NONDEGENERACY AND WEAK GLOBAL CONVERGENCE OF THE LLOYD ALGORITHM IN $\mathbb{R}^D$

MARIA EMELIANENKO*, LILI JU†, AND ALEXANDER RAND‡

Abstract. Lloyd algorithm originated in the context of optimal quantization and represents a fixed point iteration for computing an optimal quantizer. Reducing average distortion at every step, it constructs a Voronoi partition of the domain and replaces each generator with the centroid of the corresponding Voronoi cell. Optimal quantization is obtained in the case of a centroidal Voronoi tessellation (CVT), which is a special Voronoi tessellations of a domain $\Omega \in \mathbb{R}^d$ having the property that the generators of the Voronoi diagram are also the centers of mass, with respect to a given density function $\rho \geq 0$, of the corresponding Voronoi cells. Lloyd iteration is currently the most popular and elegant algorithm for computing CVTs and optimal quantizers, but many questions remain about its convergence, especially in d-dimensional spaces ($d > 1$). In this paper, we prove that any limit point of the Lloyd iteration in any dimensional spaces is non-degenerate provided that Ω is a convex and bounded set, and ρ belongs to $L^1(\Omega)$ and is positive almost everywhere. This assures that the fixed point map remains closed and hence the standard theory of descent methods guarantees weak global convergence of the Lloyd iteration to the set of non-degenerate fixed-point quantizers. While previously only conjectured, the convergence properties of the Lloyd iteration are rigorously justified under such minimal regularity assumptions on the density functional. The results presented in this paper go beyond existing convergence theories for CVT and optimal quantization related algorithms and should be of interest to both mathematical and engineering communities.

Key words. Centroidal Voronoi tessellations, quantization, Lloyd algorithm, global convergence

AMS subject classifications. 65D99,94A15,68U05

1. Introduction.

Given an bounded domain $\Omega \in \mathbb{R}^d$ and a set of distinct points $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^n \subset \Omega$. For each point $\mathbf{z}_i$, $i = 1, \ldots, n$, define the corresponding Voronoi region $V_i(\mathbf{Z})$, $i = 1, \ldots, n$, by

$$V_i(\mathbf{Z}) = \big\{\mathbf{z} \in \Omega \;\mid\; |\mathbf{z} - \mathbf{z}_i| < |\mathbf{z} - \mathbf{z}_j| \quad \text{for } j = 1, \cdots, n \quad \text{and} \quad j \neq i\big\}$$

where $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$. Clearly $V_i(\mathbf{Z}) \cap V_j(\mathbf{Z}) = \emptyset$ for $i \neq j$, and $\cup_{i=1}^n \overline{V}_i(\mathbf{Z}) = \overline{\Omega}$ so that $\{V_i(\mathbf{Z})\}_{i=1}^n$ is a tessellation of Ω. We refer to $\mathbf{V}(\mathbf{Z}) = \{V_i(\mathbf{Z})\}_{i=1}^n$ as the *Voronoi tessellation* (VT) of Ω associated with the point set $\mathbf{Z}$. A point $\mathbf{z}_i$ is called a *generator*; a subdomain $V_i(\mathbf{Z}) \subset \Omega$ is referred to as the *Voronoi region* corresponding to the generator $\mathbf{z}_i$. It is well-known that the dual tessellation (in a graph-theoretical sense) to a Voronoi tessellation of Ω is the so-called *Delaunay triangulation* (DT). All Voronoi regions $V_i(\mathbf{Z})$'s are convex (especially convex polygons/polyhedra except their part on the boundary) if Ω is convex.

Given a density function $\rho(\mathbf{z}) \geq 0$ defined on Ω, for any region $V \subset \Omega$, define $\mathbf{c}_V$, the *mass center* or *centroid* of V by

$$\mathbf{c}_V = \frac{\displaystyle\int_V \mathbf{y}\rho(\mathbf{y})\,\mathrm{d}\mathbf{y}}{\displaystyle\int_V \rho(\mathbf{y})\,\mathrm{d}\mathbf{y}}.$$

*Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA. (memelian@gmu.edu).

†Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA. (ju@math.sc.edu). This author's research is supported in part by the National Science Foundation under grant number DMS-0609575 and the Department of Energy under grant number DE-FG02-07ER64431.

‡Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. (arand@andrew.cmu.edu).

Then a special family of Voronoi tessellations can be defined in the following:

DEFINITION 1.1. [4] *We refer to a Voronoi tessellation* $(\mathbf{Z}, \mathbf{V}(\mathbf{Z}))$ *of* Ω *as a* centroidal Voronoi tessellation *(CVT) if and only if the points* $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^n$ *which serve as the generators of the associated Voronoi regions* $\mathbf{V}(\mathbf{Z}) = \{V_i(\mathbf{Z})\}_{i=1}^n$ *are also the centroids of those regions, i.e., if and only if we have that*

$$\mathbf{z}_i = \mathbf{c}_{V_i(\mathbf{Z})}, \quad \textit{for } \ i = 1, \ldots, n\,.$$

General Voronoi tessellations do not satisfy the CVT property. It is worth noting that for a given domain and density, at least one CVT exists [1] but may not be unique [4]. The CVT concept can be generalized to very broad settings that range from abstract spaces and distance metrics to discrete point sets [4]. In the latter setting, it can be recognized as being closely related to the k-means and h-means algorithms in clustering and quantization applications [10]. Its extension to general surfaces and manifolds has also been studied in [5, 8]. CVTs are very useful in many applications, including but not limited to image and data analysis, vector quantization, resource optimization, design of experiments, optimal placement of sensors and actuators, cell biology, territorial behavior of animals, numerical partial differential equations, point sampling, meshless computing, mesh generation and optimization, reduced-order modeling, computer graphics, and mobile sensing networks.

