Nearest Neighbor Decoding for a Class of Compound Channels

Francesc Molina
University of Cambridge
Universitat Pompeu Fabra
Universitat Politècnica de Catalunya
fm585@cam.ac.uk

Priyanka Patel
University of Cambridge
pp490@cantab.ac.uk

Albert Guillén i Fàbregas
University of Cambridge
Universitat Pompeu Fabra
guillen@ieee.org

***Abstract*—We study Gaussian i.i.d. codebooks and nearest neighbor decoding over continuous-alphabet channels. We define a class of compound channels that are within a small radius relative entropy ball centered at the nearest neighbor decoding metric. We derive approximations to the worst-case achievable rates and find the penalty terms proportional to the square root of the ball radius.**

I. Introduction

Most of the existing results in information theory rely on the optimistic assumption that the probability law governing channel is perfectly known to both the transmitter and receiver [1]. However, in practice, there are several environments where the channel is either unknown or difficult to estimate. It is therefore crucial to elaborate on a more general theory that accounts for these specific uncertainties [2], [3].

Mismatched decoding precisely addresses this problem: the decoder operates with a given fixed decoding metric and only the codebook can be optimized. Its close connection to practical real-world systems limited by implementation constraints and its intriguing connections to zero-error communication make it an important problem [2, Ch. 1.2].

This work focuses on Gaussian i.i.d. codebooks combined with the nearest neighbor decoder [4], [5]. The nearest neighbor decoder is optimal under additive Gaussian noise and widely adopted in non-Gaussian noise settings. Gaussian i.i.d. codebooks with nearest neighbor decoding constitute a robust transmitter-receiver pair for communicating over additive noise channels. This paper derives achievable information rates for a class of continuous-alphabet channels close to the Gaussian in terms of relative entropy.

II. System Set-up

Consider the reliable transmission of M messages over a point-to-point memoryless channel defined over the real input and output alphabets $\mathcal{X}^n = \mathbb{R}^n$ and $\mathcal{Y}^n = \mathbb{R}^n$, respectively. n is the number of channel uses used to transmit a message. For transmission, the encoder selects message $m \in \{1, \dots, M\}$ equiprobably and transmits the corresponding codeword $\boldsymbol{x}^{(m)} = (x_1^{(m)}, \dots, x_n^{(m)})$ from the codebook $\mathcal{C}_n = \{\boldsymbol{x}^{(i)}\}_{1 \le i \le M}$, in which the symbols of every codeword are distributed according to a zero-mean, variance P normal distribution $Q_X(x) = \mathcal{N}(x; P)$, where $\mathcal{N}(x;P) = \frac{1}{\sqrt{2\pi P}} \exp(-x^2/2P)$. Therefore, the codeword distribution is

$$Q_X^n(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi P}} \exp\left(-\frac{x_i^2}{2P}\right). \tag{1}$$

P should be understood as the power averaged over the codebook, in which codewords attain the per codeword power constraint only for $n \to \infty$.

The (continuous) channel probability distribution is defined from the single-letter probability density $W(y|x)$ for all pairs $(x, y) \in \mathbb{R}^2$. Conditional on $\boldsymbol{x}^{(m)} \in \mathbb{R}^n$ being transmitted, $\boldsymbol{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$ is received with probability

$$W^n(\boldsymbol{y}|\boldsymbol{x}^{(m)}) = \prod_{i=1}^{n} W(y_i|x_i^{(m)}). \tag{2}$$

The only knowledge available about W^n is that it is memoryless and that its single-letter distribution W is within a small radius relative entropy ball centered at the Gaussian distribution $\widehat{W}(y|x) = \mathcal{N}(y-x; \sigma^2)$ corresponding to the nearest neighbor decoder, as

$$W \in \mathcal{B} = \{W : \; D(\widehat{W} \| W | Q_X) \le r\}. \tag{3}$$

The decoder performs nearest neighbor decoding, i.e., it decides in favor of the codeword that is closest in Euclidean distance to the observation $\boldsymbol{y} \in \mathbb{R}^n$ as

$$\hat{m} = \underset{1 \le \bar{m} \le M}{\operatorname{argmax}} \prod_{i=1}^{n} \widehat{W}(y_i|x_i^{(\bar{m})}). \tag{4}$$

The receiver can operate without knowing σ^2 as it does not alter the ranking of codewords, but its value must be known by the code designer, since the relative entropy ball (3) depends on it. The average error probability is considered. An error occurs whenever $\hat{m} \neq m$; ties are counted as errors.

