetrizing cost, note that the set of all joint distributions $\bar{P}(x_1^L)$ with marginals equal to P is also a convex set defined by linear equality constraints. If we let

$$\tau(\bar{P}, t) = \sum_{x_1^L : T_{\mathbf{x}} = t/L} \bar{P}(x_1^L) \ , \tag{36}$$

be the probability of a type-t sequence under $\bar{P}$, it is simple to numerically evaluate

$$g(P) = \max_{\bar{P}} \min_{U \in \mathcal{U}_{\mathrm{sym}}} \sum_{s,t} l(s) U(s|t) \tau(\bar{P}, t) \ . \tag{37}$$

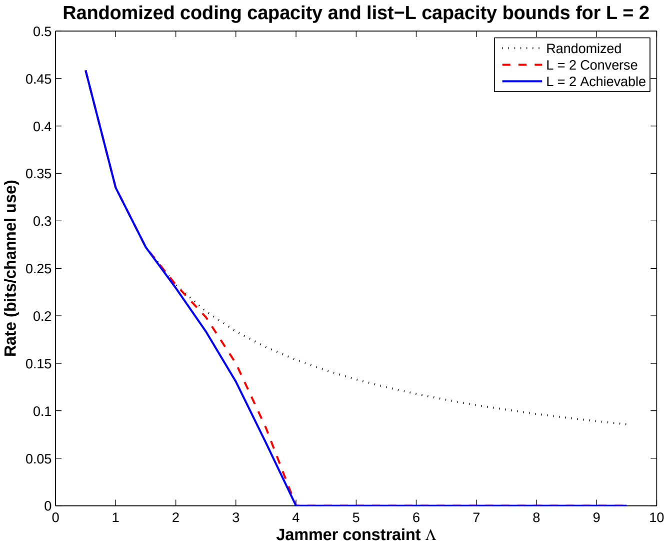

Fig. 1. Randomized coding capacity $C_r(\Lambda)$ and bounds on list-L capacity $\bar{C}_L(\Lambda)$ versus the state constraint Λ for $L = 2$.

We calculated the achievable rates and converse bounds for $\sigma = 8$, and the results are shown for list sizes $L = 2$ and $L = 4$ in Figures 1 and 2. For state constraint Λ, the randomized coding capacity $C_r(\Lambda)$ in (28) is given by the dotted line. The achievable rate of Theorem 3 is shown by the solid line, and the converse bound of Theorem 2 by the dashed line. These two curves are given by restricting the optimization over P in the right side of (28).

When $\Lambda = \infty$, the randomized coding capacity of this channel is given by (25) and is 0.0597 bits/channel use. Therefore, when $\Lambda = \infty$, the result in (26) shows that the the list-L capacity is 0 for $L < 8$ and equal to 0.0597 for $L > 8$. That is, when the jammer is unconstrained, no positive rate is achievable under average error using list decoding with list size smaller than 8. However, from Figures 1 and 2 we can see that when $\Lambda < \infty$ we can achieve positive rates for list sizes L smaller than 8. However, for a range of Λ, the randomized coding capacity is achievable using lists of size 2 or 4. Figure 1 also illustrates another fundamental difference between list-decoding with state constraints and list-decoding without constraints: for a range around $\Lambda = 3$, the list-2 capacity $\bar{C}_2(\Lambda)$ is positive but strictly smaller than the randomized coding capacity $C_r(\Lambda)$.

In general, we conjecture that the converse region of Theorem 2 is not tight and that a stronger converse could be shown. The strong symmetrizing cost in (9) allows optimization over all joint distributions with the same marginals. The converse proof uses a jamming strategy corresponding to taking a random set of L codewords from the codebook as inputs to a symmetrizing channel $U(s|x^L)$ to generate the state

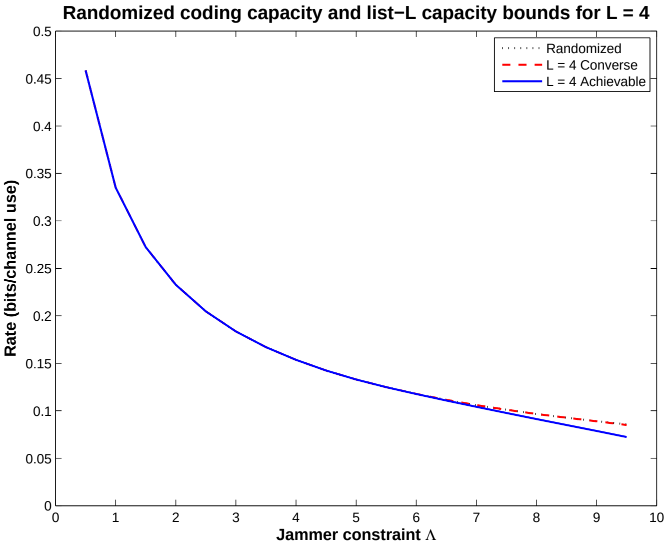

Fig. 2. Randomized coding capacity $C_r(\Lambda)$ and bounds on list-L capacity $\bar{C}_L(\Lambda)$ versus the state constraint Λ for $L = 4$.

sequence. The strong symmetrizing cost is a conservative bound on the cost of such a strategy. It may be that techniques such as [**?**] could improve this bound; we leave this for future work. Our results here establish that the behavior of list-decoding for constrained AVCs is fundamentally different than the unconstrained case, much like the situation for list size 1.

