Channel-Optimized Strategic Quantizer Design via Dynamic Programming

Anju Anand
Electrical and Computer Engineering
Binghamton University, SUNY NY, USA
https://orcid.org/0000-0002-9378-5833
aanand6@binghamton.edu

Emrah Akyol
Electrical and Computer Engineering
Binghamton University, SUNY NY, USA
https://orcid.org/0000-0002-0663-1677
eakyol@binghamton.edu

2023 IEEE Statistical Signal Processing Workshop (SSP) | 978-1-6654-5245-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/SSP53291.2023.10207995

***Abstract*—We consider the design problem of a strategic quantizer over a noisy channel, extending the classical work on channel-optimized quantization to strategic settings where the encoder and the decoder have misaligned objectives. Building on our recent work on strategic quantization over noiseless channels, we employ a random channel index assignment mapping, as done in prior work on classical channel-optimized quantizer design literature, combined with a dynamic programming approach to optimize quantization boundaries. Our analysis and numerical results demonstrate several interesting aspects of channel-optimized strategic quantization which do not appear in its classical (nonstrategic) counterpart. The codes are available at: https://tinyurl.com/ssp2023dpnoise.**

***Index Terms*—Quantization, joint source-channel coding, game theory, dynamic programming**

I. Introduction

This paper is concerned with the quantizer design problem for the setting where two agents (the encoder and the decoder) with misaligned objectives communicate over a memoryless noisy channel. The classical (non-strategic) counterparts of this problem have been investigated thoroughly in the literature, see e.g., [1]–[8]. We here carry out the analysis to strategic communication cases, see e.g., [9]–[11] where the encoder and the decoder have different objectives, as opposed to the classical communication paradigm where the encoder and the decoder form a team with identical objectives.

This problem can also be solved using gradient descent based methods, as we have analyzed in [12], [13]. However, the quantizer might be only locally optimal. Building on our recent work on strategic quantizer design over a perfect (noiseless) communication channel [11], and inspired by the prior literature on classical (nonstrategic) channel-optimized quantization via dynamic programming (DP) [8], we analyze and design the channel-optimized quantizer for strategic settings, used in conjunction with random index assignment. Our main computational design tool is dynamic programming, which was first used to implement a quantizer in [14].

The problem setting has a plethora of applications in engineering as well as Economics. This class of problems, i.e., "information design," also known as "Bayesian Persuasion," is an active research area in Economics. For an engineering application, consider the Internet of Things, where agents with misaligned objectives communicate over rate-limited communication channels with delay constraints.

This research is supported by the NSF via grants CCF #1910715 and CAREER #2048042.

II. Problem Formulation

Consider the following quantization problem: an encoder observes a realization of the source $X \in \mathcal{X}$ with a probability distribution μ and maps it to a message $Z \in \mathcal{Z}$, where $\mathcal{Z}$ is a set of discrete messages with a cardinality constraint $|\mathcal{Z}| \leq M$ using a non-injective mapping $Q : \mathcal{X} \to \mathcal{Z}$. An index mapping $\pi : [1 : M] \to [1 : M]$ is chosen uniformly at random and is applied to the message Z. The message $\pi(Z)$ is transmitted over a noisy channel with transition probability matrix $p(z_j|z_i)$. After receiving the message Z', the decoder applies a mapping $\phi : \mathcal{Z} \to \mathcal{Y}$ on the message Z' and takes an action $Y = \phi(Z')$. The encoder and the decoder minimize their respective objectives $D_E = \mathbb{E}_\pi\{\mathbb{E}\{\eta_E(X, Y)|\pi\}\}$ and $D_D = \mathbb{E}_\pi\{\mathbb{E}\{\eta_D(X, Y)|\pi\}\}$, which are misaligned ($\eta_E \neq \eta_D$). The encoder designs Q *ex-ante*, i.e., without the knowledge of the realization of X, using only the objectives η_E and η_D, the statistics of the source $\mu(\cdot)$, and the channel parameters (transition probability matrix $p(z_j|z_i)$). The objectives (η_E and η_D), the shared prior (μ), the index assignment (π), the channel transition probability matrix ($p(z_j|z_i)$), and the mapping (Q) are known to the encoder and the decoder. The problem is to design Q for the equilibrium, i.e., the encoder minimizes its distortion if used with a corresponding decoder that minimizes its own distortion. We consider the Stackelberg equilibrium as indicated by the problem formulation. The problem statement is presented in the box, and the communication setting is given in Figure 1.

