Strategic Quantization over a Noisy Channel

Anonymous Authors
Please do NOT provide authors' names and affiliations
in the paper submitted for review, but keep this placeholder.

***Abstract*—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. We study the strategic quantization over a noisy channel, extending the classical work on channel-optimized quantization to strategic settings where the encoder and the decoder have misaligned objectives. Gradient descent is used to find the optimum quantization boundaries. For the noisy channel, a random index assignment mapping is implemented. Our analysis and numerical results demonstrate several interesting aspects of channel optimized strategic quantization which do not appear in its classical (nonstrategic) counterpart. The codes for all of the experiments in this paper are available at: https://tinyurl.com/isit2023**

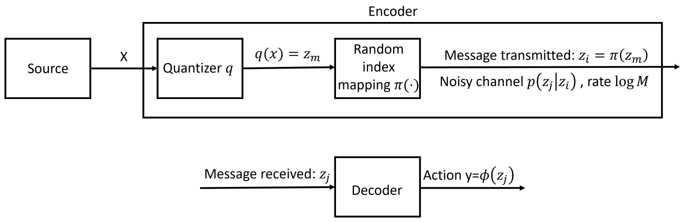


Fig. 1: Communication diagram

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 [**?**], where we study strategic quantizer design over a perfect (noiseless) communication channel (see also a recent related work [11]), we analyze and design channel-optimized strategic quantizer for two classes of distortion functions in the form of $2:1$ and $1:1$ mappings. Our main computational design tool here is gradient descent used in conjunction with random index assignment for the noisy channel. Compared to our recent related work on channel-optimized strategic quantization, the contribution of this paper is two-fold: a) We utilize a gradient-descent based optimization approach as opposed to dynamic programming, and b) we extend the analysis from scalar settings to $2:1$ distortion settings via a subterfuge by designing a set of quantizers.

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 an injective mapping, $q : \mathcal{X} \to \mathcal{Z}$. An index mapping $\pi : [1:M] \to [1:M]$ is chosen uniformly at random and is applied to the message Z. The message $\pi(Z)$ is transmitted over a noisy channel with transition probability matrix $p(j|i)$. 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 and takes an action $Y = \phi(Z)$. The encoder and decoder minimize their respective objectives $D_E = \mathbb{E}_\pi\{\mathbb{E}\{\eta_E(X,Y)|\pi\}\}$ and $D_D = \mathbb{E}_\pi\{\mathbb{E}\{\eta_D(X,Y)|\pi\}\}$, which are misaligned ($\eta_E \neq \eta_D$). The encoder designs 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 (q) are known to the encoder and the decoder. The problem is to design q for the equilibrium, i.e., the encoder minimizes its distortion if used with a corresponding decoder that minimizes its own distortion. This communication setting is given in Figure 1.

In this paper, we focus two special cases (with $\eta_D = (x-y)^2$ 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 channekreconstruction setting with the source as $(X, \theta) \sim \mu_{X,\theta}$. Specifically, we take a quadratic cost measure $\eta_E = (x + \theta - y)^2$.

Building on our recent under review paper [**?**], where we use dynamic programming solution concept for the same problem (quantizing a scalar source X with a noisy channel with rate constraint for polynomial distortion measures), here we use gradient descent to find the quantizer, and we further extend the discussion to a 2-dimensional vector source. The distortion measures are not constrained to be polynomial in this paper. 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 handle by using multiple initializations.

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 a index mapping $\pi : \{1, \ldots, M\} \to \{1, \ldots, M\}$ chosen uniformly at random, find the quantizer decision levels q, and actions $\mathbf{y}(q) = [y_1, \ldots, y_M]$ as a function of the set of quantizer decision levels that satisfy:

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

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

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 $i \in [1 : M]$.

