Conference Paper

Improving on the Cut-Set Bound via a Geometric Analysis of Typical Sets

Author(s):
Wu, Xiugang; Özgür, Ayfer

Publication Date:
2016

Permanent Link:
https://doi.org/10.3929/ethz-a-010645872

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Abstract—We consider the discrete memoryless symmetric primitive relay channel, where, a source $X$ wants to send information to a destination $Y$ with the help of a relay $Z$ and the relay can communicate to the destination via an error-free digital link of rate $R_0$, while $Y$ and $Z$ are conditionally independent and identically distributed given $X$. We develop two upper bounds on the capacity of this channel that are tighter than existing bounds, including the celebrated cut-set bound. Our approach significantly differs from the standard information-theoretic approach for proving upper bounds on the capacity of multi-user channels. We build on the blowing-up lemma to analyze the probabilistic geometric relations between the typical sets of the $n$-letter random variables associated with a reliable code for communicating over this channel. These relations translate to new entropy inequalities between the $n$-letter random variables involved.

I. INTRODUCTION

Characterizing the capacity of relay channels [1] has been a long-standing open problem in network information theory. The seminal work of Cover and El Gamal [2] has introduced two basic achievability schemes: Decode-and-Forward and Compress-and-Forward, and derived a general upper bound on the capacity, now known as the cut-set bound. Over the last decade, significant progress has been made on the achievability side; these schemes have been extended and unified to multi-relay networks [3]–[4] and many new relaying strategies have been discovered, such as Amplify-and-Forward, Compute-and-Forward, Noisy Network Coding etc. [5]–[7]. However, the progress on developing upper bounds that are tighter than the cut-set bound has been relatively limited. In particular, in most of the special cases where the capacity is known, the upper bound is given by the cut-set bound [2], [8]–[10].

In general, however, the cut-set bound is known to be not tight. Specifically, consider the primitive relay channel depicted in Fig. 1, where the source’s input $X$ is received by the relay $Z$ and the destination $Y$ through a channel $p(y,z|x)$, and the relay $Z$ can communicate to the destination $Y$ via an error-free digital link of rate $R_0$. When $Y$ and $Z$ are conditionally independent given $X$, and $Y$ is a stochastically degraded version of $Z$, Zhang [11] used the blowing up lemma to show that the inequality between the capacity and the cut-set bound is indeed strict in certain regimes of this channel.

This work was supported in part by the NSF CAREER award 1254786 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370.

However, Zhang’s result does not provide any information regarding the gap or suggest a way to compute it. For a special case of the primitive relay channel where the noise is modulo additive and $Z$ is a corrupted version of the noise for the $X$-$Y$ link, Aleksic, Razaghi and Yu characterize the capacity and show that it is strictly lower than the cut-set bound [12]. While this result provides an exact capacity characterization for a non-trivial special case, it builds strongly on the peculiarity of the channel model and in this respect its scope is more limited than Zhang’s result.

More recently, a new upper bound demonstrating an explicit gap to the cut-set bound was developed by Xue [13] for general primitive relay channels. Xue’s bound relates the gap of the cut-set bound to the reliability function of the $X$-$Y$ link. Unlike Zhang’s result, Xue’s bound can be numerically computed. While it is strictly tighter than the cut-set bound in certain regimes of the primitive relay channel, with an explicit computable gap, it can also be looser than the cut-set bound.

In [14], we presented two new upper bounds on the capacity of the primitive relay channel. The first of these bounds can be regarded as a direct improvement of Xue’s bound. It is indeed strictly tighter than both Xue’s bound and the cut-set bound and like Xue’s bound involves the reliability function of the $X$-$Y$ link. Our second bound was based on a new set of arguments and is structurally different than the first one. It can be significantly tighter than our first bound as demonstrated in [14] for the binary symmetric relay channel, however it can also be looser than it for some other channel models.