CVTs possess an optimization property that can be used as a basis for an alternate definition. Given any set of points $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^n$ in Ω and any tessellation $\mathbf{V} = \{V_i\}_{i=1}^n$ of Ω, define the *energy* by

$$\mathcal{E}(\mathbf{Z}, \mathbf{V}) = \sum_{i=1}^{n} \int_{V_i} \rho(\mathbf{y}) |\mathbf{y} - \mathbf{z}_i|^2 \, \mathrm{d}\mathbf{y}\,. \tag{1.1}$$

Then it can be shown that $\mathcal{E}$ is minimized only if $(\mathbf{Z}, \mathbf{V})$ forms a CVT of Ω [4] . Notice that $(\mathbf{Z}, \mathbf{V})$ still may not be a minimizer of $\mathcal{E}$ although $(\mathbf{Z}, \mathbf{V})$ is a CVT. For some properties of the minimizers of the energy functional $\mathcal{E}$ (i.e., CVTs or optimal quantizations), see discussions in [4, 2, 22]. Although the energy $\mathcal{E}$ may not be directly identified with the energy of some physical system, it is often naturally associated with quantities such as *error distortion*, *variance* and *cost* in many applications.

Currently the most popular and effective algorithm for construction of CVTs is the so-called Lloyd algorithm [18] that is a simple iteration between constructing Voronoi diagrams and mass centroids.

ALGORITHM 1.1. (**Lloyd Algorithm**) *Given a domain* Ω*, a density function* $\rho(\mathbf{x})$ *defined on* Ω*, and a positive integer* n*.*

0. *Select an initial set of* n *points* $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^n$ *on* Ω*;*
1. *Construct the Voronoi regions* $\mathbf{V}(\mathbf{Z}) = \{V_i(\mathbf{Z})\}_{i=1}^n$ *of* Ω *associated with* $\mathbf{Z}$*;*
2. *Determine the centroids of the Voronoi regions* $\mathbf{V}(\mathbf{Z})$*; these centroids form the new set of points* $\mathbf{Z}$*;*
3. *If the new points meet some convergence criterion, return* $(\mathbf{Z}, \mathbf{V}(\mathbf{Z}))$ *and terminate; otherwise, goto step* 1.

The Lloyd algorithm also coincides with the simplest version of the k-means algorithm in clustering and quantization fields [10, 12, 18, 16]. A commonly used stopping criterion is to test whether $\|\mathbf{Z}^k - \mathbf{Z}^{k-1}\| \leq \epsilon$ holds for some predefined tolerance ϵ at some iteration k. It is worth noting that the energy $\mathcal{E}(\mathbf{Z}, \mathbf{V}(\mathbf{Z}))$ decreases monotonically along Lloyd iteration. Some acceleration scheme based on multigrid techniques are proposed in [6]. A probabilistic version of the Lloyd method for computing CVTs and its parallel implementation have also been studied in [14]. One famous generalization in the vector quantization setting is the Linde-Buzo-Gray (LBG) algorithm [17], widely used in statistics and data mining applications.

Since Lloyd's pioneering work, many studies have been made on convergence properties of the Lloyd iteration [4, 7, 15, 19], and despite its long history and wide popularity in many applications, it is far from being complete. Many researchers addressed the question of existence and uniqueness of an optimal quantizer. Most notably, in [9] and later in [25] and [15], it was shown that Lloyd method is a local contraction (thus *locally* convergent) in the one dimensional space when the density function is logarithmically concave. This result was reiterated in [4] with a much simpler proof. In [26], this result was extended to all differential and positive densities in the one-dimensional setting of scalar quantization. A different approach was used in [7], where the fixed point map was shown to be closed and a weak global convergence result was presented for all continuous and positive density functional along with some numerical studies and precise theoretic estimates of the local convergence rate.

In higher dimensions, convergence results are much more limited. For discrete distributions, the algorithm is shown to converge and, moreover, reach a CVT in a finite number of steps [21, 11]. In the case of continuous distributions, [11, 24] show convergence of the energy functional by defining the value of the Lloyd map on degenerate points. In order to ensure that the algorithm is closed, this requires that degenerate points be **defined** as fixed points in the definition of the algorithm. So, while standard theory of descent methods ensure that this extended Lloyd algorithm will converge to a fixed point, these limit points could be degenerate. We seek to apply the theory of descent methods to the Lloyd algorithm without any extension to degenerate points. This makes sense as it will be shown that points of degeneracy maximize the energy in some sense. Moreover, this is necessary in order to ensure that limit points are in fact CVTs.

In this paper we prove that for an arbitrary, positive density function under a mild L^1 regularity assumption, the fixed point mapping remains closed, which guarantees weak global convergence of the Lloyd iteration. Although we rely on the standard global convergence theory for descent algorithms the way discussed in previous investigations ([24],[25] etc), we take a more direct approach of rigorously justifying the non-degeneracy of the limiting and intermediate iteration points, the fact that much simplifies the analysis of the method and that escaped rigorous treatment until now. These results ensure that Lloyd iterates do not approach degenerate points and that any limit point of the algorithm is a CVT with n distinct generators.

The rest of the paper is organized as follows. In section 2, we introduce some necessary notations and give an overview of some relevant results and properties. In subsection 3.1, we derive precise estimates on the distance from a centroid to the boundary of the domain and obtain an upper bound on the difference of two Voronoi regions. In subsection 3.2 we present the proof of non-degeneracy of the limit points of the Lloyd iteration, followed in Section 3.3 by the main result of the weak global convergence of Lloyd algorithm. Conclusions and some open questions are given in section 4.

