Strategic Quantization

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

Abstract

We analyze quantizer design when an encoder and a decoder with misaligned objectives communicate over a rate-constrained noiseless channel. The unconstrained version of this problem has been extensively studied under information design or Bayesian persuasion in Economics literature. However, this setting is more relevant to engineering applications when the cardinality of the message source is constrained. We motivate the problem with toy examples through which we explore the intricacies of strategic quantization, explicitly showing that the nature of the encoding policy may change with quantizer resolution, and performing an iterative optimization of the decoder and encoder mappings may not converge to a local optimum solution. As a remedy, we develop a gradient descent-based solution method instead. We analyze the poor local optima issues associated with the optimization method and further show that even for well-behaving sources, like Uniform, there are multiple local optima depending on the distortion measures chosen, in sharp contrast with classical quantization. We perform high-resolution analysis for this setting, and show that the encoder distortion may converge to a non-zero value, and that the encoder distortion may increase with rate for some cases. We finally present numerical results obtained via the proposed algorithm that suggests its validity and demonstrate strategic quantization features that differentiate it from its classical counterpart. The codes are available at: http://tinyurl.com/strategic-quantization.

Index Terms

Quantization, game theory, strategic communication.

I. Introduction

INFORMATION design or the Bayesian persuasion problem in Economics literature studies settings where a sender and a receiver with misaligned objectives communicate [1], [2]. These problems analyze how information can be used by a communication system designer (sender) to influence the action taken by the receiver. This class of problems have been an active research area in Economics due to their modeling abilities of real-life scenarios, see e.g., [1], [3]–[5]. Some applications that can be examined with this model are the design of transcripts when schools compete to improve their students' job prospects [5], and voter mobilization and gerrymandering [6], as well as various engineering applications, including in modeling misinformation spread over social networks [7], and privacy-constrained information processing [8], and many more [9]. For a comprehensive overview of the relevant literature in Economics, refer to [3], [10].

In this paper, we impose a cardinality constraint on the message space, which makes the problem more pertinent to engineering applications. This setup is called *strategic quantization*. The following is a practical problem that can be analyzed using the framework in this paper.

Consider the communication between two smart cars by competing manufacturers, e.g., Tesla and Honda, over a noiseless fixed-rate zero-delay channel. Tesla (decoder) asks for traffic congestion information from Honda (encoder) which is ahead in traffic to decide on changing its route or not. Honda's objective might be to make Tesla take a specific action, e.g., to change its route, while Tesla's objective is to make an accurate estimation of the traffic condition in order to take the correct decision. Since Honda's objective is different from Tesla's, Honda does not have an incentive to convey a truthful congestion estimation. However, Tesla is aware of Honda's motives but, if possible would still like to use Honda's information. How would these cars communicate? Problems of this nature can be handled using the strategic quantization model given in [11] and elaborated on in this paper. Note that here Honda has three different behavioral choices: it can choose not to communicate (non-revealing strategy), can communicate exactly what the Tesla wants (fully-revealing strategy), or it can craft a message that would make Tesla change its route (partially revealing strategy). Tesla can choose 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.

We now summarize some of the results on strategic quantization from Economics and Computer Science literature. In [12], the algorithmic complexity of finding the optimum strategic quantizer was shown to be NP-hard. The authors provide approximate results, in relation to another problem they solved conclusively. Aybaş and Türkel presented a constructive method for deriving optimal quantizers based on a search over possible posterior distributions over their feasible set in a recent working paper [13]. Using methods outlined in [2], they provided numerous theoretical properties of strategic quantization. In [14], [15], 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. 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 [16], to study the same problem via the engineering lens.

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

We note in passing that quantizers also arise as equilibrium strategies endogenously, i.e., without an external rate constraint, in a related but distinctly different class of signaling games, namely cheap talk, which has been introduced in the seminal work of Crawford and Sobel [17]. In [17], 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 [1], [2]), the encoder designs Q *ex-ante*, *before* seeing the source realization, and is committed to the designed Q afterward. This commitment is known to the decoder and establishes 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]

This paper is structured as follows. Section II describes the notation used and summarizes the prior work in strategic quantization. We state the problem formally in Section III. Section IV explores exciting aspects of strategic quantization. We analyze high-resolution quantization in Section V. We elaborate on the solution method in Section VI and provide the numerical results in Section VII, and Section VIII concludes. We show that the strategic quantizer might be the same as the classical quantizer with an example case in Appendix A, provide the proof for the monotonicity of the optimal quantizer for a specific type of encoder distortion in Appendix B, the gradients used in the proposed algorithm in Appendix C, proof for a theorem about a setting with quadratic costs in Appendix D, proof for the decoder's optimality in accepting encoder's message in Appendix E, a result on the form of the encoder distortion if the encoder is fully revealing in Appendix F, the high-resolution approximation of the reconstruction levels in Appendix G, and the high-resolution analysis for two sets of encoder distortion functions with decoder distortion as a Bregman loss function in Appendices H, I.

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 ($\mathcal{X}$). We denote the cumulative distributive function of a random variable X as $P_X(x)$, and the probability density function of a continuous random variable X as $p_X(x)$. The set of real numbers, non-negative real numbers, and natural numbers are denoted by $\mathbb{R}$, $\mathbb{R}_{\geq 0}$ and $\mathbb{Z}_{>0}$, 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 $[t_1, t_2]$ and the scalar Gaussian with mean μ, variance σ^2 are denoted by $U[t_1, t_2]$ and $\mathcal{N}(\mu, \sigma^2)$ respectively. The expression $t_1 \leq i \leq t_2, i \in \mathbb{Z}_{>0}$ 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 $Q : \mathcal{X} \to \mathcal{Z}$, with $|\mathcal{Z}| \leq M$ for a given quantization resolution $M \in \mathbb{Z}^+$, and given distortion measures D_E, D_D.

In this paper, building on our preliminary results [11], we show that a naive strategic variation of the Lloyd-Max algorithm may not converge to a locally optimal solution. As a remedy, we develop a gradient descent-based solution for this design problem. We then demonstrate that even for well-behaving sources, such as scalar Uniform, there are multiple local optima, depending on the distortion measures chosen, in sharp contrast with classical quantization for which the local optima is unique for the case of log-concave sources (which include Uniform sources). We also analyze the behavior of the optimal strategic quantizer for some distortion measures. The behavior can be one of the following three:

1) *Non-revealing*: the encoder does not send any information, i.e., $Q(X) =$ constant. This case happens if the encoder does not benefit from sending information to the decoder, i.e., the decoder takes action without the information from the encoder. In the particular case where the decoder objective is to minimize $\mathbb{E}\{(X - \hat{X})^2\}$, the encoder will be non-revealing, if
$$\mathbb{E}\{\eta_E(X, \mathbb{E}\{X\})\} \leq \mathbb{E}\{\eta_E(X, T(X)\}$$
for any non-injective mapping $T : X \to Z$.
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. When encoder and decoder distortions are $\eta_E(x, y) = |x - y|, \eta_D(x, y) =$

[1] These issues are well understood in the Economics literature, see, e.g., [4] 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.

Problem. *For a given rate R, scalar source X with a probability distribution function P_X, find the decision boundaries $\mathbf{q} = [x_0, x_1, \ldots, x_M]$ and actions $\mathbf{y}(\mathbf{q}) = [y_1, \ldots, y_M]$ as a function of boundaries that satisfy:*

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

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

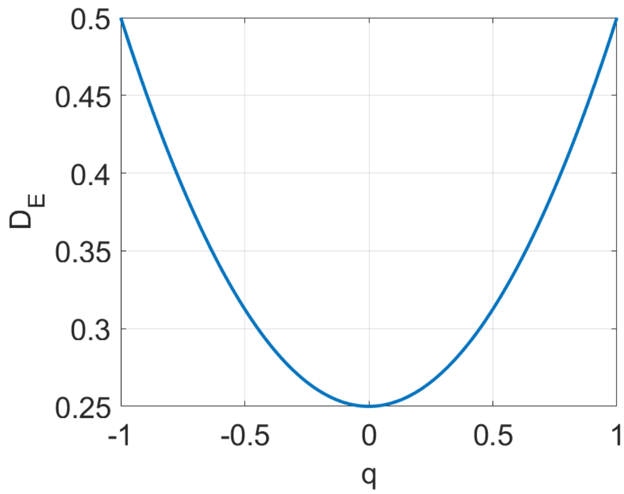


Fig. 1. Cost surface of the encoder distortion for 2-level quantization, with decision boundary q, of $X \sim U[-1, 1]$ with $\eta_D(x, y) = (x - y)^2$ and $\eta_E(x, y) = |x - y|$.

$(x - y)^2$, respectively, for a uniform source X, the encoder is fully revealing even though their distortion measures are different because they are closely aligned, as we show in Appendix A. In Fig. 1 we show the cost surface of the encoder distortion for $M = 2$ level quantization, which has the same minima as the non-strategic setting with $\eta_E(x, y) = \eta_D(x, y) = (x - y)^2$.

3) *Partially revealing*: The encoder sends some information but not exactly what the decoder wants.

In [18], [19], 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. In our recent preliminary work [19], [20], we carry 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 [21], we analyzed the problem of strategic quantization of a noisy source. We also explored 2-dimensional strategic quantization in [22].

III. Problem Definition

An encoder observes a realization of a scalar source $X \in \mathcal{X}$ with probability distribution P_X. The encoder maps X to a message $Z \in \mathcal{Z}$, where $\mathcal{Z}$ is a set of discrete messages with a cardinality constraint $|\mathcal{Z}| \le M$ using a non- injective mapping, $Q : \mathcal{X} \to \mathcal{Z}$. 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 $D_E = \mathbb{E}\{\eta_E(X, Y)\}$ and $D_D = \mathbb{E}\{\eta_D(X, Y)\}$, which are misaligned ($\eta_E \ne \eta_D$). 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 $P_X(\cdot)$. The objectives (η_E and η_D), the shared prior (P_X), and the mapping (Q) are known to the encoder and the decoder. The problem, formally described in the box, 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 Fig. 2.

The set $\mathcal{X}$ is partitioned into mutually exclusive and exhaustive sets $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_M$. The decoder's action, y_m, can be written as a quantization operator

$$y_m = Q(x) \quad \forall x \in \mathcal{V}_m$$

for all $m \in [1 : M]$. Throughout this paper, we make the following "monotonicity" assumption on the sets $\{\mathcal{V}_m\}$.

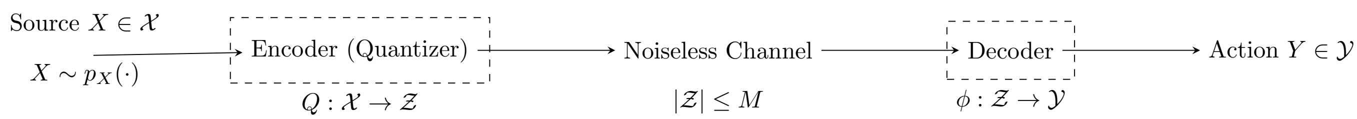


Fig. 2. Communication diagram over a noiseless channel.

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

Remark 1. *Assumption 1 is the first of the two regularity conditions commonly employed in the classical quantization literature, cf. [23]. Note that the second regularity condition, $y_m \in \mathcal{V}_m$, is not included in Assumption 1.*

In the Economics literature, Assumption 1 is referred as the "monotonicity" condition. In [14], [15], the sufficient conditions on η_E and η_D for the monotonicity of optimal encoder strategies are characterized within the unconstrained (without quantization) variation of the same problem. We note that here we have the quantization constraint in the problem formulation as an exogenous constraint on the message set, hence it is not clear apriori whether the results in [14], [15] would be applicable here.

Under Assumption 1, $\mathcal{V}_m$ is an interval, i.e.,

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

where $a = x_0 < x_1 < \ldots < x_M = b$. The encoder chooses the boundary decision levels $\mathbf{q} = [x_0, x_1, \ldots, x_M]$. The decoder determines its actions $\mathbf{y} = [y_1, \ldots, y_M]$ as a best response to $\mathbf{q}$ to minimize its cost $D_D = \mathbb{E}\{\eta_D(x, y)\}$ as follows

$$y_m^* = \underset{y_m \in \mathcal{Y}}{\arg\min}\, \mathbb{E}\{\eta_D(x, y_m) | x \in \mathcal{V}_m\} \quad \forall m \in [1 : M].$$

Hence, the decoder chooses the actions $\{y_m\}$ knowing the set of decision regions $\{\mathcal{V}_m\}$. The encoder computes what the decoder would choose as $\mathbf{y}$ given $\{\mathcal{V}_m\}$, and hence optimizes its own cost $\mathbb{E}\{\eta_E(x, y)\}$ over the choice of $\{\mathcal{V}_m\}$ accordingly:

$$\{\mathcal{V}_m^*\} = \underset{\{\mathcal{V}_m\}}{\arg\min} \sum_{m=1}^{M} \mathbb{E}\{\eta_E(x, y_m(\{\mathcal{V}_m\})) | x \in \mathcal{V}_m\},$$

or due to Assumption 1 equivalently over the choice of $\mathbf{q}$:

$$\mathbf{q}^* = \underset{\mathbf{q}}{\arg\min} \sum_{m=1}^{M} \mathbb{E}\{\eta_E(x, y_m(\mathbf{q})) | x \in \mathcal{V}_m\}.$$

Remark 2. *An important factor to be considered here is that the encoder is committed to its choice of $\mathbf{q}$, it cannot determine $\mathbf{q}$ as the best response to the action taken by the decoder $y \in \mathcal{Y}$. This implies that the encoder cannot optimize $\mathbf{q}$ as the best response to y, but the decoder can optimize its action y as the best response to $\mathbf{q}$. The encoder instead optimizes $\mathbf{q}$ as the best response to a function of itself, $\mathbf{y}(\mathbf{q})$. This nature of the problem introduces a hierarchy in the gameplay (the encoder plays first, and the decoder responds, which 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) and naturally is not a Nash equilibrium since $\mathbf{q}$ may not be the best response to $\mathbf{y}$. In cheap talk [17], Nash equilibria are sought after and the equilibria achieving strategies happen to be non-injective mappings, i.e., quantizers, without an exogenous rate constraint. It is essential to note the substantial difference between the problem formulation in this paper and the cheap talk literature [17].*

The encoder and decoder distortions are given by

$$D_s = \sum_{m=1}^{M} \mathbb{E}\{\eta_s(x, y_m) | x \in \mathcal{V}_m\}, \quad s \in \{E, D\}. \tag{1}$$

We formally define a fully revealing encoder as follows. The quantizer $\mathbf{q}_{fr} = [x_{fr,0}, \ldots, x_{fr,M}]$, $\mathcal{V}_{fr,m} = [x_{fr,m-1}, x_{fr,m})$ is obtained as:

$$\mathbf{q}_{fr}^* = \underset{\mathbf{q}}{\arg\min} \sum_{m=1}^{M} \mathbb{E}\{\eta_D(x, y_{fr,m}(\mathbf{q})) | x \in \mathcal{V}_{fr,m}\}, \tag{2}$$

where $y_{fr,m}$ is given by

$$y_{fr,m} = \underset{y}{\arg\min}\, \mathbb{E}\{\eta_D(x, y) | x \in \mathcal{V}_{fr,m}\}. \tag{3}$$

The fully revealing encoder distortion D_{fr} is the encoder distortion when the quantizer implemented is the fully revealing quantizer, $\mathbf{q}_{fr}$,

$$D_{fr}(\eta_E(x,y)) = \sum_{m=1}^{M} \mathbb{E}\{\eta_E(x, y_{fr,m}) | x \in \mathcal{V}_{fr,m}\}. \tag{4}$$

The encoder and decoder distortions for a non-revealing encoder ($M = 1$) are $D_{s,nr}$,

$$D_{s,nr} = \int_a^b \eta_s(x, y_{nr}) \mathrm{d}P_X, \quad s \in \{E, D\},$$

where y_{nr} is given by

$$y_{nr} = \min_{y \in \mathcal{Y}} \int_a^b \eta_D(x,y) \mathrm{d}P_X.$$

Remark 3. *We assume a scalar source X with cumulative distribution $P_X(x)$ generally. We make a note for the specific cases where we assume a continuous source, and we consider the density function $p_X(x) = \mathrm{d}P_X(x)/\mathrm{d}x$.*

IV. Interesting Aspects of Strategic Quantization

We present below simple numerical examples that effectively illustrate several intricacies of the strategic quantization problem, highlighting the distinctions from its classical counterpart.

