Strategic Quantization over a Noisy Channel

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 57th Asilomar Conference on Signals, Systems, and Computers | 979-8-3503-2574-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/IEEECONF59524.2023.10476788

***Abstract*—This paper is concerned with the strategic quantization setting where the encoder and the decoder have misaligned objectives and communicate over a noisy channel, extending the work on classical channel-optimized quantization. This problem without the quantization constraint has been well-studied under the theme of information design problems in Economics. It is more appealing and relevant to engineering applications with a constraint on the cardinality of the message space. We consider a scalar source X and develop a gradient-descent based solution in conjunction with random index assignment, which has been used in prior literature on classical channel-optimized quantizaton. In our prior work, we used dynamic programming for this problem. Here, we employ gradient descent to reduce the complexity of the algorithm. We finally present numerical results obtained via the proposed algorithm that suggest its validity and demonstrate the strategic quantization features that differentiate it from its classical counterpart. The codes are available at: https://tinyurl.com/asilomar2023.**

***Index Terms*—Quantization, joint source-channel coding, game theory, gradient descent**

I. Introduction

Consider the communication between two smart cars from competing manufacturers, such as Tesla and Honda. The Tesla car (decoder) solicits specific information from the Honda car (encoder) to determine whether to alter its route in response to traffic congestion. While Tesla aims to estimate the traffic congestion accurately, Honda's objective is to make Tesla take a specific action, such as changing its route. Honda car has no incentive to convey a truthful congestion estimate since its objective is different from that of Tesla. To incentivize Tesla to utilize Honda's information though Tesla is aware of Honda's motives, Honda has to ensure that Tesla gains in acting according to its information, that is, the distortion in using Honda's input is lower than that in ignoring it. Realistically, assuming a fixed-rate noisy channel, how would these cars communicate? Our analysis in [1] enables us to quantitatively study such problems. Honda has three different behavioural choices: it can choose not to communicate (non-revealing strategy), can communicate exactly what the Tesla wants (fully-revealing strategy), or it can craft a message that would make Tesla change its route (partially revealing strategy). Tesla can choose to not use Honda's message, if it is statistically too far from the truth. Hence, crafting an optimal message for Honda that would serve its own objective, knowing that Tesla's objective differs from it, is a significant research challenge.

This problem without any constraints on the cardinality of the message space has been studied extensively in Economics literature as information design or Bayesian persuasion [2], [3]. Broadly, these areas of research investigate how a communication system designer (sender) leverages information to impact the actions of the receiver [4], [5]. A related but distinctly different variation of signaling games, called cheap talk, where the encoder chooses the mapping from the source X to message Z *after* observing it, *ex-post*, showed that quantizers can arise as equilibrium strategies endogenously, without an external constraint [6], [7]. Since the encoder chooses the mapping only after observing the source realization, both agents form a strategy that is the best response to each other, resulting in a Nash equilibrium.

The classical (non-strategic) counterpart of communication over a fixed-rate noisy channel, i.e., channel-optimized quantization, has been investigated thoroughly in the literature, see e.g., [8]–[15]. We here carry out the analysis to strategic communication cases, see e.g., [3], [16], [17] 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.

We studied strategic quantizer design over a perfect (noiseless) communication channel in [1], [17], and analyzed strategic quantization of a noisy source in [18] and obtained results similar to [19]. We presented a dynamic-programming based algorithm in conjunction with random index mapping optimization method for channel-optimized strategic quantization in [20] (similar methods have been presented for classical quantization over a noisy channel by [14], [15]).

Compared to our recent related work on channel-optimized strategic quantization [20], the contribution of this paper is that we employ a gradient-descent based optimization as opposed to dynamic programming, enabling a lower-complexity design at the cost of possibly moving from global to local optimality.