Appendix I
Maximal Error

Using now-standard typicality arguments we can show the existence of list-decodable codes for maximal error with exponential list size. The codebook is the entire set of typical sequences T_P and the list is the union of ϵ-shells under the different state sequences. The decoder outputs a list that is the union of shells. Let

$$\mathcal{W}_{dep}(P,\Lambda) = \left\{ V(y|x) : V(y|x) = \sum_s W(y|x,s)U(s|x), \quad U(s|x) \in \mathcal{U}(P,\Lambda) \right\} . \tag{38}$$

Proof: [Proof of Theorem 1] The converse argument follows by choosing $\mathbf{s}$ according to the minimizing distribution $U(s|x)$ in $\mathcal{U}(P,\Lambda)$. To show the achievable rate, without loss of generality, suppose that the distribution P maximizing $C_{\text{dep}}(\Lambda)$ is in $\mathcal{P}_n(\mathcal{X})$ and consider the set T_P of all sequences of length n of type P (if not we can always approach the optimal P with large n). For any $V(y|x)$ we define $V'(x|y)$ from $V(y|x)P(x)$ via the Bayes rule. The (V',ϵ)-shell of typical $\mathbf{x}$ sequences around a

$\mathbf{y}$ is:

$$T^{\epsilon}_{V'}(\mathbf{y}) = \left\{ \mathbf{x} \in T_P : d_{\max}\left(T_{\mathbf{x}y}, V' T_{\mathbf{y}} \right) < \epsilon \right\} \ . \tag{39}$$

Then

$$\frac{1}{n} \log |T^{\epsilon}_{V'}(\mathbf{y})| \le H_{V' T_{\mathbf{y}}}(X|Y) + O(\epsilon \log \epsilon^{-1}) \ , \tag{40}$$

where the subscript on H indicates the the joint distribution under which to take the mutual information.

Now, for a fixed $\mathbf{x} \in T_P$ and $\mathbf{s}$ with $l(\mathbf{s}) \le n\Lambda$, we define an empirical forward channel

$$V_{\mathbf{xs}}(y|x) = \sum_{s} W(y|x,s) \frac{N(x,s|\mathbf{x},\mathbf{s})}{N(x|\mathbf{x})} \ . \tag{41}$$

Note that $V_{\mathbf{xs}} \in \mathcal{W}_{dep}(P,\Lambda)$. For a fixed received codeword $\mathbf{y}$, define the set of channels consistent with $\mathbf{y}$ as:

$$\mathcal{V}^{\delta}_{P}(\mathbf{y}) = \left\{ V \in \mathcal{W}_{dep}(P,\Lambda) \cap \mathcal{P}_n(\mathcal{Y}|\mathcal{X}) : d_{\max}\left(\sum_{y} V(y|x)P(x), T_{\mathbf{y}} \right) < \delta \right\} \ . \tag{42}$$

Consider the set

$$\mathcal{A}(\mathbf{y}) = \bigcup_{V \in \mathcal{V}^{\delta}_{P}(\mathbf{y})} T^{(|\mathcal{X}|+1|)\delta}_{V'}(\mathbf{y}) \ . \tag{43}$$

Standard typicality arguments show that if $\mathbf{x}$ generated $\mathbf{y}$ via some $\mathbf{s}$ satisfying the cost constraint, then with probability $1 - \exp(-nE(\delta))$, we have $\mathbf{x} \in \mathcal{A}(\mathbf{y})$. Furthermore:

$$\frac{1}{n} \log |\mathcal{A}(\mathbf{y})| \le \min_{V \in \mathcal{W}_{dep}(P,\Lambda)} H_{V(y|x)P(x)}(X|Y) + O(\delta \log \delta^{-1}) \ . \tag{44}$$

Note that we can view an encoding into all of T_P and decoding into $\mathcal{A}(\mathbf{y})$ as a list-decodable code with $2^{nH(P)}$ codewords and list size (44). To arrive at the desired code we can sample a set $\mathcal{B} = \{\mathbf{x}(i)\}$ of $2^{n(C_{\mathrm{dep}}(\Lambda)-\epsilon)}$ codewords from this T_P uniformly at random and say the decoder outputs $\mathcal{A}(\mathbf{y}) \cap \mathcal{B}$. We must show this set has at most $L = O(1/\epsilon)$ codewords with high probability.

Let $R = C_{\mathrm{dep}}(\Lambda) - \epsilon$. For each $\mathbf{y}$, the probability that any codeword of B is in $\mathcal{A}(\mathbf{y})$ is upper bounded by $|\mathcal{A}(\mathbf{y})|/|T_P|$, so from (44) we see

$$\mathbb{P}\left(\mathbf{x}(i) \in \mathcal{A}(\mathbf{y})\right) \le \exp\left(-n \left(C_{\mathrm{dep}}(\Lambda) - O(\delta \log \delta^{-1}) \right) \right) \ . \tag{45}$$

Since codewords are selected independently, we can bound the chance that a fraction $L \cdot 2^{-nR}$ of the 2^{nR} codewords end up in $\mathcal{A}(\mathbf{y})$ using Sanov's theorem [**?**, Theorem 12.4.1]

$$\mathbb{P}\left(|\mathcal{A}(\mathbf{y}) \cap \mathcal{B}| > L \right) \le \exp\left(-2^{nR} D\left(L 2^{-nR} \ \middle\| \ 2^{-n(C_{\mathrm{dep}}(\Lambda) - O(\delta \log \delta^{-1})} \right) + h \log(2^{nR}+1) \right) \tag{46}$$

Now we can bound the term $2^{nR} D\left(\cdot \ \| \ \cdot\right)$:

$$-L \log \frac{L}{2^{n(\epsilon - O(\delta \log \delta^{-1}))}} - 2^{nR}(1 - L 2^{-nR}) \log \frac{1 - L 2^{-nR}}{1 - 2^{-n(R + \epsilon - O(\delta \log \delta^{-1}))}} \tag{47}$$

$$\leq -nL\left(\epsilon - O(\delta \log \delta^{-1})\right) - L \log L + 2L \ . \tag{48}$$

