

University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

Mathematical and Statistical Sciences Faculty
Publications and Presentations

College of Sciences

10-2020

Quantization for Uniform Distributions on Hexagonal, Semicircular, and Elliptical Curves

Gabriela Pena

Hansapani Rodrigo

The University of Texas Rio Grande Valley

Mrinal Kanti Roychowdhury

The University of Texas Rio Grande Valley

Josef A. Sifuentes

The University of Texas Rio Grande Valley

Erwin Suazo

The University of Texas Rio Grande Valley

Follow this and additional works at: https://scholarworks.utrgv.edu/mss_fac



Part of the [Mathematics Commons](#)

Recommended Citation

Pena, G., Rodrigo, H., Roychowdhury, M.K. et al. Quantization for Uniform Distributions on Hexagonal, Semicircular, and Elliptical Curves. *J Optim Theory Appl* (2020). <https://doi.org/10.1007/s10957-020-01771-1>

This Article is brought to you for free and open access by the College of Sciences at ScholarWorks @ UTRGV. It has been accepted for inclusion in Mathematical and Statistical Sciences Faculty Publications and Presentations by an authorized administrator of ScholarWorks @ UTRGV. For more information, please contact justin.white@utrgv.edu, william.flores01@utrgv.edu.

To appear, Journal of Optimization Theory and Applications

QUANTIZATION FOR UNIFORM DISTRIBUTIONS ON HEXAGONAL, SEMICIRCULAR, AND ELLIPTICAL CURVES

¹GABRIELA PENA, ²HANSAPANI RODRIGO, ³MRINAL KANTI ROYCHOWDHURY,
⁴JOSEF SIFUENTES, AND ⁵ERWIN SUAZO

ABSTRACT. In this paper, first we have defined a uniform distribution on the boundary of a regular hexagon, and then investigated the optimal sets of n -means and the n th quantization errors for all positive integers n . We give an exact formula to determine them, if n is of the form $n = 6k$ for some positive integer k . We further calculate the quantization dimension, the quantization coefficient, and show that the quantization dimension is equal to the dimension of the object, and the quantization coefficient exists as a finite positive number. Then, we define a mixture of two uniform distributions on the boundary of a semicircular disc, and obtain a sequence and an algorithm, with the help of which we determine the optimal sets of n -means and the n th quantization errors for all positive integers n with respect to the mixed distribution. Finally, for a uniform distribution defined on an elliptical curve, we investigate the optimal sets of n -means and the n th quantization errors for all positive integers n .

1. INTRODUCTION

Quantization is a process of approximation with broad application in engineering and technology (see [GG, GN, Z]). For the mathematical treatment of quantization one is referred to Graf-Luschgy's book (see [GL1]). Recently, optimal quantization for uniform distributions on different regions have been investigated by several authors, for example, see [DR, R, RR1, RR2]. On the contrary, optimal quantization for uniform distributions on curves has not yet been much investigated. Such kind of problems has rigorous applications in many areas including signal processing. In this note we would like to list two such applications. The first application comes within the area of signal processing. When we drive long distances, quite often cellular signals get cut off. This happens because we are either far away from the tower, or there is no tower nearby to catch the signal. In optimal quantization one of our goals is to find the exact locations of the towers so that while driving we can get the best signal at our cell phones. The second application comes within the area of agriculture. The amount of agricultural water usage needs to be controlled by placing a minimal number of water sprinkles (or any other resources) in a way that they cover the whole cropland as required. We can use our knowledge in quantization to place the optimal number of sprinkles in order to fully cover the cropland area as these crops are usually planted on a boundary line of a k -sided polygon including the shapes that we discuss in this paper. In this regard, we find that it is important to investigate the optimal sets of n -means and the n th quantization errors for the points on a boundary of a hexagonal, a semicircular, or an elliptical curve with respect to a probability distribution.

In this paper, Section 3 deals with the quantization for a uniform distribution defined on the boundary of a regular hexagon. For this uniform distribution, in Theorem 3.8, we give an exact formula to determine the optimal sets of n -means and the n th quantization errors for all n if n is of the form $n = 6k$ for some positive integer k . We further calculate the quantization dimension, and the quantization coefficient, and show that the quantization dimension is equal to the dimension of the object, and the quantization coefficient exists as a finite positive number. In Proposition 3.10, we show that for sufficiently large n the points in an optimal set of n -means lie on the boundary of the hexagon. Section 4 deals with the quantization for a mixture of two

2010 *Mathematics Subject Classification.* 60Exx, 94A34.

Key words and phrases. Uniform distribution, optimal quantizers, quantization error, quantization dimension, quantization coefficient.

uniform distributions defined on the boundary of a semicircular disc. In this section, first we have explicitly determined the optimal sets of n -means and the n th quantization errors for all $1 \leq n \leq 9$. Then, we have proved Theorem 4.8, and defined a sequence and algorithm. With the help of the sequence and the algorithm, Theorem 4.8 gives all the optimal sets of n -means and the n th quantization errors for all $n \geq 4$. Section 5 deals with the quantization for a uniform distribution defined on the boundary of an ellipse. In Proposition 5.4, Proposition 5.5, and Proposition 5.6, we determine the optimal sets of n -means and the n th quantization errors for $n = 2, 6, 7$. Following the technique given in Proposition 5.5 and Proposition 5.6, we can obtain the optimal sets of n -means and the n th quantization errors for the uniform distribution for any positive integer n .

Finally, we would like to mention that the techniques, given in this paper, to determine the optimal sets of n -means and the n th quantization errors with respect to the probability distributions defined on three different curves will help to further investigate them for a more general curve. Such a problem with respect to any probability distribution, even for a uniform distribution, defined on any curve is not known yet.

2. PRELIMINARIES

Let \mathbb{R}^d denote the d -dimensional Euclidean space, $\|\cdot\|$ denote the Euclidean norm on \mathbb{R}^d for any $d \geq 1$, and $n \in \mathbb{N}$. Then, the n th *quantization error* for a Borel probability measure P on \mathbb{R}^d is defined by

$$V_n := V_n(P) = \inf \left\{ \int \min_{a \in \alpha} \|x - a\|^2 dP(x) : \alpha \subset \mathbb{R}^d, \text{card}(\alpha) \leq n \right\}.$$

A set α for which the infimum is achieved and contains no more than n points is called an *optimal set of n -means* for P , and the points in an optimal set are called *optimal quantizers*. Of course, this makes sense only if the mean squared error or the expected squared Euclidean distance $\int \|x\|^2 dP(x)$ is finite (see [AW, GKL, GL1, GL2]). It is known that for a continuous probability measure an optimal set of n -means always has exactly n -elements (see [GL1]). The number

$$D(P) := \lim_{n \rightarrow \infty} \frac{2 \log n}{-\log V_n(P)},$$

if it exists, is called the *quantization dimension* of the probability measure P ; on the other hand, for any $s \in (0, +\infty)$, the number $\lim_{n \rightarrow \infty} n^{\frac{s}{2}} V_n(P)$, if it exists, is called the *s -dimensional quantization coefficient* for P . For more details about the quantization dimension and the quantization coefficient, and their connections, one is referred to [GL1, P]. To know about the asymptotic quantization for probability measures on Riemannian manifolds, one can see [I].

3. QUANTIZATION FOR A UNIFORM DISTRIBUTION ON THE BOUNDARY OF A REGULAR HEXAGON

In this section, first we give the following proposition (see [GG, GL1]):

Proposition 3.1. *Let α be an optimal set of n -means for P , and $a \in \alpha$. Then,*

(i) $P(M(a|\alpha)) > 0$, (ii) $P(\partial M(a|\alpha)) = 0$, (iii) $a = E(X : X \in M(a|\alpha))$, where $M(a|\alpha)$ is the Voronoi region of $a \in \alpha$, i.e., $M(a|\alpha)$ is the set of all elements x in \mathbb{R}^d which are closest to a among all the elements in α .

Due to the above proposition, we see that if α is an optimal set and $a \in \alpha$, then a is the *conditional expectation* of the random variable X given that X takes values in the Voronoi region of a . Let i and j be the unit vectors in the positive directions of the x_1 - and x_2 -axes, respectively. In the sequel, the position vectors of the points $A_0, A_1, \dots, P, Q, R, D, E, F$, etc. will, respectively, be denoted by $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{p}, \tilde{q}, \tilde{r}, \tilde{d}, \tilde{e}, \tilde{f}$, etc, where by the position vector \tilde{a} of a point A , it is meant that $\overrightarrow{OA} = \tilde{a}$. In addition, we will identify the position vector of a

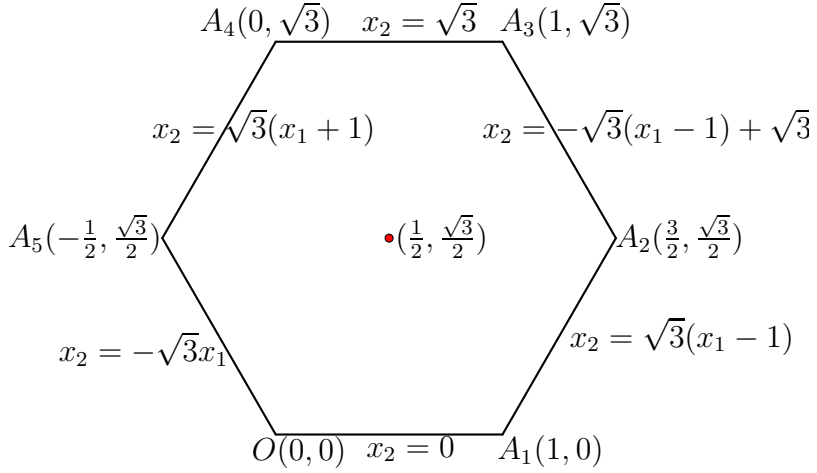


FIGURE 1. Regular hexagon with center $(\frac{1}{2}, \frac{\sqrt{3}}{2})$ and side length one.

point (a_1, a_2) by $(a_1, a_2) := a_1i + a_2j$, and apologize for any abuse in notation. For any two vectors \vec{u} and \vec{v} , let $\vec{u} \cdot \vec{v}$ denote the dot product between the two vectors \vec{u} and \vec{v} . Then, for any vector \vec{v} , by $(\vec{v})^2$, we mean $(\vec{v})^2 := \vec{v} \cdot \vec{v}$. Thus, $|\vec{v}| := \sqrt{\vec{v} \cdot \vec{v}}$, which is called the length of the vector \vec{v} . For any two position vectors $\tilde{a} := (a_1, a_2)$ and $\tilde{b} := (b_1, b_2)$, we write $\rho(\tilde{a}, \tilde{b}) := \|(a_1, b_1) - (a_2, b_2)\|^2 = (a_1 - a_2)^2 + (b_1 - b_2)^2$, which gives the squared Euclidean distance between the two points (a_1, a_2) and (b_1, b_2) . Let P and Q belong to an optimal set of n -means for some positive integer n , and let D be a point on the boundary of the Voronoi regions of the points P and Q . Since the boundary of the Voronoi regions of any two points is the perpendicular bisector of the line segment joining the points, we have $|\overrightarrow{DP}| = |\overrightarrow{DQ}|$, i.e., $(\overrightarrow{DP})^2 = (\overrightarrow{DQ})^2$ implying $(\tilde{p} - \tilde{d})^2 = (\tilde{q} - \tilde{d})^2$, i.e., $\rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0$. We call such an equation a *canonical equation*. By $E(X)$ and $V := V(X)$, we represent the expectation and the variance of a random variable X with respect to the probability distribution under consideration.