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 the quantizer q with boundary levels $[x_{\theta,0}, x_1, \ldots, x_M]$. The decoder determines a set of actions, $\mathbf{y} = [y_1, \ldots, y_M]$ as the best response to q to minimize its cost D_D for $i \in [1 : M]$ as follows

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

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

The encoder designs a quantizer q with random index mapping using only the objectives (η_E, η_D), the statistics of the source $(\mu_X(\cdot))$, and the channel transition probability matrix $(p(j|i))$, 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 to map z_m to z_i

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

where $\pi : \{1, \ldots, M\} \to \{1, \ldots, M\}$. 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

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

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 $\pi(q(x)) = z_i$ is given by $p(j|i)$, the channel transition probability.

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

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

The average distortion over all possible index assignments is

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

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

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

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

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

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

In Equation 1, equality a is given by adding and subtracting $\sum_{i=1}^{M}\int \eta_s(x,y_i(q))\mathrm{d}\mu_X(x)\Big)$, equality b is given by exchanging the summation over i and j, equality c is given by the definition of $\mathbb{E}\{\eta_s(x,y_j(q))\}$.

The average distortion and the optimum decoder reconstruction for $i \in [1:M]$ are

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

$$y_i = \arg\min_{y\in\mathcal{Y}} c_1\mathbb{E}\{\eta_s(x,y_i(q))\} + c_2\overline{D_s},$$

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

$$\overline{D_s} = \sum_{i=1}^{M}\int \eta_s(x,y_i)\mathrm{d}\mu_X. \tag{2}$$

We note that in [?], 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.

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

$$\frac{\partial D_D}{\partial y_i} = c_1\frac{\partial}{\partial y_i}\mathbb{E}\{\eta_D(x,y_i(q))\} + c_2\frac{\partial}{\partial y_i}D_D. \tag{3}$$

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

$$\frac{\partial D_D}{\partial y_i} = -2c_1\sum_{m=1}^{M}\int (x-y_i)\mathrm{d}\mu_X - 2c_2\int (x-y_i)\mathrm{d}\mu_X(x), \tag{4}$$

$$y_i = \frac{c_1\mathbb{E}\{X\} + c_2\int x\mathrm{d}\mu_X}{c_1 + c_2\int \mathrm{d}\mu_X(x)}. \tag{5}$$

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

$$\frac{\partial D_E}{\partial x_i} = c_1\frac{\partial}{\partial x_i}\mathbb{E}\{\eta_E(x,y_i(q))\} + c_2\frac{\partial}{\partial x_i}D_E. \tag{6}$$

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. This setting is given in Figure 7. 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 3 without loss of generality. If the quantizer does not cross 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 3a, 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 3b, we present an example quantizer for $\Theta=\{-2,-1,0,1,2\}$. The derivation is given in Appendix A.

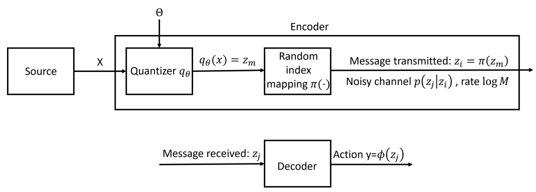
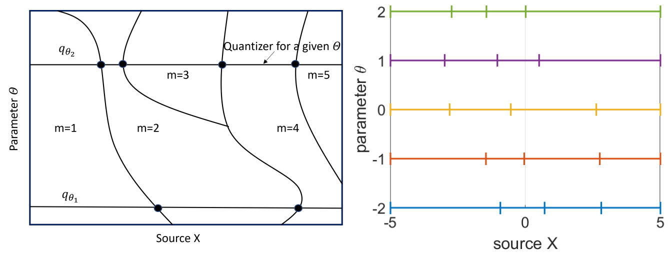


Fig. 2: Communication diagram: (X,θ)



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

Fig. 3: Quantization of X parameterized by θ.

B. Gradient descent algorithm

We first note the main difficulty associated with the design problem. The classical vector quantization solution relies on iterative Lloyd-MAx optimization, where the encoder and the decoder optimizes their mappings iteratively. Since our setup here is concerned with the Stackelberg equilibrium, i.e., a game problem as opposed to a team problem, such algorithms are not applicable. A natural optimization approach would be taking the functional gradient i.e., perturbating the quantizer mapping via an admissable perturbation function. However, the set of admissable functions have to be carefully chosen to satisfy the quantizer's properties (such as rate and convex codecell requirements) which hinders the tractibility of this more general functional optimization approach.

