Communications in Information and Systems
Volume 16, Number 2, 83–109, 2016

A randomized covering-packing duality between source and channel coding

Mukul Agarwal and Sanjoy Mitter

Let b be a *general channel*, defined as a sequence of transition probabilities in the sense of Verdu and Han [9], over which the uniform X source, denoted by U, is *directly* communicated to within distortion level D. The source U puts uniform distribution on all sequences with type *precisely* p_X as compared with the i.i.d. X source which puts 'most of' its mass on sequences with type 'close to' p_X. A *randomized* covering-packing duality is established between source-coding and channel-coding by considering the source-coding problem (covering problem) of coding the source U to within distortion level D and the channel coding problem (packing problem) of reliable communication over b, thus leading to a proof of $C \geq R_U(D)$ where C is the capacity of b and $R_U(D)$ is the rate-distortion function of U. This also leads to an operational view of source-channel separation for communication with a fidelity criterion.

1. Introduction

Let b be a general channel over which the uniform X source is *directly* communicated to within distortion level D.

This means the following:

Let the source input space be $\mathcal{X}$ and the source reproduction space be $\mathcal{Y}$. $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. Intuitively, a uniform X source, U, puts a uniform distribution on all sequences with a type p_X. This is as opposed to the i.i.d. X source which puts "most of" its mass on sequences with type "close to" p_X. See Section 3 for a precise definition. A general channel is a sequence $< b^n >_1^\infty$ where b^n is a transition probability from $\mathcal{X}^n$ to $\mathcal{Y}^n$; a precise definition of a general channel can be found in Section 3. When the block-length is n, the uniform X source is denoted by U^n. If, with input U^n

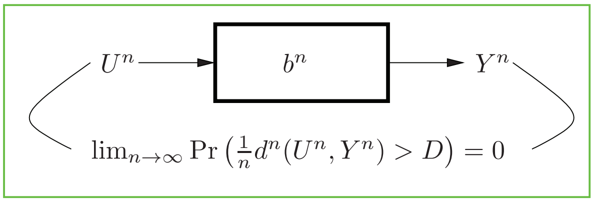

Figure 1: A channel which communicates the uniform X source to within distortion level D.

into the channel, the output is Y^n,

(1) $$\lim_{n\to\infty} \Pr\left(\frac{1}{n}d^n(U^n, Y^n) > D\right) = 0,$$

it is said that the channel $< b^n >_1^\infty$ directly communicates the uniform X source to within a distortion level D. In the above, $< d^n >_1^\infty$, $d^n : \mathcal{X}^n \times \mathcal{Y}^n \to [0, \infty)$, is a permutation-invariant (a special case is additive) distortion measure. The generality of the channel is in the sense of [9]. See Section 3 for precise definitions. See Figure 1.

Such a general channel intuitively functions as follows: when the block-length is n, with high probability, a sequence in $u^n \in \mathcal{U}^n$ is distorted within a ball of radius nD and this probability $\to 0$ as $n \to \infty$. Note that $u^n \in \mathcal{U}^n$ but the ball of radius nD exists in the output space $\mathcal{Y}^n$. Also see Figures 2 and 3. This definition of *direct* communication is reminiscent of [4].

Note that the uniform X source is not defined for all block-lengths; this point will be clarified in Section 3.

Consider the two problems:

- Covering problem: the rate-distortion source-coding problem of compressing the source U to within distortion level D, that is, computing the minimum rate needed to compress the source U to within a distortion level D. Denote the rate-distortion function by $R_U(D)$. Intuitively, the question is to find the minimum number of $y^n \in \mathcal{Y}^n$ such that balls of radius nD centered around y^n *cover* the space $\mathcal{U}^n$. By ball of radius nD centered around $y^n \in \mathcal{Y}^n$, is meant the set $\{u^n \in \mathcal{U}^n \mid \frac{1}{n}d^n(u^n, y^n) \le D\}$. Note that balls are centered around $y^n \in \mathcal{Y}^n$ but balls of radius nD exist in $\mathcal{U}^n$. Since the setting is information-theoretic, the balls should 'almost' cover the whole space.

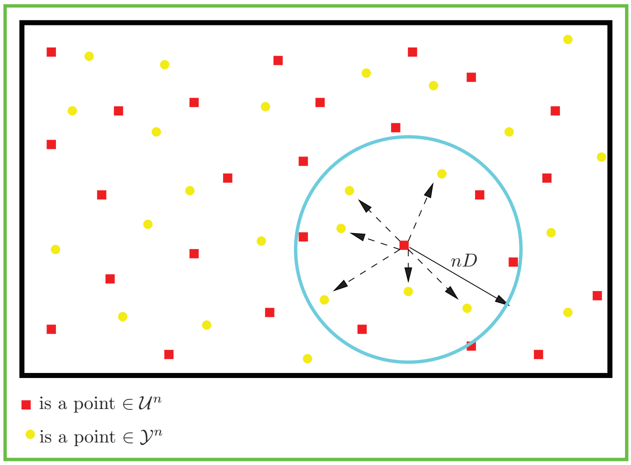

Figure 2: Intuitive action of a channel which directly communicates the uniform X source to within a distortion level D.

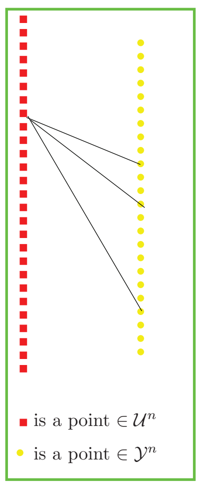

Figure 3: Intuitive action of a channel which directly communicates the uniform X source to within a distortion D, viewed as a bipartite graph. A point $u^n \in \mathcal{U}^n$ (shown in red) is mapped to a point $y^n \in \mathcal{Y}^n$ (shown in yellow) if $\frac{1}{n} d^n(u^n, y^n) \leq D$. Note that this is just an intuitive picture; the rigorous mathematics is given by (1). The mapping of only one point $\in \mathcal{U}^n$ is shown; the same happens for other points in $\mathcal{U}^n$.

- Packing problem: the channel-coding problem of communicating reliably over a general channel b which is known to directly communicate the source U to within distortion level D (packing problem). Denote the channel capacity by C. Intuitlvely, the question is to find the maximum number of $u^n \in \mathcal{U}^n$ such that balls of radius nD centered around these u^n *pack* the $\mathcal{Y}^n$ space. By ball of radius nD centered around u^n, is meant the set $\{y^n \in \mathcal{Y}^n \mid \frac{1}{n}d^n(u^n, y^n) \leq D\}$. Note that $u^n \in \mathcal{U}^n$ but balls of radius nD centered around these codewords exist in the $\mathcal{Y}^n$ space. Since the setting is information theoretic, the balls which pack the space can overlap 'a bit'.

Clearly, there is a notion of duality in these problem statements. It is unclear how to make this duality precise for these deterministic problems. However, a randomized covering-packing duality can be established between the above two problems, thus also proving that the answer to the first problem is less than or equal to the answer to the second problem, in the following way:

The codebook construction and error analysis for the source-coding problem are roughly the following: Let the block-length be n. Generate 2^{nR} codewords $\in \mathcal{Y}^n$ independently and uniformly from the set of all sequences with type q where q is an achievable type on the output space. *Roughly*, a $u^n \in \mathcal{U}^n$ is encoded via minimum distance encoding. The main error analysis which needs to be carried out is computing the probability that a certain codebook sequence does not encode a particular u^n, that is, computing

$$\Pr\left(\frac{1}{n}d^n(u^n, Y^n) > D\right) \tag{2}$$

where Y^n is a uniform random variable on sequences of type q. A best possible q is chosen in order to get an upper bound on the rate-distortion function.

