Quadratic Strategic Quantization

Anju Anand, *Student Member, IEEE* and Emrah Akyol, *Senior Member, IEEE*

***Abstract*—This paper is concerned with the quantization setting where the encoder and the decoder have misaligned quadratic objectives. Since the problem setting involves the encoder (leader) designing the quantizer decision levels first, followed by the decoder (follower) designing the actions as a function of the decision levels, we look for a Stackelberg equilibrium. The solution concept is a gradient descent based algorithm. We extend the problem to communication over a noisy channel by employing a random channel index mapping, as done in prior work on classical channel-optimized quantizer design literature, combined with a gradient descent based approach to optimize quantization boundaries. We finally present numerical results obtained via the proposed algorithms that suggest the proposed algorithms' validity and demonstrate strategic quantization features that differentiate it from its classical counterpart.**

***Index Terms*—**

I. Introduction

CONSIDER the following problem: an encoder observes a realization of source $X \in \mathcal{X}$ with a probability distribution μ and sends a message $Z \in \mathcal{Z}$ using an injective mapping $Q : \mathcal{X} \to \mathcal{Z}$, with $|\mathcal{Z}| \leq M$. After receiving the message Z, the decoder takes action $Y \in \mathcal{Y}$. The costs that the encoder and the decoder minimize are $D_E \triangleq \eta_E(X,Y)$ and $D_D \triangleq \mathbb{E}\{\eta_D(X,Y)\}$, with $\eta_E \neq \eta_D$ (misaligned objectives). The encoder designs Q *ex-ante*, i.e., without the knowledge of the realization of X, using only the functions η_E and η_D, and the statistics of the source, $\mu(\cdot)$. The functions (η_E and η_D), the shared prior (μ), and the mapping (Q) are known to the encoder and the decoder. The problem is to design Q. We call this setup *strategic quantization*.

This paper is concerned with the quantizer design problem for the setting where two agents (the encoder and the decoder) with misaligned quadratic objectives communicate over a noiseless and a noisy channel. The classical (non-strategic) counterpart of this problem, i.e., channel-optimized quantization, has 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.

The setting without the quantization aspect (in the practical sense, if M is asymptotically large) is known in the Economics literature as the information design, or the Bayesian persuasion problem [10], [12]. These problems analyze how a communication system designer (sender) can use the information to influence the action taken by a receiver. This framework has proven beneficial in analyzing a variety of real-life applications, such as the design of transcripts when schools compete to improve their students' job prospects [13] and voter mobilization and gerrymandering [14], as well as various engineering applications, including in modeling misinformation spread over social networks [15] and privacy-constrained information processing [16], and many more [9]. For an excellent survey of the related literature in Economics, see [17], [18].

The strategic quantization problem, as described above, was discussed in a few contemporary economics and computer science studies. In [19], authors analyze the problem via a computation lens and report approximate results on this problem, relating to another problem they solved conclusively. In one of their main results, the algorithmic complexity of finding the optimum strategic quantizer was shown to be NP-hard. In a recent working paper, Aybaş and Türkel [20] analyzed this problem using the methods in [10] and provided several theoretical properties of strategic quantization. A byproduct of their analysis yields a constructive method for deriving optimal quantizers based on a search over possible posterior distributions over their feasible set. Our objective here is to leverage the rich collection of results in quantization theory, e.g., the comprehensive survey of results by Gray and Neuhoff [21], to study the same problem via the engineering lens.

We note in passing that quantizers also arise as equilibrium strategies endogenously, i.e., without an external constraint, in a related but distinctly different class of signaling games, namely the cheap talk [22]. In [22], the encoder chooses the mapping from the realization of the source X to message Z *after* observing it, *ex-post*, as different source realizations indicate optimality of different mappings for the encoder. The encoder's lack of commitment power in the cheap talk setting makes the notion of equilibrium a Nash equilibrium since both agents form a strategy that is the best response to each other's mapping. However, in our strategic quantization problem (and the information design problems in general as in [10], [12]), the encoder designs Q *ex-ante*, *before* seeing the source realization, and committed to the designed Q afterward. This commitment is known to the decoder and establishes a form of trust between the sender and the receiver, resulting in possibly higher payoffs for both agents. This difference also manifests itself in the notion of equilibrium we are seeking here since the encoder does not necessarily form the best response to the decoder due to its commitment to Q[1].

The problem setting has several applications in engineering as well as Economics. For an engineering application, consider

Authors are with the Binghamton University–SUNY, Binghamton, NY, 13902 USA {aanand6, eakyol}@binghamton.edu. This research is supported by the NSF via grants CCF #1910715 and CAREER #2048042.

[1]These issues are well understood in the Economics literature, see, e.g., [23] for an excellent survey. However, we emphasize them here for a reader with an engineering background; see [9] for a detailed discussion through the engineering lens.

the Internet of Things, where agents with misaligned objectives communicate over channels with delay constraints. For a more concrete, real-life application, consider two smart cars by competing manufacturers, e.g., Tesla and Honda, where the Tesla (decoder) car asks for a piece of specific information, such as traffic congestion, from the Honda (encoder) to decide on changing its route or not. Say Honda's objective is to make Tesla take a specific action, e.g., to change its route, while Tesla's objective is to estimate the congestion to make the right decision accurately. Honda's objective is obviously different from that of Tesla, hence has no incentive to convey a truthful congestion estimate. However, Tesla is aware of Honda's motives while still would like to use Honda's information (if possible). With the realistic assumption that Honda would observe this information through a noisy sensing channel (i.e., a sensor), how would these cars communicate over a fixed-rate zero-delay channel? Such problems can be handled using our model. Note that here Honda has three different behavioral choices: it can choose not to communication (non-revealing strategy), can communicate exactly what the Tesla wants (fully-revealing strategy), or it can craft a message that would make Tesla to change its route. Note that Tesla can choose not to 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 formidable research challenge.

II. PRELIMINARIES

A. Notation

In this paper, random variables are denoted using capital letters (say X), their sample values with respective lower case letters (x), and their alphabet with respective calligraphic letters ($\mathcal{X}$). The set of real numbers, and non-negative integers are denoted by $\mathbb{R}$, and $\mathbb{Z}_{\geq 0}$, respectively. The alphabet, $\mathcal{X}$, can be finite, infinite, or a continuum, like an interval $[a, b] \subset \mathbb{R}$. The expectation operator is written as $\mathbb{E}\{\cdot\}$. The operator $|\cdot|$ denotes the absolute value if the argument is a scalar real number and the cardinality if the argument is a set. The uniform distribution over an interval $[t_1, t_2]$, the scalar Gaussian with mean t, variance σ^2, and the 2-dimensional jointly Gaussian distribution with mean $\begin{bmatrix} t_1 & t_2 \end{bmatrix}'$ and respective variances σ_1^2, σ_2^2 with a correlation ρ are denoted by $U[t_1, t_2]$, $\mathcal{N}(t, \sigma^2)$, and $\mathcal{N}\left(\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix}\right), 0 \leq \rho < 1$ respectively. The expression $t_1 \leq i \leq t_2, i \in \mathbb{Z}_{\geq 0}$ is denoted by $i \in [t_1 : t_2]$.

B. Strategic Quantization Prior Work

The strategic quantization problem can be described as follows: the encoder observes a signal $X \in \mathcal{X}$, and sends a message $Z \in \mathcal{Z}$ to the decoder, upon receiving which the decoder takes the action $Y \in \mathcal{Y}$. The encoder designs the quantizer decision levels Q to minimize its objective D_E, while the decoder designs the quantizer representative levels $\mathbf{y}$ to minimize its objective D_D. Note that the objectives of the encoder and the decoder are misaligned ($D_E \neq D_D$). The strategic quantizer is a mapping $Q : \mathcal{X} \rightarrow \mathcal{Z}$, with $|\mathcal{Z}| \leq M$ for a given quantization resolution $M \in \mathbb{Z}^+$, and given distortion measures D_E, D_D.

As mentioned earlier, our problem is a variation of the Bayesian Persuasion (or information design) class of problems where the encoder and the decoder with misaligned objectives communicate [10]. This class of the problems have been an active research area in Economics due to their modeling abilities of real-life scenarios, see e.g., [12], [13], [18], [23].