2. Preliminaries. For any set $V \subset \Omega$, let us denote by m(V) the measure of V i.e.,

$$m(V) := \int_V \, \mathrm{d}\mathbf{y},$$

and by $M(V)$ the mass of V, i.e.,

$$M(V) := \int_V \rho(\mathbf{y})\, \mathrm{d}\mathbf{y}.$$

The fact that ρ is positive almost everywhere means $M(V) > 0$ if $m(V) > 0$. We also define $\hat{\Omega}^n$ as follows:

$$\hat{\Omega}^n := \{\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n) \in \bar{\Omega}^n \mid \mathbf{z}_i \neq \mathbf{z}_j \ \ \forall\, i \neq j\}$$

The so-called Lloyd map is the function $\mathbf{T} : \hat{\Omega}^n \to \hat{\Omega}^n$ that takes a tuple of generators into the tuple of the centroids of its Voronoi regions (exactly the iteration used in the Lloyd algorithm). More precisely, let $\mathbf{T} = (\mathbf{T}_1, \cdots, \mathbf{T}_n)$, then for any $\mathbf{Z} \in \hat{\Omega}^n$,

$$\mathbf{T}_i(\mathbf{Z}) = \frac{\displaystyle\int_{V_i(\mathbf{Z})} \mathbf{y}\rho(\mathbf{y})\, \mathrm{d}\mathbf{y}}{\displaystyle\int_{V_i(\mathbf{Z})} \rho(\mathbf{y})\, \mathrm{d}\mathbf{y}}$$

for $i = 1, \cdots, n$. The Lloyd map is continuous from $\hat{\Omega}^n$ to $\hat{\Omega}^n$ [4, 7], but not defined on $\bar{\Omega}^n \setminus \hat{\Omega}^n$. The Lloyd algorithm is then the process generating a sequence of point sets:

$$\mathbf{Z}, \mathbf{T}(\mathbf{Z}), \mathbf{T}^2(\mathbf{Z}), \cdots, \mathbf{T}^k(\mathbf{Z}), \cdots$$

Clearly, the resulting Lloyd iterates are bounded.

For any $\mathbf{Z} \in \hat{\Omega}^n$, the *quantization energy* is defined as

$$\mathcal{G}(\mathbf{Z}) = \mathcal{E}(\mathbf{Z}, \mathbf{V}(\mathbf{Z})) = \sum_{i=1}^{n} \int_{V_i(\mathbf{Z})} \rho(\mathbf{y})|\mathbf{y} - \mathbf{z}_i|^2\, \mathrm{d}\mathbf{y}.$$

The quantization energy function is also continuous on $\hat{\Omega}^n$. Furthermore, this energy can be extended continuously to $\overline{\Omega}^n$. When $\mathbf{Z} \notin \hat{\Omega}^n$, the value of $\mathcal{G}$ is given by the same formula except we consider $\mathbf{Z}$ without duplicate components. For example, in the case of $n = 2$, if $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2)$ and $\mathbf{z}_1 \neq \mathbf{z}_2$, then

$$\mathcal{G}(\mathbf{Z}) = \int_{V_1(\mathbf{Z})} \rho(\mathbf{y})|\mathbf{y} - \mathbf{z}_1|^2\, \mathrm{d}\mathbf{y} + \int_{V_2(\mathbf{Z})} \rho(\mathbf{y})|\mathbf{y} - \mathbf{z}_2|^2\, \mathrm{d}\mathbf{y}.$$

In the degenerate case, $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_1)$, then

$$\mathcal{G}(\mathbf{Z}) = \mathcal{G}((\mathbf{z}_1)) = \int_{\Omega} \rho(\mathbf{y})|\mathbf{y} - \mathbf{z}_1|^2\, \mathrm{d}\mathbf{y}.$$

The derivative of $\mathcal{G}$ in $\hat{\Omega}^n$ is given [4] by

$$\frac{\partial \mathcal{G}}{\partial \mathbf{z}_i}(\mathbf{Z}) = 2M(V_i(\mathbf{Z}))(\mathbf{Z}_i - \mathbf{T}_i(\mathbf{Z})), \tag{2.1}$$

and thus the set of critical points of $\mathcal{G}$ in $\hat{\Omega}^n$ is exactly the set of fixed points of the Lloyd map and thus is the set of CVTs.

For convenience of discussion, a more general energy $\mathcal{H}$ is defined in [7] as: for any $(\mathbf{Y}, \mathbf{Z}) \in (\hat{\Omega}^n, \hat{\Omega}^n)$,

$$\mathcal{H}(\mathbf{Y}, \mathbf{Z}) = \mathcal{E}(\mathbf{Z}, \mathbf{V}(\mathbf{Y})) = \sum_{i=1}^{n} \int_{V_i(\mathbf{Y})} \rho(\mathbf{y}) |\mathbf{y} - \mathbf{z}_i|^2 \, \mathrm{d}\mathbf{y} \, .$$

Notice that $\mathcal{H}$ can not be continuously extended to $(\overline{\Omega}^n, \overline{\Omega}^n)$ like the function $\mathcal{G}$. The Lloyd algorithm is in fact a dual minimization process [7]: for any $\mathbf{Z} \in \hat{\Omega}^n$,

$$\mathcal{H}(\mathbf{Z}, \mathbf{T}(\mathbf{Z})) = \min_{\mathbf{Y} \in \hat{\Omega}^n} \mathcal{H}(\mathbf{Z}, \mathbf{Y}),$$
$$\mathcal{H}(\mathbf{Z}, \mathbf{Z}) = \min_{\mathbf{Y} \in \hat{\Omega}^n} \mathcal{H}(\mathbf{Y}, \mathbf{Z}).$$

As long as ρ is positive almost everywhere, the above minimization problems in the right hand side both have a unique solution. Thus we have that

$$\mathcal{G}(\mathbf{T}(\mathbf{Z})) = \mathcal{H}(\mathbf{T}(\mathbf{Z}), \mathbf{T}(\mathbf{Z})) \leq \mathcal{H}(\mathbf{Z}, \mathbf{T}(\mathbf{Z})) \leq \mathcal{H}(\mathbf{Z}, \mathbf{Z}) = \mathcal{G}(\mathbf{Z}). \qquad (2.2)$$

Uniqueness of the centroid implied by the positive density in a convex set ensures that if $(\mathbf{Z}, \mathbf{V}(\mathbf{Z}))$ is not CVT, then $\mathcal{G}(\mathbf{T}(\mathbf{Z})) < \mathcal{G}(\mathbf{Z})$.

The following definition of weak convergence will be used throughout the paper.