The codebook construction and error analysis for the channel-coding problem are roughly the following: Let the block length be n. Generate 2^{nR} codewords $\in \mathcal{U}^n$ independently and uniformly. Let y^n be received. The determination of which codeword is transmitted is *roughly* via minimum distortion decoding. As it will become clear later, the main error calculation in the channel-coding problem is computing the probability of correct decoding, namely:

$$\Pr\left(\frac{1}{n}d^n(U^n, y^n) > D\right) \tag{3}$$

where y^n has type q. Finally, a worst case error analysis is done by taking the worst possible q.

By symmetry, (2) and (3) are equal assuming the distortion measure is additive (more generally, permutation invariant) and this leads to a proof that $C \geq R_U(D)$. *This equality of (2) and (3) is a randomized covering-packing duality between source-coding and channel-coding.* Further, this represents an operational view.

As regards this paper, an operational view, refers to a view which uses only the operational meanings of information-theoretic concepts: for example, channel capacity is defined as the maximum rate of reliable communication or the rate-distortion function as the minimum rate needed to code a source to within a certain distortion level. No functional simplifications beyond the equality of (2) and (3) are needed.

This proof is discussed precisely in Section 4.

A source-channel separation theorem for communication with a fidelity criterion follows from the above: a channel $k =< k^n >_1^\infty$ with input space $\mathcal{I}$ and output space $\mathcal{O}$ is said to *be capable of* communicating the uniform X source to within distortion level D if there exists an encoder-decoder pair $< e^n, f^n >_1^\infty$ such that $< e^n \circ k^n \circ f^n >_1^\infty$ *directly* communicates the uniform X source to within distortion level D. It follows that the capacity of the channel $b =< e^n \circ k^n \circ f^n >_1^\infty$ is $\geq R_U(D)$. From this it follows that the capacity of the channel k is $\geq R_U(D)$. See Section 5 for details. By use of the usual argument of source-coding followed by channel-coding, proof of the optimality of source-channel separation for communication of the uniform X source over the channel k to within distortion level D follows. *This is an operational view for the same reason as above.* Both the channel capacity problem and the rate-distortion problem are infinite dimensional optimization problems. By use of this methodology, the optimality of source-channel separation is proved without reducing the problems to finite dimensional problems. This is as opposed to the proof of separation: for example, in [8] which crucially relies on the the single-letter maximum mutual information expression for channel capacity and the single-letter minimum mutual information expression for the rate-distortion function.

In Shannon's model of communication, the objective is to reliably communicate, or to communicate within a certain distortion, a message over a noisy channel. The definition of reliable communication is that the probability of error should tend to zero with increasing block lengths of the code. The fundamental result is the noisy channel coding theorem which states that provided communication takes place at a rate below the capacity of the channel, the objective, namely, communication of the message with probability of error tending to zero, can be attained. The converse theorem states that the objective cannot be attained if communication takes place at a rate above capacity. The definition of capacity is in terms of maximizing the mutual information between the input and output of the channel where maximization is performed over the space of input probability distributions. It should be emphasized that this model of communication is that of reliable communication in the sense of Shannon. There are many other problems where the channel is used for other objectives and where the characterization of the channel in terms of capacity defined in terms of the mutual information between the input and output is not appropriate. One example is where the source message is an unstable source where the appropriate characterization is in terms of anytime capacity. Another situation is where the channel is used for the purpose of control as well as communication. It is for this reason that the high-level view of not reducing reliable communication to finite dimensional problems is important. In this paper, a high-level view is taken for the problem of communication with a fidelity criterion.

In addition to considering a general channel, since the decoding rule for the channel-coding problem depends only on the end-to-end description that the channel communicates the uniform X source to within distortion level D, assuming random codes are permitted, duality also holds for a compound channel, that is, where the channel belongs to a set of channels (see for example [2] for a discussion on compound channels). For the same reason, source-channel separation for communication with a fidelity criterion also holds for a general, compound channel assuming random codes are permitted. This will be discussed in some detail, later.

The source U is ideal for this purpose because it puts mass only on the set of sequences with a particular type. If one tries to carry out the above argument for the i.i.d. X source, ϵs and δs enter the picture. In this sense, the uniform source is more fundamental than the i.i.d. source. A generalization to the i.i.d. X source can be carried out. The details for the i.i.d. X source can be found in [1] and will be reported in a later paper.

2. Literature survey

Duality between source-coding and channel-coding has been discussed in a number of settings in the information-theory literature.

Shannon [8] discussed, at a high level, a duality between source-coding and channel-coding. Quoting directly from [8],

"There is a curious and provocative duality between the properties of a source with a distortion measure and those of a channel. This duality is enhanced if we consider channels in which there is a "cost" associated with the different input letters, and it is desired to find the capacity subject to the constraint that the expected cost not exceed a certain quantity. Thus input letter i might have cost a_i and we wish to find the capacity with the side condition $\sum_i P_i a_i \leq a$, say, where P_i is the probability of using input letter i. This problem amounts, mathematically, to maximizing a mutual information under variation of the P_i with a linear inequality as constraint. The solution of this problem leads to a capacity of the P_i with a linear inequality as constraint. The solution of this problem leads to a capacity cost function $C(a)$ for the channel. It can be shown readily that this function is concave downward. Solving this problem corresponds, in a sense, to finding a source that is just right for the channel and the desired cost.

In a somewhat dual way, evaluating the rate-distortion function $R(d)$ for a source amounts, mathematically, to *minimizing* a mutual information under variation of the $q_i(j)$, again, with a linear inequality as constraint. The solution leads to a function $R(d)$ which is convex downward. Solving this problem corresponds to finding a channel that is just right for the source and allowed distortion level. This duality can be pursued further and is related to a duality between past and future and the notions of control and knowledge. Thus we may have knowledge of the past but cannot control it; we may control the future but have no knowledge of it."

A general formulation of this functional duality has been posed in [6] where the channel capacity with cost constraints problem and the rate-distortion problem, are defined as problems which are "duals" of each other in the mathematical programming sense. It is proved that channel capacity is equal to the rate-distortion function if the problems are duals to each other. The purpose of our paper is not to demonstrate a functional duality or a mathematical programming based duality, but an operational duality as defined in the previous section.

Operational duality, as defined by Ankit et al [5] refers to the property that optimal encoding/decoding schemes for one problem lead to optimal encoding/decoding schemes for the corresponding dual problem. They show that if used as a lossy compressor, the maximum-likelihood channel decoder of a randomly chosen capacity-achieving codebook achieves the rate-distortion function almost surely . *Note that the definition of operational used in [5] is different from the definition of operational used in this paper.*

Csiszar and Korner [2] prove the rate-distortion theorem by first constructing a "backward" DMC and codes for this DMC such that source-codes meeting the distortion criterion are obtained from this channel code by using the channel decoder as a source encoder and vice-versa; in this scheme, channel codes with large error probability are needed. The view-point is suggestive of a duality between source and channel coding. No backward channel is used in our paper: we use a forward channel which directly communicates the source U to within distortion level D and make use of the corresponding rate-distortion source-coding problem.

Yassaee et al [10] have studied duality between the channel coding problem and the secret-key agreement problem (in the source-model sense). They show how an achievability proof for one of these problems can be converted into an achievability proof for the other.

The decoding rule used in this paper is a variant of a minimum distortion decoding rule. For discrete memoryless channels, decoders minimizing a distortion measure have been studied as mis-match decoding and are suboptimal in general though optimal if the distortion measure is matched, that is, equal to the negative log of the channel transition probability; see for example the paper of Csiszar and Narayan [3].

3. Notation and definitions

Superscript n will denote a quantity related to block-length n. For example, x^n will be the channel input when the block-length is n. As block-length varies, $x =< x^n >_1^\infty$ will denote the sequence for various block-lengths.