We want to show reliable information rates for this setting, which can be interpreted as a mismatched compound channel

This work was supported by the European Research Council under Grant 725411 and by the Spanish Ministry of Economy and Competitiveness under Grant PID2020-116683GB-C22. Francesc Molina is also supported by the Spanish Ministry of Universities through Margarita Salas Fellowship.

[2, Ch. 2.4.4]. In particular, since the actual channel is unknown, universal (worst-case) rates are derived based on the mismatched decoding framework.

III. General Channels and Decoding Metrics

We will keep our analysis general for any i.i.d. codebook and memoryless decoding metric defined respectively by $Q_X(x)$ and $\widehat{W}(y|x)$, over the same alphabets. Gaussian i.i.d. codebooks with nearest neighbor decoding are treated in the next section.

It is well known (see e.g. [2] and references therein) that an achievable rate for continuous-alphabet channels with mismatched decoding is given by the generalized mutual information (GMI)

$$I_{\mathrm{GMI}}(Q_X, \widehat{W}) = \sup_{s\geq 0} \ \mathbb{E}_{Q_X\times W}\left[i_s(X,Y)\right] \tag{5}$$

under the mismatched information density

$$i_s(x,y) = \log \frac{\widehat{W}(y|x)^s}{\mathbb{E}_{Q_X}[\widehat{W}(y|X)^s]}. \tag{6}$$

Consequently, in order to guarantee reliable communication over every channel $W \in \mathcal{B}$, we have the following rate

$$\underline{I}_{\mathrm{GMI}}(Q_X, \widehat{W}) = \min_{W\in\mathcal{B}} \ I_{\mathrm{GMI}}(Q_X, \widehat{W}) \tag{7}$$

$$= \min_{W\in\mathcal{B}} \ \sup_{s\geq 0} \ \mathbb{E}_{Q_X\times W}\left[i_s(X,Y)\right] \tag{8}$$

henceforth denoted worst-case GMI. The minimization is over all conditional probability distributions in the small relative entropy ball $\mathcal{B}$ defined in (3).

Two considerations are stated before presenting our main result. Firstly, we will only consider continuous channels $W(y|x)$, and we assume the continuity of GMI rate and constraints as functionals of W as per the norm

$$\|W\| = \max_{x,y} |W(y|x)|. \tag{9}$$

Secondly, since we are interested in characterizing the regime of quasi-perfect estimation or small mismatch between the true channel probability law W and the decoding metric $\widehat{W}$, we expand the constraint, leading to the following Lemma.

Lemma 1. *For every pair of channel distributions* $(W, \ \widehat{W})$, *we have*

$$D(\widehat{W}\|W|Q_X) = d(\widehat{W}\|W|Q_X) + o(d(\widehat{W}\|W|Q_X)) \tag{10}$$

where

$$d(\widehat{W}\|W|Q_X) = \iint_{\mathbb{R}^2} Q_X(x) \frac{(W(y|x) - \widehat{W}(y|x))^2}{2\widehat{W}(y|x)} \mathrm{d}x\mathrm{d}y. \tag{11}$$

Proof. The proof follows from using the Taylor expansion of the logarithm $\log(1+x) = x - \frac{1}{2}x^2 + o(x^2)$ inside the relative entropy. We have followed the footsteps of [6, eq. (1)–(4)] and rearranged the approximation error terms. For simplicity of notation, we have used natural logarithms. □

The main result is given in the following Theorem.

Theorem 1. *Consider an input distribution* $Q_X(x)$, *a family of continuous-alphabet channels* $W(y|x)$ *and a mismatched decoder* (4) *employing estimated conditional distribution* $\widehat{W}(y|x)$ *satisfying* (3). *Then, for sufficiently small* r, *the worst-case GMI* (8) *can be expressed as*

$$\underline{I}_{\mathrm{GMI}}(Q_X, \widehat{W}) = \sup_{s\geq 0} \left\{ \mathbb{E}_{Q_X\times\widehat{W}}[i_s(X,Y)] - \sqrt{2rV_s} + o(r) \right\} \tag{12}$$

achieved by the worst-case channel (for a given value of parameter s*)*

$$W_\star(y|x) = \widehat{W}(y|x)\left(1 - \sqrt{2r}\frac{i_s(x,y) - \mathbb{E}_{\widehat{W}}[i_s(x,Y)]}{\sqrt{V_s}}\right) \tag{13}$$

with $V_s \triangleq \mathbb{E}_{Q_X}[\mathrm{Var}[i_s(X,Y)|X]]$. *The actual worst-case channel is the one maximizing* (12).