III. Main Results

A. Analysis

Let X take values from the source alphabet $\mathcal{X} \in [a, b]$. The set $\mathcal{X}$ is divided into mutually exclusive and exhaustive sets $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_M$. The message

$$z_i = \pi(Q(x)), \quad x \in \mathcal{V}_m$$

where $Q(x) = z_m \, \forall x \in \mathcal{V}_m$, π is a bijective index mapping $\pi : \{1, \ldots, M\} \to \{1, \ldots, M\}$, $\pi(z_m) = z_i$ is transmitted over

Problem. *Using random index assignment for a given noisy channel with rate R and bit error rate p_{err}, scalar source $\mathcal{X} \in [a,b]$ points with a given probability distribution μ, find the decision boundaries $\mathbf{q} = [x_0, x_1, \ldots, x_M]$ and actions $\mathbf{y}(\mathbf{q}) = [y_1, \ldots, y_M]$ as a function of boundaries that satisfy:*

$$\mathbf{q}^* = \arg\min_{\mathbf{q}} \sum_{m=1}^{M} \mathbb{E}_\pi\{\mathbb{E}\{\eta_E(x, y_m^*(\mathbf{q}))|\pi, x \in [x_{m-1}, x_m)\}\},$$

where actions $y_m^*(\mathbf{q}) = \arg\min_{y_m \in \mathcal{Y}} \mathbb{E}_\pi\{\mathbb{E}\{\eta_D(x, y_m)|\pi, x \in [x_{m-1}, x_m)\}\} \, \forall m \in [1:M]$, *and the rate satisfies* $\log M \leq R$.

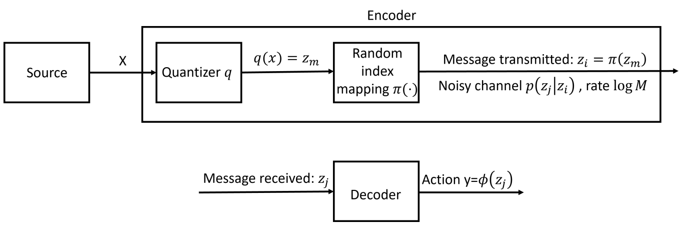


Fig. 1. Communication diagram

a noisy channel and received as z_j with probability $p(z_j|z_i)$. The decoder receives the message and takes the action

$$y = \phi(z_j),$$

which includes applying the inverse mapping π^{-1} first. We make the following "monotonicity" assumption.

Assumption 1. *$\mathcal{V}_m$ is convex for all $m \in [1:M]$.*

Under assumption 1, $\mathcal{V}_m$ is an interval since X is a scalar,

$$\mathcal{V}_m = [x_{m-1}, x_m)$$

where $x_{m-1} < x_m, x_0 = a, x_M = b$. The encoder chooses the boundary decision levels $\mathbf{q} = [x_0, x_1, \ldots, x_M]$. The decoder determines its actions $\mathbf{y} = [y_1, \ldots, y_M]$ as the best response to $\mathbf{q}$ to minimize its cost $D_D = \mathbb{E}_\pi\{\mathbb{E}\{\eta_D(X,Y)|\pi\}\}$ for $m \in [1:M]$ as follows

$$y_m^* = \arg\min_{y_m \in \mathcal{Y}} \sum_{m=1}^{M} \mathbb{E}_\pi\{\mathbb{E}\{\eta_D(x, y_m)|\pi, x \in \mathcal{V}_m\}\}.$$

