

Channels with cost constraints: strong converse and dispersion

Victoria Kostina, Sergio Verdú

Dept. of Electrical Engineering, Princeton University, NJ 08544, USA

Abstract—This paper shows the strong converse and the dispersion of memoryless channels with cost constraints. The analysis is based on a new non-asymptotic converse bound expressed in terms of the distribution of a random variable termed the b -tilted information density, which plays a role similar to that of the information density in channel coding without cost constraints. We also analyze the fundamental limits of lossy joint-source-channel coding over channels with cost constraints.

Index Terms—Converse, finite blocklength regime, channels with cost constraints, joint source-channel coding, strong converse, dispersion, memoryless sources, memoryless channels, Shannon theory.

I. INTRODUCTION

This paper is concerned with the maximum channel coding rate achievable at average error probability $\epsilon > 0$ where the cost of each codeword is constrained.

The capacity-cost function $C(\beta)$ of a channel specifies the maximum achievable channel coding rate compatible with vanishing error probability and with codeword cost not exceeding β in the limit of large blocklengths. In this paper we consider stationary memoryless channels with separable cost function, i.e.

- (i) $P_{Y^n|X^n} = P_{Y|X} \times \dots \times P_{Y|X}$, with $P_{Y|X}: \mathcal{A} \rightarrow \mathcal{B}$;
- (ii) $b_n(x^n) = \frac{1}{n} \sum_{i=1}^n b(x_i)$ where $b: \mathcal{A} \rightarrow [0, \infty]$.

In this case,

$$C(\beta) = \sup_{\mathbb{E}[b(X)] \leq \beta} I(X; Y) \quad (1)$$

A channel is said to satisfy the strong converse if $\epsilon \rightarrow 1$ as $n \rightarrow \infty$ for any code operating at a rate above the capacity. For memoryless channels without cost constraints, the strong converse was first shown by Wolfowitz: [1] treats the discrete memoryless channel (DMC), while [2] generalizes the result to memoryless channels whose input alphabet is finite while the output alphabet is the real line. Arimoto [3] showed a new converse bound stated in terms of Gallager's random coding exponent, which also leads to the strong converse for the DMC. Kemperman [4] showed that the strong converse holds for a DMC with feedback. For a particular discrete channel with finite memory, the strong converse was shown by Wolfowitz [5] and independently by Feinstein [6], a result soon generalized to a more general stationary discrete channel with finite memory [7]. In a more general setting not requiring the assumption of stationarity or finite memory, Verdú and

Han [8] showed a necessary and sufficient condition for a channel without cost constraints to satisfy the strong converse. In the special case of finite-input channels, that necessary and sufficient condition boils down to the capacity being equal to the limit of maximal normalized mutual informations. In turn, that condition is implied by the information stability of the channel [9], a condition which in general is not easy to verify.

As far as channel coding with input cost constraints, the strong converse for DMC with separable cost was shown by Csiszár and Körner [10, Theorem 6.11] and by Han [11, Theorem 3.7.2]. Regarding continuous channels, the strong converse has only been studied in the context of additive Gaussian noise channels with the cost function being the power of the channel input block, $b_n(x^n) = \frac{1}{n} |x^n|^2$. In the most basic case of the memoryless additive white Gaussian noise (AWGN) channel, the strong converse was shown by Shannon [12] (contemporaneously with Wolfowitz's finite alphabet strong converse). Yoshihara [13] proved the strong converse for the time-continuous channel with additive Gaussian noise having an arbitrary spectrum and also gave a simple proof of Shannon's strong converse result. Under the requirement that the power of each message converges stochastically to a given constant β , the strong converse for the AWGN channel with feedback was shown by Wolfowitz [14]. Note that in all those analyses of the power-constrained AWGN channel the cost constraint is meant on a per-codeword basis. In fact, the strong converse ceases to hold if the cost constraint is averaged over the codebook.