Let P be the uniform distribution defined on the boundary L of the regular hexagon with vertices $O(0,0)$, $A_1(1,0)$, $A_2(\frac{3}{2}, \frac{\sqrt{3}}{2})$, $A_3(1, \sqrt{3})$, $A_4(0, \sqrt{3})$, $A_5(-\frac{1}{2}, \frac{\sqrt{3}}{2})$ (see Figure 1). Let s represent the distance of any point on L from the origin tracing along the boundary of the hexagon in the counterclockwise direction. Then, the points $O, A_1, A_2, A_3, A_4, A_5$ are, respectively, represented by $s = 0, s = 1, s = 2, s = 3, s = 4$, and $s = 5$. For any two points A and B on the boundary L , not necessarily the vertices, by \widehat{AB} , it is meant the portion of the boundary L with initial point A and terminal point B . For example, suppose D is a point on the side OA_1 , and E is a point on the side A_2A_3 , then by \widehat{DE} it is meant

$$\widehat{DE} = DA_1 \cup A_1A_2 \cup A_2E.$$

Notice that $\widehat{ED} = EA_3 \cup A_3A_4 \cup A_4A_5 \cup A_5O \cup OD$, and so, \widehat{DE} and \widehat{ED} are not same. The probability density function (pdf) f of the uniform distribution P is given by $f(s) := f(x_1, x_2) = \frac{1}{6}$ for all $(x_1, x_2) \in L$, and zero, otherwise. Notice that $L = \bigcup_{j=1}^6 L_j$, where L_j , for $0 \leq t \leq 1$, are represented by the parametric equations as follows:

$$L_j = \begin{cases} \{t\tilde{a}_1 + (1-t)\tilde{o}\} & \text{if } j = 1 \\ \{t\tilde{a}_j + (1-t)\tilde{a}_{j-1}\} & \text{if } j = 2, 3, 4, 5, 6. \end{cases}$$

Thus, for $0 \leq t \leq 1$, we have

$$\begin{aligned} L_1 &= \{(t, 0)\}, & L_2 &= \left\{ \left(\frac{t}{2} + 1, \frac{\sqrt{3}t}{2} \right) \right\}, & L_3 &= \left\{ \left(\frac{3(1-t)}{2} + t, \frac{1}{2}\sqrt{3}(1-t) + \sqrt{3}t \right) \right\}, \\ L_4 &= \left\{ \left(1-t, \sqrt{3}(1-t) + \sqrt{3}t \right) \right\}, & L_5 &= \left\{ \left(-\frac{t}{2}, \sqrt{3}(1-t) + \frac{\sqrt{3}t}{2} \right) \right\}, \text{ and} \\ L_6 &= \left\{ \left(\frac{t-1}{2}, \frac{1}{2}\sqrt{3}(1-t) \right) \right\}. \end{aligned}$$

Again, $dP(s) = P(ds) = f(x_1, x_2)ds = \frac{1}{6}ds$. On each L_j for $1 \leq j \leq 6$, we have $(ds)^2 = (dx_1)^2 + (dx_2)^2 = (dt)^2$ yielding $ds = dt$.

Let us now prove the following lemma.

Lemma 3.2. *Let X be a continuous random variable with uniform distribution taking values on L . Then, $E(X) = (\frac{1}{2}, \frac{\sqrt{3}}{2})$ and $V := V(X) = \frac{5}{6}$.*

Proof. Recall that by (a, b) it is meant $ai + bj$, where i and j are two unit vectors in the positive directions of x_1 - and x_2 -axes, respectively. Thus, we have,

$$\begin{aligned} E(X) &= \int_L (x_1 i + x_2 j) dP = \frac{1}{6} \left(\int_0^1 \left((t, 0) + \left(\frac{t}{2} + 1, \frac{\sqrt{3}t}{2} \right) + \left(\frac{3(1-t)}{2} + t, \frac{1}{2}\sqrt{3}(1-t) + \sqrt{3}t \right) \right. \right. \\ &\quad \left. \left. + \left(1-t, \sqrt{3}(1-t) + \sqrt{3}t \right) + \left(-\frac{t}{2}, \sqrt{3}(1-t) + \frac{\sqrt{3}t}{2} \right) + \left(\frac{t-1}{2}, \frac{1}{2}\sqrt{3}(1-t) \right) \right) dt \right) \end{aligned}$$

implying $E(X) = (\frac{1}{2}, \frac{\sqrt{3}}{2})$. The variance $V := V(X)$ is given by

$$\begin{aligned} V(X) &= E\|X - E(X)\|^2 = \int_L \rho((x_1, x_2), E(X)) dP = \frac{1}{6} \int_L \rho((x_1, x_2), E(X)) ds \\ &= \frac{1}{6} \sum_{j=1}^6 \int_{L_j} \rho((x_1, x_2), E(X)) ds = \int_{L_1} \rho((t, 0), (\frac{1}{2}, \frac{\sqrt{3}}{2})) dt = \int_0^1 \rho((t, 0), (\frac{1}{2}, \frac{\sqrt{3}}{2})) dt = \frac{5}{6}. \end{aligned}$$

Hence, the proof of the lemma is complete. \square

Remark 3.3. For any $(a, b) \in \mathbb{R}^2$, we have

$$E\|X - (a, b)\|^2 = V(X) + \|(a, b) - (\frac{1}{2}, \frac{\sqrt{3}}{2})\|^2,$$

which is minimum if $(a, b) = (\frac{1}{2}, \frac{\sqrt{3}}{2})$, and the minimum value is $V(X)$. Thus, we see that the optimal set of one-mean is the set $\{(\frac{1}{2}, \frac{\sqrt{3}}{2})\}$, and the corresponding quantization error is the variance $V := V(X)$ of the random variable X .

The following proposition gives the optimal set of two-means.

Proposition 3.4. *The set $\{(\frac{13}{12}, \frac{\sqrt{3}}{2}), (-\frac{1}{12}, \frac{\sqrt{3}}{2})\}$ forms an optimal set of two-means, and the quantization error for two-means is given by $V_2 = \frac{71}{144} = 0.493056$.*

Proof. Let P and Q form an optimal set of two-means. Let ℓ be the boundary of their Voronoi regions. Let P lie in the region which contains the vertex A_1 , and Q lie in the other region. The following cases can arise:

Case 1. ℓ intersects the sides OA_1 and A_1A_2 .

Let ℓ intersect OA_1 and A_1A_2 at the points D and E , respectively. Let the parametric values for D and E be, respectively, α and β , i.e., $\tilde{d} = (\alpha, 0)$, and $\tilde{e} = (\frac{\beta}{2} + 1, \frac{\sqrt{3}}{2}\beta)$, where $0 \leq \alpha < 1$

and $0 < \beta \leq 1$. Then, we have

$$\begin{aligned}\tilde{p} &= E(X : X \in \widehat{DE}) = \frac{\int_{\widehat{DE}}(x_1, x_2)dP}{\int_{\widehat{DE}}dP} = \frac{\int_{\alpha}^1(t, 0) dt + \int_0^{\beta}(\frac{t}{2} + 1, \frac{\sqrt{3}t}{2}) dt}{\int_{\alpha}^1 1 dt + \int_0^{\beta} 1 dt} \\ &= \left(\frac{-2\alpha^2 + \beta^2 + 4\beta + 2}{-4\alpha + 4\beta + 4}, \frac{\sqrt{3}\beta^2}{-4\alpha + 4\beta + 4} \right).\end{aligned}$$

Similarly,

$$\tilde{q} = E(X : X \in \widehat{ED}) = \left(\frac{2\alpha^2 - \beta(\beta + 4) + 10}{4(\alpha - \beta + 5)}, -\frac{\sqrt{3}(\beta^2 - 12)}{4(\alpha - \beta + 5)} \right).$$

Recall that the boundary ℓ of the two Voronoi regions is the perpendicular bisector of the line segment joining P and Q , and hence, we have the two canonical equations as $\rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0$, and $\rho(\tilde{e}, \tilde{p}) - \rho(\tilde{e}, \tilde{q}) = 0$. There is no solution for these two equations yielding the fact that this case cannot happen.

Case 2. ℓ intersects the sides OA_1 and A_2A_3 .

Let ℓ intersect OA_1 and A_2A_3 at the points D and E , respectively. Let the parametric values for D and E be, respectively, α and β , i.e., $\tilde{d} = (\alpha, 0)$, and $\tilde{e} = (\frac{3(1-\beta)}{2} + \beta, \frac{1}{2}\sqrt{3}(1-\beta) + \sqrt{3}\beta)$, where $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$. Then, proceeding in the similar way as Case 1, we have

$$\tilde{p} = E(X : X \in \widehat{DE}) = \frac{\int_{\widehat{DE}}(x_1, x_2)dP}{\int_{\widehat{DE}}dP} = \left(\frac{-2\alpha^2 - \beta^2 + 6\beta + 7}{-4\alpha + 4\beta + 8}, -\frac{\sqrt{3}(\beta + 1)^2}{4(\alpha - \beta - 2)} \right),$$

and

$$\tilde{q} = E(X : X \in \widehat{ED}) = \left(\frac{2\alpha^2 + \beta^2 - 6\beta + 5}{4\alpha - 4\beta + 16}, -\frac{\sqrt{3}(\beta^2 + 2\beta - 11)}{4(\alpha - \beta + 4)} \right).$$

Solving the two canonical equations $\rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0$, and $\rho(\tilde{e}, \tilde{p}) - \rho(\tilde{e}, \tilde{q}) = 0$, we have $\alpha = 0$, and $\beta = 1$ yielding $\tilde{p} = (1, \frac{1}{\sqrt{3}})$, and $\tilde{q} = (0, \frac{2}{\sqrt{3}})$. Due to symmetry, we can obtain the distortion error due to the points \tilde{p} and \tilde{q} as

$$\begin{aligned}\int_L \min_{a \in \{\tilde{p}, \tilde{q}\}} \|x - a\|^2 dP &= \frac{2}{6} \left(\int_0^1 \rho((t, 0), (1, \frac{1}{\sqrt{3}})) dt + \int_0^1 \rho((\frac{t}{2} + 1, \frac{\sqrt{3}t}{2}), (1, \frac{1}{\sqrt{3}})) dt \right. \\ &\quad \left. + \int_0^1 \rho((\frac{3(1-t)}{2} + t, \frac{1}{2}\sqrt{3}(1-t) + \sqrt{3}t), (1, \frac{1}{\sqrt{3}})) dt \right) = \frac{1}{2}.\end{aligned}$$

Case 3. ℓ intersects the sides OA_1 and A_3A_4 .

Let ℓ intersect OA_1 and A_3A_4 at the points D and E , respectively. Let the parametric values for D and E be, respectively, α and β , i.e., $\tilde{d} = (\alpha, 0)$, and $\tilde{e} = (\beta, \sqrt{3})$, where $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$. Then, proceeding in the similar way as Case 1, we have

$$\tilde{p} = E(X : X \in \widehat{DE}) = \frac{\int_{\widehat{DE}}(x_1, x_2)dP}{\int_{\widehat{DE}}dP} = \left(\frac{\alpha^2 + \beta^2 - 2\beta - 6}{2(\alpha - \beta - 3)}, -\frac{\sqrt{3}(\beta + 1)}{\alpha - \beta - 3} \right),$$

and

$$\tilde{q} = E(X : X \in \widehat{ED}) = \left(\frac{\alpha^2 + \beta^2 - 2\beta}{2\alpha - 2\beta + 6}, -\frac{\sqrt{3}(\beta - 2)}{\alpha - \beta + 3} \right).$$

Solving the two canonical equations $\rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0$, and $\rho(\tilde{e}, \tilde{p}) - \rho(\tilde{e}, \tilde{q}) = 0$, we have the following three sets of solutions:

$$\{\alpha \rightarrow 0, \beta \rightarrow 0\}, \{\alpha \rightarrow \frac{1}{2}, \beta \rightarrow \frac{1}{2}\}, \{\alpha \rightarrow 1, \beta \rightarrow 1\}.$$

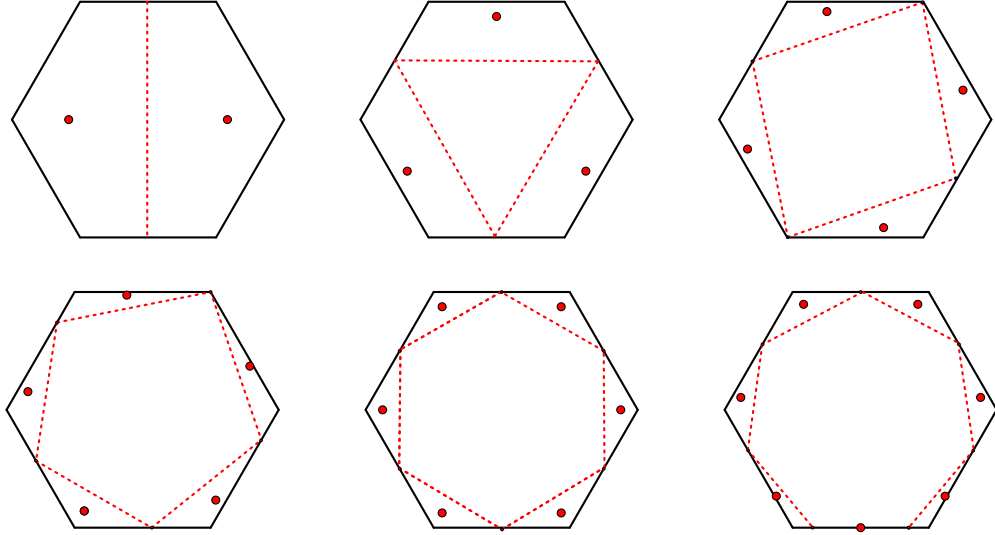


FIGURE 2. Points in an optimal set of n -means for $2 \leq n \leq 7$.