A. Monotonicity

Monotonicity property of the optimal encoding map has been an active research question in Economics, see e.g, [] , for the general Bayesian Persuasion problem. In the following, we present sufficient conditions for the monotonicity of the optimal quantizer, and show that the examples considered in this paper satisfy these sufficient conditions.

We prove the following theorem which provides sufficient conditions for monotonicity of the optimal quantizer in Appendix B.

Theorem 2. *The structure of the optimal quantizer is monotonic, that is, the quantizer regions are convex code cells, $x_{m-1} < x_m, m \in [1:M]$, when the encoder distortion is of the form $\eta_E(x,y) = l(x) f(|g(x) - h(y)|)$, where $l(\cdot) : \mathbb{R} \to \mathbb{R}_{\geq 0}$, $f(\cdot) : \mathbb{R} \to \mathbb{R}$ is increasing, $g(\cdot) : \mathbb{R} \to \mathbb{R}$ is monotonic, and $h(\cdot)$ is bijective. Note that there are no constraints on the decoder distortion $\eta_D(x,y)$.*

In this paper, we consider specific encoder distortions of the form $\eta_E(x,y) = (x^3 - y)^2$, $\eta_E(x,y) = (x - y^3)^2$ since they satisfy the sufficient conditions of monotonicity in contrast with encoder distortions $\eta_E(x,y) = (x - y^2)^2$, $\eta_E(x,y) = (x^2 - y)^2$, which we show below and formally state in Corollary 2.1.

The encoder objectives $\eta_E(x,y) = (x - y^3)^2, \eta_E(x,y) = (x^3 - y)^2, \eta_E(x,y) = (x - y^2)^2, \eta_E(x,y) = (x^2 - y)^2$ can be written in the form given in Theorem 2 as:

$\eta_E(x,y)$	$l(x)$	$g(x)$	$h(y)$	$f(\lvert g(x) - h(y)\rvert)$
$(x - y^3)^2$	1	x	y^3	$\lvert g(x) - h(y)\rvert^2$
$(x^3 - y)^2$	1	x^3	y	$\lvert g(x) - h(y)\rvert^2$
$(x - y^2)^2$	1	x	y^2	$\lvert g(x) - h(y)\rvert^2$
$(x^2 - y)^2$	1	x^2	y	$\lvert g(x) - h(y)\rvert^2$

From the above table, we note that for $\eta_E(x,y) = (x - y^2)^2$, $h(y) = y^2$ is not bijective, and for $\eta_E(x,y) = (x^2 - y)^2$, $g(x) = x^2$ is not monotonic, and hence they do not satisfy the sufficient conditions in Theorem 2. The functions $l(\cdot), g(\cdot), h(\cdot), f(\cdot)$ for $\eta_E(x,y) = (x - y^3)^2, \eta_E(x,y) = (x^3 - y)^2$ satisfy the sufficient conditions in Theorem 2.

Corollary 2.1. *The sufficient conditions for the monotonicty of the optimal quantizer in Theorem 2 are satisfied by the encoder objectives $\eta_E(x,y) = (x - y^3)^2, \eta_E(x,y) = (x^3 - y)^2$, and are not satisfied by the encoder objectives $\eta_E(x,y) = (x - y^2)^2, \eta_E(x,y) = (x^2 - y)^2$. Hence, the resulting optimal quantizer for $\eta_E(x,y) = (x - y^2)^2, \eta_E(x,y) = (x^2 - y)^2$ may or may not be monotonic, while that for $\eta_E(x,y) = (x - y^3)^2, \eta_E(x,y) = (x^3 - y)^2$ are monotonic.*

Remark 4. *Note that the distortion function expression in Theorem 2 also includes $\eta_E(x,y) = (x + c_1 - c_2 y)^2$, for $c_1, c_2 \in \mathbb{R}$ as a special case.*

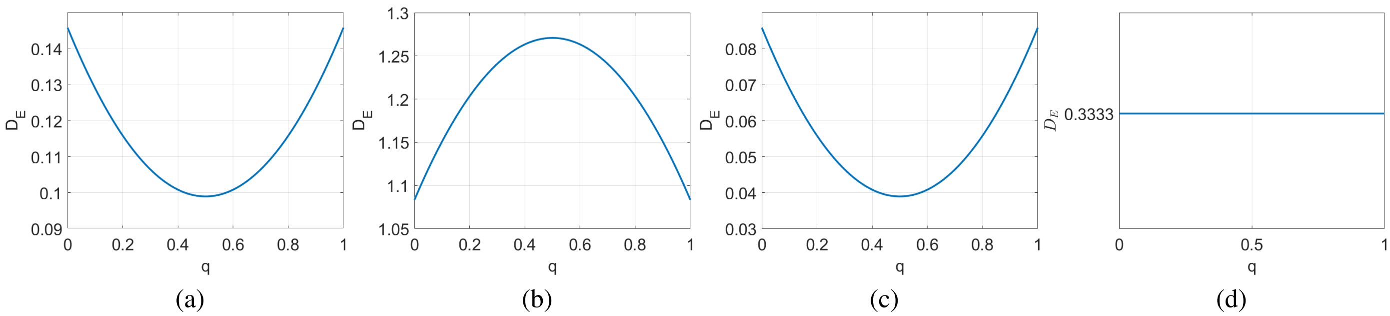


Fig. 3. Cost surface of the encoder distortion for 2−level quantization of $X \sim U[0,1]$ as $\mathbf{q} = \begin{bmatrix} a & q & b \end{bmatrix}$ with decoder distortion $\eta_D(x,y) = (x-y)^2$ and encoder distortion (a) $\eta_E(x,y) = (x-1.5y)^2$, (b) $\eta_E(x,y) = (x-3y)^2$, (c) $\eta_E(x,y) = (x+0.2-1.5y)^2$, (d) $\eta_E(x,y) = (x-2y)^2$.

B. To reveal or not to reveal... or partially reveal?

Classical quantization is a team problem where the encoder and the decoder share identical objectives, while strategic quantization is a game problem where the encoder and decoder objectives are misaligned. Depending on the misalignment between η_E and η_D, the encoding policy may be one of the following types. If the objectives are too misaligned, the encoder chooses not to send anything - a "non-revealing" policy in Economics conventions. The encoding policy may be the same as classical quantization with a common distortion η_D, i.e., the encoder may not be able to utilize the information design advantage to persuade the decoder to take a specific action, termed as "fully revealing". Alternatively, the quantizer may not be identical to the classical quantizer, a "partially revealing" policy. A simple example of a non-revealing policy is with $\eta_E(x,y) = -\eta_D(x,y)$, a zero-sum game, which results in the optimal strategy for the encoder as not sending any information. We illustrate the other cases via numerical examples given below.

Consider a continuous scalar source X, and an encoder and decoder with distortion measures $\eta_E(x,y) = (x+\alpha-\beta y)^2$ and $\eta_D(x,y) = (x-y)^2$, for given $\alpha, \beta \in \mathbb{R}$. 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 $\frac{X+\alpha}{\beta}$, both in the mean squared error (MSE) sense. Can the encoder "persuade" the decoder by carefully designing quantizer intervals $\mathcal{V}_m^*$?

In Appendix D, we prove the following result which is an extension of a result presented in [11], as well as in [14]:

Theorem 3. *For $\eta_E(x,y) = (x+\alpha-\beta y)^2$ and $\eta_D(x,y) = (x-y)^2$, $\alpha, \beta \in \mathbb{R}$, the optimal strategic quantizer Q of a continuous source X is:*

$$Q^*(x) = \begin{Bmatrix} \underset{Q}{\arg\min}\, \mathbb{E}\{(X-Q(X))^2\}, & \textit{for } 0 < \beta < 2 \\ \textit{arbitrary}, & \textit{for } \beta = 0, 2 \\ \textit{constant}, & \textit{otherwise} \end{Bmatrix}.$$

Note that the first case corresponds to the fully-revealing behavior, while the second one corresponds to encoder distortion remaining constant for all quantizers, and the third is non-revealing.

We see from Theorem 3 that for $0 < \beta < 2$, the encoder behaves like a non-strategic quantizer with a fully revealing strategy, i.e., the strategic quantizer simplifies to its classical counterpart where both the encoder and the decoder have identical objectives. However, when $\beta > 2$, it is non-revealing, and the quantizer is arbitrary for $\beta = 2$. Fig. 3 shows the cost curves of various encoder distortions $\eta_E(x,y) = (x-1.5y)^2, (x-3y)^2, (x+0.2-1.5y)^2, (x-2y)^2$ for decoder distortion $\eta_D(x,y) = (x-y)^2$ in quantizing a Uniform source $X \sim U[0,1]$ to $M = 2$ levels. We see in Figures 3a,b that $\beta = 1.5$ has an optima only for the quantizer $\begin{bmatrix} 0 & 0.5 & 1 \end{bmatrix}$ which is the non-strategic quantizer, and that $\beta = 3$ is optimal when the quantizer is $\begin{bmatrix} 0 & 1 \end{bmatrix}$ (non-revealing). Fig. 3c shows the encoder's distortion when $\eta_E(x,y) = (x+0.2-1.5y)^2$ and $\eta_D(x,y) = (x-y)^2$. We see that the optimal point does not change although the distortion values are shifted. Equation 21 in Appendix D shows that the terms containing α in the encoder's distortion are: $\alpha^2 + 2\alpha(1-\beta)\mathbb{E}\{X\}$. As seen from Figures 3a,c, this evaluates to -0.06 for our example. Fig. 3d shows the encoder's distortion for $\beta = 2$, for which setting we see from Equation 22 in Appendix D that all quantizers have the same distortion, equal to the non-revealing distortion, given in Equation 20.

We consider the quantization of a continuous source X with encoder and decoder distortion functions $\eta_E(x,y) = (f(x)-y)^2, \eta_D(x,y) = (x-y)^2$, and conclude the structure of $f(x)$ as affine from the optimality of the fully revealing strategy for the encoder in the following theorem. The proof is given in Appendix F.

Theorem 4. *If the strategic quantization of a continuous source X with encoder and decoder distortions $\eta_E(x,y) = (f(x)-y)^2, \eta_D(x,y) = (x-y)^2$ is fully revealing, $\mathbf{q}^*_{fr}$, then $f(x) = t_1x + t_2, t_1, t_2 \in \mathbb{R}$.*

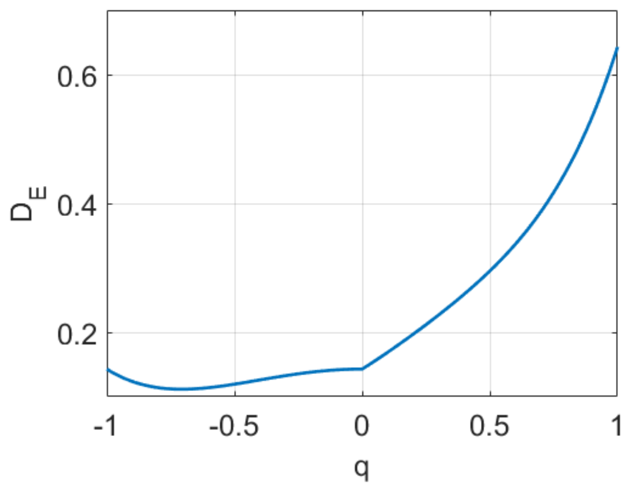


Fig. 4. Cost surface of the encoder distortion for 3-level quantization, with decision boundary q, of $X \sim U[-1,1]$ with $\eta_D(x,y) = (x-y)^2$ and $\eta_E(x,y) = (x^3-y)^2, xy \geq 0; \to \infty$ otherwise.

C. Quantizer resolution is binding

We now consider the question: does the nature of optimal encoding policy change with the quantization constraint in strategic communication? That is, is there a case where the encoder is non-revealing for $M=2$ but prefers to send a message for $M=3$? The answer is, perhaps surprisingly, affirmative.

Consider a continuous source $X \sim U[-1,1]$ with the distortion functions $\eta_D(x,y) = (x-y)^2$ and

$$\eta_E(x,y) = \begin{cases} (x^3-y)^2, & xy \geq 0 \\ \to \infty, & \text{otherwise.} \end{cases}$$

A fully non-revealing policy, i.e., the case of $R=0$ ($M=1$) yields $D_E(0) = \int_{-1}^{1} (t^3)^2 \frac{1}{2}\, dt = 1/7$.

We next consider $M=2$. From symmetry, the optimal encoding policy is simply setting the boundary at $q_1 = 0$. However, this yields $y_1 = -y_2 = -1/2$ and $D_E(1) = 1/7$, which is identical to $D_E(0)$. Hence, optimal strategic quantizer for $M=2$ does not send any information to the decoder, i.e., non-revealing.

We finally consider the case of $M=3$. Similar to the previous case, we parametrize $\mathcal{V}_m^*$ as $[-1,q], (q,0], (0,1]$ (parameterizing $\mathcal{V}_m^*$ as $[-1,0], (0,q], (q,1]$ results in an infeasible solution) and express $\{y_m\}$ as a function of q, which yields

$$y_1 = \frac{-1+q}{2}, \quad y_2 = \frac{q}{2}, \quad y_3 = \frac{1}{2}.$$

Substituting again $\{y_m\}$ in $\eta_E(x,y)$:

$$J(q) = \frac{1}{2}\left(\int_{-1}^{q} (t^3 - y_1)^2\, dt + \int_{q}^{0} (t^3 - y_2)^2\, dt + \int_{0}^{1} (t^3 - y_3)^2\, dt \right).$$

Enforcing the KKT conditions similar to (19), we obtain $q = 0, \pm\sqrt{1/2}$.[2] Since $q \leq 0$, the possible solutions are $q = 0, -\sqrt{1/2}$. Of these two choices, $q = -\sqrt{1/2}$ is the local minima (as seen in Fig. 4) yielding a distortion $D_E = 25/224$, hence is optimal. Here, the encoder uses a different quantizer, that is, the encoder does not reveal information exactly as the decoder would want, and is partially revealing.

We note here that $D_E(\log 3) = 25/224 < 1/7 = D_E(1) = D_E(0)$. Hence, at $M=2$, the strategic quantizer does not communicate any information while, at $M=3$, in sharp contrast to $M=2$, uses the quantization channel fully to send three messages, demonstrating that the quantization constraint can change the nature of the optimum encoder policy. Moreover, it shows that the operational rate-distortion function $D_E(R)$ here is not a strictly decreasing function of rate R, since $D_E(1) = D_E(0)$, unlike its classical counterpart.

D. Failure of "Strategic Lloyd-Max"

We now consider whether a simple strategic variation of the Lloyd-Max approach, based on iterative inposition of necessary conditions of optimality on one mapping whule keeping the other one fixed, results in a locally optimal quantizer, as is known

[2] There is an error in our prior work [11] in the values of q.

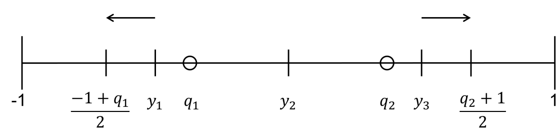


Fig. 5. Movement of the quantization boundaries through Lloyd-Max iterations in the running example.

for non-strategic quantization. We demonstrate here that such a naive "strategic Lloyd-Max" algorithm may not even converge via the following example.

Let the source $X \sim U[-1,1]$ be quantized to $M=3$ levels where encoder and decoder distortions functions are $\eta_E(x,y) = (x^3-y)^2$ and $\eta_D(x,y)=(x-y)^2$,respectively.

