Saddlepoint Approximation of the Error Probability of Binary Hypothesis Testing

Gonzalo Vazquez-Vilar[12], Albert Guillén i Fàbregas[345], Tobias Koch[12], Alejandro Lancho[12]
[1]Universidad Carlos III de Madrid, [2]Gregorio Marañón Health Research Institute
[3]Universitat Pompeu Fabra, [4]ICREA, [5]University of Cambridge
Emails: {gvazquez,guillen}@ieee.org, {koch,alancho}@tsc.uc3m.es

***Abstract*—We propose a saddlepoint approximation of the error probability of a binary hypothesis test between two i.i.d. distributions. The approximation is accurate, simple to compute, and yields a unified analysis in different asymptotic regimes. The proposed formulation is used to efficiently compute the meta-converse lower bound for moderate block-lengths in several cases of interest.**

I. Introduction

Non-Bayesian binary hypothesis testing has found a large number of applications in information theory, image processing, signal processing, social sciences and biology. The goal is, upon observing a certain random sequence, to decide which of the two possible probability distributions generated the observation. The probability of making a mistake in this decision is minimized when employing the likelihood ratio test, as shown by Neyman and Pearson [1]. While in some applications having a coarse approximation of the performance of this test is sufficient, in other applications the computation of the exact minimum error probability is important. In this case, for most testing distributions of interest, computing the minimum probability of error can become a heavy computational task especially when the observation length grows large.

In the context of reliable communication, binary hypothesis testing has been instrumental in the derivation of converse bounds to the error probability. In [2, Sec. III], Shannon, Gallager and Berlekamp derived lower bounds to the error probability in the transmission of M messages by analyzing an instance of binary hypothesis testing. Blahut emphasized the fundamental role of binary hypothesis testing in information theory in [3]. More recently, Polyanskiy, Poor and Verdú applied the Neyman-Pearson lemma to a specific binary hypothesis test to derive the *meta-converse bound* [4, Th. 27], a fundamental finite-length lower bound to the channel-coding error probability from which several converse bounds can be recovered. While this bound is surprisingly accurate in several cases of interest, it is difficult to evaluate even for simple channels.

This work has been funded in part by the European Research Council (ERC) under grants 714161 and 725411, by the Spanish Ministry of Economy and Competitiveness under grants TEC2013-41718-R, RYC-201416332, IJCI-2015-27020 and TEC2016-78434-C3 (AEI/FEDER, EU), by the Madrid Autonomous Community under grant S2103/ICE-2845 and by the Spanish Ministry of Education, Culture and Sport under grant FPU14/01274.

In this paper, we derive a saddlepoint expansion of the minimum error probability trade-off of non-Bayesian hypothesis testing between two arbitrary i.i.d. distributions satisfying mild regularity conditions. This expansion yields an approximation which is accurate, simple to compute and allows a unified analysis in different asymptotic regimes. The proposed framework is then applied to analyze the meta-converse lower bound for symmetric channels. The saddlepoint approximation can be computed for auxiliary distributions other than the capacity-achieving output distribution. In particular, we show the advantage of computing it for the tilted output distribution attaining the sphere-packing and strong-converse exponents [2], [5].

This work is related to [6], [7], where the meta-converse lower bound is evaluated by solving (and approximating) Laplace integrals using a steepest descent technique. However, the method followed in [6], [7] is tailored to specific scenarios (the AWGN channel is treated in [6], while parallel channels, binary input AWGN channels and the binary symmetric channel are investigated in [7]), and seems difficult to extend to arbitrary auxiliary distributions. Saddlepoint methods have also been used to obtain approximations of random-coding bounds in [8]–[10].

II. Binary Hypothesis Testing

Let Y be a random variable on $\mathcal{Y}$ distributed according to either P or Q, where P and Q are probability measures. We consider a binary hypothesis test deciding the underlying distribution based on n i.i.d. observations of Y. We denote the observed vector by $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$ and the n-fold distributions by P^n and Q^n, respectively. Let $T(\boldsymbol{y}) \in [0,1]$ denote the probability of the test deciding P given an observation $\boldsymbol{y}$. Then, $1 - T(\boldsymbol{y})$ is the probability of deciding Q. The best error probability under P among all tests T with error probability under Q not exceeding β is

$$\alpha_\beta\big(P^n, Q^n\big) \triangleq \inf_{T:\mathrm{E}_{Q^n}[T(\boldsymbol{Y})]\leq\beta} \big\{1 - \mathrm{E}_{P^n}[T(\boldsymbol{Y})]\big\}, \quad (1)$$

where $\mathrm{E}_A[\cdot]$ denotes expectation with respect to A.