This problem was previously studied in Economics as well as Computer Science. In [20], authors showed the existence of optimal strategic quantizers in abstract spaces. Moreover, authors provide a low-complexity method to obtain the optimal strategic quantizer. In [24], [25], authors characterize sufficient conditions for the monotonicity of the optimal strategic quantizer, and as a byproduct of their analysis, characterize its behavior (non-revealing, fully revealing, or partially revealing) for some special settings. In Computer Science, in []

In [26], we showed that a strategic variation of the Lloyd-Max algorithm does not converge to a locally optimal solution. As a remedy we develop a gradient descent based solution for this problem. We also demonstrated that even for well-behaving sources, such as scalar Gaussian, there are multiple local optima, depending on the distortion measures chosen, in sharp contrast with the classical quantization for which the local optima is unique for the case of log-concave sources (which include Gaussian sources). We also analyzed the behavior of the optimal strategic quantizer for some typical settings. The behavior can be one of the following three:

1) *Non-revealing*: the encoder does not send any information, i.e., $Q(X) =$ constant.
2) *Fully revealing*: the encoder effectively sends the information the decoder asks, which simplifies the problem into classical quantizer design with the decoder's objective.
3) *Partially revealing*: The encoder sends some information but not exactly the decoder wants.

In [27], [28], we carried out our analysis of strategic quantization to the scenario where there is a noisy communication channel between the encoder and the decoder, using random index mapping in conjunction with gradient descent based and dynamic programming solutions respectively. In [11], [28], we derived the globally optimal strategic quantizer via a dynamic programming based solution to resolve the poor local minima issues with gradient descent based solutions since the objective function is non-convex.

C. Gradient descent based Quantization Prior Work

As mentioned earlier, this problem is well-studied in the classical, i.e., non-strategic, compression literature,

III. PROBLEM FORMULATION

Consider the following quantization problem: an encoder observes realizations of the two sources $X \in \mathcal{X}, \theta \in \mathcal{T}$ with joint probability distribution $(X, \theta) \sim \mu_{X,\theta}$, and maps (X, θ) 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 an onto mapping parameterized by θ, $q_\theta : \mathcal{X} \rightarrow \mathcal{Z}$. After receiving the message

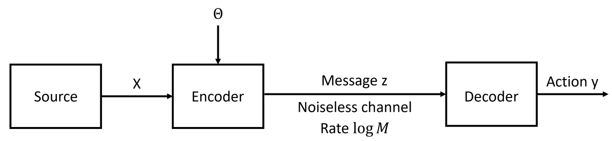


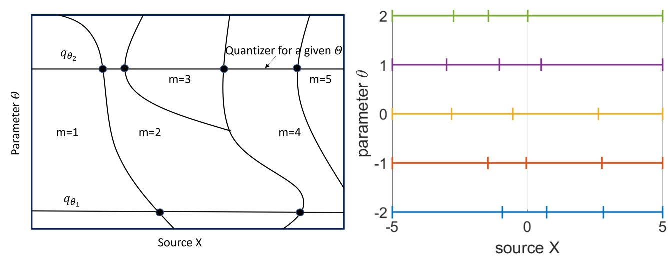
Fig. 1: Communication diagram: (X, θ) over a noiseless channel

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 decoder minimize their respective objectives $\overline{D_E} = \mathbb{E}\{\eta_E(X, \theta, Y)\}$ and $\overline{D_D} = \mathbb{E}_D\{\eta_D(X, \theta, Y)\}$, which are misaligned ($\eta_E \neq \eta_D$). The encoder designs $\mathbf{q} = \{q_\theta, \theta \in \mathcal{T}\}$ *ex-ante*, i.e., without the knowledge of the realization of (X, θ), using only the objectives η_E and η_D, and the statistics of the source $\mu_{X,\theta}(\cdot, \cdot)$. The objectives ($\eta_E$ and η_D), the shared prior (μ), 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. Since the encoder chooses the quantization decision levels $\mathbf{q}$ first, followed by the decoder choosing the quantization representative levels ($\mathbf{y}$), we look for a Stackelberg equilibrium.

Note that implementing a $2:1$ quantizer from $\mathbf{q} : (\mathcal{X}, \mathcal{T}) \to \mathcal{Z}$ can be simplified to computing a set of quantizers corresponding to each $\theta \in \mathcal{T}$ as in Figure 2 without loss of generality. If the quantizer does not include a region m for some realization of θ, the encoder never sends the message m i.e., the encoder chooses a lower rate and is less revealing for that value of θ. In Figure 2a, we see that the quantizer q_{θ_1} only includes $m = 1, 2, 4$ regions, while the quantizer q_{θ_2} contains all five regions. Figure 2b shows an example quantizer obtained as a numerical result.

This problem setting is extended to a noisy channel with channel transition probability $p(j|i)$ using random index assignment in conjunction with gradient descent. Random index assignment has been used in prior literature [7], [8] to make the design less dependent on the specific channel settings and be less computationally complexity. 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 $p(j|i) = p(z_j|z_i)$. 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 $D_E = \mathbb{E}_\pi\{\mathbb{E}\{\eta_E(X, \theta, Y)|\pi\}\}$ and $D_D = \mathbb{E}_\pi\{\mathbb{E}\{\eta_D(X, \theta, Y)|\pi\}\}$, which are misaligned ($\eta_E \neq \eta_D$). In addition to the above mentioned parameters that are common knowledge, the index assignment (π), and the channel transition probability matrix ($p(z_j|z_i)$) are also known by the encoder and the decoder. This communication setting is shown in Figure 3. In this paper, we consider the encoder's distortion as $\eta_E = (x + \theta - y)^2$ and an MSE decoder $\eta_D = (x - y)^2$. Building on [26] and the draft paper [27], where we study strategic quantizer design for distortion functions in the form of $1 : 1$ mapping (see also a recent related work [11], [28]), in this paper we analyze and design strategic quantizer for distortion functions in the form of a $2 : 1$ mapping. We analyse the communication over both noiseless and noisy channel settings. Our main computational design tool here is gradient descent, which is used in conjunction with random index assignment for the noisy channel. Compared to our recent related work on gradient descent based strategic quantization, the contribution of this paper is the extension of the analysis from scalar settings to $2 : 1$ distortion settings via a subterfuge by designing a set of quantizers. We note that in [28], we assume that $0 < p_{err} < \frac{M-1}{M}$ so that $c_1, c_2 > 0$ since dynamic programming divides the optimization problem into sub problems, however here we do not need to assume the same since we use a gradient descent based solution which optimizies the objective function as a whole. Since the objective function is non-convex, we handle the issue of local optima by using multiple initializations.



(a) M=5 level quantizer (b) Example quantizer

Fig. 2: Quantization of X parameterized by θ.

IV. Main Results

A. Analysis

Let X take values from the source alphabet $\mathcal{X} \subseteq [a_X, b_X]$, and θ take values from $\mathcal{T} \subseteq [a_\theta, b_\theta]$. The joint probability distribution function of (X, θ) is $\mu_{X,\theta}$. The set $\mathcal{X}$ is divided into mutually exclusive and exhaustive sets parameterized by the realization of θ, $\mathcal{V}_{\theta,1}, \mathcal{V}_{\theta,2}, \ldots, \mathcal{V}_{\theta,M}$.

We make the following "monotonicity" assumption.

Assumption 1 (Convex code-cells): $\mathcal{V}_{\theta,i}$ is convex for all $\theta \in \mathcal{T}, i \in [1 : M]$.

Under assumption 1, $\mathcal{V}_{\theta,i}$ is an interval since X is a scalar, i.e.,

$$\mathcal{V}_{\theta,i} = [x_{\theta,i-1}, x_{\theta,i}).$$

The integrals expressed throughout this paper are defined over the set $\mathcal{V}_{\theta,i}$ which is omitted for brevity, unless specified otherwise.

We consider the encoder and decoder objectives as $\eta_E = (x+\theta-y)^2, \eta_D = (x-y)^2$, respectively. The encoder chooses the set of quantizers $\mathbf{q} = \{q_\theta, \theta \in \mathcal{T}\}$ with boundary levels $q_\theta = [x_{\theta,0}, x_{\theta,1}, \ldots, x_{\theta,M}]$ for each $\theta \in \mathcal{T}$ to minimize its distortion,

$$\overline{D_E} = \sum_{i=1}^{M} \int_{\theta \in \mathcal{T}} \mathbb{E}\{\eta_E(x, \theta, y_i(\mathbf{q}))|x \in \mathcal{V}_{\theta,i}\}, \quad (1)$$

where y_i^* are determined by the decoder as the best response to $\mathbf{q}$ to minimize its cost $\overline{D_D} = \mathbb{E}\{\eta_D(X,\theta,\mathbf{y})\}$ for $i \in [1:M]$ as follows

$$y_i^* = \underset{y_i\in\mathcal{Y}}{\arg\min}\sum_{i=1}^{M}\mathbb{E}\{(x-y)^2|x\in\mathcal{V}_{:,i}\},$$

where $\mathcal{V}_{:,i} = \{\mathcal{V}_{\theta,i}, \forall\theta\in\mathcal{T}\}$. For MSE decoder, $\eta_D = (x-y)^2$, and the optimal reconstruction points are given by

$$y_i^* = \frac{\int\limits_{\theta\in\mathcal{T}}\int x\mathrm{d}\mu_{X,\theta}}{\int\limits_{\theta\in\mathcal{T}}\int \mathrm{d}\mu_{X,\theta}}. \tag{2}$$

The decoder determines a single set of actions $\mathbf{y}$ since it is unaware of the realization of θ.