We then initialize $\{y_m\}$ arbitrarily and iterate steps 1 and 2 below until convergence:

1) For a given y_1 and y_2, find q_1 and q_2 that minimize $D_E = \mathbb{E}\{(x^3-y)^2\}$ as

$$q_1 = \left(\frac{y_1+y_2}{2}\right)^{\frac{1}{3}}, \quad q_2 = \left(\frac{y_2+y_3}{2}\right)^{\frac{1}{3}}. \tag{5}$$

2) For a given q_1 and q_2, find $\{y_m\}$ that minimize $D_D = \mathbb{E}\{(x-y)^2\}$ as

$$y_1 = \frac{\int_{-1}^{q_1}\frac{1}{2}t\ dt}{\int_{-1}^{q_1}\frac{1}{2}\ dt} = \frac{-1+q_1}{2}, y_2 = \frac{\int_{q_1}^{q_2}\frac{1}{2}t\ dt}{\int_{q_1}^{q_2}\frac{1}{2}\ dt} = \frac{q_1+q_2}{2}, y_3 = \frac{\int_{q_2}^{1}\frac{1}{2}t\ dt}{\int_{q_2}^{1}\frac{1}{2}\ dt} = \frac{q_2+1}{2}, \tag{6}$$

We note that during these iterations, q_1 and q_2 move towards -1 and 1, respectively, i.e., the boundaries move towards the endpoints of the interval taken for quantization with each iteration as demonstrated in Fig. 5, hence the iterations converge to $M=1$ solution (non-revealing) with $D_E = 1/7$. Now, let us examine whether this solution is a local optimum.

Any admissible perturbation of with some $0<\epsilon<1$ of $q_1 = -q_2 = -1$ would result in a $M=3$ level quantizer with decision boundaries $q_1 = -1+\epsilon$, $q_2 = 1-\epsilon$ with the corresponding decoder actions $y_1 = -1+\frac{\epsilon}{2}$, $y_2 = 0$, $y_3 = 1-\frac{\epsilon}{2}$, yielding $D_E = 1/7 - \epsilon(1-\frac{\epsilon}{2})^2(1-\epsilon)^2$ which is smaller than than that of the non-revealing solution ($D_E = 1/7$), hence this is not a locally optimal solution. This observation indicates that the straightforward enforcement of optimality conditions may not yield a locally optimal solution, which contrasts sharply with the case in classical quantization. In other words, unlike its classical counterpart, a trivial extension of the Lloyd-Max algorithm adopted for strategic settings may not converge to a locally optimum solution.

E. Multiple Local Minima

For log-concave scalar sources, it is a well-known result in the literature on classical quantization that the local minima correspond with the global one [24], [25]; as a result, Lloyd-Max is guaranteed to converge to the globally optimal solution. The question that naturally arises from this is whether the same conclusion holds true in the strategic domain. We infer the solution to this inquiry via a numerical counter-example: Consider a Uniform scalar source $X \sim U[a,b]$ with $\eta_E(x,y) = (x^3-y)^2, \eta_D(x,y) = (x-y)^2$ quantized to $M=2$ levels as $\mathbf{q} = \begin{bmatrix} a & q & b \end{bmatrix}$. The decoder actions y_1, y_2,

$$y_1 = \frac{a+q}{2}, \quad y_2 = \frac{q+b}{2}.$$

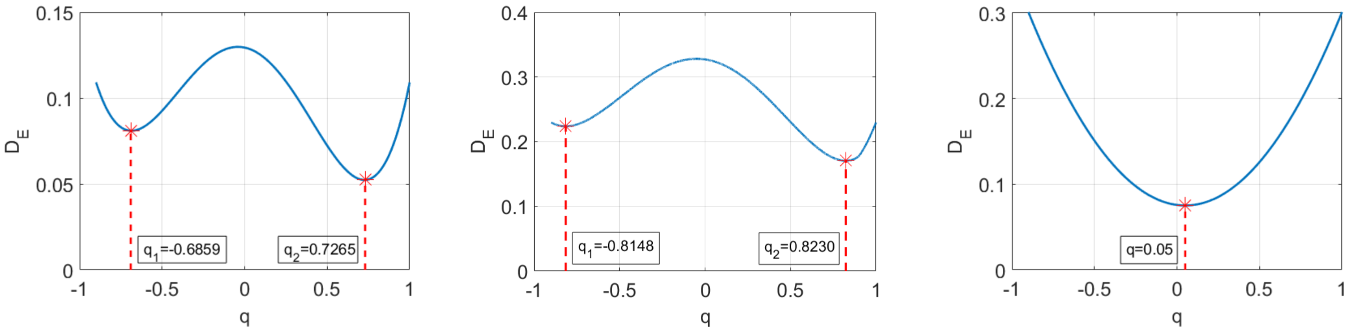
The cost function associated with two-level quantization of X and its derivative with respect to the quantization decision level q,

$$J(\mathbf{q}) = \int_a^q (x^3-y_1)^2 \mathrm{d}P_X + \int_q^b (x^3-y_2)^2 \mathrm{d}P_X,$$

$$\frac{\partial J}{\partial q} = q^3 - \frac{q}{2} + \frac{a+b}{4}(1-(b^2+a^2)).$$

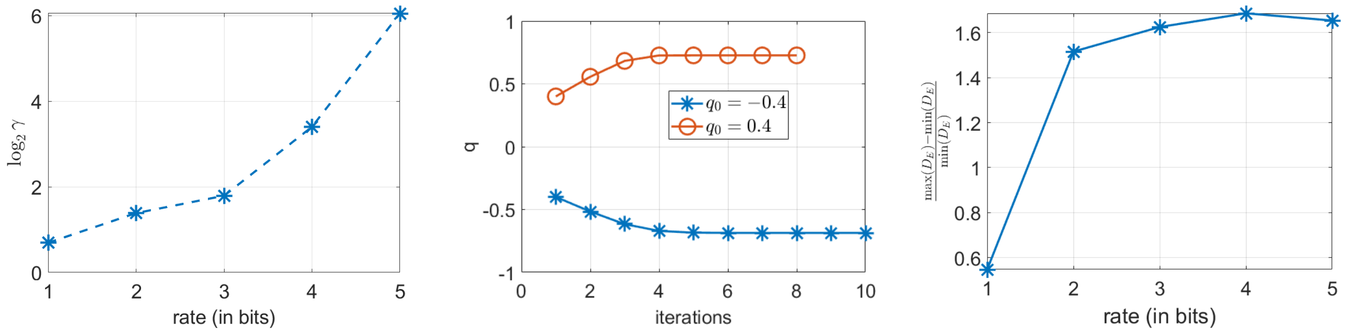
When we consider $X \sim U[0,1]$, the local minima occurs at $q = 1/\sqrt{2}$ only, while for $X \sim U[-1,1]$, there are two local minima $q = \pm 1/\sqrt{2}$, both resulting in the same distortion to the encoder due to symmetricity. We provide an alternate example here to show the existence of multiple local optima for a log-concave source as $X \sim U[-0.9,1]$ in which case, $q = -0.6859, 0.7265$ satisfy $\partial J/\partial q = 0$ and result in different distortions, as also demonstrated in 6a. Fig. 6b shows the cost curve for the same setting with a different encoder distortion measure $\eta_E(x,y) = |x^3-y|$, where we see that the local minima are different from the previous case, $q = -0.8148, 0.8230$.

In Fig. 6c, we see the corresponding cost curve for the non-strategic case which has a unique local minima as expected. Hence, the strategic problem can indeed have multiple local minima even if used in conjunction with log-concave sources.



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

Fig. 6. Cost surfaces of the encoder distortion for 2-level quantization, with decision boundary q, of $X\sim U[-0.9,1]$ with $\eta_D(x,y)=(x-y)^2$.



(a) Number of locally optimal quantizers (γ) (b) Iterations for two initializations with $\mathbf{q}=[-0.9,q,1]$ (c) Comparing D_E for locally optimal quantizers.

Fig. 7. Local optima in quantizing a source $X\sim U[-0.9,1]$ with $\eta_D(x,y)=(x-y)^2, \eta_E(x,y)=(x^3-y)^2$.

In Fig. 7a, we plot the number of locally optimal quantizers (γ) on a logarithmic scale for $\eta_E(x,y)=(x^3-y)^2$, $\eta_D(x,y)=(x-y)^2$ for a Uniform source $X\sim U[-0.9,1]$, and in Fig. 7b we see that initializing the quantizer as $\begin{bmatrix}-0.9 & -0.4 & 1\end{bmatrix}$ and $\begin{bmatrix}-0.9 & 0.4 & 1\end{bmatrix}$ results in two different quantizers, $\begin{bmatrix}-0.9 & -0.6859 & 1\end{bmatrix}$ and $\begin{bmatrix}-0.9 & 0.7265 & 1\end{bmatrix}$, as also mentioned earlier. Fig. 7 shows the fractional difference of the minimum and maximum encoder distortions amongst the locally optimal solutions. We observe that this value seems to be increasing with rate.

This poor local optima issue can be resolved by using powerful non-convex methods such as deterministic annealing [26]–[28], or by using dynamic programming for global optimality in the scalar case [29]. We implemented dynamic programming-based algorithms for quantizing a noiseless source over a noiseless and a noisy channel, respectively [18], [19].

F. Accept or reject?

As stated before, the decoder can choose to ignore the encoder's message (equivalent to non-revealing) when the resulting distortion is lower than accepting the information.

The decoder distortions for a partially revealing quantizer $\mathbf{q}=[x_0,\dots,x_M]$,

$$D_{D,pr}(\eta_D(x,y))=\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m}\eta_D(x,y_m)\mathrm{d}P_X,\quad s\in\{E,D\}.$$

In the following theorem, we consider a sufficient condition on the form of the decoder distortion measure which results in the decoder accepting the encoder's message. The proof is given in Appendix E.

Theorem 5. *If $\eta_D(x,y)=(f(x)-y)^2$, $f:\mathbb{R}\to\mathbb{R}$, $f(x)$ bounded, $\mathbb{E}\{f(X)^2\}<\infty$, then the decoder distortion $D_{D,pr}<D_{D,nr}$, irrespective of the encoder distortion $\eta_E(x,y)$ and the quantizer chosen* $\mathbf{q}$.

Remark 5. *Note that $\eta_D(x,y)=(x-y)^2$, which is considered in this paper, is a special case of the above.*

V. High-resolution analysis

In this section, we analyze the high-resolution behavior of the encoder, following a similar analysis done in classical quantization. We first summarize the high-resolution results for classical quantization below [23].

We make the following assumptions in the subsequent analysis in this section:

Assumption 6. *M is asymptotically large, $M \to \infty$.*

Let us define the step size as $\Delta_m = x_m - x_{m-1}$, and $\Delta_y = \min_{m \in [1:M]} y_m - y_{m-1}$.

Assumption 7. $\Delta_m << \Delta_y, \quad m \in [1:M]$.

Assumption 8. *Source distribution admits a density $p_X(\cdot)$ that is differentiable almost everywhere.*

Classical high-resolution analysis implicitly assumes that the encoder does not have any non-revealing regions (intervals that the encoder does not quantize). However, in strategic quantization, the encoder may choose not to quantize certain intervals. In [30], [31], authors present a bi-pooling result for the strategic communication problem which implies, in our setting that the encoder is revealing for at most two regions. We first identify these intervals, say $\mathcal{X}_{NR}$, and then perform high-resolution analysis on $\mathcal{X}_R = \mathcal{X} \backslash \mathcal{X}_{NR}$. We note that $\mathcal{X}_{NR}$ and the corresponding decoder actions for these regions can be found without performing high-resolution analysis. Let $M_{\mathcal{X}_R} = M - |\mathcal{X}_{NR}|$.

We assume the decoder distortion $\eta_D(x, y)$ is a Bregman loss function [3]. This assumption is required in converting to integrals in (25), (33), as we show below.

Remark 6. *Note that in this section, we limit the source X to be continuous, with probability density function $p_X(\cdot)$, and the decoder distortion $\eta_D(x, y)$ to be a Bregman loss function.*

The non-revealing regions for the specific cases in this paper are analyzed in Appendix I and summarized below:

$p_X(\cdot)$	$\eta_E(x,y)$	$\mathcal{X}_{NR}$
$\mathcal{N}(0,1)$	$(x-y^3)^2$	$(-\infty,-0.378] \cup (0.378,\infty)$
$U[0,1]$	$(x^3-y)^2$	$[0,1/\sqrt{3}]$
$U[-1,1]$	$(x^3-y)^2$	$[-1/\sqrt{2},1/\sqrt{2}]$
$\mathcal{N}(0,1)$	$(x^3-y)^2$	$[-1/\sqrt{2},1/\sqrt{2}]$

In the following analysis, we consider the quantization of $\mathcal{X}_R$.

At high rates, the probability density function can be approximated as uniform in each region,

$$p_X(x) \approx p_m, \quad x \in \mathcal{V}_m. \tag{7}$$

Then, the probability of the source realization being in a given region $\mathcal{V}_m$,

$$f_m = \int_{x_{m-1}}^{x_m} p_X(x) \mathrm{d}x \approx (x_m - x_{m-1}) p_m,$$

and the approximate probability density p_m can be written as,

$$p_m = \frac{f_m}{\Delta_m}. \tag{8}$$

We approximate the reconstruction levels to be the midpoint of the interval,

$$y_m = x_{m-1} + \Delta_m/2. \tag{9}$$

As we show in Appendix G, this approximation is less accurate for non-Uniform sources. For instance, a better approximation for a Gaussian source is $y_m = (1 - \Delta_m^2/12)(x_{m-1} + \Delta_m/2)$. But we use (9) for simplicity.

Let $M(x)$ be the density of the number of quantization levels, i.e., $M(x)\mathrm{d}x$ is the number of quantization levels in $[x, x+\mathrm{d}x]$ interval. The limiting point density $\lambda(x)$ is defined as

$$\lambda(x) = \lim_{M\to\infty} \frac{M(x)}{M_{\mathcal{X}_R}}.$$

Then,

$$\int_{\mathcal{X}_R} \lambda(x) \mathrm{d}x = 1. \tag{10}$$

Consider any interval $\Delta(x)$ around x. For a sufficiently large M, there will be approximately $M(x)\Delta(x) = M_{\mathcal{X}_R}\lambda(x)\Delta(x)$ quantization levels in the $\Delta(x)$ interval.

[3]Including (but not limited to, see [32]) the important case of $\eta_D(x, y) = (x - y)^2$.

We rewrite the width of cell m, Δ_m in terms of $\lambda(\cdot)$ as the ratio of the length of the interval to the number of quantization levels in the interval:

$$\triangle_m = \frac{\triangle(x)}{M_{\mathcal{X}_R}\lambda(x)\triangle(x)} \approx \frac{1}{M_{\mathcal{X}_R}\lambda(y_m)}. \tag{11}$$

Remark 7. *A simple approach to incorporate the non-revealing regions in the high-resolution analysis would be to consider $\lambda(x)$ as 0 for $x \in \mathcal{X}_{NR}$. However, this implies that the step size is not negligible for $\mathcal{V}_m \in \mathcal{X}_{NR}$, and we cannot approximate the integrals as summations in the derivations as in (25), (33).*

We now present our result on the high-resolution encoder distortion for $\eta_E(x,y) = (x-y^3)^2$ in the following Theorem and Corollary, which are proved in Appendix H.

Theorem 9. *The encoder distortion in quantizing $X \sim p_X(x)$ with encoder distortion $\eta_E(x,y) = (x-y^3)^2$ and decoder distortion any Bregman loss function at high rates is*

$$D_E = D'_M + D_M,$$

where D_M depends on M, while D'_M does not,

$$D_M = \frac{1}{12}\frac{1}{M^2}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int_{x_{m-1}}^{x_m} p_X(x)^{1/3}\Bigg)^3,$$

$$D'_M = \mathbb{E}\{X^2\} + \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int_{x_{m-1}}^{x_m}(x^6 - 2x^4)p_X(x)\mathrm{d}x + \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_{NR}}}^{M}\int_{x_{m-1}}^{x_m}(y_m^6 - 2y_m^3 x)p_X(x)\mathrm{d}x.$$

From the expression of D_E in Theorem 9, we see that the high-resolution encoder distortion converges as rate increases, which we present below as a corollary.

Corollary 9.1. *The high-resolution encoder distortion for $\eta_E(x,y) = (x-y^3)^2$ decreases with M and converges to*

$$D'_M = \mathbb{E}\{X^2\} + \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int_{x_{m-1}}^{x_m}(x^6 - 2x^4)p_X(x)\mathrm{d}x + \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_{NR}}}^{M}\int_{x_{m-1}}^{x_m}(y_m^6 - 2y_m^3 x)p_X(x)\mathrm{d}x. \tag{12}$$

Unlike classical quantization, where the encoder distortion becomes negligible at high rates, here we see that the encoder distortion may converge to some non-zero value.