Channel dispersion quantifies the backoff from capacity, unescapable at finite blocklengths due to the random nature of the channel coming into play, as opposed to the asymptotic representation of the channel as a deterministic bit pipe of a given capacity. Polyansky et al. [15] found the dispersion of the DMC without cost constraints as well as that of the AWGN channel with a power constraint. Hayashi [16] showed the dispersion of the DMC with and without cost constraints. For constant composition codes over DMC, Moulin [17] found the dispersion and refined the third order term in the expansion of the maximum achievable code rate, under regularity conditions.

In this paper, we show a new non-asymptotic converse bound for general channels with input cost constraints in terms of a random variable we refer to as the b -tilted information density, which parallels the notion of d -tilted information for lossy compression [18]. Not only does the new bound lead to a general strong converse result but it is also tight enough to

This work was supported in part by the National Science Foundation (NSF) under Grant CCF-1016625 and by the Center for Science of Information (CSol), an NSF Science and Technology Center, under Grant CCF-0939370.

find the channel dispersion-cost function, when coupled with the corresponding achievability bound. More specifically, we show that for the DMC, $M^*(n, \epsilon, \beta)$, the maximum achievable code size at blocklength n , error probability ϵ and cost β , is given by

$$M^*(n, \epsilon, \beta) = nC(\beta) - \sqrt{nV(\beta)}Q^{-1}(\epsilon) + o(\sqrt{n}) \quad (2)$$

where $V(\beta)$ is the dispersion-cost function, and $Q^{-1}(\cdot)$ is the inverse of the Gaussian complementary cdf. Satisfyingly, the capacity-cost and the dispersion-cost functions are given by the mean and the variance of the b-tilted information density, juxtaposing nicely with the corresponding results in [15] (conventional information density without cost constraints) and [18] (d-tilted information in rate-distortion theory). In addition, we perform a refined analysis of the $o(\sqrt{n})$ term in (2). In particular, we conclude that, under mild regularity assumptions, the third order term is equal to $\frac{1}{2} \log n + O(1)$, thereby refining Hayashi's result [16].

Section II introduces the b-tilted information density. Section III states the new non-asymptotic converse bound which holds for a general channel with cost constraints, without making any assumptions on the channel (e.g. alphabets, stationarity, memorylessness). An asymptotic analysis of the converse and achievability bounds, including the proof of the strong converse and the expression for the channel dispersion-cost function, is presented in Section IV. Section V generalizes the results in Sections III and IV to the lossy joint source-channel coding setup.

II. b-TILTED INFORMATION DENSITY

In this section, we introduce the concept of b-tilted information density and several relevant properties.

Fix the transition probability kernel $P_{Y|X}: \mathcal{X} \rightarrow \mathcal{Y}$ and the cost function $b: \mathcal{X} \mapsto [0, \infty]$. In the application of this single-shot approach in Section IV, \mathcal{X} , \mathcal{Y} , $P_{Y|X}$ and b will become \mathcal{A}^n , \mathcal{B}^n , $P_{Y^n|X^n}$ in (i) and b^n in (ii), respectively. Denote

$$\mathbb{C}(\beta) = \sup_{\substack{P_X: \\ \mathbb{E}[b(X)] \leq \beta}} I(X; Y) \quad (3)$$

$$\lambda^* = \mathbb{C}'(\beta) \quad (4)$$

For a pair $\{P_{Y|X}, P_{\bar{Y}}\}$, define

$$\imath_{X;\bar{Y}}(x; y) = \log \frac{dP_{Y|X=x}}{dP_{\bar{Y}}}(y) \quad (5)$$

The familiar information density $\imath_{X;Y}(x; y)$ between realizations of two random variables with joint distribution $P_X P_{Y|X}$ follows by particularizing (5) to $\{P_{Y|X}, P_Y\}$, where $P_X \rightarrow P_{Y|X} \rightarrow P_Y$ ¹. In general, however, the function in (5) does not require $P_{\bar{Y}}$ to be induced by any input distribution.