If $\{\alpha \rightarrow 0, \beta \rightarrow 0\}$, then $\tilde{p} = (1, \frac{1}{\sqrt{3}})$ and $\tilde{q} = (0, \frac{2}{\sqrt{3}})$ which falls in Case 2. If $\{\alpha \rightarrow \frac{1}{2}, \beta \rightarrow \frac{1}{2}\}$, then $\tilde{p} = (\frac{13}{12}, \frac{\sqrt{3}}{2})$, and $\tilde{q} = (-\frac{1}{12}, \frac{\sqrt{3}}{2})$, and the corresponding distortion is obtained as

$$\begin{aligned} \int_L \min_{a \in \{\tilde{p}, \tilde{q}\}} \|x - a\|^2 dP &= \frac{2}{6} \left(\int_{\frac{1}{2}}^1 \rho((t, 0), (\frac{13}{12}, \frac{\sqrt{3}}{2})) dt + \int_0^1 \rho((\frac{t}{2} + 1, \frac{\sqrt{3}t}{2}), (\frac{13}{12}, \frac{\sqrt{3}}{2})) dt \right. \\ &+ \int_0^1 \rho((\frac{3(1-t)}{2} + t, \frac{1}{2}\sqrt{3}(1-t) + \sqrt{3}t), (\frac{13}{12}, \frac{\sqrt{3}}{2})) dt \\ &\left. + \int_0^{\frac{1}{2}} \rho((1-t, \sqrt{3}(1-t) + \sqrt{3}t), (\frac{13}{12}, \frac{\sqrt{3}}{2})) dt \right) = \frac{71}{144} = 0.493056, \end{aligned}$$

which is smaller than $\frac{1}{2}$, the distortion error obtained in Case 2. If $\{\alpha \rightarrow 1, \beta \rightarrow 1\}$, then $\tilde{p} = (1, \frac{2}{\sqrt{3}})$ and $\tilde{q} = (0, \frac{1}{\sqrt{3}})$ which falls in Case 2 after giving a rotation of the hexagon with respect to its center by an angle of $\frac{\pi}{3}$. Thus, the distortion error in this case is $\frac{1}{2}$ which is same as in Case 2.

Case 4. ℓ intersects the sides OA_1 and A_4A_5 .

This case is the reflection of Case 2 with respect to the diagonal A_1A_4 , and so the same distortion error $\frac{1}{2}$ occurs in this case.

Case 5. ℓ intersects the sides OA_1 and OA_5 .

This case is the reflection of Case 1 with respect to the line $x_1 = \frac{1}{2}$, and so, this case cannot happen.

Taking into account all the distortion errors, we see that the distortion error is minimum when $\tilde{p} = (\frac{13}{12}, \frac{\sqrt{3}}{2})$, and $\tilde{q} = (-\frac{1}{12}, \frac{\sqrt{3}}{2})$. Hence, the optimal set of two-means is $\{(\frac{13}{12}, \frac{\sqrt{3}}{2}), (-\frac{1}{12}, \frac{\sqrt{3}}{2})\}$, and the quantization error for two-means is $V_2 = \frac{71}{144} = 0.493056$. Thus, the proof of the proposition is deduced (see Figure 2). \square

Proposition 3.5. *The set $\{(\frac{37}{32}, \frac{9\sqrt{3}}{32}), (\frac{1}{2}, \frac{15\sqrt{3}}{16}), (-\frac{5}{32}, \frac{9\sqrt{3}}{32})\}$ forms an optimal set of three-means with quantization error $V_3 = \frac{199}{768} = 0.259115$.*

Proof. The proof of the proposition can be deduced by considering different cases as it was done in Proposition 3.4. To avoid too much technicality, we will prove it in a different way. Recall that the probability distribution is uniform on the boundary of the regular hexagon, and so we can assume that the Voronoi regions of the elements in an optimal set of three-means will

partition the boundary of the hexagon into three equal parts. Let the points P , Q , and R form an optimal set of three-means. Let the boundaries of the Voronoi regions cut the sides OA_1 , A_2A_3 , and A_4A_5 at the points D , E , and F with parameters, respectively, given by α , β , and γ . Let P , Q , and R lie in the Voronoi regions that contain the vertices A_1 , A_3 , and A_5 , respectively. Then, $\tilde{p} = E(X : X \in \widehat{DE})$, $\tilde{q} = E(X : X \in \widehat{EF})$, and $\tilde{r} = E(X : X \in \widehat{FD})$. Then, proceeding as Case 1 of Proposition 3.4, we have

$$\begin{aligned}\tilde{p} &= \left(\frac{-2\alpha^2 - \beta^2 + 6\beta + 7}{-4\alpha + 4\beta + 8}, -\frac{\sqrt{3}(\beta + 1)^2}{4(\alpha - \beta - 2)} \right), \\ \tilde{q} &= \left(\frac{\beta^2 - 6\beta - \gamma^2 + 7}{-4\beta + 4\gamma + 8}, \frac{\sqrt{3}(\beta^2 + 2\beta + \gamma^2 - 4\gamma - 7)}{4(\beta - \gamma - 2)} \right), \\ \tilde{r} &= \left(\frac{2\alpha^2 + \gamma^2 - 2}{4(\alpha - \gamma + 2)}, \frac{\sqrt{3}(\gamma - 2)^2}{4(\alpha - \gamma + 2)} \right).\end{aligned}$$

Solving the canonical equations $\rho(\tilde{d}, \tilde{r}) - \rho(\tilde{d}, \tilde{p}) = 0$, $\rho(\tilde{e}, \tilde{p}) - \rho(\tilde{e}, \tilde{q}) = 0$, and $\rho(\tilde{f}, \tilde{q}) - \rho(\tilde{f}, \tilde{r}) = 0$, we have three sets of solutions:

$$\{\alpha \rightarrow 0, \beta \rightarrow 0, \gamma \rightarrow 0\}, \{\alpha \rightarrow \frac{1}{2}, \beta \rightarrow \frac{1}{2}, \gamma \rightarrow \frac{1}{2}\}, \text{ and } \{\alpha \rightarrow 1, \beta \rightarrow 1, \gamma \rightarrow 1\}.$$

If $\{\alpha \rightarrow 0, \beta \rightarrow 0, \gamma \rightarrow 0\}$, then $\tilde{p} = (\frac{7}{8}, \frac{\sqrt{3}}{8})$, $\tilde{q} = (\frac{7}{8}, \frac{7\sqrt{3}}{8})$, and $\tilde{r} = (-\frac{1}{4}, \frac{\sqrt{3}}{2})$ yielding the distortion error

$$\int_L \min_{a \in \{\tilde{p}, \tilde{q}, \tilde{r}\}} \|x - a\|^2 dP = \frac{3}{6} \left(\int_0^1 \rho((\tilde{t}, 0), \tilde{p}) dt + \int_0^1 \rho((\frac{t}{2} + 1, \frac{\sqrt{3}t}{2}), \tilde{p}) dt \right) = \frac{13}{48}.$$

Similarly, if $\{\alpha \rightarrow \frac{1}{2}, \beta \rightarrow \frac{1}{2}, \gamma \rightarrow \frac{1}{2}\}$, then $\tilde{p} = (\frac{37}{32}, \frac{9\sqrt{3}}{32})$, $\tilde{q} = (\frac{1}{2}, \frac{15\sqrt{3}}{16})$, and $\tilde{r} = (-\frac{5}{32}, \frac{9\sqrt{3}}{32})$ yielding the distortion error $\frac{199}{768}$; and if $\{\alpha \rightarrow 1, \beta \rightarrow 1, \gamma \rightarrow 1\}$, then $\tilde{p} = (\frac{5}{4}, \frac{\sqrt{3}}{2})$, $\tilde{q} = (\frac{1}{8}, \frac{7\sqrt{3}}{8})$, and $\tilde{r} = (\frac{1}{8}, \frac{\sqrt{3}}{8})$ with distortion error $\frac{13}{48}$. Since among the distortion errors, $\frac{199}{768} = 0.259115 < 0.270833 = \frac{13}{48}$, we can say that the set $\{(\frac{37}{32}, \frac{9\sqrt{3}}{32}), (\frac{1}{2}, \frac{15\sqrt{3}}{16}), (-\frac{5}{32}, \frac{9\sqrt{3}}{32})\}$ forms an optimal set of three-means with quantization error $V_3 = \frac{199}{768} = 0.259115$ (see Figure 2). \square

Proceeding in the similar way as Proposition 3.5, the following two propositions can be proved.

Proposition 3.6. *The set $\{(\frac{17}{24}, \frac{1}{8\sqrt{3}}), (\frac{31}{24}, \frac{5\sqrt{3}}{8}), (\frac{7}{24}, \frac{23}{8\sqrt{3}}), (-\frac{7}{24}, \frac{3\sqrt{3}}{8})\}$ forms an optimal set of four-means with quantization error $V_4 = \frac{23}{144} = 0.159722$ (see Figure 2).*

Proposition 3.7. *The set $\{(0.07095, 0.122889), (1.03737, 0.20269), (1.2881, 1.18696), (0.383892, 1.70901), (-0.343148, 0.99973)\}$ forms the optimal set of five-means with quantization error $V_5 = 0.10509$ (see Figure 2).*

The following theorem determines the optimal sets of n -means and the n th quantization errors for all positive integers n of the form $n = 6k$, where $k \in \mathbb{N}$. It also helps us to determine the quantization dimension and the quantization coefficient for the uniform distribution defined on the boundary of the regular hexagon.

Theorem 3.8. *Let $n \in \mathbb{N}$ be such that $n = 6k$ for some positive integer k . Then, the optimal set of n -means for P is given by*

$$\begin{aligned}\alpha_n &= \left\{ \left(\frac{r}{8}, \frac{\sqrt{3}r}{8} \right), \left(1 - \frac{r}{8}, \frac{\sqrt{3}r}{8} \right), \left(\frac{6-r}{4}, \frac{\sqrt{3}}{2} \right), \left(1 - \frac{r}{8}, -\frac{1}{8}\sqrt{3}(r-8) \right), \right. \\ &\quad \left. \left(\frac{r}{8}, -\frac{1}{8}\sqrt{3}(r-8) \right), \left(\frac{r-2}{4}, \frac{\sqrt{3}}{2} \right) \right\} \cup \gamma \cup \bigcup_{i=1}^5 T_i(\gamma),\end{aligned}$$

where $\gamma := \begin{cases} \{r + \frac{2j-1}{2(k-1)}(1-2r) : j = 1, 2, \dots, (k-1)\} & \text{if } k \geq 2 \\ \emptyset & k = 1 \end{cases}$, and $r = \frac{2(\sqrt{13}(k-1)-4)}{13(k-1)^2-16}$, and

T_i for $1 \leq i \leq 5$, are five affine transformations on \mathbb{R}^2 , such that $T_i(OA_1) = A_iA_{i+1}$ for $1 \leq i \leq 4$, and $T_5(OA_1) = A_5O$. The quantization error for n -means is given by

$$V_n = \frac{13(13(k-1)^2 - 8\sqrt{13}(k-1) + 16)}{12(16 - 13(k-1)^2)^2}.$$

Proof. Let us first prove the theorem for $n = 6$, i.e., when $k = 1$. Recall that the probability distribution is uniform on the boundary of the regular hexagon, and so we can assume that the Voronoi regions of the elements in an optimal set of six-means will partition the boundary of the hexagon into six equal parts. Let the boundaries of the Voronoi regions cut the sides $OA_1, A_1A_2, A_2A_3, A_3A_4, A_4A_5$, and A_5O of the hexagon given in Figure 1 at the points D_1, D_2, \dots, D_6 , respectively. Then, we must have $|OD_1| = |A_1D_2| = |A_2D_3| = \dots = |A_5D_6|$ which equals r , say, where $0 \leq r \leq 1$. Let $\{\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_6\}$ forms an optimal set of six-means, where $\tilde{p}_1 = E(X : X \in \widehat{D_1D_2})$, $\tilde{p}_2 = E(X : X \in \widehat{D_2D_3})$, \dots , $\tilde{p}_5 = E(X : X \in \widehat{D_5D_6})$, and $\tilde{p}_6 = E(X : X \in \widehat{D_6D_1})$. Then, we have

$$\tilde{p}_1 = \frac{\int_r^1 (t, 0) dt + \int_0^r (\frac{t}{2} + 1, \frac{\sqrt{3}t}{2}) dt}{\int_r^1 1 dt + \int_0^r 1 dt} = (-\frac{r^2}{4} + r + \frac{1}{2}, \frac{\sqrt{3}r^2}{4}).$$