We can pick δ such that $O(\delta \log \delta^{-1}) < \epsilon/2$ by choosing n sufficiently large. Then substituting (48) in (46), upper bounding $R < \log |\mathcal{Y}|$, and taking a union bound over all $\mathbf{y}$ we have:

$$\mathbb{P}\left(\exists \mathbf{y} \ : \ |\mathcal{A}(\mathbf{y}) \cap \mathcal{B}| > L\right) \leq \exp\left(-n\left(L\epsilon/2 + 2\log|\mathcal{Y}|\right) - L \log L + 2L\right) \ . \tag{49}$$

For sufficiently large n choosing $L > \lceil \frac{4 \log |\mathcal{Y}|}{\epsilon} \rceil$ makes the exponent negative, showing that with high probability the random selection will produce an $(n, 2^{nR}, L)$ list-decodable code under maximal error whose error is bounded by $1 - \exp(-nE(\delta))$. ■

Appendix II
Average Error

A. Facts about symmetrizability

The following theorem shows that if $I(P)$ is positive, then $\tilde{L}_{\mathrm{sym}}(P, \Lambda)$ is finite. In particular, since $I\left(P^*, \Lambda\right)$ is finite, the theorem implies that if $C_{\mathrm{std}}(\Lambda) > 0$, then $\tilde{L}_{\mathrm{sym}}(P^*, \Lambda) < \infty$. The proof follows straightforwardly from the results of [?].

Lemma 1 (Finite symmetrizability): Let $\mathcal{W}$ be an arbitrarily varying channel with state cost function $l(\cdot)$. If $C_{\mathrm{std}}(\Lambda) = 0$ then $L_{\mathrm{sym}}(P, \Lambda) = \infty$ for all P. If $C_{\mathrm{std}}(\Lambda) > 0$ then

$$\tilde{L}_{\mathrm{sym}}(P, \Lambda) \leq \frac{\log(\min(|\mathcal{Y}|, |\mathcal{S}|))}{I\left(P, \Lambda\right)} \tag{50}$$

for all P such that $I\left(P, \Lambda\right) > 0$.

B. Achievability under average error

Given a P that is not weakly L-symmetrizable, we can use the coding scheme of Hughes [?] modified in the natural way suggested by Csiszár and Narayan [?] for list size 1. The codebook consists of N constant-composition codewords drawn uniformly from the codewords of type P. In order to describe the decoding rule we will use, we define the set

$$\mathcal{G}_{\eta}(\Lambda) = \left\{P_{XSY} \in \mathcal{P}(\mathcal{X} \times \mathcal{S} \times \mathcal{Y}) : D\left(P_{XSY} \ \| \ P_X \times P_S \times W\right) \leq \eta, \ \mathbb{E}[l(s)] \leq \Lambda\right\} \ , \tag{51}$$

where

$$(P_X \times P_S \times W)(x,s,y) = P_X(x) P_S(s) W(y|x,s) \ . \tag{52}$$

The set $\mathcal{G}_\eta(\Lambda)$ contains joint distributions which are close to those generated from the AVC $\mathcal{W}$ via independent inputs with distribution P_X and P_S.

Definition 1 (Decoding rule): Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ be a given codebook and suppose $\mathbf{y}$ was received. Let $\psi(\mathbf{y})$ denote the list decoded from $\mathbf{y}$. Then put $i \in \psi(\mathbf{y})$ if and only if there exists an $\mathbf{s} \in \mathcal{S}^n(\Lambda)$ such that

1) $T_{\mathbf{x}_i \mathbf{s} \mathbf{y}} \in \mathcal{G}_\eta(\Lambda)$, and
2) for every set of L other distinct codewords $\{\mathbf{x}_j : j \in J,\ J \subset [N] \setminus \{i\},\ |J| = L\}$ such that there exists a set $\{\mathbf{s}_j : \mathbf{s}_j \in \mathcal{S}^n(\Lambda),\ j \in J\}$ with $T_{\mathbf{x}_j \mathbf{s}_j \mathbf{y}} \in \mathcal{G}_\eta(\Lambda)$ for all $j \in J$ we have

$$I\left(YX \ \wedge\ X^L \middle| S\right) \le \eta \ , \tag{53}$$

where P_{YXX^LS} is the joint type of $(\mathbf{y}, \mathbf{x}_i, \{\mathbf{x}_j : j \in J\}, \mathbf{s})$.

An interpretation of this rule is that the decoder outputs a list of codewords $\{\mathbf{x}_i\}$ each having a "good explanation" $\{\mathbf{s}_i\}$. A "good explanation" is a state sequence that plausibly could have generated the observed output $\mathbf{y}$ (condition 1) and makes all other L-tuples of codewords seem independent of the codeword and output (condition 2). The only thing to prove is that this decoding rule is unambiguous. The key is to show that no tuple of random variables (Y, X^{L+1}, S^{L+1}) can satisfy the conditions of the decoding rule. This in turn shows that for sufficiently large n, no set of $L+1$ *codewords* can satisfy the conditions of the decoding rule. Therefore, for sufficiently large blocklengths, the decoding rule will only output M or fewer codewords.