Further, define the function

$$j_{X;\bar{Y}}(x; y, \beta) = \imath_{X;\bar{Y}}(x; y) - \lambda^*(b(x) - \beta) \quad (6)$$

¹We write $P_X \rightarrow P_{Y|X} \rightarrow P_Y$ to indicate that P_Y is the marginal of $P_X P_{Y|X}$, i.e. $P_Y(y) = \sum_{x \in \mathcal{X}} P_{Y|X}(y|x) P_X(x)$.

The special case of (6) with $P_{\bar{Y}} = P_{Y^*}$, where P_{Y^*} is the unique output distribution that achieves the supremum in (3), defines b-tilted information density:

Definition 1 (b-tilted information density). *The b-tilted information density between $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ is $j_{X;Y^*}(x; y, \beta)$.*

Since P_{Y^*} is unique even if there are several (or none) input distributions P_{X^*} that achieve supremum in (3), there is no ambiguity in Definition 1. If there are no cost constraints (i.e. $b(x) = 0 \forall x \in \mathcal{X}$), then $\mathbb{C}'(\beta) = 0$ regardless of β , and

$$j_{X;Y^*}(x; y, \beta) = \imath_{X;Y^*}(x; y) \quad (7)$$

The counterpart of the b-tilted information density in rate-distortion theory is the d-tilted information [18].

Denote

$$\beta_{\min} = \inf_{x \in \mathcal{X}} b(x) \quad (8)$$

$$\beta_{\max} = \sup \{ \beta \geq 0 : \mathbb{C}(\beta) < \mathbb{C}(\infty) \} \quad (9)$$

A nontrivial generalization of the well-known properties of information density in the case of no cost constraints, the following result highlights the importance of b-tilted information density in the optimization problem (3). It will be of key significance in the asymptotic analysis in Section IV.

Theorem 1. *Fix $\beta_{\min} < \beta < \beta_{\max}$. Assume that P_{X^*} achieving (3) is such that*

$$\mathbb{E}[b(X^*)] = \beta \quad (10)$$

$$\mathbb{C}(\beta) = \sup_{P_X} \mathbb{E}[j_{X;Y}(X; Y, \beta)] \quad (11)$$

$$= \sup_{P_X} \mathbb{E}[j_{X;Y^*}(X; Y, \beta)] \quad (12)$$

$$= \mathbb{E}[j_{X;Y^*}(X^*; Y^*, \beta)] \quad (13)$$

$$= \mathbb{E}[j_{X;Y^*}(X^*; Y^*, \beta) | X^*] \quad (14)$$

where (14) holds P_{X^*} -a.s.

Corollary 2.

$$\text{Var}[j_{X;Y^*}(X^*; Y^*, \beta)] = \mathbb{E}[\text{Var}[j_{X;Y^*}(X^*; Y^*, \beta) | X^*]] \quad (15)$$

$$= \mathbb{E}[\text{Var}[\imath_{X;Y^*}(X^*; Y^*) | X^*]] \quad (16)$$

III. NEW CONVERSE BOUND

Converse and achievability bounds give necessary and sufficient conditions, respectively, on (M, ϵ, β) in order for a code to exist with M codewords and average error probability not exceeding ϵ and β , respectively. Such codes (allowing stochastic encoders and decoders) are rigorously defined next.

Definition 2 ((M, ϵ, β) code). *An (M, ϵ, β) code for $\{P_{Y|X}, b\}$ is a pair of random transformations $P_{X|S}$ (encoder) and $P_{Z|Y}$ (decoder) such that $\mathbb{P}[S \neq Z] \leq \epsilon$, where the probability is evaluated with S equiprobable on an alphabet*

of cardinality M , $S - X - Y - Z$, and the codewords satisfy the maximal cost constraint (a.s.)