The average symbol error probability of the channel is

$$p_{err} = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1, j\neq i}^{M} p(z_j|z_i).$$

Let $c_1 \triangleq p_{err}/(M-1)$, $c_2 \triangleq 1 - Mc_1$. The integrals expressed throughout this paper are defined over the set $\mathcal{V}_i$ unless specified otherwise. The probability that the receiver receives the noisy message $\hat{z} = z_j$ if z_i was transmitted using $\pi(Q(x)) = z_i$ is $p(z_j|z_i)$, the channel transition probability.

The end-to-end distortion given an index assignment π is

$$\mathbb{E}\{\eta_s|\pi\} = \sum_{i=1}^{M} \int \sum_{j=1}^{M} \eta_s(x, y_j) p(z_j|z_i) \mathrm{d}\mu.$$

The average distortion over all possible index assignments is

$$D_s = \sum_{i=1}^{M} \int \sum_{j=1}^{M} \eta_s(x, y_j) \mathbb{E}_\pi(p(z_j|z_i)) \mathrm{d}\mu = I_{j\neq i} + I_{j=i}$$

where $I_{k,j\neq i}$ and $I_{k,j=i}$ are defined as follows:

$$\begin{aligned} I_{j\neq i} &= \sum_{i=1}^{M} \int \sum_{j=1, j\neq i}^{M} \eta_s(x, y_j) \mathbb{E}_\pi(p(z_j|z_i)) \mathrm{d}\mu \\ &= c_1 \sum_{i=1}^{M} \int \sum_{j=1, j\neq i}^{M} \eta_s(x, y_j) \mathrm{d}\mu, \\ I_{j=i} &= \sum_{i=1}^{M} \int \eta_s(x, y_i) \mathbb{E}_\pi(p(z_i|z_i)) \mathrm{d}\mu \\ &= .(1 - p_{err}) \sum_{i=1}^{M} \int \eta_s(x, y_i) \mathrm{d}\mu. \end{aligned}$$

We next express the terms in a way that can be approached via dynamic programming:

$$I_{j\neq i} = c_1 \sum_{i=1}^{M} \int \left(\sum_{j=1}^{M} \eta_s(x, y_j) - \eta_s(x, y_i) \right) \mathrm{d}\mu.$$

We assume $0 < p_{err} < \frac{M-1}{M}$ so that $c_1, c_2 > 0$. The average distortion and optimum decoder reconstruction for $i \in [1:M]$,

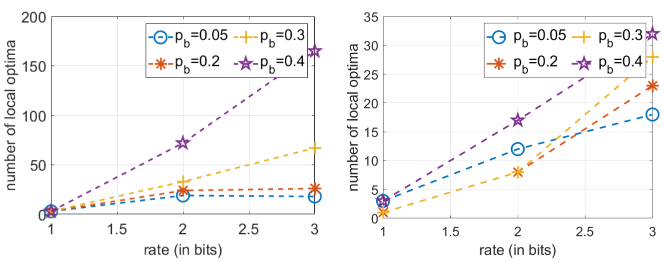
$$D_s = \sum_{i=1}^{M} \left(c_1 \mathbb{E}\{\eta_s(x, y_i)\} + c_2 \int \eta_s(x, y_i) \mathrm{d}\mu \right),$$

$$y_i = \arg\min_{y \in \mathcal{Y}} \left(c_1 \mathbb{E}\{\eta_D(x, y_i)\} + c_2 \int \eta_D(x, y_i) \mathrm{d}\mu \right).$$

The average distortion can be written including the distortion in the noiseless setting (without random index assignment) as

$$D_s = \sum_{i=1}^{M} c_1 \mathbb{E}\{\eta_s(x, y_i)\} + c_2 \overline{D^s}, \ \overline{D^s} = \sum_{i=1}^{M} \int \eta_s(x, y_i) \mathrm{d}\mu. \tag{1}$$

