Nonlinear pricing with finite information ☆

Dirk Bergemann [a,*], Edmund Yeh [b], Jinkun Zhang [b]

[a] *Department of Economics, Yale University, New Haven, CT 06511, United States of America*
[b] *Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, United States of America*

ARTICLE INFO

ABSTRACT

We analyze nonlinear pricing with finite information. We consider a multi-product environment where each buyer has preferences over a d-dimensional variety of goods. The seller is limited to offering a finite number n of d-dimensional choices. The limited menu reflects a finite communication capacity between the buyer and seller.

We identify necessary conditions that the optimal finite menu must satisfy, for either the socially efficient or the revenue-maximizing mechanism. These conditions require that information be bundled, or "quantized," optimally.

We introduce vector quantization and establish that the losses due to finite menus converge to zero at a rate of $1/n^{2/d}$. In the canonical model with one-dimensional products and preferences, this establishes that the loss resulting from using the n-item menu converges to zero at a rate proportional to $1/n^2$.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

The theory of mechanism design addresses a wide set of questions, ranging from the design of markets and exchanges to the design of constitutions and political institutions. A central result in the theory of mechanism design is the "revelation principle," which establishes that if an allocation can be implemented incentive-compatible in any mechanism, then it can be truthfully implemented in the direct revelation mechanism, whereby every agent reports his private information, or type, truthfully. Yet, when the amount of private information (the type space) of the agents is large, the direct revelation mechanism requires both the agents to have abundant capacity to communicate with the principal and the principal to have abundant capacity to process information. By contrast, the objective of this paper is to study the performance of optimal mechanisms, when the agents can communicate only limited information, and/or the principal can process only limited information. We pursue our analysis in the context of a representative, but suitably tractable, mechanism design environment: namely, the canonical problem of nonlinear pricing. Here, the principal (seller) is offering a variety of choices to the agent (buyer), who has private information about his own willingness-to-pay (preference or type) for the product.

Our distinct point of view, relative to the seminal analysis by Mussa and Rosen (1978) and Maskin and Riley (1984), is that the information conveyed by the agents, and subsequently the menu of possible choices offered by the seller, is finite, rather than uncountable as in the earlier analysis. The limits to information may arise for various reasons, direct or indirect. On the demand side, it may be too difficult or complex for the buyer, or consumer, to communicate his exact preferences and resulting willingness-to-pay to the seller. On the supply side, it may be too time-consuming for the seller to process the fine details of the buyer's preferences, or to identify the buyer's preferences across many goods with close attributes and only subtle differences.

Our analysis adopts a linear-quadratic specification (analogous to that of Mussa and Rosen (1978) and Maskin and Riley (1984)) in which the buyer's gross utility is the product of his willingness-to-pay (or type) θ and the consumed quantity (or quality) q of the product, whereas the cost of production is quadratic in the quantity (or quality). We reveal a fundamental connection between the problem of optimal nonlinear pricing with limited information and the problem of optimally quantizing a source signal by using a finite number of representation levels in information theory. In our setting, the socially efficient quantity (quality) q for a buyer should be equated to his valuation θ if a continuum of choices were available. In the case where a finite number of choices are accessible, q can take on only a finite number of values. If we interpret θ as the source signal and q as the representation level, then the total social welfare can be written in terms of the mean square error between the source signal and the representation signal. Thus, the social welfare maximization problem can be characterized by the Lloyd-Max optimality conditions, a well-established result in the theory of quantization. Furthermore, we can extend this analysis to the revenue maximization problem after replacing the buyer's true valuation with the corresponding virtual valuation. We estimate the welfare and revenue loss resulting from the use of a finite n-item menu (relative to the continuum menu). In particular, we characterize the rate of convergence for the welfare and revenue loss as a function of n. First, we examine this problem for a given distribution on the buyer's type, and then over all possible type distributions with finite support.

We establish that the maximum welfare loss and the maximum revenue loss shrink towards zero at a rate proportional to $1/n^2$. We thus use quantization theory to approach a problem of mechanism design with limited information transmission.

Our approach extends naturally via vector quantization to the multidimensional nonlinear pricing problem. Here, the seller is offering a variety of heterogeneous products to the buyer, who has private information about his preferences (types) for these products. We maintain a linear-quadratic specification in multiple dimensions as in Armstrong (1996). The advantage of the linear-quadratic model in one or many dimensions is its tractability. In particular, we can frequently compute the lower and upper bounds explicitly. We briefly discuss in the Conclusion how existing results in information theory would allow us to provide results for general non-linear environments as long as certain regularity conditions, such as concavity or convexity of the optimization program are maintained. In the multi-dimensional environment we require an additional separability condition regarding the type distribution. This condition was introduced earlier by Armstrong (1996) to guarantee the incentive-compatibility of the menu in the continuous multi-dimensional setting.

We interpret the private information (the preference or type vector) as the signal vector and the choice (quantity or quality vector) as the representation vector. The social welfare maximization problem and the revenue maximization problem can still be characterized by the Lloyd-Max optimality conditions for vector quantization. We estimate the welfare and revenue loss resulting from the use of a d-dimensional finite menu with n choices. We establish an upper bound on the welfare loss by appealing to decomposition result, see Lookabaugh and Gray (1989a). The upper bound uses subtle vector quantization methods to design the multi-product finite menus over the entire type space. The gain from vector quantization consists of three components: space-filling advantage, shape advantage, and dependence advantage. Most notably, even in the extreme case when the types are distributed independently and uniformly across all dimensions, the vector quantization method can still reduce the welfare loss and the revenue loss due to the space-filling advantage. This is the main reason why we bundle the buyer's preferences over multiple goods as a vector instead of viewing them separately as independent dimensions. We then establish the vector-quantization-based upper bound and the lower bounds on the welfare loss and the revenue loss.

The role of limited information in mechanism design has recently attracted increased attention. McAfee (2002) phrases the priority rationing problem as a two-sided matching problem (between the buyer and services) and shows that a binary priority contract ("coarse matching") can already achieve at least half of the social welfare that could be generated by a continuum of priorities. Hoppe et al. (2010) extend the matching analysis and explicitly consider monetary transfers between the agents. In particular, they present lower bounds on the revenue which can be achieved with specific, but not necessarily optimal, binary contracts. By contrast, Madarasz and Prat (2017) suggest a specific allocation—the "profit-participation" mechanism—to establish approximation results, rather than finite optimality results, in the nonlinear pricing environment. While the above contributions are concerned with single agent environments, there have been a number of contributions to multi-agent mechanisms—specifically single-item auctions among many bidders. Blumrosen et al. (2007) consider the effect of restricted communication in auctions with either two agents or binary messages for every agent. Kos (2012) generalizes the analysis by allowing for a finite number of messages and agents. In turn, their equilibrium characterization in terms of partitions shares features with the optimal information structures in auctions as derived by Bergemann and Pesendorfer (2007).

Closer to our approach is Wilson (1989), who considers the impact of a finite number of priority classes on the efficient rationing of services. His analysis is less concerned with the optimal priority ranking for a given finite class and more with the approximation properties of the finite priority classes. Wilson (1989) shows that the social welfare loss due to the use

of a finite number of priority classes converges to zero at a rate no faster than $1/n^2$, where n is the number of classes. The analysis in Wilson (1989), however, is limited to one-dimensional social welfare maximization and is not easily generalizable to the multidimensional social welfare maximization problem or the revenue maximization problem. The latter problems have remained open in general. In earlier work, Bergemann et al. (2012a, 2012b), some of use introduced the quantization technique to analyze an environment with limited information. Bergemann et al. (2012a) focused entirely on the one-dimensional environment to obtain upper and lower bounds on worst-case welfare and revenue. Their main results had a gap between lower and upper bounds in either case. In the current work, Proposition 2 and 3 eliminate the gap by making use of the high rate quantization results. Bergemann et al. (2012b) considers a multi-dimensional environment with some additional restrictions. In particular, the analysis is restricted to welfare maximization in the absence of incentive constraints. Using some foundational results in information theory, this paper represents the first systematic and comprehensive solution to these problems in many dimensions,

Even in the absence of communication constraints, the multidimensional mechanism design does not represent a trivial generalization of its one-dimensional counterpart. In many environments of interest, the preferences of an individual agent cannot be summarized by a mere scalar, but are more suitably represented as a vector. A real-life example would be a buyer who has to make choices in a supermarket where a large variety of commodities are available. Hence, designing a smart pricing strategy (e.g., product bundling by offering a combination of several distinct products for joint sale, rather than selling each item separately) is of first-order concern in practice. In this respect, Wilson (1993) and Armstrong (1996) provide two notable early contributions, with explicit solutions to specific multidimensional screening problems. Rochet and Chone (1998) develop a systematic approach, dubbed the dual approach, for a general class of environments, and pointed to the prevalence of bunching (agents with different type profiles making the same choices). We refer readers to Rochet and Stole (2003) for a detailed survey of multidimensional screening problems.

The rest of the paper is organized as follows. First, we introduce the basic nonlinear pricing model in the following section. Then, in Section 3, we establish the link to the quantization problem in information theory in the one-dimensional product space. Moreover, we introduce the Lloyd-Max conditions that the optimal finite menu must satisfy. In Section 4, we generalize our approach to the multi-product environment by using vector quantization. In Section 5, we conclude with a brief summary and note some open issues for our future research. The Appendix collects all proofs not presented in the main body of the paper.

2. Model

We consider a seller (she) who is providing d heterogeneous goods to a buyer (he) with a continuum of possible preferences. Each buyer's preferences over these goods is characterized by a d-dimensional vector $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}_+^d$, called the buyer's type vector, where for $1 \leq l \leq d$, θ_l represents his preference (type) for good l. Let $\Theta = [0, 1]^d$, where the unit hypercube is without loss of generality as long as we consider a compact d-dimensional type space. The joint probability distribution of θ, denoted by $F(\theta)$, is assumed to be commonly known. We denote by F_l the marginal distribution function of type θ_l. We assume that the joint density function f is continuous almost everywhere (a.e.) in the support (i.e., the type space):

$$\Theta = \left\{ \theta \in \mathbb{R}_+^d : f(\theta) > 0 \right\}.$$

We further assume that the buyer's preferences over d products, $\theta_1, \ldots, \theta_d$, are identically, but not necessarily independently, distributed.

A buyer with type θ who receives a quantity (or quality) vector $q = (q_1, \ldots, q_d) \in \mathbb{R}_+^d$ and makes a monetary payment t receives the following net utility:

$$U(\theta, q, t) = \theta^T \Phi q - t, \tag{1}$$

where $\Phi = (\phi_{ij})_{d \times d}$ is a $d \times d$ symmetric matrix which captures the interactions among different goods. We assume that $\phi_{ii} > 0$ for all i.

The firm has a quadratic cost function for providing the vector q:

$$c(q) = \frac{1}{2} q^T \Sigma q. \tag{2}$$

The matrix $\Sigma = (\sigma_{ij})_{d \times d}$ is a $d \times d$ symmetric positive-definite matrix which characterizes the interactions in the production cost of multiple products. All of its diagonal elements are assumed to be positive: $\sigma_{ii} > 0$ for all i. The seller's profit is given by:

$$R(q, t) = t - c(q) = t - \frac{1}{2} q^T \Sigma q. \tag{3}$$

This setting, usually called the linear-quadratic model, has been used extensively in the literature (see the seminal analysis of Mussa and Rosen (1978) for the one-dimensional case and Armstrong (1996) for the multidimensional model). For the

multidimensional model, it is helpful to notice that we can always rewrite the above model by a standard change of basis argument in the "standard form," with

$$\Phi = \Sigma = \mathbb{I}_d, \tag{4}$$

where $\mathbb{I}_d$ is the $d \times d$ identity matrix (see Lemma 2 in the Appendix).