$$b(X) \leq \beta \quad (17)$$

The non-asymptotic quantity of principal interest is $M^*(\epsilon, \beta)$, the maximum code size achievable at error probability ϵ and cost β . Blocklength will enter into the picture later when we consider (M, d, ϵ) codes for $\{P_{Y^n|X^n}, b_n\}$, where $P_{Y^n|X^n}: \mathcal{A}^n \mapsto \mathcal{B}^n$ and $b_n: \mathcal{A}^n \mapsto [0, \infty]$. We will call such codes (n, M, d, ϵ) codes, and denote the corresponding non-asymptotically achievable maximum code size by $M^*(n, \epsilon, \beta)$. For now, though, blocklength n is immaterial, as the converse and achievability bounds do not call for any Cartesian set structure of the channel input and output alphabets. Accordingly, forgoing n , we state the converse for a generic pair $\{P_{Y|X}, b\}$, rather than the less general $\{P_{Y^n|X^n}, b_n\}$.

Theorem 3 (Converse). *The existence of an (M, ϵ, β) code for $\{P_{Y|X}, b\}$ requires that*

$$\epsilon \geq \inf_X \max_{\gamma > 0} \left\{ \sup_{\bar{Y}} \mathbb{P} [J_{X;\bar{Y}}(X; Y, \beta) \leq \log M - \gamma] - \exp(-\gamma) \right\} \quad (18)$$

$$\geq \max_{\gamma > 0} \left\{ \sup_{\bar{Y}} \inf_{x \in \mathcal{X}} \mathbb{P} [J_{X;\bar{Y}}(x; Y, \beta) \leq \log M - \gamma | X = x] - \exp(-\gamma) \right\} \quad (19)$$

Proof: Fix an (M, ϵ) code $\{P_{X|S}, P_{Z|Y}\}$, $\gamma > 0$, and an auxiliary probability distribution $P_{\bar{Y}}$ on \mathcal{Y} . Since $b(X) \leq \beta$, we have

$$\begin{aligned} & \mathbb{P} [J_{X;\bar{Y}}(X; Y, \beta) \leq \log M - \gamma] \\ & \leq \mathbb{P} [\iota_{X;\bar{Y}}(X; Y) - \lambda^*(b(X) - \beta) \leq \log M - \gamma] \end{aligned} \quad (20)$$

$$\leq \mathbb{P} [\iota_{X;\bar{Y}}(X; Y) \leq \log M - \gamma] \quad (21)$$

$$\begin{aligned} & = \mathbb{P} [\iota_{X;\bar{Y}}(X; Y) \leq \log M - \gamma, Z \neq S] \\ & + \mathbb{P} [\iota_{X;\bar{Y}}(X; Y) \leq \log M - \gamma, Z = S] \end{aligned} \quad (22)$$

$$\begin{aligned} & \leq \mathbb{P} [Z \neq S] \\ & + \frac{1}{M} \sum_{m=1}^M \sum_{x \in \mathcal{X}} P_{X|S}(x|m) \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) P_{Z|Y}(m|y) \\ & \cdot \mathbb{1} \{P_{Y|X}(y|x) \leq P_{\bar{Y}}(y) M \exp(-\gamma)\} \end{aligned} \quad (23)$$

$$\begin{aligned} & \leq \epsilon \\ & + \exp(-\gamma) \sum_{y \in \mathcal{Y}} P_{\bar{Y}}(y) \sum_{m=1}^M P_{Z|Y}(m|y) \sum_{x \in \mathcal{X}} P_{X|S}(x|m) \\ & \leq \epsilon + \exp(-\gamma) \end{aligned} \quad (24)$$

Optimizing over $\gamma > 0$ and the distribution of the auxiliary random variable \bar{Y} , we obtain the best possible bound for a given P_X , which is generated by the encoder $P_{X|S}$. Choosing

P_X that gives the weakest bound to remove the dependence on the code, (18) follows.