We consider the problem setting of $\eta_E(x,y) = (x^3-y)^2$ now. In the previous case with $\eta_E(x,y) = (x-y^3)^2$, Corollary 9.1 shows that the encoder distortion is decreasing with M. However, as we show below, when $\eta_E(x,y) = (x^3-y)^2$, we find that the encoder distortion may be increasing, decreasing, or constant with M. That is, for some $M > M_0$, $M_0 \in \mathbb{Z}_{>0}$, the encoder may prefer to not utilize the rate fully.

We now present our results for $\eta_E(x,y) = (x^3-y)^2$ in the following Theorem and Corollary, which are proved in Appendix I.

Theorem 10. *The encoder distortion in quantizing $X \sim p_X(x)$ with encoder distortion $\eta_E(x,y) = (x^3-y)^2$, and decoder distortion any Bregman loss function at high rates is*

$$D_E = D'_M + D_M,$$

where D_M depends on M, while D'_M does not,

$$D_M = \frac{1}{12}\frac{1}{M^2}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int_{x_{m-1}}^{x_m} p_X(x)^{1/3}\Bigg)^3 - \frac{1}{2}\frac{1}{M^2}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int_{x_{m-1}}^{x_m}(x^2 p_X(x))^{1/3}\Bigg)^3,$$

$$\begin{aligned} D'_M \approx\ & \mathbb{E}\{X^6\} - \mathbb{E}\{X^2\} - 2\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int_{x_{m-1}}^{x_m}(x^4 - x^2)p_X(x)\mathrm{d}x \\ & + \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_{NR}}}^{M}\int_{x_{m-1}}^{x_m}(x^2 + y_m^2 - 2y_m x)p_X(x)\mathrm{d}x - 2\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_{NR}}}^{M}\int_{x_{m-1}}^{x_m}(x^3 - x)y_m p_X(x)\mathrm{d}x. \end{aligned}$$

Algorithm 1 Proposed strategic quantizer design

Parameters: ϵ, λ
Input: $p_X(\cdot), \mathcal{X}, M, \eta_E, \eta_D$
Output: $\mathbf{q}^*$, $\{y_m^*\}$, D_E, D_D
Initialization: assign a monotone $\mathbf{q}_0$ randomly, compute associated encoder distortion $D_E(0)$, $i=1$;
while $\Delta D > \epsilon$ *or until a set amount of iterations* **do**
  compute the gradients in (15), $\{\partial D_E/\partial x_m\}_i$
  compute the updated quantizer $\mathbf{q}_{i+1} \triangleq \mathbf{q}_i - \lambda\{\partial D_E/\partial x_m\}_i$
  compute actions $\mathbf{y}(\mathbf{q}_{i+1})$ via (16)
  compute encoder distortion $D_E(i+1)$ associated with quantizer values $\mathbf{q}_{i+1}$ and actions $\mathbf{y}(\mathbf{q}_{i+1})$ via (1)
  compute $\Delta D = D_E(i) - D_E(i+1)$
return quantizer $\mathbf{q}^* = \mathbf{q}_{i+1}$, actions $\mathbf{y}(\mathbf{q}^*)$, encoder and decoder distortions D_E and D_D computed for optimal quantizer and decoder actions $\mathbf{q}^*, \mathbf{y}(\mathbf{q}^*)$ via (1)

In the corollary below, we present the convergence results of D_E in Theorem 10.

Corollary 10.1. *The high-resolution encoder distortion for $\eta_E(x,y) = (x-y^3)^2$ converges monotonically to*

$$D'_M \approx \mathbb{E}\{X^6\} - \mathbb{E}\{X^2\} - 2\sum_{\substack{m=1\\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} (x^4 - x^2) p_X(x) \mathrm{d}x + \sum_{\substack{m=1\\ \mathcal{V}_m \in \mathcal{X}_{NR}}}^{M} \int_{x_{m-1}}^{x_m} (x^2 + y_m^2 - 2y_m x) p_X(x) \mathrm{d}x$$
$$- 2\sum_{\substack{m=1\\ \mathcal{V}_m \in \mathcal{X}_{NR}}}^{M} \int_{x_{m-1}}^{x_m} (x^3 - x) y_m p_X(x) \mathrm{d}x. \tag{13}$$

The high-resolution encoder distortion is increasing, decreasing, or constant as M increases depending on whether $D_M < 0, D_M > 0$, or $D_M = 0$, respectively,

$$D_M = \frac{1}{12}\frac{1}{M^2}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} p_X(x)^{1/3}\Bigg)^3 - \frac{1}{2}\frac{1}{M^2}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} (x^2 p_X(x))^{1/3}\Bigg)^3. \tag{14}$$

We observe in the numerical results in Figures 11b, 14a,b that the encoder does not quantize $\mathcal{X}_{NR}$ for $M = [2,4,8,16,32]$, which leads us to the following conjecture. Let $\mathcal{V}_{NR,t} \in \mathcal{X}_{NR}$, $t \in \mathcal{Z}_{>0}$ be the intervals that the encoder does not quantize at high resolution.

Conjecture 10.1. *For any $M-$level quantization, $Q(x) = k_t$ for $x \in \mathcal{V}_{NR,t}$, where $t \in \mathcal{Z}_{>0}, \mathcal{V}_{NR,t} \in \mathcal{X}_{NR}$, $k_t \in \mathbb{R}$, i.e., the encoder does not quantize regions in $\mathcal{X}_{NR}$ regardless of the rate.*

VI. Proposed Design Method

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, which is identical for the decoder and the encoder (team problem), is non-increasing with each iteration. However, here we consider a game problem (as opposed to a team problem) where the objectives are different, a strategic variation of these algorithms would enforce optimality with respect to a different distortion measure at each iteration and hence do not converge as illustrated in detail in [11].

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 has 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/GDnoiseless.

Since we use the gradients $\partial D_E/\partial x_m$ in the algorithm, we need to compute the gradients which we provide in Appendix C for a minimum mean squared error (MMSE) decoder $\eta_D(x,y) = (x-y)^2$ with encoder distortion measure $\eta_E(x,y)$.

Remark 8. *The proposed algorithm is for a continuous source X.*

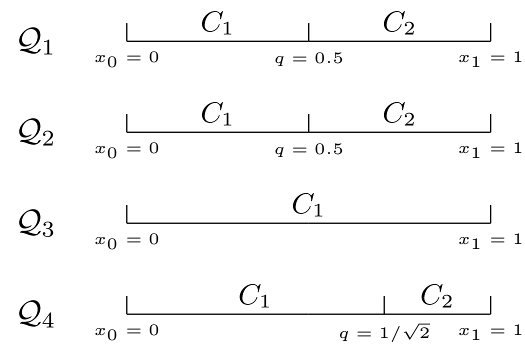


Fig. 8. Quantizers for 2-level quantization of $X \sim \mathcal{N}(0, 1)$ with $\eta_D(x, y) = (x - y)^2$ and $\eta_E(x, y) = (x - y)^2, (1.5x - y)^2, (2.5x - y)^2, (x^3 - y)^2$ as $\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3, \mathcal{Q}_4$, respectively.

VII. Numerical Results

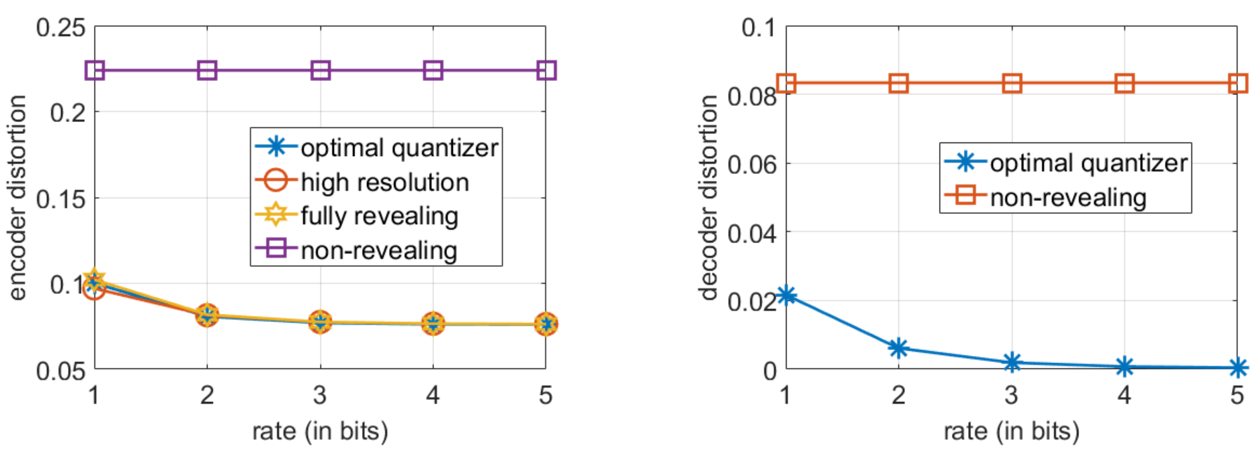
We plot the quantizers for $M = 2$ level quantization of the source $X \sim U[0, 1]$ in Fig. 8 for $\eta_D(x, y) = (x - y)^2$ with $\eta_E(x, y) = (x - y)^2, (x - 1.5y)^2, (x - 2.5y)^2, (x^3 - y)^2$ as $\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3, \mathcal{Q}_4$, respectively. As expected from Theorem 3, the encoder is fully revealing for $\eta_E(x, y) = (x - 1.5y)^2$ ($\mathcal{Q}_2$) and implements the same quantizer as in the non-strategic case with decoder distortion ($\mathcal{Q}_1$), and is non-revealing for $\eta_E(x, y) = (x-2.5y)^2$ as shown in $\mathcal{Q}_3$. The quantizer for $\eta_E(x, y) = (x^3-y)^2$ ($\mathcal{Q}_4$) is partially revealing and different from the classical quantizer.

We consider the following two encoder and decoder distortion measures

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

each with two source distributions: $(i)\, U[0, 1]$, $(ii)\, \mathcal{N}(0, 1)$. We limit the range of the Gaussian distribution to $[-5, 5]$ for tractability of the algorithm.

We plot the encoder distortion for the settings described above of the optimal strategic quantizer, the high-resolution distortion, fully-revealing (implementing a non-strategic quantizer with decoder distortion metric), and non-revealing ($M = 1$) cases in Figures 9a, 10a, 12a, 13a, and observe that the distortion in using the optimal quantizer is the minimum [4]. We also note that the high-resolution distortion approximates the encoder distortion better as the rate increases, as expected. We plot the decoder distortions for the optimal strategic quantizer and the non-revealing (decoder does not accept encoder's message) cases in Figures 9b, 10b, 12b, 13b, and we observe that the decoder gains in accepting the encoder's information (the optimal quantizer distortion is lower than that for non-revealing) as stated in Theorem 5.



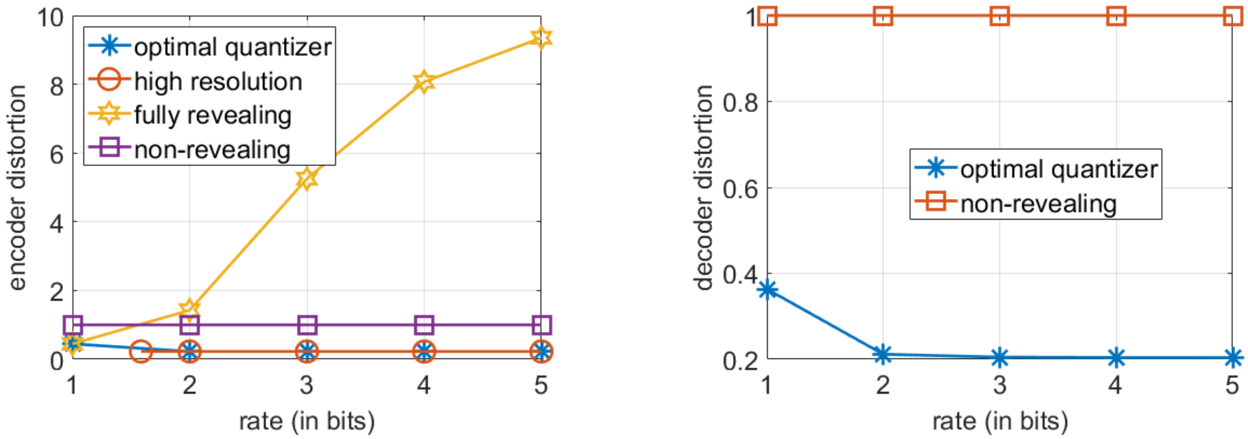
(a) Encoder distortion

(b) Decoder distortion

Fig. 9. Encoder and decoder distortions in quantizing $X \sim U[0, 1], \eta_D(x, y) = (x - y)^2, \eta_E(x, y) = (x - y^3)^2$.

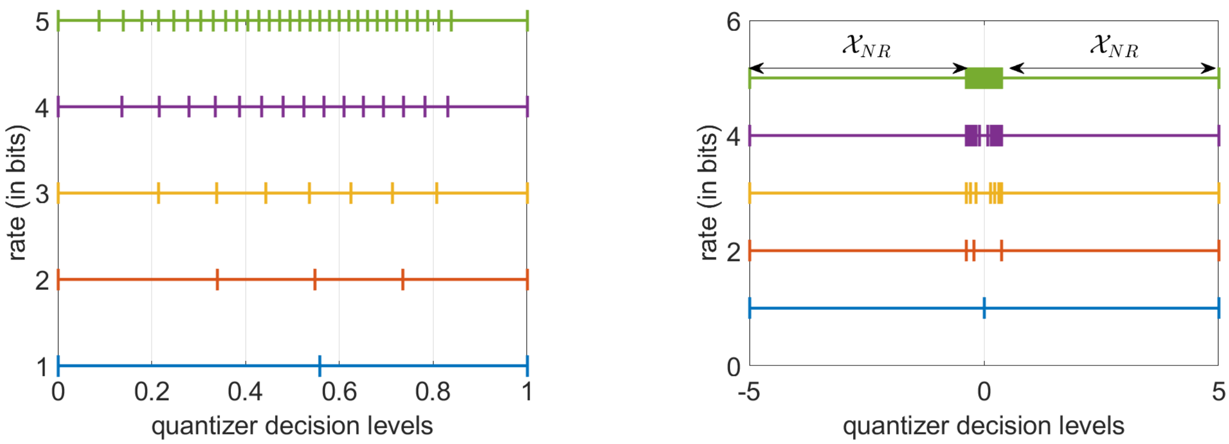
The encoder distortions may not become negligibly small at high rates as in classical quantization. This can be attributed to the disparity between the objectives of the encoder and the decoder, even if quantization was not done, the distortions would not completely vanish, see e.g., [9] for more details. This is also computed from (12) which evaluates to $0.0762, 0.2248$ for

[4]We plot the high-resolution distortion from $M = 3$ in Fig. 10a since there are at least three regions with two that are non-revealing.



(a) Encoder distortion (b) Decoder distortion

Fig. 10. Encoder and decoder distortions in quantizing $X \sim \mathcal{N}(0,1), \eta_D(x,y) = (x-y)^2, \eta_E(x,y) = (x-y^3)^2$.



(a) $X \sim U[0,1]$ (b) $X \sim \mathcal{N}(0,1)$

Fig. 11. Quantizer structures in quantizing X with $\eta_D(x,y) = (x-y)^2, \eta_E(x,y) = (x-y^3)^2$.

$X \sim U[0,1], \mathcal{N}(0,1)$, and from (13) which evaluates to $0.0698, 9.9556$, for $X \sim U[0,1], \mathcal{N}(0,1)$, respectively, as seen in Figures 9a, 10a, 12a, 13a,.

