Strategic Quantization

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

***Abstract*—This paper is concerned with the quantization setting where the encoder and the decoder have misaligned objectives. Although the unconstrained version of this problem has been extensively studied within the field of information design problems in Economics, the problem becomes more relevant for engineering applications when the cardinality of the message space is constrained. We first motivate the problem via a toy example demonstrating the strategic quantization problem's intricacies, explicitly showing that the quantization resolution can change the nature of optimal encoding policy and the iterative optimization of the decoder and the encoder mappings may not converge to a local optimum solution. As a remedy, we develop a gradient-descent based solution. We analyze the poor local optimal optima issues associated with the optimization method, and show that even for well-behaving sources, such as Gaussian, there are multiple local minima, depending on the distortion measures chosen, in sharp contrast with the classical quantization. We finally present numerical results obtained via the proposed algorithms that suggest the proposed algorithms' validity and demonstrate strategic quantization features that differentiate it from its classical counterpart.**

***Index Terms*—Quantization, Game theory, Gradient descent**

I. Introduction

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

The setting without the quantization aspect (in the practical sense, if M is asymptotically large) is known in the Economics literature as the information design, or the Bayesian persuasion problem [1], [2]. These problems analyze how a communication system designer (sender) can use the information to influence the action taken by a receiver. This framework has proven beneficial in analyzing a variety of real-life applications, such as the design of transcripts when schools compete to improve their students' job prospects [3] and voter mobilization and gerrymandering [4], as well as various engineering applications, including in modeling misinformation spread over social networks [5] and privacy-constrained information processing [6], and many more [7]. For an excellent survey of the related literature in Economics, see [8], [9].

The strategic quantization problem, as described above, was discussed in a few contemporary economics and computer science studies. In [10], authors analyze the problem via a computation lens and report approximate results on this problem, relating to another problem they solved conclusively. In one of their main results, the algorithmic complexity of finding the optimum strategic quantizer was shown to be NP-hard. In a recent working paper, Aybaş and Türkel [11] analyzed this problem using the methods in [2] and provided several theoretical properties of strategic quantization. A byproduct of their analysis yields a constructive method for deriving optimal quantizers based on a search over possible posterior distributions over their feasible set. Our objective here is to leverage the rich collection of results in quantization theory, e.g., the comprehensive survey of results by Gray and Neuhoff [12], to study the same problem via the engineering lens.

We note in passing that quantizers also arise as equilibrium strategies endegenously, i.e., without an external constraint, in a related but a distinctly different class of signaling games, namely the cheap talk [13]. In [13], 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 committed to the designed Q afterward. This commitment is known to the decoder and establishes a form of trust between the sender and the receiver, resulting in possibly higher payoffs for both agents. This difference also manifests itself in the notion of equilibrium we are seeking here since the encoder does not necessarily form the best response to the decoder due to its commitment to Q[1].

The problem setting has several applications in engineering as well as Economics. For an engineering application, consider the Internet of Things, where agents with misaligned objectives communicate over channels with delay constraints. For a more concrete, real-life application, consider two smart cars by competing manufacturers, e.g., Tesla and Honda, where the Tesla (decoder) car asks for a piece of specific information,

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

[1]These issues are well understood in the Economics literature, see, e.g., [14] for an excellent survey. However, we emphasize them here for a reader with an engineering background; see [7] for a detailed discussion through the engineering lens.

such as traffic congestion, from the Honda (encoder) to decide on changing its route or not. Say Honda's objective is to make Tesla take a specific action, e.g., to change its route, while Tesla's objective is to estimate the congestion to make the right decision accurately. Honda's objective is obviously different from that of Tesla, hence has no incentive to convey a truthful congestion estimate. However, Tesla is aware of Honda's motives while still would like to use Honda's information (if possible). With the realistic assumption that Honda would observe this information through a noisy sensing channel (i.e., a sensor), how would these cars communicate over a fixed-rate zero-delay channel? Such problems can be handled using our model. Note that here Honda has three different behavioral choices: it can choose not to communication (non-revealing strategy), can communicate exactly what the Tesla wants (fully-revealing strategy), or it can craft a message that would make Tesla to change its route. Note that Tesla can choose not to use Honda's message, if it is statistically too far from the truth. Hence, crafting an optimal message for Honda that would serve its own objective, knowing that Tesla's objective differs from it, is a formidable research challenge.