To show (19), we weaken (18) by moving \inf_X inside $\sup_{\bar{Y}}$, and write

$$\begin{aligned} & \inf_X \mathbb{P} [J_{X;\bar{Y}}(X; Y, \beta) \leq \log M - \gamma] \\ & = \inf_X \sum_{x \in \mathcal{X}} P_X(x) \mathbb{P} [J_{X;\bar{Y}}(x; Y, \beta) \leq \log M - \gamma | X = x] \end{aligned} \quad (25)$$

$$= \inf_{x \in \mathcal{X}} \mathbb{P} [J_{X;\bar{Y}}(x; Y, \beta) \leq \log M - \gamma | X = x] \quad (26)$$

Remark 1. At short blocklengths, it is possible to get a better bound by giving more freedom in (6) not restricting λ^* to be (4).

Achievability bounds for channels with cost constraints can be obtained from the random coding bounds in [15], [19] by restricting the distribution from which the codewords are drawn to satisfy $b(X) \leq \beta$ a.s. In particular, for the DMC, we may choose P_{X^n} to be equiprobable on the set of codewords of type which is closest to the input distribution P_{X^*} that achieves the capacity-cost function. As we will see in Section IV-C, owing to (16), such constant composition codes achieve the dispersion of channel coding under input cost constraints.

IV. ASYMPTOTIC ANALYSIS

In this section, we reintroduce the blocklength n into the non-asymptotic converse of Section III, i.e. let X and Y therein turn into X^n and Y^n , and perform its analysis, asymptotic in n .

A. Assumptions

The following basic assumptions hold throughout Section IV.

- (i) The channel is stationary and memoryless, $P_{Y^n|X^n} = P_{Y|X} \times \dots \times P_{Y|X}$.
- (ii) The cost function is separable, $b_n(x^n) = \frac{1}{n} \sum_{i=1}^n b(x_i)$, where $b: \mathcal{A} \mapsto [0, \infty]$.
- (iii) The codewords are constrained to satisfy the maximal power constraint (17).
- (iv) $\sup_{x \in \mathcal{A}} \text{Var} [J_{X;\bar{Y}}(x; Y, \beta) | X = x] = V_{\max} < \infty$.

Under these assumptions, the capacity-cost function $C(\beta) = \mathbb{C}(\beta)$ is given by (1). Observe that in view of assumption (i), as long as $P_{\bar{Y}^n}$ is a product distribution, $P_{\bar{Y}^n} = P_{\bar{Y}} \times \dots \times P_{\bar{Y}}$,

$$J_{X^n;\bar{Y}^n}(x^n; y^n, \beta) = \sum_{i=1}^n J_{X;\bar{Y}}(x_i; y_i, \beta) \quad (27)$$

B. Strong converse

We show that if transmission occurs at a rate greater than the capacity-cost function, the error probability must converge to 1, regardless of the specifics of the code. Toward this end, we fix some $\alpha > 0$, we choose $\log M \geq nC(\beta) + 2n\alpha$, and we weaken the bound (19) in Theorem 3 by fixing $\gamma = n\alpha$ and

$P_{Y^n} = P_{Y^*} \times \dots \times P_{Y^*}$, where Y^* is the output distribution that achieves $C(\beta)$, to obtain

$$\epsilon \geq \inf_{x^n \in \mathcal{A}^n} \mathbb{P} \left[\sum_{i=1}^n J_{X;Y^*}(x_i; Y_i, \beta) \leq nC(\beta) + n\alpha \right] - \exp(-n\alpha) \quad (28)$$

$$\geq \inf_{x^n \in \mathcal{A}^n} \mathbb{P} \left[\sum_{i=1}^n J_{X;Y^*}(x_i; Y_i, \beta) \leq \sum_{i=1}^n c(x_i) + n\alpha \right] - \exp(-n\alpha) \quad (29)$$

where for notational convenience we have abbreviated $c(x) = \mathbb{E} [J_{X;Y^*}(x; Y, \beta) | X = x]$, and (29) employs (12).