Lemma 2: Let $\beta > 0$, $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$ and constraint Λ, $P \in \mathcal{P}(\mathcal{X})$ with $I(P,\Lambda) > 0$ and $\min_x P(x) \ge \beta$, and $M = \tilde{L}_{\mathrm{sym}}(P,\Lambda) + 1$. For any $\alpha > 0$ and every collection of distributions $\{U_i \in \mathcal{P}(\mathcal{X}^M \times \mathcal{S}) : i = 1, 2, \ldots, M\}$ such that

$$\sum_{x^{M+1}, s} P(x_i) U_i(x^M_{-\{i\}}, s) l(s) \le \tilde{\lambda}_M(P) - \alpha \tag{54}$$

for all $i = 1, 2, \ldots, M+1$, there exists a $\zeta > 0$ such that

$$\max_{j \neq i} \sum_{y, x^{M+1}} \left| \sum_s W(y|x_i, s) U_i(x^{M+1}_{-\{i\}}, s) P(x_i) - \sum_s W(y|x_j, s) U_j(x^{M+1}_{-\{j\}}, s) P(x_j) \right| \ge \zeta \ . \tag{55}$$

Proof: Note that the outer sum in (55) is over all x^{M+1}. Define the function $V_k : \mathcal{X}^{M+1} \times \mathcal{S} \to \mathbb{R}$ by:

$$V_k(x^{M+1}, s) = U_k(x^{M+1}_{-\{k\}}, s) \ . \tag{56}$$

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

$$\begin{aligned}
&\max_{j \neq i} \sum_{y, x^{M+1}} \left| \sum_s W(y|x_i, s) V_i(x^{M+1}, s) P(x_i) - \sum_s W(y|x_j, s) V_j(x^{M+1}, s) P(x_j) \right| \\
&= \max_{j \neq i} \sum_{y, x^{M+1}} \left| \sum_s W(y|x_i, s) V_{\pi_i}(\pi(x^{M+1}), s) P(x_i) \right. \\
&\qquad \left. - \sum_s W(y|x_j, s) V_{\pi_j}(\pi(x^{M+1}), s) P(x_j) \right| . \qquad (57)
\end{aligned}$$

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

$$\begin{aligned}
&\max_{j \neq i} \sum_{y, x^{M+1}} \frac{1}{(M+1)!} \sum_{\pi \in \Pi_{M+1}} \left| \sum_s W(y|x_i, s) V_{\pi_i}(\pi(x^{M+1}), s) P(x_i) \right. \\
&\qquad \left. - \sum_s W(y|x_j, s) V_{\pi_j}(\pi(x^{M+1}), s) P(x_j) \right| . \qquad (58)
\end{aligned}$$

Define the average

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

Note that $\bar{V}$ is a symmetric function for all s.

Now we use the convexity of $|\cdot|$ to pull the averaging inside the absolute value to get a further lower bound on (58) by substituting in $\bar{V}$.

$$\begin{aligned}
F(\bar{V}, P) = &\max_{j \neq i} \sum_{y, x^{M+1}} \left| \sum_s W(y|x_i, s) \bar{V}(x_{-\{i\}}^{M+1}, s) P(x_i) \right. \\
&\qquad \left. - \sum_s W(y|x_j, s) \bar{V}(x_{-\{j\}}^{M+1}, s) P(x_j) \right| . \qquad (59)
\end{aligned}$$

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 $\zeta > 0$ 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 [?, Lemma A3]) that for all j:

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

Therefore

$$\sum_s W(y|x_1,s)\bar{V}^*(s|x_2^{M+1}) \ . \tag{60}$$

is symmetric in $(x_1, x_2, \dots, x_{M+1})$. Therefore $\bar{V}^*(s|x_2^{M+1}) \in \mathcal{U}_{\text{sym}}(M+1)$. From the definition of $\tilde{\lambda}_M(P)$ in (11) we see that

$$\sum_{x^{M+1},s} \bar{V}^*(x_{-\{i\}}^{M},s)P(x_i)l(s) \geq \tilde{\lambda}_M(P) \ . \tag{61}$$

But from (54), and the definition of $\bar{V}$ we see that the $\{U_i\}$ must be chosen such that

$$\sum_{x^{M+1},s} \bar{V}^*(x_{-\{i\}}^{M},s)P(x_i)l(s) \leq \tilde{\lambda}_M(P) - \alpha \ . \tag{62}$$

Therefore we have a contradiction and the minimum ζ of $F(\bar{V},P)$ must be greater than 0. Equation (55) follows. ■

The next lemma shows that for a sufficiently small choice of the threshold η in the decoding rule there are no random variables that can force the decoding rule to output a list that is too large. The proof follows from Lemma 2 in the same way as in [?].

Lemma 3: Let $\beta > 0$, $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$ and constraint Λ, $P \in \mathcal{P}(\mathcal{X})$ with $\min_x P(x) \geq \beta$, and $M = \tilde{L}_{\text{sym}}(P,\Lambda) + 1$. Then there exists an $\eta > 0$ sufficiently small such that no tuple of rv's (Y, X^{M+1}, S^{M+1}) can simultaneously satisfy

$$\min_x P(x) \geq \beta \tag{63}$$

$$P_{X_i} = P \tag{64}$$

$$P_{YX_iS_i} \in \mathcal{G}_\eta(\Lambda) \tag{65}$$

$$I\left(YX_i \ \wedge \ X_{-\{i\}}^{M+1} \middle| S_i\right) \leq \eta \quad 1 \leq i \leq M+1 \tag{66}$$

Proof: [Proof of Theorem 3] Given Lemma 3 the theorem follows from Lemma 3 of [?]. ■

C. *Converse*

The key idea in the converse is to show that for a codebook with codewords whose types are symmetrizable and close to a fixed symmetrizable type P, then the jammer has a strategy that keeps the error bounded away from 0. The rest follows from approximation and covering arguments.