Similarly, $\tilde{p}_2 = (\frac{1}{4}(-2r^2 + 2r + 5), \frac{1}{4}\sqrt{3}(2r + 1))$, and so on. Solving the canonical equations $\rho(\tilde{d}_2, \tilde{p}_1) - \rho(\tilde{d}_2, \tilde{p}_2) = 0$, we have $r = 0, \frac{1}{2}, 1$. If $r = 0$, then we see that D_1, D_2, \dots, D_6 coincide with O, A_1, A_2, \dots, A_5 , respectively. On the other hand, if $r = 1$, then D_1, D_2, \dots, D_6 coincide with A_1, A_2, \dots, A_5, O , respectively. Thus, if $r = 0$, or if $r = 1$, then the optimal set of six-means consists of the midpoints of the sides of the hexagon, and hence, the distortion error for six-means in these two cases is given by

$$6 \int_0^1 \rho((t, 0), (\frac{1}{2}, 0)) dP = \int_0^1 \rho((t, 0), (\frac{1}{2}, 0)) dt = \frac{1}{12}.$$

If $r = \frac{1}{2}$, then D_1, D_2, \dots, D_6 coincide with the midpoints of the sides of $OA_1, A_1A_2, \dots, A_5O$, respectively, yielding $\tilde{p}_1 = (\frac{15}{16}, \frac{\sqrt{3}}{16})$, $\tilde{p}_2 = (\frac{11}{8}, \frac{\sqrt{3}}{2})$, $\tilde{p}_3 = (\frac{15}{16}, \frac{15\sqrt{3}}{16})$, $\tilde{p}_4 = (\frac{1}{16}, \frac{15\sqrt{3}}{16})$, $\tilde{p}_5 = (-\frac{3}{8}, \frac{\sqrt{3}}{2})$, and $\tilde{p}_6 = (\frac{1}{16}, \frac{\sqrt{3}}{16})$ (see Figure 2), and the corresponding distortion error is given by

$$6 \left(\int_{\frac{1}{2}}^1 \rho((t, 0), \tilde{p}_1) dP + \int_0^{\frac{1}{2}} \rho((\frac{t}{2} + 1, \frac{\sqrt{3}t}{2}), \tilde{p}_1) dP \right) = \frac{13}{192}.$$

Since $\frac{13}{192} < \frac{1}{12}$, the set $\{\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_6\}$ obtained for $r = \frac{1}{2}$, gives the optimal set of six-means with quantization error $V_6 = \frac{13}{192}$. Notice that for $k = 1$, we have $r = \frac{1}{2}$, and $\gamma = \emptyset$. Thus, the statement of the proposition is true for $n = 6$, i.e., when $k = 1$. Let $n = 6k$ for $k \geq 2$. Then, as the hexagon is a regular polygon, and the probability distribution P is uniform, it is not difficult to show that the optimal set α_n contains six elements from each of the interior angles, $k - 1$ elements from each side of the hexagon. Let a, b, c, d, e, f be the six points that α_n contains from the interior of the angles $\angle O, \angle A_1, \angle A_2, \angle A_3, \angle A_4, \angle A_5$, respectively. As the optimal set α_n contains $k - 1$ elements from each side of the hexagon, and P is a uniform probability distribution, the Voronoi regions of a, b, c, d, e, f will form isosceles triangles P -almost surely. Let the length of each of the two equal sides of the isosceles triangle formed by the Voronoi regions equal r . Then,

$$a = \frac{\int_{1-r}^1 (\frac{t-1}{2}, \frac{1}{2}\sqrt{3}(1-t)) dt + \int_0^r (t, 0) dt}{\int_{1-r}^1 1 dt + \int_0^r 1 dt} = (\frac{r}{8}, \frac{\sqrt{3}r}{8}).$$

Similarly,

$$b = \left(1 - \frac{r}{8}, \frac{\sqrt{3}r}{8}\right), \quad c = \left(\frac{6-r}{4}, \frac{\sqrt{3}}{2}\right), \quad d = \left(1 - \frac{r}{8}, -\frac{1}{8}\sqrt{3}(r-8)\right),$$

$$e = \left(\frac{r}{8}, -\frac{1}{8}\sqrt{3}(r-8)\right), \quad \text{and } f = \left(\frac{r-2}{4}, \frac{\sqrt{3}}{2}\right).$$

Let γ be the set of all the $k-1$ points that α_n contains from the side OA . Then, the Voronoi regions of the points in γ covers the closed interval $[r, 1-r]$ yielding

$$\gamma = \left\{ \left(r + \frac{2j-1}{2(k-1)}(1-2r), 0 \right) : j = 1, 2, \dots, (k-1) \right\}.$$

Let T_i for $1 \leq i \leq 5$ be the affine transformations on \mathbb{R}^2 as given in the hypothesis. Then, the set of points that α_n contains from the sides A_iA_j is $T_i(\gamma)$, and the set of points that α_n contains from the side A_5O is $T_5(\gamma)$, yielding

$$\alpha_n = \{a, b, c, d, e, f\} \cup \gamma \cup \bigcup_{i=1}^5 T_i(\gamma).$$

Using the symmetry, the quantization error for n -means is obtained as

$$\begin{aligned} V_n &= 6(\text{quantization error due to the points } a \text{ and the } (k-1) \text{ points in } \gamma) \\ &= \frac{6}{6} \left(\int_{1-r}^1 \rho\left(\left(\frac{t-1}{2}, \frac{1}{2}\sqrt{3}(1-t)\right), a\right) dt + \int_0^r \rho\left((t, 0), a\right) dt \right. \\ &\quad \left. + (k-1) \left(\int_r^{r+\frac{1-2r}{k-1}} \rho\left((t, 0), \left(r + \frac{1}{2(k-1)}(1-2r), 0\right)\right) dt \right) \right) \\ &= \frac{1}{24(k-1)^2} \left(24r^2 - 12r + r^3 (13(k-1)^2 - 16) + 2 \right). \end{aligned}$$

Notice that for a given k , the quantization error V_n is a function of r . Solving $\frac{\partial V_n}{\partial r} = 0$, we have $r = \frac{2(\sqrt{13}(k-1)-4)}{13(k-1)^2-16}$. Putting $r = \frac{2(\sqrt{13}(k-1)-4)}{13(k-1)^2-16}$, we have

$$V_n = \frac{13(13(k-1)^2 - 8\sqrt{13}(k-1) + 16)}{12(16 - 13(k-1)^2)^2}.$$

Thus, the proof of the theorem is complete. \square

Remark 3.9. By Proposition 3.4, we see that the boundary of the Voronoi regions of the optimal set of two-means bisects the two opposite sides of the hexagon. Due to rotational symmetry, there are three different optimal sets of two-means. Proposition 3.5 implies that the points in an optimal set of three-means form an equilateral triangle. Due to rotational symmetry, there are two different optimal sets of three-means. Proposition 3.6 yields that the points in an optimal set of four-means form a rectangle of side lengths $\frac{7}{6}$ and $\frac{2}{\sqrt{3}}$. Due to rotational symmetry, there are three different optimal sets of four-means. Proposition 3.7 implies that there are five different optimal sets of five-means. Theorem 3.8 implies that an optimal set of n -means, where $n = 6k$ for some $k \in \mathbb{N}$, is unique. We observe that the points in an optimal set of six-means form a regular hexagon of side length $\frac{7}{8}$. On the other hand, if n is of the form $n = 6k + m$, where $1 \leq m \leq 5$, then we can show that an optimal set of n -means contains six elements from each of the interior angles, k elements from each of m sides, $k-1$ elements from each of the remaining $6-m$ sides, and the number of such sets is 6C_m . The proof being too technical, we skip the proof of it in the paper.

Proposition 3.10. *For sufficiently large n the points in an optimal set of n -means lie on the boundary of the hexagon.*

Proof. For large n there exists a unique positive integer $\ell(n)$ such that

$$(1) \quad 6\ell(n) \leq n < 6(\ell(n) + 1).$$

Let $a_n, b_n, c_n, d_n, e_n, f_n$ be the six points that an optimal set α_n contains from the interior, respectively, of the angles $\angle O, \angle A_1, \angle A_2, \angle A_3, \angle A_4, \angle A_5$. By Theorem 3.8, we have

$$a_{6\ell(n)} = \left(\frac{\sqrt{13}(\ell(n) - 1) - 4}{4(13(\ell(n) - 1)^2 - 16)}, \frac{\sqrt{3}(\sqrt{13}(\ell(n) - 1) - 4)}{4(13(\ell(n) - 1)^2 - 16)} \right) \text{ implying } \lim_{\ell(n) \rightarrow \infty} a_{6\ell(n)} = (0, 0).$$

Similarly, $\lim_{\ell(n) \rightarrow \infty} a_{6(\ell(n)+1)} = (0, 0)$. Thus, (1) implies that $\lim_{n \rightarrow \infty} a_n = (0, 0) = O$. Similarly, we can show that if $n \rightarrow \infty$, then $b_n \rightarrow A_1, c_n \rightarrow A_2, d_n \rightarrow A_3, e_n \rightarrow A_4$, and $f_n \rightarrow A_5$. The rest of the points in α_n are already on the boundary of the hexagon. Thus, the proof of the proposition is complete. \square

Proposition 3.11. *Quantization dimension $D(P)$ of the uniform distribution P defined on the boundary of the regular hexagon equals the dimension of the boundary of the hexagon. Moreover, the quantization coefficient exists as a finite positive number which equals 3.*

Proof. For $n \in \mathbb{N}, n \geq 6$, let $\ell(n)$ be the unique positive integer such that $6\ell(n) \leq n < 6(\ell(n)+1)$. Then, $V_{6(\ell(n)+1)} < V_n \leq V_{6\ell(n)}$ implying

$$(2) \quad \frac{2 \log(6\ell(n))}{-\log V_{6(\ell(n)+1)}} < \frac{2 \log n}{-\log V_n} < \frac{2 \log(6(\ell(n) + 1))}{-\log V_{6\ell(n)}}.$$

Notice that

$$\lim_{n \rightarrow \infty} \frac{2 \log(6\ell(n))}{-\log V_{6(\ell(n)+1)}} = \lim_{\ell(n) \rightarrow \infty} \frac{2 \log(6\ell(n))}{-\log \left(\frac{13(13\ell(n)^2 - 8\sqrt{13}\ell(n) + 16)}{12(16 - 13\ell(n)^2)^2} \right)} = 1,$$

and

$$\lim_{n \rightarrow \infty} \frac{2 \log(6(\ell(n) + 1))}{-\log V_{6\ell(n)}} = \lim_{\ell(n) \rightarrow \infty} \frac{2 \log(6(\ell(n) + 1))}{-\log \left(\frac{13(13(\ell(n)-1)^2 - 8\sqrt{13}(\ell(n)-1) + 16)}{12(16 - 13(\ell(n)-1)^2)^2} \right)} = 1$$

and hence, by (2), $\lim_{n \rightarrow \infty} \frac{2 \log n}{-\log V_n} = 1$ which is the dimension of the underlying space. Again,

$$(3) \quad (6\ell(n))^2 V_{6(\ell(n)+1)} < n^2 V_n < (6(\ell(n) + 1))^2 V_{6\ell(n)}.$$

We have

$$\lim_{n \rightarrow \infty} (6\ell(n))^2 V_{6(\ell(n)+1)} = \lim_{\ell(n) \rightarrow \infty} (6\ell(n))^2 \frac{13(13\ell(n)^2 - 8\sqrt{13}\ell(n) + 16)}{12(16 - 13\ell(n)^2)^2} = 3,$$

and

$$\lim_{n \rightarrow \infty} (6(\ell(n) + 1))^2 V_{6\ell(n)} = \lim_{\ell(n) \rightarrow \infty} (6(\ell(n) + 1))^2 \frac{13(13(\ell(n) - 1)^2 - 8\sqrt{13}(\ell(n) - 1) + 16)}{12(16 - 13(\ell(n) - 1)^2)^2} = 3,$$

and hence, by (3) we have $\lim_{n \rightarrow \infty} n^2 V_n = 3$, i.e., the quantization coefficient exists as a finite positive number which equals 3. Thus, the proof of the proposition is complete. \square

Remark 3.12. Proceeding in the similar way, as it is done for the uniform distribution on the boundary of the regular hexagon, we can determine the optimal sets of n -means and the n th quantization error for any positive integer n for the uniform distribution on the boundary of any regular m -sided polygon for $m \geq 3$.