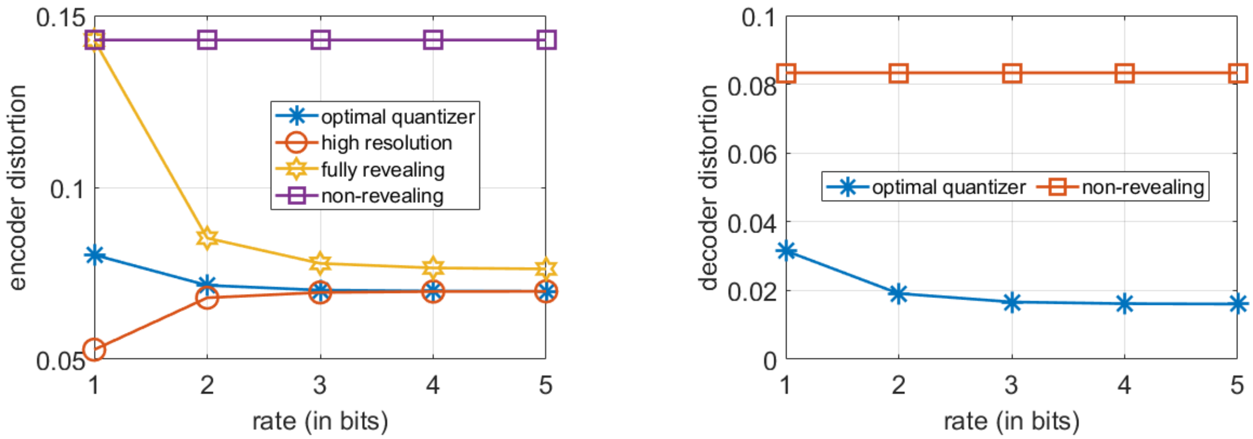
We observe that the high-resolution encoder distortion increases with rate for $X \sim U[0,1]$, i.e., for some $M_0 > M$, the encoder prefers to be less revealing. We see this also from evaluating D_M in (14) which is $-0.0019/M^2$, i.e., $D_M < 0$ for $X \sim U[0,1]$ with $\mathcal{X}_R = [1/\sqrt{3}, 1]$.

When $\eta_E(x,y) = (x-y^3)^2, X \sim U[0,1]$, there is no non-revealing region, as seen in Fig. 11a. When $\eta_E(x,y) = (x-y^3)^2, X \sim \mathcal{N}(0,1)$, the encoder does not quantize $[-5,-0.378] \cup [0.378, 5]$, as we see in Fig. 11b. We note that there can be at most two non-revealing regions. When $\eta_E(x,y) = (x^3-y)^2$, the encoder does not quantize $[0, 1/\sqrt{3}]$ for $X \sim U[0,1]$ and $[-1/\sqrt{2}, 1/\sqrt{2}]$ for $X \sim \mathcal{N}(0,1)$, as we see in Figures 14a,b. These results validate our analysis in Appendices H, I.

Remark 9. *Note that our findings apply to any general distortion measure. We take the decoder's measure as MSE, the most commonly used distortion metric. The choice of the encoder's distortion measure ($\eta_E \neq \eta_D$ due to problem formulation) is arbitrary; however, some measures yield non-interesting solutions such as nonrevealing (encoder does not send any information) or fully-revealing (the problem simplifies to non-strategic quantization with decoder distortion metric). Hence we chose measures that yield results that demonstrate the interesting aspects of strategic quantization.*

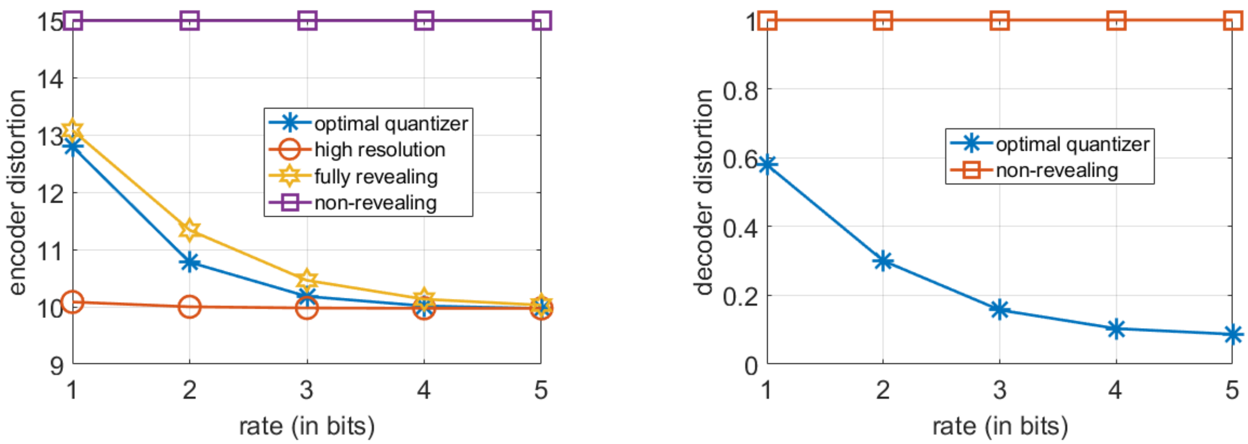
VIII. Conclusion

In this paper, we have introduced the strategic quantization problem and highlighted its distinguishing features in comparison to classical quantization. Through numerical experimentation, we have established that the strategic variation of Lloyd-Max may not converge to a local optimum. Building upon this insight, we have devised gradient descent-based solutions for addressing the strategic quantization problem. We analyzed the encoder distortion at high rates [33]. The numerical outcomes obtained



(a) Encoder distortion (b) Decoder distortion

Fig. 12. Encoder and decoder distortions in quantizing $X \sim U[0,1], \eta_D(x,y) = (x-y)^2, \eta_E(x,y) = (x^3-y)^2$.



(a) Encoder distortion (b) Decoder distortion

Fig. 13. Encoder and decoder distortions in quantizing $X \sim \mathcal{N}(0,1), \eta_D(x,y) = (x-y)^2, \eta_E(x,y) = (x^3-y)^2$.

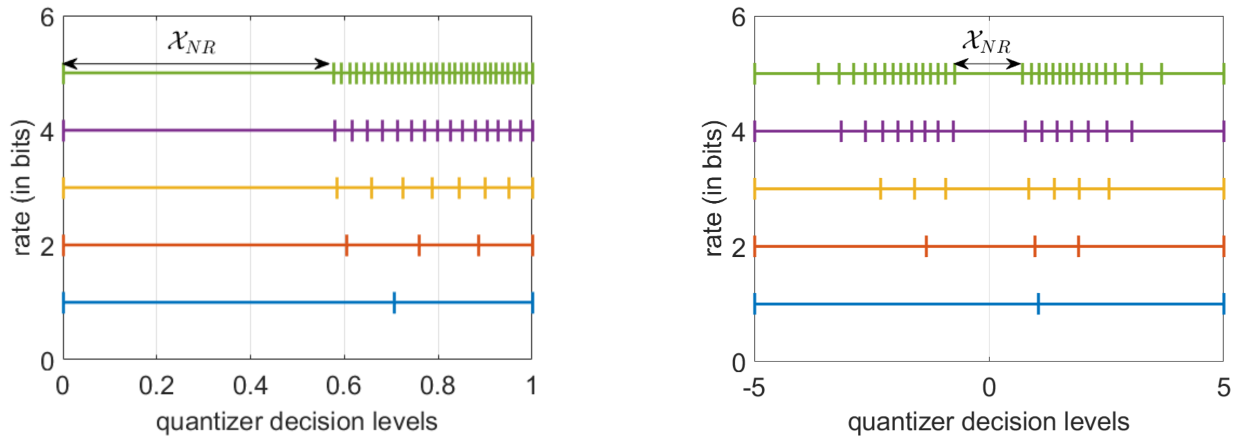
using our proposed algorithm raise several open theoretical inquiries regarding the operational distortion-rate curve of optimal strategic quantizers. We implemented a globally optimal dynamic programming-based algorithm in [18]. We considered a noisy channel as a continuation of this work, where the solution method involves random index assignment with gradient descent [20], and with dynamic programming [19]. We extended our analysis to a 2-dimensional source with quadratic costs for the encoder and decoder in [22]. We also analyzed the strategic quantization of a noisy source in [21].

Appendix A
Quantizer behavior for a uniform source X with $\eta_E(x,y) = |x-y|, \eta_D(x,y) = (x-y)^2$

We first compute the decoder reconstructions and derivatives used in subsequent equations below:

$$y_m = \frac{\int\limits_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x}{\int\limits_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x} = \frac{x_{m-1}+x_m}{2}, \quad \frac{\partial y_m}{\partial x_m} = \frac{1}{2},$$

$$y_{m+1} = \frac{\int\limits_{x_m}^{x_{m+1}} x p_X(x)\mathrm{d}x}{\int\limits_{x_m}^{x_{m+1}} p_X(x)\mathrm{d}x} = \frac{x_m+x_{m+1}}{2}, \quad \frac{\partial y_{m+1}}{\partial x_m} = \frac{1}{2}.$$



(a) $X \sim U[0,1]$ (b) $X \sim \mathcal{N}(0,1)$

Fig. 14. Quantizer structures in quantizing X with $\eta_D(x,y)=(x-y)^2, \eta_E(x,y)=(x^3-y)^2$.

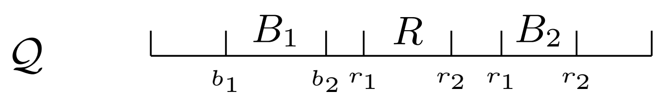


Fig. 15. Non-monotonic quantizer

The encoder distortion,

$$D_E=\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m}|x-y_m|p_X(x)\mathrm{d}x=\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m}\left|x-\frac{x_{m-1}+x_m}{2}\right|p_X(x)\mathrm{d}x$$

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

Enforcing KKT conditions of optimality, $\frac{\partial D_E}{\partial x_m}=0$, we have

$$\frac{\partial D_E}{\partial x_m}=\frac{1}{b-a}\left(\left(x_m-\frac{x_{m-1}+x_m}{2}\right)-\left(\frac{x_{m+1}+x_m}{2}-x_m\right)+\frac{1}{2}\int_{x_{m-1}}^{\frac{x_{m-1}+x_m}{2}}p_X(x)\mathrm{d}x-\frac{1}{2}\int_{\frac{x_{m-1}+x_m}{2}}^{x_m}p_X(x)\mathrm{d}x\right.$$

$$\left.+\frac{1}{2}\int_{x_m}^{\frac{x_m+x_{m+1}}{2}}p_X(x)\mathrm{d}x-\frac{1}{2}\int_{\frac{x_m+x_{m+1}}{2}}^{x_{m+1}}p_X(x)\mathrm{d}x\right)$$

$$=\frac{1}{b-a}\left(x_m-\frac{x_{m-1}+x_{m+1}}{2}+\frac{1}{2}\frac{x_m-x_{m-1}}{2}-\frac{1}{2}\frac{x_m-x_{m-1}}{2}+\frac{1}{2}\frac{x_{m+1}-x_m}{2}-\frac{1}{2}\frac{x_{m+1}-x_m}{2}\right)=0,$$

which yields $x_m=\frac{x_{m-1}+x_{m+1}}{2}$ or equivalently, $x_m-x_{m-1}=x_{m+1}-x_m$, i.e., a uniform quantizer, which implies that the encoder is fully-revealing. Let $x_{m-1}=x_m-\triangle_{m-1}, x_{m+1}=x_m+\triangle_{m+1}$. Then, the above equation becomes $\triangle_{m-1}=\triangle_{m+1}$, i.e., the source is uniformly divided (the encoder is fully revealing).

Appendix B
Monotonicity of the Quantizer

Consider encoder distortion $\eta_E(x,y)=l(x)f(|g(x)-h(y)|)$ where $l(\cdot)>0, f(\cdot)$ increasing, $g(\cdot)$ monotonic, $h(\cdot)$ bijective. Assume the quantizer structure as given in Fig. 15 with B_1 and B_2 be mapped to y_1, and R mapped to y_2. The function $g(\cdot)$ being monotonic results in order preserving or order reversing of the quantizer structure of x when mapped to $g(x)$ (the order B_1, R, B_2 may remain as it is or reverse to B_2, R, B_1). Since B_1 and B_2 are mapped to y_1, $\eta_E(x,y_1)<\eta_E(x,t)$ or $|g(x)-h(y_1)|<|g(x)-h(t)|$, $t\in\mathcal{Y}, \forall x\in[b_1,b_2]\cup[b_3,b_4]$.

Let $g_1 = \min(g(b_2), g(b_3)), g_2 = \max(g(b_2), g(b_3))$. For any $x \in [b_2, b_3]$, i.e., $g(x) \in [g_1, g_2]$, let $Q(x) = t$. Then, there are three cases:

1) $h(t) < g_1$: $|g(x) - h(t)| = |g(x) - g_1| + |g_1 - h(t)|$. The first term is fixed, the second term is minimized when $h(t) = h(y_1)$ i.e., $t = y_1$.
2) $h(t) > g_2$: $|g(x) - h(t)| = |g(x) - g_2| + |g_2 - h(t)|$. Like the previous case, the first term is fixed, the second term is minimized when $h(t) = h(y_1)$ i.e., $t = y_1$.
3) $h(t) \in [g_1, g_2]$
   a) $h(t) < h(y_1)$: B_1 should be mapped to t, contradicts the given quantizer structure.
   b) $h(t) > h(y_1)$: B_2 should be mapped to t, contradicts the given quantizer structure.

The given quantizer structure cannot happen if R is mapped to an action $y_2 \neq y_1$. If B_1 and B_2 are mapped to y_1, any $x \in [b_2, b_3]$ is also mapped to y_1, i.e., the optimal quantizer for the given encoder distortion has to be monotonic.

APPENDIX C
GRADIENTS

The gradients of the encoder's distortion,

$$D_E = \sum_{m=1}^{M} \int_{x_{m-1}}^{x_m} \eta_E(x, y_m) p_X(x) \mathrm{d}x,$$

with respect to the quantizer decision levels for decoder distortion measure $\eta_D(x, y) = (x - y)^2$ are given as:

$$\frac{\partial D_E}{\partial x_m} = p_X(x_m)(\eta_E(x_m, y_m) - \eta_E(x_m, y_{m+1})) + \frac{\partial y_m}{\partial x_m} \int_{x_{m-1}}^{x_m} \frac{\partial \eta_E(x, y_m)}{\partial y_m} p_X(x) \mathrm{d}x + \frac{\partial y_{m+1}}{\partial x_m} \int_{x_m}^{x_{m+1}} \frac{\partial \eta_E(x, y_{m+1})}{\partial y_{m+1}} p_X(x) \mathrm{d}x, \quad (15)$$

since the reconstruction levels y_m, y_{m+1} are the only ones which depend on x_m,

$$y_m = \frac{\int_{x_{m-1}}^{x_m} x p_X(x) \mathrm{d}x}{\int_{x_{m-1}}^{x_m} p_X(x) \mathrm{d}x}, \quad y_{m+1} = \frac{\int_{x_m}^{x_{m+1}} x p_X(x) \mathrm{d}x}{\int_{x_m}^{x_{m+1}} p_X(x) \mathrm{d}x}. \quad (16)$$

The gradients $\frac{\partial y_m}{\partial x_m}, \frac{\partial y_{m+1}}{\partial x_m}$ are

$$\frac{\partial y_m}{\partial x_m} = \frac{x_m p_X(x_m) \int_{x_{m-1}}^{x_m} p_X(x) \mathrm{d}x - p_X(x_m) \int_{x_{m-1}}^{x_m} x p_X(x) \mathrm{d}x}{(\int_{x_{m-1}}^{x_m} p_X(x) \mathrm{d}x)^2} = p_X(x_m) \frac{(x_m - y_m)}{\int_{x_{m-1}}^{x_m} p_X(x) \mathrm{d}x}, \quad (17)$$

$$\frac{\partial y_{m+1}}{\partial x_m} = -p_X(x_m) \frac{x_m \int_{x_m}^{x_{m+1}} p_X(x) \mathrm{d}x - \int_{x_m}^{x_{m+1}} x p_X(x) \mathrm{d}x}{(\int_{x_m}^{x_{m+1}} p_X(x) \mathrm{d}x)^2} = -p_X(x_m) \frac{(x_m - y_{m+1})}{\int_{x_m}^{x_{m+1}} p_X(x) \mathrm{d}x}. \quad (18)$$

APPENDIX D
QUANTIZER BEHAVIOUR FOR $\eta_E(x, y) = (x + \alpha - \beta y)^2, \eta_D(x, y) = (x - y)^2$