The current paper is a continuation of our work in [14]. We present a new bound which is strictly tighter than our first bound in [14] and is also structurally different from it. In particular, it does not involve the reliability function of the $X$-$Y$ link but is structurally closer to our second bound in [14]. The more important contribution of this paper is to distill a new proof technique which significantly differs from existing
converse approaches in the literature and can be potentially
useful for other multi-user problems. In general, proving an
upper bound on the capacity of a multi-user channel involves
dealing with entropy relations between the various \( n \)-letter
random variables induced by the reliable code and the channel
structure (together with using Fano’s inequality). In order to
prove the desired relations between the entropies of the \( n \)-letter
random variables involved, in this paper we consider their
\( B \)-letter i.i.d. extensions (leading to length \( B \) i.i.d. sequences of
\( n \)-letter random variables). We then use the blowing up lemma
to analyze the geometry of the typical sets associated with
these \( B \)-letter sequences. We present two different ways to
translate the (probabilistic) geometric relations between these
typical sets into new entropy relations between the random
variables involved. This leads to two different bounds on the
capacity of the primitive relay channel which do not include
each other in general. As pointed out before, the first of these
bounds is new to this paper, the second one recovers the second
bound we presented in [14].

II. PRELIMINARIES

Consider a primitive relay channel as depicted in Fig. 1. The
source’s input \( X \) is received by the relay \( Z \) and the destination
\( Y \) through a channel

\[
(\Omega_X, p(y,z|x), \Omega_Y \times \Omega_Z)
\]

where \( \Omega_X, \Omega_Y \) and \( \Omega_Z \) are finite sets denoting the alphabets
of the source, the destination and the relay, respectively, and
\( p(y,z|x) \) is the channel transition probability; the relay \( Z \) can
communicate to the destination \( Y \) via an error-free digital link
of rate \( R_0 \).

For this channel, a code of rate \( R \) for \( n \) channel uses, denoted by

\[
(C_{n,R}, f_{n}(\omega^n), g_{n}(y^n, f_{n}(\omega^n))), \text{ or simply, } (C_{n,R}, f_{n}, g_{n}),
\]

consists of the following:

1) A codebook at the source \( X \),

\[
C_{n,R} = \{ x^n(m) \in \Omega_X^n, m \in \{ 1, 2, \ldots, 2^{nR} \};
\]

2) An encoding function at the relay \( Z \),

\[
f_n : \Omega_X^n \rightarrow \{ 1, 2, \ldots, 2^{nR} \};
\]

3) A decoding function at the destination \( Y \),

\[
g_n : \Omega_Y^n \times \{ 1, 2, \ldots, 2^{nR} \} \rightarrow \{ 1, 2, \ldots, 2^{nR} \}.
\]

The average probability of error of the code is defined as

\[
P_{e}^{(n)} = \text{Pr}(g_n(Y^n, f_n(Z^n)) \neq M),
\]

where the message \( M \) is assumed to be uniformly drawn from
the message set \( \{ 1, 2, \ldots, 2^{nR} \} \). A rate \( R \) is said to be
achievable if there exists a sequence of codes

\[
\{(C_{n,R}, f_{n}, g_{n})\}_{n=1}^{\infty}
\]

such that the average probability of error \( P_{e}^{(n)} \rightarrow 0 \) as \( n \rightarrow \infty \).

The capacity of the primitive relay channel is the supremum
of all achievable rates. The well-known cut-set bound on
the capacity of the primitive relay channel is stated as the
following.

\[ R \leq I(X;Y,Z) \]

(1)

\[ R \leq I(X;Y) + R_0. \]

(2)

Inequalities (1) and (2) are generally known as the broadcast
bound and multiple-access bound, since they correspond to the
broadcast channel \( X-Y'Z \) and multiple-access channel \( XZ-Y' \),
respectively.

A. Symmetric Primitive Relay Channel

To simplify the exposition, in this paper, we only concentrate
on the symmetric case of the primitive relay channel, that is, when \( Y \) and \( Z \)
are conditionally independent and identically distributed given \( X \), however our results can be
extended to the asymmetric case by using channel simulation
arguments. Formally, a primitive relay channel is said to be
symmetric if

1) \( p(y,z|x) = p(g(x)p(z|x), \) \)

2) \( \Omega_Y = \Omega_Z := \Omega, \text{ and } \text{Pr}(Y = \omega|X = x) = \text{Pr}(Z = \omega|X = x) \text{ for any } \omega \in \Omega \text{ and } x \in \Omega_X. \)

In this case, we also use \( p(\omega|x) \) to denote the transition probability
of both the \( X-Y \) and \( X-Z \) channels.