The source input space is $\mathcal{X}$ and the source reproduction space is $\mathcal{Y}$. $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. X is a random variable on $\mathcal{X}$. Let $p_X(x)$ be rational $\forall x$. Let n_0 denote the least positive integer for which $n_0 p_X(x)$ is an integer $\forall x \in \mathcal{X}$. Let $\mathcal{U}^n$ denote the set of sequences with (*exact*) type p_X. $\mathcal{U}^n$ is nonempty if and only if n_0 divides n. Let $n' \triangleq n_0 n$. Let $U^{n'}$ denote a random variable which is uniform on $\mathcal{U}^{n'}$ and zero elsewhere. Then, $< U^{n'} >_1^\infty$ is the uniform X source and is denoted by U. The uniform X source can be defined only for those X for which $p_X(x)$ is rational $\forall x \in \mathcal{X}$.

Every mathematical entity which had a superscript n in Section 1 will have a superscript n' henceforth. This is because the uniform X source is defined only for block-lengths n'. The reader is urged not to get confused between this change of superscript between Section 1 and the rest of this paper. Further, the reader is urged to read Section 1 by replacing n with n' in mathematical entities.

Let q denote a type on the set $\mathcal{Y}$ which is achievable when the block-length is n'. $\mathcal{V}_q^{n'}$ is the set of all sequences with type q. The uniform distribution on $\mathcal{V}_q^{n'}$ is $V_q^{n'}$.

Since the uniform X source is defined only for block-lengths n', distortion measure, channels, encoders and decoders will be defined only for block-lengths n'.

$d =< d^{n'} >_1^\infty$ is the distortion measure where $d^{n'} : \mathcal{X}^{n'} \times \mathcal{Y}^{n'} \to [0, \infty)$. Let $\pi^{n'}$ be a permutation (rearrangement) of $(1, 2, \dots, n')$. That is, for $1 \leq i \leq n'$, $\pi^{n'}(i) \in \{1, 2, \dots, n'\}$ and that, $\pi^{n'}(i)$, $1 \leq i \leq n'$ are different. For $x^{n'} \in \mathcal{X}^{n'}$, denote

$$\pi^{n'} x^{n'} \triangleq (x^{n'}(\pi^{n'}(1)), x^{n'}(\pi^{n'}(2)), \dots, x^{n'}(\pi^{n'}(n'))). \tag{4}$$

For $y^{n'} \in \mathcal{Y}^{n'}$, $\pi^{n'} y^{n'}$ is defined analogously. $< d^{n'} >_1^\infty$ is said to be permutation invariant if $\forall n'$,

$$d^{n'}(\pi^{n'} x^{n'}, \pi^{n'} y^{n'}) = d^{n'}(x^{n'}, y^{n'}), \forall x^{n'} \in \mathcal{X}^{n'}, y^{n'} \in \mathcal{Y}^{n'}. \tag{5}$$

An additive distortion measure is defined as follows. Let $d : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ be a function. Define

$$d^{n'}(x^{n'}, y^{n'}) = \sum_{i=1}^{n'} d(x^{n'}(i), y^{n'}(i)). \tag{6}$$

Then, $< d^{n'} >_1^\infty$ is an additive distortion measure.

Additive distortion measures are special cases of permutation invariant distortion measure. Except at the end of the paper where conditions are derived for a certain technical condition to be true for which additive distortion measures will be required, the rest of this paper will use permutation invariant distortion measures.

A general channel $b =< b^{n'} >_1^\infty$ is defined as follows:

The input space of the channel is $\mathcal{X}$ and the output space is $\mathcal{Y}$.

$$\begin{aligned} b^{n'} : & \mathcal{X}^{n'} \to \mathcal{P}(\mathcal{Y}^{n'}) \\ & x^{n'} \to b^{n'}(y^{n'} | x^{n'}) \end{aligned} \tag{7}$$

$b^{n'}(y^{n'}|x^{n'})$ should be thought of as the probability that the output of the channel is $y^{n'}$ given that the input is $x^{n'}$.

Note that the channel model is general in the sense of [9]. In particular, no ergodicity or information stability requirements on the channel are needed.

Let

$$\mathcal{M}_R^{n'} \triangleq \{1, 2, \ldots, 2^{\lfloor n'R \rfloor}\} \tag{8}$$

$\mathcal{M}_R^{n'}$ is the message set. When the block-length is n', a rate R deterministic source encoder is $e_s^{n'} : \mathcal{X}^{n'} \to \mathcal{M}_R^{n'}$ and a rate R deterministic source decoder $f_s^{n'} : \mathcal{M}_R^{n'} \to \mathcal{Y}^{n'}$. $(e_s^{n'}, f_s^{n'})$ is the block-length n' rate R deterministic source-code. The source-code is allowed to be random in the sense that the encoder-decoder is a joint probability distribution on the space of deterministic encoders and decoders. $< e_s^{n'}, f_s^{n'} >_1^\infty$ is the rate R source-code. The classic argument used in [8] to prove the achievability part of the rate-distortion theorem uses a random source code.

When the block-length is n', a rate R deterministic channel encoder is a map $e_c^{n'} : \mathcal{M}_R^{n'} \to \mathcal{X}^{n'}$ and a rate R deterministic channel decoder is a map $f_c^{n'} : \mathcal{Y}^{n'} \to \hat{\mathcal{M}}_R^{n'}$ where $\hat{\mathcal{M}}_R^{n'} \triangleq \mathcal{M}_R^{n'} \cup \{e\}$ is the message reproduction set where 'e' denotes error. The encoder and decoder are allowed to be random in the sense discussed previously. $< e_c^{n'}, f_c^{n'} >_1^\infty$ is the rate R channel code. The classic argument used in [7] to derive the achievability of the mutual information expression for channel capacity uses a random channel code.

The source-code $< e_s^{n'}, f_s^{n'} >_1^\infty$ is said to code the source U to within a distortion level D if with input $U^{n'}$ to $e_s^{n'} \circ f_s^{n'}$, the output is $Y^{n'}$ such that

$$\lim_{n' \to \infty} \Pr\left(\frac{1}{n'} d^{n'}(U^{n'}, Y^{n'}) > D\right) = 0. \tag{9}$$

(9) is the probability of excess distortion criterion. The infimum of rates needed to code the uniform X source to within the distortion level D is the rate-distortion function $R_U^P(D)$. If lim in (9) is replaced with lim inf, the criterion is called the inf probability of excess distortion criterion and the corresponding rate-distortion function is denoted by $R_U^P(D, \inf)$. The superscript 'P' in $R_U^P(D)$ and $R_U^P(D, \inf)$ signifies the use of the probability of excess distortion criterion as opposed to the expected distortion criterion used frequently in literature. Note that (9) could alternatively be defined by using lim sup instead of lim; the two are equivalent in this case because if the lim sup of a sequence of non-negative real numbers is zero, the limit of the sequence of numbers exists and is also zero.

Denote

$$g =< g^{n'} >_1^\infty \triangleq < e_c^{n'} \circ b^{n'} \circ f_c^{n'} >_1^\infty . \tag{10}$$

Then, g is a general channel with input space $\mathcal{M}_R^{n'}$ and output space $\hat{\mathcal{M}}_R^{n'}$. Rate R is said to be reliably achievable over b if there exists a rate R channel code $< e_c^{n'}, f_c^{n'} >_1^\infty$ such that

$$\lim_{n' \to \infty} \sup_{m^{n'} \in \mathcal{M}_R^{n'}} g^{n'}(\{m^{n'}\}^c | m^{n'}) = 0. \tag{11}$$

Supremum of all achievable rates is the capacity of b.

The channel b is said to communicate the source U *directly* to within distortion level D if with input $U^{n'}$ to $b^{n'}$, the output is $Y^{n'}$ such that

$$\lim_{n' \to \infty} \Pr\left(\frac{1}{n'} d^{n'}(U^{n'}, Y^{n'}) > D\right) = 0. \tag{12}$$

See Figure 1 in Section 1 with n replaced by n'.