Since the problem is Stackelberg in nature, with the encoder (leader) choosing the quantization decision levels $\mathbf{q}$ first, followed by the decoder (follower) choosing the quantization representative levels as a function of the decision levels $\mathbf{y}(\mathbf{q})$, it allows a gradient descent based solution optimizing the quantization decision levels. We have analyzed the gradient descent based solution for both noiseless and noisy settings in [12], [13] respectively, where we also extend the noisy setting further to 2-dimensional quantization with scalar reconstruction. We show in [12] that a strategic version of Lloyd-Max where the encoder and decoder iteratively try to optimize their distortion in response to the other may not converge to even a locally optimal solution. While gradient descent can be a solution method to this problem, in Figure 2 in [12] we see that unlike classical quantization, the strategic version can have multiple local minima even when used in conjunction with log-concave sources on a noiseless channel. In Figure 2, we show the number of locally optimal quantizers with rate for a non-strategic $\eta_E = (x-y)^2$ (increases with bit error rate) and a strategic quantizer $\eta_E = (x^3-y)^2$ for a Gaussian source with MSE decoder $\eta_D = (x-y)^2$ for both cases. To mitigate the issue of local optima, in this paper we approach the problem with a different solution method by using dynamic programming which gives us the globally optimal solution. We discretize the source in order to make the algorithm tractable for DP. The condition $c_1, c_2 > 0$ is not required for a gradient based solution since it minimizes the average distortion as a whole, while it is required for a DP based algorithm since the terms in the summation over the quantization regions are individually minimized.



(a) $\eta_E = (x-y)^2$ (b) $\eta_E = (x^3-y)^2$

Fig. 2. Number of local optima quantizers with $\eta_D = (x-y)^2$ for (a), (b).

B. Dynamic Programming Algorithm

While our analysis holds for general source alphabets, for making the dynamic programming algorithm tractable, the source has to be purely discrete. In case it is not, we approximate it by uniformly quantizing $[a, b]$ to N points to obtain an ordered ascending set of N points, i.e.,

$$\mathcal{X} = \{a + (1+2t)\delta\}, \quad t = 0, \dots, N-1$$

where $\delta = (b-a)/(2N)$ with the probability mass function

$$P(t) = \int_{x_t-\delta}^{x_t+\delta} d\mu, \quad t \in [0 : N-1].$$

We carry out the analysis in a continuous setting. For the discrete setting, the integrals in the equations transform into summations over $x_t \in \mathcal{V}_i$. We define the following terms:

1) The encoder and decoder costs for source interval $[\alpha, \beta)$ for a given action y for $s \in \{E, D\}$:

$$C_s(\alpha, \beta, y) = c_1 \mathbb{E}\{\eta_s(x, y)\} + c_2 \int_\alpha^\beta \eta_s(x_t, y) \mathrm{d}\mu.$$

2) The decoder's optimal action for the interval $[\alpha, \beta)$:

$$\kappa(\alpha, \beta) = \arg\min_{y \in \mathcal{Y}} C_D(\alpha, \beta, y).$$

3) Costs for the source interval $[\alpha, \beta)$ in conjunction with the optimal action for $s \in \{E, D\}$:

$$\epsilon_s(\alpha, \beta) = C_s(\alpha, \beta, \kappa(\alpha, \beta)).$$

4) Equilibrium costs associated with the m level optimal strategic quantizer for $[x_0, \beta)$ for $s \in \{E, D\}$, where $\mathbf{q} = [x_0, \dots, x_m], a = x_0 < \dots < x_m = \beta$:

$$D_m(x_0, \beta) = \min_{\mathbf{q}} \sum_{i=1}^m \epsilon_s(x_{i-1}, x_i).$$

5) The set of all non-empty convex subsets of $\mathcal{X}$:

$$\mathcal{S} = \{[\alpha, \beta) : \alpha, \beta \in \mathcal{X}, \alpha < \beta\}.$$

