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

***Abstract*—This paper is concerned with the quantization setting where the encoder and the decoder have misaligned objectives and communicate over a noisy channel. While the unconstrained variation of this problem has been well-studied under the theme of information design problems in Economics, the problem becomes more appealing and relevant to engineering applications with a constraint on the cardinality of the message space. We first consider a scalar source X and develop a gradient-descent based solution along with random index assignment for the $1:1$ quantizer mapping. In our prior work, we use dynamic programming for this scalar setting. We use gradient descent here to reduce the complexity of the algorithm. Since the objective is non-convex, we use multiple initializations. We extend our analysis to a 2-dimensional source (X, θ), and design a locally optimal $2:1$ quantizer mapping via optimizing a quantizer for each particular realization of θ. We finally present numerical results obtained via the proposed algorithms that suggest their validity and demonstrate the strategic quantization features that differentiate it from its classical counterpart. The codes for all of the experiments in this paper are available at: https://tinyurl.com/asilomar2023.**

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

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 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.

Building on [12], [13], where we study strategic quantizer design over a perfect (noiseless) communication channel (see also our recent related work [11]), we analyze and design channel-optimized strategic quantizer for two classes of distortion functions in the form of $1:1$ and $2:1$ mappings. Our main computational design tool here is gradient descent used in conjunction with random index assignment which has been successfully used for the classical counterpart of this problem to account for the noisy channel. Compared to our recent related work on channel-optimized strategic quantization [11], the contribution of this paper is two-fold: First, 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. Second, we extend our analysis from scalar settings to $2:1$ distortion settings via a subterfuge by designing a set of quantizers.

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.

Consider a real-life scenario where two smart cars from competing manufacturers, such as Tesla and Honda, attempt to exchange information. Specifically, the Tesla car (decoder) requests specific information from the Honda car (encoder) to decide whether to change its route in response to traffic congestion. While Tesla's objective is to accurately estimate traffic congestion, 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. Assuming a fixed-rate noisy channel, how would these cars communicate? Our analysis here enables us to quantitatively study such problems.

II. Problem Formulation

Consider the following scalar quantization problem: an encoder observes a realization of the source $X \in \mathcal{X}$ with a probability distribution μ_X, and maps X 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, $\mathbf{q} : \mathcal{X} \to \mathcal{Z}$. 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

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

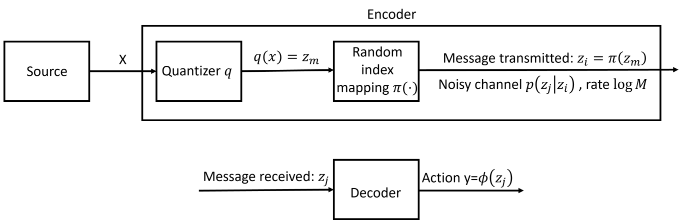


Fig. 1: Communication diagram

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 $\mu_X(\cdot)$. The objectives (η_E and η_D), the shared prior (μ), 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.

In this paper, we focus on two special cases (with MSE decoder $\eta_D = (x - y)^2$ for both):

1) Scalar source and channel such as X with $\eta_E^1 = (x^3 - y)^2, \eta_E^2 = (1.5x - y)^2$.
2) 2-dimensional vector source-scalar channel reconstruction setting with the source as $(X, \theta) \sim \mu_{X,\theta}$. Specifically, we take a quadratic cost measure $\eta_E = (x + \theta - y)^2$.

In our recent paper [13], we use dynamic programming solution concept along with random index assignment for the $1 : 1$ 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, and we further extend the discussion to a 2-dimensional vector source. 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 [13] is not enforced here, since unlike dynamic programming, gradient descent does not involve optimization of sub-problems.

III. Main Results

A. Analysis

Let X take values from the source alphabet $\mathcal{X} \in [a_X, b_X]$ with probability distribution function μ_X. The set $\mathcal{X}$ is divided into mutually exclusive and exhaustive sets, $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_M$. We make the following "monotonicity" assumption.