Lemma 4 (Approximating joint distributions): Let $\mathcal{X}$ be a finite set with $|\mathcal{X}| \geq 2$. For any $\epsilon > 0$ and probability distribution P on $\mathcal{X}$ there exists a $\delta > 0$ such that for any collection of distributions $\{P_i \in \mathcal{P}(\mathcal{X}) : i \in [L]\}$ satisfying

$$d_{\max}\left(P_i, P\right) < \delta \qquad \forall i \tag{67}$$

and any joint distribution $\bar{P}(x_1, x_2, \ldots, x_L)$ with

$$\sum_{x_j : j \neq i} \bar{P}(x_1, x_2, \ldots, x_L) = P_i(x_i) \qquad \forall i,\ x_i \in \mathcal{X} \tag{68}$$

there exists a joint distribution $\hat{P}(x_1, x_2, \ldots, x_L)$ such that

$$\sum_{x_j : j \neq i} \hat{P}(x_1, x_2, \ldots, x_L) = P(x_i) \qquad \forall i,\ x_i \in \mathcal{X} \tag{69}$$

and

$$d_{\max}\left(\bar{P}, \hat{P}\right) < \epsilon\ . \tag{70}$$

Proof: [Proof of Lemma 4] Fix $\epsilon > 0$ and P. We consider two cases depending on whether $\min_{x \in \mathcal{X}} P(x) = 0$ or not.

Case 1. First suppose $\min_{x \in \mathcal{X}} P(x) = \beta > 0$. Consider a set of distributions $\{P_i : i \in [L]\}$ satisfying (67) and let $\bar{P}(x_1^L)$ be a joint distribution satisfying (68). We treat probability distributions as vectors in $\mathbb{R}^{|\mathcal{X}|^L}$. We can construct a distribution $\hat{P}$ satisfying (69) and (70) in two steps: first we project $\bar{P}$ onto the set of all vectors whose entries sum to 1 and satisfy (69), and then we find a $\hat{P}$ close to this projection which is a proper probability distribution.

Let $\mathcal{B}$ be the subspace of $\mathbb{R}^{|\mathcal{X}|^L}$ of all vectors P' satisfying the marginal constraints (69) as well as the sum probability constraint

$$\sum_{x_1^L} P'(x_1^L) = 1\ . \tag{71}$$

We can summarize these linear constraints in the matrix form

$$AP' = b'\ , \tag{72}$$

where A contains the coefficients on the left-hand sides of the constraints (69) and (71) and b' has the right-hand sides. We can assume A has full row-rank by removing linearly dependent constraints. Note that the distribution $\bar{P}$ satisfies

$$A\bar{P} = \bar{b} \ , \tag{73}$$

where $\bar{b}$ has the right-hand sides of (68) instead of (69).

Now let $\tilde{P}$ be the Euclidean projection of $\bar{P}$ onto the subspace $\mathcal{B}$:

$$\tilde{P} = \bar{P} + A^T(AA^T)^{-1}(b' - A\bar{P}) \ . \tag{74}$$

The error in the projection is

$$\bar{P} - \tilde{P} = A^T(AA^T)^{-1}(A\bar{P} - b') \tag{75}$$

$$= A^T(AA^T)^{-1}(\bar{b} - b') \ . \tag{76}$$

From (67) we can see that all elements of $(\bar{b} - b')$ are in $(-\delta, \delta)$. Since the rows of A are linearly independent, the singular values of A are strictly positive and a function of $|\mathcal{X}|$ and L only. Therefore there is a function $\mu_1(|\mathcal{X}|, L)$ such that

$$\left\| A^T(AA^T)^{-1}(\bar{b} - b') \right\|_2 < \mu_1(|\mathcal{X}|, L) \cdot \delta \ . \tag{77}$$

Since $|\mathcal{X}|$ is finite there is a function $\mu_2(|\mathcal{X}|, L)$ such that

$$d_{\max}\left(\tilde{P}(x_1^L), \bar{P}(x_1^L)\right) < \mu_2(|\mathcal{X}|, L) \cdot \delta \ . \tag{78}$$

If the resulting $\tilde{P}$ from this first projection has all nonnegative entries, then we set $\hat{P} = \tilde{P}$ and choose δ sufficiently small so that $\mu_2(|\mathcal{X}|, L) \cdot \delta < \epsilon$.

If $\tilde{P}$ has entries that are not in $[0, 1]$ then it is not a valid probability distribution. However, since $\bar{P}$ is a probability distribution, we know that

$$\min_{x_1^L} \tilde{P}(x_1^L) > -\mu_2(|\mathcal{X}|, L) \cdot \delta \ . \tag{79}$$

Let P^L be the joint distribution on $\mathcal{X}^L$ with independent marginals P:

$$P^L(x_1, \ldots, x_L) = P(x_1) \cdots P(x_L) \ . \tag{80}$$

Since $\min_x P(x) > \beta$ we have $P^L(x_1^L) > \beta^L$ for all L. Let

$$\alpha = \frac{\mu_2(|\mathcal{X}|, L) \cdot \delta}{\beta^L} \ , \tag{81}$$

and set

$$\hat{P} = (1-\alpha)\tilde{P} + \alpha P^L \ . \tag{82}$$

Then $\hat{P}(x_1^L) > 0$ for all x_1^L and by the triangle inequality:

$$d_{\max}\left(\bar{P}, \hat{P}\right) \leq d_{\max}\left(\bar{P}, \tilde{P}\right) + d_{\max}\left(\tilde{P}, \hat{P}\right) \tag{83}$$

$$< \mu_2(|\mathcal{X}|, L) \cdot \delta + \alpha d_{\max}\left(\tilde{P}, P^L\right) \tag{84}$$

$$< \left(1 + \frac{1}{\beta^L}\right) \mu_2(|\mathcal{X}|, L) \cdot \delta \ . \tag{85}$$

Therefore for δ sufficiently small, we can choose a $\hat{P}$ such that $d_{\max}\left(\bar{P}, \hat{P}\right) < \epsilon$ for any $\epsilon > 0$.

