Multi-dimensional Mechanism Design with Limited Information

DIRK BERGEMANN, Yale University
JI SHEN, London School of Economics and Political Science
YUN XU, Yale University
EDMUND YEH, Northeastern University

We analyze a nonlinear pricing model with limited information. Each buyer can purchase a large variety, d, of goods. His preference for each good is represented by a scalar and his preference over d goods is represented by a d-dimensional vector. The type space of each buyer is given by a compact subset of R_+^d with a continuum of possible types. By contrast, the seller is limited to offer a finite number M of d-dimensional choices.

We provide necessary conditions that the optimal finite menu of the social welfare maximizing problem has to satisfy. We establish an underlying connection to the theory of quantization and provide an estimate of the welfare loss resulting from the usage of the d-dimensional M-class menu. We show that the welfare loss converges to zero at a rate proportional to $d/M^{2/d}$.

We show that in higher dimensions, a significant reduction in the welfare loss arises from an optimal partition of the d-dimensional type space that takes advantage of the correlation among the d parameters.

Categories and Subject Descriptors: J.4 [**Social and Behavioral Sciences**]: Economics

General Terms: Economics, Theory

Additional Key Words and Phrases: Mechanism Design, Multi-Dimensional Private Information, Limited Information, Nonlinear Pricing, Quantization, Information Theory

1. INTRODUCTION

The primary focus in the theory of mechanism design has been to model and analyze the role of private information in economic environments. The optimal solution of mechanism design problem typically resolves trade-off between the socially efficient or revenue-maximizing allocation and the constraints imposed by the private information of the agents. However, when putting the theory of mechanism design into practice, other theoretically important and practically important issues come into consideration, in particular the cost of operating the mechanism. An important, but implicit, assumption in the overwhelming majority of earlier work is the assumption that the revelation of the information and the implementation of the associated allocation is realized with zero cost. The emphasis of the present contribution is to analyze a canonical mechanism design problem when it is costly to reveal or to transmit the private information.

In an earlier work ([Bergemann et al. 2012]), we analyzed the canonical nonlinear pricing model in which a seller offers a menu with a finite number of choices to a buyer

Dirk Bergemann acknowledges financial support through NSF Grant SES 0851200.
Author's addresses: D. Bergemann, Department of Economics, Yale University; J. Shen, Department of Finance, London School of Economics and Political Science; Y. Xu, Department of Electrical Engineering, Yale University; E. Yeh, Department of Electrical and Computer Engineering, Northeastern University.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
EC'12, June 4–8, 2012, Valencia, Spain. Copyright 2012 ACM 978-1-4503-1415-2/12/06...$10.00.

with a continuum of possible valuations. Within the linear-quadratic model (following a tradition established by the seminal papers in this area, e.g. [Mussa and Rosen 1978] and [Maskin and Riley 1984]), we establish a link between the classic screening model and the theory of quantization. In particular, we bound the loss that we incur from using discretized contracts, both in terms of the social welfare and the seller's expected revenue. A key insight, that we shall use in present context as well, is that we can view the private information (i.e. the individual taste parameter) as the source signal and his choice (quantity or quality available according to the menu) as the representation level. It then follows that the relationship between type and choice can be described in terms of the Lloyd-Max optimality conditions, a well-established result in the theory of quantization. In the welfare maximization problem, where the private information is either publicly observable or can be elicited by means of the Vickrey-Clarke-Groves mechanism, the central objective is to determine the socially optimal allocation. It then follows by the above insight that the total social welfare can be written as the mean square error between the source signal and the representation level. A similar technique can be applied to the revenue maximization problem by reformulating it as a welfare maximization problem, using the representation of the objective function by means of the virtual utility. In both situations, we show that a contract with n choices, an n-class contract, converges to its continuous counterpart at a rate proportional to $1/n^2$. It is worth mentioning that this is an exact result rather an asymptotic one, i.e., the result holds for any n, small or large.

In the present contribution, we apply this approach to a (particular class of) multi-dimensional screening problems where the consumer is facing a variety of goods and his preference over each good can be characterized by a scalar (so that his preference over all varieties is summarized by a vector). The one-dimensional Lloyd-Max conditions can be extended to their multi-dimensional counterparts, as represented vividly by the celebrated Voronoi diagram. Thus, even in the multi-dimensional environment, we can provide an upper bound on the convergence rate and show that it is consistent with its one-dimensional case if we quantize each dimension by a scalar quantizer separately. Given this, each dimension contributes the same amount and the overall loss is represented by the summation over all d independent scalar quantization.

Yet, we can considerably improve the upper bound by using vector quantization over the entire multi-dimensional type space. Here we rely on results from the theory of vector quantization and analyze the advantage of vector quantization over scalar quantization when d becomes sufficiently large. The total gain is shown to consist of three parts: (i) space-filling advantage, (ii) shape advantage and (iii) 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 significantly, due to the space-filling advantage.

In the past decade a number of notable contributions have analyzed nearly-optimal contracts in the context of communication constraints. In the context of the public good provision, [Ledyard and Palfrey 2002] find that simple voting rules can perform sufficiently close to the fully exact one in the presence of large populations. In a two-sided matching market, [McAfee 2002] and [Hoppe et al. 2010] compare the performance of coarse matching (binary segmentation of whole population on each side) versus an exact assortative matching scheme, with or without monetary transfers or information asymmetry. To determine the optimal rationing of service, [Wilson 1989] pioneered this literature by emphasizing the use of a finite number of priority classes. Importantly, these contributions confine their analysis to a one-dimensional space of private information. By contrast, we are explicitly focussing on the role of multi-dimensional private information.

Recently, the effects of limited communication have been investigated in auction environments, another canonical model in mechanism design. [Blumrosen et al. 2007] analyze limited communication, in a single-item independent-private-value environment, by assuming that the bidders, each endowed with a continuously distributed valuation can only use message spaces of finite cardinality. A noteworthy result is that the welfare optimizing protocols treat the ex ante symmetric agents asymmetrically, and recently [Kos 2011] provided some generalizations. [Bergemann and Pesendorfer 2007] analyze the joint design of optimal allocation and information structures in a single item auction. They establish that coarse partitions of the type space and asymmetry in the bidders' information structure are part of the optimal auction design. A number of related papers, including [Rothkopf and Harstad 1994], [Blumrosen and Feldman 2006], and [Blumrosen et al. 2007] show that the welfare loss incurred by limited communication in a single-parameter environment is of the order $O(1/n^2)$ where n is the number of choices available. Our paper, in the framework of nonlinear pricing, achieves a similar bound in the one-dimensional case and extends the convergence rate to the multi-dimensional case.

We should emphasize that the multi-dimensional screening problem does not represent a trivial extension of its one-dimensional counterpart. In many environments of interest, the preference of an individual agent cannot be summarized by a mere scalar but is more suitably represented as a vector. A real-life example would be a customer who has to make his 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] are two notable early contributions with explicit solutions to specific multi-dimensional screening problems. [Rochet and Chone 1998] developed a systematic approach, coined the dual approach, to a general class of environments and pointed to the prevalence of bunching (agents with different type profile making the same choices). We refer readers to [Rochet and Stole 2003] for a detailed survey of multi-dimensional screening problems.

Our analysis bypasses the issues related to an exact solution of the multi-dimensional screening problem. We estimate the welfare loss for any arbitrarily high-dimensional case with continuously distributed types. [Armstrong 1999] is the related to this issue. He established the asymptotic optimality of a single cost-based two-part tariff contract where all consumer surplus can be extracted as the number of varieties goes to infinity. The key assumption for his method is that the tastes are (almost) independently distributed across multiple products. By contrast, our contribution can accommodate any form of dependence among valuations of different products. In fact, the correlation among products implies that it is sub-optimal to price each commodity separately. The main focus of this paper is then to design and price a finite number of bundles, composed of a diversity of goods.