Proof. The proof is detailed in Appendix A. □

IV. Nearest Neighbor Decoding

We next analyze the case of Gaussian i.i.d. codebooks and nearest neighbor decoding for general and additive channels, and analytically characterize the corresponding penalty terms.

A. General Continuous-Alphabet Channels

We analyze general channels by particularizing the results in Section III to Gaussian input and nearest neighbor decoding. We characterize the penalty term appearing in Theorem 1.

Theorem 2. *Consider the input distribution* $Q_X(x)=\mathcal{N}(x;P)$, *a family of continuous-alphabet channels* $W(y|x)$ *and a mismatched decoder* (4) *employing estimated conditional distribution* $\widehat{W}(y|x)=\mathcal{N}(y-x;\sigma^2)$ *satisfying* (3). *Then, for sufficiently small* r, *the worst-case GMI rate satisfies*

$$\underline{I}_{\mathrm{GMI}}(Q_X, \widehat{W}) = \sup_{s\geq 0} \left\{ \frac{1}{2}\log(1+s\Gamma) + \frac{\Gamma}{2}\frac{1-s}{s^{-1}+\Gamma} - \sqrt{r\cdot\frac{\Gamma(s^2\Gamma+2)}{(s^{-1}+\Gamma)^2}} + o(r) \right\} \tag{14}$$

with $\Gamma \triangleq P/\sigma^2$.

Proof. The proof follows from particularizing the results of Theorem 1. Appendix B details the computations. □

Corollary 2.1. *As* $r \to 0$, *the worst-case GMI rate can be accurately approximated by setting* $s = 1$, *to yield*

$$\underline{I}_{\mathrm{GMI}}(Q_X, \widehat{W}) \approx \frac{1}{2}\log(1+\Gamma) - \sqrt{r\cdot\frac{\Gamma(\Gamma+2)}{(1+\Gamma)^2}}. \tag{15}$$

Proof. The proof follows from the fact that as $r \to 0$ the first two terms dominate the summation. Therefore, the result tends to be the one corresponding to the dominant terms; in this case, $s = 1$. □

B. Additive Non-Gaussian Noise Channels

We next analyze the case where the channel is assumed to be noise-additive, i.e., $W(y|x) = W(y-x)$, but the noise distribution W remains unknown. We redefine the small relative entropy ball to contain only additive noise channels $W(y|x) = W(y-x)$ as

$$W \in \mathcal{B} = \{W : \; D(\widehat{W}\|W) \leq r\} \tag{16}$$

where thanks to the additive structure of both channel and metric, the relative entropy becomes independent of the input:

$$D(\widehat{W}\|W) = \int_{\mathbb{R}} \widehat{W}(z) \log \frac{\widehat{W}(z)}{W(z)} \mathrm{d}z. \tag{17}$$

For additive noise channels with a known second-order moment $\mathbb{E}[W^2]$, Lapidoth [4] showed that the rate

$$I_{\mathrm{GMI}}(Q_X) = \frac{1}{2} \log \left(1 + \frac{P}{\mathbb{E}[W^2]}\right) \tag{18}$$

is achievable by Gaussian codebooks and nearest neighbor decoding. The result is relevant in practice, since $\mathbb{E}[W^2]$ turns out to be the only knowledge needed about any noise-additive W to design a reliable transmission scheme attaining (18). Our approach assumes instead an available estimate σ^2 for the unknown second-order moment $\mathbb{E}[W^2]$ and an uncertainty level given by the ball radius r. The worst-case GMI rate is

$$\underline{I}_{\mathrm{GMI}}(Q_X, \widehat{W}) = \min_{W \in \mathcal{B}} \frac{1}{2} \log \left(1 + \frac{P}{\mathbb{E}[W^2]}\right) \tag{19}$$

where the dependence on the noise distribution appears only as a function of its second-order moment.

The following Theorem summarizes the result.