The gradient of the encoder's distortion with respect to the decision levels for $\eta_E = (x+\theta-y)^2, \eta_D = (x-y)^2$ are

$$\begin{aligned}\frac{\partial\overline{D_E}}{\partial x_{\theta,m}} &= (x_{\theta,m}+\theta-y_m)^2\frac{\mathrm{d}\mu_{X,\theta}}{\mathrm{d}x\mathrm{d}\theta}\\ &\quad-(x_{\theta,m}+\theta-y_{m+1})^2\frac{\mathrm{d}\mu_{X,\theta}}{\mathrm{d}x\mathrm{d}\theta}\\ &\quad-2\frac{\mathrm{d}y_m}{\mathrm{d}x_{\theta,m}}\int\limits_{\theta\in\mathcal{T}}\int(x+\theta-y_m)\mathrm{d}\mu_{X,\theta}\\ &\quad-2\frac{\mathrm{d}y_{m+1}}{\mathrm{d}x_{\theta,m}}\int\limits_{\theta\in\mathcal{T}}\int\limits_{x_{\theta,m-1}}^{x_{\theta,m}}(x+\theta-y_{m+1})\mathrm{d}\mu_{X,\theta},\end{aligned} \tag{3}$$

where

$$\frac{\mathrm{d}y_m}{\mathrm{d}x_{\theta,m}} = \frac{\mathrm{d}\mu_{X,\theta}(x_{\theta,m},\theta)}{\mathrm{d}x\mathrm{d}\theta}\frac{x_{\theta,m}-y_m}{\int\limits_{\theta\in\mathcal{T}}\int\mathrm{d}\mu_{X,\theta}}, \tag{4}$$

$$\frac{\mathrm{d}y_{m+1}}{\mathrm{d}x_{\theta,m}} = -\frac{\mathrm{d}\mu_{X,\theta}(x_{\theta,m},\theta)}{\mathrm{d}x\mathrm{d}\theta}\frac{x_{\theta,m}-y_m}{\int\limits_{\theta\in\mathcal{T}}\int\limits_{x_{\theta,m-1}}^{x_{\theta,m}}\mathrm{d}\mu_{X,\theta}}. \tag{5}$$

V. Noisy channel

We extend the above analysis to communication over a noisy channel with channel parameters (transition probability matrix $p(j|i)$) using random index assignment in conjunction with gradient descent.

An index mapping $\pi : [1:M] \to [1:M]$ is chosen uniformly at random and is applied to the message Z. The message $\pi(Z) \in \mathcal{Z}$ is transmitted over a noisy channel with transition probability matrix $p(j|i)$. After receiving the message $Z' \in \mathcal{Z}$, the decoder applies a mapping $\phi : \mathcal{Z} \to \mathcal{Y}$, where $|\mathcal{Y}| = |\mathcal{Z}|$, on the message Z' and takes an action $Y = \phi(Z')$ (which includes the inverse mapping π^{-1} applied on Z' first).

The encoder minimizes

$$D_E = \sum_{i=1}^{M}\int\limits_{\theta\in\mathcal{T}}\mathbb{E}_\pi\{\mathbb{E}\{(x+\theta-y_i)^2|\pi, x\in\mathcal{V}_{\theta,i}\}\} \tag{6}$$

with the choice of a set of quantizers $\mathbf{q} = \{q_\theta, \theta\in\mathcal{T}\}$ with boundary levels $q_\theta = [x_{\theta,0}, x_{\theta,1}, \ldots, x_{\theta,M}]$ for each $\theta\in\mathcal{T}$. The decoder determines a single set of actions since it is unaware of the realization of θ, $\mathbf{y} = [y_1,\ldots,y_M]$, as the best

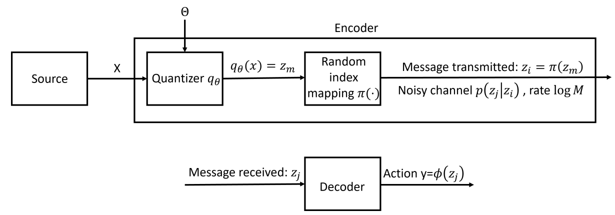


Fig. 3: Communication diagram: (X,θ) over a noisy channel

response to $\mathbf{q}$ to minimize its cost $D_D = \mathbb{E}_\pi\{\mathbb{E}\{(x-y)^2)|\pi\}\}$ for $i\in[1:M]$ as follows

$$y_i^* = \underset{y_i\in\mathcal{Y}}{\arg\min}\sum_{i=1}^{M}\mathbb{E}_\pi\{\mathbb{E}\{\eta_D(x,\theta,\mathbf{y})|\pi, x\in\mathcal{V}_{:,i}\}\},$$

where $\mathcal{V}_{:,i} = \{\mathcal{V}_{\theta,i}, \forall\theta\in\mathcal{T}\}$. The integrals expressed throughout this paper are defined over the set $\mathcal{V}_{\theta,i}$ which is omitted for brevity, unless specified otherwise.

The encoder designs a set of quantizers $\mathbf{q}$ with random index mapping using only the objectives (η_E, η_D), the statistics of the source $(\mu_{X,\theta}(\cdot,\cdot))$, the channel transition probability matrix $(p(j|i))$, and the index mapping (π) without the knowledge of the realization of (X,θ). After observing (x,θ), the encoder chooses the quantizer q_θ, quantizes the source as

$$z_m = q_\theta(x), \quad x\in\mathcal{V}_{\theta,m},$$

and uses random index mapping to map z_m to z_i

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

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 probability $p(j|i)$. The decoder receives the message 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 $c_1 = \frac{p_{err}}{M-1}, c_2 = 1-Mc_1$. The probability that the receiver receives the noisy message $\hat{z} = z_j$ if z_i was transmitted using $\pi(q_\theta(x)) = z_i$ is given by $p(j|i)$, the channel transition probability.

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

$$\mathbb{E}\{\eta_s|\pi\} = \sum_{i=1}^{M}\sum_{j=1}^{M}\int\limits_{\theta\in\mathcal{T}}\int\eta_s(x,\theta,y_j(\mathbf{q}))p(j|i)\mathrm{d}\mu_{x,\theta}.$$

The average distortion over all possible index assignments is

$$\begin{aligned}D_s &= \mathbb{E}_\pi\{\mathbb{E}\{\eta_s|\pi\}\}\\ &= \sum_{i=1}^{M}\sum_{j=1}^{M}\int\limits_{\theta\in\mathcal{T}}\int\eta_s(x,\theta,y_j(\mathbf{q}))\mathbb{E}_\pi\{p(j|i)\}\mathrm{d}\mu_{x,\theta}\\ &= I_{j\neq i} + I_{j=i},\end{aligned}$$

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

$$\begin{aligned}
I_{j\neq i} &= \sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M}\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_j(\mathbf{q}))\mathbb{E}_\pi\{p(j|i)\}\mathrm{d}\mu_{x,\theta}\\
&= \sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M}\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_j(\mathbf{q}))\frac{p_{err}}{M-1}\mathrm{d}\mu_{x,\theta},\\
I_{j=i} &= \sum_{i=1}^{M}\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_i(\mathbf{q}))\mathbb{E}_\pi\{p(j|i)\}\mathrm{d}\mu_{x,\theta}\\
&= \sum_{i=1}^{M}\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_i(\mathbf{q}))(1-p_{err})\mathrm{d}\mu_{X,\theta}.
\end{aligned}$$