Definition 2.1. *Denote by $\sigma(x, S)$ a distance between a point $x \in \Omega$ and a set S and by Γ the set of fixed points of $\mathbf{T}$. We will say that the sequence of Lloyd iterates $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ is* weakly convergent *if $\lim_{i \to +\infty} \nabla \mathcal{G}(\mathbf{Z}^i) = \mathbf{0}$ and $\lim_{i \to +\infty} \sigma(\mathbf{Z}^i, \Gamma) = 0$.*

In other words, the sequence of Lloyd iterates converges in a weak sense, if the energy values asymptotically approach some limit and the sequence of Lloyd iteration points approaches the set of fixed points of $\mathbf{T}$.

In order to analyze *global* convergence of the Lloyd algorithm in general dimensional spaces, we recall the following result.

Theorem 2.1. [7, 20] *If the iterates in the Lloyd algorithm stay in a compact set where the Lloyd map $\mathbf{T}$ is continuous, then the algorithm is weakly globally convergent, i.e., $\lim_{i \to +\infty} \nabla \mathcal{G}(\mathbf{Z}^i) = \mathbf{0}$ for the Lloyd iterates $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ starting from any initial guess (in other words any limit point of $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ is a critical point of $\mathcal{G}$).*

Theorem 2.1 was proved in [7] using the properties of $\mathcal{H}$ and the (Weak) Global Convergence Theorem stated in [20]. It was also shown in [7] and [26] that the Lloyd algorithm has to converge to the set of critical points of the energy in one dimensional space provided that ρ is strictly positive and smooth, but for higher dimensional cases, this question is still open, as discussed earlier in the introduction.

It has also never been rigorously justified that the set of limit points cannot contain more than one point, except for the case of log-concave density in one-dimensional case, restricting the convergence in the aforementioned result to the "weak" sense as given in Definition 2.1. Although we conjecture that a stronger convergence result can be obtained in most practical cases, this question requires a separate rigorous treatment, while at present we focus our attention on extending Theorem 2.1 to higher dimensions.

More precisely, the goal of this paper is to present a proof for the weak global convergence of the Lloyd algorithm in any dimensional space. We do this by showing that Lloyd method in $\mathbb{R}^d$ is non-degenerate and hence it converges in the sense of Definition 2.1 for any L^1 density functions in any dimension.

3. Weak Global Convergence of Lloyd Iteration. In the following part of this paper, we assume the following conditions hold.

ASSUMPTION 3.1.
1) The domain $\Omega \in \mathbb{R}^d$ is a convex and bounded set with the diameter

$$\mathrm{diam}(\Omega) := \sup_{\mathbf{z},\mathbf{y}\in\Omega} |\mathbf{z}-\mathbf{y}| = R_\Omega < +\infty.$$

2) The density function ρ belongs to $L^1(\Omega)$ and is positive almost everywhere. Consequently, we have that

$$0 < M(\Omega) = \|\rho\|_{L^1(\Omega)} = \int_\Omega \rho(\mathbf{y})\mathrm{d}\mathbf{y} < +\infty.$$

For the proof of weak global convergence of the Lloyd algorithm, we will show that any limit point of the algorithm is non-degenerate (in $\hat{\Omega}^n$) and thus the iterates stay inside a compact set in $\hat{\Omega}^n$. The weak global convergence of the algorithm will then follow from Theorem 2.1. This ensures that any limit point of the algorithm does in fact correspond to a CVT. Before these results can be shown, a number of technical intermediate lemmas must be shown. The goal of these lemmas is to estimate how the quantization energy changes between iterations of the Lloyd map when iterates approach the boundary of $\hat{\Omega}^n$. As iterates approach the boundary of $\hat{\Omega}^n$, some generator approaches the boundary of its Voronoi region. As all Voronoi regions are convex sets, it becomes important to bound the distance between the boundary and the centroid of a generic convex set in Ω with a prescribed mass.

3.1. Some lemmas. First, let us present a classic result from measure theory.

LEMMA 3.1. *Let Assumption 3.1 be satisfied. Then for any $\epsilon > 0$, there exists a constant $\eta > 0$ such that*

$$M(V) = \int_V \rho(\mathbf{y})\,\mathrm{d}\mathbf{y} \le \epsilon. \tag{3.1}$$

for any set $V \subset \Omega$ with $m(V) \le \eta$.

The proof of this Lemma is given in the Appendix. For any $\epsilon \ge 0$, define

$$E_\epsilon = \{\eta \in [0, m(\Omega)] \mid M(V) \le \epsilon \ \text{ if } \ V \subset \Omega \text{ with } m(V) \le \eta\},$$

and Lemma 3.1 tells us that E_ϵ is a nonempty set, so that

$$\eta_\epsilon := \sup_{\eta\in E_\epsilon} \eta \tag{3.2}$$

is well-defined and satisfies that $\eta_{\epsilon_1} \le \eta_{\epsilon_2}$ if $\epsilon_1 \le \epsilon_2$. It is also obvious that $\eta_\epsilon > 0$ if $\epsilon > 0$. For a given $\epsilon \ge 0$, η_ϵ is often difficult to compute, but here we only assert its existence.

Now, we estimate the distance from the boundary to the centroid of a general convex set. The important feature of the estimate below is that lower bound on the distance between the centroid and the boundary of a general convex set only depends on the set through the mass of the set.