3. One-dimensional product space

We begin the analysis with the one-dimensional version of the model; thus, $d = 1$. This will allow us to introduce some key notions in the classic setting of nonlinear pricing, first analyzed by Mussa and Rosen (1978). We begin with the social welfare maximization problem and then proceed to the revenue maximization problem.

3.1. Welfare maximization

In the presence of private information, the socially efficient allocation can be implemented by a direct mechanism $(q(\theta), t(\theta))$. In the efficient direct mechanism the buyer is offered a menu $\{q(\theta)\}_{\theta \in [0,1]}$ in which type θ is allocated the quantity (quality) $q(\theta)$ against a (monetary) transfer $t(\theta)$. An efficient mechanism maximizes the expected social welfare as the sum of the buyer's net utility and the seller's profit:

$$\mathbb{E}_\theta \left[\theta q(\theta) - \frac{1}{2} q(\theta)^2 \right].$$

For a buyer with type θ, it is socially optimal to provide a production level equal to his type: $q^*(\theta) = \theta$. The maximized social welfare for a given distribution F is:

$$W_F \triangleq \mathbb{E}_\theta \left[\theta q^*(\theta) - \frac{1}{2} q^*(\theta)^2 \right] = \frac{1}{2} \mathbb{E} \left[\theta^2 \right]. \tag{5}$$

The socially optimal menu thus requires an uncountable infinity, or continuum, of reports $\theta \in [0, 1]$ and a corresponding continuum of allocations, hence our definition of W_F.

The mechanism, as a special case of the Vickrey-Clark-Groves mechanism, must satisfy two sets of constraints: the individual rationality (or participation) constraint, $\theta q(\theta) - t(\theta) \geq 0$, for all $\theta \in [0, 1]$; and the incentive constraint, $\theta q(\theta) - t(\theta) \geq \theta q(\theta') - t(\theta')$, for all $\theta, \theta' \in \Theta$.

By contrast, we are interested in finding the optimal menu when the buyer can communicate his type only in a finite language; or equivalently, when the seller can process only finitely many different messages; or equivalently, when the seller can produce only finitely many alternative versions of her product.

An n-item menu is composed of $n < \infty$ different allocations $\{q_k\}$, $k = 1, \ldots, n$, where q_k is the quantity (quality) of the k-th item of the menu. Let $\left\{ P_k = [\theta_{k-1}, \theta_k) \right\}_{k=1}^n$ represent a corresponding partition of the set of buyer types, $[0, 1]$, where $0 = \theta_0 \leq \theta_1 \leq \ldots \leq \theta_n = 1$.[1] A buyer with type $\theta \in P_k$ is assigned quantity $q(\theta) = q_k$. A finite menu (and its associated assignments) given by $\{P_k, q_k\}_{k=1}^n$ is called an n-item menu henceforth. We shall refer to menu as the partition of the type space and the allocation. At this stage, we omit the transfers $\{t_k\}_{k=1}^n$ which follow directly from incentive compatibility. Let M be the set of all n-item menus:

$$M \triangleq \left\{ \{P_k, q_k\}_{k=1}^n : P_k = [\theta_{k-1}, \theta_k), 0 = \theta_0 \leq \theta_1 \leq \ldots \leq \theta_n = 1 \right\}. \tag{6}$$

We choose a finite menu $\{P_k, q_k\}_{k=1}^n$ from M to maximize the expected social welfare for a given distribution F:

$$W_F(n) \triangleq \max_{\{P_k, q_k\}_{k=1}^n \in M} \left\{ \mathbb{E}_\theta \left[\theta q - \frac{1}{2} q^2 \right] \right\}. \tag{7}$$

We ask how well the optimal n-item menu $\left\{ P_k^*, q_k^* \right\}_{k=1}^n$ can approximate the performance of the optimal continuous menu $\{q^*(\theta)\}_{\theta \in [0,1]}$ as measured by the welfare loss: $W_F - W_F(n)$.

It is easy to see that a tight lower bound on the welfare loss over all distributions is zero, i.e., $\inf_F \{W_F - W_F(n)\} = 0$. This can be achieved by the discrete uniform distribution $\Pr\left[\theta = \frac{k}{n} \right] = \frac{1}{n}$ for $k = 1, \ldots, n$. In what follows we will therefore focus on the upper bound on the welfare loss over all distributions with finite support. Thus, let $\mathcal{F}$ be the set of all distribution functions on $[0, 1]$. Our main task is to estimate the worst-case behavior of the welfare loss over all distributions $F \in \mathcal{F}$.

---

[1] In general, a partition need not consist of intervals (only) due to the nature of the optimization problem in (7); however, it can be shown that the optimal partition consists of intervals.

Definition 1 *(Welfare loss).* For a given distribution F, the welfare loss of the optimal n-item menu relative to the continuous menu given is:

$$L_F(n) \triangleq W_F - W_F(n);$$

and across all distributions $F \in \mathcal{F}$ the maximum welfare loss is:

$$L(n) \triangleq \sup_{F \in \mathcal{F}} L_F(n).$$

If we view θ as a continuous signal that must be represented by a representation point q_k in the interval P_k, then this is an instance of the *quantization problem* in information theory. The intervals $\{P_k\}_{k=1}^n$ and the corresponding representation points $\{q_k\}_{k=1}^n$ are jointly chosen to minimize the mean square error when we view the quantity q as the predictor of type θ. In information theory, this error is often referred to as the *distortion* D due to the quantization:

$$D \triangleq \min_{\{P_k, q_k\}_{k=1}^n \in M} \left\{ \mathbb{E}_\theta \left[(\theta - q)^2 \right] \right\}. \tag{8}$$

Given the distribution F, the distortion D is equivalent to the welfare $W_F(n)$ as defined above in (7). With this perspective, we can interpret the finite menu $\{P_k^*, q_k^*\}_{k=1}^n$ as the solution to a scalar quantization problem. Henceforth, we use the terms quantizer and finite menu interchangeably.

The optimal scalar quantizer $\{P_k^*, q_k^*\}_{k=1}^n$ must satisfy the following optimality conditions:

$$\theta_k^* = \frac{1}{2}\left(q_k^* + q_{k+1}^*\right), \quad q_k^* = \mathbb{E}\left[\theta | \theta \in P_k^*\right], \quad k = 1, \dots, n; \tag{9}$$

with $\theta_0^* = 0$ and $\theta_n^* = 1$. Thus, θ_k^*, which separates two neighboring intervals P_k^* and P_{k+1}^*, must be the arithmetic average of q_k^* and q_{k+1}^*. At the same time, q_k^*, the representation level for the interval $P_k^* = \left[\theta_{k-1}^*, \theta_k^*\right)$, must be the conditional mean for θ, given that θ falls in this interval. These two necessary conditions are often referred to as *Lloyd-Max optimality conditions* as they were independently established by Lloyd (1957) (and published in 1982) and later by Max (1960). The Lloyd-Max optimality conditions remain valid in many dimensions. In Section 4, we therefore interpret $\{P_k^*\}_{k=1}^n$ as a Voronoi partition (a set of the nearest-neighbor regions) with respect to $\{q_k^*\}_{k=1}^n$, and q_k^* is chosen as the conditional mean of θ, given that θ lies in the region P_k^* (see Lloyd (1982)).[2]

A commonly used scalar quantizer is the *uniform quantizer*, by which is meant that (i) the boundary points are equally spaced, $\theta_k - \theta_{k-1} = \Delta$, and (ii) the representation points are the midpoints of the quantization interval. For specific distributions it is possible to obtain the closed form of the optimal finite menus from the Lloyd-Max conditions. Here we consider the uniform distribution $\theta \sim \mathcal{U}[0, 1]$, for which the welfare loss can be exactly established, and a fortiori the resulting convergence rate as n increases.

Proposition 1 *(Welfare loss for uniform distribution).* *The optimal n-item menu* $\{P_k^*, q_k^*\}_{k=1}^n$, $P_k^* = \left[\theta_{k-1}^*, \theta_k^*\right)$ *is given by:*

$$\theta_k^* = \frac{k}{n}, \quad q_k^* = \frac{k - \frac{1}{2}}{n}, \quad k = 0, \dots, n; \tag{10}$$

and the associated welfare loss is $L_{\mathcal{U}}(n) = 1/24n^2$.

The optimal scalar quantization for the uniform distribution illustrates how the Lloyd-Max conditions are used to obtain the optimal finite menu. The resulting boundary points $\{\theta_k^*\}_{k=0}^n$, as well as the representation points $\{q_k^*\}_{k=0}^n$, share the uniformity with the underlying distribution of the values. In particular, the optimal quantizer is a uniform quantizer. The partition point $\{\theta_k^*\}$ and the probability weight of the representation points are illustrated in Fig. 1 for the uniform distribution with $n = 5$.

The uniform distribution intuitively makes quantization difficult as (i) the uncertainty is uniformly spread and (ii) the optimal allocation is a uniform response to the true underlying state. We now show that this intuition can be made precise in the sense that the welfare loss associated with the uniform distribution is indeed an upper bound on welfare loss across all possible distributions as n becomes large. The asymptotic results in quantization theory are often referred to as high-rate quantization, as the number of representation points is allowed to become large.

---

[2] The optimality conditions (9) are not sufficient conditions, and Lloyd (1957) provides counterexamples. Trushkin (1982) provides general conditions on the probability density function, in particular that the density is logconcave, for the necessary conditions to lead to a unique solution.
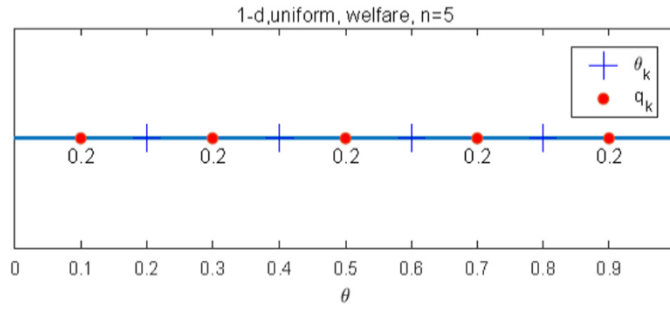
Fig. 1. Quantizer and representation points for welfare maximization with uniform distribution and $n = 5$.

3.2. High-rate quantization

For a general class of distributions, analytical solutions to the Lloyd-Max conditions are not available. We therefore design a sequence of finite menus to obtain an upper bound on the welfare loss. In these menus the quantities $\{q_k\}$ are consistent with the Lloyd-Max conditions (9). This construction estimates how fast the maximum welfare loss converges to zero as the number of classes n tends to infinity.