III. MAIN RESULTS

This section presents two new upper bounds on the capacity
of symmetric primitive relay channels that are generally tighter
than the cut-set bound. Before stating our main theorems, in
the following section we first explain the relation of our new
bounds to the cut-set bound.

A. Improving on the Cut-Set Bound

Let the relay’s transmission be denoted by \( I_n = f_n(Z^n). \)
Let us recall the derivation of the cut-set bound. The first step in
deriving (1)–(2) is to use Fano’s inequality to conclude that

\[ nR \leq I(X^n;Y^n, I_n) + n\epsilon. \]

We can then either proceed as

\[ nR \leq I(X^n;Y^n, I_n) + n\epsilon \leq I(X^n;Y^n, Z^n) + n\epsilon \leq nI(X;Y, Z) + n\epsilon \]

to obtain the broadcast bound (1), where the second inequality
follows from the data processing inequality and the single
letterization in the third line can be either done with a time-
sharing or fixed composition code argument\(^1\); or we can

\(^1\)Note that the time-sharing or the fixed composition code argument
for single letterization is needed to preserve the coupling to the second inequality
in (4) via \( X \).
Proceed as
\[ nR \leq I(X^n; Y^n, I_n) + n\epsilon \]
\[ \leq I(X^n; Y^n) + H(I_n|Y^n) - H(I_n|X^n) + n\epsilon \quad (3) \]
\[ \leq nI(X; Y) + nR_0 + n\epsilon \quad (4) \]
to obtain the multiple-access bound (2), where to obtain the last inequality we upper bound \( H(I_n|Y^n) \) by \( nR_0 \) an the use the fact that \( H(I_n|X^n) \) is non-negative.

Instead of simply lower bounding \( H(I_n|X^n) \) by 0 in the last step, our bounds presented in the next two subsections are based on letting \( \mathcal{H}(I_n|X^n) = n\alpha_n \), and prove a third inequality that forces \( \alpha_n \) to be strictly positive. Intuitively, it is easy to see that \( \alpha_n \) cannot be arbitrarily small. Specifically, suppose \( \alpha_n \approx 0 \), then roughly speaking, this implies that given the transmitted codeword \( X^n \), there is no ambiguity about \( I^n \), or equivalently, all the \( Z^n \) sequences jointly typical with \( X^n \) are mapped to the same \( I_n \). Since \( Y^n \) and \( Z^n \) are statistically equivalent given \( X^n \) (they share the same typical set given \( X^n \)) this would further imply that \( I_n \) can be determined based on \( Y^n \) and therefore the transmitted codeword can be decoded based solely on \( Y^n \), which forces the rate to be even smaller than \( I(X; Y) \). In general, there is a trade-off between how close the rate can get to the multiple-access bound \( I(X; Y) + R_0 \) and how much it can exceed the point-to-point capacity \( I(X; Y) \) of the \( X-Y \) link.

In our new bounds, we capture this trade-off by leaving \( \alpha_n \) as it is in (3), yielding
\[ R \leq I(X; Y) + R_0 - \alpha_n + \epsilon \]
and proving a new constraint on the rate involving \( \alpha_n \). This new constraint is obtained by writing
\[ nR \leq I(X^n; Y^n, I_n) + n\epsilon \]
\[ = H(Y^n, I_n) - H(Y^n|X^n) - H(I_n|X^n) + n\epsilon \quad (5) \]
and upper bounding \( H(Y^n, I_n) \) in terms of \( \alpha_n \). We do this in two different ways corresponding to the two different ways of expanding \( H(Y^n, I_n) \), i.e.
\[ H(Y^n, I_n) = H(Y^n) + H(I_n|Y^n) \]
\[ = H(I^n) + H(Y_n|I^n). \]