Neyman and Pearson (NP) gave an explicit form for the test achieving this trade-off [1]. Defining the log-likelihood ratio

$$\jmath(\boldsymbol{y}) \triangleq \frac{1}{n} \log \frac{\mathrm{d}P^n(\boldsymbol{y})}{\mathrm{d}Q^n(\boldsymbol{y})}, \quad (2)$$

the NP test minimizing (1) is given by

$$T_{\text{NP}}(\boldsymbol{y}) = \mathbb{1}\big[\jmath(\boldsymbol{y}) > \gamma\big] + p\mathbb{1}\big[\jmath(\boldsymbol{y}) = \gamma\big] \tag{3}$$

where $\mathbb{1}[\cdot]$ is the indicator function and γ and $p \in [0,1]$ are parameters chosen such that $\mathrm{E}_{Q^n}\big[T_{\text{NP}}(\boldsymbol{Y})\big] = \beta$. This test yields the following characterization for $\alpha_\beta\big(P^n, Q^n\big)$.

Lemma 1 (NP error trade-off): The optimal error probability tradeoff for testing between P^n and Q^n can be expressed as

$$\alpha_\beta\big(P^n, Q^n\big) = \max_{\gamma}\Big\{\mathbb{P}\big[\jmath(\boldsymbol{Y}) \le \gamma\big] + e^{n\gamma}\big(\mathbb{Q}\big[\jmath(\boldsymbol{Y}) > \gamma\big] - \beta\big)\Big\}, \tag{4}$$

where the probabilities $\mathbb{P}[\cdot]$ and $\mathbb{Q}[\cdot]$ are computed with respect to $\boldsymbol{Y} \sim P^n$ and $\boldsymbol{Y} \sim Q^n$, respectively.

Proof: To obtain (4), we optimize [11, eq. (95)] over γ'. Equality is attained for the threshold appearing in the NP test. ∎

III. Saddlepoint Approximation

We consider the following expansion of a tail probability, where $\mathcal{O}(f_n)$ summarizes terms of order no larger than f_n.

Lemma 2 (Saddlepoint expansion): Let $\{Z_\ell\}_{\ell=1}^n$ be a sequence of i.i.d. real-valued non-lattice random variables[1] with positive variance and define $\bar{Z}_n \triangleq \frac{1}{n}\sum_{\ell=0}^n Z_\ell$. Let $\kappa(s) = \log \mathrm{E}[e^{sZ_1}]$ be the cumulant generating function of Z_1 with region of convergence $\mathcal{S}_\kappa$, and denote by $\kappa'(s)$ and $\kappa''(s)$ its first and second derivatives. Assume that $0 \in \mathcal{S}_\kappa$, and that the mapping $t \to \exp(\kappa(it))$, where $i = \sqrt{-1}$, has a finite ξ-norm for some $\xi \ge 1$. If there exists $s \in \mathcal{S}_\kappa$ such that $\kappa'(s) = \gamma$ then, for such s,

$$\begin{aligned}\Pr\big[\bar{Z}_n \ge \gamma\big] &= \mathbb{1}[s<0] \\ &\quad + \operatorname{sgn}(s)\Big(\Psi(\lambda_n) + \mathcal{O}\big(\Psi(\lambda_n) n^{-\frac{1}{2}}\big)\Big) e^{n(\kappa(s) - s\kappa'(s))},\end{aligned} \tag{5}$$

where $\operatorname{sgn}(x)$ is equal to 1 for $x \ge 0$ and -1 for $x < 0$,

$$\Psi(\lambda_n) \triangleq \mathsf{Q}(|\lambda_n|) e^{\frac{\lambda_n^2}{2}}, \tag{6}$$

$$\lambda_n \triangleq |s|\sqrt{n\kappa''(s)}, \tag{7}$$

and $\mathsf{Q}(\cdot)$ denotes the Gaussian Q-function.

Proof: See [12, eq. (2.2.6)] and [12, Prop. 2.3.1]. ∎

The function $\Psi(\lambda_n)$ defined in (6) satisfies

$$\Psi(\lambda_n) = \begin{cases} \big(\sqrt{2\pi}\lambda_n\big)^{-1} + \mathcal{O}(\lambda_n^{-3}), & \lambda_n \to \infty, \\ \Psi(\hat{\lambda}) + \mathcal{O}(\lambda_n - \hat{\lambda}), & \lambda_n \to \hat{\lambda} < \infty. \end{cases} \tag{8}$$

Suposse that $s_n \to \hat{s}$ as $n \to \infty$ and let $\gamma = \gamma_n$ satisfy $\gamma_n = \kappa'(s_n)$. Under mild regularity conditions, the error term in (5) is uniformly bounded in a compact set around $s = \hat{s}$ [12, Prop. 2.3.1]. Then, according to (8), for $\hat{s} > 0$, $\Psi(\lambda_n) = \mathcal{O}\big(n^{-\frac{1}{2}}\big)$ and the error term in (5) becomes $\mathcal{O}(n^{-1})$. In contrast, if $s_n \to 0$ as $s_n = \mathcal{O}\big(n^{-\frac{1}{2}}\big)$, then $\Psi(\lambda_n) = \mathcal{O}(1)$ and the error term in (5) becomes $\mathcal{O}\big(n^{-\frac{1}{2}}\big)$.