Let the probability density function of the continuous source X be $p_X(\cdot)$. Let us parameterize $\mathcal{V}_m^*$ as $[x_{m-1}, x_m)$, where $x_m \in \mathcal{X}$, $x_{m-1} < x_m$, i.e., $\mathbf{q} = [x_0, \ldots, x_M]$. The decoder's actions $\mathbf{y}$ is a function of $\mathbf{q}$. For a given $\mathbf{q} = [x_0, \ldots, x_m]$, the decoder with $\eta_D(x, y) = (x - y)^2$ determines $\mathbf{y} = [y_1, \ldots, y_m]$ as follows:

$$y_m = \frac{\int_{x_{m-1}}^{x_m} x p_X(x) \mathrm{d}x}{\int_{x_{m-1}}^{x_m} p_X(x) \mathrm{d}x}.$$

The encoder's distortion and its derivative with respect to x_m

$$D_E = \sum_{m=1}^{M} \int_{x_{m-1}}^{x_m} (x + \alpha - \beta y_m)^2 p_X(x) \mathrm{d}x,$$

$$\frac{\partial D_E}{\partial x_m} = \big((x_m+\alpha-\beta y_m)^2 - (x_m+\alpha-\beta y_{m+1})^2\big)p_X(x_m) - 2\beta\frac{\mathrm{d}y_m}{\mathrm{d}x_m}\int\limits_{x_{m-1}}^{x_m}(x+\alpha-\beta y_m)p_X(x)\mathrm{d}x$$
$$-2\beta\frac{\mathrm{d}y_{m+1}}{\mathrm{d}x_m}\int\limits_{x_m}^{x_{m+1}}(x+\alpha-\beta y_{m+1})p_X(x)\mathrm{d}x.$$

Enforcing the KKT conditions for optimality

$$\frac{\partial D_E}{\partial x_m} = 0, \quad m\in[1:M].$$

We re-write $\frac{\partial D_E}{\partial x_m}$ as follows for ease of computation,

$$\frac{\partial D_E}{\partial x_m} = I_1+I_2+I_3+I_4, \tag{19}$$

where

$$I_1 = (x_m+\alpha-\beta y_m)^2 p_X(x_m), \quad I_2 = -(x_m+\alpha-\beta y_{m+1})^2 p_X(x_m),$$
$$I_3 = -2\beta\frac{\mathrm{d}y_m}{\mathrm{d}x_m}\int\limits_{x_{m-1}}^{x_m}(x+\alpha-\beta y_m)p_X(x)\mathrm{d}x, \quad I_4 = -2\beta\frac{\mathrm{d}y_{m+1}}{\mathrm{d}x_m}\int\limits_{x_m}^{x_{m+1}}(x+\alpha-\beta y_{m+1})p_X(x)\mathrm{d}x.$$

Simplifying I_1, I_2, I_3, I_4, with derivatives $\frac{\mathrm{d}y_m}{\mathrm{d}x_m}, \frac{\mathrm{d}y_{m+1}}{\mathrm{d}x_m}$ from 17, 18, respectively,

$$I_1+I_2 = \beta p_X(x_m)(y_{m+1}-y_m)(2(x_m+\alpha)-\beta(y_m+y_{m+1})),$$

$$I_3 = -2\beta p_X(x_m)(x_m-y_m)(y_m(1-\beta)+\alpha), \quad I_4 = 2\beta p_X(x_m)(x_m-y_{m+1})(y_{m+1}(1-\beta)+\alpha),$$

$$\begin{aligned} I_3+I_4 &= 2\beta p_X(x_m)\big(-y_m((1-\beta)x_m-\alpha)+(1-\beta)y_m^2+y_{m+1}((1-\beta)x_m-\alpha)-(1-\beta)y_{m+1}^2\big)\\ &= 2\beta p_X(x_m)(y_{m+1}-y_m)\big(((1-\beta)x_m-\alpha)-(1-\beta)(y_m+y_{m+1})\big).\end{aligned}$$

Substituting these terms in 19,

$$\begin{aligned}\frac{\partial D_E}{\partial x_m} &= \beta p_X(x_m)(y_{m+1}-y_m)\big(2(x_m+\alpha)-\beta(y_m+y_{m+1})+2(1-\beta)((x_m-(y_m+y_{m+1})-2\alpha)\big)\\ &= \beta(2-\beta)p_X(x_m)(y_{m+1}-y_m)(2x_m-(y_m+y_{m+1})).\end{aligned}$$

The solutions that satisfy (19) are $\beta=0,2$, or $x_m=\frac{y_m+y_{m+1}}{2}$ (the other condition, $y_{m+1}=y_m$, 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 will send a message if its distortion decreases in doing so, $D_{E,nr}>D_E$. We now analyze the encoder distortion for the same.

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

$$D_E = \int_a^b x^2 p_X(x)\mathrm{d}x + \alpha^2 + 2\alpha(1-\beta)\int_a^b x p_X(x)\mathrm{d}x + \beta(\beta-2)\sum_{m=1}^M y_m\int\limits_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x.$$

The distortion for a non-informative quantizer:

$$D_{E,nr} = \int_a^b (x+\alpha-\beta y)^2 p_X(x)\mathrm{d}x = \int_a^b x^2 p_X(x)\mathrm{d}x + \alpha^2 + 2\alpha(1-\beta)\int_a^b x p_X(x)\mathrm{d}x + \beta(\beta-2)y\int_a^b x p_X(x)\mathrm{d}x. \tag{20}$$

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

$$D_E = \int_a^b x^2 p_X(x)\mathrm{d}x + \alpha^2 + 2\alpha(1-\beta)\int_a^b x p_X(x)\mathrm{d}x + \beta(\beta-2)\sum_{m=1}^{M} y_m \int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x \tag{21}$$

$$= D_{E,nr} + \beta(\beta-2)\left(\sum_{m=1}^{M} y_m \int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x - y\int_a^b x p_X(x)\mathrm{d}x\right) \tag{22}$$

$$= D_{E,nr} + \beta(\beta-2)\left(\sum_{m=1}^{M} \frac{\int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x}{\int_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x}\int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x - \frac{\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x}{\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x}\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x\right)$$

$$= D_{E,nr} + \beta(\beta-2)T,$$

where $T = \sum_{m=1}^{M} \frac{\left(\int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x\right)^2}{\int_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x} - \frac{\left(\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x\right)^2}{\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x}$. In order for the quantizer to be informative ($M > 1$), the term $\beta(\beta-2)T$ has to be negative. 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 Cauchy-Schwarz inequality, we have

$$\left(\sum_{m=1}^{M} u'_m v'_m\right)^2 \le \sum_{i=1}^{M}(u'_m)^2 \sum_{i=1}^{M}(v'_m)^2.$$

With a substitution, $u'_m = u_m/\sqrt{v_m}, v'_m = \sqrt{v_m}$, we get $\left(\sum_{m=1}^{M} u_m\right)^2 \le \sum_{m=1}^{M}\frac{u_m^2}{v_m}\sum_{m=1}^{M} v_m$

$$\frac{\left(\sum_{m=1}^{M} u_m\right)^2}{\sum_{m=1}^{M} v_m} \le \sum_{m=1}^{M}\frac{u_m^2}{v_m}.[5] \tag{23}$$

Let us define $u_m = \int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x$, $v_m = \int_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x$. Then, we get

$$\sum_{m=1}^{M} \frac{\left(\int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x\right)^2}{\int_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x} - \frac{\left(\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x\right)^2}{\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x} \ge 0,$$

with equality if $u_m = k v_m$, where k is a constant, which implies $y_m = u_m/v_m = k$, a constant, or that the actions are not unique, which is not possible. So, $T > 0$.

Therefore, the only possible case is case 2 with $0 < \beta < 2$, and the encoder chooses a non-strategic quantizer (as we show earlier in (19) that the only solution when $\beta \neq 0, 2$ is a non-strategic encoder if the encoder sends some message). From (22), 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.

Appendix E
Decoder: accept or reject?

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

$$D_D = \sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} (f(x) - y_m)^2 \mathrm{d}P_X.$$

[5]This is also called Sedrakyan's lemma.

We obtain the decoder actions y_m by enforcing KKT conditions,

$$\frac{\partial D_D}{\partial y_m} = -2 \int_{x_{m-1}}^{x_m} (f(x) - y_m)\mathrm{d}P_X = 0,$$

that is,

$$y_m = \frac{\int_{x_{m-1}}^{x_m} f(x)\mathrm{d}P_X}{\int_{x_{m-1}}^{x_m} \mathrm{d}P_X}.$$

The non-revealing distortion to the decoder,

$$D_{D,nr} = \int_a^b (f(x) - y)^2 \mathrm{d}P_X = \int_a^b f(x)^2 \mathrm{d}P_X - \frac{\Big(\sum_{m=1}^{M} \int_{x_{m-1}}^{x_m} f(x)\mathrm{d}P_X\Big)^2}{\sum_{m=1}^{M} \int_{x_{m-1}}^{x_m} \mathrm{d}P_X},$$

where we use $y = \int_a^b f(x)\mathrm{d}P_X / \int_a^b \mathrm{d}P_X$. The decoder's distortion for $\log M$ rate communication,

$$\begin{aligned} D_D &= \sum_{m=1}^{M} \int_{x_{m-1}}^{x_m} (f(x) - y_m)^2 \mathrm{d}P_X = \int_a^b f(x)^2 \mathrm{d}P_X + \sum_{m=1}^{M} y_m^2 \int_{x_{m-1}}^{x_m} \mathrm{d}P_X - 2\sum_{m=1}^{M} y_m \int_{x_{m-1}}^{x_m} f(x)\mathrm{d}P_X \\ &= \int_a^b f(x)^2 \mathrm{d}P_X - \sum_{m=1}^{M} \frac{\Big(\int_{x_{m-1}}^{x_m} f(x)\mathrm{d}P_X\Big)^2}{\int_{x_{m-1}}^{x_m} \mathrm{d}P_X} = D_{D,nr} - T \end{aligned}$$

where we define $T = \sum_{m=1}^{M} \frac{\Big(\int_{x_{m-1}}^{x_m} f(x)\mathrm{d}P_X\Big)^2}{\int_{x_{m-1}}^{x_m} \mathrm{d}P_X} - \frac{\Big(\sum_{m=1}^{M} \int_{x_{m-1}}^{x_m} f(x)\mathrm{d}P_X\Big)^2}{\sum_{m=1}^{M} \int_{x_{m-1}}^{x_m} \mathrm{d}P_X}$ similar to Appendix D, with $u_m = \int_{x_{m-1}}^{x_m} f(x)\mathrm{d}P_X$, $v_m = \int_{x_{m-1}}^{x_m} \mathrm{d}P_X$, we get $T \geq 0$, with equality if $u_m = kv_m$, i.e., $k = \int_{x_{m-1}}^{x_m} f(x)\mathrm{d}P_X / \int_{x_{m-1}}^{x_m} \mathrm{d}P_X = \mathbb{E}\{f(x)|x \in \mathcal{V}_m\}$ is a constant for $m \in [1:M]$ (not feasible since the actions y_m, $m \in [1:M]$ are unique). Hence we have,

$$D_D < D_{D,nr},$$

i.e., the decoder benefits from using the information provided by the encoder.

Appendix F
Fully revealing encoder

We rewrite the encoder's distortion as follows:

$$\begin{aligned} D_E &= \mathbb{E}\{(f(X) - Y)^2\} = \mathbb{E}\{(X - Y + f(X) - X)^2\} = \mathbb{E}\{(X - Y)^2\} + \mathbb{E}\{(f(X) - X)^2\} + 2\mathbb{E}\{(X - Y)(f(X) - X)\} \\ &= \mathbb{E}\{(X - Y)^2\} + \mathbb{E}\{(f(X) - X)^2\} + 2\sum_{m=1}^{M} \int_{x_{m-1}}^{x_m} (x - y_m)(f(x) - x)p_X(x)\mathrm{d}x. \end{aligned}$$

Enforcing KKT conditions of optimality, $\frac{\partial D_E}{\partial x_m} = 0$ for all $m \in [1, M]$:

$$\begin{aligned} \frac{\partial D_E}{\partial x_m} &= \frac{\partial \mathbb{E}\{(X - Y)^2\}}{\partial x_m} + 2(x_m - y_m)(f(x_m) - x_m)p_X(x_m) - 2(x_m - y_{m+1})(f(x_m) - x_m)p_X(x_m) \\ &\quad - 2\frac{\partial y_m}{\partial x_m} \int_{x_{m-1}}^{x_m} (f(x) - x)p_X(x)\mathrm{d}x - 2\frac{\partial y_{m+1}}{\partial x_m} \int_{x_m}^{x_{m+1}} (f(x) - x)p_X(x)\mathrm{d}x. \end{aligned}$$

Substituting $\partial y_m/\partial x_m, \partial y_{m+1}/\partial x_m$ from (17), (18),

$$\begin{aligned}
\frac{\partial D_E}{\partial x_m} &= \frac{\partial \mathbb{E}\{(X-Y)^2\}}{\partial x_m} + 2(x_m - y_m)(f(x_m) - x_m)p_X(x_m) - 2(x_m - y_{m+1})(f(x_m) - x_m)p_X(x_m) \\
&\quad - 2\frac{x_m - y_m}{\int\limits_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x}\int\limits_{x_{m-1}}^{x_m}(f(x)-x)p_X(x)\mathrm{d}x + 2\frac{x_m - y_{m+1}}{\int\limits_{x_m}^{x_{m+1}} p_X(x)\mathrm{d}x}\int\limits_{x_m}^{x_{m+1}}(f(x)-x)p_X(x)\mathrm{d}x \\
&= \frac{\partial \mathbb{E}\{(X-Y)^2\}}{\partial x_m} + p_X(x_m)\big(2(x_m - y_m)(f(x_m) - x_m) - 2(x_m - y_{m+1})(f(x_m) - x_m) \\
&\quad - 2(x_m - y_m)(\mathbb{E}\{f(X)|x \in \mathcal{V}_m\} - y_m) + 2(x_m - y_{m+1})(\mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\} - y_{m+1})\big) \\
&= \frac{\partial \mathbb{E}\{(X-Y)^2\}}{\partial x_m} + p_X(x_m)\big(2(f(x_m) - x_m)(y_{m+1} - y_m) + 2x_m(\mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\} - \mathbb{E}\{f(X)|x \in \mathcal{V}_m\}) \\
&\quad - 2x_m(y_{m+1} - y_m) + 2(y_{m+1}^2 - y_m^2) - 2(y_{m+1}\mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\} - y_m\mathbb{E}\{f(X)|x \in \mathcal{V}_m\})\big).
\end{aligned}$$

If the encoder is fully revealing, $x_m = \frac{y_m + y_{m+1}}{2}$, $\frac{\partial \mathbb{E}\{(X-Y)^2\}}{\partial x_m} = 0$,

$$\begin{aligned}
\frac{\partial D_E}{\partial x_m} &= 2p_X(x_m)\Bigg(\bigg(f(x_m) - \frac{y_m + y_{m+1}}{2}\bigg)(y_{m+1} - y_m) + 2\frac{y_m + y_{m+1}}{2}\bigg(\mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\} - \mathbb{E}\{f(X)|x \in \mathcal{V}_m\}\bigg) \\
&\quad - 2\frac{y_m + y_{m+1}}{2}(y_{m+1} - y_m) + 2(y_{m+1}^2 - y_m^2) - 2(y_{m+1}\mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\} - y_m\mathbb{E}\{f(X)|x \in \mathcal{V}_m\})\Bigg) \\
&= p_X(x_m)(y_{m+1} - y_m)(2f(x_m) - \mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\} - \mathbb{E}\{f(X)|x \in \mathcal{V}_m\}).
\end{aligned}$$

The encoder is fully revealing if (necessary condition),

$$f\bigg(\frac{y_m + y_{m+1}}{2}\bigg) = \frac{\mathbb{E}\{f(X)|x \in \mathcal{V}_m\} + \mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\}}{2}.$$

The term $f\big(\frac{y_m + y_{m+1}}{2}\big)$ is a function of y_m and y_{m+1}, but the term $(\mathbb{E}\{f(X)|x \in \mathcal{V}_m\} + \mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\})/2$ will simplify to being a function of only $\{y_m\}$ only if $\mathbb{E}\{\cdot\}$ and $f(\cdot)$ can be interchanged, which happens only when $f(\cdot)$ is an affine function, i.e., if the encoder is fully revealing, $f(x)$ has to be affine.