II. PRELIMINARIES

A. Notation

In this paper, random variables are denoted using capital letters (say X), their sample values with respective lower case letters (x), and their alphabet with respective calligraphic letters ($\mathcal{X}$). The set of real numbers is denoted by $\mathbb{R}$. 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 $t_1 \leq i \leq t_2, i \in \mathbb{Z}_{\geq 0}$ is denoted by $i \in [t_1 : t_2]$.

B. Strategic Quantization Prior Work

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

As mentioned earlier, our problem is a variation of the Bayesian Persuasion (or information design) class of problems where the encoder and the decoder with misaligned objectives communicate [2]. This class of the problems have been an active research area in Economics due to their modeling abilities of real-life scenarios, see e.g., [1], [3], [9], [14].

This problem was previously studied in Economics as well as Computer Science. In [11], authors showed the existence of optimal strategic quantizers in abstract spaces. Moreover, authors provide a low-complexity method to obtain the optimal strategic quantizer. In [15], [16], authors characterize sufficient conditions for the monotonicity of the optimal strategic quantizer, and as a byproduct of their analysis, characterize its behavior (non-revealing, fully revealing, or partially revealing) for some special settings. In Computer Science, in []

In [17], we showed that a strategic variation of the Lloyd-Max algorithm does not converge to a locally optimal solution. As a remedy we develop a gradient descent based solution for this problem. We also demonstrated that even for well-behaving sources, such as scalar Gaussian, there are multiple local optima, depending on the distortion measures chosen, in sharp contrast with the classical quantization for which the local optima is unique for the case of log-concave sources (which include Gaussian sources). We also analyzed the behavior of the optimal strategic quantizer for some typical settings. The behavior can be one of the following three:

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

In [18], [19], we carried out our analysis of strategic quantization to the scenario where there is a noisy communication channel between the encoder and the decoder, using random index mapping in conjunction with gradient descent based and dynamic programming solutions respectively. In [19], [20], we derived the globally optimal strategic quantizer via a dynamic programming based solution to resolve the poor local minima issues with gradient descent based solutions since the objective function is non-convex.

III. PROBLEM DEFINITION

Consider the following quantization problem: an encoder observes a realization of a scalar source $X \in \mathcal{X}$ with probability distribution μ. The encoder maps X to a message $Z \in \mathcal{Z}$, where $\mathcal{Z}$ is a set of discrete messages with a cardinality constraint $|\mathcal{Z}| \leq M$ 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 ($\eta_E \neq \eta_D$). The encoder designs Q *ex-ante*, i.e., without the knowledge of the realization of X, using only the objectives η_E and η_D, and the statistics of the source $\mu_X(\cdot)$. The objectives (η_E and η_D), the shared prior (μ), and the mapping (Q) are known to the encoder and the decoder. The problem is to design Q for the equilibrium, i.e., the encoder minimizes its distortion if used with a corresponding decoder that minimizes its own distortion. This communication setting is given in Figure 1.

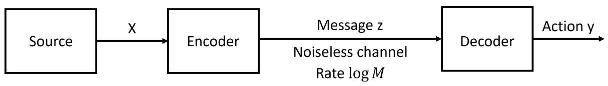


Fig. 1. Communication diagram over a noiseless channel

IV. Model

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

$$y_m = Q(x) \quad \forall x \in \mathcal{V}_m \tag{1}$$

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

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. [21]. Note that the second regularity condition, $y_m \in \mathcal{V}_m$, is not included in Assumption 1.