To show that the right side of (29) converges to 1, we invoke the following law of large numbers for non-identically distributed random variables.

Theorem 4 (e.g. [20]). *Suppose that W_i are uncorrelated and $\sum_{i=1}^{\infty} \text{Var} \left[\frac{W_i}{b_i} \right] < \infty$ for some strictly positive sequence (b_n) increasing to $+\infty$. Then,*

$$\frac{1}{b_n} \left(\sum_{i=1}^n W_i - \mathbb{E} \left[\sum_{i=1}^n W_i \right] \right) \rightarrow 0 \text{ in } L^2 \quad (30)$$

Let $W_i = J_{X;Y^*}(x_i; Y_i, \beta)$ and $b_i = i$. Since (recall (iv))

$$\sum_{i=1}^{\infty} \text{Var} \left[\frac{1}{i} J_{X;Y^*}(x_i; Y_i, \beta) | X_i = x_i \right] \leq V_{\max} \sum_{i=1}^{\infty} \frac{1}{i^2} \quad (31)$$

by virtue of Theorem 4 the right side of (29) converges to 1, so any channel satisfying (i)–(iv) also satisfies the strong converse.

As noted in [21, Theorem 77] in the context of the AWGN channel, the strong converse does not hold if the power constraint is averaged over the codebook, i.e. if, in lieu of (17), the cost requirement is

$$\frac{1}{M} \sum_{m=1}^M \mathbb{E} [b(X) | S = m] \leq \beta \quad (32)$$

To see why, fix a code of rate $C(\beta) < R < C(2\beta)$ none of whose codewords costs more than 2β and whose error probability vanishes as n increases, $\epsilon \rightarrow 0$. Since $R < C(2\beta)$, such a code exists. Now, replace half of the codewords with the all-zero codeword (assuming $b(0) = 0$) while leaving the decision regions of the remaining codewords untouched. The average cost of the new code satisfies (32), its rate is greater than the capacity-cost function, $R > C(\beta)$, yet its average error probability does not exceed $\epsilon + \frac{1}{2} \rightarrow \frac{1}{2}$.

C. Dispersion

First, we give the operational definition of the dispersion-cost function of any channel.

Definition 3 (Dispersion-cost function). *The channel dispersion-cost function, measured in squared information*

units per channel use, is defined by

$$V(\beta) = \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{(nC(\beta) - \log M^*(n, \epsilon, \beta))^2}{2 \log_e \frac{1}{\epsilon}} \quad (33)$$

An explicit expression for the dispersion-cost function of a memoryless channel is given in the next result (the proof is omitted due to space limitations).

Theorem 5. *In addition to assumptions (i)–(iv), assume that the capacity-achieving input distribution P_{X^*} is unique and that the channel has finite input and output alphabets.*

$$M^*(n, \epsilon, \beta) = nC(\beta) - \sqrt{nV(\beta)} Q^{-1}(\epsilon) + \theta(n) \quad (34)$$

$$C(\beta) = \mathbb{E} [J_{X;Y^*}(X^*; Y^*, \beta)] \quad (35)$$

$$V(\beta) = \text{Var} [J_{X;Y^*}(X^*; Y^*, \beta)] \quad (36)$$

where $P_{X^*Y^*} = P_{X^*}P_{Y|X}$, and the remainder term $\theta(n)$ satisfies:

a) If $V(\beta) > 0$,

$$-\frac{1}{2} (|\text{supp}(P_{X^*})| - 1) \log n + O(1) \leq \theta(n) \leq \frac{1}{2} \log n + O(1) \quad (37)$$

b) If $V(\beta) = 0$, (37) holds, and (38) is replaced by

$$\theta(n) \leq O\left(n^{\frac{1}{3}}\right) \quad (39)$$

Remark 2. According to a recent result of Moulin [17], the achievability bound on the remainder term in (38) can be tightened to match the converse bound in (38), thereby establishing that $\theta(n) = \frac{1}{2} \log n + O(1)$, provided that the following regularity assumptions hold:

- The random variable $\iota_{X;Y^*}(X^*; Y^*)$ is of nonlattice type;
- $\text{supp}(P_{X^*}) = \mathcal{A}$;
- $\text{Cov} [\iota_{X;Y^*}(X^*; Y^*), \iota_{X;Y}(\bar{X}^*; Y^*)] < \text{Var} [\iota_{X;Y^*}(X^*; Y^*)]$ where $P_{\bar{X}^*X^*Y^*}(\bar{x}, x, y) = \frac{1}{P_{Y^*}(y)} P_{X^*}(\bar{x}) P_{Y|X}(y|\bar{x}) P_{Y|X}(y|x) P_{X^*}(x)$.

V. JOINT SOURCE-CHANNEL CODING

In this section we state the counterparts of Theorems 3 and 5 in the lossy joint source-channel coding setting. Proofs of the results in this section are obtained by fusing the proofs in Sections III and IV and those in [19].

A joint source-channel coding problem arises if the source is no longer equiprobable on an alphabet of cardinality M , as in Definition 2, but is rather arbitrarily distributed on an abstract alphabet \mathcal{M} . Further, instead of reproducing the transmitted S under a probability of error criterion, we might be interested in approximating S within a certain distortion, so that a decoding failure occurs if the distortion between the source and its reproduction exceeds a given distortion level d , i.e. if $d(S, Z) > d$. A (d, ϵ, β) code is a code for a fixed source-channel pair such that the probability of exceeding distortion d is no larger than ϵ and no channel codeword costs more than β . A (d, ϵ, β) code in a block coding setting, when

a source block of length k is mapped to a channel block of length n , is called a $(k, n, d, \epsilon, \beta)$ code. The counterpart of the b-tilted information density in lossy compression is the d-tilted information, $J_S(s, d)$, which, in a certain sense, quantifies the number of bits required to reproduce the source outcome $s \in \mathcal{M}$ within distortion d . For rigorous definitions and further details we refer the reader to [19].

Theorem 6 (Converse). *The existence of a (d, ϵ, β) code for S and $P_{Y|X}$ requires that*

$$\epsilon \geq \inf_{P_{X|S}} \max_{\gamma > 0} \left\{ \sup_{\bar{Y}} \mathbb{P} [J_S(S, d) - J_{X;\bar{Y}}(X; Y, \beta) \geq \gamma] - \exp(-\gamma) \right\} \quad (40)$$

$$\geq \max_{\gamma > 0} \left\{ \sup_{\bar{Y}} \mathbb{E} \left[\inf_{x \in \mathcal{X}} \mathbb{P} [J_S(S, d) - J_{X;\bar{Y}}(x; Y, \beta) \geq \gamma \mid S] \right] - \exp(-\gamma) \right\} \quad (41)$$

where the probabilities in (40) and (41) are with respect to $P_S P_{X|S} P_{Y|X}$ and $P_{Y|X=x}$, respectively.

Under the usual memorylessness assumptions, applying Theorem 4 to the bound in (41), it is easy to show that the strong converse holds for lossy joint source-channel coding over channels with input cost constraints. A more refined analysis leads to the following result.

Theorem 7 (Gaussian approximation). *Assume the channel has finite input and output alphabets. Under restrictions (i)–(iv) of [19] and (ii)–(iv) of Section IV-A, the parameters of the optimal (k, n, d, ϵ) code satisfy*

$$nC(\beta) - kR(d) = \sqrt{nV(\beta) + kV(d)} Q^{-1}(\epsilon) + \theta(n) \quad (42)$$

where $V(d) = \text{Var}[J_S(S, d)]$, $V(\beta)$ is given in (36), and the remainder $\theta(n)$ satisfies, if $V(\beta) > 0$,

$$-\frac{1}{2} \log n + O\left(\sqrt{\log n}\right) \leq \theta(n) \leq \bar{\theta}(n) + \left(\frac{1}{2} |\text{supp}(P_{X^*})| - 1\right) \log n \quad (43)$$

where $\bar{\theta}(n)$ denotes the upper bound on $\theta(n)$ given in [19, Theorem 10]. If $V(\beta) = 0$, the upper bound on $\theta(n)$ stays the same, and the lower one becomes (39).