Our first bound attacks the first conditional entropy term and is based on proving that
\[ H(I_n|Y^n) \leq H\left( \sqrt{\frac{\alpha_n \ln 2}{2}} + \sqrt{\frac{\alpha_n \ln 2}{2}} \log(|\Omega| - 1), \right. \] (6)
while our second bound attacks the second conditional entropy term and is based on proving that
\[ H(Y_n|I^n) \]
\[ \leq H(X^n|I^n) \]
\[ = H(X^n) - H(X^n|Z^n) + n(H(Z|X) + \Delta(p(x), \alpha_n)), \]
where \( \Delta(p(x), \alpha_n) \) is a quantity that depends on the input distribution \( p(x) \) and \( \alpha_n \), which we formally define in Section III-C. Once these entropy relations are proved, it is not difficult to plug them in (5) and see how they lead to the theorems stated in the next two sections. The heart of our argument is therefore to prove these two entropy inequalities.

To accomplish this, we suggest a new set of proof techniques. In particular, we look at the \( B \)-letter i.i.d. extensions of the random variables \( X^n, Y^n \) and \( I_n \) and study the geometric relations between their typical sets by using the generalized blowing-up lemma. While we use this same general approach for bounding the two entropy terms, we build on different arguments in each case, which eventually leads to two different bounds on the capacity of the relay channel that do not include each other in general.

B. Bounding \( H(I_n|Y^n) \)

Our first bound builds on bounding \( H(I_n|Y^n) \) and it is given by the following theorem that will be proved in Section IV. This bound is new and in particular strictly tighter than our first bound in [14].

**Theorem 3.1:** For the symmetric primitive relay channel, if a rate \( R \) is achievable, then there exists some \( p(x) \) and
\[ a \in \left[ 0, \min\left\{ R_0, H(Z|X), \ln \frac{2}{|\Omega|} \left( \frac{1}{\Omega} - 1 \right)^2 \right\} \right] \] (7)
such that
\[ R \leq I(X; Y, Z) \]
\[ R \leq I(X; Y) + R_0 - a \]
\[ R \leq I(X; Y) + H\left( \sqrt{\frac{a \ln 2}{2}} \right) \]
\[ + \sqrt{\frac{a \ln 2}{2}} \log(|\Omega| - 1) - a. \] (10)

Clearly our bound in Theorem 3.1 implies the cut-set bound in Proposition 2.1. In fact, it can be checked that our bound is **strictly** tighter than the cut-set bound for any \( R_0 > 0 \). For this, note that (9) will reduce to (2) only if \( a = 0 \); however, if \( a = 0 \) then (10) will constrain \( R \) by the rate \( I(X; Y) \) which is lower than the cut-set bound.

C. Bounding \( H(Y^n|I_n) \)

Before presenting our second upper bound, we first define a parameter \( \Delta(p(x), a) \) that will be used in stating the theorem. This bound is equivalent our second bound in [14], however we provide an alternative definition for \( \Delta(p(x), a) \) in terms of information-theoretic quantities.

**Definition 3.1:** Given a fixed channel transition probability \( p(\omega|x) \), for any \( p(x) \) and \( a \geq 0 \), \( \Delta(p(x), a) \) is defined as
\[ \Delta(p(x), a) := \max_{\tilde{p}(\omega|x)} H(\tilde{p}(\omega|x)|p(x)) + D(\tilde{p}(\omega|x)||p(\omega|x)|p(x)) \]
\[ - H(p(\omega|x)|p(x)) \]
\[ - H(\tilde{p}(\omega|x)|p(x)) \]
\[ s.t. \sum_{(x, \omega)} |p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)| \leq 2\sqrt{a \ln 2}. \]

In the above, we adopt the notation in [16]. Specifically, \( D(\tilde{p}(\omega|x)||p(\omega|x)|p(x)) \) is the conditional relative entropy.
defined as
\[
D(\tilde{p}(\omega|x)||p(\omega|x)) := \sum_{(x,\omega)} p(x)\tilde{p}(\omega|x) \log \frac{\tilde{p}(\omega|x)}{p(\omega|x)},
\]
and \(H(\tilde{p}(\omega|x)||p(x))\) is the conditional entropy defined with respect to the joint distribution \(p(x)p(\omega|x)\), i.e.,
\[
H(\tilde{p}(\omega|x)||p(x)) := -\sum_{(x,\omega)} p(x)\tilde{p}(\omega|x) \log \tilde{p}(\omega|x),
\]
and \(H(p(\omega|x)||p(x))\) is the conditional entropy similarly defined with respect to \(p(x)p(\omega|x)\).