[1] A random variable Z is said to be lattice if, and only if $\sum_{k=-\infty}^{\infty} \Pr[Z = a + k\delta] = 1$ for some $a, \delta \ge 0$. Otherwise it is said to be non-lattice.

A. Saddlepoint approximation of $\alpha_\beta\big(P^n, Q^n\big)$

For the binary hypothesis test considered in Section II, computing the tail probabilities in (4) requires solving two n-dimensional integrals, which is not feasible in general, even for moderate values of n. To compute the trade-off $\alpha_\beta(P^n, Q^n)$ in an efficient manner, we shall apply the expansion in Lemma 2 to $\mathbb{P}\big[\jmath(\boldsymbol{Y}) \le \gamma\big]$ and $\mathbb{Q}\big[\jmath(\boldsymbol{Y}) > \gamma\big]$.

We define

$$\kappa(s) \triangleq \log \int \frac{\mathrm{d}P(y)^s}{\mathrm{d}Q(y)^s}\,\mathrm{d}Q(y) = (s-1)D_s(P\|Q), \tag{9}$$

where $D_s(P\|Q)$ denotes the Rényi divergence of order s. This function is well defined provided that P is absolutely continuous with respect to Q, i.e., $P \ll Q$. In the rest of the paper, we shall also assume that

(A1) P and Q are distinct and absolutely continuous with respect to each other, i.e., $P \ll Q$, and $Q \ll P$,

(A2) the random variables $Z_P \triangleq -\jmath(Y_P)$, $Y_P \sim P$, and $Z_Q \triangleq \jmath(Y_Q)$, $Y_Q \sim Q$, satisfy the conditions in Lemma 2 and [12, Prop. 2.3.1].

Theorem 1: The NP trade-off $\alpha_\beta\big(P^n, Q^n\big)$ as a function of the auxiliary parameter $s \in \mathcal{S}_\kappa$ is given by

$$\alpha(s) = \mathbb{1}[s>1] + a_n(s) e^{n(\kappa(s) + (1-s)\kappa'(s))}, \tag{10}$$

$$\beta(s) = \mathbb{1}[s<0] + b_n(s) e^{n(\kappa(s) - s\kappa'(s))}, \tag{11}$$

where the sub-exponential factors $a_n(s)$ and $b_n(s)$ satisfy

$$a_n(s) = \operatorname{sgn}(1-s)\Psi\big(|1-s|\sqrt{n\kappa''(s)}\big)\big(1 + \mathcal{O}\big(n^{-\frac{1}{2}}\big)\big), \tag{12}$$

$$b_n(s) = \operatorname{sgn}(s)\Psi\big(|s|\sqrt{n\kappa''(s)}\big)\big(1 + \mathcal{O}\big(n^{-\frac{1}{2}}\big)\big), \tag{13}$$

where $\operatorname{sgn}(x)$ is equal to 1 for $x \ge 0$ and -1 for $x < 0$.

Proof: According to the NP test (3), the trade-off between α and β as a function of an auxiliary parameter s satisfies

$$\mathbb{P}\big[\jmath(\boldsymbol{Y}) < \kappa'(s)\big] \le \alpha(s) \le \mathbb{P}\big[\jmath(\boldsymbol{Y}) \le \kappa'(s)\big], \tag{14}$$

$$\mathbb{Q}\big[\jmath(\boldsymbol{Y}) > \kappa'(s)\big] \le \beta(s) \le \mathbb{Q}\big[\jmath(\boldsymbol{Y}) \ge \kappa'(s)\big]. \tag{15}$$

Then, the result follows by applying Lemma 2 to the random variables Z_P and Z_Q, defined in (A2) above, which have cumulant generating functions

$$\kappa_P(s) = \log \mathrm{E}_P\Big[\Big(\tfrac{\mathrm{d}P(y)}{\mathrm{d}Q(y)}\Big)^{-s}\Big] = \kappa(1-s), \tag{16}$$

$$\kappa_Q(s) = \log \mathrm{E}_Q\Big[\Big(\tfrac{\mathrm{d}P(y)}{\mathrm{d}Q(y)}\Big)^{s}\Big] = \kappa(s). \tag{17}$$

For random variables satisfying (A2) the corresponding expansions of the upper and lower bounds in (14) and (15) coincide and yield (10) and (11), respectively. ∎

Combining Lemmas 1 and 2 we obtain an alternative formulation of the optimal trade-off of binary hypothesis testing.

Theorem 2: The NP trade-off $\alpha_\beta\big(P^n, Q^n\big)$ satisfies

$$\begin{aligned}\alpha_\beta\big(P, Q\big) = \max_s \Big\{ &\big(a_n(s) + b_n(s)\big) e^{n(\kappa(s) + (1-s)\kappa'(s))} \\ &+ \mathbb{1}[s>1] + \big(\mathbb{1}[s<0] - \beta\big) e^{n\kappa'(s)}\Big\},\end{aligned} \tag{18}$$

with $a_n(s)$ and $b_n(s)$ defined in (12) and (13), respectively.