II. Problem Formulation

Consider the following scalar quantization problem: an encoder observes a realization of the source $X \in \mathcal{X} \in [a, b]$ with a probability distribution f, and maps X to a message $Z \in \mathcal{Z}$, where $\mathcal{Z}$ is a set of discrete messages with a

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

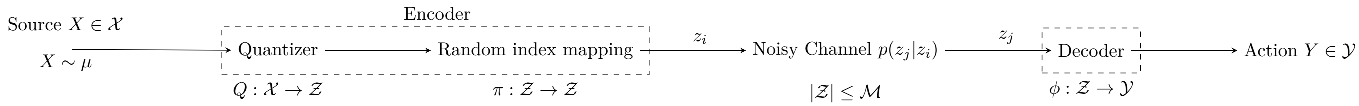


Fig. 1: Communication diagram

Problem. *Using a noisy channel with rate R and probability transition matrix $p(j|i)$, with a scalar source $X \in \mathcal{X}$ with a probability distribution $f(x)$, and an index mapping $\pi : \{1, \dots, M\} \to \{1, \dots, M\}$ chosen uniformly at random, find the quantizer decision levels $\mathbf{q}$, and actions $\mathbf{y}(\mathbf{q}) = [y_1, \dots, y_M]$ as a function of the set of quantizer decision levels that satisfy:*

$$\mathbf{q}^* = \arg\min_{\mathbf{q}} \sum_{m=1}^{M} \mathbb{E}_\pi\{\mathbb{E}\{\eta_E(X, \mathbf{y}^*)|\pi, x \in \mathcal{V}_m\}\},$$

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

cardinality constraint $|\mathcal{Z}| \leq M$ using a non-injective mapping, $\mathbf{q} : \mathcal{X} \to \mathcal{Z}$. 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 $p(z_j|z_i) = p(j|i)$. After receiving the message Z', the decoder applies a mapping $\phi : \mathcal{Z} \to \mathcal{Y}$, where $|\mathcal{Y}| = |\mathcal{Z}|$, on the message Z' (which includes the inverse mapping $\pi^{-1}(Z')$ first) and takes an action $Y = \phi(Z')$. The encoder and 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 $\mathbf{q}$ *ex-ante*, i.e., without the knowledge of the realization of X, using only the objectives η_E and η_D, and the statistics of the source $f(\cdot)$. The objectives (η_E and η_D), the shared prior (f), the channel parameters (transition probability matrix $p(j|i)$), the index assignment (π), and the mapping ($\mathbf{q}$) are known to the encoder and the decoder. The problem is to design $\mathbf{q}$ for the equilibrium, i.e., the encoder minimizes its distortion if used with a corresponding decoder that minimizes its own distortion. This communication setting is given in Figure 1. The quantizer $\mathbf{q}$ divides the set $\mathcal{X}$ into mutually exclusive and exhaustive sets as $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_M$, $\mathbf{q}(x) = z_m, x \in \mathcal{V}_m$.

In our recent paper [20], we use dynamic programming solution concept along with random index assignment for this problem (quantizing a scalar source X for a rate-constrained noisy channel with misaligned objectives for encoder and decoder). Here, we use gradient descent to find the quantizer Since we use gradient descent instead of dynamic programming, approximation of a continuous source by discretization is not required here. However, there is an issue of local optima which we address by using multiple initializations. The constraint on the average symbol error probability $0 < p_{err} < (M-1)/M$ in [20] is not enforced here, since unlike dynamic programming, gradient descent does not involve optimization of sub-problems.

III. Main Results

In this section, we present our results on the derivation of the distortions for communication over a noisy channel using random index mapping, and a gradient-descent based algorithm to compute the strategic quantizer.

A. Analysis

We make the following "monotonicity" assumption.

Assumption 1 (Convex code-cells)**.** *$\mathcal{V}_m$ is convex for all $m \in [1 : M]$.*

Remark 1. *Assumption 1 is the first of the two regularity conditions commonly employed in the classical quantization literature, cf. [13]. Note that the second regularity condition, $y_m \in \mathcal{V}_m$, is not included in Assumption 1.*

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

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

The encoder chooses a non-injective mapping, $Q : \mathcal{X} \to \mathcal{Z}$ which is the quantizer $\mathbf{q}$ with boundary levels $[x_0, x_1, \dots, x_M]$ to minimize its cost D_E

$$D_E = \sum_{m=1}^{M} \mathbb{E}_\pi\{\mathbb{E}\{\eta_E(x, \mathbf{y}^*(\mathbf{q}))|\pi, x \in \mathcal{V}_m\}\},$$

where the decoder determines a set of actions, $\mathbf{y}^*(\mathbf{q}) = [y_1, \dots, y_M]$ as the best response to $\mathbf{q}$ to minimize its cost D_D for $m \in [1 : M]$ as follows

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

The encoder designs a quantizer $\mathbf{q}$ using only the objectives ($\eta_s, s \in \{E, D\}$), the statistics of the source ($f(\cdot)$), the channel transition probability matrix ($p(j|i)$), and the index assigment (π) without the knowledge of the realization of X. After observing x, the encoder quantizes the source as