In this paper, only the end-to-end description of a channel $< b^{n'} >_1^\infty$ that it communicates the uniform X source directly to within distortion level D is used. Exact knowledge of the kernel $b^{n'}$ is not used. For this reason, the general channel should be thought of as a *black-box* which communicates the uniform X source to within distortion level D.

4. Randomized covering-packing duality

Theorem 1. *Let b directly communicate source U to within distortion level D under a permutation invariant distortion measure d. Assume that $R_U^P(D) = R_U^P(D, \inf)$. Then, reliable communication can be accomplished over b at rates $< R_U^P(D)$. In other words, the capacity of b, $C \geq R_U^P(D)$.*

Note that the technical condition $R_U^P(D) = R_U^P(D, \inf)$ can be proved for an additive distortion measure. See the discussion following the proof of the theorem.

Proof. Since b directly communicates source U to within distortion level D, there exists a sequence $\omega_{n'}$, $\omega_{n'} \to 0$ as $n' \to \infty$ such that

$$\lim_{n' \to \infty} \Pr\left(\frac{1}{n'} d^{n'}(U^{n'}, Y^{n'}) > D\right) \leq \omega_{n'} \tag{13}$$

The proof of the theorem will be carried out via parallel random-coding arguments for two problems:

- *Channel-coding problem:* Rates of reliable communication over b.

- *Source-coding problem:* Rates of coding for the uniform X source with a distortion D under the inf probability of excess distortion criterion.

Codebook generation:

- *Codebook generation for the channel-coding problem:* Let reliable communication be desired at rate R. Generate $2^{\lfloor n'R \rfloor}$ sequences independently and uniformly from $\mathcal{U}^{n'}$. This is the codebook.
- *Codebook generation for the source-coding problem:* Let source-coding be desired at rate R. Generate $2^{\lfloor n'R \rfloor}$ codewords independently and uniformly from $\mathcal{V}_q^{n'}$ for some type q on $\mathcal{Y}$ which is achievable for block-length n'.

Mathematical description of codebook:

- *Mathematical description of the codebook for the channel coding problem:* Denote by $\mathfrak{K}^{n'}$, the set of all sequences of sequences $(u_1^{n'}, u_2^{n'}, \dots, u_{2^{\lfloor n'R \rfloor}}^{n'})$ where $u_i^{n'} \in \mathcal{U}^{n'}$ $\forall i \in \{1, 2, \dots, 2^{\lfloor n'R \rfloor}\}$. The uniform distribution on this set is the rate R codebook. The particular realization of this codebook will be denoted by $\mathfrak{k}^{n'}$. The j^{th} coordinate of $\mathfrak{k}^{n'}$, that is, the j^{th} codeword, will be denoted by $\mathfrak{k}^{n'}(j)$, $j \in \{1, 2, \dots, 2^{\lfloor n'R \rfloor}\}$.
- *Mathematical description of the codebook for the source coding problem:* Denote by $\mathfrak{L}^{n'}$, the set of all sequences of sequences $(y_1^{n'}, y_2^{n'}, \dots, y_{2^{\lfloor n'R \rfloor}}^{n'})$ where $y_i^{n'} \in \mathcal{V}_q^{n'}$ $\forall i \in \{1, 2, \dots, 2^{\lfloor n'R \rfloor}\}$. The uniform distribution on this set is the rate R codebook. The particular realization of this codebook will be denoted by $\mathfrak{l}^{n'}$. The j^{th} coordinate of $\mathfrak{l}^{n'}$, that is, the j^{th} codeword will be denoted by $\mathfrak{l}^{n'}(j)$, $j \in \{1, 2, \dots, 2^{\lfloor n'R \rfloor}\}$.

Joint typicality:

Joint typicality for both the channel-coding and source-coding problems is defined as follows: $(u^{n'}, y^{n'}) \in \mathcal{U}^{n'} \times \mathcal{Y}^{n'}$ jointly typical if

$$\frac{1}{n'} d^{n'}(u^{n'}, y^{n'}) \leq D. \tag{14}$$

Decoding and encoding:

- *Decoding for the channel-coding problem:* Let $y^{n'}$ be received. If there exists unique $i \in \{1, 2, \dots, 2^{\lfloor n'R \rfloor}\}$ such that $(\mathfrak{k}^{n'}(i), y^{n'})$ jointly typical, declare that $\mathfrak{k}^{n'}(i)$ is transmitted, else declare error.
- *Encoding for the source-coding problem:* Let $u^{n'} \in \mathcal{U}^{n'}$ need to be source-coded. If there exists some $i \in \{1, 2, \dots, 2^{\lfloor n'R \rfloor}\}$ for which

$(u^{n'}, \mathfrak{l}^{n'}(i))$ jointly typical, encode $u^{n'}$ to one such $\mathfrak{l}^{n'}(i)$, else declare error.

Some notation:

- *Notation for the channel-coding problem:* Let message $m^{n'} \in \mathcal{M}_R^{n'}$ be transmitted. Codeword corresponding to $m^{n'}$ is $u_c^{n'}$. Non-transmitted codewords are $u'^{n'}_1, u'^{n'}_2, \ldots, u'^{n'}_{2^{\lfloor n'R \rfloor}-1}$. $u_c^{n'}$ is a realization of $U_c^{n'}$. $U_c^{n'}$ is uniform on $\mathcal{U}^{n'}$. $u'^{n'}_i$ is a realization of $U'^{n'}_i$. $U'^{n'}_i$ is uniform on $\mathcal{U}^{n'}$, $1 \le i \le 2^{\lfloor n'R \rfloor} - 1$. $U_c^{n'}, U'^{n'}_i, 1 \le i \le 2^{\lfloor n'R \rfloor} - 1$ are independent of each other. The channel output is $y^{n'}$. $y^{n'}$ is a realization of $Y^{n'}$. $y^{n'}$ may depend on $u_c^{n'}$ but does not depend on $u'^{n'}_i, 1 \le i \le 2^{\lfloor n'R \rfloor} - 1$. As random variables, $Y^{n'}$ and $U_c^{n'}$ might be dependent but $Y^{n'}, U'^{n'}_i, 1 \le i \le 2^{\lfloor n'R \rfloor} - 1$ are independent. If the type q of the sequence $y^{n'}$ needs to be explicitly denoted, the sequence is denoted by $y_q^{n'}$. $\mathcal{G}^{n'}$ is the set of all achievable types q on $\mathcal{Y}$ for block-length n'.
- *Notation for the source-coding problem:* $u_s^{n'}$ is the sequence which needs to be source-coded. $u_s^{n'}$ is a realization of $U_s^{n'}$ which is uniformly distributed on $\mathcal{U}^{n'}$. The codewords are $y^{n'}_{q,i}, 1 \le i \le 2^{\lfloor n'R \rfloor}$ where q denotes the type. $y^{n'}_{q,i}$ is a realization of $V^{n'}_{q,i}, 1 \le i \le 2^{\lfloor n'R \rfloor}$ where $V^{n'}_{q,i}$ is uniformly distributed on the subset of $\mathcal{Y}^{n'}$ consisting of all sequences with type q. $u_s^{n'}, y^{n'}_{q,i}, 1 \le i \le 2^{\lfloor n'R \rfloor}$ are independently generated; as random variables, $U_s^{n'}, Y^{n'}_{q,i}, 1 \le i \le 2^{\lfloor n'R \rfloor}$ are independent. $\mathcal{G}^{n'}$ is the set of all achievable types q on $\mathcal{Y}$ for block-length n'

Error analysis: For the channel-coding problem, the probability of correct decoding is analyzed and for the source-coding problem, the probability of error is analyzed.