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

Remark 1: Assumption 1 is the first of the two regularity conditions commonly employed in the classical quantization literature, cf. [6]. Note that the second regularity condition, $y_m \in \mathcal{V}_m$, is not included in Assumption 1.

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

$$\mathcal{V}_i = [x_{i-1}, x_i).$$

The encoder chooses a non-injective mapping, $Q : \mathcal{X} \to \mathcal{Z}$ which is the quantizer $\mathbf{q}$ with boundary levels $[x_0, x_1, \ldots, x_M]$. The decoder determines a set of actions, $\mathbf{y} = [y_1, \ldots, y_M]$ as the best response to $\mathbf{q}$ to minimize its cost D_D for $i \in [1 : M]$ as follows

$$y_i^* = \arg\min_{y_i \in \mathcal{Y}} \sum_{i=1}^{M} \mathbb{E}_\pi\{\mathbb{E}\{\eta_D(x, \mathbf{y}(\mathbf{q}))|\pi, x \in \mathcal{V}_i\}\}.$$

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

Remark 2: A crucial factor to consider is that the decoder can optimize its action ($\mathbf{y}$), as the best response to the information provided by the encoder ($\mathbf{q}$). However, the encoder is constrained by its commitment to its original choice of $\mathbf{q}$ and cannot determine $\mathbf{q}$ as the best response to $\mathbf{y}$. Instead, it can only optimize as the best response to a function of itself, $\mathbf{y}(\mathbf{q})$. This introduces a hierarchy in the game play, where the encoder acts first and the decoder responds, resulting in what is referred to as the "Stackelberg equilibrium" in the computer science and control literature, and more formally constitutes an instance of subgame perfect Bayesian Nash equilibrium. It is not a Nash equilibrium since $\mathbf{q}$ may not be the best response to $\mathbf{y}$. In cheap talk [14], Nash equilibria are sought after and the equilibria achieving strategies happen to be non-injective mappings (quantizers) without an exogenous rate constraint. It is essential to note that the problem formulation in this paper differs substantially from the cheap talk literature [14].

The encoder designs a quantizer $\mathbf{q}$ using only the objectives $(\eta_s, s \in \{E, D\})$, the statistics of the source $(\mu_X(\cdot))$, and the channel transition probability matrix $(p(j|i))$, and the index assigement (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

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

Problem 1: 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 $\mu_X(x)$, 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:

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

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

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

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

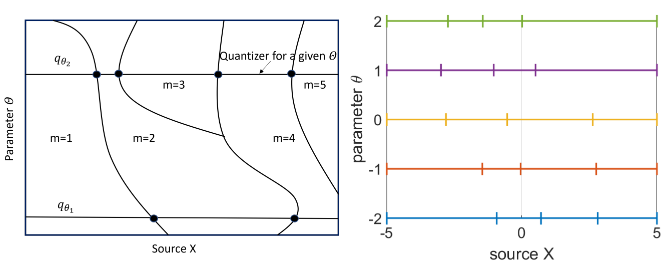
The average distortion over all possible index assignments are

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

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

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

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



(a) M=3 level quantizer (b) Example quantizer

Fig. 2: Quantization of X parameterized by θ.

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

In the above equation, (a)is obtained via adding and subtracting $\sum_{i=1}^{M}\int \eta_s(x, y_i(q))\mathrm{d}\mu_X$, ($b$) follows from exchanging the summations over i and j, using $\mathbb{E}\{\eta_s(x, y_j(q))\} = \sum_{i=1}^{M}\int \eta_s(x, y_j(q))\mathrm{d}\mu_X$, and changing the summation index of the first term to i. The average distortions and the optimum decoder reconstruction for $i \in [1 : M]$ are :