2. MODEL SETUP

2.1. Multi-Product Model

We consider a monopolistic firm facing a continuum of consumers and providing d heterogeneous goods. Each consumer's preferences over these goods is characterized by a d-dimensional vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d) \in R_+^d$, called the consumer's type vector, where for $1 \leq l \leq d$, θ_l represents his preference for good l. Let $\Theta \subseteq R_+^d$ denote a compact d-dimensional type space. We assume that the joint probability distribution of $\boldsymbol{\theta}$, denoted by $F(\boldsymbol{\theta})$, is commonly known. If a type $\boldsymbol{\theta}$ consumes the bundle of goods with quantity (or quality) vector $\boldsymbol{q} = (q_1, \ldots, q_d)$ by transferring a payment $t(\boldsymbol{q}) = \sum_{l=1}^{d} t_l(q_l)$, where

$t_l(q_l)$ is the payment for good l with quantity (or quality) q_l. Assuming the consumer has linear utility, his net utility is:

$$u(\boldsymbol{\theta}, \boldsymbol{q}) - t(\boldsymbol{q}) = \boldsymbol{\theta}^T \Phi \boldsymbol{q} - t(\boldsymbol{q})$$

where the superscript T represents the transpose of the vector and $\Phi = (\phi_{ij})_{d\times d}$ is a $d \times d$ matrix which captures the interactions among different goods. We assume that $\phi_{ii} > 0$ for all i so that:

$$\frac{\partial^2 u}{\partial \theta_i \partial q_i} > 0.$$

It turns out that no further assumptions, such as invertibility, symmetry or positive-definiteness of Φ, are needed for the analysis which follows.

The firm incurs a quadratic cost $c(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^T \Sigma \boldsymbol{q}$ by providing the bundle $\boldsymbol{q}$. Here, $\Sigma = (\sigma_{ij})_{d\times d}$ is a $d \times d$ symmetric positive-definite matrix which characterizes the interactions in the production of multiple products. All of its diagonal elements must be positive: $\sigma_{ii} > 0$ for all i. If producing good i raises (reduces) the marginal cost of producing good j, then we set $\sigma_{ij} = \sigma_{ji} > (<) 0$ and call these two goods substitutes (complements). If $\sigma_{ij} = \sigma_{ji} = 0$, the technologies of producing good i and j are independent.

2.2. Multi-Agent Model

An alternative to the multi-product model is a multi-agent model, where the firm serves one product to d heterogeneous customers with one-dimensional linear utilities. In this case, $\boldsymbol{\theta}$ can be viewed as the vector of all customers' tastes for the one product. Customer i's utility from consuming quantity q_i is $\phi_{ii}\theta_i q_i$, and his utility is also affected by others' consumption. If customer j consumes quantity q_j, he imposes an externality on customer i by raising i's utility by $\phi_{ij}\theta_i q_j$. If $\phi_{ij} > (<) 0$, the externality is positive (negative). If $\phi_{ij} = 0$, then customer j does not affect i's utility. Thus, ϕ_{ij}/ϕ_{ii} measures the strength of the externality imposed by customer j relative to customer i. Note that Φ need not to be symmetric. In this case, $\mathbb{E}\left[\boldsymbol{\theta}^T \Phi \boldsymbol{q}\right]$ quantifies the total consumers' surplus.

Since the multi-product and the multi-agent model are mathematically equivalent, we will focus on the multi-product interpretation in this paper.

3. WELFARE MAXIMIZATION

In the absence of information constraints, $M = \infty$, the social welfare is determined by maximizing

$$SW(\infty) = \mathbb{E}_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}^T \Phi \boldsymbol{q} - \frac{1}{2}\boldsymbol{q}^T \Sigma \boldsymbol{q}\right].$$

This represents a natural extension of the linear-quadratic model to the multi-dimensional case. We say the social welfare $SW(\infty)$ has a standard form if $\Phi = \Sigma = I_d$ (the identity matrix of size d). In fact, we show that we can always transform the social welfare into the standard form.

We can diagonalize the positive-definite matrix Σ: $\Sigma = P^T \Lambda P$, where $\Lambda = diag(\lambda_1, ..., \lambda_d)$, $\lambda_i > 0$ 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 $\Lambda^i = diag\left(\lambda_1^i, ..., \lambda_d^i\right)$, $i = \pm\frac{1}{2}$. Then it is easy to show $A^T B = \Phi$ and $B^T B = \Sigma$. If we introduce the new type and quantity (or quality) vectors: $\hat{\boldsymbol{\theta}} = A\boldsymbol{\theta}$ and $\hat{\boldsymbol{q}} = B\boldsymbol{q}$, then the utility and cost function

can be written in the standard form in terms of $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{q}}$:

$$\begin{aligned} u(\boldsymbol{\theta}, \boldsymbol{q}) &= \boldsymbol{\theta}^T \Phi \boldsymbol{q} = \boldsymbol{\theta}^T A^T B \boldsymbol{q} = \hat{\boldsymbol{\theta}}^T \hat{\boldsymbol{q}}, \\ c(\boldsymbol{q}) &= \frac{1}{2} \boldsymbol{q}^T \Sigma \boldsymbol{q} = \frac{1}{2} \boldsymbol{q}^T B^T B \boldsymbol{q} = \frac{1}{2} \hat{\boldsymbol{q}}^T \hat{\boldsymbol{q}}. \end{aligned}$$

Thus, without loss of generality, we focus on the social welfare in the standard form (i.e., assuming that $\Phi = \Sigma = I_d$).

When the consumer's type vector is publicly known, it is socially optimal to provide a production vector equal to the type vector for every consumer: $\boldsymbol{q}^*(\boldsymbol{\theta}) = \boldsymbol{\theta}$. The maximum social welfare equals:

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

By contrast, we assume that, due to information constraints, the customer faces a discretized contract, i.e., a finite number M of pairs $\{(\boldsymbol{q}_m, t_m)\}_{m=1}^M$, where $\boldsymbol{q}_m = (q_{m,1}, \ldots, q_{m,d})$ is the m-th quantity (or quality) vector of goods provided by the seller, $t_{m,l}$ is the price paid for $q_{m,l}$, and $t_m = \sum_{l=1}^d t_{m,l}$ is the total price charged for $\boldsymbol{q}_m$. Such a discretized contract or menu is called a d-dimensional M-class contract.

Let $\{B_m\}_{m=1}^M$ represent a partition of the consumer's d-dimensional type space Θ, i.e., $B_i \cap B_j = \emptyset$ if $i \neq j$, and $\cup_{m=1}^M B_m = \Theta$. A consumer with type vector $\boldsymbol{\theta} \in B_m$ will choose the quantity (or quality) vector $\boldsymbol{q}(\boldsymbol{\theta}) = \boldsymbol{q}_m$ and pay the total price $t(\boldsymbol{q}(\boldsymbol{\theta})) = t_m$.