4. QUANTIZATION FOR A MIXED UNIFORM DISTRIBUTION ON A SEMICIRCULAR CURVE

The basic definitions and notations that were defined in the first paragraph of Section 3 are also true in this section.

We need the following proposition which generalizes a theorem in [RR2].

Proposition 4.1. *Let α_n be an optimal set of n -means for a uniform distribution on the unit circular arc S given by*

$$S := \{(\cos \theta, \sin \theta) : \alpha \leq \theta \leq \beta\},$$

where $0 \leq \alpha < \beta \leq 2\pi$. Then,

$$\alpha_n := \left\{ \frac{2n}{\beta - \alpha} \sin\left(\frac{\beta - \alpha}{2n}\right) \left(\cos\left(\alpha + (2j-1)\frac{\beta - \alpha}{2n}\right), \sin\left(\alpha + (2j-1)\frac{\beta - \alpha}{2n}\right) \right) : j = 1, 2, \dots, n \right\}$$

forms an optimal set of n -means, and the corresponding quantization error is given by

$$V_n = 1 - \frac{4n^2}{(\alpha - \beta)^2} \sin^2 \frac{\alpha - \beta}{2n}.$$

Proof. Notice that S is an arc of the unit circle $x_1^2 + x_2^2 = 1$ which subtends a central angle of $\beta - \alpha$ radian, and the probability distribution is uniform on S . Hence, the density function is given by $f(x_1, x_2) = \frac{1}{\beta - \alpha}$ if $(x_1, x_2) \in S$, and zero, otherwise. Thus, the proof follows in the similar way as a similar theorem in [RR2]. \square

Let L be the boundary of the semicircular disc $x_1^2 + x_2^2 = 1$, where $x_2 \geq 0$. Let the base of the semicircular disc be AOB , where A and B have the coordinates $(-1, 0)$ and $(1, 0)$, and O is the origin $(0, 0)$. Let s represent the distance of any point on L from the origin tracing along the boundary L in the counterclockwise direction. Notice that $L = L_1 \cup L_2$, where

$$\begin{aligned} L_1 &= \{(x_1, x_2) : x_1 = t, x_2 = 0 \text{ for } -1 \leq t \leq 1\}, \text{ and} \\ L_2 &= \{(x_1, x_2) : x_1 = \cos t, x_2 = \sin t \text{ for } 0 \leq t \leq \pi\}. \end{aligned}$$

Let P be the mixed uniform distribution defined on the boundary of the semicircular disc such that $P := \frac{1}{2}P_1 + \frac{1}{2}P_2$, where P_1 is the uniform distribution on the base L_1 of the semicircular disc, and P_2 is the uniform distribution on the semicircular arc L_2 . Thus, if f_1 and f_2 are the probability density functions for P_1 and P_2 , we have

$$f_1(x_1, x_2) = \begin{cases} \frac{1}{2} & \text{if } (x_1, x_2) \in L_1, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad f_2(x_1, x_2) = \begin{cases} \frac{1}{2\pi} & \text{if } (x_1, x_2) \in L_2, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, if f is the probability density function for the mixed distribution P , then we have $f = \frac{1}{2}f_1 + \frac{1}{2}f_2$, i.e., f is defined by

$$f(x_1, x_2) = \begin{cases} \frac{1}{4} & \text{if } (x_1, x_2) \in L_1, \\ \frac{1}{2\pi} & \text{if } (x_1, x_2) \in L_2, \\ 0 & \text{otherwise.} \end{cases}$$

On both L_1 and L_2 , we have $ds = \sqrt{\left(\frac{dx_1}{dt}\right)^2 + \left(\frac{dx_2}{dt}\right)^2} dt = dt$ yielding $dP(s) = P(ds) = f(x_1, x_2)ds = f(x_1, x_2)dt$.

Lemma 4.2. *Let X be a continuous random variable with the mixed distribution P taking values on L . Then,*

$$E(X) = \left(0, \frac{1}{\pi}\right) \text{ and } V := V(X) = \frac{2}{3} - \frac{1}{\pi^2}.$$

Proof. We have,

$$E(X) = \int_L (x_1 i + x_2 j) dP = \frac{1}{4} \int_{L_1} (t, 0) dt + \frac{1}{2\pi} \int_{L_2} (\cos t, \sin t) dt = \left(0, \frac{1}{\pi}\right).$$

To calculate the variance, we proceed as follows:

$$\begin{aligned} V(X) &= E\|X - E(X)\|^2 = \int_L \left((x_1 - 0)^2 + (x_2 - \frac{1}{\pi})^2 \right) dP \\ &= \frac{1}{4} \int_{-1}^1 \left((t - 0)^2 + (0 - \frac{1}{\pi})^2 \right) dt + \frac{1}{2\pi} \int_0^\pi \left((\cos t - 0)^2 + (\sin t - \frac{1}{\pi})^2 \right) dt \\ &= \frac{2}{3} - \frac{1}{\pi^2}. \end{aligned}$$

Hence the lemma. \square

Remark 4.3. Proceeding similarly as Remark 3.3, we see that the optimal set of one-mean is the set $\{(0, \frac{1}{\pi})\}$, and the corresponding quantization error is the variance $V := V(X)$ of the random variable X .

In the following proposition we give the optimal sets of two-means.

Proposition 4.4. *Let P be the mixed distribution on the boundary of the semicircle. Then, the set $\{(-\frac{1}{4} - \frac{1}{\pi}, \frac{1}{\pi}), (\frac{1}{4} + \frac{1}{\pi}, \frac{1}{\pi})\}$ forms the optimal set of two-means, and the quantization error for two-means is given by*

$$V_2 = 2 \left(\frac{96 - 24\pi + 7\pi^2}{192\pi^2} + \frac{-96 - 8\pi + 17\pi^2}{64\pi^2} \right) = 0.242369.$$

Proof. Let the points P and Q form an optimal set of two-means. Let ℓ be the boundary of their Voronoi regions. The following two cases can arise:

Case 1. ℓ intersects both L_1 and L_2 .

Let ℓ intersect L_1 and L_2 at the points D and E , respectively. Let the points D and E be given by the parameters $t = \alpha$ and $t = \beta$. Let P and Q be the conditional expectations of the random variable X given that X takes values on the boundaries $DB \cup \widehat{BE}$ and $\widehat{EA} \cup AD$, respectively. Then, after some calculations, we have

$$\begin{aligned} \tilde{p} &= E(X : X \in DB \cup \widehat{BE}) = \frac{\frac{1}{4} \int_\alpha^1 (t, 0) dt + \frac{1}{2\pi} \int_0^\beta (\cos t, \sin t) dt}{\frac{1}{4} \int_\alpha^1 1 dt + \frac{1}{2\pi} \int_0^\beta 1 dt} \\ &= \left(\frac{\frac{1}{4}(\frac{1}{2} - \frac{\alpha^2}{2}) + \frac{\sin \beta}{2\pi}}{\frac{1-\alpha}{4} + \frac{\beta}{2\pi}}, \frac{1 - \cos \beta}{2\pi(\frac{1-\alpha}{4} + \frac{\beta}{2\pi})} \right), \end{aligned}$$

and similarly

$$\tilde{q} = E(X : X \in \widehat{EA} \cup AD) = \left(\frac{\frac{1}{4}(\frac{\alpha^2}{2} - \frac{1}{2}) - \frac{\sin \beta}{2\pi}}{\frac{\alpha+1}{4} + \frac{\pi-\beta}{2\pi}}, \frac{\cos \beta + 1}{2\pi(\frac{\alpha+1}{4} + \frac{\pi-\beta}{2\pi})} \right).$$

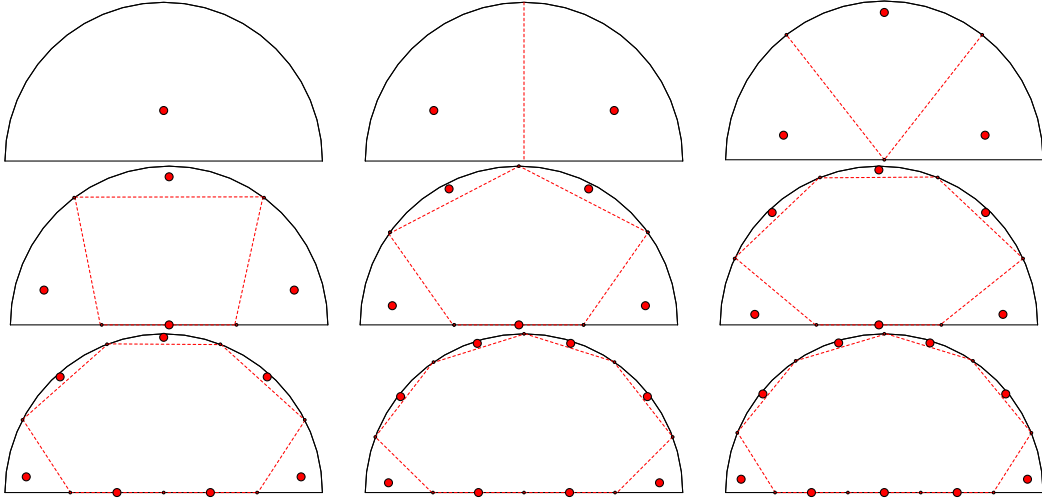
Since P and Q form an optimal set of two-means, and DE is the boundary of their corresponding Voronoi regions, we have the canonical equations as $\rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0$, and $\rho(\tilde{e}, \tilde{p}) - \rho(\tilde{e}, \tilde{q}) = 0$. Put the values of $\tilde{p}, \tilde{q}, \tilde{d}$ and \tilde{e} , and then solving the two equations in α and β , we have $\alpha = 0$ and $\beta = \frac{\pi}{2}$ implying

$$\tilde{p} = \left(\frac{1}{4} + \frac{1}{\pi}, \frac{1}{\pi} \right), \text{ and } \tilde{q} = \left(-\frac{1}{4} - \frac{1}{\pi}, \frac{1}{\pi} \right),$$

and the corresponding distortion V_2 (Case 1) error, due to a symmetry, is given by

$$\begin{aligned} V_2(\text{Case 1}) &= 2 \left(\frac{1}{4} \int_\alpha^1 \rho((t, 0), \tilde{p}) dt + \frac{1}{2\pi} \int_\beta^\pi \rho((\cos t, \sin t), \tilde{p}) dt \right) \\ &= 2 \left(\frac{96 - 24\pi + 7\pi^2}{192\pi^2} + \frac{-96 - 8\pi + 17\pi^2}{64\pi^2} \right) = 0.242369. \end{aligned}$$

Case 2. ℓ intersects L_2 at two points.

FIGURE 3. Points in an optimal set of n -means for $1 \leq n \leq 9$.

Let ℓ intersect L_2 at the points D and E , respectively. As before, there exist parameters $t = \alpha$ and $t = \beta$, for which we have

$$\tilde{d} = (\cos \alpha, \sin \alpha), \text{ and } \tilde{e} = (\cos \beta, \sin \beta).$$

Let P and Q be the conditional expectations of the random variable X given that X takes values on the boundary above and below the line ℓ , respectively. Then,

$$\tilde{p} = E(X : X \in \widehat{DE}), \text{ and } \tilde{q} = E(X : X \in \widehat{EA} \cup \widehat{AB} \cup \widehat{BD}).$$

Proceeding in the similar way as Case 1, we calculate \tilde{p} and \tilde{q} , and obtain the canonical equations $\rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0$, and $\rho(\tilde{e}, \tilde{p}) - \rho(\tilde{e}, \tilde{q}) = 0$. Solving the above two equations in α and β , we have $\alpha = 0.436587$, and $\beta = \pi - \alpha$, i.e., the line ℓ is parallel to the base AOB of the semicircle, yielding

$$\tilde{p} = \left(0, \frac{2 \cos \alpha}{\pi - 2\alpha}\right) = (0, 0.798971), \text{ and } \tilde{q} = \left(0, \frac{1 - \cos \alpha}{\pi(\frac{\alpha}{\pi} + \frac{1}{2})}\right) = (0, 0.0467274),$$

and the corresponding distortion $V_2(\text{Case 2})$ is given by

$$V_2(\text{Case 2}) = 2 \left(\frac{1}{2\pi} \int_{\beta}^{\frac{\pi}{2}} \rho((\cos t, \sin t), \tilde{p}) dt + \frac{1}{4} \int_0^1 \rho((t, 0), \tilde{q}) dt + \frac{1}{2\pi} \int_0^{\beta} \rho((\cos t, \sin t), \tilde{q}) dt \right)$$

implying $V_2(\text{Case 2}) = 0.434806$.