For any $V \in \Omega$, let $R_V := \mathrm{diam}(V)$ denote the diameter of V (Clearly $R_V \le R_\Omega$) and define the V-dependent constant δ_V by

$$\delta_V := \frac{\eta^2_{M(V)/4}}{64 R_V^{2d-1}}$$

where $\eta_{M(V)/4}$ is the constant defined in (3.2) with $\epsilon = \frac{M(V)}{4}$. Then the following result holds:

LEMMA 3.2. *Let Assumption 3.1 be satisfied. Then for any convex set $V \subset \Omega$, it holds that*

$$\mathrm{dist}(\mathbf{c}_V, \partial V) \geq \delta_V \tag{3.3}$$

where $\mathrm{dist}(\mathbf{c}_V, \partial V)$ *denotes the shortest distance from the centroid of V denoted by $\mathbf{c}_V$ to the boundary of V.*

Proof. Without loss of generality, we assume that the origin is the nearest boundary point to the centroid $\mathbf{c}_V$ and let $\mathbf{c}_V$ have the coordinate $(\mathbf{0}, x)$ for some $x > 0$ where $\mathbf{0}$ denotes the zero vector in $\mathbb{R}^{d-1}$. Since V is convex and $\sup_{\mathbf{z},\mathbf{y}\in\Omega} |\mathbf{z}-\mathbf{y}| = R_V$, we must have $R_V \geq y_d \geq 0$ for any $\mathbf{y} = (y_1, \cdots, y_d) \in V$. Further, the d-th coordinate of $\mathbf{c}_V$ exactly equals $\mathrm{dist}(\mathbf{c}_V, \partial V)$. Then, from the definition of the centroid, we have that

$$\begin{aligned} \mathrm{dist}(\mathbf{c}_V, \partial V) &= \frac{\int_V \rho(\mathbf{y}) y_d \, \mathrm{d}\mathbf{y}}{\int_V \rho(\mathbf{y}) \, \mathrm{d}\mathbf{y}} \\ &= \frac{\int_0^{R_V} \int_{A(y_d)} \rho(\tilde{\mathbf{y}}, y_d) y_d \, \mathrm{d}\tilde{\mathbf{y}} \mathrm{d}y_d}{M(V)} \\ &= \frac{\int_0^{R_V} y_d [\int_{A(y_d)} \rho(\tilde{\mathbf{y}}, y_d) \, \mathrm{d}\tilde{\mathbf{y}}] \, \mathrm{d}y_d}{M(V)} \end{aligned} \tag{3.4}$$

where $\tilde{\mathbf{y}} = (y_1, \cdots, y_{d-1})$ and $A(y_d) = \{\tilde{\mathbf{y}} \in \mathbb{R}^{d-1} \mid (\tilde{\mathbf{y}}, y_d) \in V\}$. Let us define

$$f_\rho(y_d) := \int_{A(y_d)} \rho(\tilde{\mathbf{y}}, y_d) \, \mathrm{d}\tilde{\mathbf{y}}.$$

The representation above has essentially reduced the problem to a similar one dimensional problem. To bound (3.4) from below, follow the following steps. In the first step, we find a set such that $\int_S f_\rho(y_d) \, \mathrm{d}y_d$ is sufficiently large and f_ρ is bounded below away from 0. Second, a lower bound on this Lebesgue measure of this set is found. Finally, the third step is to combine these estimate to get a lower bound on (3.4) in terms of the mass of V.

STEP 1: For any number $\mu > 0$, define $S_\mu = \{y_d \mid f_\rho(y_d) > \mu R_V^{d-1}\}$ and let $L(S_\mu)$ denotes the measure of S_μ. First we have

$$\begin{aligned} M(V) &= \int_V \rho(\mathbf{y}) \, \mathrm{d}\mathbf{y} \\ &= \int_0^{R_V} f_\rho(y_d) \, \mathrm{d}y_d \\ &= \int_{S_\mu} f_\rho(y_d) \, \mathrm{d}y_d + \int_{[0,R_V]-S_\mu} f_\rho(y_d) \, \mathrm{d}y_d \\ &\leq \int_{S_\mu} f_\rho(y_d) \, \mathrm{d}y_d + \mu R_V^d. \end{aligned}$$

Setting $\mu_1 = \dfrac{M(V)}{2R_V^d}$, then the above inequality gives us

$$\int_{S_{\mu_1}} f_\rho(y_d) \, \mathrm{d}y_d \geq M(V) - \mu_1 R_V^d = M(V) - \frac{M(V)}{2} = \frac{M(V)}{2}. \tag{3.5}$$

STEP 2: From the fact

$$M(V) \geq \int_{S_\mu} f_\rho(y_d)\ \mathrm{d}y_d \geq L(S_\mu)\mu R_V^{d-1},$$

it is easy to deduce that

$$L(S_\mu) \leq \frac{M(V)}{\mu R_V^{d-1}},$$

and so that

$$m(V_\mu) \leq L(S_\mu) R_V^{d-1} \leq \frac{M(V)}{\mu}$$

where $V_\mu = \{(\tilde{\mathbf{y}}, y_d) \in V \mid y_d \in S_\mu,\ \tilde{\mathbf{y}} \in A(y_d)\}$. Setting $\mu_2 = \dfrac{M(V)}{\eta_{M(V)/4}}$, we have

$$m(V_{\mu_2}) \leq \frac{M(V)}{\mu_2} = \eta_{M(V)/4}.$$

Then, by Lemma 3.1, we obtain that

$$\int_{S_{\mu_2}} f_\rho(y_d)\ \mathrm{d}y_d = \int_{S_{\mu_2}} \int_{A(y_d)} \rho(\tilde{\mathbf{y}}, y_d)\ \mathrm{d}\tilde{\mathbf{y}}\mathrm{d}y_d = \int_{V_{\mu_2}} \rho(\mathbf{y})\ \mathrm{d}\mathbf{y} \leq \frac{M(V)}{4}. \tag{3.6}$$

According to (3.5) and (3.6), we get

$$\int_{S_{\mu_1} - S_{\mu_2}} f_\rho(y_d)\ \mathrm{d}y_d \geq \frac{M(V)}{2} - \frac{M(V)}{4} \geq \frac{M(V)}{4}. \tag{3.7}$$

Notice that we also have

$$\int_{S_{\mu_1} - S_{\mu_2}} f_\rho(y_d)\ \mathrm{d}y_d \leq \mu_2 R_V^{d-1} L(S_{\mu_1}). \tag{3.8}$$