In this case, we choose the M-class contract $\{B_m, \boldsymbol{q}_m\}_{m=1}^M$ so as to maximize the expected social welfare:

$$\max_{\{B_m, \boldsymbol{q}_m\}_{m=1}^M \in L_F} SW(M) = \max_{\{B_m, \boldsymbol{q}_m\}_{m=1}^M \in L_F} \mathbb{E}_{\boldsymbol{\theta}} \left[\boldsymbol{\theta}^T \boldsymbol{q} - \frac{1}{2} \boldsymbol{q}^T \boldsymbol{q} \right], \tag{2}$$

where the set of all M-class contract for a given distribution F is given by:

$$L_F = \left\{ \{B_m, \boldsymbol{q}_m\}_{m=1}^M : B_i \cap B_j = \emptyset \text{ if } i \neq j, \text{ and } \cup_{m=1}^M B_m = \Theta \right\}.$$

3.1. Connection to Vector Quantization

When the joint probability distribution of $\boldsymbol{\theta}$ is known, maximizing the social welfare is equivalent to minimizing:

$$\mathbb{E}_{\boldsymbol{\theta}} \left[\boldsymbol{\theta}^T \boldsymbol{\theta} - 2\boldsymbol{\theta}^T \boldsymbol{q} + \boldsymbol{q}^T \boldsymbol{q} \right] = \mathbb{E}_{\boldsymbol{\theta}} \left[(\boldsymbol{\theta} - \boldsymbol{q})^T (\boldsymbol{\theta} - \boldsymbol{q}) \right] = \mathbb{E}_{\boldsymbol{\theta}} \left[\|\boldsymbol{\theta} - \boldsymbol{q}\|^2 \right]$$

where $\|\cdot\|$ is the Euclidean norm. In the appendix, we show that if we view $\boldsymbol{\theta}$ as the input and $\boldsymbol{q}_m$ as the representation point of $\boldsymbol{\theta}$ in the region B_m, then this becomes the d-dimensional M-region vector quantization problem, where the partition $\{B_m,\}_{m=1}^M$ and the set of representation points $\{\boldsymbol{q}_m\}_{m=1}^M$ are chosen to minimize the mean square error (MSE):

$$\min_{\{B_m, \boldsymbol{q}_m\}_{m=1}^M \in L_F} MSE(M) = \min_{\{B_m, \boldsymbol{q}_m\}_{m=1}^M \in L_F} \mathbb{E}_{\boldsymbol{\theta}} \left[\|\boldsymbol{\theta} - \boldsymbol{q}\|^2 \right]. \tag{3}$$

In this case, $\{B_m, \boldsymbol{q}_m\}_{m=1}^M$ can be viewed as a d-dimensional M-region vector quantizer. Therefore, the optimal solution must satisfy the following Lloyd-Max conditions for vector quantization, see [Gersho and Gray 1992].

THEOREM 3.1. *(Lloyd-Max conditions for vector quantization) Consider the vector quantization problem* (3). *The optimal partition* $\{B_m^*\}_{m=1}^M$ *of the type space and the set*

of representation points $\{\boldsymbol{q}_m^*\}_{m=1}^M$ *must satisfy:*

$$\boldsymbol{q}_m^* = \mathbb{E}_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}|\boldsymbol{\theta} \in B_m^*\right], \tag{4}$$

$$B_m^* = \{\boldsymbol{\theta} \in S : \|\boldsymbol{\theta} - \boldsymbol{q}_m^*\| \leq \|\boldsymbol{\theta} - \boldsymbol{q}_l^*\| \textit{ for all } l\}. \tag{5}$$

In other words, $\boldsymbol{q}_m^*$ is chosen as the conditional mean of $\boldsymbol{\theta}$ given that $\boldsymbol{\theta}$ lies in the region B_m^*, and $\{B_m^*\}_{m=1}^M$ is chosen as a Voronoi partition (see Definition 3.10) with respect to $\{\boldsymbol{q}_m^*\}_{m=1}^M$.

We now consider how the optimal d-dimensional M-class contract can approximate the performance of the optimal continuous contract for a general joint distribution function F. Specifically, we quantify the welfare loss in terms of the distribution function F, the number of classes M, and the dimension d.

Definition 3.2. For any joint distribution function F, the welfare loss induced by the optimal d-dimensional M-class contract compared with the optimal continuous contract is defined by:

$$\Delta\left(F; M; d\right) \equiv SW^*\left(\infty\right) - SW^*\left(M\right) = \inf_{\{B_m, \boldsymbol{q}_m\}_{m=1}^M \in L_F} \left[SW^*\left(\infty\right) - SW\left(M\right)\right].$$

We are interested in the worst-case behavior of the welfare loss over all joint distributions over a d-dimensional support set with positive and finite volume. Without loss of generality, we may assume the type space $\Theta \subseteq [0,1]^d$.[1] Let $\mathcal{F}$ be the set of all joint distribution functions in type space $\Theta \subseteq [0,1]^d$. Our main task is to quantify the worst-case behavior of $\Delta\left(F; M; d\right)$ over all distributions $F \in \mathcal{F}$.

Definition 3.3. The maximum welfare loss induced by the optimal d-dimensional M-class contract over all $F \in \mathcal{F}$ is defined by:

$$\Delta\left(M; d\right) \equiv \sup_{F \in \mathcal{F}} \Delta\left(F; M; d\right)$$

3.2. Welfare Loss of One-Dimensional M-Class Contract

Before delving into higher dimensions, we review some basic results of the one-dimensional case as a reference for comparison. More detailed discussion, together with rigorous proofs, can be found in our earlier work, [Bergemann et al. 2012]. Note that when $d = 1$, q_m is a scalar and $B_m = [\theta_{m-1}, \theta_m)$ is an interval. The associated Lloyd-Max conditions in (4) and (5) now reduce to:

$$q_m^* = \mathbb{E}\left[\theta|\theta \in \left[\theta_{m-1}^*, \theta_m^*\right)\right],$$

and

$$\theta_m^* = \frac{q_m^* + q_{m+1}^*}{2}.$$

That is, q_m^* is the conditional mean in the interval B_m, and θ_m^*, which separates two neighboring intervals B_m and B_{m+1}, is the arithmetic average of q_m^* and q_{m+1}^*. In [Bergemann et al. 2012] we show that the convergence rate of the welfare loss induced by the optimal one-dimensional M-class contract is of order $1/M^2$. Specifically,

$$\Delta\left(F; M; 1\right) \leq \frac{1}{8M^2}$$

[1]For any set $\Theta \subseteq R_+^d$ with positive and finite volume, let $b = \sup_{\boldsymbol{\theta} \in S} \|\boldsymbol{\theta}\|$. Then $0 < b < \infty$. We normalize all type vectors in Θ by the factor b so that $\Theta \subseteq [0,1]^d$.

for all F defined on $[0,1]$, and $M \geq 1$. The maximum welfare loss $\Delta(M;1)$ is upper bounded by $\frac{1}{8M^2}$, and lower bounded by $\frac{1}{24M^2}$.

[Wilson 1989] arrived at a related result by using a different technique. He implicitly quantized the distribution function of θ uniformly, and then expanded the social welfare by the Taylor series around zero up to the order of $1/M^2$. By contrast, we use quantization theory to solve the problem directly, by choosing a scalar quantizer $\{(q'_m, \theta'_m)\}_{m=1}^M$ in the type space with q'_m consistent with the Lloyd-Max conditions and θ'_m being equally distributed. We use such a quantizer to provide an upper bound on the welfare loss. Our quantization approach is straightforward, and has the significant advantage that it extends naturally, via vector quantization, to the multi-dimensional case. In the following sections, we established that our earlier results in one dimension can be viewed as special case of a general quantization approach in higher dimensions.

3.3. Welfare Loss of d-Dimensional M-Class Contract

In this section, we provide our main results on how the d-dimensional M-class contract can approximate the performance of the optimal continuous contract for a general joint distribution on the type space. We estimate the convergence rate of the welfare loss induced by discretized contracts as the number of classes tends to infinity.