- *Event of correct decoding for the channel-coding problem:* From the encoding-decoding rule, it follows that the event of correct decoding given that a particular message is transmitted is the set

$$\begin{aligned}(15)\quad &\Big\{((u_c^{n'}, u'^{n'}_1, u'^{n'}_2, \ldots, u'^{n'}_{2^{\lfloor n'R \rfloor}-1}), y^{n'}) \mid \\ &\frac{1}{n'} d^{n'}(u_c^{n'}, y^{n'}) \le D \text{ and } \frac{1}{n'} d^{n'}(u'^{n'}_i, y^{n'}) > D, 1 \le i \le 2^{\lfloor n'R \rfloor} - 1\Big\}.\end{aligned}$$

Henceforth, this event will be denoted by

$$(16)\qquad \left\{\frac{1}{n'} d^{n'}(U_c^{n'}, Y^{n'}) \le D\right\} \cap \cap_{i=1}^{2^{\lfloor n'R \rfloor}-1} \left\{\frac{1}{n'} d^{n'}(U'^{n'}_i, Y^{n'}) > D\right\}.$$

- *Error event for the source-coding problem:* From the encoding-decoding rule, it follows that the error event given that a particular message $u^{n'}$ needs to be source-coded is

$$\left\{(y_{q,1}, y_{q,2}, \ldots, y^{n'}_{q,2^{\lfloor nR \rfloor}}) \mid \frac{1}{n'} d^{n'}(u^{n'}, y^{n'}_{q,i}) > D, 1 \le i \le 2^{\lfloor nR \rfloor}\right\}. \tag{17}$$

Henceforth, this event will be denoted by

$$\cap_{i=1}^{2^{\lfloor n'R \rfloor}} \left\{\frac{1}{n'} d^{n'}(u^{n'}, V^{n'}_{q,i}) > D\right\}. \tag{18}$$

Note that there is freedom in the way q is chosen for codebook generation.

Probability distributions with respect to which error analyses need to be carried out:

- *Probability distributions with respect to which error analysis need to be carried out for the channel-coding problem:* From the codebook generation and the action of the channel, it follows that the probability distribution with respect to which we need to do calculations is

$$\Pr(((u^{n'}_c, u'^{n'}_1, u'^{n'}_2, \ldots, u'^{n'}_{2^{\lfloor n'R \rfloor}-1}), y^{n'})) = \frac{1}{|\mathcal{U}^{n'}|^{2^{\lfloor n'R \rfloor}}} c^{n'}(y^{n'}|u^{n'}_c). \tag{19}$$

- *Probability distributions with respect to which error analysis need to be carried out for the source-coding problem:* From the codebook generation, it follows that the probability distribution with respect to which we need to do calculation is

$$\Pr((y^{n'}_{q,1}, y^{n'}_{q,2}, \ldots, y^{n'}_{2^{\lfloor n'R \rfloor}})) = \frac{1}{|\mathcal{V}^{n'}_q|^{2^{\lfloor n'R \rfloor}}}. \tag{20}$$

Calculation:

- *Calculation of the probability of correct decoding for the channel-coding problem:*

  Bound for probability of event (16):

(21)

$$
\begin{aligned}
&\Pr\left(\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'}) \le D\right\} \cap \cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\}\right)\\
&= \Pr\left(\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'}) \le D\right\}\right) + \Pr\left(\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\}\right)\\
&\quad - \Pr\left(\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'}) \le D\right\} \cup \cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\}\right)\\
&\ge (1-\omega_{n'}) + \Pr\left(\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\}\right) - 1\\
&= -\omega_{n'} + \Pr\left(\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\}\right)\\
&= -\omega_{n'} + \prod_{i=1}^{2^{\lfloor n'R\rfloor}-1}\Pr\left(\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\}\right)\\
&\text{(since } U'^{n'}_i, 1 \le i \le 2^{\lfloor n'R\rfloor}-1,\ Y^{n'} \text{ are independent random variables)}\\
&= -\omega_{n'} + \left[\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},Y^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R\rfloor}-1}\\
&\text{(where } U^{n'} \text{ is uniform on } \mathcal{U}^{n'} \text{ and is independent of } Y^{n'})\\
&= -\omega_{n'} + \left[\sum_{y^{n'}\in\mathcal{Y}^{n'}} p_{Y^{n'}}(y^{n'})\Pr\left(\frac{1}{n'}d^{n'}(U^{n'},Y^{n'}) > D \,\middle|\, Y^{n'}=y^{n'}\right)\right]^{2^{\lfloor n'R\rfloor}-1}\\
&= -\omega_{n'} + \left[\sum_{y^{n'}\in\mathcal{Y}^{n'}} p_{Y^{n'}}(y^{n'})\Pr\left(\frac{1}{n'}d^{n'}(U^{n'},y^{n'}) > D \,\middle|\, Y^{n'}=y^{n'}\right)\right]^{2^{\lfloor n'R\rfloor}-1}\\
&= -\omega_{n'} + \left[\sum_{y^{n'}\in\mathcal{Y}^{n'}} p_{Y^{n'}}(y^{n'})\Pr\left(\frac{1}{n'}d^{n'}(U^{n'},y^{n'}) > D\right)\right]^{2^{\lfloor n'R\rfloor}-1}\\
&\text{(since } U^{n'} \text{ and } Y^{n'} \text{ are independent)}\\
&\ge -\omega_{n'} + \left[\inf_{y^{n'}\in\mathcal{Y}^{n'}}\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R\rfloor}-1}\\
&= -\omega_{n'} + \left[\inf_{q\in\mathcal{G}^{n'}}\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y_q^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R\rfloor}-1}.
\end{aligned}
$$

The last equality above follows because

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y^{n'}) > D\right\}\right) \tag{22}$$

depends only on the type of $y^{n'}$; see the symmetry argument later.

Rate R is achievable if

$$-\,\omega_{n'} + \left[\inf_{q\in\mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y_q^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R\rfloor}-1} \to 1 \text{ as } n' \to \infty. \tag{23}$$

Since $\omega_{n'} \to 0$ as $n' \to \infty$, rate R is achievable if

$$\left[\inf_{q\in\mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y_q^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R\rfloor}-1} \to 1 \text{ as } n' \to \infty. \tag{24}$$

- *Calculation of probability of error for the source-coding problem:*

  Bound for probability of event (18) is calculated using standard arguments:

$$\begin{aligned}
&\Pr\left(\cap_{i=1}^{2^{\lfloor n'R\rfloor}} \left\{\frac{1}{n'}d^{n'}(u^{n'}, V_{q,i}^{n'}) > D\right\}\right) \\
&= \prod_{i=1}^{2^{\lfloor n'R\rfloor}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_{q,i}^{n'}) > D\right\}\right) \\
&= \left[\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_{q,i}^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R\rfloor}}
\end{aligned} \tag{25}$$

where $V_q^{n'}$ is uniform on $\mathcal{V}_q^{n'}$.