Case 2. We turn now to the second case. Suppose that $\min_{x \in \mathcal{X}} P(x) = 0$. Let $\mathcal{X}_0 = \{x \in \mathcal{X} : P(x) = 0\}$ and $\mathcal{Z} = \mathcal{X} \setminus \mathcal{X}_0$. Let $Q \in \mathcal{P}(\mathcal{Z})$ be the restriction of P to $\mathcal{Z}$. Then Q is a probability distribution on $\mathcal{Z}$. First suppose that $|\mathcal{Z}| = 1$. Then $P(x) = 1$ for some $x \in \mathcal{X}$. Let

$$\hat{P}(x_1^L) = P(x_1) \cdots P(x_L) \ . \tag{86}$$

Since all the marginal distributions P_i of $\bar{P}$ satisfy $d_{\max}(P, P_i) < \delta$ we know that $d_{\max}\left(\bar{P}, \hat{P}\right) < \delta$.

Now suppose $|\mathcal{Z}| \geq 2$. We can construct $\hat{P}$ by first finding a a joint distribution $\bar{Q}$ that is close to $\bar{P}$ and then invoking the first case of this proof on $\bar{Q}$. From (67) we know that for some $c > 0$ we have

$$\sum_{x_1^L \notin \mathcal{Z}^L} \bar{P}(x_1, x_2, \ldots, x_L) \triangleq c\delta \tag{87}$$

$$< |\mathcal{X}|^L \delta \ . \tag{88}$$

Define $\bar{Q}$ by

$$\bar{Q}(x_1^L) = \begin{cases} \bar{P}(x_1^L) + |\mathcal{Z}|^{-L} c\delta & x_1^L \in \mathcal{Z}^L \\ 0 & x_1^L \notin \mathcal{Z}^L \end{cases} \tag{89}$$

Since $\bar{Q}$ has support only on $\mathcal{Z}^L$ we can think of it either as a distribution on $\mathcal{X}^L$ or on $\mathcal{Z}^L$. Note that

$$d_{\max}\left(\bar{P}, \bar{Q}\right) < c\delta \ . \tag{90}$$

Let $\{Q_i : i \in [L]\}$ be the i-th marginal distributions of $\bar{Q}$:

$$Q_i(x_i) = \sum_{x_j : j \neq i} \bar{Q}(x_1, x_2, \ldots, x_L) = Q_i(x_i) \qquad \forall i, \ x_i \in \mathcal{Z} \ . \tag{91}$$

Then we have for some $c' > 0$

$$d_{\max}(Q, Q_i) < c'\delta \ . \tag{92}$$

Now we can apply Case 1 of this proof using the set $\mathcal{Z}$ and distributions Q, $\{Q_i\}$, and $\bar{Q}$. For any $\epsilon_1 > 0$ we can find a $\delta_1 > 0$ such that if $\{Q_i\}$ satisfy

$$d_{\max}\left(Q, Q_i\right) < \delta_1 \ , \tag{93}$$

then there exists a $\hat{Q}$ with marginals equal to Q such that

$$d_{\max}\left(\bar{Q}, \hat{Q}\right) < \epsilon_1 \ . \tag{94}$$

Let $\hat{P}$ be the extension of $\hat{Q}$ to a distribution on $\mathcal{X}^L$ by setting $\hat{P}(x_1^L) = \hat{Q}(x_1^L)$ for $x_1^L \in \mathcal{Z}^L$ and 0 elsewhere. By the triangle inequality we have

$$d_{\max}\left(\bar{P}, \hat{Q}\right) \le d_{\max}\left(\bar{P}, \bar{Q}\right) + d_{\max}\left(\bar{Q}, \hat{Q}\right) \tag{95}$$

$$< c\delta + \epsilon_1 \ . \tag{96}$$

We can choose δ sufficiently small so that δ_1 and ϵ_1 are sufficiently small to guarantee that this distance is less than ϵ. ■

Lemma 5: Let $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$ and constraint Λ and let L be a positive integer. Let $\epsilon > 0$ be arbitrary and suppose P is a distribution with $\lambda_L(P) < \Lambda - \epsilon$. Then there exists a $\delta > 0$ and n_0 such that for any (n, N, L) list code with $n \ge n_0$ and $N \ge L + 1$ whose codewords $\{\mathbf{x}(i) : i \in [N]\}$ satisfy

$$d_{\max}\left(T_{\mathbf{x}(i)}, P\right) < \delta \qquad \forall i \in [N] \tag{97}$$

$$\lambda_L(T_{\mathbf{x}(i)}) < \Lambda - \epsilon \qquad \forall i \in [N] \ , \tag{98}$$

the average error for the code is lower bounded:

$$\max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \bar{\varepsilon}_L(\mathbf{s}) > \frac{1}{L+1} - \frac{L}{N(L+1)} \ . \tag{99}$$