For any $F \in \mathcal{F}$, we have

$$SW(M) = \mathbb{E}_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}^T\boldsymbol{q} - \frac{1}{2}\boldsymbol{q}^T\boldsymbol{q}\right] = \frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}^T\boldsymbol{\theta}\right] - \frac{1}{2}MSE(M).$$

Recall that the optimal continuous contract offers the social welfare $SW^*(\infty) = \frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}^T\boldsymbol{\theta}\right]$, and thus

$$SW^*(\infty) - SW(M) = \frac{1}{2}MSE(M), \tag{6}$$

where

$$MSE(M) = \mathbb{E}_{\boldsymbol{\theta}}\left[\|\boldsymbol{\theta} - \boldsymbol{q}\|^2\right] = \sum_{m=1}^{M}\int_{B_m}\|\boldsymbol{\theta} - \boldsymbol{q}_m\|^2 dF(\boldsymbol{\theta}).$$

Therefore, we have

$$\begin{aligned}\Delta(F;M;d) &= \inf_{\{B_m,\boldsymbol{q}_m\}_{m=1}^M \in L_F}\left[SW^*(\infty) - SW(M)\right] \\ &= \inf_{\{B_m,\boldsymbol{q}_m\}_{m=1}^M \in L_F}\frac{1}{2}\sum_{m=1}^{M}\int_{B_m}\|\boldsymbol{\theta} - \boldsymbol{q}_m\|^2 dF(\boldsymbol{\theta}),\end{aligned} \tag{7}$$

and correspondingly:

$$\Delta(M;d) = \sup_{F\in\mathcal{F}}\inf_{\{B_m,\boldsymbol{q}_m\}_{m=1}^M \in L_F}\frac{1}{2}\sum_{m=1}^{M}\int_{B_m}\|\boldsymbol{\theta} - \boldsymbol{q}_m\|^2 dF(\boldsymbol{\theta}). \tag{8}$$

PROPOSITION 3.4. *For any $F \in \mathcal{F}$, and any $M \geq 1$, $d \geq 1$, $\Delta(F;M;d) \leq \frac{d}{2M^{2/d}}$.*

Proof. We can construct a vector quantizer with K^d representation points by using the same scalar quantizer with K representation points in each of the d dimensions. It is easy to see that in this case, we simply choose the set of regions as orthotopes, defined as the Cartesian product of intervals in d dimensions. Such a vector quantizer is called the d-dimensional repeated scalar quantizer. We will use it to prove the upper bound. Let $K = \lfloor M^{1/d} \rfloor$. For any given $F \in \mathcal{F}$, consider the K-level scalar quantizer

$\{A_k, r_k\}_{k=1}^K$ for a random variable $X \in [0,1]$ (see the Appendix for an introduction to the K-level scalar quantizer) as follows:

$$\begin{aligned} A_k &= \left[\frac{k-1}{K}, \frac{k}{K}\right), \\ r_k &= \mathbb{E}_X\left[X | X \in A_k\right], \end{aligned}$$

where $\{A_k\}_{k=1}^K$ forms the uniform grid on $[0,1]$, and r_k is the conditional mean on A_k. Construct the corresponding repeated scalar quantizer $\{B'_m, \boldsymbol{q}'_m\}_{m=1}^{K^d}$ over the d-dimensional type space $[0,1]^d$:

$$\begin{aligned} \{B'_m\}_{m=1}^{K^d} &= \{A_{k_1} \times \ldots \times A_{k_d} : k_l \in \{1, \ldots, K\}, l = 1, \ldots, d\}, \\ \{\boldsymbol{q}'_m\}_{m=1}^{K^d} &= \{(r_{k_1}, \ldots, r_{k_d}) : k_l \in \{1, \ldots, K\}, l = 1, \ldots, d\}. \end{aligned}$$

Note that $B'_i \cap B'_j = \emptyset$ if $i \neq j$, and $\cup_{m=1}^{K^d} B'_m = \Theta = [0,1]^d$, and thus $\{B'_m, \boldsymbol{q}'_m\}_{m=1}^{K^d} \in L_F$. Since $M \geq K^d$, and since $\Delta(\cdot; M; \cdot)$ is a decreasing function of M according to Definition 3.2,

$$\begin{aligned} \Delta(F; M; d) &\leq \Delta\left(F; K^d; d\right) \leq \frac{1}{2} \sum_{m=1}^{K^d} \int_{B'_m} \|\boldsymbol{\theta} - \boldsymbol{q}'_m\|^2 \, dF(\boldsymbol{\theta}) \\ &= \frac{1}{2} \sum_{l=1}^{d} \left\{ \sum_{m=1}^{K^d} \int_{B'_m} \left(\theta_l - q'_{m,l}\right)^2 dF(\boldsymbol{\theta}) \right\}. \end{aligned}$$

Based on the construction of $\{B'_m\}_{m=1}^{K^d}$ and $\{\boldsymbol{q}'_m\}_{m=1}^{K^d}$, we have

$$\begin{aligned} &\sum_{m=1}^{K^d} \int_{B'_m} \left(\theta_l - q'_{m,l}\right)^2 dF(\boldsymbol{\theta}) \\ &= \sum_{k_1=1}^{K} \cdots \sum_{k_d=1}^{K} \int_{A_{k_1} \times \ldots \times A_{k_d}} \left(\theta_l - r_{k_l}\right)^2 dF(\boldsymbol{\theta}) \\ &= \sum_{k_1=1}^{K} \cdots \sum_{k_d=1}^{K} \int_{A_{k_l}} \left(\theta_l - r_{k_l}\right)^2 \left\{ \int_{A_{-k_l}} dF_{-l}(\boldsymbol{\theta}_{-l} | \theta_l) \right\} dF_l(\theta_l) \\ &= \sum_{k_l=1}^{K} \int_{A_{k_l}} \left(\theta_l - r_{k_l}\right)^2 \left\{ \sum_{k_1=1}^{K} \cdots \sum_{k_{l-1}=1}^{K} \sum_{k_{l+1}=1}^{K} \cdots \sum_{k_d=1}^{K} \int_{A_{-k_l}} dF_{-l}(\boldsymbol{\theta}_{-l} | \theta_l) \right\} dF_l(\theta_l), \end{aligned}$$

where

$$\begin{aligned} A_{-k_l} &= A_{k_1} \times \ldots \times A_{k_{l-1}} \times A_{k_{l+1}} \ldots \times A_{k_d}, \\ \boldsymbol{\theta}_{-l} &= (\theta_1, \ldots, \theta_{l-1}, \theta_{l+1}, \ldots, \theta_d), \end{aligned}$$

and $F_{-l}(\cdot|\cdot)$ is the conditional distribution function of $\boldsymbol{\theta}_{-l}$ given θ_l.

Note that

$$\sum_{k_1=1}^{K} \cdots \sum_{k_{l-1}=1}^{K} \sum_{k_{l+1}=1}^{K} \cdots \sum_{k_d=1}^{K} \int_{A_{-k_l}} dF_{-l}(\boldsymbol{\theta}_{-l} | \theta_l) = 1,$$

for any θ_l. Then

$$\sum_{m=1}^{K^d} \int_{B'_m} \left(\theta_l - q'_{m,l}\right)^2 dF(\boldsymbol{\theta}) = \sum_{k_l=1}^{K} \int_{A_{k_l}} \left(\theta_l - r_{k_l}\right)^2 dF_l(\theta_l).$$

Therefore,

$$\Delta(F; M; d) \leq \frac{1}{2} \sum_{l=1}^{d} \left\{ \sum_{k_l=1}^{K} \int_{A_{k_l}} \left(\theta_l - r_{k_l}\right)^2 dF_l(\theta_l) \right\}.$$

Since $A_k = \left[\frac{k-1}{K}, \frac{k}{K}\right)$ and $r_{k_l} = \mathbb{E}_{\theta_l}\left[\theta_l | \theta_l \in A_{k_l}\right]$, based on the previous analysis for one-dimensional case (see Proposition 2 in [Bergemann et al. 2012]),

$$\begin{aligned}
\Delta(F; M; d) &\leq \frac{1}{2} \sum_{l=1}^{d} \left\{ \sum_{k_l=1}^{K} \left[F_l\left(\frac{k_l}{K}\right) - F_l\left(\frac{k_l - 1}{K}\right) \right] var\left(\theta_l | \theta_l \in A_{k_l}\right) \right\} \\
&\leq \frac{1}{2} \sum_{l=1}^{d} \frac{1}{4K^2} \leq \frac{1}{2} \sum_{l=1}^{d} \frac{1}{(K+1)^2} \leq \frac{d}{2M^{2/d}},
\end{aligned}$$

because $K = \left\lfloor M^{1/d} \right\rfloor \geq 1$, and $2K \geq K + 1 \geq M^{1/d}$.■

In order to obtain the convergence rate for the maximum welfare loss, we establish the following lemma.