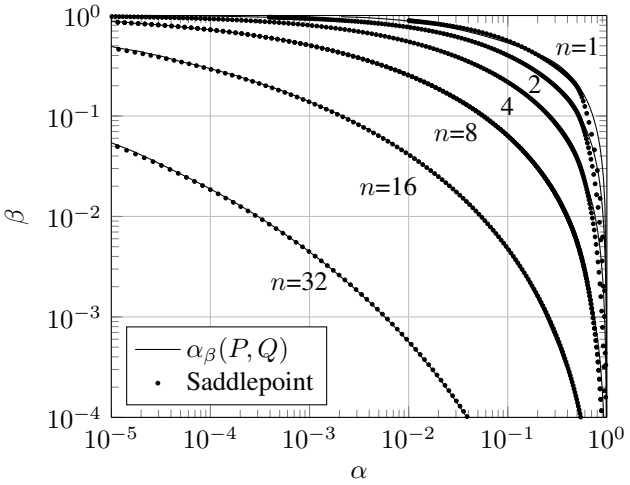

Fig. 1. NP error trade-off of a binary hypothesis test between a Laplacian distribution versus a mixture of two normal distributions.

Proof: The proof follows by applying in Lemma 1 the saddlepoint expansion (5) to the probabilities $\mathbb{P}\big[\jmath(\boldsymbol{Y}) \le \gamma\big]$ and $\mathbb{Q}\big[\jmath(\boldsymbol{Y}) > \gamma\big]$, and by using the mapping $\kappa'(s) = \gamma$. The proof is then analogous to that of Theorem 1. ∎

For a sufficiently large n, Theorems 1 and 2 allow us to approximate the trade-off $\alpha_\beta\big(P, Q\big)$ by disregarding the $\mathcal{O}\big(n^{-\frac{1}{2}}\big)$ terms in (12) and (13). We shall refer to the resulting expression as *saddlepoint approximation*. In this approximation, the probabilities $\alpha(s), \beta(s) \le \frac{1}{2}$ correspond to $s \in [0, 1]$.

While obtaining a closed-form expression for $\kappa(s)$ is difficult in general, evaluating the cumulant generating function $\kappa(s)$ and its first and second derivatives only requires to solve three one-dimensional integrals, as shown in the next result.

Proposition 1: Let $\kappa(s)$ be defined in (9) and let

$$J_\ell(s) \triangleq \mathrm{E}_Q\big[e^{s\jmath(Y)}\jmath(Y)^\ell\big] = \mathrm{E}_P\big[e^{(s-1)\jmath(Y)}\jmath(Y)^\ell\big]. \quad (19)$$

Then $\kappa(s) = \log J_0(s)$ and

$$\kappa'(s) = \frac{J_1(s)}{J_0(s)}, \quad \kappa''(s) = \frac{J_2(s)}{J_0(s)} - \frac{J_1(s)^2}{J_0(s)^2}. \quad (20)$$

Proof: The derivative of $J_\ell(s)$ is $J'_\ell(s) = J_{\ell+1}(s)$. Noting that $\kappa(s) = \log J_0(s)$, then (20) follows by computing its two first derivatives. ∎

We illustrate the accuracy of the saddlepoint approximation with an example. We let P be a Laplacian distribution with unit variance and unit mean, and Q be the equiprobable mixture of two normal distributions with unit variance and mean equal to -1 and $+1$, respectively. Fig. 1 depicts the trade-off α versus β for different values of n. We show the exact trade-off $\alpha_\beta(P^n, Q^n)$ computed using Monte Carlo methods and the saddlepoint approximation that follows from Theorem 1 by ignoring the $\mathcal{O}\big(n^{-\frac{1}{2}}\big)$ terms in (12)-(13). The functions $\kappa(s)$, $\kappa'(s)$ and $\kappa''(s)$ have been obtained for a grid of values of $s \in [-1, 2]$ via numerical integration as described in Proposition 1. We can see that the precision of the approximation is remarkable for a number of observations $n \ge 8$ in the whole range of values. In the range $\alpha, \beta \le \frac{1}{2}$ the approximation is accurate even for $n \ge 1$.

B. Asymptotic analysis

We next study the asymptotic behavior of the error probability of binary hypothesis testing as $n \to \infty$. To this end, we explicitly write the dependence of α and β with n, as α_n and β_n, respectively. We shall also assume that (A1)-(A2) hold.