The problem setting requires the encoder to decide the set of quantizers q ($\mathbf{q}$ in case of 2-dimensional source) first, followed by the decoder computing its set of actions $\mathbf{y}$ as a response to q ($\mathbf{q}$), so q ($\mathbf{q}$) is the variable to be optimized. The encoder performs gradient descent on the set of quantizers q ($\mathbf{q}$). The algorithm is described below. The codes are made available at https://tinyurl.com/isit2023.

Function main():

Input: $\mu_X(\cdot), \mathcal{X}, M, \eta_E(\cdot,\cdot), \eta_D(\cdot,\cdot), p_b$
Output: q^*, $\mathbf{y}^*$, D_E, D_D
Initialization: $\{q_{init}(i), i\in[1:R]\}, tol=1, iter=1$
Parameters: $\epsilon, \triangle, R$
$p_{err} \leftarrow 1-(1-p_b)^{\log_2 M}$
$c_1 \leftarrow \frac{p_{err}}{M-1}$
$c_2 \leftarrow 1-Mc_1$
for $i\in[1:R]$ **do**
  $q^0 \leftarrow q_{init}(i)$
  $\mathbf{y} \leftarrow reconstruction(q^0,\mu_X)$
  $D_E^0 \leftarrow distortion(q^0, ym, \mu_X, \eta_E, E)$
  while $tol > \epsilon$ **do**

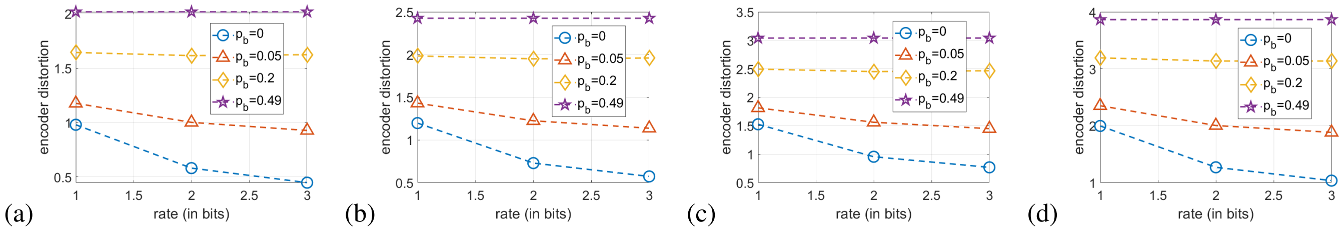


Fig. 4: Encoder distortions for jointly Gaussian (X, θ) with varying correlation: (a) $\rho = 0$ (b) $\rho = 0.2$ (c) $\rho = 0.5$ (d) $\rho = 0.9$

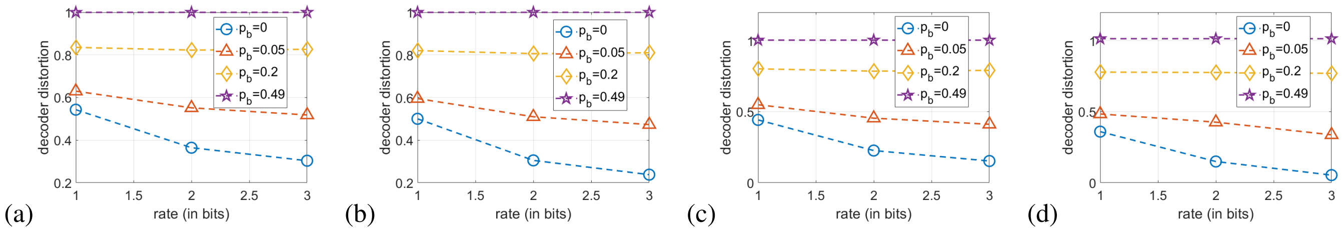


Fig. 5: Decoder distortions for jointly Gaussian (X, θ): (a) $\rho = 0$ (b) $\rho = 0.2$ (c) $\rho = 0.5$ (d) $\rho = 0.9$