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

$$\begin{aligned}
I_{j\neq i} &= c_1\Bigg(\sum_{i=1}^{M}\sum_{j=1}^{M}\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_j(\mathbf{q}))\mathrm{d}\mu_{X,\theta}\\
&\quad -\sum_{i=1}^{M}\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_i(\mathbf{q}))\mathrm{d}\mu_{X,\theta}\Bigg)\\
&= c_1\Bigg(\sum_{j=1}^{M}\sum_{i=1}^{M}\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_j(\mathbf{q}))\mathrm{d}\mu_{X,\theta}\\
&\quad -\sum_{i=1}^{M}\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_i(\mathbf{q}))\mathrm{d}\mu_{X,\theta}\Bigg)\\
&\overset{a}{=} c_1\Bigg(\sum_{j=1}^{M}\mathbb{E}\{\eta_s(x,\theta,y_j(\mathbf{q}))\}\\
&\quad -\sum_{i=1}^{M}\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_i(\mathbf{q}))\mathrm{d}\mu_{X,\theta}\Bigg)\\
&= c_1\sum_{i=1}^{M}\Bigg(\mathbb{E}\{\eta_s(x,\theta,y_i(\mathbf{q}))\}\\
&\quad -\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_i(\mathbf{q}))\mathrm{d}\mu_{X,\theta}\Bigg).
\end{aligned}$$

where equality a in the above equation is obtained by $\mathbb{E}\{\eta_s(x,\theta,y_j(\mathbf{q}))\} = \int\limits_{\theta\in\mathcal{T}}\int\limits_{a_X}^{b_X}\eta_s(x,\theta,y_j(\mathbf{q}))\mathrm{d}\mu_{X,\theta} = \sum_{i=1}^{M}\int\limits_{\theta\in\mathcal{T}}\int\eta_s(x,\theta,y_j(\mathbf{q}))\mathrm{d}\mu_{X,\theta}$. Then, the average distortion and the optimum decoder reconstruction for $i\in[1:M]$ are

$$D_s = c_1\sum_{i=1}^{M}\Bigg(\mathbb{E}\{\eta_s(x,\theta,y_i(\mathbf{q}))\} \tag{7}$$

$$+ c_2\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_i(\mathbf{q}))\mathrm{d}\mu_{X,\theta}\Bigg), \tag{8}$$

$$y_i = \underset{y\in\mathcal{Y}}{\arg\min}\, c_1(\mathbb{E}\{\eta_s(x,\theta,y_i(\mathbf{q}))\} \tag{9}$$

$$+ c_2\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,\theta,y_i(\mathbf{q}))\mathrm{d}\mu_{X,\theta}. \tag{10}$$

The above terms D_s and y_i can be written in terms of the distortion in the noiseless setting as

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

$$y_i = \underset{y\in\mathcal{Y}}{\arg\min}\, c_1\mathbb{E}\{\eta_s(x,\theta,y)\} + c_2\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,y,\theta)\mathrm{d}\mu_{X,\theta},$$

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

$$\overline{D_s} = \sum_{i=1}^{M}\int\limits_{\theta\in\mathcal{T}}\int \eta_s(x,y_i,\theta)\mathrm{d}\mu_{X,\theta}. \tag{11}$$

The reconstruction levels $\mathbf{y}$ are found using the KKT conditions of optimality:

$$\frac{\partial D_D}{\partial y_i} = c_1\frac{\partial}{\partial y_i}\mathbb{E}\{\eta_D(x,\theta,y_i(\mathbf{q}))\} + c_2\frac{\partial}{\partial y_i}\overline{D_D}. \tag{12}$$

When the decoder distortion is MMSE, $\eta_D(x,\theta,y)=(x-y)^2$,

$$\frac{\partial D_D}{\partial y_i} = -2c_1\int\limits_{\theta\in\mathcal{T}}\int\limits_{a_X}^{b_X}(x-y_i)\mathrm{d}\mu_{X,\theta} - 2c_2\int\limits_{\theta\in\mathcal{T}}\int(x-y_i)\mathrm{d}\mu_{X,\theta},$$

$$y_i = \frac{c_1\mathbb{E}\{X\} + c_2\int\limits_{\theta\in\mathcal{T}}\int x\mathrm{d}\mu_{X,\theta}}{c_1 + c_2\int\limits_{\theta\in\mathcal{T}}\int \mathrm{d}\mu_{X,\theta}},$$

similar to the expression in [8], [28].

The gradient of the encoder's distortion with respect to the decision levels are

$$\frac{\partial D_E}{\partial x_{\theta,i}} = c_1\frac{\partial}{\partial x_{\theta,i}}\sum_{j=1}^{M}(\mathbb{E}\{\eta_E(x,\theta,y_j(\mathbf{q}))\}) + c_2\frac{\partial}{\partial x_{\theta,i}}\overline{D_E}, \tag{13}$$

where $\frac{\partial\overline{D_E}}{\partial x_{\theta,i}}$ is given in Equation 3 for $\eta_E=(x+\theta-y)^2$ and

$$\frac{\partial}{\partial x_{\theta,i}}\mathbb{E}\{\eta_E(x,\theta,y_i(\mathbf{q}))\} = \frac{\partial}{\partial x_{\theta,i}}\int\limits_{\theta\in\mathcal{T}}\int\limits_{a_X}^{b_X}\eta_E(x,\theta,y_i(\mathbf{q}))\mathrm{d}\mu_{X,\theta}. \tag{14}$$

For $\eta_E=(x+\theta-y)^2, \eta_D=(x-y)^2$, this term evaluates to

$$\begin{aligned}
I_1 &= \frac{\partial}{\partial x_{\theta,i}}\sum_{j=1}^{M}\mathbb{E}\{(x+\theta-y_j)^2\}\\
&= \frac{\partial}{\partial x_{\theta,i}}\mathbb{E}\{(x+\theta-y_i)^2\} + \mathbb{E}\{(x+\theta-y_{i+1})^2\}\\
&= -2\frac{\partial y_i}{\partial x_{\theta,i}}\int\limits_{\theta\in\mathcal{T}}\int(x+\theta-y_i) \qquad (15)\\
&\quad -2\frac{\partial y_{i+1}}{\partial x_{\theta,i}}\int\limits_{\theta\in\mathcal{T}}\int\limits_{x_{\theta,i}}^{x_{\theta,i+1}}(x+\theta-y_{i+1}), \qquad (16)
\end{aligned}$$

where

$$\frac{\mathrm{d}y_i}{\mathrm{d}x_{\theta,i}} = c_2 \frac{x_{\theta,i} - y_i}{(c_1 + c_2 \int\limits_{\theta \in \mathcal{T}} \int \mathrm{d}\mu_{X,\theta})} \frac{\mathrm{d}\mu_{X,\theta}(x_{\theta,i}, \theta)}{\mathrm{d}x\mathrm{d}\theta},$$

$$\frac{\mathrm{d}y_{i+1}}{\mathrm{d}x_{\theta,m}} = -c_2 \frac{x_{\theta,i} - y_{i+1}}{(c_1 + c_2 \int\limits_{\theta \in \mathcal{T}} \int\limits_{x_{\theta,i}}^{x_{\theta,i+1}} \mathrm{d}\mu_{X,\theta})} \frac{\mathrm{d}\mu_{X,\theta}(x_{\theta,i}, \theta)}{\mathrm{d}x\mathrm{d}\theta}.$$

A. Deterministic θ

We simplify the above analysis to a 1-dimensional source X with a probability distribution μ_X by considering a deterministic θ. The encoder minimizes D_E with the choice of $\mathbf{q} = [x_0, x_1, \ldots, x_{M+1}]$. The decoder minimizes D_D with the choice of $\mathbf{y} = [y_1, \ldots, y_M]$. Then, the encoder's distortion and the decoder's reconstruction levels are given by

$$D_s = c_1 \sum_{i=1}^{M} \left(\mathbb{E}\{\eta_s(x, y_i(\mathbf{q}))\} + c_2 \int\limits_{x_{i-1}}^{x_i} \eta_s(x, y_i(\mathbf{q})) \mathrm{d}\mu_X \right),$$

$$y_i = \underset{y \in \mathcal{Y}}{\arg\min} \left(c_1(\mathbb{E}\{\eta_s(x, y)\} + c_2 \int\limits_{x_{i-1}}^{x_i} \eta_s(x, y) \mathrm{d}\mu_X \right).$$

For an MSE decoder,

$$y_i = \frac{c_1 \mathbb{E}\{X\} + c_2 \int_\theta \int x \mathrm{d}\mu_{X,\theta}}{c_1 + c_2 \int_\theta \int \mathrm{d}\mu_{X,\theta}}, \tag{17}$$