Proof: From Lemma 4 we can see that for any $\epsilon_1 > 0$ there exists a $\delta_1 > 0$ such that for any set $J \subset [N]$ of codewords with $|J| = L$ and $d_{\max}\left(T_{\mathbf{x}(j)}, P\right) < \delta_1$, we can find a joint type $\bar{P} \in \mathcal{P}(\mathcal{X}^L)$ with marginals equal to P such that the joint type $T_{\mathbf{x}(J)}$ satisfies

$$d_{\max}\left(T_{\mathbf{x}(J)}, \bar{P}\right) < \epsilon_1 \ . \tag{100}$$

Now let U achieve the minimum in the definition of $\lambda_L(P)$. Since $\lambda_L(P) < \Lambda - \epsilon$ we have

$$\sum_{s, x_1^L} l(s) U(s|x_1^L) T_{\mathbf{x}(J)}(x_1^L) \le \sum_{s, x_1^L} l(s) U(s|x_1^L) \bar{P}(x_1^L) + \epsilon_1 \lambda^* |\mathcal{X}|^L \tag{101}$$

$$< \Lambda - \epsilon + \epsilon_1 \lambda^* |\mathcal{X}|^L \ , \tag{102}$$

where $\lambda^* = \max_{s \in \mathcal{S}} l(s)$. Now choose $\epsilon_1 = \epsilon/(2\lambda^* |\mathcal{X}|^L)$ so that

$$\sum_{s,x_1^L} l(s) U(s|x_1^L) T_{\mathbf{x}(J)}(x_1^L) < \Lambda - \epsilon/2 \ , \tag{103}$$

and choose $\delta = \delta_1$ according to Lemma 4.

The jammer will pick a $J \subset [N]$ with $|J| = L$ uniformly from all such subsets and select its state sequence according to the random variable $\mathbf{S}(J)$ with distribution

$$Q^n(\mathbf{s}) = \prod_{t=1}^{n} U(s_t | \{x_t(j) : j \in J\}) \ . \tag{104}$$

The expected cost of $\mathbf{S}(J)$ is

$$\begin{aligned} \frac{1}{n}\mathbb{E}[l(\mathbf{S}(J))] &= \frac{1}{n} \sum_{t=1}^{n} \sum_{\mathbf{s}} l(s_t) U(s_t | \{x_t(j) : j \in J\}) && (105) \\ &= \sum_{s,\tilde{x}^L} l(s) U(s|\tilde{x}_1, \ldots, \tilde{x}_L) \frac{|\{t : x_t(j) = \tilde{x}_j \ \forall j\}|}{n} && (106) \\ &= \sum_{s,\tilde{x}^L} l(s) U(s|\tilde{x}_1^L) T_{\mathbf{x}(J)} && (107) \\ &< \Lambda - \epsilon/2 \ . && (108) \end{aligned}$$

We can also bound the variance of $l(\mathbf{S}(J))$:

$$\mathrm{Var}\left(l(\mathbf{S}(J))\right) \le \frac{(\lambda^*)^2}{n} \ . \tag{109}$$

Then Chebyshev's inequality gives the bound:

$$\begin{aligned} \mathbb{P}(l(\mathbf{S}(U_J, J)) > \Lambda) &\le \frac{(\lambda^*)^2}{n(\Lambda - (\Lambda - \epsilon/2))^2} && (110) \\ &\le \frac{4(\lambda^*)^2}{n\epsilon^2} \ . && (111) \end{aligned}$$

We now need some properties of symmetrizing channels used with the random variables $\mathbf{S}(J)$. Firstly, we have:

$$\begin{aligned} \mathbb{E}\left[W^n(\mathbf{y}|\mathbf{x}(i), \mathbf{S}(J))\right] &= \sum_{\mathbf{s}} W^n(\mathbf{y}|\mathbf{x}(i), \mathbf{s}) U^n(\mathbf{s}|\{x(j) : j \in J\}) && (112) \\ &= \mathbb{E}\left[W^n(\mathbf{y}|\mathbf{x}(j), \mathbf{S}(J \setminus \{j\} \cup \{i\}))\right] \ . && (113) \end{aligned}$$

Using (113) we can see that for some subset $G \subset [N]$ with $|G| = L + 1$:

$$\begin{aligned} \sum_{i \in G} \mathbb{E}\left[\bar{\varepsilon}_L(i, \mathbf{S}(G \setminus \{i\}))\right] &= \sum_{i \in G} \left(1 - \sum_{\mathbf{y}: i \in \psi(\mathbf{y})} \mathbb{E}\left[W^n(\mathbf{y}|\mathbf{x}_i, \mathbf{S}(G \setminus \{i\}))\right] \right) && (114) \\ &= L + 1 - \sum_{i \in G} \sum_{\mathbf{y}: i \in \psi(\mathbf{y})} \mathbb{E}\left[W^n(\mathbf{y}|\mathbf{x}_{i_0}, \mathbf{S}_{G \setminus \{i_0\}})\right] \ . && (115) \end{aligned}$$

Because each $\mathbf{y}$ can be decoded to a list of size at most L , we can get a lower bound

$$\begin{aligned}\sum_{i\in G} \mathbb{E}\left[\varepsilon(i, \mathbf{S}_{G\setminus\{i\}})\right] &\geq L+1-L \sum_{\mathbf{y}\in\mathcal{Y}^n} \mathbb{E}\left[W^n(\mathbf{y}|\mathbf{x}_{i_0}, \mathbf{S}_{G\setminus\{i_0\}})\right] \\ &= 1 \ . \qquad (116)\end{aligned}$$

We can now begin to bound the probability of error for this jamming strategy. Let $\mathcal{J}$ be the set of all subsets of $[N]$ of size L, and let $\mathbf{J}$ be a random variable uniformly distributed on $\mathcal{J}$. We can write the expected error as

$$\mathbb{E}_{\mathbf{J},\mathbf{S}(\mathbf{J})}\left[\bar{\varepsilon}_L(\mathbf{S}(\mathbf{J}))\right] = \frac{1}{\binom{N}{L}} \frac{1}{N} \sum_{J\in\mathcal{J}} \sum_{i=1}^{N} \mathbb{E}\left[\bar{\varepsilon}_L(i, \mathbf{S}(J))\right] \ . \qquad (117)$$