  $q^{iter} \leftarrow q^{iter-1} - \triangle\{\frac{\partial D_E}{\partial x_{i,m}}\}$
  $\mathbf{y} \leftarrow reconstruction(q^{iter}, \mu_X, \eta_D)$
  $D_E^{iter} \leftarrow distortion(q^{iter}, \mathbf{y}, \mu_X, \eta_E, E)$
  $tol \leftarrow \frac{D_E^{iter} - D_E^{iter-1}}{D_E^{iter-1}}$
 end while
end for
$q^* \leftarrow q$
$\mathbf{y}^* \leftarrow \mathbf{y}$
$D_E^* \leftarrow D_E$
$D_D^* \leftarrow distortion(q^*, \mathbf{y}^*, \mu_X, \eta_D, D)$

Function distortion():

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

Function reconstruction():

Input: $q, \mu_X, \eta_D, p_{err}, c_1, c_2$
Output: $\mathbf{y}$
for $i \in [1 : M]$ **do**
 $y_i \leftarrow \frac{c_1 \mathbb{E}\{X\} + c_2 \int_{x_{i-1}}^{x_i} x \mathrm{d}\mu_X(x)}{c_1 + c_2 \int_{x_{i-1}}^{x_i} \mathrm{d}\mu_X(x)}$
end for

IV. Numerical Results

We consider the following settings:

A jointly Gaussian source $(X, \theta) \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_X^2 & \rho\\ \rho & \sigma_\theta^2\end{bmatrix}\right), 0 \leq \rho < 1$ and an encoder and decoder with distortion measures $\eta_E(x, \theta, y) = (x + \theta - y)^2, \eta_D(x, \theta, y) = (x - y)^2$.

We plot the encoder and decoder distortions associated with the above settings for correlation coefficient $\rho = [0, 0.2, 0.5, 0.9]$, bit error rate $p_b = [0, 0.05, 0.2, 0.49]$ in Figures 4 and 5 respectively. We observe that the encoder's distortion increases with correlation.

We observe in Figures 4, 5, and 6 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 8.

V. 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.

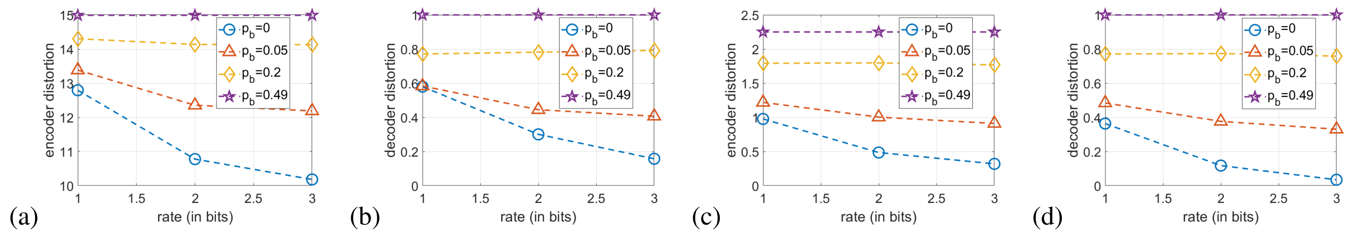


Fig. 6: Encoder and decoder distortions for Gaussian X with $\eta_D = (x-y)^2$: (a,b) $\eta_E = (x^3-y)^2$ (c,d) $\eta_E = (1.5x-y)^2$

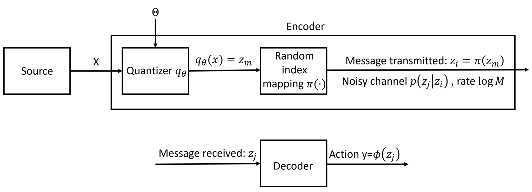
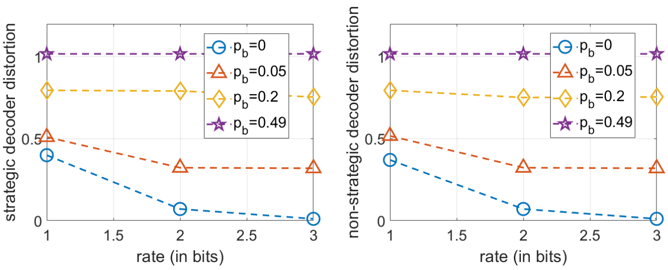


Fig. 7: Communication diagram: (X, θ)



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

Fig. 8: Comparison of strategic decoder and non-strategic distortions for $\rho = 0.99$.