We use gradient descent to optimize $\mathbf{q}$. The gradients for MSE decoder with encoder distortion η_E,

$$\frac{\partial D_E}{\partial x_i} = c_1 \frac{\partial y_i}{\partial x_i} \frac{\partial}{\partial y_i} \mathbb{E}\{\eta_s(x, y_i)\} + c_1 \frac{\partial y_{i+1}}{\partial x_i} \frac{\partial}{\partial y_{i+1}} \mathbb{E}\{\eta_s(x, y_{i+1})\} + c_2 \frac{\partial \overline{D_E}}{\partial x_i}$$

where

$$\overline{D_E} = \sum_{i=1}^{M} \int\limits_{x_{i-1}}^{x_i} \eta_s(x, y_i(\mathbf{q})) \mathrm{d}\mu_X \Big) \tag{18}$$

$$\begin{aligned} \frac{\partial \overline{D_E}}{\partial x_i} &= \eta_E(x_i, y_i) \frac{\mathrm{d}\mu_X(x_i)}{\mathrm{d}x} + c_2 \eta_E(x_i, y_{i+1}) \frac{\mathrm{d}\mu_X(x_i)}{\mathrm{d}x} \\ &+ c_2 \frac{\partial y_i}{\partial x_i} \frac{\partial}{\partial y_i} \int \eta_E(x, y_i) \mathrm{d}\mu_X \qquad (19) \\ &+ c_2 \frac{\partial y_{i+1}}{\partial x_i} \frac{\partial}{\partial y_i} \int \eta_E(x, y_i) \mathrm{d}\mu_X \qquad (20) \end{aligned}$$

$$\frac{\partial y_i}{\partial x_i} = c_2 \frac{\mathrm{d}\mu_X(x_i)}{\mathrm{d}x_i} \frac{x_i - y_i}{(c_1 + c_2 \int \mathrm{d}\mu_X)} \tag{21}$$

$$\frac{\partial y_{i+1}}{\partial x_i} = -c_2 \frac{\mathrm{d}\mu_X(x_i)}{\mathrm{d}x_i} \frac{x_i - y_{i+1}}{(c_1 + c_2 \int_{x_i}^{x_{i+1}} \mathrm{d}\mu_X)} \tag{22}$$

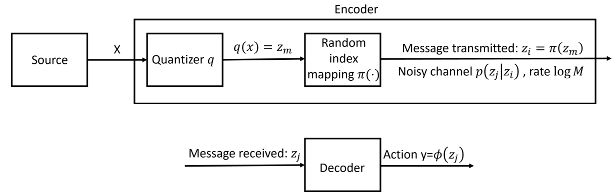


Fig. 4: Communication diagram over a noisy channel

VI. Gradient descent algorithm

The classical vector quantization solution relies on iterative Lloyd-Max optimization, where the encoder and the decoder optimize their mappings iteratively. Since the encoder and the decoder's objectives as the same (team problem), the distortion is non-increasing with each iteration and converges to a locally optimal solution. On the contrary, we consider a game problem (as opposed to a team problem) where the objectives differ. A variation of these algorithms for the strategic settings would enforce optimality alternatively between the distortion measures of the encoder and the decoder. Hence, the iterative optimization for strategic cases may not converge, as demonstrated in [26]. While the functional gradient approach of perturbing the quantizer mapping using an admissible perturbation function is a natural optimization method, selecting an appropriate set of admissible functions can be challenging as they need to satisfy the quantizer's properties such as rate and convex codecell requirements. This can limit the tractability of the more general functional optimization approach. Instead, we optimize by performing gradient descent on $\mathbf{q}$. However, if the cost function is non-convex, the algorithm may get stuck at a poor local optima. As a simple remedy, we perform gradient descent with multiple initializations and pick the best local optima among them. The algorithm is summarized below. The codes are available at https://tinyurl.com/quadraticsq.

VII. Algorithm: quadratic

Function main():

Input: $\mu_{X,\theta}(\cdot,\cdot), \mathcal{X}, \mathcal{T}, M, \eta_E, \eta_D, p_b$
Output: $\{q^*_\theta\}$, $\{y^*_m\}$, D_E, D_D
Initialization: $\mathbf{q}_{init}, iter = 1$
Parameters: $\epsilon, \triangle$
$p_{err} \leftarrow 1 - (1 - p_b)^{\log_2 M}$
$c_1 \leftarrow p_{err}/(M - 1)$
$c_2 \leftarrow 1 - Mc_1$
$\mathbf{q} \leftarrow \mathbf{q}_{init}$
$\mathbf{y} \leftarrow reconstruction(\mathbf{q}, \mu_{X,\theta}, p_{err}, \mathbb{E}\{X\})$
$flag \leftarrow 1$
while $flag \neq 0$ **do**
  $distenc \leftarrow distortion(\mathbf{q}, \mathbf{y}, \mu_{X,\theta}, \eta_E, E, p_{err})$
  for $\theta \in \mathcal{T}$ **do**
    for $i \in [1 : M - 1]$ **do**
      $\Delta \leftarrow 1$
      $der_{\theta,i} \leftarrow derivative(\mathbf{q}, \mathbf{y}, \mu_{X,\theta}, \theta, i, p_{err})$
      $temp \leftarrow q_{\theta,i} - \triangle\, der_{\theta,i}$
      $\mathbf{qt} \leftarrow \mathbf{q}$
      $qt_{\theta,i} \leftarrow temp$
      $\mathbf{y} \leftarrow reconstruction(\mathbf{qt}, \mu_{X,\theta}, p_{err}, \mathbb{E}\{X\})$

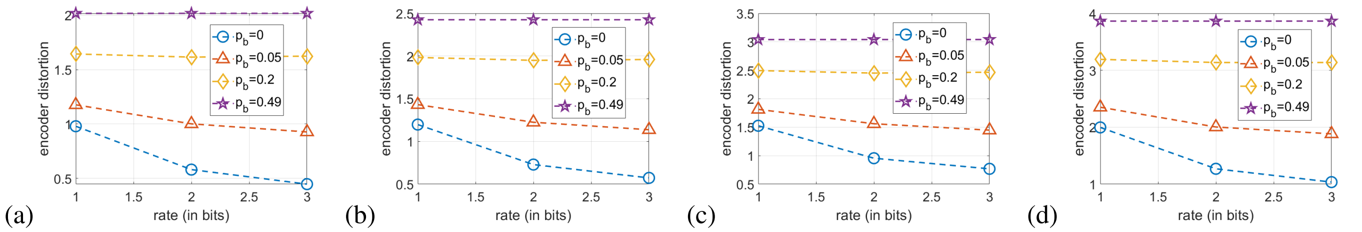


Fig. 5: Encoder distortions for jointly Gaussian (X, θ) with varying correlation: (a) $\rho = 0$ (b) $\rho = 0.2$ (c) $\rho = 0.5$ (d) $\rho = 0.9$

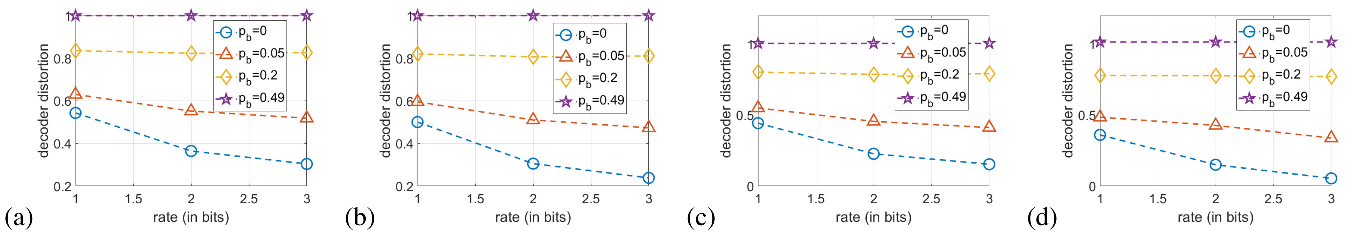


Fig. 6: Decoder distortions for jointly Gaussian (X, θ): (a) $\rho = 0$ (b) $\rho = 0.2$ (c) $\rho = 0.5$ (d) $\rho = 0.9$

$dt \leftarrow distortion(\mathbf{qt}, \mathbf{y}, \mu_{X,\theta}, \eta_E, E, p_{err})$
if $temp > q_{\theta,i-1}$ && $temp < q_{\theta,i}$ && $dt < distenc$ **then**
$\mathbf{q} \leftarrow \mathbf{qt}$
else
$\mathbf{q} \leftarrow check(\mathbf{q}, \mu, p_{err}, \mathbb{E}\{X\}, \triangle, der_{\theta,i}, distenc, \theta, i)$
end if
end for
end for
$\mathbf{y} \leftarrow reconstruction(\mathbf{q}, \mu, p_{err}, \mathbb{E}\{X\})$
$dt \leftarrow distortion(\mathbf{q}, \mathbf{y}, \mu, \eta_E, E, p_{err})$
if $iter > 1$ **then**
if $all(\mathbf{der}) < \epsilon \,||\, dt == distenc$ **then**
$flag = 0$
end if
end if
$iter \leftarrow iter + 1$
end while
$\mathbf{q}^* \leftarrow \mathbf{q}$
$\mathbf{y}^* \leftarrow reconstruction(\mathbf{q}^*, \mu, p_{err}, \mathbb{E}\{X\})$
$D_E \leftarrow distortion(\mathbf{q}^*, \mathbf{y}^*, \mu, \eta_E, E, p_{err})$
$D_D \leftarrow distortion(\mathbf{q}^*, \mathbf{y}^*, \mu, \eta_D, E, p_{err})$