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

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

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

$$\overline{D_s} = \sum_{i=1}^{M}\int \eta_s(x, y_i)\mathrm{d}\mu_X.$$

The reconstruction levels $\mathbf{y}$ are found using the first-order KKT optimality condition. If the decoder distortion is MMSE, $\eta_D(x, y) = (x - y)^2$,

$$\frac{\partial D_D}{\partial y_i} = -2c_1\sum_{j=1}^{M}\int_{x_{j-1}}^{x_j}(x - y_i)\mathrm{d}\mu_X - 2c_2\int(x - y_i)\mathrm{d}\mu_X,$$

$$y_i = \frac{c_1\mathbb{E}\{X\} + c_2\int x\mathrm{d}\mu_X}{c_1 + c_2\int \mathrm{d}\mu_X}. \tag{1}$$

B. $2-$*dimensional source*

We extend this problem to a 2-dimensional vector source (X, θ) with joint probability distribution $\mu_{X,\theta}$ where $\theta \in \Theta$ as finding a set of quantizers $\mathbf{q} = \{q_\theta, \theta \in \Theta\}$. The decoder decides a single set of actions $\mathbf{y}$ since it is unaware of the realization of θ. The derivation is given in Appendix A. Note that implementing a $2:1$ quantizer from $\mathbf{q} : (\mathcal{X}, \Theta) \to \mathcal{Z}$ can be simplified to computing a set of quantizers corresponding to each $\theta \in \Theta$ as in Figure 2 without loss of generality. If the quantizer does not cross region m for some realization of θ, the

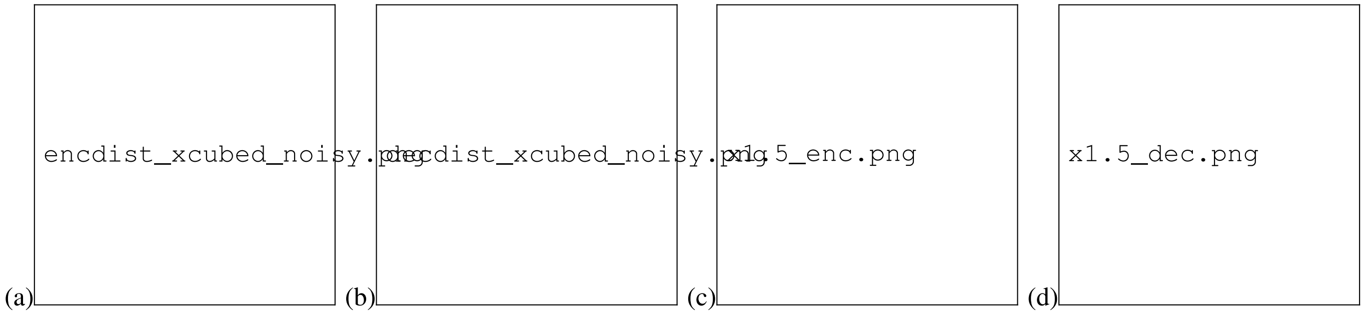


Fig. 3: Encoder and decoder distortions for $X \sim \mathbb{N}(0,1)$ with $\eta_D = (x-y)^2$: (a,b) $\eta_E = (x^3-y)^2$ (c,d) $\eta_E = (1.5x-y)^2$

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. We also see that the encoder chooses a rate of $\log 3, \log 4, \log 5$ bits depending on θ. In Figure 2b, we present an example quantizer. The average distortions and the optimum decoder reconstruction for $i \in [1:M]$ are

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

$$y_i = \arg\min_{y \in \mathcal{Y}} c_1 \mathbb{E}\{\eta_s(x, \theta, y_i)\} + c_2 \int_{\theta \in \Theta} \int \eta_s(x, \theta, y_i) \mathrm{d}\mu_{X,\theta},$$

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

$$\overline{D_s} = \sum_{i=1}^{M} \int_{\theta \in \Theta} \int \eta_s(x, y_i, \theta) \mathrm{d}\mu_{X,\theta}.$$

C. 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, hence do not converge as illustrated in detail in [12]. A natural optimization approach would be taking the functional gradient i.e., perturbating 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. Instead, we perform gradient descent on the quantizer decision levels $\mathbf{q}$. One prominent problem in such optimization procedures is that if the cost surface is non-convex, which is the case here, the algorithm can be stuck at a poor local optima. As a simple remedy, we use multiple initializations and pick the best local optima among them. The algorithm is summarized below. The codes are made available at https://tinyurl.com/asilomar2023.