We first introduce the so-called *high-rate non-uniform* quantizing scheme, which provides an asymptotic optimal quantization result (see Gersho and Gray (2007)). Specifically, consider a distribution $f(\theta)$ on $[0, 1]$ and define the overall quantization distortion, namely the mean square error due to the quantization, to be

$$D = \sum_{k=1}^{n} \int_{\theta_{k-1}}^{\theta_k} (\theta - q_k)^2 f(\theta) d\theta,$$

where $\theta_0 = 0$ and $\theta_n = 1$. When (a) n is sufficiently large and (b) the input distribution f is sufficiently smooth, the conditional type distribution in each quantization interval is approximately uniform. In turn, the distortion can be approximated by:

$$D \approx \sum_{k=1}^{n} f(\theta_k) \frac{\left(\theta_k - \theta_{k-1}\right)^2}{12}.$$

Meanwhile, when n is sufficiently large, we denote by $N(\theta)\Delta_\theta$ the number of quantization levels that lie between θ and $\theta + \Delta_\theta$ for a small increment Δ_θ, where Δ_θ is larger than any quantization interval length $(\theta_k - \theta_{k-1})$, and the term $\lambda(\theta) = N(\theta)/n$ is referred to as the *point density* of the quantizer. When Δ_θ is sufficiently small, the quantizer density integral satisfies

$$\int_0^1 \lambda(\theta) d\theta = \frac{\int_0^1 N(\theta) d\theta}{n} = 1.$$

It follows that there are approximately $n\lambda(\theta)\Delta_\theta$ uniformly spaced intervals between θ_k and $\theta_k + \Delta_\theta$, and thus the length of each quantization interval is

$$\theta_k - \theta_{k-1} \approx \frac{\text{length of increment } \Delta_\theta}{\text{number of intervals between } \theta_k \text{ and } \theta_k + \Delta_\theta}$$

$$\approx \frac{\Delta_\theta}{n\lambda(\theta_k)\Delta_\theta} = \frac{1}{n\lambda(\theta_k)}.$$

Therefore, an approximation of the total distortion can be given as:

$$D \approx \frac{1}{12} \sum_{k=1}^{n} f(\theta_k) \left(\frac{1}{n\lambda(\theta_k)}\right)^2,$$

which can be further approximated by an integral, as n is large:

$$D \approx \frac{1}{12n^2} \int_0^1 \frac{f(\theta)}{\lambda^2(\theta)} d\theta.$$

By applying Hölder's inequality, this distortion integral can be further lower-bounded (see Theorem 2 in Zador (1982)):

$$\liminf_{n \to \infty} n^2 D \geq \frac{1}{12} \left(\int_0^1 f(\theta)^{1/3} d\theta \right)^3, \tag{11}$$

and the equality holds with the optimal quantizer density distribution, given by:

$$\lambda^*(\theta) = \frac{f(\theta)^{1/3}}{\int_0^1 f(t)^{1/3} dt}.$$

Under the assumption that n is sufficiently large, the *high-rate non-uniform distortion* (11) provides a lower bound of the quantization distortion, thus:

$$D \approx \frac{1}{12n^2} \left(\int_0^1 f(\theta)^{1/3} d\theta \right)^3.$$

Note that the *high-rate non-uniform* quantizing above assumes a sufficiently large n, yielding an arbitrarily high quantization rate, and quantizing points $\{\theta_k\}$ that are not necessarily uniformly distributed in the whole region $[0,1]$. We then derive an asymptotic bound for welfare loss using the high-rate non-uniform quantizing result.

Proposition 2 (*Bound on welfare loss*). *If n is sufficiently large, then*

$$L(n) \leq \frac{1}{24n^2}. \tag{12}$$

Combining the above result with Proposition 1, we obtain a sharp characterization of the asymptotic behavior of $L(n)$. A version of the one-dimensional social welfare maximization problem in (7) was considered earlier in Wilson (1989) and in Bergemann et al. (2012a). Wilson (1989) obtained an approximate solution to (7) by a version of a uniform quantizer of the distribution function of θ, and then by expanding the social welfare on each quantization interval by the Taylor series around zero up to the order of $1/n^3$, where n is the total number of intervals. His version of the uniform quantizer requires that each element of the partition contains the same probability (rather than maintaining the distance between the partition points). Proposition 3 in Wilson (1989) establishes that the efficiency loss resulting from an n-item menu is of an order no more than $1/n^2$; i.e., $W_F(n) \geq W_F - O(1/n^2)$. In Bergemann et al. (2012a), Proposition 3 obtains a larger, and a fortiori weaker, upper bound on the welfare loss by also relying on a uniform quantizer.

3.3. Revenue maximization

We now analyze the problem of revenue maximization. In contrast to the social welfare problem, here, the seller wishes to design a menu $\{q(\theta), t(\theta)\}_{\theta \in [0,1]}$ to maximize her expected net revenue, i.e., the difference between the gross revenue that she receives from the buyer minus the cost of providing the demanded quantity (quality) for given distribution F:

$$R_F = \max_{\{q(\theta), t(\theta)\}} \left\{ \mathbb{E}_\theta \left[t(\theta) - \frac{1}{2} q(\theta)^2 \right] \right\}.$$

As before, the contract offered has to satisfy two sets of constraints, namely the incentive constraints and the individual rationality (or participation) constraints. The revenue maximization problem is transformed into a welfare maximization problem (without incentive constraints) after replacing the valuation θ with the corresponding virtual valuation:

$$\psi(\theta) \triangleq \theta - \frac{1 - F(\theta)}{f(\theta)}. \tag{13}$$

The virtual valuation is below the true valuation, and the inverse of the hazard rate $(1 - F(\theta))/f(\theta)$ accounts for the information rent. This problem has been analyzed first by Mussa and Rosen (1978) and Maskin and Riley (1984). The expected revenue of the seller (without communication constraints) is then:

$$R_F^* \triangleq \mathbb{E}_\theta \left[q^*(\theta) \psi(\theta) - \frac{1}{2} (q^*(\theta))^2 \right],$$

and the resulting optimal contract exhibits:

$$q^*(\theta) \triangleq \max \{\psi(\theta), 0\}. \tag{14}$$

We identify the lowest value θ at which the virtual valuation attains a nonnegative value as $\underline{\theta} \triangleq \min \{\theta | \psi(\theta) \geq 0\}$ and hence the corresponding revenue is

$$R_F = \mathbb{E}_\theta \left[q^*(\theta) \psi(\theta) - \frac{1}{2} q^*(\theta)^2 \right] = \frac{1}{2} \int_{\underline{\theta}}^{1} \psi^2(\theta) \, dF(\theta). \tag{15}$$

With the monotonicity of $\psi(\theta)$, we can relabel the type θ directly in terms of the corresponding virtual valuation $\widehat{\theta}$: $\widehat{\theta} \triangleq \psi(\theta)$, and define the associated intervals $\{P_k\}_{k=1}^{n}$ directly in terms of the new variable $\widehat{\theta}$:

$$\theta \in P_k = [\theta_{k-1}, \theta_k) \Leftrightarrow \widehat{\theta} \in \widehat{P}_k = [\widehat{\theta}_{k-1}, \widehat{\theta}_k), 1 \leq k \leq n,$$

where $\widehat{\theta}_k = \psi(\theta_k), 1 \leq k \leq n$. After this change of variable, we define a distribution function $G(\widehat{\theta})$ in terms of the original distribution function $F(\theta) : G(\widehat{\theta}) = G(\psi(\theta)) \triangleq F(\theta)$.[3] Then, the revenue of an n-item menu can be written in terms of the virtual type $\widehat{\theta}$:

$$R_F(n) = \mathbb{E}_\theta \left[q\psi(\theta) - \frac{1}{2} q^2 \right] = \mathbb{E}_{\widehat{\theta}} \left[q\widehat{\theta} - \frac{1}{2} q^2 \right]. \tag{16}$$

Definition 2 *(Revenue loss).* For a given distribution F, the revenue loss of the optimal n-item menu relative to the continuous menu is:

$$\widehat{L}_F(n) \triangleq R_F - R_F(n);$$

and across all distributions $F \in \mathcal{F}$ the maximum revenue is:

$$\widehat{L}(n) \triangleq \sup_{F \in \mathcal{F}} \widehat{L}_F(n).$$

We can now give the worst-case distribution for one-dimensional revenue maximization with the asymptotic upper bound for revenue loss.

Proposition 3 *(One-dimensional revenue bound).* If n is sufficiently large, then

$$\widehat{L}(n) \leq \frac{1}{24n^2}. \tag{17}$$

Similar to the welfare maximization case, the above upper bound is indeed attained by a specific distribution, namely the uniform distribution. But in contrast to the welfare maximization problem, it is the *uniform distribution on the upper half of the unit interval*; thus, $[1/2, 1]$. In this interval, the virtual utility is guaranteed to be positive everywhere.

Thus, the convergence rate of the revenue loss induced by the optimal n-item menu for the uniform distribution is of order $1/n^2$. It follows that the convergence rate for revenue maximization is identical to the one we established for the social welfare maximization environment in Proposition 2.

Proposition 4 *(One-dimensional revenue loss).* If n is sufficiently large, then

$$\widehat{L}(n) = \frac{1}{24n^2}.$$

Additionally, we find that for any finite menu the seller tends to serve fewer consumers when compared to the case of a continuous menu. Thus, for example, the Lloyd-Max optimality conditions in the case of the uniform distribution yield that

$$\theta_0^*(n) = \frac{n+1}{2n+1} > \frac{1}{2} = \theta_0^*.$$

The difference $\theta_0^*(n) - \theta_0^*$ shrinks to 0 as n goes to infinity. This is a consequence of the fact that the seller's ability to extract revenue is more limited in the case of finite menus. To compensate, the seller would like to reduce the service coverage in order to pursue higher profits. This is illustrated in Fig. 2.

---

[3] We note that the critical bounds are established by distributions that generate monotone virtual valuations. Thus the restriction to monotone, or "regular environments" in the language of Myerson (1979), is without loss of generality.
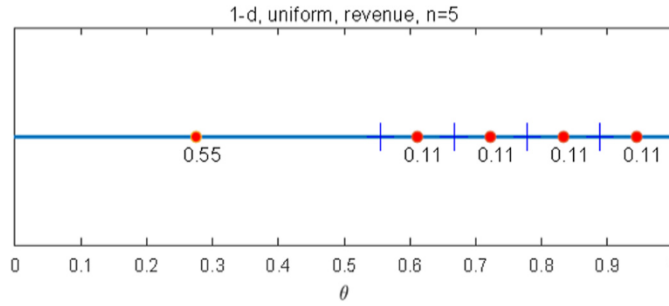
Fig. 2. Quantizer and representation points for revenue maximization with uniform distribution on $[0, 1]$ and $n = 5$.

4. Multidimensional product space

In this section, we consider the multidimensional version of the nonlinear pricing problem, which leads to the design of finite menus over multiple products. We demonstrate that our quantization view generalizes to the multidimensional environment. The optimal design of finite menus requires the technique of vector quantization. We present bounds on the welfare and revenue loss arising from the communication constraints. In particular, we show that in many cases it is beneficial to bundle the buyer's preferences over multiple goods as a vector, instead of treating them separately as independent quantities as repeated scalar quantization would suggest, thereby enabling the true joint design of finite menus over multiple goods.

4.1. Multidimensional welfare maximization

With a continuous menu, the social welfare of (1)-(3) is maximized by solving the d-dimensional linear-quadratic program:

$$W_F = \max_{q(\theta)} \mathbb{E}_\theta \left[\theta^T q - \frac{1}{2} q^T q \right];$$

and it is socially optimal to provide a quantity (quality) vector equal to the type vector $q^*(\theta) = \theta$. The maximal social welfare for a given distribution F therefore equals:

$$W_F = \mathbb{E}_\theta \left[\theta^T q^*(\theta) - \frac{1}{2} q^*(\theta)^T q^*(\theta) \right] = \frac{1}{2} \mathbb{E}_\theta \left[\theta^T \theta \right]. \tag{18}$$

Any finite menu is now defined as a partition of the $d-$dimensional buyer's type space Θ, i.e., $P_i \cap P_j = \emptyset$ if $i \neq j$, and $\cup_{k=1}^n P_k = \Theta$. All consumers with type vector $\theta \in P_k$ will be allocated the kth quantity (quality) vector q_k. Now, $\{P_k, q_k\}_{k=1}^n$ describes a finite multi-product menu, called the n-item menu. If we view θ as the signal vector and q_k as the representation vector of θ in the region P_k, then this becomes the d-dimensional n-region vector quantization problem, where the partition $\{P_k\}_{k=1}^n$ and the set of representation points $\{q_k\}_{k=1}^n$ are jointly chosen to minimize the mean square error or distortion:

$$\min_{\{P_k, q_k\}_{k=1}^n \in M} \left\{ \mathbb{E}_\theta \left[\|\theta - q\|^2 \right] \right\}.$$

In this manner, any multi-product finite menu $\{P_k, q_k\}_{k=1}^n$ can be viewed as a vector quantizer. We can therefore use the two terms "vector quantizer" and "finite multi-product menu" interchangeably.