The encoder minimizes D_E with the choice of the quantizer decision levels $\mathbf{q}^* = [x_0^*, \dots, x_M^*]$, and the decoder chooses the representative levels y_m^* to minimize its distortion D_D,

$$\mathbf{q}^* = \arg\min_{\mathbf{q}} \sum_{i=1}^M \int \sum_{j=1}^M \eta_E(x, y_j^*) \mathbb{E}_\pi(p(z_j|z_i)) \mathrm{d}\mu,$$

$$y_m^* = \arg\min_{y} \sum_{i=1}^M \int \eta_D(x, y) \mathbb{E}_\pi(p(z_j|z_i)) \mathrm{d}\mu.$$

The encoder's distortion in quantizing the interval $[\alpha, \beta)$ with one representation level, $\epsilon_E(\alpha, \beta)$, is computed for each $[\alpha, \beta) \in \mathcal{S}$. We set the $1-$level distortion $D_1(x_0, \beta) = \epsilon(x_0, \beta), \beta \in \mathcal{X} \backslash x_0$. The $m-$level distortion for an interval $[x_0, \beta) \in \mathcal{S}$ due to quantizing the interval with m representative levels can be written in terms of the $1-$level distortions,

$$D_m(x_0, \beta) = \min_{\substack{x_0, \dots, x_m \in \mathcal{X} \\ a = x_0 < x_1 < \dots < x_m = \beta}} \sum_{i=1}^m D_1(x_{i-1}, x_i).$$

The optimization for $m-$level quantization of $[x_0, \beta)$ can be written as the sum of $(m-1)$ level quantization of $[x_0, \alpha)$ and $1-$level quantization of $[\alpha, \beta)$ as the Bellman equations,

$$D_m(x_0, \beta) = D_{m-1}(x_0, r_{m-1}(x_0, \beta) + D_1(r_{m-1}(x_0, \beta), \beta)$$
$$r_{m-1}(x_0, x_m) \triangleq \arg\min_{\alpha \in \mathcal{X}} \{D_{m-1}(x_0, \alpha) + D_1(\alpha, \beta)\}.$$

Dynamic programming requires a forward and a backward pass. During the forward pass, we compute and store $r_{m-1}(x_0, \beta)$ and $D_m(x_0, \beta)$ for each pair $(m, \beta), m \in [2 : M], \beta \in \mathcal{X} \backslash x_0$ recursively starting from $m = 2$ using the pre-computed values of $D_1(\alpha, \beta)$. In the backward pass, we set $x_0^* = a, x_M^* = b$ and compute optimal decision (x_m^*) and representative levels (y_m^*) recursively as

$$x_{m-1}^* = r_{m-1}(x_0^*, x_m^*), \quad m = M, \dots, 2,$$
$$y_m^* = \kappa(x_{m-1}^*, x_m^*), \quad m = [1 : M].$$

Remark 1. *The required accuracy can be achieved when the source is discretized by increasing N.*

Remark 2. *The worst-case complexity of the DP algorithm in terms of the three parameters $N, M, |\mathcal{Y}|$ is $\mathcal{O}(N^3|\mathcal{Y}| + N^2 M)$.*

C. Special Distortion Functions

If the distortion functions are of the specific form

$$\eta_E(x,y) = (f(x) - y)^2, \quad \eta_D(x,y) = (x-y)^2$$

where $f(\cdot)$ is any (Borel) measurable function, then the analysis simplifies as follows. Using (1), the average encoder distortion over all possible index assignments

$$D_E = c_1 \sum_{i=1}^{M} (f_2 - 2y_i f_1 + y_i^2) + c_2 \overline{D^E},$$

where we denote $\mathbb{E}\{f(x)\}$ and $\mathbb{E}\{f^2(x)\}$ as f_1 and f_2 respectively. The terms $C_E(\alpha, \beta, y)$ and $D_E(n,m)$ then simplify as