It can be easily seen that \(\Delta(p(x), a) \geq 0\) for all \(p(x)\) and \(a \geq 0\), and \(\Delta(p(x), a) = 0\) when \(a = 0\). Moreover, for any fixed \(p(x)\) and \(a > 0\), \(\Delta(p(x), a) = \infty\) if and only if there exists some \(x\) with \(p(x) > 0\), and some \(\omega\) and \(\tilde{\omega}\) such that \(p(\omega|x) = 0\) and \(\tilde{p}(\omega|x) > 0\). Thus, a sufficient condition for \(\Delta(p(x), a) < \infty\) for all \(p(x)\) and \(a > 0\) is that the channel transition matrix is fully connected, i.e., \(p(\omega|x) > 0, \forall x, \omega \in \Omega \times \Omega\). In this case, \(\Delta(p(x), a) \rightarrow 0\) as \(a \rightarrow 0\) for any \(p(x)\).

We are now ready to state our second new upper bound, which is proved by bounding \(H(Y^n|I_a)\).

**Theorem 3.2:** For the symmetric primitive relay channel, if a rate \(R\) is achievable, then there exists some \(p(x)\) and \(a \in [0, \min\{R_0, H(Z|X)\}]\) such that
\[
\begin{align*}
R &\leq I(X;Y,Z) \\
R &\leq I(X;Y) + R_0 - a \\
R &\leq I(X;Y) + \Delta(p(x), a)
\end{align*}
\]
\[(13)-(15)\]

Theorem 3.2 also implies the cut-set bound in Proposition 2.1. In particular, when the channels \(X-Y\) and \(X-Z\) have a fully connected transition matrix, our new bound is strictly tighter than the cut-set bound since \(\Delta(p(x), a) \rightarrow 0\) as \(a \rightarrow 0\) for any \(p(x)\) in this case.

In the remaining space we provide the proof of Theorem 3.1.

**IV. PROOF OF THEOREM 3.1**

Based on the discussion in Section III-A, to show Theorem 3.1, it suffices to prove the entropy inequality (6) between various \(n\)-letter random variables. For this, we go to the higher dimensional, say \(nB\) dimensional space, to invoke the concepts of typical sets, and resort to a result on measure concentration, namely, the generalized blowing-up lemma.

Specifically, consider the \(B\)-length i.i.d. extensions of the random variables \(X^n, Y^n, Z^n\) and \(I_a^n\), i.e.,
\[
\{(X^n(b), Y^n(b), Z^n(b), I_a^n(b))\}_{b=1}^B,
\]
\[(16)\]
where for any \(b \in [1 : B]\), \((X^n(b), Y^n(b), Z^n(b), I_a^n(b))\) has the same distribution as \((X^n, Y^n, Z^n, I_a^n)\). For notational convenience, in the sequel we write the \(B\)-length vector \([X^n(1), X^n(2), \ldots, X^n(B)]\) as \(X^n\) and similarly \(Y^n, Z^n\) and \(I^n\); note here we have \(I^n = [f_\omega(Z^n(1)), f_\omega(Z^n(2)), \ldots, f_\omega(Z^n(B))] = f(Z^n)\).

The following lemma is critical for establishing inequality (6). Its own proof is given at the end of the paper.