As in the scalar case, we need to guarantee that the allocations are incentive compatible. With multiple dimensions, this requirement is much more complex than the monotonicity condition in the scalar case. There, the monotonicity condition is a necessary and sufficient condition for incentive compatibility. In the following we will follow the approach introduced by Armstrong (1996) and require a separability condition of the type distribution.

Armstrong (1996) introduced a multiplicative separability condition in the analysis of many-item nonlinear pricing. This condition allows him to establish the incentive compatibility of the menu which is an intractable problem in more general environments. This condition states that the "average taste" across all dimensions—in our context L_2 norm of the taste vector $\|\theta\|_2$ of a consumer—does not provide any additional information about which of her taste parameters θ_i, are likely to be greater than others, for any one of the dimensions, $i = 1, ..., d$. Thus, the value of the average states does not convey any information on which the ray from the origin the vector θ lies.

Definition 3 (*Separability condition*). A joint distribution F is separable if the density f satisfies:

$$f(\theta) = f_1(\|\theta\|_2) \times f_2(\theta)$$

and f_2 is homogeneous of degree zero in θ.

We denote the class of separable density functions by $\mathcal{F}_S$. The above separability condition implies that we can write the density in terms of polar coordinates $f = f(r, \phi)$, where $r = \|\theta\|_2$ and ϕ is the vector of angles with respect to the given orthonormal basis. We describe this more formally in (25) further down.

A second implication of the separability condition is that for a distribution F to belong to $\mathcal{F}_S$ it has to have support on the positive orthant of the d-dimension unit ball rather than the d-dimensional unit cube.

Since the separability condition requires the distribution to be supported on a d-dimensional ball, our main task is to quantify the worst-case behavior of the welfare loss, given earlier in Definition 1 and now augmented by the dimension d, namely $L_F(n, d)$, over all separable distributions $F \in \mathcal{F}_S$. Given the separability condition, a natural generalization of the uniform distribution in the multidimensional type space is the uniform distribution on the positive orthant of the d-dimensional unit ball, denoted as the *uniball* distribution (or $\mathcal{U}_b$), namely $\mathcal{U}_b(\theta) = 1/V_+(d)$, where

$$V_+(d) = \frac{1}{2^d} \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2} + 1\right)}$$

is the volume of the positive orthant of the d-dimension unit ball and $\Gamma(\cdot)$ is the gamma function.

A possible quantization scheme in the multidimensional environment would be to separately and repeatedly perform the scalar quantization we investigated in the previous section in every dimension. Such repeated scalar quantization would create a set of regions that are orthotopes, i.e., the Cartesian product of intervals in d dimensions.

However, as one might have anticipated, in general repeated scalar quantization does not result in the optimal n-item menu. Indeed, in higher dimensions, we can use more subtle vector quantization methods to design better finite menus. To achieve this, we bundle the buyer's preferences over multiple goods as a vector, instead of viewing them separately as independent choices.

For an arbitrary multidimensional distribution F, the optimal vector quantization typically cannot be established explicitly. In some special cases it can be iteratively approximated with the multidimensional version of the Lloyd-Max optimality condition.

Nevertheless, Lookabaugh and Gray (1989a) establish that when the number of items per dimension is sufficiently large, then the gain from vector quantization gain can be decomposed into three distinct terms that allow for an explicit calculation of the loss. We use this decomposition approach in the next result.

Proposition 5 (*Welfare loss for* uniball *distribution*). *For the* uniball *distribution, assuming d and* $n^{\frac{1}{d}}$ *are sufficiently large, then*

$$L_{\mathcal{U}_b}(n, d) = \frac{1}{8n^{\frac{2}{d}}}.$$

We illustrate the optimal vector quantization for the case of the uniform distribution on the unit-ball in Fig. 3. By contrast, in Fig. 4 we display the solution for the repeated scalar quantization. The vector quantization bundles across the dimensions, whereas the scalar quantization creates orthotopes of varying size along each dimension.

As we consider the optimal quantization in the positive orthant, one might guess that the "worst scenario" (i.e., achieving the largest welfare loss) would be the case when the type probability is equally distributed in the d-dimensional unit ball— thus, the *uniball* distribution. In fact, we now show that the *uniball* case gives an asymptotic upper bound of the welfare loss in the presence of the separability condition.

Proposition 6 (*Multidimensional welfare loss bound*). *Given both d and* $n^{\frac{1}{d}}$ *are sufficiently large, for any distribution* $F \in \mathcal{F}_S$

$$L(n, d) \leq \frac{1}{8n^{\frac{2}{d}}}. \tag{19}$$

Combining Proposition 5 and Proposition 6, establishes the asymptotic behavior of the welfare loss $L(n, d)$. We note that our analysis here assumed the separability condition (see Definition 3). We have separately developed a similar analysis for multidimensional welfare loss in the absence of the separability condition. The upper bound for the welfare loss without the separability condition is a fortiori larger than the bound with the separability condition. The details are presented in Proposition 9 in the Appendix. Proposition 9 presents the corresponding result to Proposition 5 in the absence of the separability condition. We show that the upper bound in then attained by the multidimensional uniform distribution on the unit cube. Now, the quantization region is not restricted anymore to the positive orthant of the unit ball, and given instead by the unit cube. Without the restriction to the separability condition, the bound shows a linear degradation in the dimensionality d of the allocation problem, namely $d/n^{2/d}$.

4.2. Multidimensional revenue maximization

We complete our analysis by considering the revenue maximization problem in many dimensions. The problem for the seller in the direct mechanism without communication constraints is given by maximizing
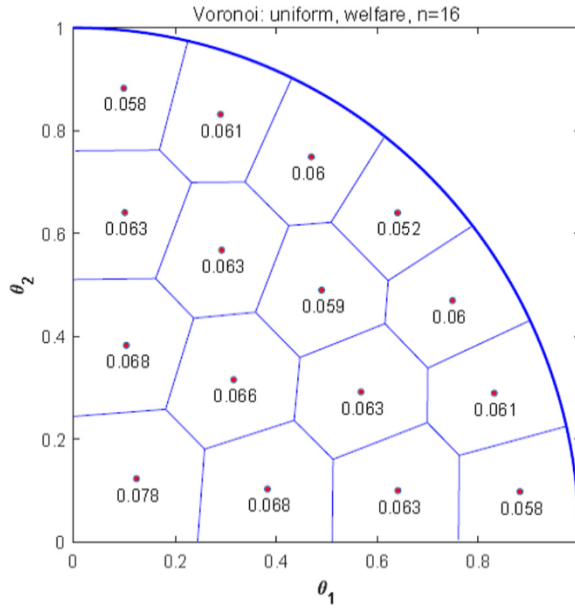
Fig. 3. Quantizer and representation points for welfare maximization with *uniball* distribution for $d = 2$, $n = 16$.
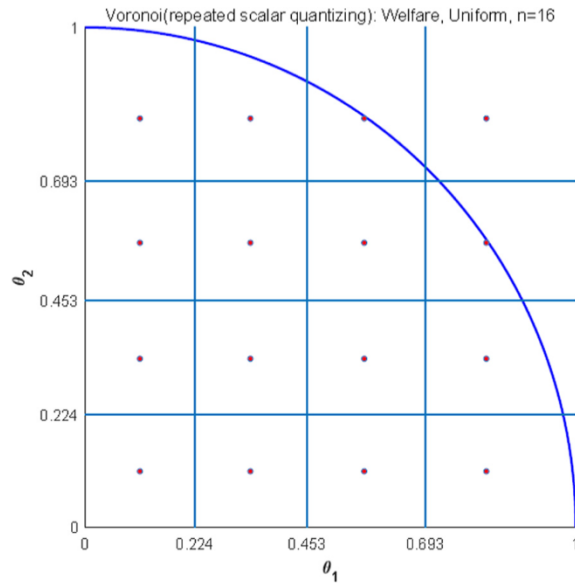

Fig. 4. Repeated scalar quantizer and representation points for welfare maximization with uniball distribution and $d = 2$, $n = 16$.

$$R_F = \max_{\{q(\theta), t(\theta)\}} \mathbb{E}_\theta \left[t(\theta) - \frac{1}{2} q(\theta)^T q(\theta) \right],$$

subject to the individual rationality and incentive constraints.

In an important contribution, Armstrong (1996) shows that the firm's revenue—given the separability condition (see Definition 3)— can be written as:

$$R_F = \mathbb{E}_\theta \left[\psi(\theta)^T q(\theta) - \frac{1}{2} q(\theta)^T q(\theta) \right],$$

where

$$\psi(\theta) = h(\theta) \theta, \quad h(\theta) = 1 - \frac{\beta(\theta)}{f(\theta)}, \quad \beta(\theta) = \int_1^{+\infty} f(r\theta) r^{d-1} dr. \tag{20}$$

The optimal continuous menu satisfies:

$$q^* (\theta) = \begin{cases} \psi (\theta) & \text{if } \theta \in \Theta^+ \\ 0 & \text{if } \theta \in \Theta \backslash \Theta^+ \end{cases},$$

where $\Theta^+ = \{\theta \in \Theta : h(\theta) \geq 0\}$ is the active type space. The maximum revenue can therefore be expressed as:

$$R_F = \mathbb{E}_\theta \left[\psi (\theta)^T q^* (\theta) - \frac{1}{2} q^* (\theta)^T q^* (\theta) \right] = \frac{1}{2} \int_\Theta \psi (\theta)^T \psi (\theta) \, dF (\theta). \tag{21}$$

The finite version of the revenue maximization problem specifies a menu which contains $n < \infty$ different items. Armstrong (1996) already observed that some consumers with low type vectors in the active type space Θ will leave the market when a finite menu is offered. Thus, there exists a region $P_0 \subseteq \Theta$, determined endogenously, such that all consumers with $\theta \in P_0$ will choose $q_0 = 0, t_0 = 0$. The seller chooses $\{P_k, q_k\}_{k=0}^n$ to maximize the expected revenue:

$$R_F (n) = \max_{\{P_k, q_k\}_{k=0}^n \in M} \left\{ \mathbb{E}_\theta \left[\psi (\theta)^T q - \frac{1}{2} q^T q \right] \right\}. \tag{22}$$

Virtual type space   Define for $\theta \in \Theta$, the virtual type vector

$$\widehat{\theta} = \psi (\theta) = h (\theta) \theta. \tag{23}$$

As in the one-dimensional analysis, we can transform the partition $\{P_k\}_{k=0}^n$ of the active type space into a partition $\{\widehat{P}_k\}_{k=0}^n$ of the virtual type space Θ as follows:

$$\theta \in P_k \Leftrightarrow \widehat{\theta} \in \widehat{P}_k = \{\psi (\theta) : \theta \in P_k\}.$$

In the virtual space, the expected revenue for an n-item menu can be written as:

$$\mathbb{E}_\theta \left[\psi (\theta)^T q - \frac{1}{2} q^T q \right] = \mathbb{E}_{\widehat{\theta}} \left[\widehat{\theta}^T q - \frac{1}{2} q^T q \right],$$

and the expected revenue of the optimal n-item menu is given as:

$$R_F (n) = \max_{\{\widehat{P}_k, q_k\}_{k=0}^n \in M} \mathbb{E}_{\widehat{\theta}} \left[\widehat{\theta}^T q - \frac{1}{2} q^T q \right].$$

The problem is now formally equivalent to the earlier multidimensional welfare maximization problem (7). We now consider how well the optimal n-item menu can approximate the performance of the optimal continuous menu. We can further write the revenue loss from quantization as:

$$\widehat{L}_F (n, d) \triangleq \frac{1}{2} \mathbb{E}_{\widehat{\theta}} \left[(\widehat{\theta} - q)^T (\widehat{\theta} - q) \right],$$

where we extend the expression $\widehat{L}_F (n)$ given by the earlier Definition 2 to account for the dimension d.