When both α_n and β_n decay exponentially in n, the hypothesis test is in the so-called *large deviations regime*. Let $\beta_n = \hat{b}_n e^{-nB}$ where $B \ge 0$ and $\hat{b}_n$ satisfies $\lim_{n\to\infty} \frac{1}{n} \log \hat{b}_n = 0$. Theorem 1 provides a direct approach to study the existing trade-off between the error exponents $A \triangleq \lim_{n\to\infty} -\frac{1}{n} \log \alpha_n$, where $\alpha_n = \alpha_{\beta_n}(P^n, Q^n)$, and B. Using (10) and (11), we obtain the following parametric representation of A and B as a function of $s \in [0, 1]$:

$$A(s) = -\kappa(s) - (1 - s)\kappa'(s), \quad (21)$$

$$B(s) = -\kappa(s) + s\kappa'(s). \quad (22)$$

This representation appears, e.g., in [2, Th. 5] and [3, Th. 4]. A non-parametric relation between A and B is given by the next result, corresponding to [3, Th. 7]:

Corollary 1 (Error exponent trade-off): The error exponent A as a function of B is given by

$$A(B) = \max_{0\le s\le 1} \Big\{ -\tfrac{1}{s}\kappa(s) - \tfrac{1-s}{s} B \Big\} \quad (23)$$

$$= \max_{0\le s\le 1} \Big\{ \tfrac{1-s}{s} \big(D_s(P\|Q) - B \big) \Big\}. \quad (24)$$

Proof: Solving for $\kappa'(s)$ in (22), and substituting the resulting expression in (21) yields

$$A(s) = -\tfrac{1}{s}\kappa(s) - \tfrac{1-s}{s} B(s). \quad (25)$$

This equation expresses $A(s)$ as a function of $B(s)$. However, the dependence between A and B is not explicit since the value of s appearing in (25) needs to be obtained from (22) by setting $B(s) = B$ and solving for s. We next show that this value of s is precisely the one optimizing (23). Since the term in braces in (23) coincides with (25), we then conclude that the optimization in (23) yields $A(B) = A(s)$. The identity (24) follows from (9) and (23).

To show that the value of s optimizing (23) satisfies (22), we take the derivative of the bracketed term in (23) with respect to s. Equating the result to 0 yields (22) when $B(s) = B$. ∎

When $\beta_n = \beta > 0$, the corresponding error exponent is $B = 0$ and the error probability of the hypothesis test is governed by the *small deviations (or Stein's) regime*. In this setting, the error probability α_n becomes [13, Th. 1.1]

$$-\frac{1}{n} \log \alpha_n = -\kappa'(0) - \sqrt{\tfrac{1}{n}\kappa''(0)}\,\mathsf{Q}^{-1}(\beta) + \mathcal{O}\big(\tfrac{\log n}{n}\big), \quad (26)$$

which is usually referred to as normal approximation.

This approximation can be recoved from a refined version of Theorem 2, where the error terms are specified in more detail. Indeed, to recover (26) one requires a sequence $s = s_n$ that satisfies $s_n \to 0$ at a rate $n^{-\frac{1}{2}}$. In this case, the error term in (13) becomes $\mathcal{O}\big(\Psi(\lambda_n) n^{-\frac{1}{2}}\big) = \mathcal{O}\big(n^{-\frac{1}{2}}\big)$. As the sub-exponential factor in (18) behaves as $n^{-\frac{1}{2}}$ in this regime, it is masked by this error term.

IV. Application to Channel Coding

We consider the problem of transmitting M equiprobable messages over a length-n memoryless channel $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_i^n P_{Y|X}(y_i|x_i)$. The encoder maps a message $v \in \{1,\dots,M\}$ to a codeword $\boldsymbol{x}(v)$ using a codebook $\mathcal{C}$. Based on the channel output $\boldsymbol{y}$, the decoder guesses the transmitted message $\hat{v}$ with error probability $P_{\mathrm{e}}(\mathcal{C}) \triangleq \Pr\{\hat{V} \neq V\}$.

The meta-converse bound [4, Th. 27] lower-bounds $P_{\mathrm{e}}(\mathcal{C})$ by the error probability of a binary hypothesis test. Under certain symmetry conditions, it yields [4, Th. 28]

$$P_{\mathrm{e}}(\mathcal{C}) \geq \alpha_{\frac{1}{M}}\big(P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}, Q_{\boldsymbol{Y}}\big). \tag{27}$$

In particular, (27) holds when $\Pr\big[\frac{\mathrm{d}P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{x})}{\mathrm{d}Q_{\boldsymbol{Y}}(\boldsymbol{Y})} \geq \tau\big]$ is independent of $\boldsymbol{x} \in \mathcal{X}^n$ for every τ,- where the probability is computed with respect to $P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}$. This condition is satisfied, e.g., for symmetric memoryless channels when $Q_{\boldsymbol{Y}} = Q_Y^n$ with Q_Y the *capacity-achieving* or the *exponent-achieving output distribution*, defined next.