Theorem 3. *Consider a family of additive noise channels $W(y|x) = W(y-x)$ and a mismatched decoder* (4) *employing estimated conditional distribution $\widehat{W}(y|x) = \mathcal{N}(y-x; \sigma^2)$ satisfying* (16)*. Then, for sufficiently small r, Gaussian i.i.d. codebooks achieve the following worst-case GMI rate*

$$\underline{I}_{\mathrm{GMI}}(Q_X, \widehat{W}) = \frac{1}{2} \log \left(1 + \frac{\Gamma}{1 + 2\sqrt{r}}\right) + o(r) \tag{20}$$

with $\Gamma = P/\sigma^2$ the estimated signal-to-noise ratio.

Proof. The proof, detailed in Appendix C, seeks the maximum variance noise distribution in $\mathcal{B}$ defined in (16). □

Corollary 3.1. *As $r \to 0$, the worst-case GMI rate can be expanded as*

$$\underline{I}_{\mathrm{GMI}}(Q_X, \widehat{W}) \approx \frac{1}{2} \log (1 + \Gamma) - \sqrt{r} \cdot \frac{\Gamma}{1 + \Gamma}. \tag{21}$$

C. Example

Figure 1 exemplifies the study for Gaussian i.i.d. codewords and the nearest neighbor decoder. In contrast to the discrete alphabet case [7], computations of the actual worst-case rates are not possible due to computational complexity constraints. Instead, we show results from theorems obviating approximation error terms and compare them against corollaries that make the further approximation of $s = 1$.

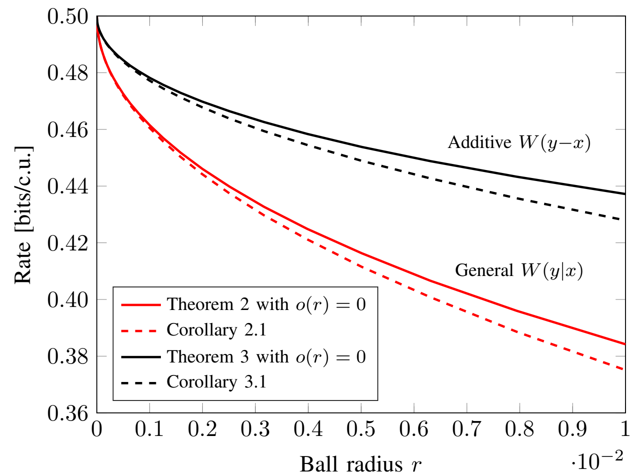

Fig. 1. Worst-case GMI rate versus ball radius r for Gaussian i.i.d. codewords $Q_X(x) = \mathcal{N}(x; P)$ and nearest neighbor decoding $\widehat{W}(y|x) = \mathcal{N}(y-x; \sigma^2)$. Simulation parameters are $\Gamma = 1$ with $P = \sigma^2 = 1$.

The worst-case GMI decays infinitely steeply at $r = 0$, showing that even a very small mismatch between W and $\widehat{W}$ can significantly impact achievable rates. Knowing the channel operation leads to an improved rate. In particular, an approximate measure can be obtained from the ratio between penalty terms of the approximations, $\sqrt{1+\Gamma^{-1}} \geq 1$ achieved at high Γ. The benefit of knowing the channel operation is significant at low Γ.

Appendix A
Proof of Theorem 1

The following problem needs to be addressed:

$$\min_{W \in \mathcal{B}} \iint_{\mathbb{R}^2} Q_X(x) W(y|x) i_s(x,y) \mathrm{d}x\mathrm{d}y \tag{22}$$

where the constraint $W \in \mathcal{B}$ considers continuous conditional distributions $W(y|x) \geq 0$ that satisfy

$$\int_{\mathbb{R}} W(y|x) \mathrm{d}y = 1 \quad , \quad x \in \mathbb{R} \tag{23}$$

$$d(\widehat{W}\|W|Q_X) + o(d(\widehat{W}\|W|Q_X)) \leq r. \tag{24}$$

When r is sufficiently small, the $o(\cdot)$ term can be omitted from the constraint, penalizing the cost function as $o(d(\widehat{W}\|W|Q_X)) = o(r)$ using the new constraint $d(\widehat{W}\|W|Q_X) \leq r$. Therefore, the following variational calculus problem needs to be addressed

$$\min_{\substack{d(\widehat{W}\|W|Q_X) \leq r \\ \int_{\mathbb{R}} W(y|x)\mathrm{d}y = 1, \; x \in \mathbb{R}}} \iint_{\mathbb{R}^2} Q_X(x) W(y|x) i_s(x,y) \mathrm{d}x\mathrm{d}y + o(r). \tag{25}$$