One can see that the above expression is indeed the multidimensional welfare loss in terms of the virtual type $\widehat{\theta}$. Let $G \left(\psi (\widehat{\theta}) \right) = F (\theta)$ denote the virtual type distribution and $g(\cdot)$ denote the density function; then,

$$\widehat{L}_F (n, d) = L_G (n, d). \tag{24}$$

Similar to the case for welfare maximization, we are interested in the worst-case behavior of the revenue loss over all joint distributions with a d-dimensional support set (i.e., the type space with positive and finite volume, and with the separability condition, thus

$$\widehat{L} (n, d) \triangleq \sup_{F \in \mathcal{F}_S} \widehat{L}_F (n, d).$$

Bounds on revenue loss   Using the relationship between the revenue maximization problem and the vector quantization problem, we can obtain upper and lower bounds on the revenue loss, as in the social welfare case.

Before investigating the multidimensional revenue loss upper and lower bounds, one may ask the following question: given that the original type distribution F satisfies the separability condition, does the virtual type distribution G also satisfy it? The answer is positive and we will show this important property in the following lemmas, which describe the behavior of the virtual type under the separability condition.

The separability condition requires the independence of the angle and the norm of the type distribution. A hyperspherical coordinate system, (see Blumenson (1960)) naturally captures this property. In this way, the type distribution can be written

in the product form as follows. We first introduce the hyperspherical coordinate transformation to the type vector space. Namely, for a d-dimensional type $\theta \in [0, 1]^d$, let $r \in [0, 1]$ and $\phi \in [0, \frac{\pi}{2}]^{d-1}$, such that:

$$
\begin{aligned}
\theta_1 &= r\cos(\phi_1), \\
\theta_2 &= r\sin(\phi_1)\cos(\phi_2), \\
&\cdots \\
\theta_{d-2} &= r\sin(\phi_1)\cdots\sin(\phi_{d-3})\cos(\phi_{d-2}), \\
\theta_{d-1} &= r\sin(\phi_1)\cdots\sin(\phi_{d-2})\cos(\phi_{d-1}).
\end{aligned}
\tag{25}
$$

We then denote the type distribution density function after the coordinate transformation as $f_T(r, \phi)$; i.e. $f(\theta) = f_T(r, \phi)$. The separability condition (3) can then equivalently be written as:

$$
f(\theta) = f_T(r, \phi) = f_r(r) \times f_\phi(\phi).
\tag{26}
$$

We can similarly transform the density of the virtual utility as the next lemma establishes.

Lemma 1 *(Separability of virtual utility function). Let $g(\cdot)$ be the probability density function of the virtual type vector distribution. If $g(\theta)$ is transformed to $g_T(r, \phi)$ with the hyperspherical coordinate transformation, then*

$$
g_T(r, \phi) = g_r(r) \times g_\phi(\phi).
$$

Lemma 1 shows that the distribution after the virtual valuation transformation (namely $G(\cdot)$) also satisfies the separability condition, and thus the *active type space* Θ to be quantized is essentially the area between the positive orthant of a d-dimensional sphere with unit radius and a d-dimensional sphere with radius $\underline{r}$, where $\underline{r} \in (0, 1)$—both centered at 0. Our previous analysis of multidimensional welfare maximization therefore applies to $G(\cdot)$.

Proposition 7 *(Upper bound on multidimensional revenue loss). For any $F \in \mathcal{F}_S$, given both d and $n^{\frac{1}{d}}$ are sufficiently large, the revenue loss satisfies*

$$
\widehat{L}_F(n, d) \le \frac{1}{8n^{\frac{2}{d}}}.
\tag{27}
$$

Proof. Given distribution $F \in \mathcal{F}_S$, by (24), $\widehat{L}_F(n, d) = L_G(n, d)$. By Lemma 1, it follows that the distribution G also satisfies the separability condition, and thus by Proposition 6:

$$
\widehat{L}_F(n, d) \le 1/8n^{\frac{2}{d}},
$$

which completes the proof. $\square$

In the one-dimensional revenue maximization scenario, the upper bound of $\widehat{L}(n)$ given by the high-rate non-uniform quantizer is indeed tight, since a type distribution that meets the bound is presented. However, in the multidimensional case a distribution that attains the upper bound in Proposition 7 may not exist, as Armstrong (1996) has shown that there always exist some customers that are not served by the finite menu when $d \ge 2$.

Nonetheless, we next show that the upper bound on the revenue loss is indeed achieved when the virtual type probability is equally distributed in the quantization area. While an explicit form of the original type distribution leading to a *uniball*-distributed virtual type is hard to find, we provide a family of special truncated beta distributions whose virtual type distribution asymptotically approaches the *uniball* distribution. Thus, the corresponding revenue loss asymptotically approximates the upper bound stated in Proposition 8.

Proposition 8 *(Revenue loss for truncated beta distribution). Let F_b be a special truncated beta distribution, specifically,*

$$
f_b(\theta) =
\begin{cases}
\frac{(d+3)2^{d-1}\Gamma(d/2)}{2\pi^{d/2}}\|\theta\|^{\frac{1-d}{2}}, & \text{if } \left(\frac{2}{d+3}\right)^{\frac{2}{d+1}} \le \|\theta\| \le 1; \\
0, & \text{otherwise.}
\end{cases}
$$

Then, if d is sufficiently large, and $n^{1/d}$ is sufficiently large, we have

$$
\widehat{L}_{F_b}(n, d) \approx \frac{1}{8n^{\frac{2}{d}}}.
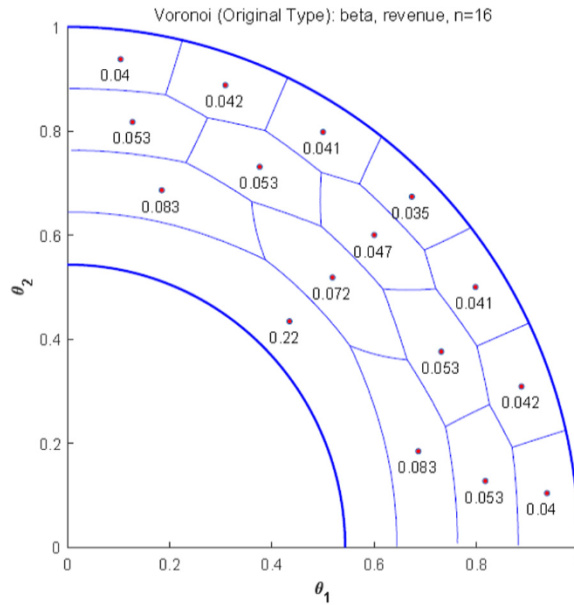\tag{28}
$$

Fig. 5. Quantizer and representation points for revenue maximization with beta distribution and $d = 2$, $n = 16$.

In Fig. 5 we illustrate the resulting quantizer. Thus, by combining the results of Proposition 7 and 8, we have indeed obtained the following asymptotic bound on the multidimensional revenue loss, namely:

$$\widehat{L}(n, d) \approx \frac{1}{8n^{\frac{2}{d}}}.$$

5. Conclusion

We explored the consequences of economic transactions with limited information within the concrete setting of the nonlinear pricing model. Using the linear-quadratic specification, we relate both social welfare maximization and revenue maximization to the quantization problem in information theory. Using this link, we introduce the Lloyd-Max conditions that the optimal finite menu for the socially efficient and the revenue-maximizing mechanism must satisfy. Additionally, we study the performance of the finite menus relative to the optimal continuous menu. Our analysis shows that for both social welfare and the seller's revenue, the losses due to the usage of the n-item finite menu converge to zero at a rate proportional to $1/n^2$.

Based on the information-theoretic approach in the one-dimensional environment, we generalize our results to the multi-product environment. We introduce the vector quantization gain and a decomposition method, and obtain a vector-quantization-based upper bound and a lower bound on the welfare loss and the revenue loss. The vector-quantization-based upper bound is tighter than the repeated scalar upper bound, and the improvement becomes significant in higher dimensions, and/or when a correlation among the buyer's preferences over multiple products exists. This shows that it is beneficial to bundle the buyer's preferences over multiple goods, and then design the finite menus jointly in multiple dimensions.

We restricted attention to a linear-quadratic payoff environment throughout. A natural question is therefore to what extent our results would generalize (or require modification) in a broader class of non-linear environments. We briefly discuss how existing results in information theory may assist us.

A first observation is that given a general utility $u(\theta, q)$ and cost function $c(q)$, the finite menu that maximizes social welfare:

$$W = \mathbb{E}_\theta[u(\theta, q) - c(q)]$$

is equivalent to the optimal quantizer for minimizing the expected distortion $\mathbb{E}_\theta[d(\theta, q)]$, where

$$d(\theta, q) = c(q) - u(\theta, q).$$

Thus, a generalization to nonlinear payoff environments is equivalent to a distortion function different from the mean squared error used in this paper. Indeed, the high-rate non-uniform quantization results were generalized in various directions, for example to a locally quadratic distortion in Li et al. (1999), to the p-th power distortion $d(x, y) = |x - y|^p$ in Zador (1982) and more recently to the Orlicz-norm distortion in Dereich and Vormoor (2011). In particular, Lookabaugh and Gray (1989b), developed an argument for vector quantization gain under the p-th power distortion which we could use to obtain

approximation results similar to the ones obtained here. The information-theoretic arguments appear therefore portable to non-linear environments. It follows that the critical and open challenge would seem to be a general characterization of incentive compatible allocation in many dimensions.

While the nonlinear pricing environment is of interest by itself, it also represents an elementary instance of the general mechanism design environment. The simplicity of the nonlinear pricing problem arises from the fact that it can be viewed as a relationship between the principal (here, the seller) and a single agent (here, the buyer), even in the presence of many buyers. The reason for this simplicity is that the principal does not have to solve allocative externalities. By contrast, in auctions and other multi-agent allocation problems, the allocation (and hence the relevant information), with respect to a given agent, constrains and is constrained by the allocation to the other agents.

Finally, the current analysis focused on limited information and the ensuing problem of efficient source coding. But clearly, from an information-theoretic as well as an economic viewpoint, it is natural to augment the analysis to reliable communication between the agent and principal over noisy channels—the problem of channel coding—which we plan to address in future work.

Appendix

The appendix collects auxiliary results and the remaining proofs of the results in the main body of the text.