Function main()

Input: $\mu(\cdot), \mathcal{X}, M, \eta_E(\cdot,\cdot), \eta_D(\cdot,\cdot), p_b$
Output: q^*, $\mathbf{y}^*$, D_E, D_D
Initialization: $\{q_{init}\}, tol = 1, iter = 1, flag = 1$
Parameters: $\epsilon, \triangle$
$p_{err} \leftarrow 1 - (1-p_b)^{\log_2 M}$
$\mathbf{q} \leftarrow q_{init}$
$\mathbf{y} \leftarrow reconstruction(\mathbf{q}, \mu, p_{err}, \mathbb{E}\{X\})$
$flag \leftarrow 1$
while $flag \neq 0$ **do**
  $distenc \leftarrow distortion(\mathbf{q}, \mathbf{y}, \mu, \eta_E, E, p_{err})$
  for $i \in [1 : M-1]$ **do**
    $\Delta \leftarrow 1$
    $der_i \leftarrow derivative(\mathbf{q}, \mathbf{y}, \mu, i, p_{err})$
    $temp \leftarrow q_i - \triangle\, der$
    $\mathbf{qt} \leftarrow \mathbf{q}$
    $qt_i \leftarrow temp$
    $\mathbf{y} \leftarrow reconstruction(\mathbf{qt}, \mu, p_{err}, \mathbb{E}\{X\})$
    $dt \leftarrow distortion(\mathbf{qt}, \mathbf{y}, \mu, \eta_E, E, p_{err})$
    if $temp > q_{i-1} \,\&\&\, temp < q_i \,\&\&\, dt < distenc$ **then**
      $\mathbf{q} \leftarrow \mathbf{qt}$
    else
      while $\triangle >$ **do**
        $\triangle \leftarrow \triangle/10$
        $temp \leftarrow q_i - \triangle\, der$
        $\mathbf{qt} \leftarrow \mathbf{q}$
        $qt_i \leftarrow temp$
        $\mathbf{y} \leftarrow reconstruction(\mathbf{qt}, \mu, p_{err}, \mathbb{E}\{X\})$
        $dt \leftarrow encoderdistortion(\mathbf{qt}, \mathbf{y}, \mu, p_{err})$
        if $temp > q_{i-1} \,\&\&\, temp < q_i \,\&\&\, dt < distenc$ **then**
          $\mathbf{q} \leftarrow \mathbf{qt}$
          break
        end if
      end while
    end if
  end for

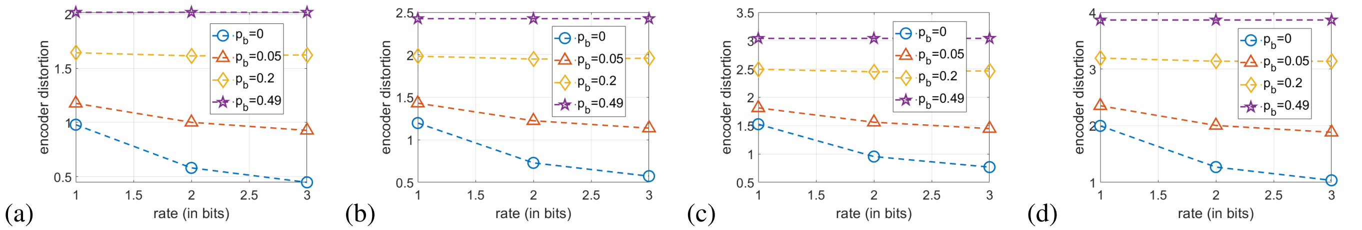


Fig. 4: Encoder distortions for jointly Gaussian (X, θ) with varying correlation: (a) $\rho = 0$ (b) $\rho = 0.2$ (c) $\rho = 0.5$ (d) $\rho = 0.9$