For simplicity, we assume that $P_{Y|X}$ is absolutely continuous with respect to the Lebesgue measure and denote its pdf by $p_{Y|X}$. We also assume that the channel input has bounded support and denote by $\bar{P}_X$ the uniform distribution. We define

$$q_\rho(y) \triangleq \frac{1}{\mu(\rho)}\left(\int p_{Y|X}(y|x)^{\frac{1}{1+\rho}}\,\mathrm{d}\bar{P}_X(x)\right)^{1+\rho}, \tag{28}$$

where $\mu(\rho) \triangleq \int\left(\int p_{Y|X}(y|x)^{\frac{1}{1+\rho}}\,\mathrm{d}\bar{P}_X(x)\right)^{1+\rho}\mathrm{d}y$. Then, the *Gallager E_0-function* [14, eq. (5.6.14)] is $E_0(\rho) \triangleq -\log\mu(\rho)$.

We consider the hypothesis test in (27) and $Q_{\boldsymbol{Y}}$ with pdf $q_{\boldsymbol{Y}}(\boldsymbol{y}) = \prod_i^n q_\rho(y_i)$. For $\rho = 0$ this distribution is the capacity-achieving output distribution. When ρ is appropriately chosen this distribution optimizes the exponential behavior of the bound, as shown next.

Theorem 3 (Meta-converse, auxiliary distribution q_ρ): Let $\log\frac{p_{Y|X}(Y|x)}{q_\rho(Y)}$ be non-lattice and let $p_{Y|X=x}$ and q_ρ share the same support, $\rho > -1$. Then, any code $\mathcal{C}$ with $M = e^{nR} \geq 2$ codewords satisfies

$$P_{\mathrm{e}}(\mathcal{C}) \geq \max_{\rho>-1}\Big\{\mathbb{1}[\rho<0] + \Big(\eta_n(\rho) - e^{-n(R-E_0'(\rho))}\Big)e^{-n(E_0(\rho)-\rho E_0'(\rho))}\Big\}, \tag{29}$$

where, for $\Psi(\cdot)$ defined in (6),

$$\eta_n(\rho) \triangleq \Psi\big(\sqrt{nV(\rho)}\big)\big(1+\mathcal{O}\big(n^{-\frac{1}{2}}\big)\big) + \mathrm{sgn}(\rho)\Psi\big(\sqrt{n\rho^2V(\rho)}\big)\big(1+\mathcal{O}\big(n^{-\frac{1}{2}}\big)\big), \tag{30}$$

and $V(\rho) \triangleq \frac{1}{(1+\rho)^2}\kappa''\big(\frac{1}{1+\rho}\big)$ for $\kappa(s) = \log\int\frac{p_{Y|X}(y|x)^s}{q_\rho(y)^{s-1}}\,\mathrm{d}y$.

Proof: Let us consider the bound that follows from (27) when $q_{\boldsymbol{Y}}(\boldsymbol{y}) = \prod_i^n q_\rho(y_i)$, $\rho > -1$. Applying Theorem 2 with $s > 0$, we obtain that, for any $\rho > -1$,

$$P_{\mathrm{e}}(\mathcal{C}) \geq \max_{s>0}\Big\{\big(a_n(s)+b_n(s)\big)e^{n(\kappa(s)+(1-s)\kappa'(s))} + \mathbb{1}[s>1] - e^{n(\kappa'(s)-R)}\Big\}. \tag{31}$$

In (31), we can let ρ be a function of s provided that the derivatives $\kappa'(s)$ and $\kappa''(s)$ are computed assuming ρ to be independent of s. We fix $\rho = \frac{1-s}{s}$.[2] Using that $\kappa(s)$ does not depend on x due to the channel symmetry, a tedious but straightforward calculation yields

$$\kappa(s) = -sE_0\big(\tfrac{1-s}{s}\big), \tag{32}$$

$$\kappa'(s) = \tfrac{1}{s}E_0'\big(\tfrac{1-s}{s}\big) - E_0\big(\tfrac{1-s}{s}\big), \tag{33}$$

where $E_0'(\rho) = \frac{\partial E_0(\rho)}{\partial\rho}$.

Although (33) coincides with the derivative of the right-hand side of (32), this is not to be expected in general. Indeed, (33) follows from Proposition 1 by assuming q_ρ independent of s and by then fixing $\rho = \frac{1-s}{s}$ in the resulting expression. In contrast, on the right-hand side of (32), the mapping $\rho = \frac{1-s}{s}$ is implicit. For the second derivative of $\kappa(s)$, there exists no direct correspondence between $\kappa''(s)$ and $E_0''(\rho)$.