Lemma 2 (Standard form). *For every utility and revenue function given by (1) and (3) there is an equivalent standard form given by (4).*

Proof. We say that the utility and the cost function have the standard form if $\Phi = \Sigma = I_d$ (the $d \times d$ identity matrix). We can transform the utility and the cost function into the standard form as follows: We diagonalize the symmetric positive-definite matrix Σ: $\Sigma = P^T \Lambda P$, where $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_d)$, $\lambda_i > 0$ is the i-th eigenvalue of Σ, and P is a unitary matrix (i.e., $P^T P = I_d$). Let $B = \Lambda^{1/2} P$ and $A = \Lambda^{-1/2} P \Phi^T$, where $\Lambda^i = \mathrm{diag}(\lambda_1^i, ..., \lambda_d^i)$, $i = \pm \frac{1}{2}$. Then, it is easy to show that $A^T B = \Phi$ and $B^T B = \Sigma$. If we introduce the new type vector $\widehat{\theta} = A\theta$ and the new quantity (quality) vector $\hat{q} = Bq$, then the buyer's net utility and the cost function can be written in the standard form in terms of $\widehat{\theta}$ and $\hat{q}$:

$$u(\theta, q) = \theta^T \Phi q = \theta^T A^T B q = \widehat{\theta}^T \hat{q},$$

$$c(q) = \frac{1}{2} q^T \Sigma q = \frac{1}{2} q^T B^T B q = \frac{1}{2} \hat{q}^T \hat{q}.$$

Thus, without loss of generality, we focus on the standard form with $\Phi = \Sigma = I_d$. $\square$

Proof of Proposition 1. The conditional mean in any interval P_k is $\mathbb{E}[\theta | \theta \in P_k] = (\theta_k + \theta_{k-1})/2$. From (9), the optimal menu $\{P_k^*, q_k^*\}_{k=1}^n$ must satisfy:

$$\theta_k^* = \frac{q_k^* + q_{k+1}^*}{2}, \quad q_k^* = \frac{\theta_{k-1}^* + \theta_k^*}{2}, \quad k = 1, \ldots, n.$$

Hence, $\theta_{k+1}^* - 2\theta_k^* + \theta_{k-1}^* = 0$. Note that $\theta_0^* = 0, \theta_n^* = 1$, and thus we have a unique solution to the Lloyd-Max conditions given by (10). The expected social welfare is

$$W_{\mathcal{U}}(n) = \mathbb{E}_\theta \left[\theta q^*(\theta) - \frac{1}{2} q^*(\theta)^2 \right] = \sum_{k=1}^n \int_{\theta_{k-1}^*}^{\theta_k^*} \left[\theta q_k^* - \frac{1}{2} q_k^{*2} \right] d\theta = \frac{1}{6} - \frac{1}{24n^2}.$$

By contrast, the social welfare realized by the optimal continuous menu is $W_{\mathcal{U}} = \frac{1}{2} \mathbb{E}[\theta^2] = 1/6$, which yields as welfare loss: $L_{\mathcal{U}}(n) = 1/24n^2$. $\square$

Proof of Proposition 2. Let F denote the type distribution supported on $[0, 1]$, and $f(\cdot)$ the density function. As n goes to infinity, by Gersho and Gray (2007), p. 186, the quantization loss of the Lloyd-Max quantizer is approached by a high-rate non-uniform quantizer. Specifically, the minimum quantization distortion is given by:

$$\underline{D} = \frac{1}{12n^2} \left(\int_0^1 f(x)^{1/3} \right)^3.$$

Thus, we have the corresponding welfare loss

$$L_F(n) = \frac{1}{2}\underline{D} = \frac{1}{24n^2}\left(\int_0^1 f(x)^{1/3}dx\right)^3.$$

Note that by Hölder's inequality, we have

$$\int_0^1 f(x)^{1/3}dx \le \left[\int_0^1 f(x)dx\right]^{\frac{1}{3}}\left[\int_0^1 1dx\right]^{\frac{2}{3}}$$

$$= \left[\int_0^1 f(x)dx\right]^{\frac{1}{3}}$$

$$= 1,$$

where the equality holds when $f(x) = 1$, $x \in [0, 1]$. We then arrive at the upper bound of welfare loss given n is large,

$$L_F(n) \le \frac{1}{24n^2},$$

which completes the proof. □

Proof of Proposition 3. Let $g(\cdot)$ denote the virtual type distribution, and $L_G(n)$ denote the quantization loss by the Lloyd-Max algorithm applied on the $g(\cdot)$ distribution in interval $[0, 1]$ with n quantizing levels (i.e., the Lloyd-Max applied on the virtual type), then we have $\widehat{L}_F(n) = L_G(n)$. Thus, as in the welfare maximization case, we have:

$$L_G(n) = \frac{1}{24n^2}\left(\int_0^1 g(x)^{1/3}dx\right)^3.$$

Evaluating the RHS, we have

$$\int_0^1 g(x)^{1/3}dx \le \left[\int_0^1 g(x)dx\right]^{\frac{1}{3}}\left[\int_0^1 1dx\right]^{\frac{2}{3}}$$

$$= \left[\int_0^1 g(x)dx\right]^{\frac{1}{3}}$$

$$\overset{(a)}{\le} 1$$

where (a) holds since $\int_0^1 g(x)dx \le \int_{-\infty}^1 g(x)dx = \int_0^1 f(x)dx = 1$. Thus, we have the upper bound of revenue loss given n is large,

$$\widehat{L}_F(n) \le \frac{1}{24n^2},$$

which completes the proof. □

Proof of Proposition 4. Consider the uniform distribution with support on $[1/2, 1]$:

$$F(\theta) = \begin{cases} 0 & \text{if } \theta \in [0, \frac{1}{2}). \\ 2, & \text{if } \theta \in [\frac{1}{2}, 1]. \end{cases}$$

The virtual type is given by:

$$\psi(\theta) \begin{cases} < 0 & \text{if } \theta \in [0, \frac{1}{2}); \\ = 2\theta - 1 & \text{if } \theta \in [\frac{1}{2}, 1], \end{cases}$$

with its inverse function for the positive input:

$$\psi^{-1}(\widehat{\theta}) = \frac{\widehat{\theta}+1}{2}, \widehat{\theta} \in [0,1]$$

and the virtual type distribution

$$g(\widehat{\theta}) = f(\psi^{-1}(\widehat{\theta}))(\psi^{-1})'(\widehat{\theta}) = 1 \quad, \widehat{\theta} \in [0,1].$$

Thus,

$$\widehat{L}(\mathcal{U}_p, n) = \frac{1}{24n^2} \left(\int_0^1 g(x)^{1/3} dx \right)^3 = \frac{1}{24n^2}.$$

The revenue loss for distribution $\mathcal{U}_p$ thus attains the bound exactly:

$$\widehat{L}(n) = \frac{1}{24n^2},$$

which completes the proof. □

Given the joint distribution $F \in \mathcal{F}$, for each dimension θ_l, $1 \leq l \leq d$, consider a K-level scalar quantizer $\{P_{l,k}, q_{l,k}\}_{k=1}^K \in M$ on $[0,1]$ where M is the set of all scalar quantizers for the marginal distribution F_l:

$$M = \left\{ \{P_{l,k}, q_{l,k}\}_{k=1}^K : P_{l,k} = [\theta_{l,k-1}, \theta_{l,k}), 0 = \theta_{l,0} \leq \theta_{l,1} \leq \ldots \leq \theta_{l,n} = 1 \right\}.$$

We construct the corresponding d-dimensional K^d-region repeated scalar quantizer $\{P_k', q_k'\}_{k=1}^{K^d}$ in the type space $[0,1]^d$ as:

$$\{P_k'\}_{k=1}^{K^d} = \{P_{1,k_1} \times \ldots \times P_{d,k_d} : k_l \in \{1, \ldots, K\}, 1 \leq l \leq d\},$$

$$\{q_k'\}_{k=1}^{K^d} = \{(q_{1,k_1}, \ldots, q_{d,k_d})^T : k_l \in \{1, \ldots, K\}, 1 \leq l \leq d\}.$$

One can see that the set of regions $\{P_k'\}_{k=1}^{K^d}$ are orthotopes, i.e., the Cartesian product of intervals in d dimensions. A simple upper bound on the welfare loss in multiple dimensions is the repeated scalar quantizer.

Given an identical type distribution F, with F_l being the marginal distribution, the social welfare loss of the repeated scalar quantizer with n quantization levels each dimension is given by $\sum_{l=1}^d L_{F_l}(n)$.

Let the welfare loss of the optimal vector quantizer with $n \cdot d$ quantization levels be $L_F(n \cdot d, d)$, then $L_F(n \cdot d, d) \leq \sum_{l=1}^d L_{F_l}(n)$ since the repeated scalar quantizer is a special case of vector quantizer.

Moreover, the vector quantization gain for social welfare is defined by the ratio of the welfare loss induced by the repeated scalar quantizer to the welfare loss induced by the optimal vector quantizer, namely

$$G_W = \frac{\sum_{l=1}^d L_{F_l}(n)}{L_F(n \cdot d, d)}. \tag{29}$$

Denote by f and f_l the joint density function and the marginal density function, respectively. When the number n of items in each dimension becomes sufficiently large, the vector quantization gain can be decomposed as follows as, established by Lookabaugh and Gray (1989a), p. 1022:

$$G_W \approx F(d) \times S(f_l, d) \times M(f_l, f, d), \tag{30}$$

where $F(d)$, $S(f_l, d)$, and $M(f_l, f, d)$ are called the space-filling advantage, shape advantage, and dependence advantage, respectively.

Specifically, when $d \geq 3$, it is optimal to choose the admissible polytopes as close as possible to the d-dimensional sphere, leading to an asymptotic space-filling advantage, $\lim_{d \to \infty} F(d) = \pi e/6$, as established by Conway and Sloane (1985).

Given the dimension d, the shape advantage $S(f_l, d)$ depends solely on f_l (e.g., the uniform distribution does not provide any shape advantage):

$$S(f_l, d) = \frac{\left[\int (f_l(\theta_l))^{1/3} d\theta_l \right]^3}{\left[\int (f_l(\theta_l))^{d/d+2} d\theta_l \right]^{d+2}}. \tag{31}$$

The dependence advantage $M(f_l, f, d)$ is determined by both f and f_l, (e.g., any independent and identical distribution does not provide any dependence advantage):

$$M(f_l, f, d) = \frac{\left[\int (f_l(\theta_l))^{d/d+2} d\theta_l\right]^{d+2}}{\left[\int \cdots \int (f(\theta_1, \ldots, \theta_d))^{d/d+2} d\theta_1 \cdots d\theta_d\right]^{(d+2)/d}}. \tag{32}$$

Proof of Proposition 5. The welfare loss is computed using the decomposition given by (30):

$$L_{\mathcal{U}_b}(n, d) \approx \frac{d \times L_{\hat{\mathcal{U}}_b}(n)}{F(d) \times S\left(\hat{\mathcal{U}}_b, d\right) \times M\left(\hat{\mathcal{U}}_b, \mathcal{U}_b, d\right)}$$

$$\overset{(a)}{\approx} \frac{d\left[\int_0^1 \hat{\mathcal{U}}_b(\theta)^{\frac{1}{3}} d\theta\right]^3}{24 n^{\frac{2}{d}}} \times \frac{6}{\pi e} \times \frac{\left[\int_0^1 \cdots \int_0^1 \left(\frac{1}{V_+(d)}\right)^{\frac{d}{d+2}} d\theta_1 \cdots d\theta_d\right]^{\frac{d+2}{d}}}{\left[\int_0^1 \hat{\mathcal{U}}_b(\theta)^{\frac{1}{3}} d\theta\right]^3}$$

$$= \frac{d \times V_+(d)^{\frac{2}{d}}}{4\pi e n^{\frac{2}{d}}}$$

$$\overset{(b)}{\approx} \frac{1}{8 n^{\frac{2}{d}}},$$

where $\hat{\mathcal{U}}_b$ denotes the marginal distribution of $\mathcal{U}_b$, which need not be computed explicitly; (a) used the one-dimensional optimal high-rate non-uniform quantizer and (b) holds by Stirling's approximation. □

Proof of Proposition 6. Under the assumption that both d and $n^{1/d}$ are sufficiently large, let $f(x)$ denotes the type probability density distribution. Then, by Gersho and Gray (2007), pp. 339, generalizing the non-uniform high-rate quantizing result into d-dimension, the optimal quantization distortion is given by:

$$D \geq \frac{d}{(d+2) n^{\frac{2}{d}}} \left(\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\right)^{-\frac{2}{d}} \left[\int_\theta f(\theta)^{\frac{d}{d+2}} d\theta\right]^{\frac{d+2}{d}} \tag{33}$$

where $\Gamma(\cdot)$ is the common gamma function. The equality holds if n is sufficiently large, and the optimal high-rate non-uniform quantizer is used, denoted as D_{min}.