In the Economics parlance, Assumption 1 is referred as the "monotonicity" condition. In [15], [16] 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 [15], [16] would be applicable here. Under Assumption 1, $\mathcal{V}_m$ is an interval, i.e.,

$$\mathcal{V}_m = [q_{m-1}, q_m) \tag{2}$$

where $a = q_0 < q_1 < \ldots < q_M = b$. The encoder chooses the boundary decision levels $\mathbf{q} = [q_0, q_1, \ldots, q_M]$. 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

$$y_m^* = \underset{y_m \in \mathcal{Y}}{\arg\min}\, \mathbb{E}\{\eta_D(x, y_m) | x \in \mathcal{V}_m\} \quad \forall m \in [1:M] \tag{3}$$

Hence, the decoder chooses the actions $\{y_m\}$ knowing the set of decision sets $\{\mathcal{V}_m\}$. 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:

$$\{\mathcal{V}_m^*\} = \underset{\{\mathcal{V}_m\}}{\arg\min} \sum_{m=1}^{M} \mathbb{E}\{\eta_E(x, y_m(\{\mathcal{V}_m\}) | x \in \mathcal{V}_m\} \tag{4}$$

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

$$\mathbf{q}^* = \underset{\mathbf{q}}{\arg\min} \sum_{m=1}^{M} \mathbb{E}\{\eta_E(x, y_m(\mathbf{q})) | x \in \mathcal{V}_m\} \tag{5}$$

Remark 2: A key consideration here is that the encoder is committed to its choice of $\mathbf{q}$, it cannot determine $\mathbf{q}$ as the best response to $\mathbf{y}$. Hence, while the decoder can optimize its action $\mathbf{y}$ as the best response to $\mathbf{q}$, the encoder cannot choose $\mathbf{q}$ as the best response to $\mathbf{y}$, but that to a function of itself, i.e., the best response to $\mathbf{y}(\mathbf{q})$. This aspect of the problem introduces a hierarchy in the game play (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 [13], 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 [13].

V. Interesting Aspects of Strategic Quantization

In this subsection, we provide simple numerical examples to demonstrate few intricacies of the strategic quantization problem that differentiate it from its classical analogue.

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). Hence we chose measures that yield results that demonstrate the interesting aspects of strategic quantization.

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

As mentioned earlier, the classical quantization problem is a team problem, i.e., the encoder and the decoder share identical objectives. In strategic quantization, this is no longer the case. Hence, depending on the misalignment between η_E and η_D, the encoder might choose not to send any information to the decoder (e.g., if they are too misaligned), which we refer to as "non-revealing" policy following the convention in the Economics literature. Alternatively, the strategic quantizer problem might simplify to classical quantization with a common distortion measure η_D, i.e., the encoder cannot utilize its information design advantage to persuade the decoder to take a specific action (referred to here as "fully-revealing" encoder policy). Finally, the encoder might choose to employ a quantizer that is not identical to the classical one, i.e., "partially-revealing" policy. It is relatively straightforward to provide an example for the first case; consider, e.g., $\eta_E(x,y) = -\eta_D(x,y)$, $\forall x, y$ which makes the problem a zero-sum game, hence, the optimal strategy for the encoder is not to send any information (see also the analysis below). We demonstrate the latter cases via a simple numerical example.

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

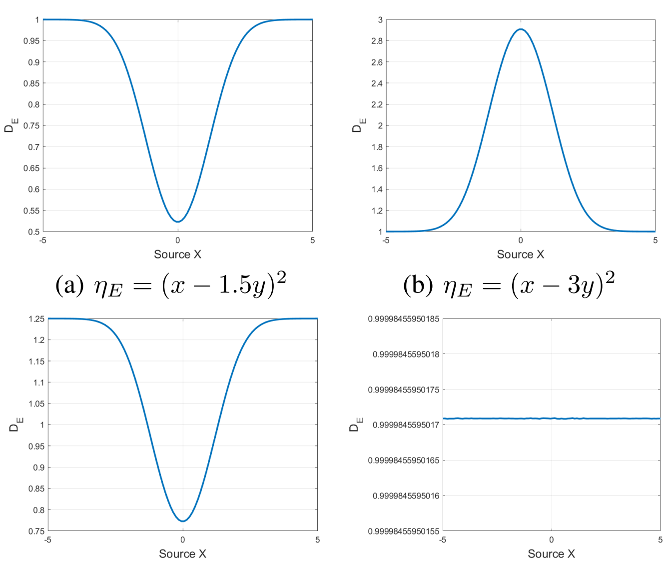
In Appendix B we prove the following result, which is an extension of a result presented in [17], as well as in [15]:

Theorem 2: For $\eta_E(x, y) = (x + \alpha - \beta y)^2$ and $\eta_D(x, y) = (x - y)^2$, the optimal strategic quantizer Q is:

$$Q(x) = \left\{ \begin{array}{ll} \arg\min \mathbb{E}\{(X - Q(X))^2\}, & \text{for } 0 < \beta < 2 \\ \textit{arbitrary}, & \text{for } \beta = 0, 2 \\ \textit{constant}, & \text{otherwise} \end{array} \right\}.$$

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

We see from Theorem 2 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 $\beta = 2$. Figures 2a,b show the distortion to the encoder for $\beta = 1.5, 3$ in quantizing a Gaussian source $X \sim \mathcal{N}(0, 1)$ to $M = 2$ levels. We see that $\beta = 1.5$ has an optima only for the quantizer $[a, 0, b]$ which is the non-strategic quantizer, and that $\beta = 3$ is optimal when the quantizer is $[a, b]$ (non-revealing). Figure 2c shows the encoder's distortion when $\eta_E = (x + 0.5 - 1.5y)^2$ and $\eta_D = (x - y)^2$. We see that the optimal point does not change although the distortion values are shifted. Equation 12 in Appendix B shows that the terms containing α in the encoder's distortion comes from the following terms: $\alpha^2 + 2\alpha(1-\beta)\mathbb{E}\{X\}$. As seen from Figures 2a,c, this evaluates to 0.25 for our example. Figure 2d shows the encoder's distortion for $\beta = 2$, for which setting we see from Equation 13 in Appendix B that all quantizers have the same distortion, equal to the non-revealing distortion, given in Equation 11.



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

(d) $\eta_E = (x + 0.5 - 1.5y)^2$ (c) $\eta_E = (x - 2y)^2$

Fig. 2. Encoder distortion for 2-level quantization of $X \sim \mathcal{N}(0, 1), \eta_D = (x - y)^2$.

B. Quantizer resolution is binding

We next focus on the question: can the quantization constraint change the nature of optimal encoding policy in strategic communication? For example, is there a case where for $M = 2$ the encoder is non-revealing but for $M = 3$ the encoder prefers to send a message? The answer is, perhaps surprisingly, affirmative.

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

$$\eta_E(x, y) = \begin{cases} (x^3 - y)^2, & xy \geq 0 \\ \to \infty, & \text{otherwise.} \end{cases} \tag{6}$$

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

$$y_1 = \frac{-1 + q}{2}, \quad y_2 = \frac{q}{2}, \quad y_3 = \frac{1}{2}. \tag{7}$$

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

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

Enforcing the KKT conditions similar to (10), 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, $q = -\sqrt{1/2}$ yields a lower distortion $D_E = 25/224$, hence is optimal.

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.

C. Failure of "Strategic Lloyd-Max"

In this subsection, we investigate whether a simple strategic variation of Lloyd-Max approach would yield a locally optimal quantizer, as in the case of classical quantization. Consider a continuous source $X \sim U[-1, 1]$ quantized with $M = 3$ messages where $\eta_E(x, y) = (x^3 - y)^2$ and $\eta_D(x, y) = (x - y)^2$.

We initialize $\{y_m\}$ arbitrarily and find q_1 and q_2 that minimize $D_E = \mathbb{E}\{(x^3 - y)^2\}$, as

$$q_1 = \left(\frac{y_1 + y_2}{2}\right)^{\frac{1}{3}}, q_2 = \left(\frac{y_2 + y_3}{2}\right)^{\frac{1}{3}}. \tag{8}$$

We then find the decoder actions $\{y_m\}$ as

[2] There is an error in [17] in the values of q.