Combination of (3.7) and (3.8) gives us

$$L(S_{\mu_1}) \geq \frac{M(V)}{4\mu_2 R_V^{d-1}}. \tag{3.9}$$

STEP 3: Using (3.4) and (3.9), we get

$$\begin{aligned}
\mathrm{dist}(\mathbf{c}_V, \partial V) &= \frac{\int_0^{R_V} y_d f_\rho(y_d)\ \mathrm{d}y_d}{M(V)} \\
&\geq \frac{\int_{S_{\mu_1}} y_d \mu_1 R_V^{d-1}\ \mathrm{d}y_d}{M(V)} \\
&\geq \frac{\mu_1 R_V^{d-1} \int_0^{L(S_{\mu_1})} y_d\ \mathrm{d}y_d}{M(V)} \\
&\geq \frac{\mu_1 R_V^{d-1} [L(S_{\mu_1})]^2}{2M(V)} \\
&\geq \frac{\mu_1 M(V)}{32\mu_2^2 R_V^{d-1}} \\
&\geq \frac{\eta_{M(V)/4}^2}{64 R_V^{2d-1}}
\end{aligned}$$

that completes the proof. ◻

REMARK 3.1. *If $\rho \in L^q(\Omega)$ for some $q > 1$, then δ_V can be selected such that it does not depend on the constant $\eta_{M(V)/4}$ such as*

$$\delta_V := \frac{M(V)^{2q/(q-1)}}{4^{(5q-3)/(q-1)}\|\rho\|_{L^q(V)}^{2q/(q-1)} R_V^{2d-1}}.$$

This follows from that fact that applying Hölder inequality to $\int_{S_\mu} f_\rho(y_d)\,\mathrm{d}y_d$ gives us

$$\int_{S_\mu} f_\rho(y_d)\,\mathrm{d}y_d \le \frac{\|\rho\|_{L^q(V)}[M(V)]^{(q-1)/q}}{\mu^{(q-1)/q}}.$$

and so that

$$\int_{S_{\mu_2}} f_\rho(y_d)\,\mathrm{d}y_d \le \frac{M(V)}{4}.$$

where $\mu_2 = \dfrac{(4\|\rho\|_{L^q(V)})^{q/(q-1)}}{[M(V)]^{1/(q-1)}}$.

The next lemma is a local estimate on the difference in the quantization energy over a particular set with respect to different generators. For our purposes, it is important to compare the energy with respect to the centroid to the energy generated by a point near the boundary. To prove this, the previous lemma is essential: it ensures that the centroid is sufficiently far away from the boundary. This again leads to an estimate that depends on the particular convex set V through $M(V)$.

For any convex set $V \subset \Omega$, let us define the V-dependent constant γ_V by

$$\gamma_V := \frac{M(V)\delta_V^2}{4}.$$

LEMMA 3.3. *Let Assumption 3.1 be satisfied and $V \subset \Omega$ be a convex set. Then for any $\mathbf{z} \in V$ with $\mathrm{dist}(\mathbf{z}, \partial V) \le \frac{\delta_V}{2}$, it holds that*

$$\int_V \rho(\mathbf{y})|\mathbf{y} - \mathbf{c}_V|^2\,\mathrm{d}\mathbf{y} \le \int_V \rho(\mathbf{y})|\mathbf{y} - \mathbf{z}|^2\,\mathrm{d}\mathbf{y} - \gamma_V, \tag{3.10}$$

where $\mathbf{c}_V$ is the centroid of V.

Proof. First, from Lemma 3.2, it is obvious that

$$|\mathbf{z} - \mathbf{c}_V| \ge \mathrm{dist}(\mathbf{c}_V, \partial V) - \mathrm{dist}(\mathbf{z}, \partial V) \ge \frac{\delta_V}{2}.$$

We have

$$\begin{aligned}
&\int_V \rho(\mathbf{y})|\mathbf{y} - \mathbf{z}|^2\,\mathrm{d}\mathbf{y} - \int_V \rho(\mathbf{y})|\mathbf{y} - c_V|^2\,\mathrm{d}\mathbf{y}\\
&\quad= \int_V \rho(\mathbf{y}) < \mathbf{y} - \mathbf{z}, \mathbf{y} - \mathbf{z} > \mathrm{d}\mathbf{y} - \int_V \rho(\mathbf{y}) < \mathbf{y} - \mathbf{c}_V, \mathbf{y} - \mathbf{c}_V > \mathrm{d}\mathbf{y}\\
&\quad= \int_V \rho(\mathbf{y})(|\mathbf{y}|^2 + |\mathbf{z}|^2 - 2 < \mathbf{y}, \mathbf{z} >)\mathrm{d}\mathbf{y}\\
&\qquad - \int_V \rho(\mathbf{y})(|\mathbf{y}|^2 + |\mathbf{c}_V|^2 - 2 < \mathbf{y}, \mathbf{c}_V >)\mathrm{d}\mathbf{y}\\
&\quad= \int_V \rho(\mathbf{y})(|\mathbf{z}|^2 - |\mathbf{c}_V|^2)\,\mathrm{d}\mathbf{y} + 2\int_V \rho(\mathbf{y}) < \mathbf{y}, \mathbf{c}_V - \mathbf{z} > \mathrm{d}\mathbf{y}.
\end{aligned}$$

Notice that

$$\int_V \rho(\mathbf{y})(|\mathbf{z}|^2 - |\mathbf{c}_V|^2)\ \mathrm{d}\mathbf{y} = M(V)(|\mathbf{z}|^2 - |\mathbf{c}_V|^2), \tag{3.11}$$

and

$$\begin{aligned}\int_V \rho(\mathbf{y}) < \mathbf{y}, \mathbf{c}_V - \mathbf{z} > \mathrm{d}\mathbf{y} &= < \int_V \mathbf{y}\rho(\mathbf{y})\ \mathrm{d}\mathbf{y}, \mathbf{c}_V - \mathbf{z} > \\ &= < M(V)\mathbf{c}_V, \mathbf{c}_V - \mathbf{z} > \\ &= M(V)(|\mathbf{c}_V|^2 - < \mathbf{c}_V, \mathbf{z} >).\end{aligned} \tag{3.12}$$

Using (3.11) and (3.12), we go further and get

$$\begin{aligned}&\int_V \rho(\mathbf{y})|\mathbf{y}-\mathbf{z}|^2\ \mathrm{d}\mathbf{y} - \int_V \rho(\mathbf{y})|\mathbf{y}-\mathbf{c}_V|^2\ \mathrm{d}\mathbf{y} \\ &= M(V)(|\mathbf{z}|^2 + |\mathbf{c}_V|^2 - 2 < \mathbf{c}_V, \mathbf{z} >)) \\ &= M(V)|\mathbf{z}-\mathbf{c}_V|^2 \\ &\geq \frac{M(V)\delta_V^2}{4}\end{aligned}$$

that deduces (3.10) and completes the proof. ☐

Before presenting the next lemma, we need to define some notation. For any point set $\mathbf{Z} \in \Omega^n$, denote by $\mathbf{Z}_i$ the i-th point of $\mathbf{Z}$.