The following Lagrangian needs to be solved

$$\mathcal{L}[W] = \iint_{\mathbb{R}^2} F(x, y, W(y|x)) \mathrm{d}x\mathrm{d}y \tag{26}$$

with

$$F(x,y,W) = Q_X(x) W i_s(x,y) - \lambda(x) W - \rho \cdot Q_X(x) \frac{(W - \widehat{W}(y|x))^2}{2\widehat{W}(y|x)} \quad (27)$$

and where we have obviated W-independent constant terms.

Now, we expand $\mathcal{L}[W]$ by adopting continuous variations $v(x,y)$, leading to the following general variation truncated to the second order

$$\Delta\mathcal{L}[W] = \iint_{\mathbb{R}^2} (F_W v(x,y) + \tfrac{1}{2} F_{WW} v^2(x,y)) \mathrm{d}x\mathrm{d}y \quad (28)$$

$$F_W = Q_X(x) i_s(x,y) - \lambda(x) - \rho \cdot Q_X(x) \frac{W - \widehat{W}(y|x)}{\widehat{W}(y|x)} \quad (29)$$

$$F_{WW} = -\frac{\rho Q_X(x)}{\widehat{W}(y|x)}. \quad (30)$$

We have used subscripts on F to denote the partial derivatives: $F_W = \frac{\partial F}{\partial W}$ and $F_{WW} = \frac{\partial^2 F}{\partial^2 W}$. The stationary point equation is found by setting the first term of (28) to zero and by invoking the Fundamental Lemma of the Calculus of Variations [8]. Specifically, we find the stationary point $W_\star(y|x)$ by solving $F_W(x,y,W_\star(y|x)) = 0$ in $(x,y) \in \mathbb{R}^2$. This gives

$$W_\star(y|x) = \widehat{W}(y|x) \left(1 + \frac{Q_X(x) i_s(x,y) - \lambda(x)}{\rho Q_X(x)}\right). \quad (31)$$

By using the constraints, we obtain:

$$\lambda(x) = Q_X(x) \mathbb{E}_{\widehat{W}}[i_s(x,Y)] \quad (32)$$

$$\rho = -\sqrt{V_s/(2r)} \quad (33)$$

$$V_s = \mathbb{E}_{Q_X}[\mathrm{Var}[i_s(X,Y)|X]]. \quad (34)$$

The worst-case GMI is shown in (12). It is easy to check that $W_\star$ corresponds to a minimum as $\int_{\mathbb{R}^2} F_{WW} v^2(x,y) \mathrm{d}x\mathrm{d}y > 0$.

Appendix B
Auxiliary Computations

For Gaussian i.i.d. codebooks $Q_X(x) = \mathcal{N}(x; P)$ and the nearest neighbor decoder $\widehat{W}(y|x) = \mathcal{N}(y-x; \sigma^2)$, we have

$$\widehat{W}(y|x)^s = \frac{\sqrt{2\pi\sigma^2 s^{-1}}}{\sqrt[s]{2\pi}} \cdot \mathcal{N}(y-x; \sigma^2 s^{-1}) \quad (35)$$

$$\mathbb{E}_{Q_X}[\widehat{W}(y|X)^s] = \frac{\sqrt{2\pi\sigma^2 s^{-1}}}{\sqrt[s]{2\pi}} \cdot \mathcal{N}(y; \sigma^2 s^{-1} + P). \quad (36)$$

The mismatched information density is (*cf.* [9, eq. (28)-(29)])

$$i_s(x,y) = C - \frac{s}{2} \frac{(y-x)^2}{\sigma^2} + \frac{1}{2} \frac{y^2}{\sigma^2 s^{-1} + P} \quad (37)$$

with

$$C = \frac{1}{2} \log\left(1 + \frac{sP}{\sigma^2}\right), \quad (38)$$

and the respective expectations are

$$\mathbb{E}_{\widehat{W}}[i_s(x,Y)] = C + \frac{1}{2} \frac{x^2 - sP}{\sigma^2 s^{-1} + P} \quad (39)$$

$$\mathbb{E}_{Q_X \times \widehat{W}}[i_s(X,Y)] = C + \frac{P}{2} \frac{1-s}{\sigma^2 s^{-1} + P}. \quad (40)$$