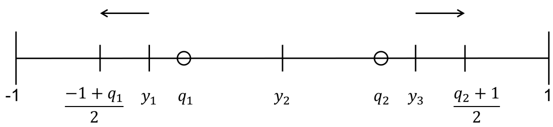


Fig. 3. Movement of the quantization boundaries through Lloyd-Max iterations in the running example.

$$y_1 = \frac{\int_{-1}^{q_1} \frac{1}{2} t \ dt}{\int_{-1}^{q_1} \frac{1}{2} \ dt} = \frac{-1+q_1}{2}, y_2 = \frac{\int_{q_1}^{q_2} \frac{1}{2} t \ dt}{\int_{q_1}^{q_2} \frac{1}{2} \ dt} = \frac{q_1+q_2}{2},$$
$$y_3 = \frac{\int_{q_2}^{1} \frac{1}{2} t \ dt}{\int_{q_2}^{1} \frac{1}{2} \ dt} = \frac{q_2+1}{2}, \tag{9}$$

and iterate between (8) and (9) until convergence. 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 Figure 3, hence the iterations converge to $M = 1$ solution which is 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.

D. Multiple Local Minima

For log-concave scalar sources, it is a well-known result in the literature on classical quantization that the local minima corresponds with the global one; 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 $\eta_E = (x^3 - y)^2, \eta_D = (x - y)^2$. 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 ($\mathbf{q} = \begin{bmatrix} a & q & b \end{bmatrix}$) and it's derivative with respect to the quantization decision level q,

$$J(\mathbf{q}) = \int_a^q (x^3 - y_1)^2 \mathrm{d}\mu + \int_q^b (x^3 - y_2)^2 \mathrm{d}\mu,$$
$$\frac{\partial J}{\partial q} = q^3 - \frac{q}{2} + \frac{a+b}{4}(1 - (b^2 + a^2)).$$

For $a = -0.9$, $b = 1$, $q = -0.6859, 0.7265$ satisfy $\frac{\partial J}{\partial q} = 0$, as also demonstrated in Fig 4. Hence, the strategic problem can indeed have multiple local minima even if used in conjunction with log-concave sources.

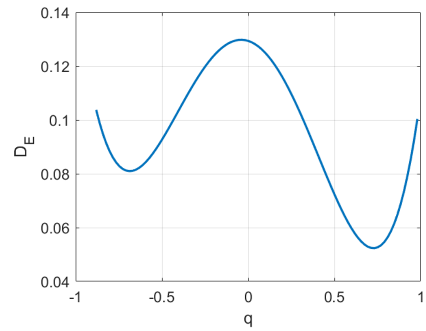
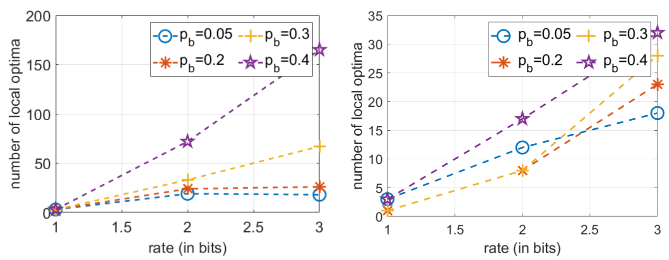


Fig. 4. Local minima of the cost surface of the encoder distortion for 2-level quantization, with decision boundary q, of $X \sim U[-0.9, 1]$ with $\eta_E = (x^3 - y)^2, \eta_D = (x - y)^2$.



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

Fig. 5. Number of local optima quantizers with $\eta_D = (x - y)^2$.

Figures 5a and 5b show the number of locally optimal quantizers for the non-strategic and the strategic case respectively, for a Gaussian source $X \sim \mathcal{N}(0, 1)$ over a binary symmetric channel with crossover probability p_b.