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

and uses an index mapping chosen uniformly at random,

$$z_i = \pi(z_m),$$

where $\pi : \{1, \ldots, M\} \to \{1, \ldots, M\}$. The message z_i is transmitted over a noisy channel and received as z_j with the channel transition probability $p(j|i)$. The decoder receives z_j and takes the action

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

The average symbol error probability of the channel is

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

Let $c_1 = p_{err}/(M-1), c_2 = 1 - Mc_1$. The end-to-end distortion given an index assignment π is

$$\mathbb{E}\{\eta_s|\pi\} = \sum_{i=1}^{M}\sum_{j=1}^{M}\int\limits_{x_{i-1}}^{x_i} \eta_s(x, y_j)p(j|i)f(x)\mathrm{d}x.$$

The average distortion over all possible index assignments is

$$\begin{aligned} D_s &= \sum_{i=1}^{M}\sum_{j=1}^{M}\int\limits_{x_{i-1}}^{x_i} \eta_s(x, y_j)\mathbb{E}_\pi\{p(j|i)\}f(x)\mathrm{d}x \\ &= I_{j\neq i} + I_{j=i}, \end{aligned}$$

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

$$\begin{aligned} I_{j\neq i} &= \sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M}\int\limits_{x_{i-1}}^{x_i} \eta_s(x, y_j)\mathbb{E}_\pi\{p(j|i)\}f(x)\mathrm{d}x \\ &= \sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M}\int\limits_{x_{i-1}}^{x_i} \eta_s(x, y_j)\frac{p_{err}}{M-1}f(x)\mathrm{d}x, \\ I_{j=i} &= \sum_{i=1}^{M}\int\limits_{x_{i-1}}^{x_i} \eta_s(x, y_i)\mathbb{E}_\pi\{p(i|i)\}f(x)\mathrm{d}x \\ &= \sum_{i=1}^{M}\int\limits_{x_{i-1}}^{x_i} \eta_s(x, y_i)(1-p_{err})f(x)\mathrm{d}x. \end{aligned}$$

$I_{j\neq i}$ can be further simplified as follows:

$$\begin{aligned} I_{j\neq i} &\overset{a}{=} c_1\sum_{i=1}^{M}\int\limits_{x_{i-1}}^{x_i}\left(\sum_{j=1}^{M}\eta_s(x, y_j) - \eta_s(x, y_i)\right)f(x)\mathrm{d}x \\ &\overset{b}{=} c_1\sum_{i=1}^{M}\left(\mathbb{E}\{\eta_s(x, y_i(\mathbf{q}))\} - \int\limits_{x_{i-1}}^{x_i}\eta_s(x, y_i(\mathbf{q}))f(x)\mathrm{d}x\right). \end{aligned}$$

In the above equation, (a) is obtained via adding and subtracting $\sum_{i=1}^{M}\int_{x_{i-1}}^{x_i} \eta_s(x, y_i(\mathbf{q}))f(x)\mathrm{d}x$, (b) follows from exchanging the summations over i and j, using

$$\mathbb{E}\{\eta_s(x, y_j(\mathbf{q}))\} = \sum_{i=1}^{M}\int\limits_{x_{i-1}}^{x_i}\eta_s(x, y_j(\mathbf{q}))f(x)\mathrm{d}x,$$

and changing the summation index of the first term to i. The average distortions and the optimum decoder reconstruction

$$D_s = c_1\sum_{i=1}^{M}\mathbb{E}\{\eta_s(x, y_i(\mathbf{q}))\} + c_2\overline{D_s}, \quad i \in [1:M], \quad (1)$$

$$y_i = \underset{y\in\mathcal{Y}}{\arg\min}\, c_1\mathbb{E}\{\eta_D(x, y)\} + c_2\int\limits_{x_{i-1}}^{x_i}\eta_D(x, y)f(x)\mathrm{d}x, \quad (2)$$

where $\overline{D_s}$ is the distortion in the noiseless setting

$$\overline{D_s} = \sum_{i=1}^{M}\int\limits_{x_{i-1}}^{x_i}\eta_s(x, y_i)f(x)\mathrm{d}x.$$

The actions $\mathbf{y}$ are found using the first-order KKT optimality condition $\partial D_D/\partial y_i = 0$. If the decoder distortion is mean squared error, i.e., $\eta_D(x, y) = (x-y)^2$, then