The term

$$V_s = \mathbb{E}_{Q_X \times \widehat{W}}[i_s^2(X,Y)] - \mathbb{E}_{Q_X}[\mathbb{E}^2_{\widehat{W}}[i_s(X,Y)|X]] \quad (41)$$

$$= \frac{P}{2} \frac{2\sigma^2 + s^2 P}{(\sigma^2 s^{-1} + P)^2} \quad (42)$$

is computed from

$$\mathbb{E}_{Q_X \times \widehat{W}}[i_s^2(X,Y)] = C^2 + \frac{CP(1-s)}{(\sigma^2 s^{-1} + P)^2} + \frac{3s^2}{4} + \frac{3(\sigma^2 + P)^2}{4(\sigma^2 s^{-1} + P)^2} - \frac{1}{2} \frac{s(3\sigma^2 + P)}{(\sigma^2 s^{-1} + P)^2} \quad (43)$$

$$\mathbb{E}_{Q_X}[\mathbb{E}^2_{\widehat{W}}[i_s(X,Y)|X]] = C^2 + \frac{CP(1-s)}{\sigma^2 s^{-1} + P} + \frac{P^2(3 - 2s + s^2)}{4(\sigma^2 s^{-1} + P)^2} \quad (44)$$

Appendix C
Maximum Variance Noise Distribution

We aim to find the noise distribution with maximum second-order moment in $\mathcal{B}$ by solving

$$\mathbb{E}[W_\star^2] = \max_{W \in \mathcal{B}} \int_{\mathbb{R}} z^2 W(z) \mathrm{d}z \quad (45)$$

$$= \max_{\substack{d(\widehat{W} \| W) \le r \\ \int_{\mathbb{R}} W(z)\mathrm{d}z = 1}} \int_{\mathbb{R}} z^2 W(z) \mathrm{d}z + o(r). \quad (46)$$

We replicate the analysis in Appendix A to solve

$$\mathcal{L}[W] = \int_{\mathbb{R}} F(z, W(z)) \mathrm{d}z \quad (47)$$

$$F(z,W) = z^2 W - \lambda W - \rho \cdot \frac{(W - \widehat{W}(z))^2}{2\widehat{W}(z)}. \quad (48)$$

The worst-case channel is found by solving $F_W(z, W_\star(z)) = 0$ in $z \in \mathbb{R}$. This is solved for $\lambda = \sigma^2$ and $\rho = \sigma^2/\sqrt{r}$, yielding

$$W_\star(z) = \widehat{W}(z) \left(1 + \sqrt{r} \cdot \frac{z^2 - \sigma^2}{\sigma^2}\right) \quad (49)$$

$$\mathbb{E}[W_\star^2] = \sigma^2(1 + 2\sqrt{r}) + o(r). \quad (50)$$

References

[1] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons Inc., 1968.

[2] J. Scarlett, A. Guillén i Fàbregas, A. Somekh-Baruch, and A. Martinez, "Information-Theoretic Foundations of Mismatched Decoding," *Foundations and Trends® in Communications and Information Theory*, vol. 17, no. 2–3, pp. 149–401, 2020.

[3] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 35–43, 1995.

[4] A. Lapidoth, "Nearest neighbor decoding for additive non-gaussian noise channels," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1520–1529, 1996.

[5] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2148–2177, 1998.

[6] S. Borade and L. Zheng, "Euclidean information theory," in *2008 IEEE International Zurich Seminar on Communications*, 2008, pp. 14–17.

[7] P. Patel, F. Molina, and A. Guillén i Fàbregas, "Mismatched decoding: Generalized mutual information under small mismatch," in *2024 IEEE International Zurich Seminar on Information and Communication*, 2024.

[8] I. M. Gelfand, R. A. Silverman *et al.*, *Calculus of variations*. Courier Corporation, 2000.

[9] J. Scarlett, V. Tan, and G. Durisi, "The dispersion of nearest-neighbor decoding for additive non-gaussian channels," *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 81–92, 2017.