VI. Proposed Solution

We first note a significant research challenge associated with the design problem. The classical vector quantization design relies on the Lloyd-Max optimization, where the encoder and the decoder optimize their mappings iteratively. These iterations converge to a locally optimal solution because the distortion, identical for the decoder and the encoder (team problem), is nonincreasing with each iteration. However, here we consider a game problem (as opposed to a team problem) where the objectives are different, a strategic variation of these algorithms would enforce optimality with respect to a different distortion measure at each iteration, hence do not converge as illustrated in detail in [17]. 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.

Having showed that a strategic variation of the Lloyd-Max algorithm cannot be used since it does not converge to even a local optima, we propose a gradient descent based solution. Due to the Stackelberg equilibrium nature of the problem, the encoder (leader) decides the quantizer decision levels $\mathbf{q}$ first, and then the decoder (follower) decides the quantizer reconstruction levels as a function of $\mathbf{q}$, $\mathbf{y}(\mathbf{q})$. Hence, the optimization can be done on $\mathbf{q}$. The gradient descent approach involves optimizing $\mathbf{q}$ along the direction of the gradients $\frac{\partial D_E}{\partial x_m}$. The algorithm is presented in Appendix A.

We present below the derivation of gradients for an MMSE decoder $\eta_D = (x-y)^2$ with an encoder with distortion measure $\eta_E(x,y)$. The gradients of the encoder's distortion with respect to the quantizer decision levels are given as:

$$\begin{aligned}\frac{\partial D_E}{\partial x_m} &= \eta_E(x_m, y_m)\frac{\mathrm{d}\mu_X(x_m)}{\mathrm{d}x} - \eta_E(x_m, y_{m+1})\frac{\mathrm{d}\mu_X(x_m)}{\mathrm{d}x} \\ &+ \frac{\partial y_m}{\partial x_m}\int \frac{\partial \eta_E(x,y_m)}{\partial y_m}\mathrm{d}\mu_X \\ &+ \frac{\partial y_{m+1}}{\partial x_m}\int_{x_m}^{x_{m+1}} \frac{\partial \eta_E(x,y_{m+1})}{\partial y_{m+1}}\mathrm{d}\mu_X,\end{aligned}$$

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

$$y_m = \frac{\int x\mathrm{d}\mu_X}{\int \mathrm{d}\mu_X}, \quad y_{m+1} = \frac{\int_{x_m}^{x_{m+1}} x\mathrm{d}\mu_X}{\int_{x_m}^{x_{m+1}} \mathrm{d}\mu_X}.$$

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

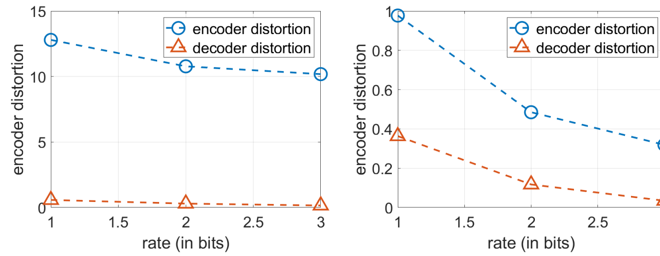
$$\begin{aligned}\frac{\partial y_m}{\partial x_m} &= \frac{x_m \frac{\mathrm{d}\mu_X(x_m)}{\mathrm{d}x}\int \mathrm{d}\mu_X - \frac{\mathrm{d}\mu_X(x_m)}{\mathrm{d}x}\int x\mathrm{d}\mu_X}{(\int \mathrm{d}\mu_X)^2} \\ &= \frac{\mathrm{d}\mu_X(x_m)}{\mathrm{d}x}\frac{x_m - y_m}{\int \mathrm{d}\mu_X},\end{aligned}$$

$$\begin{aligned}\frac{\partial y_{m+1}}{\partial x_m} &= -\frac{x_m \frac{\mathrm{d}\mu_X(x_m)}{\mathrm{d}x}\int_{x_m}^{x_{m+1}} \mathrm{d}\mu_X - \frac{\mathrm{d}\mu_X(x_m)}{\mathrm{d}x}\int_{x_m}^{x_{m+1}} x\mathrm{d}\mu_X}{(\int_{x_m}^{x_{m+1}} \mathrm{d}\mu_X)^2} \\ &= -\frac{\mathrm{d}\mu_X(x_m)}{\mathrm{d}x}\frac{x_m - y_{m+1}}{\int_{x_m}^{x_{m+1}} \mathrm{d}\mu_X}.\end{aligned}$$

