Non-Asymptotic Covering Lemmas

Sergio Verdú
Princeton University, NJ 08544, USA

Abstract—In information theory, the packing and covering lemmas are conventionally used in conjunction with the typical sequence approach in order to prove the asymptotic achievability results for discrete memoryless systems. In contrast, the single-shot approach in information theory provides non-asymptotic achievability and converse results, which are useful to gauge the backoff from the asymptotic fundamental limits due to fixed blocklength, and which do not rely on discrete/memoryless assumptions. This paper reviews the non-asymptotic covering lemmas we have obtained recently and their application in single-user and multiuser information theory.

Index Terms—Shannon theory, achievability, finite blocklength regime, random coding, lossy compression, Wyner-Ziv compression, broadcast channels, data transmission with encoder side information, almost-lossless compression with a helper.

I. INTRODUCTION

Since “A Mathematical Theory of Communication” [1] the constructive side of coding theorems giving the fundamental limits of data transmission and/or compression in terms of the statistical laws of channels and sources have largely relied on two central notions:

- random coding
- typical sequences

In [1], Shannon used random coding to show that memoryless channel capacity is lower bounded by the maximal mutual information, while in [2] he used a random selection of the reproduction points in order to show that the rate-distortion function of a memoryless source is upper bounded by the minimal mutual information. Also conceived by Shannon in [1], the typical sequence approach relies on the law of large numbers to construct encoders/decoders that neglect the possibility that the source or channel may deviate appreciably from the average behavior predicted by their statistical laws. Finite-alphabet embodiments of the typical sequence achievability approach reign supreme in both single-user information theory (e.g. Wolfowitz [3], Csiszár and Körner [4], Cover and Thomas [5]) and multiuser information theory (e.g. Csiszár and Körner [4] and El Gamal and Kim [6]). In that context, key tools for the typical sequence approach are the so-called packing and covering lemmas. Packing lemmas ensure that if there are not too many randomly generated codewords, the probability that the channel output behaves as if it were generated by (or, equivalently, is jointly typical with) an “impostor” codeword is small. Covering lemmas ensure that if there are enough randomly generated codewords, independently from a source realization, at least one of them will behave as if it were jointly typical with the source. In those lemmas, “not to many” and “enough” are made precise by an exponential growth below/above the mutual information.

II. NOTATION AND AUXILIARY RESULTS

In contrast to the conventional packing and covering lemmas [6, Lemma 3.1 and 3.3] and [4, Lemma 10.1] as well as in contrast to the asymptotic fundamental limits, the non-asymptotic versions do not involve the mutual information but the distribution function of the information density.

Definition 1. The relative information of \( a \in A \) corresponding to probability measures \( P \) and \( Q \) on \( (A, F) \) is

\[
\iota_{P||Q}(a) = \log \frac{dP}{dQ}(a) \tag{1}
\]

We often abbreviate \( \iota_{P_X \| P_Y} = \iota_{X \| Y} \).

Definition 2. Given a distribution \( P_Y \) on \( B \) and a random transformation \( P_{Y|X} : A \rightarrow B \) the information density of \((x,y) \in A \times B \) is

\[
\iota_{X,Y}(x; y) = \iota_{P_{Y|X} = x, P_Y}(y) \tag{2}
\]

A simple property satisfied by the relative information is

Lemma 1 (\( \beta, P_X, P_Y, g \)).

\[
\beta \mathbb{E}[g(Y) \mathbb{1}_{\{\iota_{X,Y}(Y) \geq \log \beta\}}] \\
\leq \mathbb{E}[g(X) \mathbb{1}_{\{\iota_{X,Y}(X) \geq \log \beta\}}] \tag{3}
\]

The origin of the following result is uncertain, but its use in rate-distortion theory is well-known.