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 (DCC)*. IEEE, 2022, pp. 173–181.

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 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 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, $\mathbf{y} = [y_1, \ldots, y_M]$ 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

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

where $\mathcal{V}_{:,i} = \{\mathcal{V}_{\theta,i}, \forall\theta \in \Theta\}$. 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 $(\mu_{X,\theta}(\cdot,\cdot))$, and the channel transition probability matrix $(p(z_j|z_i))$, 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 $\pi : \{1,\ldots,M\} \to \{1,\ldots,M\}$, 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

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

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

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

The average distortion over all possible index assignments is

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

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

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

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

$$\begin{aligned} I_{j\neq i} &\overset{a}{=} c_1 \Bigg(\sum_{i=1}^{M}\sum_{j=1}^{M} \int_\theta \int \eta_s(x,\theta,y_j(\mathbf{q})) \mathrm{d}\mu_{X,\theta} \\ &\quad - \sum_{i=1}^{M} \int_\theta \int \eta_s(x,\theta,y_i(\mathbf{q})) \mathrm{d}\mu_{X,\theta} \Bigg) \\ &\overset{b}{=} c_1 \Bigg(\sum_{j=1}^{M}\sum_{i=1}^{M} \int_\theta \int \eta_s(x,\theta,y_j(\mathbf{q})) \mathrm{d}\mu_{X,\theta} \\ &\quad - \sum_{i=1}^{M} \int_\theta \int \eta_s(x,\theta,y_i(\mathbf{q})) \mathrm{d}\mu_{X,\theta} \Bigg) \\ &\overset{c}{=} c_1 \Bigg(\sum_{j=1}^{M} \mathbb{E}\{\eta_s(x,\theta,y_j(\mathbf{q}))\} \\ &\quad - \sum_{i=1}^{M} \int_\theta \int \eta_s(x,\theta,y_i(\mathbf{q})) \mathrm{d}\mu_{X,\theta} \Bigg). \end{aligned} \tag{7}$$

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

Then, we have

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

$$y_i = \arg\min_{y \in \mathcal{Y}} c_1 \mathbb{E}\{\eta_s(x,\theta,y_i(\mathbf{q}))\} + c_2 \overline{D_s},$$

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

$$\overline{D_s} = \sum_{i=1}^{M} \int_\theta \int \eta_s(x, y_i, \theta) \mathrm{d}\mu_{X,\theta}. \tag{8}$$

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

$$\frac{\partial D_D}{\partial y_i} = c_1 \frac{\partial}{\partial y_i} \mathbb{E}\{\eta_D(x, \theta, y_i(\mathbf{q}))\} + c_2 \frac{\partial}{\partial y_i} D_D. \tag{9}$$

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

$$\begin{aligned} \frac{\partial D_D}{\partial y_i} = & -2c_1 \sum_{m=1}^{M} \int_\theta \int (x - y_i) \mathrm{d}\mu_{X,\theta} \\ & - 2c_2 \int_\theta \int (x - y_i) \mathrm{d}\mu_{X,\theta}, \end{aligned} \tag{10}$$

$$y_i = \frac{c_1 \mathbb{E}\{X\} + c_2 \int_\theta \int x \mathrm{d}\mu_{X,\theta}}{c_1 + c_2 \int_\theta \int \mathrm{d}\mu_{X,\theta}}, \tag{11}$$

similar to 11. The gradient of the encoder's distortion with respect to the decision levels are

$$\frac{\partial D_E}{\partial x_{\theta,i}} = c_1 \frac{\partial}{\partial x_{\theta,i}} \mathbb{E}\{\eta_E(x, \theta, y_i(\mathbf{q}))\} + c_2 \frac{\partial}{\partial x_{\theta,i}} D_E. \tag{12}$$

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 $\int_\theta$ transform into summations over $\theta_t \in \Theta_s$.