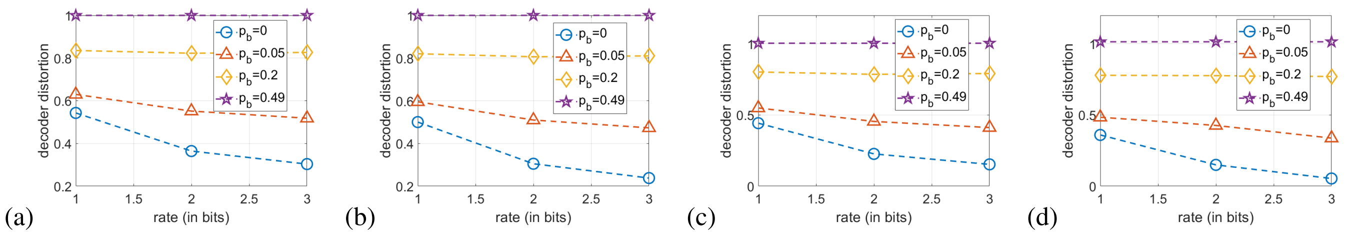
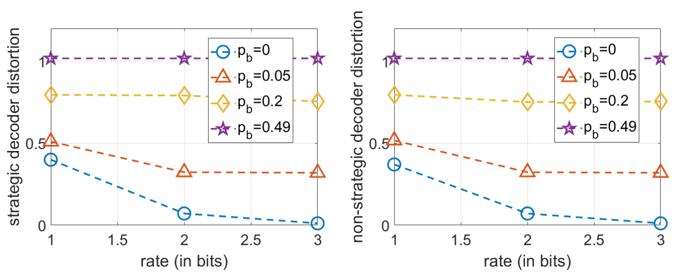


Fig. 5: Decoder distortions for jointly Gaussian (X, θ) with varying correlation: (a) $\rho = 0$ (b) $\rho = 0.2$ (c) $\rho = 0.5$ (d) $\rho = 0.9$



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

Fig. 6: Comparison of strategic decoder and non-strategic distortions for $\rho = 0.99$.

$\quad\mathbf{y} \leftarrow reconstruction(\mathbf{q}, \mu, p_{err}, \mathbb{E}\{X\})$
$\quad dt \leftarrow distortion(\mathbf{q}, \mathbf{y}, \mu, \eta_E, E, p_{err})$
$\quad$ **if** $iter > 1$ **then**
$\quad\quad$ **if** $all(\mathbf{der}) < \epsilon \,||\, dt == distenc$ **then**
$\quad\quad\quad flag = 0$
$\quad\quad$ **end if**
$\quad$ **end if**
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 reconstruction()

Input: $\mathbf{q}, \mu, p_{err}, \mathbb{E}\{X\}$
Output: $\mathbf{y}$
$c_1 \leftarrow p_{err}/(M-1)$
$c_2 \leftarrow 1 - Mc_1$
for $i \in [1 : M]$ **do**
$\quad y_i \leftarrow \frac{c_1 \mathbb{E}\{X\} + c_2 \int x \mathrm{d}\mu}{c_1 + c_2 \int \mathrm{d}\mu}$
end for

Function distortion()

Input: $\mathbf{q}, \mathbf{y}, \mu, \eta_s, s, p_{err}$
Output: D_s
Initialization: $D_s = 0$
$c_1 \leftarrow p_{err}/(M-1)$
$c_2 \leftarrow 1 - Mc_1$
for $i \in [1 : M]$ **do**
$\quad D_s \leftarrow D_s + c_1 \int_a^b \eta_s(x, y_i) \mathrm{d}\mu_X$
$\quad D_s \leftarrow D_s + c_2 \int_{x_{i-1}}^{x_i} \eta_s(x, y_i) \mathrm{d}\mu_X$
end for

IV. Numerical Results

Before delving into the details of simulations, we first note that our analysis holds for any general distortion measure. For numerical results, we adopt the widely used Mean Squared Error (MSE) metric as the decoder's distortion measure. 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 uninteresting solutions, such as non-revealing (where the encoder does not transmit any information) or fully-revealing (where the problem simplifies to non-strategic quantization).