Function check():

Input: $\mathbf{q}, \mu, p_{err}, \mathbb{E}\{X\}, \triangle, der_{\theta,i}, distenc, \theta, i$
Output: $\mathbf{q}$
Parameters: ϵ
while $\triangle > eps$ **do**
$\triangle \leftarrow \triangle/10$
$temp \leftarrow q_{\theta,i} - \triangle\, der_{\theta,i}$
$\mathbf{qt} \leftarrow \mathbf{q}$
$qt_{\theta,i} \leftarrow temp$
$\mathbf{y} \leftarrow reconstruction(\mathbf{qt}, \mu, p_{err}, \mathbb{E}\{X\})$
$dt \leftarrow distortion(\mathbf{qt}, \mathbf{y}, \mathcal{T}, \mu, \eta_s, s, p_{err})$
if $temp > q_{\theta,i-1}$ && $temp < q_{\theta,i}$ && $dt < distenc$ **then**
$\mathbf{q} \leftarrow \mathbf{qt}$
break
end if
end while

Function distortion():

Input: $\mathbf{q}, \mathbf{y}, \mathcal{T}, \mu_{X,\theta}, \eta_s, s, p_{err}$
Output: D_s
Initialization: $D_s = 0$
for $i \in [1:M]$ **do**
for $\theta \in \mathcal{T}$ **do**
$D_s = D_s + c_2 \int_{x_{\theta,end}} \eta_s(x, \theta, y_i) d\mu_{X,\theta}(x, \theta)$
$D_s \leftarrow D_s + c_1 \int_{x_{\theta,1}} \eta_s(x, \theta, y_i) d\mu_{X,\theta}(x, \theta)$
end for
end for

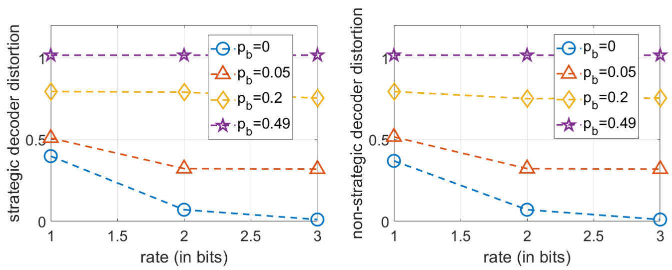
Function reconstruction():

Input: $\mathbf{q}, \mathcal{T}, \mu_{X,\theta}, \eta_D, p_{err}$
Output: $\mathbf{y}$
for $i \in [1:M]$ **do**
for $y \in \mathcal{Y}$ **do**
$dist_i(y) = c_1 \int_{\theta \in \mathcal{T}} \int_{x_{\theta,1}}^{x_{\theta,end}} \eta_s(x, \theta, y) d\mu_{X,\theta} + c_2 \int_{\theta \in \mathcal{T}} \int \eta_s(x, \theta, y) d\mu_{X,\theta}$
end for
$y_i \leftarrow \arg\min_{y \in \mathcal{Y}} dist_i(y)$
end for

VIII. Numerical Results

We present numerical results for the following settings:

1) $\eta_E = (x + \theta - y)^2$
2) $\eta_E = (x^3 - y)^2$
3) $\eta_E = (x^2 - y)^2$
4) $\eta_E = (1.5x - y)^2$



(a) $\rho = 0.99$ (b) $\rho = 0.99$

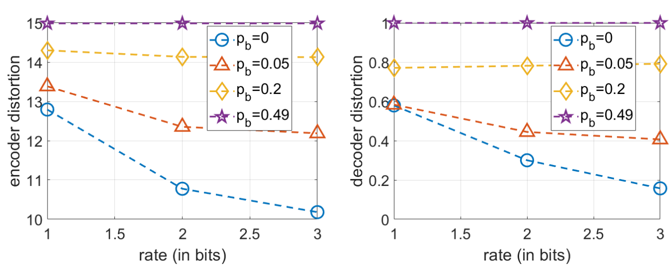
Fig. 7: Comparison of strategic decoder and non-strategic distortions for $\rho = 0.99$.

with an MSE decoder $\eta_D = (x-y)^2$ for all cases, communicating over a binary symmetric channel with crossover probability $p_b \le \frac{1}{2}$, which yields

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

We take the following bit error rates$p_b = \begin{bmatrix} 0 & 0.05 & 0.2 & 0.49 \end{bmatrix}$. For the first setting we consider a jointly Gaussian source $(X, \theta) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho \\ \rho & \sigma_\theta^2 \end{bmatrix} \right), 0 \le \rho < 1$ and plot the results for a correlation of $\rho = \begin{bmatrix} 0 & 0.2 & 0.5 & 0.9 \end{bmatrix}$ in Figures 5 and 6. We observe that the encoder's distortion increases with correlation. When $\rho = 0.99$, i.e., the encoder's distortion is essentially $\eta_E = (x + \theta - y)^2 = (2x - y)^2$, the decoder distortion is negligibly close to the non-strategic distortion ($\eta_E = (x-y)^2$) as seen in Figure 7.

We observe in Figures **??** that the encoder (and the decoder) distortions remain the same for bit error rate $p_b = 0.49$, and is equal to the distortion in the non-informative setting (equivalent to $M = 1$, the encoder does not send any message).



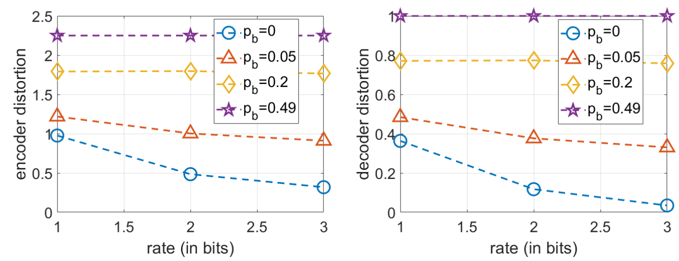
(a) Encoder distortion (b) Decoder distortion

Fig. 8: $\eta_E = (x^3 - y)^2$

For the other settings, we consider a Gaussian source $X \sim \mathcal{N}(0,1)$. All Gaussians are considered with a limited range $[-5, 5]$ for numerical ease.

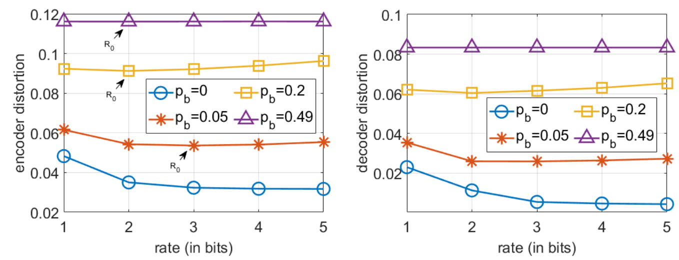
IX. Conclusion

In this paper, we propose a gradient descent based solution for the problem of strategic quantization with channel noise. Obtained numerical results suggest the validity of the proposed algorithm.



(a) Encoder distortion (b) Decoder distortion

Fig. 9: $X \sim \theta \sim \mathcal{N}(0,1), \eta_E^2 = (1.5x - y)^2, \eta_D = (x-y)^2$



(a) Encoder distortion (b) Decoder distortion

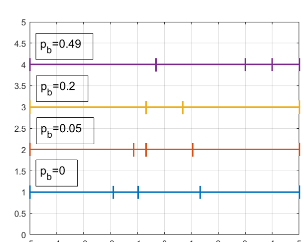
Fig. 10: $X \sim \theta \sim \mathcal{N}(0,1), \eta_E^2 = (x^2 - y)^2, \eta_D = (x-y)^2$

Appendix A
Quantizer behaviour for $\eta_E = (x + \theta - \beta y)^2, \eta_D = (x-y)^2$

Consider a continuous 2-dimensional source $(X, \theta) \sim \mu_{X,\theta}$, $\eta_E(x, \theta, y) = (x + \theta - \beta y)^2$, $\eta_D(x,y) = (x-y)^2$ for a given $\alpha, \beta \in \mathbb{R}$ quantized to M levels. In other words, the decoder wants to reconstruct X as closely as possible, while the encoder wants the decoder's construction to be as close as possible to $\frac{X+\theta}{\beta}$, both in the MSE sense. Can the encoder "persuade" the decoder by carefully designing quantizer intervals $\mathcal{V}_{\theta,m}^*$?