VII. Numerical Results

We consider a Gaussian source $\mathcal{N}(0,1)$ with a limited range $[-5,5]$ for numerical implementation. We take the decoding distortion measure as $\eta_D(x,y) = (x-y)^2$, and consider four different cases of encoder distortion, $\eta_E^1(x,y) = (x^3-y)^2$, $\eta_E^2(x,y) = (1.5x-y)^2$, $\eta_E = (x^2-y)^2$, and the non-strategic encoder $\eta_E = (x-y)^2$.

We plot the encoder and the decoder distortions associated with η_D in conjunction with η_E^1 and η_E^2 in Figures 6a and 6b respectively.



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

Fig. 6. $X \sim \mathcal{N}(0,1), \eta_D = (x-y)^2$

VIII. Conclusion

In this paper, we have first formulated the strategic quantization problem and shown its features that differentiate it from classical quantization. We have numerically demonstrated that strategic variation of Lloyd-Max may not converge to a local optimum. Motivated by this observation, we have developed gradient descent based solutions for the strategic quantization problem. Numerical results obtained via the proposed algorithm suggest several open theoretical questions about the behavior of the operational distortion-rate curve of the optimal strategic quantizers.

Appendix A
Gradient Descent Algorithm

Function main():

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

$flag = 0$
end if
end if
end while
$\mathbf{q}^* \leftarrow \mathbf{q}$
$\mathbf{y}^* \leftarrow reconstruction(\mathbf{q}^*, \mu, p_{err}, \mathbb{E}\{X\})$
$D_E \leftarrow distortion(\mathbf{q}^*, \mathbf{y}^*, \mu, \eta_E, E, p_{err})$
$D_D \leftarrow distortion(\mathbf{q}^*, \mathbf{y}^*, \mu, \eta_D, E, p_{err})$

Function check():

Input: $\mathbf{q}, \mu_X, p_{err}, \mathbb{E}\{X\}, \triangle, der, distenc, i$
Output: $\mathbf{q}$
while $\triangle \neq 0$ **do**
$\triangle \leftarrow \triangle/10$
$temp \leftarrow q_i - \triangle\, der$
$\mathbf{qt} \leftarrow \mathbf{q}$
$qt_i \leftarrow temp$
$\mathbf{y} \leftarrow reconstruction(\mathbf{qt}, \mu, p_{err}, \mathbb{E}\{X\})$
$dt \leftarrow encoderdistortion(\mathbf{qt}, \mathbf{y}, \mu, p_{err})$
if $temp > q_{i-1}\,\&\&\, temp < q_i \,\&\&\, dt < distenc$ **then**
$\mathbf{q} \leftarrow \mathbf{qt}$
break
end if
end while

Function reconstruction():

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

Function distortion():

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

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

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

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 determines $\mathbf{y} = [y_1, \ldots, y_m]$ as follows:

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

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

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

$$\begin{aligned}\frac{\partial J}{\partial x_m} &= (x_m + \alpha - \beta y_m)^2 \frac{\mathrm{d}\mu(x_m)}{\mathrm{d}x} \\ &\quad - (x_m + \alpha - \beta y_{m+1})^2 \frac{\mathrm{d}\mu(x_m)}{\mathrm{d}x} \\ &\quad - 2\beta \frac{\mathrm{d}y_m}{\mathrm{d}x_m} \int_{x_{m-1}}^{x_m} (x + \alpha - \beta y_m) \mathrm{d}\mu \\ &\quad - 2\beta \frac{\mathrm{d}y_{m+1}}{\mathrm{d}x_m} \int_{x_m}^{x_{m+1}} (x + \alpha - \beta y_{m+1}) \mathrm{d}\mu.\end{aligned}$$