Lemma 2. (\( p, M, \alpha \)). For any \( 0 \leq p \leq 1, M \geq 1, \alpha \leq \frac{M}{p} \),

\[
\left(1 - \frac{p \alpha}{M}\right)^M < 1 - p + e^{-\alpha} \tag{4}
\]

Proof. Since \( \frac{p \alpha}{M} \leq 1 \), for any \( n > M \), the left side of (4) is bounded by

\[
\left(1 - \frac{p \alpha}{M}\right)^M < \left(1 - \frac{p \alpha}{n}\right)^n < e^{-p\alpha} \tag{5}
\]

Therefore, it is enough to show that for all \( 0 \leq p \leq 1, e^{-p\alpha} \leq 1 - p + e^{-\alpha} \) (6) which, in turn, is equivalent to \( f(p) \leq f(1) \) with \( f(p) = e^{-p\alpha} + p \). But since \( f \) is convex, and \( f(0) < f(1) \)

\[
f(p) \leq (1 - p)f(0) + pf(1) \leq f(1). \tag{7}
\]
III. BASIC NON-ASYMPTOTIC COVERING LEMMAS

Lemma 3 \((P_{\overline{Z}V}, M, \gamma)\). \[\text{[7]}\]

\[
\mathbb{P} \left[ \min_{m=1, \ldots, M} I_{Z; V}(Z; V_m) > \log M - \gamma \right] \\
\leq \mathbb{P} \left[ I_{Z; V}(Z; V) > \log M - \gamma \right] + e^{-\exp(\gamma)} \tag{8}
\]

with \(P_{V_1 \cdots V_m V Z} = P_V \times \cdots \times P_V \times P_{V Z}\).

Proof. Denote

\[S = \{(z, v) \in Z \times V, I_{Z; V}(z, v) \leq \log M - \gamma\}\]

and consider the following function of \(Z\):

\[
\mathbb{E} \left[ \prod_{m=1}^{M} 1\{(Z, V_m) \notin S\}|Z\right] \\
= \prod_{m=1}^{M} \mathbb{E} \left[ 1\{(Z, V_m) \notin S\}|Z\right] \\
= (1 - \mathbb{P}((Z, V_1) \in S|Z))^M \tag{9}
\]

where

\[\text{(10) \& (11) } \leftarrow (V_1, \ldots, V_m) \text{ are independent identically distributed (also conditioned on } Z)\];

\[\text{(12) } \leftarrow \text{Lemma 1 } \left(\frac{\exp(\gamma)}{M} P_{V Z} = z, P_V, 1\{(z, \cdot) \in S\}\right)\];

\[\text{(13) } \leftarrow \text{Lemma 2}\]

Averaging (11)-(13) with respect to \(P_Z\), we obtain

\[
\mathbb{P} \left[ \bigcap_{m=1}^{M} \left\{ (Z, V_m) \notin S \right\} \right] \leq \mathbb{P}((Z, V) \notin S) + e^{-\exp(\gamma)} \tag{14}
\]

which is the desired result.

The following generalization finds wide applicability in the proof of non-asymptotic achievable results in data transmission and lossy compression.

Lemma 4 \((P_{\overline{Z}V}, M, \gamma, F)\). \[\text{[7]}\]

\[
\text{For any event } F,
\mathbb{P} \left[ \bigcap_{m=1}^{M} \left\{ (Z, V_m) \notin F \right\} \right] \\
\leq \mathbb{P}((Z, V) \notin F) + e^{-\exp(\gamma)} + \mathbb{P} \left[ I_{Z; V}(Z; V) > \log M - \gamma \right] \tag{15}
\]

Proof. It is easy to check that the bound in (14) still holds if the event \(S\) is a subset of the right side of (9). Taking

\[S = F \cap \{(z, v) \in Z \times V, I_{Z; V}(z, v) \leq \log M - \gamma\}\] \tag{16}

we get

\[
\mathbb{P} \left[ \bigcap_{m=1}^{M} \left\{ (Z, V_m) \notin F \right\} \right] \\
\leq \mathbb{P} \left[ \bigcap_{m=1}^{M} \left\{ (Z, V_m) \notin S \right\} \right] \tag{17}
\]

\[
\leq \mathbb{P}((Z, V) \notin S) + e^{-\exp(\gamma)} \tag{18}
\]

where (18) is (14). Finally, (15) follows applying the union bound to the complement of (16).