Using the definitions of a_n and b_n in (12) and (13), substituting (32) and (33) in (31), and changing the optimization variable from $s = \frac{1}{1+\rho} \in (0,\infty)$ to $\rho = \frac{1-s}{s} \in (-1,\infty)$, we obtain the desired result. ∎

Theorem 3 recovers the the *sphere-packing exponent* [2]

$$E_{\mathrm{sp}}(R) \triangleq \max_{\rho\geq 0}\big\{E_0(\rho) - \rho R\big\} \tag{34}$$

and the *strong-converse exponent* [5]

$$E_{\mathrm{sc}}(R) \triangleq \max_{-1<\rho<0}\big\{E_0(\rho) - \rho R\big\}. \tag{35}$$

Indeed, let $\rho_\star$ satisfy $E_0'(\rho_\star) = R$, and note that $E_0'(0) = I(X;Y)$ with $X \sim \bar{P}_X$. Then, $\rho_\star > 0$ corresponds to $R < I(X;Y)$ and $\rho_\star < 0$ corresponds to $R > I(X;Y)$. Setting in (29) $\rho = \hat{\rho}$ such that $E_0'(\hat{\rho}) = R - \delta$, $\delta > 0$, yields

$$P_{\mathrm{e}}(\mathcal{C}) \geq \mathbb{1}[\hat{\rho}<0] + \big(\eta_n(\hat{\rho}) - e^{-n\delta}\big)e^{-n(E_0(\hat{\rho})-\hat{\rho}(R-\delta))}. \tag{36}$$

It follows that $\lim_{n\to\infty}\frac{1}{n}\log\big(\eta_n(\hat{\rho}) - e^{-n\delta}\big) = 0$ for any $\delta > 0$. Thus, by first letting $n \to \infty$ and then $\delta \to 0$, (36) recovers the sphere-packing exponent $E_{\mathrm{sp}}(R)$ when $\rho_\star > 0$ and the strong-converse exponent $E_{\mathrm{sc}}(R)$ when $\rho_\star < 0$.

Furthermore, by letting $\rho = \hat{\rho}_n$ tend to $\rho_\star$ with n as

$$\hat{\rho}_n = \rho_\star - \frac{\log\big(\sqrt{2\pi nV(\rho_\star)}\big)}{nE_0''(\rho_\star)}, \tag{37}$$

Theorem 3 recovers not only the sphere-packing exponent when $\rho_\star > 0$, but also the sub-exponential behavior of the error probability $P_{\mathrm{e}}(\mathcal{C})$ in a certain regime.

Corollary 2: Let $\rho_\star$ satisfy $E_0'(\rho_\star) = R$ and $\rho_\star > 0$. Then,

$$P_{\mathrm{e}}(\mathcal{C}) \geq \Big(\tfrac{1}{\rho_\star}\big(2\pi nV(\rho_\star)\big)^{-\frac{1+\rho_\star}{2}} + \mathcal{O}\big(n^{-(1+\frac{\rho_\star}{2})}\big)\Big)e^{-nE_{\mathrm{sp}}(R)}. \tag{38}$$

Proof: The proof is omitted due to space constraints. ∎

The lower bound (38) has a sub-exponential factor of the order $n^{-\frac{1+\rho_\star}{2}}$. For rates between the critical rate of the channel and capacity, this order coincides with that of the random-coding upper bound to the error probability in [9, Th. 2].

[2] This choice correspond to $s = \frac{1}{1+\rho}$, also used in the derivation of Gallager random-coding bound [14, Ch. 5]. See also [14, p. 529, Prob. 5.6].

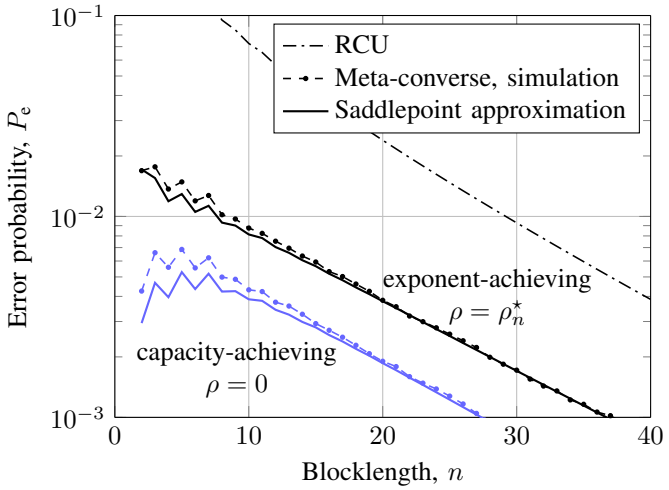

Fig. 2. Error probability bounds for a BIAWGN channel with parameters $R = 0.5$, SNR$=$ 3.5 dB (i.e., capacity $C \approx 0.76$ bits/channel use).