Thus, the welfare loss is upper bounded by

$$L_F(n, d) = \frac{1}{2} D_{min} = \frac{d}{2(d+2) n^{\frac{2}{d}}} \left(\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\right)^{-\frac{2}{d}} \left[\int_\theta f(\theta)^{\frac{d}{d+2}} d\theta\right]^{\frac{d+2}{d}}$$

Applying Hölder's inequality, we have

$$\int_\theta f(\theta)^{\frac{d}{d+2}} d\theta \leq \left(\int_\theta f(\theta) d\theta\right)^{\frac{d}{d+2}} \left(\int_\theta 1 d\theta\right)^{\frac{2}{d+2}}$$

$$\overset{(a)}{\leq} (V_+(d))^{\frac{2}{d+2}}$$

where (a) holds, since given the separability condition the quantization space is at most the positive orthant of the d-dimension unit ball.

Thus, when d is sufficiently large the welfare loss is upper-bounded by

$$L_F(n, d) \leq \frac{d}{2(d+2) n^{\frac{2}{d}}} \left(\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\right)^{-\frac{2}{d}} \left(\frac{1}{2^d} \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\right)^{\frac{2}{d}}$$

$$= \frac{d}{8(d+2) n^{\frac{2}{d}}}$$

$$\approx \frac{1}{8 n^{\frac{2}{d}}},$$

which completes the proof. □

Proof of Lemma 1. We define the cumulative density function of original type vector $\theta = (r_\theta, \phi_\theta)$ for polar coordinates to be

$$F_T(r, \phi) = \Pr\{r_\theta \leq r, \phi_\theta \leq \phi\},$$

where $\phi_\theta \leq \phi$ denotes the inequality for all $d-1$ components of vectors ϕ_θ and ϕ.

Since $f(\theta)$ satisfies the separability condition, we have:

$$F_T(r_\theta, \phi_\theta) = \int_0^{r_\theta} \int_{\phi \leq \phi_\theta} f_T(r, \phi) d\phi dr$$

$$= \int_0^{r_\theta} f_r(r) dr \times \int_{\phi \leq \phi_\theta} f_\phi(\phi) d\phi$$

$$= \Pr\{r_\theta \leq r\} \times \Pr\{\phi_\theta \leq \phi\}.$$

Similarly, we define the cumulative density function of virtual type vector $\widehat{\theta} = (r_{\widehat{\theta}}, \phi_{\widehat{\theta}})$ for polar coordinates to be

$$G_T(r, \phi) = \Pr\{r_{\widehat{\theta}} \leq r, \phi_{\widehat{\theta}} \leq \phi\},$$

with

$$G_T(r_{\widehat{\theta}}, \phi_{\widehat{\theta}}) = \int_0^{r_{\widehat{\theta}}} \int_{\phi \leq \phi_{\widehat{\theta}}} g_T(r, \phi) d\phi dr.$$

Because $\widehat{\theta} = \phi(\theta) = h(\theta)\theta$, with $h(\theta) \in \mathbb{R}$, we know that the virtual valuation transformation preserves the angle $\hat{\phi} = \phi$. Thus, we have

$$\Pr\{r_{\widehat{\theta}} \leq rh(\theta), \phi_{\widehat{\theta}} \leq \phi\} = \Pr\{r_\theta \leq r, \phi_\theta \leq \phi\}.$$

On the other hand, we know that $h(\theta)$ only depends on $\|\theta\|$ (namely r). Specifically,

$$h(\theta) = 1 - \frac{\int_1^{+\infty} f_r(tr) f_\phi(\phi) t^{d-1} dt}{f_r(r) f_\phi(\phi)}$$

$$= 1 - \frac{\int_1^{+\infty} f_r(tr) t^{d-1} dt}{f_r(r)}.$$

Thus, we can denote $h_r(r) = h(\theta)$ and then:

$$G_T(rh_r(r), \phi) = \Pr\{r_{\widehat{\theta}} \leq rh_r(r), \phi_{\widehat{\theta}} \leq \phi\}$$

$$= \Pr\{r_\theta \leq r, \phi_\theta \leq \phi\}$$

$$= \Pr\{r_\theta \leq r\} \times \Pr\{\phi_\theta \leq \phi\}.$$

Taking the derivative, we have

$$\frac{\partial^2 G_T(rh_r(r), \phi)}{\partial r \partial \phi} = g_T(rh_r(r), \phi)) \left(h_r(r) + rh_r'(r)\right).$$

And thus

$$g_T(rh_r(r), \phi)) = \frac{f_r(r)}{\left(h_r(r) + rh_r'(r)\right)} \times f_\phi(\phi).$$

It follows that the virtual type distribution g_T could be written as the product of two functions which solely depend on r and ϕ, respectively. Thus, the virtual type distribution satisfies the separability condition. $\square$

Proof of Proposition 8. First, we show that $f_b(\theta)$ is indeed a valid probability density distribution. For convenience of the integration, we denote by $S(r)$ the area of a $(d-1)$-dimensional sphere with radius r in the positive orthant, namely

$$S(r) = \frac{1}{2^d} \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1}.$$

Define the radius direction probability $f_r(\cdot)$ such that $f_r(\|\theta\|) = f_b(\theta)$; then we have

$$
\int_\theta f_b(\theta)d\theta = \int_0^1 f_r(r)S(r)dr
$$

$$
= \int_{\left(\frac{2}{d+3}\right)^{\frac{2}{d+1}}}^1 \frac{(d+3)2^{d-1}\Gamma(d/2)}{2\pi^{d/2}} r^{\frac{1-d}{2}} \frac{1}{2^d} \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1} dr
$$

$$
= \int_{\left(\frac{2}{d+3}\right)^{\frac{2}{d+1}}}^1 \frac{d+3}{2} r^{\frac{d-1}{2}} dr
$$

$$
= 1
$$

which shows that $f_b(\theta)$ is a valid probability density distribution. It is obvious that $F_b \in \mathcal{F}_S$; thus by Lemma 1 the virtual type distribution also satisfies the separability condition. Specifically corresponding to (20),

$$
\beta(\theta) = \int_1^\infty f_b(r\theta)r^{d-1}dr = \frac{\int_{\|\theta\|}^1 f_r(t)t^{d-1}dt}{\|\theta\|^d}
$$

$$
= \begin{cases} \|\theta\|^{-d}\int_{\|\theta\|}^1 At^{\frac{1-d}{2}}t^{d-1}dt = \frac{2A}{d+1}\|\theta\|^{-d}\left(1-\|\theta\|^{\frac{d+1}{2}}\right) &, \quad \text{if } \left(\frac{2}{d+3}\right)^{\frac{2}{d+1}} \leq \|\theta\| \leq 1 \\ \|\theta\|^{-d}\int_{\left(\frac{2}{d+3}\right)^{\frac{2}{d+1}}}^1 At^{\frac{1-d}{2}}t^{d-1}dt = \frac{2A}{d+3}\|\theta\|^{-d} &, \quad \text{if } \|\theta\| < \left(\frac{2}{d+3}\right)^{\frac{2}{d+1}} \end{cases}
$$

where $A = \frac{(d+3)2^{d-1}\Gamma(d/2)}{2\pi^{d/2}}$ is the coefficient in $f_b(\theta)$. Thus, we have

$$
h(\theta) = 1 - \frac{\beta(\theta)}{f_b(\theta)}
$$

$$
= \begin{cases} 1 - \frac{\frac{2A}{d+1}\|\theta\|^{-d}\left(1-\|\theta\|^{\frac{d+1}{2}}\right)}{A\|\theta\|^{\frac{1-d}{2}}} = \frac{d+3}{d+1} - \frac{2}{d+1}\|\theta\|^{-\frac{d+1}{2}} &, \quad \text{if } \left(\frac{2}{d+3}\right)^{\frac{2}{d+1}} \leq \|\theta\| \leq 1; \\ < 0 &, \quad \text{if } \|\theta\| < \left(\frac{2}{d+3}\right)^{\frac{2}{d+1}}. \end{cases}
$$

Let $\underline{\theta} = \left(\frac{2}{d+3}\right)^{\frac{2}{d+1}}$, since $h(\theta) < 0$ when $\|\theta\| < \underline{\theta}$, and $h(\theta) \geq 0$ when $\|\theta\| \geq \underline{\theta}$, we know that $\underline{\theta}$ is indeed the radius threshold for a positive virtual valuation. The virtual type $\psi(\theta)$ corresponding to positive $h(\theta)$ is defined as

$$
\widehat{\theta} = \psi(\theta) = \left(\frac{d+3}{d+1} - \frac{2}{d+1}\|\theta\|^{-\frac{d+1}{2}}\right)\theta, \quad \text{for } \underline{\theta} \leq \|\theta\| \leq 1.
$$

Let $g(\widehat{\theta})$ denote the distribution of virtual type $\widehat{\theta}$. Since $\widehat{\theta}$ has the same angle as θ, but a different magnitude, and $f_b(\theta)$ only depends on the magnitude $\|\theta\|$, we know $g(\widehat{\theta})$ only depends on $\|\widehat{\theta}\|$, i.e., $g(\widehat{\theta}) = g_r(\|\widehat{\theta}\|)$.