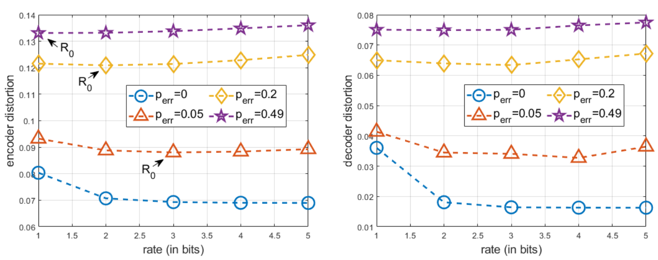
$$C_E = c_1(-2y_i f_1 + y_i^2) + c_2 \int (f(x)-y)^2 \mathrm{d}\mu,$$

$$D_m(x_0, \beta) = \min_{\mathbf{q}} \sum_{j=1}^{m} \epsilon_E(x_{j-1}, x_j) + m c_1 f_2.$$

The actions associated with $\mathcal{V}_i$ are given by minimizing $J_i(y)$ using the KKT optimality condition, $\partial J_i / \partial y = 0$,

$$J_i(y) = c_1(y - \mathbb{E}\{x\})^2 + c_2 \int (x-y)^2 \mathrm{d}\mu,$$

$$y_i = \arg\min_{y \in \mathcal{Y}} J_i(y) = \frac{c_1 \mathbb{E}\{x\} + c_2 \int x \mathrm{d}\mu}{c_1 + c_2 \int d\mu}.$$



(a) Encoder distortion (b) Decoder distortion

Fig. 3. $\eta_E = (x^3 - y)^2, \eta_D = (x-y)^2$

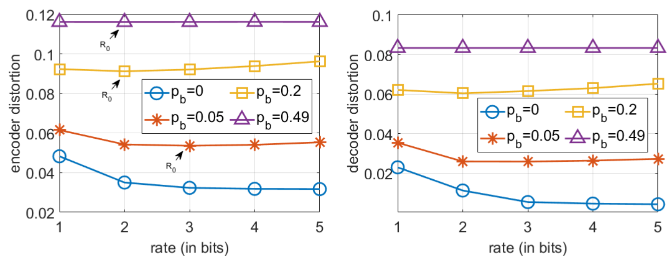
IV. Numerical Results

We consider a source uniformly distributed on $[0,1]$, i.e., $X \sim U(0,1)$, and a binary symmetric channel with crossover probability $p_b \leq \frac{1}{2}$, which yields

$$p_{err} = 1 - (1-p_b)^{\log M}.$$

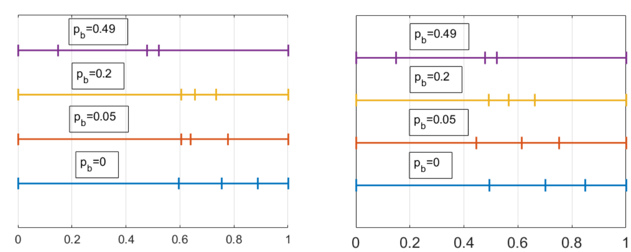
We take the decoder distortion measure as $\eta_D(x,y) = (x-y)^2$, and consider two different cases of encoder distortion, $\eta_E^1(x,y) = (x^3-y)^2$ and $\eta_E^2(x,y) = (x^2-y)^2$.

We plot the encoder and the decoder distortions associated with η_D in conjunction with η_E^1 and η_E^2 in Figures 3 and 4 respectively. One surprising aspect of these results is that the encoder distortion may be increasing with rate at high resolution. This is due to the strategic aspect of the problem, i.e., at high resolution, the encoder is forced to be more revealing than its optimal choice. The impact of channel noise can be seen in the rate threshold, i.e., the cutoff rate, the smallest R_0 for which $R > R_0$ implies $D(R) \geq D(R_0)$. Numerical results shown in Figures 3 and 4 suggest that as p_b increases, the cutoff rate R_0 gets smaller. This observation indicates that in the high-rate regime, the optimal strategic encoder might choose not to utilize the channel rate fully. We plot the obtained quantizers in Figure 5. The numerical results depicted below suggest that the encoder is less revealing with increasing p_b. While due to space constraints, we only present the results for a Uniform source here, that for a Gaussian source, as well as the codes that produced these results are available at https://tinyurl.com/ssp2023dpnoise.