Evidently, (15) is equivalent to

\[
\mathbb{P} \left[ \bigcup_{m=1}^{M} \{(Z, V_m) \in F\} \right] \\
\geq \mathbb{P} ((Z, V) \notin F) - \mathbb{P} \left[ I_{Z; V}(Z; V) > \log M - \gamma \right] - e^{-\exp(\gamma)} \tag{19}
\]

Suppose that \(\gamma\) is large enough that \(e^{-\exp(\gamma)}\) is negligible, and that \(M\) is large enough that \(\mathbb{P} \left[ I_{Z; V}(Z; V) > \log M - \gamma \right]\) is also negligible. Then (19) states that the probability that at least one of the \(M\) random variables \(V_m\) (although independent of \(Z\)) satisfies a certain property is essentially lower bounded by the corresponding probability if it were generated under \(P_{V|Z}\).

IV. MUTUAL COVERING LEMMA

The following lemma finds applications in multiuser information theory. Note that even though there are a total of \(M+N\) independent random variables, the bound behaves essentially as if in Lemma 4 there were \(M \leftarrow NM\) independent random variables.

Lemma 5. \((P_{Z V}, N, M, \gamma, F)\) \[\text{[8]}\]

\[
P_{Z_1 \cdots Z_N V_1 \cdots V_m Z V} = P_Z \times \cdots \times P_Z \times P_V \times \cdots \times P_V \times P_{Z V} \tag{20}
\]

Then, for any event \(\mathcal{F}\) and \(\gamma > 0\),

\[
\mathbb{P} \left[ \bigcap_{n=1}^{N} \bigcap_{m=1}^{M} \{(Z_n, V_m) \notin \mathcal{F}\} \right] \\
\leq \mathbb{P} ((Z, V) \notin \mathcal{F}) + \mathbb{P} \left[ I_{Z; V}(Z; V) > \log NM - 2\gamma \right] \\
+ \min \{ M, N \} - 1 \frac{1}{NM} (\exp(-\gamma) - \exp(-2\gamma)) + e^{-\exp(\gamma)} \tag{21}
\]

Proof. In view of the symmetry of the result, we may suppose, without loss of generality, that \(N \leq M\). We introduce the auxiliary random variables \((V_1, \ldots, V_N) \in \mathcal{V}^N\) such that for all \(k \in \{1, \ldots, N\}\), \((Z_k, V_k) \sim P_{Z V}\), are independent of all other random variables. Furthermore, we let \(W\) be equiprobable on \(\{1, \ldots, N\}\) and we abbreviate the random variable \(V = \bar{V}_W\), in other words,

\[
P_{\overline{V}_W} (v|v^N) = \frac{1}{N} \sum_{i=1}^{N} \delta_{v_i} \tag{22}
\]

Note that \(P_V = P_{\overline{V}}\) and

\[
\exp (I_{Z; \overline{V}}(z^N; v)) = \frac{1}{N} \sum_{i=1}^{N} \exp (I_{Z; V}(z_i; v)) \tag{23}
\]
Define the following functions

\[ \phi(z^N) = \mathbb{P} \left( \bigcup_{n=1}^{N} \{ (z_n, V_1) \in \mathcal{F} \} \right) \quad (24) \]
\[ \pi(z^N, v) = 1 \left\{ \bigcup_{n=1}^{N} \{ (z_n, v) \in \mathcal{F} \} \times 1 \{ i_{Z^N; V}(z^N; v) \leq \log M - \gamma \} \right\} \quad (25) \]

Note that, by change of measure,
\[ \mathbb{E}[\pi(Z^N, \bar{V})|Z^N = z^N] \leq M \exp(-\gamma) \mathbb{E}[\pi(z^N, V_1)] \]
\[ \leq M \exp(-\gamma) \phi(z^N) \quad (26) \]
\[ (27) \]