Considering a fully revealing encoder with $f(x)$ affine, we now do the second derivative test using the following. If $f(x) = \alpha + \beta x$, $2f(x_m) = \mathbb{E}\{f(X)|x \in \mathcal{V}_m\} + \mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\}$, $f'(x_m) = \beta$,

$$\begin{aligned}
\frac{\partial^2 D_E}{\partial x_m^2} &= \frac{\partial^2 \mathbb{E}\{(X-Y)^2\}}{\partial x_m^2} - p_X(x_m)\bigg(\frac{\partial y_{m+1}}{\partial x_m}(2f(x_m) - \mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\} - \mathbb{E}\{f(X)|x \in \mathcal{V}_m\}) \\
&\quad - \frac{\partial y_m}{\partial x_m}(2f(x_m) - \mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\} - \mathbb{E}\{f(X)|x \in \mathcal{V}_m\}) + 2f'(x_m)(y_{m+1} - y_m) \\
&\quad + p_X(x_m)(y_{m+1} - y_m)\frac{f(x_m) - \mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\}}{\int\limits_{x_m}^{x_{m+1}} p_X(x)\mathrm{d}x} - p_X(x_m)(y_{m+1} - y_m)\frac{f(x_m) - \mathbb{E}\{f(X)|x \in \mathcal{V}_m\}}{\int\limits_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x}\bigg) \\
&= \frac{\partial^2 \mathbb{E}\{(X-Y)^2\}}{\partial x_m^2} + \beta p_X(x_m)(y_{m+1} - y_m)(2 + p_X(x_m)\frac{x_m - y_{m+1}}{\int\limits_{x_m}^{x_{m+1}} p_X(x)\mathrm{d}x} - p_X(x_m)\frac{x_m - y_m}{\int\limits_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x}) \\
&= \frac{\partial^2 \mathbb{E}\{(X-Y)^2\}}{\partial x_m^2} + \beta p_X(x_m)(y_{m+1} - y_m)(2 + \frac{\partial y_{m+1}}{\partial x_m} - \frac{\partial y_m}{\partial x_m}) \\
&= \frac{\partial^2 \mathbb{E}\{(X-Y)^2\}}{\partial x_m^2}.
\end{aligned}$$

where $\frac{\partial \mathbb{E}\{f(X)|x \in \mathcal{V}_m\}}{\partial x_m}$ and $\frac{\partial \mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\}}{\partial x_m}$ are computed as follows:

$$\frac{\partial \mathbb{E}\{f(X)|x \in \mathcal{V}_m\}}{\partial x_m} = p_X(x_m)\frac{f(x_m) - \mathbb{E}\{f(X)|x \in \mathcal{V}_m\}}{\int\limits_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x},$$

$$\frac{\partial \mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\}}{\partial x_m} = -p_X(x_m)\frac{f(x_m) - \mathbb{E}\{f(X)|x \in \mathcal{V}_{m+1}\}}{\int\limits_{x_m}^{x_{m+1}} p_X(x)\mathrm{d}x}.$$

For a fully revealing encoder $\frac{\partial^2 \mathbb{E}\{(X-Y)^2\}}{\partial x_m^2} > 0$. That is, if the encoder is fully revealing, then $f(x) = \alpha + \beta x$.

APPENDIX G
HIGH-RESOLUTION APPROXIMATION FOR DECODER'S ACTION IF $\eta_D(x,y)$ IS A BREGMAN LOSS FUNCTION

In the following, we approximate y_m associated with the region $[x_{m-1}, x_m]$ for an absolutely continuous source (wrt to Lebesque measure) with a density p_X. We take the midpoint of the region $\gamma_m = \frac{x_{m+1}+x_m}{2}$ and $\Delta_m = x_{m+1} - x_m$. Since $\eta_D(x,y)$ is a Bregman loss function, we have,

$$y_m = \frac{\int\limits_{\gamma_m-\Delta_m/2}^{\gamma_m+\Delta_m/2} x p_X(x)\mathrm{d}x}{\int\limits_{\gamma_m-\Delta_m/2}^{\gamma_m+\Delta_m/2} p_X(x)dx}$$

Taylor series expansion of p_X around γ_m:

$$p_X(x) = p_X(\gamma_m) + (x-\gamma_m)p'_X(\gamma_m) + \dots,$$

where $p'_X = \mathrm{d}p_X/\mathrm{d}x$. For approximation purposes, we take only the first two terms and plug $p_X(x)$ in y_m expression above, we obtain

$$\begin{aligned}
y_m &= \frac{\int\limits_{\gamma_m-\Delta_m/2}^{\gamma_m+\Delta_m/2} x(p_X(\gamma_m) + (x-\gamma_m)p'_X(\gamma_m))dx}{\int\limits_{\gamma_m-\Delta_m/2}^{\gamma_m+\Delta_m/2} (p_X(\gamma_m) + (x-\gamma_m)p'_X(\gamma_m))dx} = \frac{\int\limits_{\gamma_m-\Delta_m/2}^{\gamma_m+\Delta_m/2} ((p_X(\gamma_m) - \gamma_m p'_X(\gamma_m))x + x^2 p'_X(\gamma_m))\mathrm{d}x}{\int\limits_{\gamma_m-\Delta_m/2}^{\gamma_m+\Delta_m/2} (p_X(\gamma_m) - \gamma_m p'_X(\gamma_m) + x p'_X(\gamma_m))dx} \\
&= \frac{(p_X(\gamma_m) - \gamma_m p'_X(\gamma_m))\frac{x^2}{2} + p'_X(\gamma_m)\frac{x^3}{3}|_{\gamma_m-\Delta_m/2}^{\gamma_m+\Delta_m/2}}{(p_X(\gamma_m) - \gamma_m p'_X(\gamma_m))x + p'_X(\gamma_m)\frac{x^2}{2}|_{\gamma_m-\Delta_m/2}^{\gamma_m+\Delta_m/2})dx} = \frac{(p_X(\gamma_m) - \gamma_m p'_X(\gamma_m))\gamma_m + p'_X(\gamma_m)(\gamma_m^2 + \frac{\Delta_m^2}{12})}{p_X(\gamma_m) - \gamma_m p'_X(\gamma_m) + p'_X(\gamma_m)\gamma_m} \\
&= \frac{\gamma_m p_X(\gamma_m) + p'_X(\gamma_m)\frac{\Delta_m^2}{12}}{p_X(\gamma_m)},
\end{aligned}$$

that is,

$$y_m \approx \gamma_m + \frac{p'_X(\gamma_m)}{p_X(\gamma_m)}\frac{\Delta_m^2}{12}$$

We consider two densities as in our running examples: Uniform over $[0,1]$ for which this approximation yields $y_m \approx \gamma_m$ and unit-variance Gaussian for which it yields $y_m \approx (1 - \frac{\Delta_m^2}{12})\gamma_m$.

APPENDIX H
HIGH-RESOLUTION ANALYSIS FOR $\eta_E(x,y) = (x - y^3)$ WITH $\eta_D(x,y)$ A BREGMAN LOSS FUNCTION

We first show an example computation of $\mathcal{X}_{NR}$ for the setting with $X \sim \mathcal{N}(0,1)$ below.

Let $\mathcal{X}_R = [t_1, t_2]$. Then, $\mathcal{V}_1 = (-\infty, t_1), \mathcal{X}_{NR} = [t_1, t_2), \mathcal{V}_{M+1} = [t_2, \infty)$, $y_1 = \mathbb{E}\{X|x \in \mathcal{V}_1\}, y_{M+1} = \mathbb{E}\{X|x \in \mathcal{V}_{M+1}\}$. In the intervals of the source that is quantized at high rates, $y_m \approx x$ since the step size is negligible and decoder reconstruction is $\mathbb{E}\{X|x \in \mathcal{V}_m\}$. The encoder distortion at high rates can be approximated as,

$$D(t_1, t_2) = \int\limits_{-\infty}^{t_1} (x^3 - y_1)^2 p_X(x)\mathrm{d}x + \int\limits_{t_1}^{t_2} (x^3 - x)^2 p_X(x)\mathrm{d}x + \int\limits_{t_2}^{\infty} (x^3 - y_{M+1})^2 p_X(x)\mathrm{d}x \tag{24}$$

It is not possible to find t_1, t_2 in closed form by enforcing KKT conditions since t_1, t_2 are in the limits of integrals in y_1, y_3, respectively. Instead, we plot (24) and find that $t_1 = -0.378, t_2 = 0.378$ minimizes $D(t_1, t_2)$.

We now consider the high-resolution analysis of a general source $X \sim p_X(x)$.

The encoder distortion,

$$D_E = D_{\mathcal{X}_{NR}} + D_{\mathcal{X}_R},$$

where $D_{\mathcal{X}_{NR}}$ is the distortion in not quantizing $\mathcal{V}_m \in \mathcal{X}_{NR}$ and $D_{\mathcal{X}_R}$ is the high-resolution distortion in quantizing $\mathcal{X}_R$ at high rates,

$$D_{\mathcal{X}_{NR}} = \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_{NR}}}^{M} \int\limits_{x_{m-1}}^{x_m} (x - y_m^3)^2 p_X(x)\mathrm{d}x = \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_{NR}}}^{M} \int\limits_{x_{m-1}}^{x_m} x^2 p_X(x)\mathrm{d}x + \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_{NR}}}^{M} y_m^6 \int\limits_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x - 2\sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_{NR}}}^{M} y_m^3 \int\limits_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x.$$

The term $D_{\mathcal{X}_R}$ can be simplified by approximations resulting from Assumptions 6, 7, 8 as given in Section V.

At high-resolution, the probability of each region can be assumed constant (7),

$$D_{\mathcal{X}_R} = \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} (x-y_m^3)^2 p_X(x)\mathrm{d}x \approx \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} (x-y_m^3)^2 p_m \mathrm{d}x.$$

The term p_m can be approximated using (8),

$$D_{\mathcal{X}_R} \approx \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} f_m \int_{x_{m-1}}^{x_m} \frac{(x-y_m^3)^2}{\triangle_m}\mathrm{d}x.$$

We approximate the reconstruction levels as $y_m = x_{m-1} + \Delta_m/2$ from (9). The encoder distortion,

$$\begin{aligned}
D_{\mathcal{X}_R} &= \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} f_m \frac{1}{\Delta_m} \int_{y_m-\Delta_m/2}^{y_m+\Delta_m/2} (x^2 - 2xy_m^3 + y_m^6)\mathrm{d}x = \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} f_m \frac{1}{\triangle_m}\left(\frac{x^3}{3} - y_m^3 x^2 + y_m^6 x)\right|_{y_m-\Delta_m/2}^{y_m+\Delta_m/2} \\
&= \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} f_m \frac{1}{\triangle_m}\Bigg(\frac{(y_m+\frac{\Delta_m}{2})^3 - (y_m-\frac{\Delta_m}{2})^3}{3} - y_m^3\bigg(\Big(y_m+\frac{\Delta_m}{2}\Big)^2 - \Big(y_m-\frac{\Delta_m}{2}\Big)^2\bigg) \\
&\quad + y_m^6\bigg(\Big(y_m+\frac{\Delta_m}{2}\Big) - \Big(y_m-\frac{\Delta_m}{2}\Big)\bigg)\Bigg) \\
&= \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} f_m \frac{1}{\triangle_m}\Bigg(\frac{1}{3}\Delta_m\bigg(\Big(y_m+\frac{\Delta_m}{2}\Big)^2 + \Big(y_m-\frac{\Delta_m}{2}\Big)^2 + \Big(y_m+\frac{\Delta_m}{2}\Big)\Big(y_m-\frac{\Delta_m}{2}\Big)\bigg) - 2y_m^4\Delta_m + y_m^6\Delta_m\Bigg) \\
&= \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} f_m \frac{1}{\triangle_m}\Bigg(\frac{1}{3}\Delta_m\bigg(y_m^2 + y_m\Delta_m + \frac{\Delta_m^2}{4} + y_m^2 - \frac{\Delta_m^2}{4} + y_m^2 - \Delta_m y_m + \frac{\Delta_m^2}{4}\bigg) - 2y_m^4\Delta_m + y_m^6\Delta_m\Bigg) \\
&= \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} f_m \frac{1}{\triangle_m}\Bigg(\frac{1}{3}\Delta_m\bigg(3y_m^2 + \frac{\Delta_m^2}{4}\bigg) - 2y_m^4\Delta_m + y_m^6\Delta_m\Bigg) = \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} f_m\bigg(y_m^2 + \frac{\Delta_m^2}{12} - 2y_m^4 + y_m^6\bigg).
\end{aligned}$$

Approximating Δ_m as $1/(M\lambda)$, as given in (11), the encoder distortion,

$$D_{\mathcal{X}_R} \approx \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} f_m\bigg(y_m^2 + y_m^6 - 2y_m^4 + \frac{(M\lambda(y_m))^{-2}}{12}\bigg).$$

Converting the above summation to integrals over $\mathcal{X}_R$,

$$D_{\mathcal{X}_R} \approx \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} p_X(y)(y^2+y^6-2y^4)\mathrm{d}y + \frac{1}{12}\frac{1}{M^2} \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} p_X(y)\lambda(y)^{-2}\mathrm{d}y. \tag{25}$$

Then, the encoder distortion

$$\begin{aligned}
D_E \approx & \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_{NR}}}^{M} y_m^6 \int_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x - 2\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_{NR}}}^{M} y_m^3 \int_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x + \int_a^b y^2 p_X(y)\mathrm{d}y \\
& + \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} p_X(y)(y^6-2y^4)\mathrm{d}y + \frac{1}{12}\frac{1}{M^2} \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} p_X(y)\lambda(y)^{-2}\mathrm{d}y.
\end{aligned}$$

The above equation can be approximated by applying Holder's inequality,

$$\bigg(\int u(y)v(y)\mathrm{d}y\bigg) \le \bigg(\int u(y)^{t_1}\mathrm{d}y\bigg)^{1/t_1}\bigg(\int v(y)^{t_2}\mathrm{d}y\bigg)^{1/t_2} \tag{26}$$

where $1/t_1 + 1/t_2 = 1$, with $u(y) = (p_X(y)/\lambda^2(y))^{1/3}, v(y) = \lambda(y)^{2/3}$ and $t_1 = 3, t_2 = 3/2$,

$$\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int\limits_{x_{m-1}}^{x_m} p_X(y)^{1/3}\mathrm{d}y\Bigg) \leq \Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int\limits_{x_{m-1}}^{x_m} \frac{p_X(y)}{\lambda(y)^2}\mathrm{d}y\Bigg)^{1/3}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int\limits_{x_{m-1}}^{x_m} \lambda(y)\mathrm{d}y\Bigg)^{2/3} \overset{a}{=} \Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int\limits_{x_{m-1}}^{x_m} \frac{p_X(y)}{\lambda(y)^2}\mathrm{d}y\Bigg)^{1/3}, \quad (27)$$

where equality a above is from (10). Then, the encoder's distortion at high-resolution can be approximated as,

$$D_E = D_M + D'_M, \quad (28)$$

where D_M depends on M and D'_M does not, and,

$$D_M = \frac{1}{12}\frac{1}{M^2}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int\limits_{x_{m-1}}^{x_m} p_X(y)^{1/3}\mathrm{d}y\Bigg)^3,$$

$$D'_M = \mathbb{E}\{X^2\} + \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M}\int\limits_{x_{m-1}}^{x_m} p_X(y)(y^6 - 2y^4)\mathrm{d}y + \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_{NR}}}^{M} y_m^6\int\limits_{x_{m-1}}^{x_m} p_X(x)\mathrm{d}x - 2\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_{NR}}}^{M} y_m^3\int\limits_{x_{m-1}}^{x_m} x p_X(x)\mathrm{d}x.$$

Note that D'_M does not depend on M since the set of non-revealing regions $\mathcal{V}_m \in \mathcal{X}_{NR}$ does not change with M. The regions $\mathcal{V}_m \in \mathcal{X}_{NR}$ and their corresponding decoder actions $\{y_m|\mathcal{V}_m \in \mathcal{X}_{NR}\}$ can be found without using high-resolution analysis, as shown in the example above.

The encoder distortion converges to D'_M as M increases since D_M decreases by a factor of $1/M^2$ as M increases.