There is choice of $q \in \mathcal{G}^{n'}$. Thus, a bound for the probability of the event is

$$\left[\inf_{q\in\mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R\rfloor}}. \tag{26}$$

Since the inf probability of excess distortion criterion is used, it follows that rate R is achievable if

$$\left[\inf_{q\in\mathcal{G}^{n'_i}} \Pr\left(\left\{\frac{1}{n'_i}d^{n'_i}(u^{n'_i}, V_q^{n'_i}) > D\right\}\right)\right]^{2^{\lfloor n'_iR\rfloor}} \to 0 \text{ for some } n'_i = n_0 n_i,\ n_i \to \infty. \tag{27}$$

Connection/Duality between channel-coding and source-coding:

The calculation required in the channel-coding problem is

$$\inf_{q \in \mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, y_q^{n'}) > D\right\}\right) \tag{28}$$

and the calculation required in the source-coding problem is

$$\inf_{q \in \mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) \tag{29}$$

It will be proved that (28) and (29) are equal. It will be proved more generally that

$$\Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, y_q^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'} d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right). \tag{30}$$

This is a symmetry argument and requires the assumption of permutation invariant distortion measure. The idea is that the left hand side of (30) depends only on the type of $y_q^{n'}$. From this it follows that the left hand side of (30) is equal to

$$\Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, V_q^{n'}) > D\right\}\right) \tag{31}$$

where $V_q^{n'}$ is independent of $U^{n'}$. Similarly, the right hand side of (30) depends only on the type of $u^{n'}$ and from this it follows that the right hand side of (30) is also equal to (31). (30) follows. Details are as follows:

The first step is to prove that

$$\Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, y_q^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, {y'_q}^{n'}) > D\right\}\right) \tag{32}$$

for sequences $y_q^{n'}$ and ${y'_q}^{n'}$ with type q. Since $U^{n'}$ is the uniform distribution on $\mathcal{U}^{n'}$, it follows that it is sufficient to prove that the sets

$$\left\{u^{n'} : \frac{1}{n'} d^{n'}(u^{n'}, y_q^{n'}) > D\right\} \text{ and } \left\{u^{n'} : \frac{1}{n'} d^{n'}(u^{n'}, {y'_q}^{n'}) > D\right\} \tag{33}$$

have the same cardinality. ${y'_q}^{n'} = \pi^{n'} y_q^{n'}$ for some permutation $\pi^{n'}$ since ${y'_q}^{n'}$ and $y_q^{n'}$ have the same type. Denote the sets

$$\mathcal{B}_{y_q^{n'}} \triangleq \left\{u^{n'} : \frac{1}{n'} d^{n'}(u^{n'}, y_q^{n'}) > D\right\} \tag{34}$$

Set $\mathcal{B}_{y_q'^{n'}}$ is defined analogously.

Let $u^{n'} \in \mathcal{B}_{y_q^{n'}}$. Since the distortion measure is permutation invariant, $d^{n'}(\pi^{n'}u^{n'}, \pi^{n'}y_q^{n'}) = d^{n'}(u^{n'}, y_q^{n'})$. Thus, $\pi^{n'}u^{n'} \in \mathcal{B}_{y_q'^{n'}}$. If $u^{n'} \neq u'^{n'}$, $\pi^{n'}u^{n'} \neq \pi^{n'}u'^{n'}$. It follows that $|\mathcal{B}_{y_q'^{n'}}| \geq |\mathcal{B}_{y_q^{n'}}|$. Interchanging $y_q^{n'}$ and $y_q'^{n'}$ in the above argument, $|\mathcal{B}_{y_q^{n'}}| \geq |\mathcal{B}_{y_q'^{n'}}|$. It follows that $|\mathcal{B}_{y_q{}^{n'}}| = |\mathcal{B}_{y_q'^{n'}}|$. (32) follows.

Let V_q^n be independent of $U^{n'}$. From (32) it follows that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, V_q^{n'}) > D\right\}\right). \tag{35}$$

By an argument identical with the one used to prove (32), it follows that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(u'^{n'}, V_q^{n'}) > D\right\}\right) \tag{36}$$

for $u^{n'}, u'^{n'} \in \mathcal{U}^{n'}$. From (36) it follows that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, V_q^{n'}) > D\right\}\right). \tag{37}$$

From (35) and (37), (30) follows.

Proof that a channel which is capable of communicating the uniform X source to within a certain distortion level is also capable of communicating bits reliably at any rate less than the infimum of the rates needed to code the uniform X source with the same distortion level under the inf *probability of excess distortion criterion:*

Denote

$$\begin{aligned} A_{n'} &\triangleq \inf_{q \in \mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y_q^{n'}) > D\right\}\right) \\ &= \inf_{q \in \mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right). \end{aligned} \tag{38}$$

From (24), it follows that rate R is achievable for the channel-coding problem if

$$(A_{n'})^{2^{\lfloor n'R \rfloor}-1} \to 1 \text{ as } n' \to \infty. \tag{39}$$

From (27), it follows that rate R is achievable for the source-coding problem if

$$(40) \qquad (A_{n_i'})^{2^{\lfloor n_i' R \rfloor}} \to 0 \text{ as } n_i' \to \infty \text{ for some } n_i' = n_0 n_i \text{ for some } n_i \to \infty.$$

Let

$$(41) \qquad \alpha \triangleq \sup\{R \mid (39) \text{ holds}\}.$$

Then, if $R' > \alpha$,

$$(42) \qquad \lim_{n_i' \to \infty} (A_{n_i'})^{2^{\lfloor n_i R' \rfloor} - 1} < 1 \ \forall\ R' > \alpha \text{ for some sequence } n_i' \to \infty$$

n_i' may depend on R'.

Then,

$$(43) \qquad \lim_{n_i' \to \infty} (A_{n_i'})^{2^{\lfloor n_i' R'' \rfloor} - 1} = 0 \text{ for } R'' > R'$$

(42) and (43) hold for all $R'' > R' > \alpha$. It follows that rates larger than α are achievable for the source-coding problem.

Thus, a channel which is capable of communicating the uniform X source to within a certain distortion level is also capable of communicating bits reliably at any rate less than the infimum of the rates needed to code the uniform X source with the same distortion level under the inf probability of excess distortion criterion.

Wrapping up the proof of the theorem:

It follows that if source U is directly communicated over b to within distortion level D, then reliable communication can be accomplished over b at rates $< R_U^P(D, \inf)$. By use of the assumption $R_U^P(D) = R_U^P(D, \inf)$, it follows that reliable communication can be accomplished over b at rates $< R_U^P(D)$. In other words, the capacity of b, $C \geq R_U^P(D)$. □

5. Discussion and recapitulation

Randomized code constructions were made for a source-coding problem and a channel-coding problem and relation drawn between source-coding rates and channel coding-rates for the two problems. The source-coding problem is a covering problem and the channel-coding problem is a packing problem. For this reason, the connection is a randomized covering-packing connection. This duality between source-coding and channel coding is captured in (30).

Note that by Berger's lemma or the type covering lemma [2], that at least for additive distortion measures, there exist source codes of rates approaching $R^P(D)$ such that "balls" around codewords cover all sequences of type p_X, not only a large fraction of them. Thus, in (9), one does not need to take a limit; in other words, in the source-coding problem, one may not need to take a limit. Thus, a deterministic version of the source-coding problem is possible; however it is unclear, how to do the same for the channel-coding problem. For this reason, the randomized versions of the problems are needed.

The technical condition $R_U^P(D) = R_U^P(D, \inf)$ is made on the rate-distortion function. This technical condition holds for additive distortion measures, and an *operational* proof which uses code constructions and various properties and relations between code constructions is provided in Chapter 5 of [1].

A proof of source-channel separation for communication with fidelity criterion follows: Let k be a channel with input space $\mathcal{I}$ and output space $\mathcal{O}$. $\mathcal{I}$ and $\mathcal{O}$ are finite sets. An encoder $< e^n >_1^\infty$ is a sequence where the input space is $\mathcal{X}$ and output space is $\mathcal{I}$ and $< f^n >_1^\infty$ is a decoder with input space $\mathcal{O}$ and output space $\mathcal{Y}$. In general, the encoder and decoder are allowed to be random. If there exists encoder-decoder $< e^n, f^n >_1^\infty$ such that by use of this encoder-decoder, communication of source U to within distortion level D happens over a channel k, then, $b =< e^{n'} \circ k \circ f^{n'} >_1^\infty$ is a channel with input space $\mathcal{X}$ and output space $\mathcal{Y}$ which communicates the source U directly to within distortion level D. Thus, rates $< R_U^P(D)$ are achievable over b by use of some encoder-decoder $< E^{n'}, F^{n'} >_1^\infty$. For this reason, reliable communication is possible over k at rates $< R_U^P(D)$ by use of encoder-decoder $< E^{n'} \circ e^{n'}, f^{n'} \circ F^{n'} >_1^\infty$. In other words, the capacity of k is $\geq R_U^P(D)$. By use of the standard argument of source-coding followed by channel coding, if capacity of k is $> R_U^P(D)$, the uniform X Source can be communicated over k by source coding followed by channel coding. Optimality of source-channel separation for communication with a fidelity criterion follows. The proof uses only the operational meanings of capacity (maximum rate of reliable communication) and rate-distortion function (minimum rate needed to compress a source with certain distortion), and randomized code constructions for these problems instead of using finite-dimensional functional simplifications or finite dimensional information theoretic definitions, for example, capacity as maximum mutual information and rate-distortion function as minimum mutual information, unlike in the traditional proof of Shannon [8]. Functional simplifications are carried out to the extent of (30).