We consider the following settings: $\eta_E = (x^3 - y)^2, \eta_E = (1.5x - y)^2, \eta_E(x, \theta, y) = (x + \theta - y)^2$. We take the source as Gaussian $X \sim \mathcal{N}(0, 1)$ for the former setting, jointly Gaussian

$$(X, \theta) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), 0 \leq \rho < 1$$

for the latter.

The support of the random variables is limited to $[-5, 5]$. We plot the encoder and decoder distortions associated with the above settings for bit error rate $p_b = [0, 0.05, 0.2, 0.49]$. We consider correlation coefficient $\rho = [0, 0.2, 0.5, 0.9]$ for the third setting.

In Figures 3, 4, and 5, we see 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 ($M = 1$, the encoder does not send any message). In the 2-dimensional case, 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 6. We observe that the encoder's distortion increases with correlation.

V. Conclusion

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

We note that the gradient descent solutions are in general only locally optimal, hence we ran the same algorithm with multiple initializations and picked the best solution, i.e., the one with the minimum encoder distortion among these local optima. In principle, this approach does not guarantee global optimality which can be achieved via dynamic programming solutions, as done in [11], [13], however at the cost of increased complexity.

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

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

Fig. 7: Communication diagram: (X, θ)

Appendix A
(X, θ) strategic quantization derivation

Let X take values from the source alphabet $\mathcal{X} \in [a_X, b_X]$, and θ take values from $\Theta \subseteq [a_\theta, b_\theta]$. 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 2 (Convex code-cells): $\mathcal{V}_{\theta,i}$ is convex for all $\theta \in \Theta, i \in [1:M]$.

Under assumption 2, $\mathcal{V}_{\theta,i}$ is an interval since X is a scalar, i.e.,

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

The encoder chooses the quantizer $\mathbf{q} = \{q_\theta, \theta \in \Theta\}$ with boundary levels $q_\theta = [x_{\theta,0}, x_{\theta,1}, \ldots, x_{\theta,M}]$ for each $\theta \in \Theta$. 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 response to $\mathbf{q}$ to minimize its cost $D_D = \mathbb{E}_\pi\{\mathbb{E}\{\eta_D(X, \theta, \mathbf{y})|\pi\}\}$ 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 $\mathcal{V}_{:,i} = \{\mathcal{V}_{\theta,i}, \forall \theta \in \Theta\}$. This setting is given in Figure 7. The integrals expressed in Appendices A and B are defined over the set $\mathcal{V}_{\theta,i}$ which is omitted for brevity, unless specified otherwise.

The encoder designs a set of quantizers $\mathbf{q}$ using only the objectives $(\eta_s, s \in \{E, D\})$, the statistics of the source $(\mu_{X,\theta}(\cdot,\cdot))$, the channel transition probability matrix $(p(z_j|z_i))$, 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

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

uses random index mapping to map z_m to z_i

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

where $\pi : \{1, \ldots, M\} \to \{1, \ldots, M\}$, and transmits z_i over a noisy channel. The message z_i is 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

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

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 $p(z_j|z_i)$, the channel transition probability.

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

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

The average distortion over all possible index assignments is

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

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

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

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

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

Similar to the 1-dimensional case, in the above equation, equality a is given by adding and subtracting $\sum_{i=1}^{M}\int\limits_{\theta\in\Theta}\int \eta_s(x,\theta,y_i(\mathbf{q}))\mathrm{d}\mu_{X,\theta}\Big)$, equality b is given by exchanging the summation over i and j and using the definition of $\mathbb{E}\{\eta_s(x,\theta,y_j(\mathbf{q}))\}$.