Let us parameterize $\mathcal{V}_{\theta,m}^*$ as $[x_{\theta,m-1}, x_{\theta,m})$, where $x_{\theta,m} \in \mathcal{X}$, $x_{\theta,m-1} < x_{\theta,m}$, i.e., $\mathbf{q} = \{q_\theta, \theta \in \mathcal{T}\}$, $q_\theta = [x_{\theta,0}, \ldots, x_{\theta,M}]$. For a given $\mathbf{q}$, the decoder determines $\mathbf{y} = [y_1, \ldots, y_m]$ as follows:

$$y_m = \frac{\int\limits_{\theta \in \mathcal{T}} \int x \mathrm{d}\mu}{\int\limits_{\theta \in \mathcal{T}} \int \mathrm{d}\mu}.$$



(a) $\eta_E = (x^3 - y)^2$ (b) $\eta_E = (1.5x - y)^2$

Fig. 11: Quantizers for $M = 4$.

The encoder's distortion and its derivative with respect to $x_{\theta,m}$

$$J(\mathbf{q}) = \sum_{m=1}^{M} \int_{\theta\in\mathcal{T}} \int (x+\theta-\beta y_m)^2 \mathrm{d}\mu,$$

$$\begin{aligned}\frac{\partial J}{\partial x_{\theta',m}} &= (x_{\theta',m}+\theta'-\beta y_m)^2 \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta'} \\ &\quad - (x_{\theta',m}+\theta'-\beta y_{m+1})^2 \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta'} \\ &\quad - 2\beta \frac{\mathrm{d}y_m}{\mathrm{d}x_{\theta',m}} \int_\theta \int (x+\theta-\beta y_m)\mathrm{d}\mu \\ &\quad - 2\beta \frac{\mathrm{d}y_{m+1}}{\mathrm{d}x_{\theta',m}} \int_\theta \int_{\mathcal{V}_{m+1}} (x+\theta-\beta y_{m+1})\mathrm{d}\mu \\ &= J_1 + J_2 + I_1 + I_2.\end{aligned}$$

$$\begin{aligned}J_1 + J_2 = (2x_{\theta',m} + 2\theta' - \beta(y_m + y_{m+1}))(\beta(y_{m+1}-y_m)) \ldots \\ \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta}\end{aligned}$$

The terms $\frac{\mathrm{d}y_m}{\mathrm{d}x_{\theta',m}}$ and $\frac{\mathrm{d}y_{m+1}}{\mathrm{d}x_{\theta',m}}$ are

$$\begin{aligned}\frac{\mathrm{d}y_m}{\mathrm{d}x_{\theta',m}} &= \frac{x_{\theta',m}\frac{\mathrm{d}\mu(x_{\theta',m})}{\mathrm{d}x\mathrm{d}\theta}\int\limits_{\theta\in\mathcal{T}}\int \mathrm{d}\mu - \frac{\mathrm{d}\mu(x_{\theta',m})}{\mathrm{d}x\mathrm{d}\theta}\int\limits_{\theta\in\mathcal{T}}\int x\mathrm{d}\mu}{(\int_\theta\int \mathrm{d}\mu)^2} \\ &= \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta} \frac{x_{\theta',m}-y_m}{\int\limits_{\theta\in\mathcal{T}}\int \mathrm{d}\mu_X}\end{aligned}$$

$$\frac{\mathrm{d}y_{m+1}}{\mathrm{d}x_{\theta',m}} = -\frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta} \frac{x_{\theta',m}-y_{m+1}}{\int\limits_{\theta\in\mathcal{T}}\int\limits_{\mathcal{V}_{\theta,m+1}} \mathrm{d}\mu}$$

$$\begin{aligned}I_1 &= -2\beta \frac{\mathrm{d}y_m}{\mathrm{d}x_{\theta',m}} \int\limits_{\theta\in\mathcal{T}}\int (x+\theta-\beta y_m)\mathrm{d}\mu \\ &= -2\beta \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta} \frac{x_{\theta',m}-y_m}{\int\limits_{\theta\in\mathcal{T}}\int \mathrm{d}\mu} \int_\theta\int (x+\theta-\beta y_m)\mathrm{d}\mu \\ &= -2\beta \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta}(x_{\theta',m}-y_m)\left(y_m + \frac{\int\limits_{\theta\in\mathcal{T}}\int \theta\mathrm{d}\mu}{\int\limits_{\theta\in\mathcal{T}}\int \mathrm{d}\mu} - \beta y_m\right) \\ &= -2\beta \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta}(x_{\theta',m}-y_m)\left(y_m(1-\beta) + \mathbb{E}\{\theta|\mathcal{V}_{:,m}\}\right)\end{aligned}$$

$$\begin{aligned}I_2 &= -2\beta \frac{\mathrm{d}y_{m+1}}{\mathrm{d}x_{\theta',m}} \int\limits_{\theta\in\mathcal{T}}\int\limits_{\mathcal{V}_{\theta,m+1}} (x+\theta-\beta y_{m+1})\mathrm{d}\mu \\ &= 2\beta \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta} \frac{x_{\theta',m}-y_{m+1}}{\int\limits_{\theta\in\mathcal{T}}\int\limits_{\mathcal{V}_{\theta,m+1}} \mathrm{d}\mu} \int\limits_{\theta\in\mathcal{T}}\int\limits_{\mathcal{V}_{\theta,m+1}} (x+\theta-\beta y_{m+1})\mathrm{d}\mu \\ &= 2\beta \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta}(x_{\theta',m}-y_{m+1})\Big(y_{m+1}(1-\beta) \\ &\quad + \mathbb{E}\{\theta|\mathcal{V}_{:,m+1}\}\Big)\end{aligned}$$

$$\begin{aligned}I_1 + I_2 &= 2\beta \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta}\Big((x_{\theta',m}-y_{m+1})(y_{m+1}(1-\beta) \\ &\quad + \mathbb{E}\{\theta|\mathcal{V}_{:,m+1}\}) - (x_{\theta',m}-y_m)(y_m(1-\beta) \\ &\quad + \mathbb{E}\{\theta|\mathcal{V}_{:,m}\})\Big) \\ &= 2\beta \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta}\Big((1-\beta)(y_{m+1}-y_m)(x_{\theta',m} \\ &\quad - (y_{m+1}+y_m)) - y_{m+1}\mathbb{E}\{\theta|\mathcal{V}_{:,m+1}\} \\ &\quad + y_m\mathbb{E}\{\theta|\mathcal{V}_{:,m}\} + x_{\theta',m}(\mathbb{E}\{\theta|\mathcal{V}_{:,m+1}\} - \mathbb{E}\{\theta|\mathcal{V}_{:,m}\})\Big)\end{aligned}$$

$$\begin{aligned}P &= J_1 + J_2 + I_1 + I_2 \\ &= \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta}\beta\Big((2\theta' + x_{\theta',m} - \beta y_m + x_{\theta',m} \\ &\quad - \beta y_{m+1})(y_{m+1}-y_m) + 2(y_{m+1}-y_m)(1-\beta) \\ &\quad (x_{\theta',m} - (y_{m+1}+y_m)) - 2y_{m+1}\mathbb{E}\{\theta|\mathcal{V}_{:,m+1}\} \\ &\quad + 2y_m\mathbb{E}\{\theta|\mathcal{V}_{:,m}\} + 2x_{\theta,m}(\mathbb{E}\{\theta|\mathcal{V}_{:,m+1}\} - \mathbb{E}\{\theta|\mathcal{V}_{:,m}\})\Big) \\ &= \frac{\mathrm{d}\mu(x_{\theta',m},\theta')}{\mathrm{d}x\mathrm{d}\theta}\beta\Big(2x_{\theta',m}\Big((2-\beta)(y_{m+1}-y_m) \\ &\quad + \mathbb{E}\{\theta|\mathcal{V}_{:,m+1}\} - \mathbb{E}\{\theta|\mathcal{V}_{:,m}\}\Big) + 2\theta'(y_{m+1}-y_m) \\ &\quad + (\beta-2)(y_{m+1}^2 - y_m^2) - 2(y_{m+1}\mathbb{E}\{\theta|\mathcal{V}_{m+1}\} \\ &\quad - y_m\mathbb{E}\{\theta|\mathcal{V}_m\})\Big)\end{aligned}$$