**Lemma 4.1:** Let \(f^{-1}(i) := \{\omega \in \Omega^n : f(\omega) = i\}\) and \(\Gamma_n(\sqrt{\frac{\Delta_n}{2}}+\delta)\) be its blown-up set defined as
\[
\Gamma_n(\sqrt{\frac{\Delta_n}{2}}+\delta) = \left\{\omega \in \Omega^n : \exists \omega' \in f^{-1}(i) \text{ s.t. } d(\omega, \omega') \leq nB\left(\sqrt{\frac{\Delta_n}{2}} + \delta\right)\right\}
\]
where \(d(\omega, \omega')\) denotes the Hamming distance between \(\omega\) and \(\omega'\). Then for any \(\delta > 0\) and \(B\) sufficiently large,
\[
\Pr(Y \in \Gamma_n(\sqrt{\frac{\Delta_n}{2}}+\delta)) \geq 1 - \delta.
\]

With the above lemma, we now upper bound \(H(I^n|Y^n)\). Let
\[
E = \mathbb{1}(Y \in \Gamma_n(\sqrt{\frac{\Delta_n}{2}}+\delta))
\]
where \(\mathbb{1}(\cdot)\) is the indicator function defined as
\[
\mathbb{1}(A) = \begin{cases} 1 & \text{if } A \text{ holds} \\ 0 & \text{otherwise} \end{cases}
\]
We have
\[
H(I^n|Y^n) \leq H(I^n, E|Y^n) = H(E|Y^n) + H(I^n|Y^n, E) \leq H(I^n|Y^n, E = 1)
\]
\[
= \Pr(E = 1)H(I^n|Y^n, E = 1) + \Pr(E = 0)H(I^n|Y^n, E = 0) + 1 \leq H(I^n|Y^n, E = 1) + \delta nBR_0 + 1.
\]

To bound \(H(I^n|Y^n, E = 1)\), consider a Hamming ball\(^2\) centered at \(Y^n\) of radius \(nB\left(\sqrt{\frac{\Delta_n}{2}} + \delta\right)\). The condition \(E = 1\), i.e., \(Y^n \in \Gamma_n(\sqrt{\frac{\Delta_n}{2}}+\delta)\), ensures that there is at least one point \(\omega \in f^{-1}(I)\) belonging to this ball, and therefore, given \(E = 1\) and \(Y^n\) there are at most \(\Gamma_n(\sqrt{\frac{\Delta_n}{2}}+\delta)\) possibilities of \(I^n\), leading to the following upper bound on \(H(I^n|Y^n, E = 1)\),
\[
H(I^n|Y^n, E = 1) \leq \log \left|\Gamma_n(\sqrt{\frac{\Delta_n}{2}}+\delta)\right| = \log \left[\Gamma_n(\sqrt{\frac{\Delta_n}{2}}+\delta)\right] \leq nB \left[H(\sqrt{\frac{\Delta_n}{2}} + \delta) + \sqrt{nB \log(\Omega^n - 1) + \delta_1}\right]
\]
\[(18)\]
for some \(\delta_1 \rightarrow 0\) as \(\delta \rightarrow 0\), where (18) follows from the characterization of the volume of a Hamming ball. Plugging (18) into (17), we have
\[
H(I^n|Y^n) \leq nB \left[H(\sqrt{\frac{\Delta_n}{2}} + \delta) + \sqrt{nB \log(\Omega^n - 1) + \delta_1}\right] + \delta nBR_0 + 1.
\]

\(^2\)A Hamming ball centered at \(c\) of radius \(r\), denoted by \(\text{Ball}(c, r)\), is defined as the set of points that are within Hamming distance \(r\) of \(c\). The center \(c\) can be omitted in the notation when it becomes irrelevant.
Dividing $B$ at both sides of the above inequality and noting that
\[ H(I|Y) = \sum_{b=1}^{B} H(I_n(b)|Y_n(b)) = BH(I_n|Y_n), \]
we have
\[ H(I_n|Y_n) \leq n \left[ H \left( \sqrt{\frac{a_n \ln 2}{2}} + \sqrt{\frac{a_n \ln 2}{2} \log(|\Omega| - 1)} \right) + \delta_1 + \delta R_0 + \frac{1}{nB} \right]. \tag{19} \]

Since $\delta, \delta_1$ and $\frac{1}{nB}$ in (19) can all be made arbitrarily small by choosing $B$ sufficiently large, we obtain (6).