Note that whether a view or a proof is operational (in the sense used in this paper) cannot be defined mathematically precisely. However, the same can be sensed intuitively from the context in which it is used.

By use of a perturbation argument, the results can be generalized to the i.i.d. X source (general p_X, not necessarily those for which $p_X(x)$ is rational) for additive distortion measures as discussed in Chapter 5 of [1].

Finally, note that the argument to prove Theorem 1 uses random codes. However, if the channel is a single channel, existence of a random code implies the existence of a deterministic code. Note further, that in the decoding rule in Theorem 1, only the end-to-end description that the channel communicates the uniform X source to within distortion level D is used, and not the particular $< b^{n'}{}_1^\infty$. For this reason, even if the channel belongs to a set, that is, the channel is compound in the sense of [2], Theorem 1 still holds. However, random codes would be needed since the argument to go from a random code to a deterministic code does not hold for a compound channel. For the same reason, a universal source channel separation theorem for communication with a fidelity criterion where universality is over the channel (channel is compound) holds if random codes are permitted. Precise details of a general, compound channel, what it means for a general, compound channel to communicate the uniform X source to within distortion level D, the capacity of a general, compound channel, and a precise statement of the theorem which incorporates universality over the channel are omitted here. The details can be found in [1], and will be reported in a later paper.

References

[1] M. Agarwal, *A universal, operational theory of multi-user communication with fidelity criteria*, Ph.D. thesis, Massachusetts Institute of Technology, (2012).

[2] I. Csiszár and J. Korner, *Information theory: coding theorems for discrete memoryless systems*, Akadémiai Kiadó, (1997).

[3] I. Csiszar and P. Narayan, *Channel capacity for a given decoding metric*, IEEE Transactions on Information Theory **41** (1995), no. 1, 35–43.

[4] M. Gastpar, *To code or not to code*, Ph.D. thesis, École Polytechnique Fédérale de Lausanne, (2002).

[5] A. Gupta and S. Verdu, *Operational duality between lossy compression and channel coding*, IEEE Transactions on Information Theory **57** (2011), no. 6, 3171–3179.

[6] S. S. Pradhan, J. Chou, and K. Ramchandran, *Duality between source and channel coding and its extension to the side information case*, IEEE Transactions on Information Theory **49** (2003), no. 5, 1181–1203.

[7] C. E. Shannon, *A mathematical theory of communication*, Bell System Technical Journal **27** (1948), 379–423 (Part 1) and 623–656 (Part 2).

[8] C. E. Shannon, *Coding theorems for a discrete source with a fidelity criterion*, Institute of Radio Engineers, National Convention Record **7** (part 4) (1959), 142–163.

[9] S. Verdu and T. S. Han, *A general formula for channel capacity*, IEEE Transactions on Information Theory **40** (1994), no. 4, 1147–1157.

[10] M. H. Yassaee, M. R. Aref, and P. A. Gohari, *Achievability proof via output statistics of random binning*, IEEE Transactions on Information Theory **60** (2014), no. 6, 6760–6786.

Appendix A. Intuitive explanation of the randomized covering-packing duality

This appendix explains on an intuitive level, the covering-packing duality. The authors emphasize that mathematically this section is imprecise and is only for the purpose of developing intuition.

A general channel which directly communicates the uniform X source to within a distortion level D can be intuitively thought of as follows: with high probability, a sequence in $\mathcal{U}^{n'}$ is communicated with distortion $\leq nD'$ and this probability $\to 0$ as $n' \to \infty$. See Figure 2 in Section 1 with n replaced with n'.

The *deterministic* (as opposed to randomized) covering-packing, or the source coding-channel coding problem in our setting, on an intuitive level is pictured in Figure A1. For the covering problem, with reference to this figure, $y^{n'} \in \mathcal{Y}^{n'}$ but balls of radius $n'D$ around $y^{n'}$ are made in the $\mathcal{U}^{n'}$ space. The source-coding question is: what is the minimum number of balls of radius $n'D$ which cover the $\mathcal{U}^{n'}$ space. In other words, what is the minimum number of $y^{n'} \in \mathcal{Y}^{n'}$ such that balls of radius nD around these $y^{n'} \in \mathcal{Y}^{n'}$ cover the $\mathcal{U}^{n'}$ space. For the packing problem, first recall the intuitive action of the channel depicted in Figure 2. With reference to the figure, for the packing problem, the question is to find the minimum number of what is the maximum number of $u^{n'} \in \mathcal{U}^{n'}$ such that balls of radius $n'D$ around these $u^{n'}$ pack the $\mathcal{Y}^{n'}$ space.

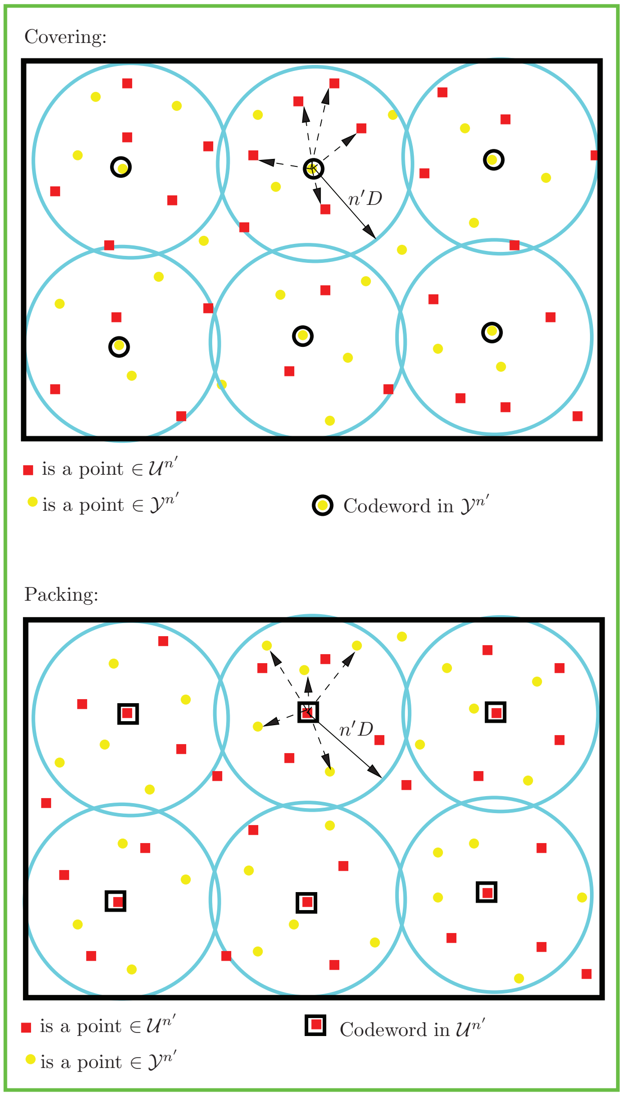

Figure A1: Covering: what is the minimum number of balls (equivalently, the number of codewords $\in \mathcal{Y}^{n'}$) with centers around certain $y^{n'} \in \mathcal{Y}^{n'}$ and balls in $\mathcal{U}^{n'}$ which cover the whole $\mathcal{U}^{n'}$ space. Packing: what is the maximum number of balls with (equivalently, the number of codewords $\in \mathcal{U}^{n'}$) centers around certain $u^{n'} \in \mathcal{U}^{n'}$ such that these balls pack the $\mathcal{Y}^{n'}$ space. Note that balls in the covering problem have centers $\in \mathcal{Y}^{n'}$ but the balls are in $\mathcal{U}^{n'}$ whereas balls in the packing problem have centers $\in \mathcal{U}^{n'}$ but the balls are in $\mathcal{Y}^{n'}$.