Enforcing the KKT conditions for optimality

$$\begin{aligned}x_{\theta',m} = \Big(&-2\theta'(y_{m+1}-y_m) + (2-\beta)(y_{m+1}^2 - y_m^2) \\ &+ 2(y_{m+1}\mathbb{E}\{\theta|\mathcal{V}_{m+1}\} - y_m\mathbb{E}\{\theta|\mathcal{V}_m\})\Big)\Big/ \\ &2\Big((2-\beta)(y_{m+1}-y_m) + \mathbb{E}\{\theta|\mathcal{V}_{m+1}\} - \mathbb{E}\{\theta|\mathcal{V}_m\}\Big)\end{aligned} \tag{23}$$

the other condition is $y_{m+1} = y_m$ which is not possible since the actions are considered unique - if not, the corresponding regions could be combined

If X and θ are independent,$\mathbb{E}\{\theta|\mathcal{V}_{m+1}\} = \mathbb{E}\{\theta|\mathcal{V}_m\} = \mathbb{E}\{\theta\}$,

$$\frac{(y_{m+1}+y_m)}{2} + \frac{(\mathbb{E}\{\theta\} - \theta')}{2-\beta} \tag{24}$$

This implies that the quantizer is a shifted version of the non-strategic quantizer if $\beta \neq 0, 2$, if the encoder decides to send something.

$$\frac{\partial J}{\partial x_m} = 0, \quad m \in [1:M] \tag{25}$$

we obtain, after some straightforward algebra, that the solutions that satisfies (25) are $\beta = 0, 2$, or $x_m = \frac{y_m + y_{m+1}}{2}$ (the other condition is $y_{m+1} = y_m$ which is not possible since the actions are considered unique - if not, the corresponding regions could be combined). This implies that the quantizer is the same as the non-strategic quantizer if $\beta \neq 0, 2$, if the encoder decides to send something.

The encoder's distortion can be simplified to the following form:

$$J = \int_a^b x^2 \mathrm{d}\mu + \alpha^2 + 2\alpha(1-\beta)\int_a^b x \mathrm{d}\mu + \beta(\beta-2)\sum_{m=1}^{M} y_m \int_{x_{m-1}}^{x_m} x \mathrm{d}\mu.$$

The distortion for a non-informative quantizer:

$$\begin{aligned} D_{non-revealing} &= \int_a^b (x+\alpha-\beta y)^2 \mathrm{d}\mu \\ &= \int_a^b x^2 \mathrm{d}\mu + \alpha^2 + 2\alpha(1-\beta)\int_a^b x \mathrm{d}\mu \\ &\quad + \beta(\beta-2) y \int_a^b x \mathrm{d}\mu. \end{aligned} \tag{26}$$

The encoder's distortion can be re-written in terms of the non-revealing distortion and some other terms as

$$\begin{aligned} J &= \int_a^b x^2 \mathrm{d}\mu + \alpha^2 + 2\alpha(1-\beta)\int_a^b x \mathrm{d}\mu \\ &\quad + \beta(\beta-2)\sum_{m=1}^{M} y_m \int_{x_{m-1}}^{x_m} x \mathrm{d}\mu && (27) \\ &= D_{non-revealing} + \beta(\beta-2)\Bigg(\sum_{m=1}^{M} y_m \int_{x_{m-1}}^{x_m} x \mathrm{d}\mu \\ &\quad - y\int_a^b x \mathrm{d}\mu\Bigg) && (28) \\ &= D_{non-revealing} + \beta(\beta-2)\Bigg(\sum_{m=1}^{M} \frac{\int_{x_{m-1}}^{x_m} x \mathrm{d}\mu}{\int_{x_{m-1}}^{x_m} \mathrm{d}\mu} \int_{x_{m-1}}^{x_m} x \mathrm{d}\mu \\ &\quad - \frac{\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} x \mathrm{d}\mu}{\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} \mathrm{d}\mu} \sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} x \mathrm{d}\mu\Bigg) \\ &= D_{non-revealing} + \beta(\beta-2)\Bigg(\sum_{m=1}^{M} \frac{\left(\int_{x_{m-1}}^{x_m} x \mathrm{d}\mu\right)^2}{\int_{x_{m-1}}^{x_m} \mathrm{d}\mu} \\ &\quad - \frac{\left(\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} x \mathrm{d}\mu\right)^2}{\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} \mathrm{d}\mu}\Bigg). \end{aligned}$$

Let $T = \left(\sum_{m=1}^{M} \frac{\left(\int_{x_{m-1}}^{x_m} x \mathrm{d}\mu\right)^2}{\int_{x_{m-1}}^{x_m} \mathrm{d}\mu} - \frac{\left(\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} x \mathrm{d}\mu\right)^2}{\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} \mathrm{d}\mu}\right)$. In order for the quantizer to be informative ($M > 1$), the second term has to be less than 0. This happens in three cases:

1) $\beta < 0$ and $T < 0$
2) $0 < \beta < 2$ and $T > 0$
3) $\beta > 2$ and $T < 0$

From Sedrakyan's lemma (derived from cauchy-shwarz inequality), we have that for real numbers $u_1, u_2, \ldots, u_n$ and positive real numbers $v_1, v_2, \ldots, v_n$:

$$\frac{(\sum_{i=1}^{n} u_i)^2}{\sum_{i=1}^{n} v_i} \leq \sum_{i=1}^{n} \frac{u_i^2}{v_i}.$$

Consider $u_i = \int_{x_{m-1}}^{x_m} x \mathrm{d}\mu, v_i = \int_{x_{m-1}}^{x_m} \mathrm{d}\mu$ which satisfies the conditions of the lemma of real u_i and positive real v_i. We get

$$\frac{\left(\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} x \mathrm{d}\mu\right)^2}{\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} \mathrm{d}\mu} < \sum_{m=1}^{M} \frac{\left(\int_{x_{m-1}}^{x_m} x \mathrm{d}\mu\right)^2}{\int_{x_{m-1}}^{x_m} \mathrm{d}\mu}$$

that is T is always non-negative.

This implies that the only possible case is case 2 with $0 < \beta < 2$, and the encoder chooses a non-strategic quantizer (as we show earlier in Equation 25 that the only solution when $\beta \neq 0, 2$ is a non-strategic encoder if the encoder sends some message). From Equation 28, we see that for $\beta = 0, 2$ the encoder distortion is the same as non-revealing distortion regardless of the quantizer used.

The optimal policy for the encoder is to be fully revealing for $\beta \in (0, 2)$, and the distortion remains the same for any M level quantization when $\beta = 0, 2$, and non-revealing otherwise.

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

[13] M. Ostrovsky and M. Schwarz, "Information Disclosure and Unraveling in Matching Markets," *American Economic Journal: Microeconomics*, vol. 2, no. 2, pp. 34–63, 2010.

[14] R. Alonso and O. Câmara, "Persuading Voters," *American Economic Review*, vol. 106, no. 11, pp. 3590–3605, 2016.

[15] O. Candogan and K. Drakopoulos, "Optimal Signaling of Content Accuracy: Engagement vs. Misinformation," *Operations Research*, vol. 68, no. 2, pp. 497–515, 2020.

[16] E. Akyol, C. Langbort, and T. Başar, "Privacy Constrained Information Processing," in *54th IEEE conference on decision and control (CDC)*. IEEE, 2015, pp. 4511–4516.

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

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

[19] S. Dughmi and H. Xu, "Algorithmic Bayesian Persuasion," *SIAM Journal on Computing*, no. 0, pp. STOC16–68, 2019.

[20] Y. C. Aybaş and E. Türkel, "Persuasion with Coarse Communication," *arXiv preprint arXiv:1910.13547*, 2019.

[21] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2383, 1998.

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

[23] J. Sobel, "Giving and Receiving Advice," *Advances in Economics and Econometrics*, vol. 1, pp. 305–341, 2013.

[24] J. Mensch, "Monotone Persuasion," *Games and Economic Behavior*, vol. 130, pp. 521–542, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0899825621001263

[25] P. Dworczak and G. Martini, "The Simple Economics of Optimal Persuasion," *Journal of Political Economy*, vol. 127, no. 5, pp. 1993–2048, Oct. 2019.

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

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

[28] ——, "Channel-optimized strategic quantizer design via dynamic programming," *IEEE Statistical Signal Processing Workshop*, 2023. To appear, available at https://tinyurl.com/asilomar2023.