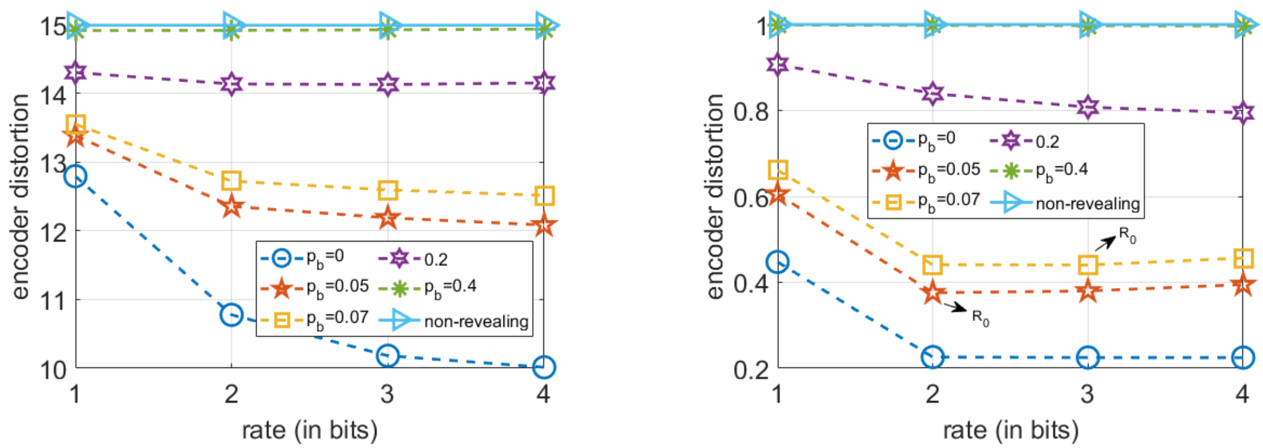
$$y_i = \frac{c_1\mathbb{E}\{X\} + c_2\int_{x_{i-1}}^{x_i} xf(x)\mathrm{d}x}{c_1 + c_2\int_{x_{i-1}}^{x_i} f(x)\mathrm{d}x}.$$

B. Gradient descent algorithm

We first note a significant research challenge associated with the design problem. The classical vector quantization design relies on the Lloyd-Max optimization, where the encoder and the decoder optimize their mappings iteratively. These iterations converge to a locally optimal solution because the distortion, identical for the decoder and the encoder (team problem), is nonincreasing with each iteration. However, here we consider a game problem (as opposed to a team problem) where the objectives are different. A strategic variation of these algorithms would enforce optimality with respect to a different distortion measure at each iteration, and hence do not converge as illustrated in detail in [1]. A natural optimization approach would be taking the functional gradient i.e., perturbing the quantizer mapping via an admissible perturbation function. However, the set of admissible functions have to be carefully chosen to satisfy the quantizer's properties (such as rate and convex codecell requirements) which hinders the tractability of this more general functional optimization approach. We note that the encoder's distortion is a function of the quantizer decision levels $(\mathbf{q})$ and quantizer representative levels which are a function of $\mathbf{q}$, $\mathbf{y}(\mathbf{q})$. This permits a gradient-descent based optimization as a function of $\mathbf{q}$ with encoder distortion as the objective.

Remark 2. *The method proposed inherits the convergence guarantees of gradient-descent based algorithms, thus ensuring local optimality. However, the resulting quantizer may not necessarily be globally optimal.*

The local optima issue can be resolved by techniques explored in literature [12], [21]–[23]. Here, we implement a simple remedy where we perform gradient descent with multiple initializations and choose the best local optimum



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

Fig. 2: Encoder distortion for a Gaussian source $X \sim \mathbb{N}(0,1)$ with $\eta_D(x,y) = (x-y)^2$ and for two different $\eta_E(x,y)$.

amongst them. The MATLAB codes are provided at https://tinyurl.com/asilomar2023 for research purposes. A sketch of the proposed method is given in Algorithm 1 below.