LEMMA 3.5. *Suppose the elements of the type vector, $\theta_1, \ldots, \theta_d$ are i.i.d. uniform random variables, i.e., $F(\theta_1, \ldots, \theta_d) = \prod_{l=1}^{d} U(\theta_l)$ where U is the uniform distribution function on $[0, 1]$. If $\hat{K} = \left\lceil M^{1/d} \right\rceil$ is sufficiently large, then $\Delta(F; M; d) \geq \frac{1}{8\pi e} \frac{d}{M^{2/d}}$ for any $d \geq 1$.*

We will prove this lemma using the analysis in Section 3.4. Proposition 3.4 provides a general upper bound on the convergence rate for any joint distribution $F \in \mathcal{F}$, and Lemma 3.5 provides a lower bound on the convergence rate for the i.i.d. uniform distribution, which can also be viewed as a lower bound on $\Delta(M; d)$. Hence, we have the following result.

PROPOSITION 3.6. *If $\hat{K} = \left\lceil M^{1/d} \right\rceil$ is sufficiently large, then $\frac{1}{8\pi e} \frac{d}{M^{2/d}} \leq \Delta(M; d) \leq \frac{1}{2} \frac{d}{M^{2/d}}$ for any $d \geq 1$.*

Hence, the maximal welfare loss induced by the d-dimensional M-class contract converges to zero at a rate proportional to $\frac{d}{M^{2/d}}$.

We use the scalar quantization repeatedly to obtain the upper bound of the convergence rate for general distributions in Proposition 3.4. However, in repeated scalar quantization, we simply partition the space with hyperrectangles, orthotopes, and treat each dimension independently, leading to a possibly weak bound. A natural question arises as to whether we can reduce the convergence rate if we use the optimal vector quantization. In fact, in higher dimensions ($d > 1$), a significant reduction of the welfare loss can be obtained by using more subtle vector quantization methods which allows us to minimize the loss in a manner that is impossible in a single dimension. For instance, we can choose quantization regions other than orthotopes, and we can take advantage of the dependence among the different entries of the type vector. This reduction in the welfare loss represents the advantage of vector quantization advantage and

is the main reason why we bundle the consumer's preferences over d goods across the d-dimensional type vector, instead of viewing them separately as d (one-dimensional) types.

3.4. Advantages of Vector Quantization

To simplify our analysis, we assume in this section that the elements of the type vector, $\theta_1, \ldots, \theta_d$, are identically, but not necessarily independently distributed. Let F and f denote the joint distribution and joint density respectively, and let $\hat{F}$ and $\hat{f}$ denote the marginal distribution and marginal density.

We consider two distinct scenarios. In the first scenario, we ignore the dependence among the consumer's preferences over d goods as if they were d independent scalar (one-dimensional) types $\theta_1, \ldots, \theta_d$. Since in this section, $\theta_1, \ldots, \theta_d$ are assumed to be identically distributed according to the marginal distribution function $\hat{F}$, the seller will offer d optimal one-dimensional K-class contracts which are identical and independent. Each contract will result in the same welfare loss $\Delta\left(\hat{F}; K; 1\right)$. In this case, the welfare maximization problem can be viewed as a scalar quantization problem. In the second scenario, we view the consumer's preferences over d goods as a d-dimensional type vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. In this case, the seller offers a d-dimensional K^d-class contract for the type vector $\boldsymbol{\theta}$ with the joint distribution function F. This contract will result in the total welfare loss $\Delta\left(F; K^d; d\right)$ over d dimensions, or equivalently, the average welfare loss $\frac{1}{d}\Delta\left(F; K^d; d\right)$. Recall in this case, the welfare maximization problem can be viewed as a vector quantization problem. To determine the vector quantization advantage, we can compare the average welfare loss induced by the optimal d-dimensional K^d-class contract with the welfare loss induced by the one-dimensional K-class contract.

Definition 3.7. For any given joint distribution F and its marginal $\hat{F}$, the vector quantization advantage for the social welfare G_{SW} in d dimensions is defined as the ratio of the welfare loss induced by the optimal K-class contract to the average welfare loss over d dimensions induced by the optimal d-dimensional K^d-class contract:

$$G_{SW} \equiv \frac{\Delta\left(\hat{F}; K; 1\right)}{\frac{1}{d}\Delta\left(F; K^d; d\right)} \tag{9}$$

From the above definitions, we can see the larger G_{SW} is, the more we gain from using vector quantization. For sufficiently large K, [Lookabaugh and Gray 1989] decompose the gain into three categories as follows.

THEOREM 3.8. *[Lookabaugh and Gray 1989] If the number of regions per dimension K becomes sufficiently large, then the quantization advantage for the social welfare can be decomposed into three factors:*

$$G_{SW} \approx SF\left(d\right) \times S\left(\hat{f}, d\right) \times DP\left(\hat{f}, f, d\right)$$

where $SF\left(d\right) \geq 1$, $S\left(\hat{f}, d\right) \geq 1$, and $DP\left(\hat{f}, f, d\right) \geq 1$ are called the space-filling advantage, shape advantage and dependence advantage, given by (11), (15) *and* (16), *respectively.*

Space-filling Advantage. As mentioned before, we have the freedom to select more complex region shapes besides orthotopes in higher dimensions ($d > 1$). This leads to the space-filling advantage $SF\left(d\right)$. Unlike the shape and dependence advantages, the space-filling advantage is a function only of the dimension, and provides the same gain

for all distributions with the same dimension. To better understand this advantage, we first introduce the following concepts.

Definition 3.9. A convex polytope H is said to be a space partition polytope if R^d can be partitioned by using the translated and rotated copies of H.

Definition 3.10. A Voronoi partition with respect to a set of points $X = \{x_1, x_2, \ldots\}$ is a partition whose regions are nearest-neighbor regions with respect to X, i.e., a point x is in the region belonging to x_i if $\|x - x_i\| \leq \|x - x_j\|$ for all j, where $\|\cdot\|$ is the Euclidean norm.

Definition 3.11. The geometric centroid of a convex polytope H is defined as

$$\tilde{x}(H) = \underset{y}{argmin} \int_H \|x - y\|^2 \, dx$$

Definition 3.12. An admissible polytope is a space partition polytope that can generate a Voronoi partition of R^d with respect to the set of geometric centroids of the regions in the Voronoi partition.

Definition 3.13. The normalized inertia of a polytope H is defined as

$$I(H) = \frac{\int_H \|x - \tilde{x}(H)\|^2 \, dx}{[V(H)]^{1+2/d}}$$

where $\tilde{x}$ is the geometric centroid of H, and $V(H)$ is its d-dimensional volume.