Since $V_2(\text{Case 2}) > V_2(\text{Case 1})$, the points in Case 1 form the optimal set of two-means, and $V_2(\text{Case 1})$ is the quantization error for two-means (see Figure 3). Thus, the proof of the proposition is complete. \square

Proposition 4.5. *The set $\{(-0.634868, 0.15471), (0, 0.92798), (0.634868, 0.15471)\}$ forms an optimal set of three-means, and the quantization error for three-means is given by $V_3 = 0.147821$.*

Proof. Let the set $\alpha := \{p, q, r\}$ be an optimal set of three-means. Two cases can arise:

Case 1. α contains a point from the base AOB .

Recall that P is defined by $P = \frac{1}{2}P_1 + \frac{1}{2}P_2$, i.e., the probability distribution P is the uniform mixture of two uniform distributions, and the semicircle is symmetric about its vertical axis. Thus, if α contains a point from the base, we can assume that $O(0, 0) \in \alpha$, and the boundaries of the other two points cut the boundary of the semicircle at the points $F(-a, 0)$, $G(a, 0)$, and $H(0, 1)$, where $0 < a < 1$. Thus, we can assume that $p = E(X : X \in \widehat{FG}) = (0, 0)$, and $q = E(X : X \in \widehat{GB} \cup \widehat{BH})$, and $r = E(X : X \in \widehat{HA} \cup \widehat{AF})$. Solving the canonical

equation $\rho(\tilde{g}, p) - \rho(\tilde{g}, q) = 0$, we obtain $a = 0.462946$ yielding $q = (0.669762, 0.414182)$, and $r = (-0.669762, 0.414182)$. If $V_3(\text{Case 1})$ is the corresponding distortion error, we have

$$V_3(\text{Case 1}) = 2\left(\frac{1}{4} \int_0^a \rho((t, 0), p) dt + \frac{1}{4} \int_a^1 \rho((t, 0), q) dt + \frac{1}{2\pi} \int_0^{\frac{\pi}{2}} \rho((\cos t, \sin t), q) dt\right)$$

yielding $V_3(\text{Case 1}) = 0.190082$.

Case 2. α does not contain any point from the base AOB .

Due to symmetry and the uniform mixture of two uniform distributions, in this case we can assume that the points in α cut the boundary of the semicircle at the points $O(0, 0)$, $G(\cos b, \sin b)$, and $H(-\cos b, \sin b)$, where $0 < b < \frac{\pi}{2}$. Thus, we can assume that $p = E(X : X \in OB \cup \widehat{BG})$, and $q = E(X : X \in \widehat{GH})$, and $r = E(X : X \in \widehat{HA} \cup AO)$. Solving the canonical equation $\rho(\tilde{g}, p) - \rho(\tilde{g}, q) = 0$, we obtain $b = 0.906133$ yielding $p = (0.634868, 0.15471)$, $q = (0, 0.92798)$, and $r = (-0.634868, 0.15471)$. If $V_3(\text{Case 2})$ is the corresponding distortion error, we have

$$V_3(\text{Case 2}) = 2\left(\frac{1}{4} \int_0^1 \rho((t, 0), p) dt + \frac{1}{2\pi} \int_0^b \rho((\cos t, \sin t), p) dt\right) + \frac{1}{2\pi} \int_b^{\pi-b} \rho((\cos t, \sin t), q) dt$$

yielding $V_3(\text{Case 2}) = 0.147821$.

Since $V_3(\text{Case 1}) > V_3(\text{Case 2})$, the points in Case 2 form the optimal set of three-means, and $V_3(\text{Case 2})$ is the quantization error for three-means (see Figure 3). Thus, the proof of the proposition is complete. \square

Let us now state the following proposition which gives the optimal sets of n -means for $4 \leq n \leq 9$ for the mixed distribution on the boundary of the semicircle. The proof follows in the similar way as the previous lemma by considering the different cases.

Proposition 4.6. *Let P be the mixed distribution on the boundary of the semicircle. Then:*

- (i) *the set $\{(0, 0), (0.788235, 0.219171), (0, 0.932871), (-0.788235, 0.219171)\}$ forms an optimal set of four-means with quantization error $V_4 = 0.098412$;*
- (ii) *the set $\{(0, 0), (0.79719, 0.120767), (0.439705, 0.856689), (-0.439705, 0.856689), (-0.79719, 0.120767)\}$ forms an optimal set of five-means with quantization error $V_5 = 0.0654358$;*
- (iii) *the set $\{(0, 0), (0.781728, 0.0661158), (0.672351, 0.707636), (0, 0.976117), (-0.672351, 0.707636), (-0.781728, 0.0661158)\}$ forms an optimal set of six-means with quantization error $V_6 = 0.0499565$;*
- (iv) *the set $\{(-0.294267, 0), (0.294267, 0), (0.865678, 0.0989137), (0.65226, 0.728637), (0, 0.977935), (-0.65226, 0.728637), (-0.865678, 0.0989137)\}$ forms an optimal set of seven-means with quantization error $V_7 = 0.0366668$;*
- (v) *the set $\{(-0.286766, 0), (0.286766, 0), (0.853609, 0.0615721), (0.777386, 0.604469), (0.294109, 0.939793), (-0.294109, 0.939793), (-0.777386, 0.604469), (-0.853609, 0.0615721)\}$ forms an optimal set of eight-means with quantization error $V_8 = 0.0290573$;*
- (vi) *the set $\{(-0.458992, 0), (0, 0), (0.458992, 0), (0.902056, 0.0840085), (0.764954, 0.621235), (0.287671, 0.942514), (-0.287671, 0.942514), (-0.764954, 0.621235), (-0.902056, 0.0840085)\}$ forms an optimal set of nine-means with quantization error $V_9 = 0.0233983$ (see Figure 3).*

Remark 4.7. Let α_n be an optimal set of n -means for P for $n \geq 4$. From the above proposition we see that for $4 \leq n \leq 9$ there exists a positive integers k , and two positive numbers a and b depending on k , such that if $m = n - k - 2$, then α_n contains k quantizers which occur due to the uniform distribution on the closed interval $[-a, a]$, and m quantizers which occur due to the uniform distribution on the semicircular arc $\{(\cos \theta, \sin \theta) : b \leq \theta \leq \pi - b\}$. It can be proved that this fact is also true for any positive integer $n \geq 10$. In addition, α_n contains two quantizers which are in the interior of the angles formed by the base AOB and the semicircular arc \widehat{BA} , the Voronoi regions of these two points contain elements from both the base and the

semicircular arc. Due to too much technicality, we skip the proof of it in the paper. The two real numbers a and b are obtained by solving the two canonical equations

$$(4) \quad \rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0, \text{ and } \rho(\tilde{e}, \tilde{q}) - \rho(\tilde{e}, \tilde{r}) = 0,$$

where

$$\left\{ \begin{array}{l} \tilde{d} = (a, 0), e = (\cos b, \sin b), \tilde{p} = (a - \frac{a}{k}, 0), \\ \tilde{q} = \frac{\frac{1}{4} \int_a^1 (t, 0) dt + \frac{1}{2\pi} \int_0^b (\cos t, \sin t) dt}{\frac{1}{4} \int_a^1 1 dt + \frac{1}{2\pi} \int_0^b 1 dt} = \left(\frac{-\pi a^2 + 4 \sin b + \pi}{8\pi \left(\frac{1-a}{4} + \frac{b}{2\pi}\right)}, \frac{\sin^2(\frac{b}{2})}{\pi \left(\frac{1-a}{4} + \frac{b}{2\pi}\right)} \right), \text{ and} \\ \tilde{r} = \frac{\int_b^{\frac{b(m-2)+\pi}{m}} (\cos t, \sin t) dt}{\int_b^{\frac{b(m-2)+\pi}{m}} 1 dt} = \left(\frac{m(\sin(\frac{b(m-2)+\pi}{m}) - \sin b)}{\pi - 2b}, \frac{m(\cos b - \cos(\frac{b(m-2)+\pi}{m}))}{\pi - 2b} \right). \end{array} \right.$$

Let us now give the following theorem.

Theorem 4.8. *Let α_n be an optimal set of n -means for $n \geq 4$ such that α_n contains $k := k(n)$ elements from the base of the semicircular disc. Then,*

$$\begin{aligned} \alpha_n := & \left\{ \left(-a + \frac{2j-1}{k}a, 0\right) : 1 \leq j \leq k \right\} \\ & \cup \left\{ \frac{2m}{\pi-2b} \sin\left(\frac{\pi-2b}{2m}\right) \left(\cos\left(b + (2j-1)\frac{\pi-2b}{2m}\right), \sin\left(b + (2j-1)\frac{\pi-2b}{2m}\right) \right) : 1 \leq j \leq m \right\} \\ & \cup \{(r, s), (-r, s)\}, \end{aligned}$$

where $m = n - k - 2$, $r = \frac{-\pi a^2 + 4 \sin b + \pi}{8\pi \left(\frac{1-a}{4} + \frac{b}{2\pi}\right)}$, $s = \frac{\sin^2(\frac{b}{2})}{\pi \left(\frac{1-a}{4} + \frac{b}{2\pi}\right)}$, and the two positive real numbers a and b are determined by the equations in (4), and the quantization error for n -means is given by

$$\begin{aligned} V_n = & \frac{1}{24} \left(\frac{4a^3}{k^2} + \frac{12(2m^2 \cos(\frac{\pi-2b}{m}) + (\pi-2b)^2 - 2m^2)}{\pi(\pi-2b)^2} \right. \\ & - \frac{1}{\pi(\pi(a-1)-2b)} \left(\pi^2 a^4 - 8\pi a^3 b - 4\pi^2 a^3 + 24\pi(a^2-1) \sin b + 6\pi^2 a^2 - 24\pi ab \right. \\ & \left. \left. - 4\pi^2 a + 48b^2 + 32\pi b + 96 \cos b + \pi^2 - 96 \right) \right). \end{aligned}$$

Proof. Let γ_n be the set of $k := k(n)$ quantizers that α_n contains from the closed interval $[-a, a]$. Then, by [RR2],

$$\gamma_n = \left\{ \left(-a + \frac{2i-1}{k}a, 0\right) : 1 \leq i \leq k \right\},$$

and the distortion error due to the set γ_n is given by $\frac{a^3}{6k^2}$. Let δ_n be the set of m quantizers which occur due to the uniform distribution on the circular arc $\{(\cos \theta, \sin \theta) : b \leq \theta \leq \pi - b\}$. Then, by Proposition 4.1, we have

$$\delta_n := \left\{ \frac{2m}{\pi-2b} \sin\left(\frac{\pi-2b}{2m}\right) \left(\cos\left(b + (2j-1)\frac{\pi-2b}{2m}\right), \sin\left(b + (2j-1)\frac{\pi-2b}{2m}\right) \right) : 1 \leq j \leq m \right\},$$

and the corresponding distortion error is given by

$$\begin{aligned} & \frac{m}{2\pi} \int_b^{b+\frac{\pi-2b}{m}} \rho\left((\cos t, \sin t), \frac{2m}{\pi-2b} \sin\left(\frac{\pi-2b}{2m}\right) \left(\cos\left(b + \frac{\pi-2b}{2m}\right), \sin\left(b + \frac{\pi-2b}{2m}\right) \right)\right) dt \\ & = \frac{2m^2 \cos\left(\frac{\pi-2b}{m}\right) + (\pi-2b)^2 - 2m^2}{2\pi(\pi-2b)^2}. \end{aligned}$$