APPENDIX B
ALGORITHM: QUADRATIC

Function main():

Input: $\mu_{X,\theta}(\cdot), \mathcal{X}, \Theta_s, M, \eta_E, \eta_D, p_b$
Output: $\{q_\theta^*\}$, $\{y_m^*\}$, D_E, D_D
Initialization: $\{\mathbf{q}_{init}(i), i \in [1:R]\}, tol = 1, iter = 1$
Parameters: $\epsilon, \triangle, R$
$p_{err} \leftarrow 1 - (1-p_b)^{\log_2 M}$
$c_1 \leftarrow \frac{p_{err}}{M-1}$
$c_2 \leftarrow 1 - Mc_1$
for $i \in [1:R]$ **do**
  $\mathbf{q}^0 \leftarrow \mathbf{q}_{init}(i)$
  $\mathbf{y} \leftarrow reconstruction(\mathbf{q}^0, \mu_{X,\theta})$
  $D_E^0 \leftarrow distortion(\mathbf{q}^0, ym, \mu_{X,\theta}, \eta_E, E)$
  while $tol > \epsilon$ **do**
    $\mathbf{q}^{iter} \leftarrow \mathbf{q}^{iter-1} - \triangle\{\frac{\partial D_E}{\partial x_{i,m}}\}$
    $\mathbf{y} \leftarrow reconstruction(\mathbf{q}^{iter}, \mu_{X,\theta}, \eta_D)$
    $D_E^{iter} \leftarrow distortion(\mathbf{q}^{iter}, \mathbf{y}, \mu_{X,\theta}, \eta_E, E)$
    $tol \leftarrow \frac{D_E^{iter} - D_E^{iter-1}}{D_E^{iter-1}}$
  end while
end for
$\mathbf{q}^* \leftarrow \mathbf{q}$
$\mathbf{y}^* \leftarrow \mathbf{y}$
$D_E^* \leftarrow D_E$
$D_D^* \leftarrow distortion(\mathbf{q}^*, \mathbf{y}^*, \mu_{X,\theta}, \eta_D, D)$

Function distortion():

Input: $\mathbf{q}, \mathbf{y}, \Theta_s, \mu_{X,\theta}, \eta_s, s, c_1, c_2$
Output: D_s
Initialization: $D_s = 0$
for $i \in [1:M]$ **do**
  for $t \in [1:T]$ **do**
    $D_s = D_s + c_2 \int_{x_{\theta_t,i-1}}^{x_{\theta_t,i}} \eta_s(x, \theta_t, y_i) \mathrm{d}\mu_{X,\theta}(x, \theta_t)$
    for $j \in [1:M]$ **do**
      $D_s \leftarrow D_s + c_1 \int_{x_{\theta_t,i-1}}^{x_{\theta_t,i}} \eta_s(x, \theta_t, y_j) \mathrm{d}\mu_{X,\theta}(x, \theta_t)$
    end for
  end for
end for

Function reconstruction():

Input: $\mathbf{q}, \Theta_s, \mu_{X,\theta}, \eta_D, p_{err}, c_1, c_2$
Output: $\mathbf{y}$
for $i \in [1:M]$ **do**
  $y_i \leftarrow \frac{c_1 \mathbb{E}\{X\} + c_2 \sum_{t=1}^{T} \int_{x_{\theta_t,i-1}}^{x_{\theta_t,i}} x \mathrm{d}\mu_{X,\theta}(x,\theta_t)}{c_1 + c_2 \sum_{t=1}^{T} \int_{x_{\theta_t,i-1}}^{x_{\theta_t,i}} \mathrm{d}\mu_{X,\theta}(x,\theta_t)}$
end for