ALGORITHM 1

Proposed strategic quantizer design algorithm

Input: $f(\cdot), \mathcal{X}, M, \eta_E(\cdot,\cdot), \eta_D(\cdot,\cdot), p_b$
Output: $\mathbf{q}^*, \mathbf{y}^*, D_E, D_D$
Initialization: assign a monotone $\mathbf{q}$ randomly, compute associated encoder distortion $D_E(0)$, iteration index $i = 1$
Parameters: ϵ, λ, N
Compute symbol error probability $p_{err} \leftarrow 1 - (1 - p_b)^{log_2 M}$
while $\Delta D > \epsilon$ or $i < N$ **do**
    Compute the gradients, $\{\partial D_E / \partial x_m\}_i$
    Compute the updated quantizer $\mathbf{q}_{i+1}$ by gradient descent with the above gradients, $\mathbf{q}_{i+1} \triangleq \mathbf{q}_i - \lambda \{\partial D_E / \partial x_m\}_i$
    Compute actions $\mathbf{y}(\mathbf{q}_{i+1})$ with (2)
    Compute encoder distortion $D_E(i+1)$ associated with quantizer $\mathbf{q}_{i+1}$ and actions $\mathbf{y}(\mathbf{q}_{i+1})$ via (1), (2)
    Compute $\Delta D = D_E(i) - D_E(i+1)$
end while
return Quantizer $\mathbf{q}^* = \mathbf{q}_{i+1}$, actions $\mathbf{y}(\mathbf{q}^*)$, encoder and decoder distortions D_E and D_D computed for the optimal quantizer and decoder actions $\mathbf{q}^*, \mathbf{y}(\mathbf{q}^*)$ via 1.

IV. NUMERICAL RESULTS

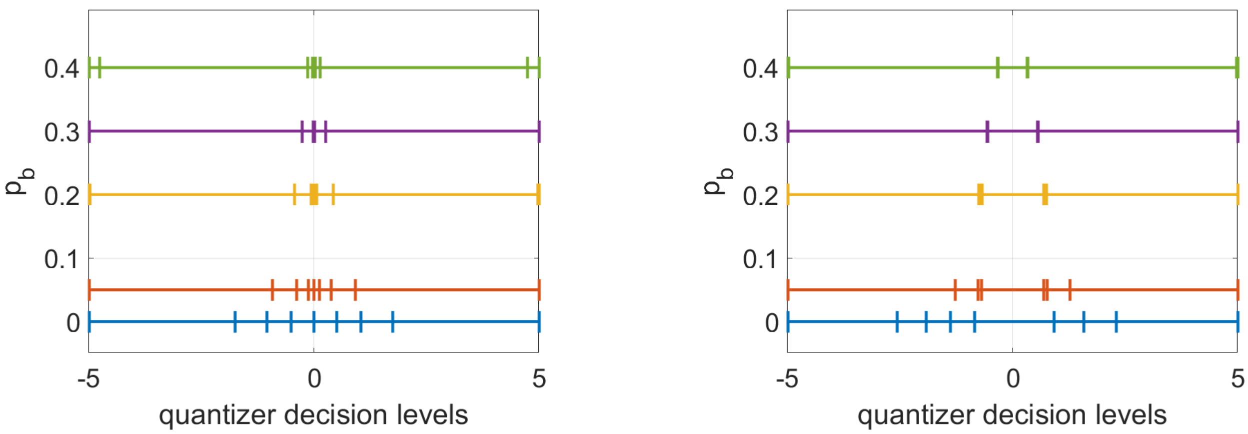
We present results for two settings, both with decoder distortion $\eta_D(x,y) = (x-y)^2$, for a Gaussian source $X \sim \mathbb{N}(0,1)$ with encoder distortions (i) $\eta_E(x,y) = (x^3 - y)^2$, and (ii) $\eta_E(x,y) = (x - y^3)^2$. We take the support of X as $[-5, 5]$ for computational ease.

In Figures 2a,b, we plot the encoder distortion associated with the above settings for bit error rates $p_b = [0, 0.05, 0.07, 0.2, 0.4]$, respectively. The encoder distortions increase with the bit error rate, as expected. The encoder distortions are smaller than the non-revealing encoder distortion, i.e., the encoder prefers to send a message for these parameters. However, we observe that as the bit error rate increases, the encoder distortion approaches non-revealing encoder distortion, which implies that ability of the encoder to utilize its informational advantage to "persuade" the decoder decreases with the bit error rate.