As mentioned in Remark 4.7, let (r, s) be the point in α_n which lies in the interior of the right hand angle formed by the base and the semicircular arc. Then,

$$(r, s) = \frac{\frac{1}{4} \int_a^1 (t, 0) dt + \frac{1}{2\pi} \int_0^b (\cos t, \sin t) dt}{\frac{1}{4} \int_a^1 1 dt + \frac{1}{2\pi} \int_0^b 1 dt} = \left(\frac{-\pi a^2 + 4 \sin b + \pi}{8\pi \left(\frac{1-a}{4} + \frac{b}{2\pi}\right)}, \frac{\sin^2(\frac{b}{2})}{\pi \left(\frac{1-a}{4} + \frac{b}{2\pi}\right)} \right).$$

Due to symmetry, the point in α_n which lies in the interior of the left hand angle formed by the base and the semicircular arc is given by $(-r, s)$, where $r = \frac{-\pi a^2 + 4 \sin b + \pi}{8\pi(\frac{1-a}{4} + \frac{b}{2\pi})}$, and $s = \frac{\sin^2(\frac{b}{2})}{\pi(\frac{1-a}{4} + \frac{b}{2\pi})}$. The distortion error due to these two corner points is given by

$$\begin{aligned} & 2\left(\frac{1}{4} \int_a^1 \rho((t, 0), (r, s)) dt + \frac{1}{2\pi} \int_0^b \rho((\cos t, \sin t), (r, s)) dt\right) \\ &= -\frac{1}{24\pi(\pi(a-1) - 2b)}(\pi^2 a^4 - 8\pi a^3 b - 4\pi^2 a^3 + 24\pi(a^2 - 1) \sin b + 6\pi^2 a^2 - 24\pi ab \\ &\quad - 4\pi^2 a + 48b^2 + 32\pi b + 96 \cos b + \pi^2 - 96) \end{aligned}$$

Taking the union of γ_n , δ_n , and the set $\{(r, s), (-r, s)\}$, we obtain α_n , and summing up the corresponding distortion errors, we obtain the quantization error V_n . Thus, the proof of the theorem is complete. \square

Remark 4.9. Let $n \in \mathbb{N}$ be such that $n \geq 4$. Then, the positive integer k , given in Theorem 4.8, depends on n . If k is known, using Theorem 4.8, one can easily determine the optimal set α_n and the corresponding quantization error.

Let us now give the following definition.

Definition 4.10. Define the sequence $\{a(n)\}$ such that $a(n) = \lfloor n(\sqrt{2} - 1) \rfloor$ for $n \geq 1$, i.e.,

$$\{a(n)\}_{n=1}^\infty = \{0, 0, 1, 1, 2, 2, 2, 3, 3, 4, 4, 4, 5, 5, 6, 6, 7, 7, 7, 8, 8, 9, 9, 9, 10, 10, 11, 11, 12, 12, \dots\},$$

where $\lfloor x \rfloor$ represents the greatest integer not exceeding x .

The following algorithm helps us to determine the exact value of k mentioned in Theorem 4.8.

4.11. **Algorithm.** Let $n \geq 4$, and let $V(n, k) := V_n$, as given by Theorem 4.8, denote the distortion error if α_n contains k elements from the base of the semicircular disc. Let $\{a(n)\}$ be the sequence defined by Definition 4.10. Then, the algorithm runs as follows:

- (i) Write $k := a(n)$.
- (ii) If $k = 1$ go to step (v), else step (iii).
- (iii) If $V(n, k - 1) < V(n, k)$ replace k by $k - 1$ and go to step (ii), else step (iv).
- (iv) If $V(n, k + 1) < V(n, k)$ replace k by $k + 1$ and return, else step (v).
- (v) End.

When the algorithm ends, then the value of k , obtained, is the exact value of k that an optimal set α_n contains from the base of the semicircular disc.

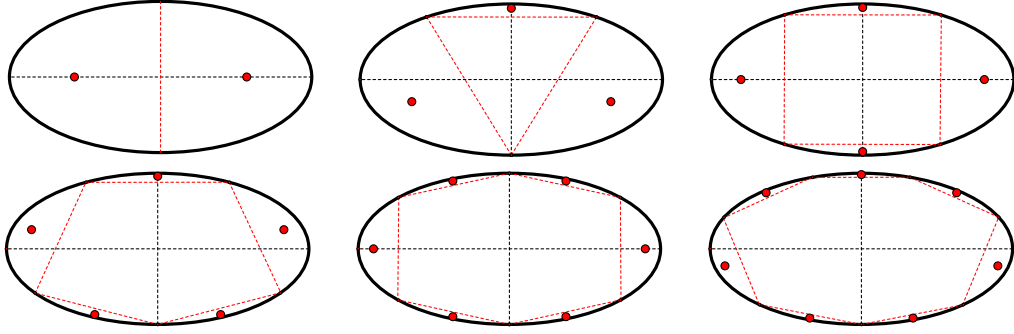
Remark 4.12. If $n = 40$, then $a(n) = 16$, and by the algorithm we also obtain $k = 16$; if $n = 51$, then $a(n) = 21$, and by the algorithm we also obtain $k = 21$. If $n = 1000$, then $a(n) = 414$, and by the algorithm, we obtain $k = 424$; if $n = 2500$, then $a(n) = 1035$, and by the algorithm, we obtain $k = 1042$; and if $n = 5000$, then $a(n) = 2071$, and by the algorithm, we obtain $k = 2083$. Thus, we see that with the help of the sequence and the algorithm we can easily determine the exact value of k for any positive integer $n \geq 4$.

The following questions still remain open.

4.13. **Open.** Can one estimate the positive integer k in terms of n for all large n ? Further, from such estimate can one obtain the asymptotics of the quantization error, and even obtain the existence of quantization dimension and quantization coefficient?

5. QUANTIZATION FOR A UNIFORM DISTRIBUTION ON AN ELLIPTICAL CURVE

The basic definitions and notations that were defined in the first paragraph of Section 3 are also used in this section. As a prototype, we take the equation of the ellipse as $x_1^2 + 4x_2^2 = 4$, whose center is $O(0, 0)$, and the lengths of the major and the minor axes are, respectively, 2 and 1. By the elliptical curve, denoted by L , we mean the boundary of the ellipse $x_1^2 + 4x_2^2 = 4$.

FIGURE 4. Points in an optimal set of n -means for $2 \leq n \leq 7$.

Let L intersect the positive and negative directions of the x_1 -axis at the points A_1 and A_3 , and the positive and negative directions of the x_2 -axis at the points A_2 and A_4 , respectively. Let P be the uniform distribution defined on L . Notice that the parametric equations of L are given by $x_1 = 2 \cos \theta$, and $x_2 = \sin \theta$ for $0 \leq \theta \leq 2\pi$. Let s represent the distance of any point on L from the point A_1 tracing along the boundary L in the counterclockwise direction. Then, $ds = \sqrt{dx_1^2 + dx_2^2} = \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta$. Thus, the length of L is given by

$$\int_L ds = \int_0^{2\pi} \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta = 9.6884482205 \text{ (up to ten decimal places).}$$

In the sequel, write $A := 9.6884482205$. Hence, the probability density function (pdf) f of the uniform distribution P is given by $f(s) := f(x_1, x_2) = \frac{1}{A}$ for all $(x_1, x_2) \in L$, and zero, otherwise. Again, $dP(s) = P(ds) = f(x_1, x_2)ds = \frac{1}{A} \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta$.

Let us now prove the following lemma.

Lemma 5.1. *Let X be a continuous random variable with uniform distribution on L . Then, $E(X) = (0, 0)$ and $V := V(X) = 2.260230080$.*

Proof. We have,

$$E(X) = \int_L (x_1 i + x_2 j) dP = \frac{1}{A} \int_0^{2\pi} \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta = (0, 0),$$

and

$$\begin{aligned} V(X) &= E\|X - E(X)\|^2 = \int_L \rho((x_1, x_2), E(X)) dP = \frac{1}{A} \int_L \rho((2 \cos \theta, \sin \theta), (0, 0)) ds \\ &= \frac{1}{A} \int_0^{2\pi} \sqrt{4 \sin^2 \theta + \cos^2 \theta} \rho((2 \cos \theta, \sin \theta), (0, 0)) d\theta = 2.260230080. \end{aligned}$$

Hence, the proof of the lemma is complete. \square

Remark 5.2. Proceeding similarly as Remark 3.3, we see that the optimal set of one-mean is the set $\{(0, 0)\}$, and the corresponding quantization error is the variance $V := V(X)$ of the random variable X .

Remark 5.3. The ellipse has two lines of symmetry: the major axis and the minor axis, and the probability distribution is uniform. To calculate the optimal sets of n -means for any positive integer $n \geq 2$, we will use this information. This will help us to avoid too much technicality in the proof of the following propositions.

Proposition 5.4. *The optimal set of two-means is $\{(1.13964, 0), (-1.13964, 0)\}$ with quantization error $V_2 = 0.961441$.*

Proof. Let $\alpha := \{\tilde{p}, \tilde{q}\}$ be an optimal set of two-means. Due to Remark 5.3, we can assume that the boundary of their Voronoi regions passes through the center of the ellipse, in other words, we can assume that the boundary of the Voronoi regions cut the ellipse at the two points D and E given by the parameters $\theta = b$ and $\theta = \pi + b$, respectively, where $0 \leq b \leq \pi$. Then, we have $\tilde{d} = (2 \cos b, \sin b)$, and

$$\tilde{p} = \frac{\int_b^{\pi+b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_b^{\pi+b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}, \text{ and } \tilde{q} = \frac{\int_{\pi+b}^{2\pi+b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_{\pi+b}^{2\pi+b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}.$$

Solving the canonical equation $\rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0$, we have $b = 0, \frac{\pi}{2}, \pi$. Notice that $b = 0$ and $b = \pi$ are reflections of each other about the origin. Thus, the following two cases can arise:

Case 1. $b = 0$.

In this case, we have $\tilde{p} = (0, 0.705665)$, and $\tilde{q} = (0, -0.705665)$ with the distortion error

$$\int \min_{a \in \alpha} \|x - a\|^2 dP = \frac{2}{A} \int_0^\pi \rho((2 \cos \theta, \sin \theta), (0, 0.705665)) ds = 1.76227.$$

Case 2. $b = \frac{\pi}{2}$.

In this case, we have $\tilde{p} = (-1.13964, 0)$, and $\tilde{q} = (1.13964, 0)$ with the distortion error

$$\int \min_{a \in \alpha} \|x - a\|^2 dP = \frac{2}{A} \int_{\frac{\pi}{2}}^{\pi+\frac{\pi}{2}} \rho((2 \cos \theta, \sin \theta), (-1.13964, 0)) ds = 0.961441.$$

Comparing the distortion errors, we see that the set $\{(1.13964, 0), (-1.13964, 0)\}$ forms the optimal set of two-means (see Figure 4) with quantization error $V_2 = 0.961441$, which yields the proposition. \square

In the following two propositions we state and prove the optimal sets of six- and seven-means.

Proposition 5.5. *The optimal set of six-means is*

$$\begin{aligned} & \{(1.7975, 0), (0.748205, 0.897609), (-0.748205, 0.897609), \\ & (-1.7975, 0), (-0.748205, -0.897609), (0.748205, -0.897609)\} \end{aligned}$$

with quantization error $V_6 = 0.198794$.

Proof. Let α be an optimal set of six-means. Due to Remark 5.3, the following two cases can arise:

Case 1. Two points of α are on the major axis, two are above and two are below the major axis.

Again, due to Remark 5.3, the two points which are on the major axis are reflections of each other with respect to the origin, and the set of points below are the reflections of the set of points above with respect to the major axis. Let the point which lie on the positive direction of the major axis be \tilde{p} , and the two points above the major axis be given by \tilde{q} , and \tilde{r} . Let the boundary of the Voronoi regions of \tilde{p} and \tilde{q} be given by \tilde{d} with parameter $\theta = b$, and the boundary of the Voronoi regions of \tilde{q} and \tilde{r} be given by \tilde{e} with parameter $\theta = c$. Then, we have

$$\begin{aligned} \tilde{d} &= (2 \cos b, \sin b), \quad \tilde{e} = (2 \cos c, \sin c), \quad \tilde{p} = \frac{\int_{2\pi-b}^{2\pi+b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_{2\pi-b}^{2\pi+b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}, \\ \tilde{q} &= \frac{\int_b^c \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_b^c \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}, \text{ and } \tilde{r} = \frac{\int_c^{\pi-b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_c^{\pi-b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}. \end{aligned}$$

Solving the canonical equations $\rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0$ and $\rho(\tilde{e}, \tilde{q}) - \rho(\tilde{e}, \tilde{r}) = 0$, we obtain $b = 0.745647$, and $c = 1.5708$, yielding

$$\{\tilde{p}, \tilde{q}, \tilde{r}\} = \{(1.7975, 0), (0.748205, 0.897609), (-0.748205, 0.897609)\}.$$

Thus, due to reflection, we can obtain all the elements of α , and the corresponding distortion error is given by

$$\int \min_{a \in \alpha} \|x - a\|^2 dP = 2(\text{distortion error due to } \tilde{p}, \tilde{q}, \tilde{r}) = 0.198794.$$

Case 2. Two points of α are on the minor axis, two are to the right and two are to the left of the minor axis.