The randomized covering-packing picture is figuratively described in Figure A2.

In the covering problem, let the block-length be n'. Suppose $u^{n'}$ needs to be compressed. Suppose a codeword of type precisely q is generated uniformly from the set of all sequences with type precisely q. Denote this uniform distribution by $V_q^{n'}$ and a realization of $V_q^{n'}$ by $y^{n'}$. Probability that $y^{n'}$ will code $U^{n'}$ is

$$\Pr\left(\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) \le D\right). \tag{A.1}$$

This probability is independent of $u^{n'}$ by symmetry because the distortion metric is

$$\Pr\left(\frac{1}{n'}d^{n'}(U^{n'}, V_q^{n'}) \le D\right). \tag{A.2}$$

The way things *intuitively* work for increasing block-lengths, the number of sequences needed to code the source $U^{n'}$ if codewords of type q are used is approximately

$$\frac{1}{\Pr\left(\frac{1}{n'}d^{n'}(U^{n'}, V_q^{n'}) \le D\right)}. \tag{A.3}$$

q is arbitrary and thus, with this coding scheme, the number of codewords to code the uniform X source is approximately

$$\inf_q \frac{1}{\Pr\left(\frac{1}{n'}d^{n'}(U^{n'}, V_q^{n'}) \le D\right)} \triangleq \beta. \tag{A.4}$$

In general, there may be a scheme for which number of codewords needed is $\le \beta$.

In the packing problem, generate $2^{n'R}$ codewords independently and uniformly from $\mathcal{U}^{n'}$. Suppose $u^{n'}$ is transmitted. By the action of the channel, it follows that with high probability, $y^{n'}$ is received such that

$$\frac{1}{n'}d^{n'}(u^{n'}, y^{n'}) \le D. \tag{A.5}$$

Let the type of the received sequence $y^{n'}$ be q. Let $u^{n'}$ be another non-transmitted codeword which is generated using $U'^{n'}$. Note that $U^{n'}$ and $U'^{n'}$ are the same in distribution. Probability that there might be a mistake to say that $u'^{n'}$ is transmitted is

$$\frac{1}{n'}d^{n'}(U'^{n'}, y^{n'}) \leq D. \tag{A.6}$$

The above probability is the same for all $y^{n'}$ by symmetry because the distortion metric is permutation invariant, and hence, is equal to

$$\Pr\left(\frac{1}{n'}d^{n'}(U'^{n'}, V_q^{n'}) \leq D\right). \tag{A.7}$$

where $V_q^{n'}$ is defined in the above discussion on covering. Note that q is arbitrary and in order to get a bound on the total number of allowed codewords, the worst possible q needs to be considered. The way union bound works and the way things work for large block-lengths, the number of sequences which can be chosen as codewords for the channel-coding problem is

$$\inf_q \frac{1}{\Pr\left(\frac{1}{n'}d^{n'}(U'^{n'}, V_q^{n'}) \leq D\right)} = \beta. \tag{A.8}$$

In general, there may be a scheme for which number of codewords is $\geq \beta$.

Finally, note that the β in the covering and packing problem are the same. It follows that $C \geq R_U^P(D)$ where C is the capacity of the channel.

This is the intuitive base behind the proof of Theorem 1 and the resulting duality. Note further that this section is only for the sake of intuition and is mathematically imprecise. Precise proof have been provided in the proof of Theorem 1.

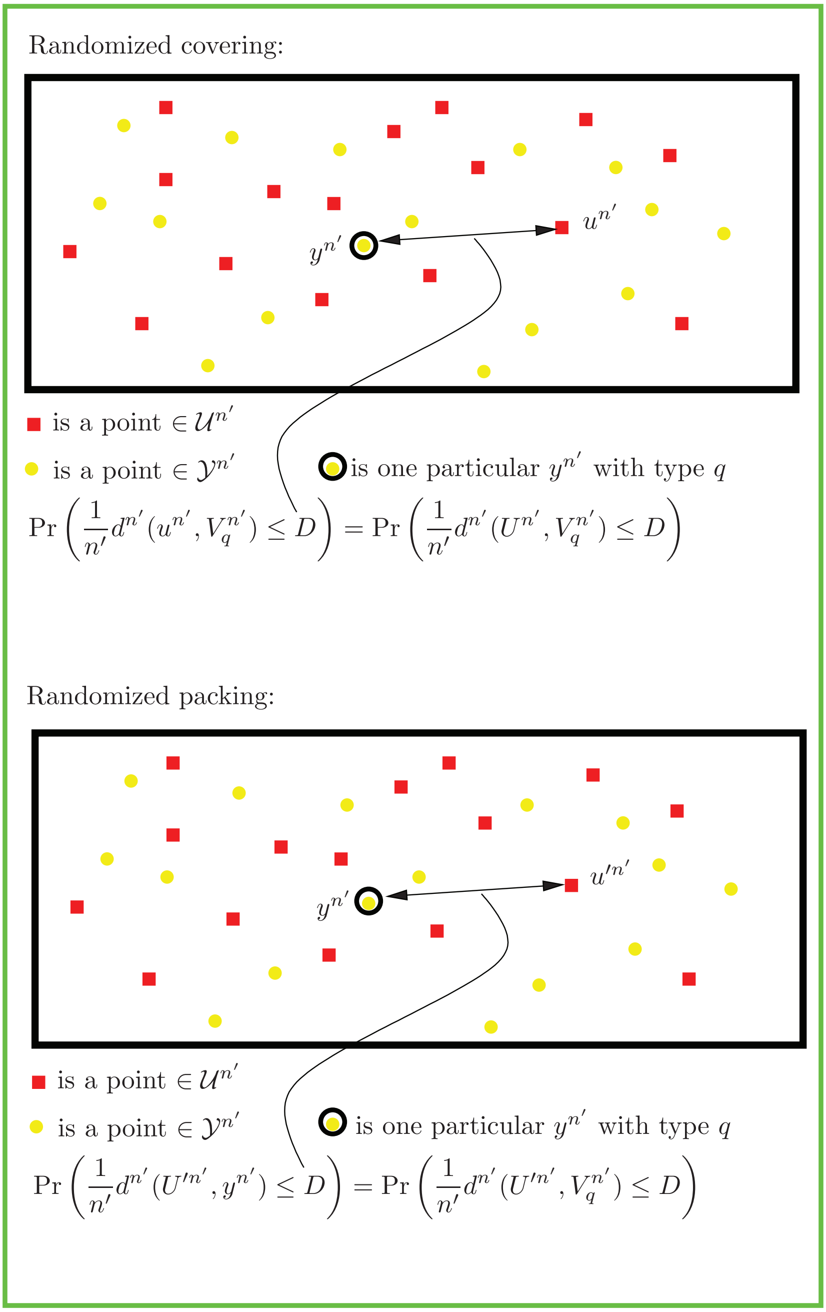

Figure A2: The randomized covering-packing picture for the problem of communication with a fidelity criterion.

Department of Electrical and Computer Engineering
Boston University, Boston, MA 02215, USA
E-mail address: magar@alum.mit.edu

Laboratory for information and decision systems
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology, MA 02139, USA
E-mail address: mitter@mit.edu

Received May 22, 2015