Let

$$
G_R(r) = \Pr(\|\widehat{\theta}\| \leq r)
$$

denote the probability that the virtual type magnitude is no more that r, with $r \in [0, 1]$. Similarly, define

$$
F_R(r) = \Pr(\|\theta\| \leq r).
$$

We then have

$$
F_R(r) = \Pr(\|\theta\| \leq r)
$$

$$
= \Pr\left(\frac{d+3}{d+1}\|\theta\| - \frac{2}{d+1}\|\theta\|^{-\frac{d-1}{2}} \leq \frac{d+3}{d+1}r - \frac{2}{d+1}r^{-\frac{d-1}{2}}\right)
$$

$$
= G_R(\psi_r(r)), \tag{34}
$$

where

$$\psi_r(r) = \frac{d+3}{d+1}r - \frac{2}{d+1}r^{-\frac{d-1}{2}}.$$

On the other hand, we could calculate $G_R(\psi_r(r))$ and $F_R(r)$ separately for $r \geq \underline{\theta}$, where

$$F_R(r) = \int_0^r f_r(t)S(t)dt$$

$$= \frac{A}{2^d}\frac{2\pi^{d/2}}{\Gamma(d/2)}\int_{\underline{\theta}}^r t^{-\frac{d-1}{2}}t^{d-1}dt$$

$$= \frac{d+3}{d+1}r^{\frac{d+1}{2}} - \frac{2}{d+1},$$

and

$$G_R(\psi_r(r)) = \int_0^{\psi_r(r)} g_r(t)S(t)dt.$$

Thus by (34),

$$\int_0^{\psi_r(r)} g_r(t)S(t)dt = \frac{d+3}{d+1}r^{\frac{d+1}{2}} - \frac{2}{d+1}.$$

Taking the derivative of both sides, we have

$$g_r(\psi_r(r))S(\psi_r(r))\psi_r'(r) = \frac{d+3}{2}r^{\frac{d-1}{2}}. \tag{35}$$

Recall that the minimum quantization distortion for a multidimensional variable $\widehat{\theta}$ achieved with the high-rate non-uniform quantizer is given by

$$D_{min} = \frac{1}{n^{2/d}}\frac{d}{d+2}\left(\frac{2\pi^{d/2}}{d\Gamma(d/2)}\right)^{-\frac{2}{d}}\left[\int_{\underline{\theta}} g(\widehat{\theta})^{\frac{d}{d+2}}d\widehat{\theta}\right]^{\frac{d+2}{d}}.$$

Given the correspondence in (24), we can compute the revenue loss as:

$$\widehat{L}_{F_b}(n,d) = \frac{1}{2}D_{min}$$

$$= \frac{1}{n^{2/d}}\frac{d}{2(d+2)}\left(\frac{2\pi^{d/2}}{d\Gamma(d/2)}\right)^{-\frac{2}{d}}\left[\int_0^1 g_r(\widehat{r})^{\frac{d}{d+2}}S(\widehat{r})d\widehat{r}\right]^{\frac{d+2}{d}}.$$

The integral term can be further written as

$$\int_0^1 g_r(\widehat{r})^{\frac{d}{d+2}}S(\widehat{r})d\widehat{r} = \int_{\underline{\theta}}^1 g_r(\psi_r(r))^{\frac{d}{d+2}}S(\psi_r(r))\psi_r'(r)dr$$

$$\overset{(a)}{=} \int_{\underline{\theta}}^1 \left(\frac{d+3}{2}r^{\frac{d-1}{2}}\right)^{\frac{d}{d+2}}S(\psi_r(r))^{\frac{2}{d+2}}\left(\psi_r'(r)\right)^{\frac{2}{d+2}}dr$$

$$= \left(\frac{1}{2^d}\frac{2\pi^{d/2}}{\Gamma(d/2)}\right)^{\frac{2}{d+2}}M,$$

where (a) holds by (35) and

$$M = \int_{\left(\frac{2}{d+3}\right)^{\frac{2}{d+1}}}^{1} \left(\frac{d+3}{2}r^{\frac{d-1}{2}}\right)^{\frac{d}{d+2}} \left(\frac{d+3}{d+1}r - \frac{2}{d+1}r^{-\frac{d-1}{2}}\right)^{\frac{2(d-1)}{d+2}} \left(\frac{d+3}{d+1} + \frac{d-1}{d+1}r^{-\frac{d+1}{2}}\right)^{\frac{2}{d+2}} dr$$

Thus, we have

$$\widehat{L}_{F_b}(n,d) = \frac{1}{n^{2/d}} \frac{d}{2(d+2)} \left(\frac{2\pi^{d/2}}{d\Gamma(d/2)}\right)^{-\frac{2}{d}} \left(\frac{1}{2^d} \frac{2\pi^{d/2}}{\Gamma(d/2)}\right)^{\frac{2}{d}} M^{\frac{d+2}{d}}$$

$$= \frac{1}{8n^{2/d}} \frac{d}{d+2} d^{\frac{2}{d}} M^{\frac{d+2}{d}}. \tag{36}$$

Note that all terms in (36) are computable and one can take this as an accurate lower bound of $\widehat{L}(n,d)$. Nevertheless, we now make further approximation to (36) and show that it asymptotically approaches the upper bound $1/8n^{2/d}$, given by Proposition 7.

Provided d is sufficiently large, we can approximate M to be

$$M \approx \int_{\theta}^{1} \frac{d}{2}r^{\frac{d}{2}} \left(r - \frac{2}{d}r^{-\frac{d}{2}}\right)^2 \left(1 + r^{-\frac{d}{2}}\right)^{\frac{2}{d}} dr$$

$$\overset{(b)}{\approx} \int_{\theta}^{1} \frac{d}{2}r^{\frac{d}{2}} \left(r - \frac{2}{d}r^{-\frac{d}{2}}\right)^2 dr$$

$$\approx \int_{\theta}^{1} \frac{d}{2}r^{\frac{d}{2}} \left[r^2 + \frac{4}{d^2}r^{-d} - \frac{4}{d}r^{-\frac{d}{2}}\right] dr$$

$$\approx \int_{\theta}^{1} \frac{d}{2}r^{\frac{d}{2}} dr + \int_{\theta}^{1} \frac{2}{d}r^{-\frac{d}{2}} dr + \int_{\theta}^{1} 2dr$$

$$\overset{(c)}{\approx} \int_{\theta}^{1} \frac{d}{2}r^{\frac{d}{2}} dr$$

$$= \frac{d}{d+3}$$

where (b) holds since $\left(1 + r^{-\frac{d}{2}}\right)^{\frac{2}{d}} \approx 1$ given d is large. (c) holds since $\int_{\theta}^{1} \frac{2}{d}r^{-\frac{d}{2}} dr \approx 0$ and $\int_{\theta}^{1} 2dr \approx 0$.

Thus, the revenue loss can be approximated as

$$\widehat{L}_{F_b}(n,d) \approx \frac{1}{8n^{2/d}} \frac{d}{d+2} d^{\frac{2}{d}} \left(\frac{d}{d+3}\right)^{\frac{d+2}{d}}$$

$$\approx \frac{1}{8n^{2/d}},$$

which completes the proof. $\square$

Proposition 9 (*Welfare loss for multi-dimensional uniform distribution*). *Given d and $n^{\frac{1}{d}}$ are large, for the d-dimensional uniform distribution $\mathcal{U}$ (i.e., θ_l are i.i.d. uniformly distributed on $[0,1]$ for $l = 1, \cdots, d$), we have*

$$L_{\mathcal{U}}(n,d) \approx \frac{d}{4\pi e n^{\frac{2}{d}}}.$$

Proof. Following the decomposition given by (30), we know that the welfare loss for the multidimensional uniform distribution can be calculated as

$$L_{\mathcal{U}}(K^d, d) = \frac{d \times \frac{1}{24K^2}}{G_W},$$

where $K = \lfloor n^{\frac{1}{d}} \rfloor$, and

$$G_W \approx SF \times S \times DP \approx \frac{\pi e}{6} \times 1 \times 1,$$

thus,

$$L_{\mathcal{U}}(K^d, d) = \frac{d}{24K^2} \times \frac{6}{\pi e} = \frac{d}{4\pi e K^2}.$$

We then have that given d and $n^{\frac{1}{d}}$ are sufficiently large,

$$L_{\mathcal{U}}(n, d) \approx \frac{d}{4\pi e n^{\frac{2}{d}}},$$

which completes the proof. $\square$

References

Armstrong, M., 1996. Multiproduct nonlinear pricing. Econometrica 64, 51–76.

Bergemann, D., Pesendorfer, M., 2007. Information structures in optimal auctions. J. Econ. Theory 137, 580–609.

Bergemann, D., Shen, J., Xu, Y., Yeh, E., 2012a. Mechanism design with limited information: the case of nonlinear pricing. In: Jain, R., Kannan, R. (Eds.), Game Theory for Networks. In: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 75. Springer Verlag, Berlin, pp. 1–10.

Bergemann, D., Shen, J., Xu, Y., Yeh, E., 2012b. Multi-dimensional mechanism design with limited information. In: Proceedings of the 13th ACM Conference on Electronic Commerce. EC' 12. ACM, New York, NY, pp. 162–179.

Bergemann, D., Shen, J., Xun, Y., Yeh, E., 2015. Nonlinear Pricing with Finite Information. Discussion paper. Yale University.

Blumenson, L., 1960. A derivation of n-dimensional spherical coordinates. Am. Math. Mon.

Blumrosen, L., Nisan, N., Segal, I., 2007. Auctions with severely bounded communication. J. Artif. Intell. Res. 28, 233–266.

Conway, J., Sloane, N., 1985. A lower bound on the average error of vector quantizers. IEEE Trans. Inf. Theory 31, 106–109.

Dereich, S., Vormoor, C., 2011. The high resolution vector quantization problem with Orlicz norm distortion. J. Theor. Probab. 24 (2), 517–544.

Gersho, A., Gray, R., 2007. Vector Quantization and Signal Compression. Kluwer Academic Press, Amsterdam.

Hoppe, H., Moldovanu, B., Ozdenoren, E., 2010. Coarse matching with incomplete information. Econ. Theory 47, 73–104.

Kos, N., 2012. Communication and efficiency in auctions. Games Econ. Behav. 75, 232–249.

Li, J., Chaddha, N., Gray, R.M., 1999. Asymptotic performance of vector quantizers with a perceptual distortion measure. IEEE Trans. Inf. Theory 45 (4), 1082–1091.

Lloyd, S., 1957. Least Squares Quantization in PCM. Bell Laboratories Technical Note.

Lloyd, S., 1982. Least square quantization in PCM. IEEE Trans. Inf. Theory 28, 127–135.

Lookabaugh, T., Gray, R., 1989a. High resolution quantization theory and the vector quantizer advantage. IEEE Trans. Inf. Theory 35, 1020–1033.

Lookabaugh, T.D., Gray, R.M., 1989b. High-resolution quantization theory and the vector quantizer advantage. IEEE Trans. Inf. Theory 35 (5), 1020–1033.

Madarasz, K., Prat, A., 2017. Sellers with misspecified models. Rev. Econ. Stud. 84, 790–815.

Maskin, E., Riley, J., 1984. Monopoly with incomplete information. Rand J. Econ. 15, 171–196.

Max, J., 1960. Quantizing for minimum distortion. IEEE Trans. Inf. Theory 6, 7–12.

McAfee, P., 2002. Coarse matching. Econometrica 70, 2025–2034.

Mussa, M., Rosen, S., 1978. Monopoly and product quality. J. Econ. Theory 18, 301–317.

Myerson, R., 1979. Incentive compatibility and the bargaining problem. Econometrica 47, 61–73.

Rochet, J., Chone, P., 1998. Ironing, sweeping and multidimensional screening. Econometrica 66, 783–826.

Rochet, J., Stole, L., 2003. The economics of multidimensional screening. In: Advances in Economics and Econometrics: Theory and Applications - Eight World Congress. Cambridge University Press, Cambridge, pp. 150–197.

Trushkin, A., 1982. Sufficient conditions for uniqueness of a locally optimal quantizere for a class of convex error weighting functions. IEEE Trans. Inf. Theory 28, 187–198.

Wilson, R., 1989. Efficient and competitive rationing. Econometrica 57 (1), 1–40.

Wilson, R., 1993. Nonlinear Pricing. Oxford University Press, Oxford.

Zador, P., 1982. Asymptotic quantization error of the continuous signals and the quantization dimension. IEEE Trans. Inf. Theory 28, 139–149.