Due to Remark 5.3, the two points which are on the minor axis are reflections of each other with respect to the origin, and the set of points to the right are the reflections of the set of points to the left with respect to the minor axis. Let the point which lie on the positive direction of the minor axis be \tilde{p} , and the two points to the left of the minor axis be given by \tilde{q} , and \tilde{r} . Let the boundary of the Voronoi regions of \tilde{p} and \tilde{q} be given by \tilde{d} with parameter $\theta = b$, and the boundary of the Voronoi regions of \tilde{q} and \tilde{r} be given by \tilde{e} with parameter $\theta = c$. Then, we have

$$\begin{aligned} \tilde{d} &= (2 \cos b, \sin b), \quad \tilde{e} = (2 \cos c, \sin c), \quad \tilde{p} = \frac{\int_{\pi-b}^b \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_{\pi-b}^b \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}, \\ \tilde{q} &= \frac{\int_c^b \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_c^b \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}, \quad \text{and} \quad \tilde{r} = \frac{\int_c^{2\pi-b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_c^{2\pi-b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}. \end{aligned}$$

Solving the canonical equations $\rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0$ and $\rho(\tilde{e}, \tilde{q}) - \rho(\tilde{e}, \tilde{r}) = 0$, we obtain $b = 1.97683$, and $c = 3.14159$, yielding

$$\{\tilde{p}, \tilde{q}, \tilde{r}\} = \{(0, 0.973196), (-1.50317, 0.57481), (-1.50317, -0.57481)\}.$$

Thus, due to reflection, we can obtain all the elements of α , and the corresponding distortion error is given by

$$\int \min_{a \in \alpha} \|x - a\|^2 dP = 2(\text{distortion error due to } \tilde{p}, \tilde{q}, \tilde{r}) = 0.209898.$$

Comparing the distortion errors in Case 1 and Case 2, we see that the set α in Case 1 forms the optimal set of six-means (see Figure 4) with quantization error $V_6 = 0.198794$. Thus, the proof of the proposition is complete. \square

Proposition 5.6. *There are two different optimal sets of seven-means, one of them is*

$$\begin{aligned} \{(0, 0.98224), (-1.26066, 0.741926), (-1.80513, -0.225347), (-0.68335, -0.916091), \\ (0.68335, -0.916091), (1.80513, -0.225347), (1.26066, 0.741926)\} \end{aligned}$$

with quantization error $V_7 = 0.152179$.

Proof. Let α be an optimal set of seven-means. Two cases can arise:

Case 1. α contains a point from the major axis, three points from above the major axis, and the other three points from below the major axis.

Due to Remark 5.3, in this case, we can assume that the three points in α which are below the major axis are the reflections of the three points above the major axis. Let α contain a point, denoted \tilde{p} , from the positive direction of the major axis. Let the points in α which are above the major axis be \tilde{q}, \tilde{r} , and \tilde{s} . Due to symmetry $(-2, 0)$ is a boundary point of the Voronoi regions. Let the boundary points of the Voronoi regions of \tilde{q}, \tilde{r} , and \tilde{s} be $\tilde{d}, \tilde{e}, \tilde{f}$, and $(-2, 0)$,

respectively, given by the parametric values $\theta = b, \theta = c, \theta = d$, and $\theta = \pi$. Then, we have

$$\begin{aligned}\tilde{d} &= (2 \cos b, \sin b), \quad \tilde{e} = (2 \cos c, \sin c), \quad \tilde{f} = (2 \cos d, \sin d), \\ \tilde{p} &= \frac{\int_{2\pi-b}^{2\pi+b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_{2\pi-b}^{2\pi+b} \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}, \quad \tilde{q} = \frac{\int_b^c \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_b^c \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}, \\ \tilde{r} &= \frac{\int_c^d \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_c^d \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}, \quad \text{and } \tilde{s} = \frac{\int_d^\pi \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_d^\pi \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}.\end{aligned}$$

Solving the canonical equations $\rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0$, $\rho(\tilde{e}, \tilde{q}) - \rho(\tilde{e}, \tilde{r}) = 0$, and $\rho(\tilde{f}, \tilde{r}) - \rho(\tilde{f}, \tilde{s}) = 0$ we obtain $b = 0.661475, c = 1.42154$, and $d = 2.11384$, which give the set $\{\tilde{p}, \tilde{q}, \tilde{r}, \tilde{s}\}$ equals

$$\{(1.84197, 0), (0.953451, 0.852294), (-0.372194, 0.962574), (-1.6113, 0.517322)\},$$

and the other three points in α are the reflections of $\tilde{q}, \tilde{r}, \tilde{s}$ with respect to the major axis. The corresponding distortion error is given by

$$\int \min_{a \in \alpha} \|x - a\|^2 dP = \text{distortion error due to } \tilde{p} + 2(\text{distortion error due to } \tilde{q}, \tilde{r}, \tilde{s}) = 0.152488.$$

Case 2. α contains a point from the minor axis, three points are to the left of the minor axis, and the other three points are the reflections with respect to the minor axis.

Let α contain a point, denoted by \tilde{p} , from the positive direction of the minor axis. Due to symmetry $(0, -1)$ is a boundary point of the Voronoi regions. Let the points in α which are to the left of the minor axis be \tilde{q}, \tilde{r} , and \tilde{s} . Let the boundary points of their Voronoi regions be $\tilde{d}, \tilde{e}, \tilde{f}$, and $(0, -1)$, respectively, given by the parametric values $\theta = b, \theta = c, \theta = d$, and $\theta = \frac{3\pi}{2}$. Then, we have

$$\begin{aligned}\tilde{d} &= (2 \cos b, \sin b), \quad \tilde{e} = (2 \cos c, \sin c), \quad \tilde{f} = (2 \cos d, \sin d), \\ \tilde{p} &= \frac{\int_{\pi-b}^b \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_{\pi-b}^b \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}, \quad \tilde{q} = \frac{\int_b^c \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_b^c \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}, \\ \tilde{r} &= \frac{\int_c^d \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_c^d \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}, \quad \text{and } \tilde{s} = \frac{\int_d^{\frac{3\pi}{2}} \sqrt{4 \sin^2 \theta + \cos^2 \theta} (2 \cos \theta, \sin \theta) d\theta}{\int_d^{\frac{3\pi}{2}} \sqrt{4 \sin^2 \theta + \cos^2 \theta} d\theta}.\end{aligned}$$

Solving the canonical equations $\rho(\tilde{d}, \tilde{p}) - \rho(\tilde{d}, \tilde{q}) = 0$, $\rho(\tilde{e}, \tilde{q}) - \rho(\tilde{e}, \tilde{r}) = 0$, and $\rho(\tilde{f}, \tilde{r}) - \rho(\tilde{f}, \tilde{s}) = 0$, we obtain $b = 1.8999, c = 2.71024$, and $d = 3.97294$, which give the set $\{\tilde{p}, \tilde{q}, \tilde{r}, \tilde{s}\}$ equals

$$\{(0, 0.98224), (-1.26066, 0.741926), (-1.80513, -0.225347), (-0.68335, -0.916091)\},$$

and thus, due to reflection, we can obtain all the points in α , and the corresponding distortion error is given by

$$\int \min_{a \in \alpha} \|x - a\|^2 dP = \text{distortion error due to } \tilde{p} + 2(\text{distortion error due to } \tilde{q}, \tilde{r}, \tilde{s}) = 0.152179.$$

Comparing the distortion errors in Case 1 and Case 2, we see that the set α in Case 2 forms the optimal set of seven-means with quantization error $V_7 = 0.152179$ (see Figure 4). Instead of choosing the point \tilde{p} from the positive direction of the minor axis, we can choose it from the negative direction of the minor axis. This will give another optimal set of seven-means. Thus, the proof of the proposition is complete. \square

Remark 5.7. Following the technique given in Proposition 5.5, we can obtain the optimal set of n -means for any even positive integer $n \geq 4$; on the other hand, if n is odd, following the technique given in Proposition 5.6, we can obtain the optimal sets of n -means for any odd positive integer $n \geq 3$. Notice that if n is even, there are two points in the optimal set which lie on the major axis and so, the optimal set is unique; if n is odd, there is one point in the optimal set which lies on the minor axis and so, there are two different optimal sets of n -means. Thus, we see that

an optimal set of three-means is $\{(0, 0.943319), (-1.3166, -0.29241), (1.3166, -0.29241)\}$ with quantization error $V_3 = 0.661148$; the optimal set of four-means is

$$\{(1.60986, 0), (0, 0.953301), (-1.60986, 0), (0, -0.953301)\}$$

with quantization error $V_4 = 0.393732$; and an optimal set of five-means is

$$\{(0, 0.960639), (-1.669, 0.253562), (-0.832677, -0.869617), (0.832677, -0.869617), (1.669, 0.253562)\}$$

with quantization error $V_5 = 0.28329$ (see Figure 4).

We now end the paper with the following conjecture.

Conjecture 6. By Proposition 4.1, we see that the quantization coefficient for the uniform distribution on a unit circle is $\frac{\pi^2}{3}$; on the other hand, by Proposition 3.11, the quantization coefficient for the uniform distribution on the boundary of a regular hexagon inscribed in a unit circle is 3. Notice that a regular m -sided polygon inscribed in a circle tends to the circle as m tends to infinity. We conjecture that the quantization coefficient for the uniform distribution on the boundary of a regular m -sided polygon inscribed in a circle is an increasing function of m , and approaches to the quantization coefficient for the uniform distribution on the circle as m tends to infinity.

REFERENCES

- [AW] E.F. Abaya and G.L. Wise, *Some remarks on the existence of optimal quantizers*, Statistics & Probability Letters, Volume 2, Issue 6, December 1984, Pages 349-351.
- [DR] C.P. Dettmann and M.K. Roychowdhury, *Quantization for uniform distributions on equilateral triangles*, Real Analysis Exchange, Vol. 42(1), 2017, pp. 149-166.
- [GG] A. Gersho and R.M. Gray, *Vector quantization and signal compression*, Kluwer Academy publishers: Boston, 1992.
- [GKL] R.M. Gray, J.C. Kieffer and Y. Linde, *Locally optimal block quantizer design*, Information and Control, 45 (1980), pp. 178-198.
- [GL1] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics 1730, Springer, Berlin, 2000.
- [GL2] A. György and T. Linder, *On the structure of optimal entropy-constrained scalar quantizers*, IEEE transactions on information theory, vol. 48, no. 2, February 2002.
- [GN] R. Gray and D. Neuhoff, *Quantization*, IEEE Trans. Inform. Theory, 44 (1998), pp. 2325-2383.
- [I] M. Iacobelli, *Asymptotic quantization for probability measures on Riemannian manifolds*, ESAIM: COCV 22 (2016) 770-785.
- [P] K. Pötzlberger, *The quantization dimension of distributions*, Math. Proc. Camb. Phil. Soc., 131, 507-519 (2001).
- [R] M.K. Roychowdhury, *Optimal quantizers for some absolutely continuous probability measures*, Real Analysis Exchange, Vol. 43(1), 2017, pp. 105-136.
- [RR1] J. Rosenblatt and M.K. Roychowdhury, *Optimal quantization for piecewise uniform distributions*, Uniform Distribution Theory 13 (2018), no. 2, 23-55.
- [RR2] J. Rosenblatt and M.K. Roychowdhury, *Uniform distributions on curves and quantization*, arXiv:1809.08364 [math.PR].
- [Z] R. Zam, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*, Cambridge University Press, 2014.

SCHOOL OF MATHEMATICAL AND STATISTICAL SCIENCES, UNIVERSITY OF TEXAS RIO GRANDE VALLEY, 1201 WEST UNIVERSITY DRIVE, EDINBURG, TX 78539-2999, USA.

Email address: ¹gabriela.penamtz@gmail.com

Email address: {²hansapani.rodrido,³mrinal.roychowdhury, ⁴josef.sifuentes, ⁵erwin.suazo}@utrgv.edu