Then we have:

$$\mathbb{E}_{\mathbf{J},\mathbf{S}(U_{\mathbf{J}},\mathbf{J})}\left[\bar{\varepsilon}_L(\mathbf{S}(U_{\mathbf{J}},\mathbf{J}))\right] \geq \frac{1}{\binom{N}{L}} \frac{1}{N} \sum_{G\subset[N]:|G|=L+1} \sum_{i\in G} \mathbb{E}\left[\bar{\varepsilon}_L(i, \mathbf{S}(G\setminus\{i\}))\right] \ . \qquad (118)$$

Now we can rewrite the inner sum using (113):

$$\begin{aligned}\mathbb{E}_{\mathbf{J},\mathbf{S}(\mathbf{J})}\left[\bar{\varepsilon}_L(\mathbf{S}(\mathbf{J}))\right] &\geq \frac{\binom{N}{L+1}}{\binom{N}{L}\cdot N} && (119)\\ &= \frac{\binom{N}{L}\frac{N-L}{L+1}}{\binom{N}{L}\cdot N} && (120)\\ &= \frac{N-L}{(L+1)N} && (121)\\ &= \frac{1}{L+1} - \frac{L}{N(L+1)} \ . && (122)\end{aligned}$$

Finally, we can add in the bound (111) to obtain

$$\begin{aligned}\frac{1}{L+1} - \frac{L}{N(L+1)} &\leq \mathbb{E}_{\mathbf{J},\mathbf{S}(\mathbf{J})}\left[\bar{\varepsilon}_L(\mathbf{S}(\mathbf{J}))\right] && (123)\\ &\leq \max_{\mathbf{s}\in\mathcal{S}^n(\Lambda)} \bar{\varepsilon}_L(\mathbf{s}) + \mathbb{P}\left(l(\mathbf{S}(\mathbf{J})) > \Lambda\right) && (124)\\ &\leq \max_{\mathbf{s}\in\mathcal{S}^n(\Lambda)} \bar{\varepsilon}_L(\mathbf{s}) \frac{4(\lambda^*)^2}{n\epsilon^2} \ . && (125)\end{aligned}$$

Now, we can choose n_0 large enough such that

$$\max_{\mathbf{s}\in\mathcal{S}^n(\Lambda)} \bar{\varepsilon}_L(\mathbf{s}) > \frac{1}{L+2} - \frac{L}{N(L+1)} \ . \qquad (126)$$

■

Lemma 6: Let $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$ and constraint Λ and let L be a positive integer. For any $\epsilon > 0$ there exists a $\nu(L, \mathcal{W}, \epsilon) > 0$ and n_0 such that for any (n, N, L) list code (ϕ, ψ) with $n \geq n_0$ and $N > L + 1$ whose codewords $\{\mathbf{x}(i) : i \in [N]\}$ satisfy

$$\lambda_L(T_{\mathbf{x}(i)}) < \Lambda - \epsilon \qquad \forall i \in [N] \ , \tag{127}$$

the error must satisfy

$$\max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \bar{\varepsilon}_L(\mathbf{s}) > \nu(L, \mathcal{W}, \epsilon) \ . \tag{128}$$

Proof: Fix $\epsilon > 0$. For each $P \in \mathcal{P}(\mathcal{X})$ from Lemma 4 we know there is a $\delta(P) > 0$ such that any joint distribution $\bar{P}$ with marginals within $\delta(P)$ of P can be approximated by a $\hat{P}$ with marginals equal to P such that $d_{\max}\left(\bar{P}, \hat{P}\right) < \epsilon$. Let

$$\mathcal{B}(P) = \left\{ P' \in \mathcal{P}(\mathcal{X}) : d_{\max}\left(P, P'\right) < \delta(P) \right\} \ . \tag{129}$$

Then $\{\mathcal{B}(P) : P \in \mathcal{P}(\mathcal{X})\}$ is an open cover of $\mathcal{P}(\mathcal{X})$. Since $\mathcal{P}(\mathcal{X})$ is compact there is a constant r and finite subcover $\{\mathcal{B}(P_j) : j \in [r]\}$. From this finite cover we can create a partition $\{A_j : j \in [r]\}$ of $\mathcal{P}$ such that $A_j \subseteq \mathcal{B}(P_j)$ for all j.

Now consider an (n, N, L) code whose codewords $\mathcal{C}$ satisfy (127). Let $F_j = \{i \in [N] : T_{\mathbf{x}(i)} \in A_j\}$. We can bound the error

$$\bar{\varepsilon}_L(\mathbf{s}) = \frac{1}{Nr} \sum_{j=1}^{r} \sum_{i \in F_j} \bar{\varepsilon}_L(i, \mathbf{s}) \geq \frac{|F_j|}{Nr} \left(\frac{1}{|F_j|} \sum_{i \in F_j} \bar{\varepsilon}_L(i, \mathbf{s}) \right) \ . \tag{130}$$

Since $\{F_j\}$ partition the codebook, for some j we have $|F_j| \geq N/r$. From Lemma 5 the jammer can force the error to be lower bounded by

$$\max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \bar{\varepsilon}_L(\mathbf{s}) \geq \frac{1}{r^2} \left(\frac{1}{L+1} - \frac{L}{N(L+1)} \right) \ . \tag{131}$$

Since the constant r is a function of ϵ, $\mathcal{W}$ and L, we are done. ■

Theorem 2 follows from the preceding Lemma.