We next prove Lemma 4.1.

**Proof of Lemma 4.1:** Consider any $(x, i) \in T_e^B(X^n, I_n)$, where $T_e^B(X^n, I_n)$ denotes the $\epsilon$-jointly typical sets with respect to $(X^n, I_n)$. From [17, Sec. 2.5], we have for some $\epsilon_1 \to 0$ as $\epsilon \to 0$,
\[ p(i|x) \geq 2^{-BH(I_n(X^n)+\epsilon_1)} \geq 2^{-nB(a_n+\epsilon_1)}, \]
\[ \text{i.e.,} \]
\[ \Pr(Z \in f^{-1}(i)|x) \geq 2^{-nB(a_n+\epsilon_1)}. \]

We now apply the generalized blowing-up lemma as stated in the following.

**Lemma 4.2 (Generalized Blowing-Up Lemma):** Let $U_1, U_2, \ldots, U_n$ be $n$ independent random variables taking values in a finite set $\mathcal{U}$. Then, for any $A \subseteq \mathcal{U}^n$ with $\Pr(U^n \in A) \geq 2^{-n U_n}$,
\[ \Pr(U^n \in \Gamma_{nB(\sqrt{\frac{\alpha}{nB}} + 2\sqrt{\epsilon_1})(f^{-1}(i))}) \geq 1 - e^{-2n^2}, \forall \epsilon_1 > 0. \]

With Lemma 4.2, we have
\[ \Pr(Z \in \Gamma_{nB(\sqrt{\frac{\alpha}{nB}} + 2\sqrt{\epsilon_1})(f^{-1}(i))}|x) \]
\[ = \Pr(Z \in \Gamma_{nB(\sqrt{\frac{X_n+1}{2nB}} + 2\sqrt{\epsilon_1})(f^{-1}(i))}|x) \]
\[ \geq \Pr(Z \in \Gamma_{nB(\sqrt{\frac{X_n+1}{2nB}} + 2\sqrt{\epsilon_1})(f^{-1}(i))}|x) \]
\[ \geq \Pr(Z \in \Gamma_{nB(\sqrt{\frac{X_n+1}{2nB}} + \sqrt{\epsilon_1})(f^{-1}(i))}|x) \]
\[ \geq 1 - e^{-2nB_1} \]
\[ \geq 1 - \sqrt{\epsilon_1} \]
for sufficiently large $B$. Noting that $Y$ and $Z$ are identically distributed given $X$, we obtain
\[ \Pr(Y \in \Gamma_{nB(\sqrt{\frac{\alpha}{nB}} + 2\sqrt{\epsilon_1})(f^{-1}(i))}|x) \geq 1 - \sqrt{\epsilon_1}, \]
\[ \text{and thus,} \]
\[ \Pr(Y \in \Gamma_{nB(\sqrt{\frac{\alpha}{nB}} + 2\sqrt{\epsilon_1})(f^{-1}(i))}) \]
\[ = \sum_{(x,i)} \Pr(Y \in \Gamma_{nB(\sqrt{\frac{\alpha}{nB}} + 2\sqrt{\epsilon_1})(f^{-1}(i))}|x,i)p(x, i) \]
\[ = \sum_{(x,i)} \Pr(Y \in \Gamma_{nB(\sqrt{\frac{\alpha}{nB}} + 2\sqrt{\epsilon_1})(f^{-1}(i))}|x)p(x, i) \]
\[ \geq (1 - \sqrt{\epsilon_1}) \sum_{(x,i)} p(x, i) \]
\[ \geq (1 - \sqrt{\epsilon_1})^2 \]
\[ \geq 1 - 2\sqrt{\epsilon_1} \]
for sufficiently large $B$, where (20) follows due to the Markov chain: $Y \leftrightarrow X \leftrightarrow Z \leftrightarrow I$. Finally, choosing $\delta$ to be $2\sqrt{\epsilon_1}$ concludes the proof of Lemma 4.1. \[ \blacksquare \]

**References**