(a) Encoder distortion (b) Decoder distortion

Fig. 4. $\eta_E = (x^2 - y)^2, \eta_D = (x-y)^2$



(a) $\eta_E = (x^3 - y)^2$ (b) $\eta_E = (x^2 - y)^2$

Fig. 5. Quantizers for $M = 4$ with $\eta_D = (x-y)^2$ for (a), (b).

Remark 3. *Note that our findings apply to any general distortion measure. We take the decoder's measure as MSE, the most commonly used distortion metric. The choice of the encoder's distortion measure ($\eta_E \neq \eta_D$ due to problem formulation) is arbitrary; however, some measures yield non-interesting solutions such as nonrevealing (encoder does not send any information) or fully-revealing (the problem simplifies to non-strategic quantization). Hence we chose a measure that yields results that demonstrate the interesting aspects of strategic quantization.*

V. Conclusions

In this paper, we extended our DP-based strategic quantizer design approach to the problem of strategic quantization with channel noise. Our analysis and the obtained numerical results have uncovered several aspects of the optimal strategic quantizers in this noisy channel setting that are not observed in its classical (nonstrategic) counterpart.

References

[1] J. Dunham and R. Gray, "Joint Source and Noisy Channel Trellis Encoding (Corresp.)," *IEEE Transactions on Information Theory*, vol. 27, no. 4, pp. 516–519, 1981.

[2] E. Ayanoglu and R. Gray, "The Design of Joint Source and Channel Trellis Waveform Coders," *IEEE Transactions on Information Theory*, vol. 33, no. 6, pp. 855–865, 1987.

[3] N. Farvardin, "A Study of Vector Quantization for Noisy Channels," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 799–809, 1990.

[4] D. Miller and K. Rose, "Combined Source-Channel Vector Quantization Using Deterministic Annealing," *IEEE Transactions on Communications*, vol. 42, no. 234, pp. 347–356, 1994.

[5] S. Gadkari and K. Rose, "Robust Vector Quantizer Design by Noisy Channel Relaxation," *IEEE Transactions on Communications*, vol. 47, no. 8, pp. 1113–1116, 1999.

[6] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Springer Sci. & Business Media, 2012, vol. 159.

[7] X. Yu, H. Wang, and E.-H. Yang, "Design and Analysis of Optimal Noisy Channel Quantization with Random Index Assignment," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5796–5804, 2010.

[8] S. Dumitrescu, "On the Design of Optimal Noisy Channel Scalar Quantizer with Random Index Assignment," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 724–735, 2016.

[9] E. Akyol, C. Langbort, and T. Başar, "Information-Theoretic Approach to Strategic Communication as a Hierarchical Game," *Proceedings of the IEEE*, vol. 105, no. 2, pp. 205–218, 2016.

[10] E. Kamenica and M. Gentzkow, "Bayesian Persuasion," *American Economic Review*, vol. 101, no. 6, pp. 2590–2615, 2011.

[11] A. Anand and E. Akyol, "Optimal Strategic Quantizer Design via Dynamic Programming," in *Proceedings of the IEEE Data Compression Conference*. IEEE, 2022, pp. 173–181.

[12] E. Akyol and A. Anand, "Strategic Quantization," *IEEE International Symposium on Information Theory*, 2023. Available at https://tinyurl.com/GDnoiseless.

[13] A. Anand and E. Akyol, "Strategic Quantization over a Noisy Channel," 2023. Available at https://tinyurl.com/ssp2023dpnoise.

[14] J. D. Bruce, "Optimum Quantization," Ph.D. dissertation, MIT Research Laboratory of Electronics, 1965.