Invoking (23) we can bound for any \( \delta > 0 \),
\[ \mathbb{P} \left[ i_{Z^N; V}(Z^N; \bar{V}_1) > \log M - \gamma \right] \]
\[ = \mathbb{P} \left[ \frac{1}{N} \sum_{i=1}^{N} \exp(i_{Z^N; V}(Z_i; \bar{V}_1)) > M \exp(-\gamma) \right] \quad (28) \]
\[ \leq \mathbb{P} \left[ \exp(i_{Z^N; V}(Z_1; \bar{V}_1)) > NM \exp(-\gamma) - \delta \right] \]
\[ + \mathbb{P} \left[ \sum_{i=2}^{N} \exp(i_{Z^N; V}(Z_i; \bar{V}_1)) > \delta \right] \quad (29) \]
\[ \leq \mathbb{P} \left[ \exp(i_{Z^N; V}(Z_1; \bar{V}_1)) > NM \exp(-\gamma) - \delta \right] + \frac{N-1}{\delta} \quad (30) \]

where in (30) we have used Markov’s inequality and
\[ \mathbb{E}[\exp(i_{Z^N; V}(Z_i; \bar{V}_1))] = 1 \text{ if } i \neq 1. \]
Furthermore, we have
\[ 1 - \mathbb{E}[\pi(Z^N, \bar{V})] \]
\[ = \mathbb{P} \left[ \bigcap_{n=1}^{N} \{ (Z_n, \bar{V}) \notin \mathcal{F} \} \cup \{ i_{Z^N; \bar{V}}(Z^N; \bar{V}) > \log M - \gamma \} \right] \quad (31) \]
\[ = \frac{1}{N} \sum_{w=1}^{N} \mathbb{P} \left[ \bigcap_{n=1}^{N} \{ (Z_n, V_w) \notin \mathcal{F} \} \cup \{ i_{Z^N; \bar{V}}(Z^N; \bar{V}) > \log M - \gamma \} | W = w \right] \quad (32) \]
\[ = \mathbb{P} \left[ \bigcap_{n=1}^{N} \{ (Z_n, \bar{V}_1) \notin \mathcal{F} \} \cup \{ i_{Z^N; \bar{V}}(Z^N; \bar{V}_1) > \log M - \gamma \} \right] \quad (33) \]
\[ \leq \mathbb{P} \left[ (Z_1, \bar{V}_1) \notin \mathcal{F} \right] + \mathbb{P} \left[ i_{Z^N; \bar{V}}(Z^N; \bar{V}_1) > \log M - \gamma \right] \quad (34) \]

By conditioning on \( Z^N \) we can write the left side of (21) as
\[ \mathbb{P} \left[ \bigcap_{n=1}^{N} \bigcap_{m=1}^{M} \{ (Z_n, V_m) \notin \mathcal{F} \} \right] \]
\[ = \mathbb{E} \left[ (1 - \phi(Z^N))^M \right] \quad (35) \]
\[ = \mathbb{E} \left[ \left( 1 - \frac{M \exp(-\gamma) \phi(Z^N) \exp(\gamma)}{M} \right)^M \right] \quad (36) \]
\[ \leq \mathbb{E} \left[ \left( 1 - \frac{\mathbb{E}[\pi(Z^N, \bar{V})|Z^N] \exp(\gamma)}{M} \right)^M \right] \quad (37) \]
\[ \leq 1 - \mathbb{E} \left[ \pi(Z^N, \bar{V}) \right] + e^{-\exp(\gamma)} \quad (38) \]

where
\[ (37) \leftrightarrow (27); \quad (38) \leftrightarrow \text{Lemma 2 with} \]
\[ (p, M, \alpha) \leftrightarrow (\mathbb{E} \left[ \pi(Z^N, \bar{V})|Z^N \right], M, \exp(\gamma)) \quad (39) \]

The desired result now follows from (30), (34) and (38) choosing
\[ \delta = MN (\exp(-\gamma) - \exp(-2\gamma)). \quad (40) \]

A one-shot counterpart of Marton’s achievability result for the two-user broadcast channel with public message is obtained in [8].

REFERENCES