We observe in Fig. 2b for a given bit error rate p_b, beyond some rate, say R_0, increasing the rate does not decrease the encoder distortion. We refer to the threshold rate, R_0 as the cutoff rate, defined as the smallest R_0 for which $R > R_0$ implies that $D_E(p_b, R) \geq D_E(p_b, R_0)$, where $D_E(p_b, R)$ is the encoder distortion at rate R with bit error rate p_b. Depending on how the encoder distortion aligns with decoder distortion, the cutoff rate may increase (more aligned) or decrease (less aligned) with the bit error rate [20]. In Fig. 2b, we observe that the cutoff rate R_0 increases with the bit error rate.

In Figures 3a,b, we plot quantizers for strategic ($\eta_E(x,y) = (x^3-y)^2, \eta_D(x,y) = (x-y)^2$) and non-strategic ($\eta_E(x,y) = \eta_D(x,y) = (x-y)^2$) cases for various bit error rate values. We observe that the quantization intervals are finer around the mean for the non-strategic setting. This is in contrast with the strategic case where we observe that the encoder prefers not to disclose much information, i.e., the quantization intervals are relatively sparse around the mean (where most of the probability measure is contained).

We note that our analysis is not limited to the specific distortion measures that we use to illustrate the numerical results. We take the decoder's distortion measure as the conventionally used MSE metric, $\eta_D(x,y) = (x-y)^2$, for numerical results. The choice of distortion measure for the encoder (η_E) is arbitrary, ($\eta_E \neq \eta_D$) due to the problem formulation. However, certain choices of distortion measures can lead to less interesting solutions, such as non-revealing (where the encoder does not transmit any information) or



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

Fig. 3: Quantizers for $X \sim \mathbb{N}(0, 1)$ with $\eta_D = (x - y)^2$ for non-strategic ($\eta_E(x, y) = (x - y)^2$) and strategic encoder ($\eta_E(x, y) = (x^3 - y)^2$).

fully-revealing (where the problem simplifies to non-strategic quantization with the decoder distortion measure).

V. Conclusion

In this paper, we analyzed the problem of strategic quantization over a noisy channel. For our design method, we implemented a gradient-descent based algorithm based on our prior work for the noiseless setting [1]. The obtained numerical results confirm our theoretical analysis.

Gradient-descent based algorithms converge to a local optimum, which may not be the globally optimal solution. As a simple remedy, we used multiple initializations and chose the best solution among them. Global optimality can be achieved by dynamic-programming based algorithms, as done in our prior work [17], [20], however at the cost of increased complexity.

References

[1] E. Akyol and A. Anand, "Strategic Quantization," in *IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 543–548.

[2] L. Rayo and I. Segal, "Optimal Information Disclosure," *Journal of Political Economy*, vol. 118, no. 5, pp. 949–987, 2010.

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

[4] E. Kamenica, "Bayesian Persuasion and Information Design," *Annual Review of Economics*, vol. 11, pp. 249–272, 2019.

[5] D. Bergemann and S. Morris, "Information Design: A Unified Perspective," *Journal of Economic Literature*, vol. 57, no. 1, pp. 44–95, 2019.

[6] V. P. Crawford and J. Sobel, "Strategic Information Transmission," *Econometrica: Journal of the Econometric Society*, pp. 1431–1451, 1982.

[7] H. Kono and M. Kandori, "Corrigendum to Crawford and Sobel (1982) "Strategic Information Transmission"," *Econometrica*, vol. 89, no. 4, pp. 1–10, 2021.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] ——, "Strategic Quantization of a Noisy Source," in *59th Annual Allerton Conference on Communication, Control, and Computing*, 2023, pp. 1–7.

[19] R. Dobrushin and B. Tsybakov, "Information Transmission with Additional Noise," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 293–304, 1962.

[20] A. Anand and E. Akyol, "Channel-Optimized Strategic Quantizer Design via Dynamic Programming," in *2023 IEEE Statistical Signal Processing Workshop (SSP)*, 2023, pp. 621–625.

[21] K. Rose, "Deterministic Annealing for Clustering, Compression, Classification, Regression, and Related Optimization Problems," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2210–2239, 1998.

[22] A. Gamal, L. Hemachandra, I. Shperling, and V. Wei, "Using Simulated Annealing to Design Good Codes," *IEEE Transactions on Information Theory*, vol. 33, no. 1, pp. 116–123, 1987.

[23] D. Bertsimas and J. Tsitsiklis, "Simulated Annealing," *Statistical Science*, vol. 8, no. 1, pp. 10 – 15, 1993. [Online]. Available: https://doi.org/10.1214/ss/1177011077