VI. CONCLUSION

We introduced the concept of b-tilted information density (Definition 1), a random variable whose distribution plays the key role in the analysis of optimal channel coding under input cost constraints. We showed a new converse bound (Theorem 3), which gives a lower bound on the average error probability in terms of the cdf of the b-tilted information density. The properties of b-tilted information density listed in Theorem 1 play a key role in the asymptotic analysis of the bound in Theorem 3 in Section IV, which does not only lead to the strong converse and the dispersion-cost function when coupled with the corresponding achievability bound, but it also proves

that the third order term in the asymptotic expansion (2) is upper bounded (in the most common case of $V(\beta) > 0$) by $\frac{1}{2} \log n + O(1)$. In addition, we showed in Section V that the results of [19] generalize to coding over channels with cost constraints and also tightened the estimate of the third order term in [19]. As propounded in [22], the gateway to refined analysis of the third order term is an apt choice of a non-product distribution $P_{\bar{Y}^n}$ in the bounds in Theorems 3 and 6.

REFERENCES

- [1] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois Journal of Mathematics*, vol. 1, no. 4, pp. 591–606, 1957.
- [2] —, "Strong converse of the coding theorem for semicontinuous channels," *Illinois Journal of Mathematics*, vol. 3, no. 4, pp. 477–489, 1959.
- [3] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 19, no. 3, pp. 357–359, 1973.
- [4] J. H. B. Kemperman, "Strong converses for a general memoryless channel with feedback," in *Proceedings 6th Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes*, 1971, pp. 375–409.
- [5] J. Wolfowitz, "The maximum achievable length of an error correcting code," *Illinois Journal of Mathematics*, vol. 2, no. 3, pp. 454–458, 1958.
- [6] A. Feinstein, "On the coding theorem and its converse for finite-memory channels," *Information and Control*, vol. 2, no. 1, pp. 25 – 44, 1959.
- [7] J. Wolfowitz, "A note on the strong converse of the coding theorem for the general discrete finite-memory channel," *Information and Control*, vol. 3, no. 1, pp. 89 – 93, 1960.
- [8] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.
- [9] M. Pinsker, *Information and information stability of random variables and processes*. San Francisco: Holden-Day, 1964.
- [10] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge Univ Press, 2011.
- [11] T. S. Han, *Information spectrum methods in information theory*. Springer, Berlin, 2003.
- [12] C. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, 1959.
- [13] K. Yoshihara, "Simple proofs for the strong converse theorems in some channels," *Kodai Mathematical Journal*, vol. 16, no. 4, pp. 213–222, 1964.
- [14] J. Wolfowitz, "Note on the Gaussian channel with feedback and a power constraint," *Information and Control*, vol. 12, no. 1, pp. 71 – 78, 1968.
- [15] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [16] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4947–4966, 2009.
- [17] P. Moulin, "The log-volume of optimal constant-composition codes for memoryless channels, within $O(1)$ bits," in *Proceedings 2012 IEEE International Symposium on Information Theory*, Cambridge, MA, July 2012, pp. 826–830.
- [18] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, June 2012.
- [19] —, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.
- [20] E. Çinlar, *Probability and Stochastics*. Springer, 2011.
- [21] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton University, 2010.
- [22] M. Tomamichel and V. Y. F. Tan, "A Tight Upper Bound for the Third-Order Asymptotics of Discrete Memoryless Channels," *ArXiv eprint arXiv:1212.3689*, Dec. 2012.
- [23] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *The Annals of Probability*, pp. 146–158, 1975.
- [24] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed. John Wiley & Sons, 1971, vol. II.