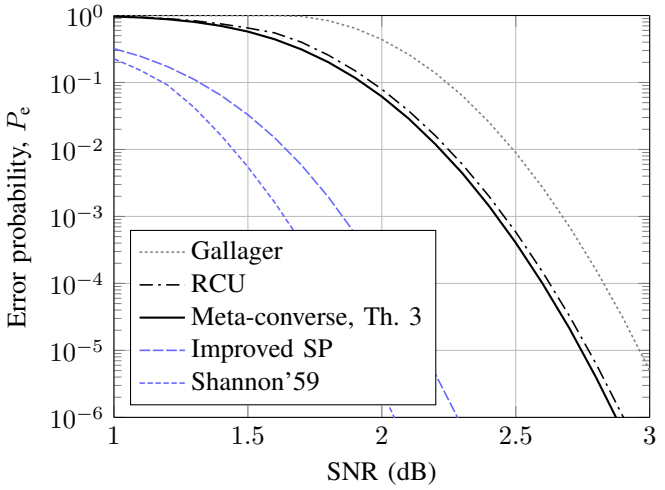

Fig. 3. Error probability bounds for a BIAWGN channel with parameters $R = 0.75$, $n = 1024$.

A. *Example: The binary-input AWGN (BIAWGN) channel*

We consider the converse bounds that follow from (27) with $q_{\boldsymbol{Y}}(\boldsymbol{y}) = \prod_i^n q_\rho(y_i)$ when $\rho = 0$ (capacity-achieving) and when $\rho = \rho_n^\star$ is the value of ρ maximizing (29) in Theorem 3 (exponent-achieving). In order to show the accuracy of the proposed saddlepoint approximations, even for extremely low values of n, Fig. 2 compares the converse bounds evaluated via Monte Carlo simulation, and the saddlepoint approximations obtained by disregarding the $\mathcal{O}\big(n^{-\frac{1}{2}}\big)$ terms in Theorem 2 (for $\rho = 0$) and in Theorem 3 (for $\rho = \rho_n^\star$). For reference, we also plot the RCU upper bound [4, Th. 16] evaluated according to [10, eq. (61)]. We can see how the saddlepoint approximations become very accurate for $n \geq 20$ and still provide a good approximation of the bounds for $n < 20$. Additionally, Fig. 2 also shows that by using the exponent-achieving $\rho = \rho_n^\star$ we obtain a tighter meta-converse than by using $\rho = 0$. In fact, as discussed after Theorem 3, the meta-converse bound with $\rho = \rho_n^\star$ attains the sphere-packing error exponent, while this is not true in general when $\rho = 0$. Therefore, the gap between the two bounds is not only significant, but it may even grow exponentially with n.

Figure 3 compares different bounds in a scenario with $R = 0.75$ bits/channel use and $n = 1024$. The lower bounds in the figure are Shannon's sphere-packing (SP) bound for the AWGN channel [15], the improved SP bound for symmetric channels [16, Th. 3.1], and Theorem 3. As upper bounds we plot the RCU bound [4, Th. 16] and Gallager's random coding bound [14, Th. 5.6.2]. From Fig. 3 we conclude that, as perhaps expected, the meta-converse lower bound is much tighter than both Shannon's and the improved SP bounds. The error probability of the best code is precisely characterized between the RCU and the meta-converse bound from Th. 3. For $n = 1024$ and the exponent-achieving auxiliary distribution, evaluation of (27) using previous results in the literature was computationally unfeasible. In contrast, evaluating the saddlepoint approximation has a computational complexity similar (if not smaller) to that of the other bounds.

REFERENCES

[1] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. R. Soc. Lond. A*, vol. 231, no. 694-706, p. 289, 1933.

[2] C. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65 – 103, 1967.

[3] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 405–417, 1974.

[4] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[5] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels (corresp.)," *IEEE Trans. Inf. Theory*, vol. 19, no. 3, pp. 357–359, 1973.

[6] T. Erseghe, "On the evaluation of the Polyanskiy-Poor-Verdú converse bound for finite block-length coding in AWGN," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6578–6590, Dec 2015.

[7] ——, "Coding in the finite-blocklength regime: Bounds based on Laplace integrals and their asymptotic approximations," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6854–6883, Dec 2016.

[8] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2647–2666, May 2014.

[9] J. Honda, "Exact asymptotics for the random coding error probability," in *2015 IEEE Int. Symp. on Inf. Theory (ISIT)*, June 2015, pp. 91–95.

[10] J. Font-Segura, G. Vazquez-Vilar, A. Martinez, A. Guillén i Fàbregas, and A. Lancho, "Saddlepoint approximations of lower and upper bounds to the error probability in channel coding," in *52nd Conf. on Inf. Syst. and Sci. (CISS)*, Mar. 2018.

[11] G. Vazquez-Vilar, A. T. Campo, A. Guillén i Fàbregas, and A. Martinez, "Bayesian M-ary hypothesis testing: The meta-converse and Verdú-Han bounds are tight," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2324–2333, May 2016.

[12] J. L. Jensen, *Saddlepoint Approximations*. Oxford: Oxford University Press, 1995.

[13] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," in *Trans. 3rd Prague Conf. Inf. Theory*, Prague, 1962, pp. 689–723.

[14] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, Inc., 1968.

[15] C. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, pp. 611–656, 1959.

[16] G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes over symmetric memoryless channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1962–1990, May 2008.