Then, the average distortions and the optimum decoder reconstruction for $i\in[1:M]$ are

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

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

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

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

The reconstruction levels $\mathbf{y}$ are found using the first order derivative condition:

$$\begin{aligned} \frac{\partial D_D}{\partial y_i} &= \frac{\partial}{\partial y_i}\Bigg(c_1\mathbb{E}\{\eta_D(x,\theta,y_i)\} \\ &\quad + c_2\int\limits_{\theta\in\Theta}\int \eta_s(x,\theta,y_i)\mathrm{d}\mu_{X,\theta}\Bigg). \end{aligned}$$

When the decoder distortion is measured via MMSE, i.e., $\eta_D(x,\theta,y) = (x-y)^2$,

$$\begin{aligned} \frac{\partial D_D}{\partial y_i} &= -2c_1\int\limits_{\theta\in\Theta}\int_{a_X}^{b_X}(x-y_i)\mathrm{d}\mu_{X,\theta} \\ &\quad - 2c_2\int\limits_{\theta\in\Theta}\int (x-y_i)\mathrm{d}\mu_{X,\theta}, \\ y_i &= \frac{c_1\mathbb{E}\{X\} + c_2\int\limits_{\theta\in\Theta}\int x\mathrm{d}\mu_{X,\theta}}{c_1 + c_2\int\limits_{\theta\in\Theta}\int \mathrm{d}\mu_{X,\theta}}, \end{aligned}$$

similar to Equation 1. The gradient of the encoder's distortion with respect to the decision levels are

$$\frac{\partial D_E}{\partial x_{\theta,i}} = c_1\frac{\partial}{\partial x_{\theta,i}}\mathbb{E}\{\eta_E(x,\theta,y_i(\mathbf{q}))\} + c_2\frac{\partial}{\partial x_{\theta,i}}\overline{D_E}.$$

While our analysis holds for general distributions of θ, to make the implementation of the algorithm tractable, θ has to be purely discrete. In case it is not, we approximate it by quantizing Θ to T points. Let $\Theta_s = \{\theta_1,\theta_2,\ldots,\theta_T\}$ be the ordered set of the quantized points with probability mass function $\{P_t\}$. The encoder designs a set of quantizers corresponding to each $\theta\in\Theta_s$. We carry out the analysis in discrete setting, which yields that the integrals $\int\limits_{\theta\in\Theta}$ transform into summations over $\theta_t\in\Theta_s$.

Appendix B
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**
$\Delta \leftarrow 1$
$der_{\theta,i} \leftarrow derivative(\mathbf{q}, \mathbf{y}, \mu_{X,\theta}, \theta, i, p_{err})$
$temp \leftarrow q_{\theta,i} - \triangle\, der_{\theta,i}$
$\mathbf{qt} \leftarrow \mathbf{q}$
$qt_{\theta,i} \leftarrow temp$
$\mathbf{y} \leftarrow reconstruction(\mathbf{qt}, \mu_{X,\theta}, p_{err}, \mathbb{E}\{X\})$
$dt \leftarrow distortion(\mathbf{qt}, \mathbf{y}, \mu_{X,\theta}, \eta_E, E, p_{err})$
if $temp > q_{\theta,i-1}$ && $temp < q_{\theta,i}$ && $dt < distenc$ **then**
$\mathbf{q} \leftarrow \mathbf{qt}$
else
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
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 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**
$D_s = D_s + c_2 \int\limits_{x_{\theta,end}} \eta_s(x, \theta, y_i) \mathrm{d}\mu_{X,\theta}(x, \theta)$
$D_s \leftarrow D_s + c_1 \int\limits_{x_{\theta,1}} \eta_s(x, \theta, y_i) \mathrm{d}\mu_{X,\theta}(x, \theta)$
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**
$dist_i(y) = c_1 \int\limits_{\theta \in \mathcal{T}} \int\limits_{x_{\theta,1}}^{x_{\theta,end}} \eta_s(x, \theta, y) \mathrm{d}\mu_{X,\theta} + c_2 \int\limits_{\theta \in \mathcal{T}} \int \eta_s(x, \theta, y) \mathrm{d}\mu_{X,\theta}$
end for
$y_i \leftarrow \arg\min\limits_{y \in \mathcal{Y}} dist_i(y)$
end for