Definition 3.1. *Let $k, n \in \mathbb{N}$. For any points $\mathbf{Z} \in \hat{\Omega}^k$ and $\mathbf{Y} \in \hat{\Omega}^n$, define the d_*-distance from $\mathbf{Z}$ to $\mathbf{Y}$ by*

$$d_*(\mathbf{Y}, \mathbf{Z}) = \max_{1\leq j\leq k} \min_{1\leq i\leq n} |\mathbf{Z}_j - \mathbf{Y}_i|.$$

Then the natural definition of the distance between $\mathbf{Z}$ and $\mathbf{Y}$ is given by

$$d(\mathbf{Y}, \mathbf{Z}) = \max(d_*(\mathbf{Y}, \mathbf{Z}), d_*(\mathbf{Z}, \mathbf{Y})).$$

This is a generalization of the regular Euclidean difference between to elements of Ω^n to the case of vectors of a different length. This definition will come useful for us when we consider a degenerate set of generators in the proof of our main theorem. Provided this definition, let us consider the set theoretic difference of the two Voronoi tessellations generated by the vectors $\mathbf{Z}$ and $\mathbf{Y}$ as

$$\sum_{i=1}^{k} M\left(V_i(\mathbf{Z}) - \cup_{j\in S_i} V_j(\mathbf{Y})\right).$$

This definition is illustrated in Figure 3.1 below.

The next lemma for quantifies the fact that if two possibly degenerate points in Ω^n are close together in the sense of Definition 3.1, then the corresponding Voronoi regions of these points are also close together in some sense. The appropriate measurement of how far apart these Voronoi regions are is the total mass of the difference of corresponding Voronoi regions.

Lemma 3.4. *Let Assumption 3.1 be satisfied. Fix $\mathbf{Z} \in \hat{\Omega}^k$ for $k \leq n$. For any $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ such that if $\mathbf{Y} \in \hat{\Omega}^n$ and $d(\mathbf{Z}, \mathbf{Y}) \leq \delta_\epsilon$, then*

$$\sum_{i=1}^{k} M\left(V_i(\mathbf{Z}) - \cup_{j\in S_i} V_j(\mathbf{Y})\right) \leq \epsilon$$

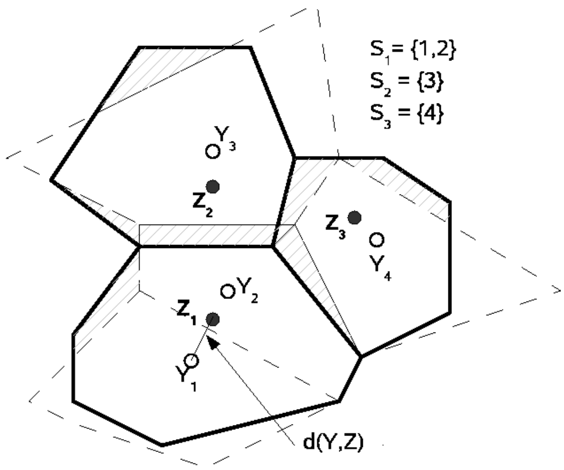

FIG. 3.1. *An illustration of the definition of* $d(\mathbf{Y},\mathbf{Z})$ *and the difference of Voronoi tessellations generated by* $\mathbf{Z}$ *and* $\mathbf{Y}$*, where* $\mathbf{Z} \in \hat{\Omega}_3$ *and* $\mathbf{Y} \in \hat{\Omega}_4$ *(shaded).*

where $\{S_i\}_{i=1}^k$ *is any partition of* $\{1,\cdots,n\}$ *satisfying* $S_i \in \{j \mid |\mathbf{Z}_i - \mathbf{Y}_j| \le |\mathbf{Z}_m - \mathbf{Y}_j|,\ \forall\, m \in \{1,\cdots,k\}\}$.

Proof. Let us suppose that $\mathbf{Y} \in \hat{\Omega}^n$ and $d(\mathbf{Z},\mathbf{Y}) < \delta$ for some $\delta > 0$. Since $\mathbf{Z} \in \hat{\Omega}^k$ is fixed, let us define the constant $d_{\mathbf{Z}}$ by

$$d_{\mathbf{Z}} = \min_{i\neq j} \frac{|\mathbf{Z}_i - \mathbf{Z}_j|}{2}.$$

We let δ be small enough such that $\delta \le \dfrac{d_{\mathbf{Z}}}{2}$, then it is not difficult to verify that S_i is a nonempty set for each $1 \le i \le k$. From the Definition 3.1, it is also obvious that for any $1 \le i \le k$,

$$|\mathbf{y} - \mathbf{Z}_i| - \delta \le |\mathbf{y} - \mathbf{Y}_j| \le |\mathbf{y} - \mathbf{Z}_i| + \delta, \quad \forall\, \mathbf{y} \in \Omega,\ j \in S_i.$$