Definition 3.14. The coefficient of the (d-dimensional) optimal vector quantization is defined as

$$C(d) = \frac{1}{d} \inf_{H_d \in \mathcal{H}_d} I(H_d) = \frac{1}{d} \inf_{H_d \in \mathcal{H}_d} \frac{\int_{H_d} \|x - \tilde{x}(H_d)\|^2 \, dx}{[V(H_d)]^{1+2/d}} \tag{10}$$

where $\mathcal{H}_d$ is the set of all admissible polytopes in R^d.

Definition 3.15. The optimal admissible polytope is an admissible polytope which has the minimum inertia of all admissible polytopes, i.e., attains the coefficient of the optimal vector quantization.

[Lookabaugh and Gray 1989] showed that the space-filling advantage can be written as

$$SF(d) = \frac{C(1)}{C(d)} = \frac{1}{12 \times C(d)}. \tag{11}$$

Example 3.16. (1) For $d = 1$, the optimal admissible polytope is trivially the interval, so it is easy to calculate $C(1) = \frac{1}{12}$, and the space-filling advantage $SF(1) = 1$. Thus, there is no space-filling advantage for one dimensional space, where we can use only scalar quantization.

(2) For $d = 2$, we can show the equilateral triangle, the rectangle, and the regular hexagon are all admissible polytopes. Furthermore, the infimum in (10) is achieved when the regular hexagon is used, yielding $C(2) = \frac{5}{36\sqrt{3}}$, and the space-filling advantage $SF(2) = \frac{3\sqrt{3}}{5} \approx 1.0392$ [Gersho 1979]. In other words, in a two-dimensional space, even if we consider only the space-filling advantage, the welfare loss can be reduced by $1 - 1.0392^{-1} \approx 3.77\%$ by choosing the partition based on a set of regular hexagons, instead of a partition based on rectangles, as in the repeated scalar quantization.

To see intuitively why hexagons are better than rectangles in two dimensions, consider the i.i.d. uniform distribution. Suppose we use the same number of hexagons or rectangles to partition the space, each of which has the same area. Note that when the distances between points on the boundary of a region to its centroid are more equalized, as in the hexagon, the MSE in two dimensions becomes lower. This is because the MSE is a convex function of the distance between the boundary points and the centroid.

For $d \geq 3$, it is quite hard to find the optimal admissible polytope. However, it is a classic result that the d-dimensional sphere has smaller normalized inertia than any d-dimensional convex polytope. Therefore, if the sphere were an admissible polytope, the infimum in (10) would be achieved. Unfortunately, spheres cannot be used to cover the space. However, a lower bound on $C(d)$, or equivalently an upper bound on $SF(d)$ can be obtained by using the sphere [Gersho 1979]:

$$SF(d) \leq SF_U(d) = \frac{d+2}{12}(V_d)^{2/d}, \tag{12}$$

where V_d is the volume of a unit sphere in d-dimensional Euclidean space.

[Zador 1982] developed an upper bound on $C(d)$, or equivalently a lower bound on $SF(d)$ using random quantization where the representation points are picked at random, and the partition is a Voronoi partition with respect to the set of representation points:

$$SF(d) \geq SF_L(d) = \frac{d}{12\Gamma(1+2/d)}(V_d)^{2/d}. \tag{13}$$

[Conway and Sloane 1985] further showed that

$$1 \leq SF_L(d) \leq SF_U(d) \leq \frac{\pi e}{6},$$

for all $d \geq 1$, and

$$\lim_{d\to\infty} SF_L(d) = \lim_{d\to\infty} SF_U(d) = \lim_{d\to\infty} SF(d) = \frac{\pi e}{6}. \tag{14}$$

The above result indicates that we can choose admissible polytopes which are closer geometrically to the sphere as the dimension d becomes larger, and the optimal admissible polytope indeed approaches the sphere in infinite dimensional space, with the space-filling advantage $SF(d)$ asymptotically approaching $\frac{\pi e}{6} \approx 1.423$.

From (11), we can see that the space-filling advantage depends only on the coefficient of vector quantization, and hence by (10), only on the efficiency with which admissible polytopes can fill the space. Specifically, it does not depend on the probability distribution of the type or the dependence among the elements of the type vector.

Although the set of optimal admissible polytopes and their centroids determine the space-filling advantage, they do not generate the optimal vector quantizer in general. Recall that the optimal quantizer must satisfy the Lloyd-Max conditions and are affected by the distribution of the type and the dependence among the elements of the type. These effects are captured by the shape and dependence advantages. For the i.i.d. uniform distribution, however, optimal admissible polytopes and their centroids do form the optimal vector quantizer because the entire gain of vector quantization is captured by the space-filling advantage.

Shape Advantage. [Lookabaugh and Gray 1989] showed that the shape advantage can be written as

$$S\left(\hat{f},d\right)=\frac{\left[\int\left(\hat{f}\left(\theta_l\right)\right)^{1/3}d\theta_l\right]^3}{\left[\int\left(\hat{f}\left(\theta_l\right)\right)^{d/d+2}d\theta_l\right]^{d+2}}. \tag{15}$$

For any given dimension, $S\left(\hat{f},d\right)$ depends solely on the shape of the marginal density function $\hat{f}$ and does not depend on how the random vector θ is scaled (see the following example of the Gaussian density). [Royden 1968] proved that $S\left(\hat{f},d\right)\geq 1$ for all $\hat{f}$ and $d\geq 1$. We consider two examples.

Example 3.17. (1) Suppose the marginal density is uniform on $[0,1]$. Then it is easy to calculate $S\left(\hat{f},d\right)=1$ for all $d\geq 1$. In other words, the vector quantizer cannot provide any shape advantage for the uniform distribution.

(2) Suppose the marginal density is zero mean Gaussian with variance σ^2, i.e., $\hat{f}\left(\theta_l\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\theta_l^2/2\sigma^2\right)$. Then we can calculate $S\left(\hat{f},d\right)=\sqrt{27}/\left(\frac{d+2}{d}\right)^{\frac{d+2}{2}}$ which is independent of the variance σ^2 [Lookabaugh and Gray 1989]. It is easy to show $S\left(\hat{f},1\right)=1$, $S\left(\hat{f},d+1\right)\geq S\left(\hat{f},d\right)$ for all $d\geq 1$, and $\lim_{d\to\infty}S\left(\hat{f},d\right)=\sqrt{27}/e\approx 1.912$.

Dependence Advantage. [Lookabaugh and Gray 1989] showed that the dependence advantage can be written as

$$DP\left(\hat{f},f,d\right)=\frac{\left[\int\left(\hat{f}\left(\theta_l\right)\right)^{d/d+2}d\theta_l\right]^{d+2}}{\left[\int\cdots\int\left(f\left(\theta_1,\ldots,\theta_d\right)\right)^{d/d+2}d\theta_1\ldots d\theta_d\right]^{(d+2)/d}}. \tag{16}$$

Given the dimension, $DP\left(\hat{f},f,d\right)$ depends on the joint density function and its marginal density function, and thus implicitly on the dependence among $\theta_1,\ldots,\theta_d$. It is easy to show $DP\left(\hat{f},f,d\right)=1$ if $\theta_1,\ldots,\theta_d$ are i.i.d. random variables, i.e.,

$$f\left(\theta_1,\ldots,\theta_d\right)=\prod_{l=1}^{d}\hat{f}\left(\theta_l\right).$$

In this case, the gain over the scalar quantizer is entirely attributed to the space-filling and shape advantages. In other cases, however, it may be quite difficult to calculate it analytically, since the joint density function and the calculation of d-dimensional integral are required. Nevertheless, we can still obtain some intuition how vector quantization takes advantage of the probabilistic dependence from the following example.