Appendix I
High-resolution analysis for $\eta_E(x, y) = (x^3 - y)^2$ with $\eta_D(x, y)$ a Bregman loss function

We first consider examples of computation of $\mathcal{X}_{NR}$ for three cases: $X \sim U[0, 1]$, $X \sim U[-1, 1]$, $X \sim \mathcal{N}(0, 1)$.

Consider $X \sim U[0, 1]$. Let $\mathcal{X}_{NR} = [0, t]$. Then, $\mathcal{V}_{M+1} = [t, 1)$. In the intervals of the source that is quantized at high rates, $y_m \approx x$ since the step size is negligible and decoder reconstruction is $\mathbb{E}\{X|x \in \mathcal{V}_m\}$. The encoder distortion at high rates can be approximated by,

$$D(t) = \int\limits_0^t (x^3 - \frac{t}{2})^2\mathrm{d}x + \int\limits_t^1 (x^3 - x)^2\mathrm{d}x. \quad (29)$$

Enforcing KKT conditions of optimality, we get $t = 0, t = 1/\sqrt{3} = 0.5773$ as possible solutions. Substituting these values in (29), $D(t = 0) = 0.0762$, $D(t = 1/\sqrt{3}) = 0.0698$, i.e., the encoder does not quantize $[0, 1/\sqrt{3}]$.

Consider $X \sim U[-1, 1]$. Let the non-revealing region be $[t_1, t_2] \subset [-1, 1]$. The encoder distortion can be approximated at high rates as,

$$D = \frac{1}{2}\Bigg(\int\limits_{-1}^{t_1} (x^3 - x)^2\mathrm{d}x + \int\limits_{t_1}^{t_2} (x^3 - \frac{t_1 + t_2}{2})^2\mathrm{d}x + \int\limits_{t_2}^{1} (x^3 - x)^2\mathrm{d}x\Bigg). \quad (30)$$

Enforcing KKT conditions, $\partial D/\partial t_1 = \partial D/\partial t_2 = 0$,

$$\frac{\partial D}{\partial t_1} = \frac{1}{2}\Bigg((t_1^3 - t_1)^2 - (t_1^3 - \frac{t_1 + t_2}{2})^2 - \int\limits_{t_1}^{t_2} (x^3 - \frac{t_1 + t_2}{2})\mathrm{d}x\Bigg)$$

$$\frac{\partial D}{\partial t_2} = \frac{1}{2}\Bigg(-(t_2^3 - t_2)^2 + (t_2^3 - \frac{t_1 + t_2}{2})^2 - \int\limits_{t_1}^{t_2} (x^3 - \frac{t_1 + t_2}{2})\mathrm{d}x\Bigg),$$

which gives the following equations

$$(t_1^3 - t_1)^2 - (t_1^3 - \frac{t_1 + t_2}{2})^2 - \frac{t_2^4 - t_1^4}{4} + \frac{t_2^2 - t_1^2}{2} = 0,$$
$$-(t_2^3 - t_2)^2 + (t_2^3 - \frac{t_1 + t_2}{2})^2 - \frac{t_2^4 - t_1^4}{4} + \frac{t_2^2 - t_1^2}{2} = 0.$$

Solving the above set of two equations, we obtain the solutions: $t_1 = -1/\sqrt{2}, t_2 = 1/\sqrt{2}$, $t_1 = t_2$, $t_1 = t_2 = 1/\sqrt{6}$, $t_1 = t_2 = -1/\sqrt{6}$, and a non-viable solution $t_1 = 1/\sqrt{2}, t_2 = -1/\sqrt{2}$ ($t_1 < t_2$ is not satisfied). The solutions that involve $t_1 = t_2$ imply there is no non-revealing region.

Substituting these values of t_1, t_2 in the (30), $D(t_1 = -1/\sqrt{2}, t_2 = 1/\sqrt{2}) = 0.0290$, $D(t_1 = t_2) = 0.0762$, i.e., the encoder does not quantize $[-1/\sqrt{2}, 1/\sqrt{2}]$.

Consider $X \sim \mathcal{N}(0, 1)$. Let $\mathcal{X}_{NR} = [t_1, t_2]$. Then, the encoder distortion at high rates can be approximated as,

$$D(t_1, t_2) = \int\limits_{-\infty}^{t_1} (x^3 - x)^2 p_X(x)\mathrm{d}x + \int\limits_{t_1}^{t_2} (x^3 - y)^2 p_X(x)\mathrm{d}x + \int\limits_{t_2}^{\infty} (x^3 - x)^2 p_X(x)\mathrm{d}x, \tag{31}$$

where $y = \mathbb{E}\{X | x \in \mathcal{X}_{NR}\}$.

In the previous example, KKT conditions were enforced and we found t in closed form. However, in this example, since y_2 involves t_1, t_2 in limits of integrals, it is not possible to compute the values of t_1, t_2 in closed form. Instead we find the values by plotting the cost curve, and we find $t_1 = -1/\sqrt{2}, t_2 = 1/\sqrt{2}$ minimizes (31).

We now consider the high-resolution analysis of a general source $X \sim p_X(x)$.

The encoder's distortion can be simplified as follows

$$\begin{aligned} D &= \mathbb{E}\{(X^3 - Y)^2\} = \mathbb{E}\{(X^3 - X + X - Y)^2\} = \mathbb{E}\{(X^3 - X)^2\} + \mathbb{E}\{(X - Y)^2\} + 2\mathbb{E}\{(X^3 - X)(X - Y)\} \\ &= \mathbb{E}\{X^6\} - 2\mathbb{E}\{X^4\} + \mathbb{E}\{X^2\} + 2\mathbb{E}\{X^4\} - 2\mathbb{E}\{X^2\} + \mathbb{E}\{(X - Y)^2\} - 2\mathbb{E}\{(X^3 - X)Y\} \\ &= \mathbb{E}\{X^6\} - \mathbb{E}\{X^2\} + \mathbb{E}\{(X - Y)^2\} - 2\mathbb{E}\{(X^3 - X)Y\} \\ &= \mathbb{E}\{X^6\} - \mathbb{E}\{X^2\} + D_{mse} - 2D_1 \end{aligned} \tag{32}$$

where D_{mse}, D_1 are

$$D_{mse} = \mathbb{E}\{(X - Y)^2\} = D_{mse,\mathcal{X}_R} + D_{mse,\mathcal{X}_{NR}}, \quad D_1 = \mathbb{E}\{(X^3 - X)Y\} = D_{1,\mathcal{X}_R} + D_{1,\mathcal{X}_{NR}},$$

and $D_{mse,\mathcal{X}_R}, D_{mse,\mathcal{X}_{NR}}, D_{1,\mathcal{X}_R}, D_{1,\mathcal{X}_{NR}}$ are

$$D_{mse,\mathcal{X}_R} = \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} \int\limits_{x_{m-1}}^{x_m} (x - y_m)^2 p_X(x)\mathrm{d}x, \quad D_{mse,\mathcal{X}_{NR}} = \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_{NR}}}^{M} \int\limits_{x_{m-1}}^{x_m} (x - y_m)^2 p_X(x)\mathrm{d}x,$$

$$D_{1,\mathcal{X}_R} = \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} \int\limits_{x_{m-1}}^{x_m} (x^3 - x) y_m p_X(x)\mathrm{d}x, \quad D_{1,\mathcal{X}_{NR}} = \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_{NR}}}^{M} \int\limits_{x_{m-1}}^{x_m} (x^3 - x) y_m p_X(x)\mathrm{d}x.$$

To approximate $D_{1,\mathcal{X}_R}$ and $D_{mse,\mathcal{X}_R}$, we follow a similar analysis as detailed in Appendix H, with Assumptions 6, 7, 8.

The probability of each region can be assumed constant (7),

$$D_{1,\mathcal{X}_R} = \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} \int\limits_{x_{m-1}}^{x_m} (x^3 - x) y_m p_X(x)\mathrm{d}x \approx \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} \int\limits_{x_{m-1}}^{x_m} (x^3 - x) y_m p_m \mathrm{d}x.$$

The term p_m can be approximated using (8),

$$D_{1,\mathcal{X}_R} \approx \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} f_m \int\limits_{x_{m-1}}^{x_m} \frac{(x^3 - x) y_m}{\Delta_m} \mathrm{d}x.$$

Approximating $y_m = x_m - \Delta_m/2$ as in (9),

$$\begin{aligned} D_{1,\mathcal{X}_R} &= \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} f_m \frac{1}{\Delta_m} y_m \int\limits_{y_m - \frac{\Delta_m}{2}}^{y_m + \frac{\Delta_m}{2}} (x^3 - x)\mathrm{d}x = \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} f_m \frac{1}{\Delta_m} y_m \left(\frac{x^4}{4} - \frac{x^2}{2}\right)\Bigg|_{y_m - \frac{\Delta_m}{2}}^{y_m + \frac{\Delta_m}{2}} \\ &= \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} f_m \frac{1}{\Delta_m} y_m \left(\frac{(y_m + \frac{\Delta_m}{2})^4 - (y_m - \frac{\Delta_m}{2})^4}{4} - \frac{(y_m + \frac{\Delta_m}{2})^2 - (y_m - \frac{\Delta_m}{2})^2}{2}\right) \\ &= \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_R}}^{M} f_m y_m \left(y_m \left(y_m^2 + \frac{\Delta_m^2}{4}\right) - y_m\right). \end{aligned}$$

Approximating Δ_m as $1/(M\lambda(y_m))$, as given in (11), and converting back to integrals,

$$D_{1,\mathcal{X}_R} \approx \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} p_X(y)(y^4 + y^2\frac{(M\lambda(y))^{-2}}{4} - y^2)\mathrm{d}y \tag{33}$$

$$\approx \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} p_X(y)(y^4 - y^2)\mathrm{d}y + \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} p_X(y)(y^2\frac{(M\lambda(y))^{-2}}{4})\mathrm{d}y.$$

Using Holder's inequality (26) with $u(y) = (y^2 p_X(y)/\lambda^2(y))^{1/3}, v(y) = \lambda(y)^{2/3}, t_1 = 3, t_2 = 3/2$,

$$\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} (y^2 p_X(y))^{1/3}\mathrm{d}y\Bigg) \leq \Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} \frac{y^2 p_X(y)}{\lambda(y)^2}\mathrm{d}y\Bigg)^{1/3}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} \lambda(y)\mathrm{d}y\Bigg)^{2/3} \overset{a}{=} \Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} \frac{y^2 p_X(y)}{\lambda(y)^2}\mathrm{d}y\Bigg)^{1/3},$$

where equality a above comes from (10).

This expression can be approximated as

$$D_{1,\mathcal{X}_R} \approx \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} p_X(y)(y^4 - y^2)\mathrm{d}y + \frac{1}{4M^2}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} (y^2 p_X(y))^{1/3}\mathrm{d}y\Bigg)^3.$$

Following a similar analysis for $D_{mse,\mathcal{X}_R}$,

$$D_{mse,\mathcal{X}_R} = \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} (x - y_m)^2 p_X(x)\mathrm{d}x.$$

The probability of each region is assumed as constant (7),

$$D_{mse,\mathcal{X}_R} = \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} (x - y_m)^2 p_m\mathrm{d}x.$$

The term p_m is approximated as (8),

$$D_{mse,\mathcal{X}_R} = \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \frac{f_m}{\Delta_m} \int_{x_{m-1}}^{x_m} (x - y_m)^2\mathrm{d}x.$$

Using the approximation for y_m in (9), $y_m = x_m - \Delta_m/2$,

$$D_{mse,\mathcal{X}_R} = \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \frac{f_m}{\Delta_m} \int_{y_m-\frac{\Delta_m}{2}}^{y_m+\frac{\Delta_m}{2}} (x - y_m)^2\mathrm{d}x = \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \frac{f_m}{\Delta_m}\frac{1}{3}(x - y_m)^3\Bigg|_{y_m-\frac{\Delta_m}{2}}^{y_m+\frac{\Delta_m}{2}} = \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \frac{1}{12} f_m \Delta_m^2$$

$$= \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} \frac{1}{12} p_X(x)\Delta_m^2\mathrm{d}x \overset{a}{=} \sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} \frac{1}{12} p_X(x)(M\lambda(x))^{-2}\mathrm{d}x \overset{b}{=} \frac{1}{12}\frac{1}{M^2}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} p_X(y)^{1/3}\mathrm{d}y\Bigg)^3,$$

where equality a is from (11), and equality b is from (27).

Substituting D_1 and D_{mse} back in (32), the high-resolution approximation of the encoder's distortion is

$$D_E = D_M + D'_M, \tag{34}$$

where D_M depends on M and D'_M does not, and,

$$D_M = \frac{1}{12}\frac{1}{M^2}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} p_X(x)^{1/3}\mathrm{d}x\Bigg)^3 - \frac{1}{2}\frac{1}{M^2}\Bigg(\sum_{\substack{m=1\\ \mathcal{V}_m\in\mathcal{X}_R}}^{M} \int_{x_{m-1}}^{x_m} (x^2 p_X(x))^{1/3}\mathrm{d}x\Bigg)^3,$$

$$D'_M = \mathbb{E}\{X^6\} - \mathbb{E}\{X^2\} + \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_{NR}}}^{M} \int_{x_{m-1}}^{x_m} (x-y_m)^2 p_X(x)\mathrm{d}x - 2 \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_{R}}}^{M} \int_{x_{m-1}}^{x_m} (x^4-x^2) p_X(x)\mathrm{d}x$$
$$- 2 \sum_{\substack{m=1 \\ \mathcal{V}_m \in \mathcal{X}_{NR}}}^{M} \int_{x_{m-1}}^{x_m} (x^3-x) y_m p_X(x)\mathrm{d}x$$

Note that D'_M does not depend on M since the set of non-revealing regions $\mathcal{V}_m \in \mathcal{X}_{NR}$ does not change with M. The regions $\mathcal{V}_m \in \mathcal{X}_{NR}$ and their corresponding decoder actions $\{y_m|\mathcal{V}_m \in \mathcal{X}_{NR}\}$ can be found without using high-resolution analysis, as shown in the example above.

The encoder distortion converges to D'_M as M increases, however, the encoder distortion may be increasing, decreasing, or constant as M increases depending on whether $D_M < 0, D_M > 0$ or $D_M = 0$.

References

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

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

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

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

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

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

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

[8] 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.

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

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

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

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

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

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

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

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

[18] 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.

[19] ——, "Channel-Optimized Strategic Quantizer Design via Dynamic Programming," in *IEEE Statistical Signal Processing Workshop (SSP)*, 2023, pp. 621–625.

[20] ——, "Strategic Quantization over a Noisy Channel," in *Conference Record of the Fifty Seventh Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2023.

[21] ——, "Strategic quantization of a noisy source," in *2023 59th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2023, pp. 1–7.

[22] ——, "Strategic Quantization with Quadratic Distortion Measures," in *IEEE International Conference on Communications Workshops (ICC Workshops), in review*, 2024.

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

[24] A. Trushkin, "Sufficient Conditions for Uniqueness of a Locally Optimal Quantizer for a Class of Convex Error Weighting Functions," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 187–198, 1982.

[25] J. Kieffer, "Uniqueness of Locally Optimal Quantizer for Log-concave Density and Convex Error Weighting Function," *IEEE Transactions on Information Theory*, vol. 29, no. 1, pp. 42–47, 1983.

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

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

[28] 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.

[29] D. Sharma, "Design of Absolutely Optimal Quantizers for a Wide Class of Distortion Measures," *IEEE Transactions on Information Theory*, vol. 24, no. 6, pp. 693–702, 1978.

[30] I. Arieli, Y. Babichenko, R. Smorodinsky, and T. Yamashita, "Optimal persuasion via bi-pooling," *Theoretical Economics*, vol. 18, no. 1, pp. 15–36, 2023.

[31] O. Candogan, "Optimality of Double Intervals in Persuasion: A Convex Programming Framework," *Available at SSRN 3452145*, 2019.

[32] A. Banerjee, X. Guo, and H. Wang, "On the Optimality of Conditional Expectation as a Bregman predictor," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2664–2669, 2005.

[33] E. Akyol and A. Anand, "Strategic Quantization at High Resolution," in *IEEE International Symposium on Information Theory (ISIT), in review*, 2024.