Enforcing the KKT conditions for optimality

$$\frac{\partial J}{\partial x_m} = 0, \quad m \in [1:M] \tag{10}$$

we obtain, after some straightforward algebra, that the solutions that satisfies (10) are $\beta = 0, 2$, or $x_m = \frac{y_m + y_{m+1}}{2}$ (the other condition is $y_{m+1} = y_m$ which is not possible since the actions are considered unique - if not, the corresponding regions could be combined). This implies that the quantizer is the same as the non-strategic quantizer if $\beta \neq 0, 2$, if the encoder decides to send something.

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

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

The distortion for a non-informative quantizer:

$$\begin{aligned}D_{nr} &= \int_a^b (x + \alpha - \beta y)^2 \mathrm{d}\mu \\ &= \int_a^b x^2 \mathrm{d}\mu + \alpha^2 + 2\alpha(1-\beta)\int_a^b x \mathrm{d}\mu \\ &\quad + \beta(\beta-2) y \int_a^b x \mathrm{d}\mu.\end{aligned} \tag{11}$$

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

$$\begin{aligned} J &= \int_a^b x^2 \mathrm{d}\mu + \alpha^2 + 2\alpha(1-\beta)\int_a^b x \mathrm{d}\mu \\ &\quad + \beta(\beta-2)\sum_{m=1}^{M} y_m \int_{x_{m-1}}^{x_m} x \mathrm{d}\mu \qquad (12) \\ &= D_{nr} + \beta(\beta-2)\Bigg(\sum_{m=1}^{M} y_m \int_{x_{m-1}}^{x_m} x\mathrm{d}\mu - y\int_a^b x\mathrm{d}\mu\Bigg) \qquad (13) \\ &= D_{nr} + \beta(\beta-2)\Bigg(\sum_{m=1}^{M} \frac{\int_{x_{m-1}}^{x_m} x\mathrm{d}\mu}{\int_{x_{m-1}}^{x_m} \mathrm{d}\mu}\int_{x_{m-1}}^{x_m} x\mathrm{d}\mu \\ &\quad - \frac{\sum\limits_{m=1}^{M}\int_{x_{m-1}}^{x_m} x\mathrm{d}\mu}{\sum\limits_{m=1}^{M}\int_{x_{m-1}}^{x_m} \mathrm{d}\mu}\sum_{m=1}^{M}\int_{x_{m-1}}^{x_m} x\mathrm{d}\mu\Bigg) \\ &= D_{nr} + \beta(\beta-2)\Bigg(\sum_{m=1}^{M} \frac{\left(\int_{x_{m-1}}^{x_m} x\mathrm{d}\mu\right)^2}{\int_{x_{m-1}}^{x_m} \mathrm{d}\mu} \\ &\quad - \frac{\left(\sum\limits_{m=1}^{M}\int_{x_{m-1}}^{x_m} x\mathrm{d}\mu\right)^2}{\sum\limits_{m=1}^{M}\int_{x_{m-1}}^{x_m} \mathrm{d}\mu}\Bigg). \end{aligned}$$

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

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

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

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

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

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

that is T is always non-negative.

This implies that the only possible case is case 2 with $0 < \beta < 2$, and the encoder chooses a non-strategic quantizer (as we show earlier in Equation 10 that the only solution when $\beta \neq 0, 2$ is a non-strategic encoder if the encoder sends some message). From Equation 13, we see that for $\beta = 0, 2$ the encoder distortion is the same as non-revealing distortion regardless of the quantizer used.

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

References

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

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

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

[6] 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.

[7] ——, "Information-Theoretic Approach to Strategic Communication as a Hierarchical Game," *Proceedings of the IEEE*, vol. 105, no. 2, pp. 205–218, 2016.

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

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

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

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

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

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

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

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

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

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

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

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

[20] ——, "Optimal Strategic Quantizer Design via Dynamic Programming," in *Proceedings of the IEEE Data Compression Conference*. IEEE, 2022, pp. 173–181.

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