Example 3.18. Suppose $d=2$, and the joint density is $f\left(\boldsymbol{\theta}\right)=n\in\mathbb{N}$ if $\boldsymbol{\theta}=\left(\theta_1,\theta_2\right)\in\cup_{i=1}^{n}\left[\frac{i-1}{n},\frac{i}{n}\right]^2$, and $f\left(\boldsymbol{\theta}\right)=0$ otherwise. This indicates θ_1 and θ_2 are positively correlated, representing the consumer's types for two complement commodities. We can show that both θ_1 and θ_2 are marginally uniformly distributed on $[0,1]$ with the correlation coefficient $\rho=1-\frac{1}{n^2}$. We can calculate $DP\left(\hat{f},f,2\right)=n=\frac{1}{\sqrt{1-\rho}}$. Note that

$DP\left(\hat{f}, f, 2\right) = 1$ when θ_1 and θ_2 are uncorrelated, i.e., $\rho = 0$ (or $n = 1$). The more correlated θ_1 and θ_2 are, i.e., the larger ρ (or n) is, the larger $DP\left(\hat{f}, f, 2\right)$ becomes. It becomes arbitrarily large as θ_1 and θ_2 become complete positive correlated, i.e., $\rho \to 1$ (or $n \to \infty$).

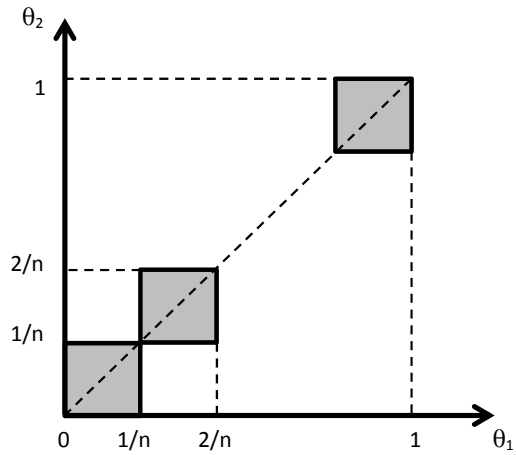

Fig. 1. An example of the Dependence Advantage

Now we compare the two-dimensional vector quantization with the scalar quantization for the above example. Since the marginal density is uniform, we do not have any shape advantage with vector quantization. Suppose we were only allowed to use rectangles to partition the space so that we would not have the space-filling advantage either. In this case, the above example says we can still reduce the MSE (or welfare loss) by a factor n via the dependence advantage. This is accomplished by partitioning only the region $\cup_{i=1}^{n}\left[\frac{i-1}{n}, \frac{i}{n}\right]^2$ (with the positive density) with vector quantization, instead of partitioning the whole space $[0,1]^2$ with repeated scalar quantization. Thus, the reduction results from exploiting the dependence between θ_1 and θ_2 by vector quantization.

Based on the above analysis, we now prove Lemma 3.5.

Proof of Lemma 3.5. Let $\hat{K} = \lceil M^{1/d} \rceil$. Recall that there are no shape and dependence advantages for i.i.d. uniform random variables $\theta_1, \ldots, \theta_d$. Thus $S\left(\hat{f}, d\right) = DP\left(\hat{f}, f, d\right) = 1$. When $\hat{K}$ is sufficiently large,

$$G_{SW} = SF(d) \times S\left(\hat{f}, d\right) \times DP\left(\hat{f}, f, d\right) = SF(d) \leq SF_U(d) \leq \frac{\pi e}{6}.$$

Note that $G_{SW} = \frac{\Delta\left(F;\hat{K};1\right)}{\frac{1}{d}\Delta\left(F;\hat{K}^d;d\right)}$, and $\Delta\left(F;\hat{K};1\right) = \frac{1}{24\hat{K}^2}$, as established in [Bergemann et al. 2012], so we have

$$\begin{aligned}\Delta\left(F;M;d\right) &\geq \Delta\left(F;\hat{K}^d;d\right) \geq \frac{6d \times \Delta\left(F;\hat{K};1\right)}{\pi e} = \frac{1}{4\pi e}\frac{d}{\hat{K}^2} \\ &\geq \frac{1}{4\pi e}\frac{d}{2\left(\hat{K}-1\right)^2} \geq \frac{1}{8\pi e}\frac{d}{M^{2/d}},\end{aligned}$$

which concludes the proof.■

In summary, vector quantization can take advantage of dimensionality, the shape of the marginal density, and the dependence among different elements of the type vector, whereas this is impossible with scalar quantization. Even for the i.i.d. uniform distribution for which there are no shape and dependence advantages, vector quantization can still offer the space-filling advantage approaching to $\frac{\pi e}{6} \approx 1.423$. That is, vector quantization can reduce the welfare loss by roughly $1 - \frac{6}{\pi e} \approx 29.7\%$ as the dimension d and the number of regions per dimension K become sufficiently large. Even though vector quantization might not improve the convergence rate of the welfare loss, it improves the coefficient significantly. This means that vector quantization can provide lower welfare losses per dimension compared with the scalar quantization.

4. CONCLUSIONS

Based on the information-theoretic approach developed in [Bergemann et al. 2012] in a one-dimensional environment, we analyzed the welfare maximizing problem in a multi-product environment. We offered two approaches to estimate a bound on the welfare loss. The first approach dealt with each dimension separately and then applied scalar quantization to each dimension, similar to the one-dimensional analysis. Such treatment ignores the dependence among the profile of all types. The second approach explicitly used vector quantization to introduce an additional advantage which in turn improved the coefficient of the convergence rate of the welfare loss. This improvement becomes significant when the number of choices along each dimension becomes large. Our analysis has illustrated that a simple contract with few choices can achieve a significantly high level of welfare.

APPENDIX

In this appendix, we provide a brief and self-contained introduction to the theory of quantization. We can view scalar (one-dimensional) quantization as a process of approximating a continuous random variable (called the input) X on $\Theta = [a,b] \subset R$ by a finite set of discrete values $Y \equiv \{y_k\}_{k=1}^n \subset R$. In other words, we can define an n-level scalar (one-dimensional) quantizer as $\{\mathcal{A}, Y\} = \{A_k, y_k\}_{k=1}^n$, where

(i) $\mathcal{A}$ is a partition of the input set $\Theta = [a,b]$ into n intervals: $\mathcal{A} \equiv \{A_k\}_{k=1}^n$, where $A_k = [x_{k-1}, x_k)$, and $\{x_k\}_{k=1}^n$ are often called the boundary points or endpoints which form an increasing sequence with $x_0 = a$, and $x_n = b$;

(ii) Y is a set of representation points: $Y = \{y_k\}_{k=1}^n \subset R$.

In this manner, the quantizer can be viewed as a mapping $y : \Theta \to Y$, so that $y\left(x\right) = y_k$ if $x \in A_k$, or equivalently:

$$y(x) = \sum_{k=1}^{n} y_k 1_{A_k}\left(x\right)$$

where the indicator function $1_{A_k}\left(x\right) = 1$ if $x \in A_k$ and 0 otherwise.

The quality of a quantizer is usually measured by the squared error $e(x,y) = |x-y|^2$. If the random variable X is drawn from a probability distribution function $F(x)$, then the n-level scalar (one-dimensional) quantization problem can be viewed as choosing the partition $\mathcal{A} = \{A_k\}_{k=1}^n$ and the set of representation points $Y = \{y_k\}_{k=1}^n$ to minimize the Mean Squared Error (MSE):

$$\min_{\mathcal{A},Y} MSE(n) = \min_{\{A_i,y_i\}_{i=1}^n} \mathbb{E}_X [X - y(X)]^2 = \min_{\{A_i,y_i\}_{i=1}^n} \sum_{k=1}^n \int_{A_k} (x - y_k)\, dF(x)$$

Now the question is how to determine the partition $\{A_k\}_{k=1}^n$ and the representation points $\{y_k\}_{k=1}^n$ and to minimize the mean squared error. In 1957, however, Lloyd proposed optimality conditions (called Lloyd-Max conditions) that any optimal quantizer (one with the smallest MSE) must satisfy, which can be stated in the following way: (1) given the representation points $Y = \{y_k\}_{k=1}^n$, the boundary point x_k is chosen to be the midpoint between the two representation points y_k and y_{k+1}, i.e., $x_k = \frac{1}{2}(y_k + y_{k+1})$; (2) given the partition $\mathcal{A} = \{A_k\}_{k=1}^n$, the representation point y_k corresponding to a given interval A_k must be the conditional mean of X on that interval, i.e., $y_k = \mathbb{E}_X [X | X \in A_k]$.

Similarly, we can view d-dimensional vector quantization as a process of approximating a d-dimensional continuous random vector (called the input) $\boldsymbol{X}$ on $\Theta \subset R^d$ by a finite set of discrete values $Y \equiv \{\boldsymbol{y}_m\}_{m=1}^M \subset R^d$. In other words, we can define an d-dimensional M-region vector quantizer as $\{\mathcal{B}, Y\} = \{B_m, \boldsymbol{y}_m\}_{m=1}^M$, where

(i) $\mathcal{B}$ is a partition of the input set Θ into M regions: $\mathcal{B} \equiv \{B_m\}_{m=1}^M$, where $B_i \cap B_j = \emptyset$ if $i \neq j$, and $\cup_{m=1}^M B_m = \Theta$;

(ii) Y is a set of representation points: $Y = \{\boldsymbol{y}_m\}_{m=1}^M \subset R^d$.

In this manner, the vector quantizer can be viewed as a mapping $\boldsymbol{y} : \Theta \to Y$, so that $\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{y}_m$ if $\boldsymbol{x} \in B_m$, or equivalently:

$$\boldsymbol{y}(\boldsymbol{x}) = \sum_{m=1}^M \boldsymbol{y}_m 1_{B_m}(\boldsymbol{x}),$$

where the indicator function $1_{B_m}(\boldsymbol{x}) = 1$ if $\boldsymbol{x} \in B_m$ and 0 otherwise.

The quality of a quantizer can be usually measured by the squared error $e(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|^2 = (\boldsymbol{x}-\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y})$, where $\|\cdot\|$ is the Euclidean norm. If the random vector $\boldsymbol{X}$ is drawn from a joint probability distribution function $F(\boldsymbol{x})$, then the d-dimensional M-region vector quantization problem can be viewed as choosing the partition $\mathcal{B} = \{B_m\}_{m=1}^M$ and the set of representation points $Y = \{\boldsymbol{y}_m\}_{m=1}^M$ to minimize the Mean Squared Error (MSE):

$$\min_{\mathcal{B},Y} MSE(M) = \min_{\{B_m,\boldsymbol{y}_m\}_{m=1}^M} \mathbb{E}_{\boldsymbol{X}} [\boldsymbol{X} - y(\boldsymbol{X})]^2 = \min_{\{B_m,\boldsymbol{y}_m\}_{m=1}^M} \sum_{m=1}^M \int_{B_m} (\boldsymbol{x} - \boldsymbol{y}_k)\, dF(\boldsymbol{x})$$

The Lloyd-Max conditions that any optimal vector quantizer must satisfy can be stated in the following way: (1) given the representation points $\{\boldsymbol{y}_m\}_{m=1}^M$, the partition $\mathcal{B} = \{B_m\}_{m=1}^M$ is chosen to be the Voronoi partition with respect to $\{\boldsymbol{y}_m\}_{m=1}^M$, i.e., $B_m = \{\boldsymbol{X} \in \Theta : \|\boldsymbol{X} - \boldsymbol{y}_m\| \leq \|\boldsymbol{X} - \boldsymbol{y}_l\| \text{ for all } l\}$; (2) given the partition $\mathcal{B} = \{B_m\}_{m=1}^M$, the representation points $\boldsymbol{y}_m$ corresponding to a given region B_m must be the conditional mean of $\boldsymbol{X}$ on that region, i.e., $\boldsymbol{y}_m = \mathbb{E}_{\boldsymbol{X}} [\boldsymbol{X} | \boldsymbol{X} \in B_m]$.

REFERENCES

ARMSTRONG, M. 1996. Multiproduct nonlinear pricing. *Econometrica 64*, 51–75.

ARMSTRONG, M. 1999. Price discrimination by a many-product firm. *Review of Economic Studies*.

BERGEMANN, D. AND PESENDORFER, M. 2007. Information structures in optimal auctions. *Journal of Economic Theory 137*, 580–609.

BERGEMANN, D., SHEN, J., XU, Y., AND YEH, E. 2012. Mechanism design with limited information: The case of nonlinear pricing. *Game Theory for Networks: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 0075*.

BLUMROSEN, L. AND FELDMAN, M. 2006. Implementation with a bounded action space. In *The 7th ACM Conference on Electronic Commerce*.

BLUMROSEN, L., NISAN, N., AND SEGAL, I. 2007. Auctions with severely bounded communicatione. *Journal of Artificial Intelligence Research 28*, 233–266.

CONWAY, J. H. AND SLOANE, N. J. A. 1985. A lower bound on the average error of vector quantizers. *IEEE Transactions on Information Theory 31,* 1, 106–109.

GERSHO, A. 1979. Asymptotically optimal block quantization. *IEEE Transactions on Information Theory 25,* 4, 373–380.

GERSHO, A. AND GRAY, R. M. 1992. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, Norwell, Massachusetts.

HOPPE, H., MOLDOVANU, B., AND OZDENOREN, E. 2010. Coarse matching with incomplete information. *Economic Theory*.

KOS, N. 2011. Communication and efficiency in auction. *Game and Economic Behavior*.

LEDYARD, J. O. AND PALFREY, T. R. 2002. The approximation of efficient public good mechanisms by simple voting schemes. *Journal of Public Economics 83*, 939–979.

LOOKABAUGH, T. D. AND GRAY, R. M. 1989. High-resolution quantization theory and the vector quantizer advantage. *IEEE Transactions on Information Theory 35,* 5, 1020–1033.

MASKIN, E. AND RILEY, J. 1984. Monopoly with incomplete information. *Rand Journal of Economics 15*, 171–196.

MCAFEE, P. 2002. Coarse matching. *Econometrica 70*, 2025–2034.

MUSSA, M. AND ROSEN, S. 1978. Monopoly and product quality. *Journal of Economic Theory 18*, 301–317.

ROCHET, J. C. AND CHONE, P. 1998. Ironing, sweeping and multidimensional screening. *Econometrica 66*, 783–826.

ROCHET, J. C. AND STOLE, L. 2003. The economics of multidimensional screening. *in Advances in Economic and Econometrics (edited by M. Dewatripont and L. P. Hansen and S. J. Turnovsky)*, 115–150.

ROTHKOPF, M. H. AND HARSTAD, R. M. 1994. On the role of discrete bid levels in oral auctions. *European Journal of Operational Research 74*, 572–581.

ROYDEN, H. L. 1968. *Real Analysis*. Collier Macmillan, New York.

WILSON, R. 1989. Efficient and competitive rationing. *Econometrica 57*, 1–40.

WILSON, R. 1993. *Nonlinear Pricing*. Oxford University Press, Oxford.

ZADOR, P. 1982. Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Transactions on Information Theory 28*, 134–149.