For any $\mathbf{y}$, if $\mathbf{y} \in V_i(\mathbf{Z}) - \cup_{j\in S_i} V_j(\mathbf{Y})$ (i.e., $\mathbf{y} \in V_i(\mathbf{Z})$ and $\mathbf{y} \notin \cup_{j\in S_i} V_j(\mathbf{Y})$), then we will argue by contradiction that $\operatorname{dist}(\mathbf{y}, \partial V_i(\mathbf{Z}))$, the shortest distance from $\mathbf{y}$ to the boundary of $V_i(\mathbf{Z})$, satisfies

$$\operatorname{dist}(\mathbf{y}, \partial V_i(\mathbf{Z})) < \sigma = \frac{3\delta}{\alpha_{\mathbf{Z}}}, \tag{3.13}$$

where $\alpha_{\mathbf{Z}} = \dfrac{2}{\sqrt{(1 + R_\Omega/d_{\mathbf{Z}})^2 + (R_\Omega/d_{\mathbf{Z}})^2}}$ is a constant depending only on $\mathbf{Z}$ and R_Ω. Let us first assume that $\operatorname{dist}(\mathbf{y}, \partial V_i(\mathbf{Z})) \ge \sigma$. We will show that $\mathbf{y} \in \cup_{j\in S_i} V_j(\mathbf{Y})$. Take any $1 \le j \le k$ with $j \ne i$, without loss of generality, assume $V_i(\mathbf{Z})$ and $V_j(\mathbf{Z})$ are neighbor Voronoi cells. Notice that $V_i(\mathbf{Z})$ and $V_j(\mathbf{Z})$ are both convex polygons/polyhedra except the part belonging to the boundary of Ω, see Figure 3.2 for illustration in two dimensional cases.

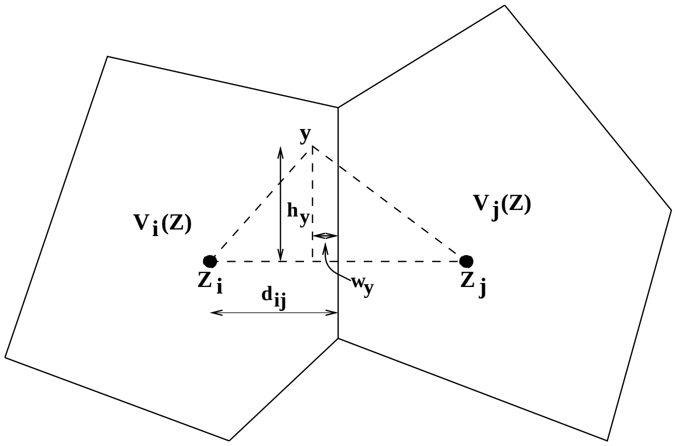

FIG. 3.2. *Two neighboring Voronoi regions.*

One can easily verify that

$$\begin{aligned}|\mathbf{y}-\mathbf{Z}_j|-|\mathbf{y}-\mathbf{Z}_i| &= \sqrt{(d_{ij}+w_{\mathbf{y}})^2+h_{\mathbf{y}}^2}-\sqrt{(d_{ij}-w_{\mathbf{y}})^2+h_{\mathbf{y}}^2}\\ &\geq \frac{2d_{ij}w_{\mathbf{y}}}{\sqrt{(d_{ij}+w_{\mathbf{y}})^2+h_{\mathbf{y}}^2}}\\ &= \frac{2w_{\mathbf{y}}}{\sqrt{(1+w_{\mathbf{y}}/d_{ij})^2+(h_{\mathbf{y}}/d_{ij})^2}}.\end{aligned}$$

Since $d_{ij} \geq d_{\mathbf{Z}}$, $R_\Omega \geq w_{\mathbf{y}} > \sigma$, and $h_{\mathbf{y}} \leq R_\Omega$, we obtain

$$\begin{aligned}|\mathbf{y}-\mathbf{Z}_j|-|\mathbf{y}-\mathbf{Z}_i| &\geq \frac{2\sigma}{\sqrt{(1+R_\Omega/d_{\mathbf{Z}})^2+(R_\Omega/d_{\mathbf{Z}})^2}}\\ &= 3\delta,\end{aligned}$$

Then, for any $j \in S_i$ and any $m \in S_l$ with $l \neq i$, it holds

$$\begin{aligned}|\mathbf{y}-\mathbf{Y}_j| &\leq |\mathbf{y}-\mathbf{Z}_i|+\delta\\ &\leq |\mathbf{y}-\mathbf{Z}_l|-3\delta+\delta\\ &\leq |\mathbf{y}-\mathbf{Y}_m|+\delta-2\delta\\ &\leq |\mathbf{y}-\mathbf{Y}_m|-\delta\\ &< |\mathbf{y}-\mathbf{Y}_m|.\end{aligned}$$

This means $\mathbf{y} \in \cup_{j\in S_i}V_j(\mathbf{Y})$ and gives us a contradiction with $\mathbf{y} \notin \cup_{j\in S_i}V_j(\mathbf{Y})$, so that the inequality (3.13) holds. Since $V_i(\mathbf{Z}) \subset \Omega$ is convex and $\mathrm{diam}(\Omega) = R_\Omega < +\infty$, it holds that

$$m(\partial V_i(\mathbf{Z})) \leq S_d\left(\frac{R_\Omega}{2}\right)^{d-1} \tag{3.14}$$

where $S_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ denotes the hyper-surface area of an d-sphere of unit radius, see [3] for details. Then the inequality (3.13) with (3.14) implies that

$$m(V_i(\mathbf{Z})-\cup_{j\in S_i}V_j(\mathbf{Y})) \leq \frac{\sigma S_d R_\Omega^{d-1}}{2^{d-1}}.$$

Then selecting

$$\delta = \delta_\epsilon := \min\left(\frac{d_{\mathbf{Z}}}{2},\ \frac{2^{d-1}\eta_{\epsilon/k}\alpha_{\mathbf{Z}}}{3S_d R_\Omega^{d-1}}\right)$$

ensures that $m(V_i(\mathbf{Z}) - \cup_{j\in S_i} V_j(\mathbf{Y})) \leq \eta_{\epsilon/k}$ Due to Lemma 3.1, this means that

$$\begin{aligned}\sum_{i=1}^{k} M\left(V_i(\mathbf{Z}) - \cup_{j\in S_i} V_j(\mathbf{Y})\right) &= \sum_{i=1}^{k} \int_{V_i(\mathbf{Z})-\cup_{j\in S_i} V_j(\mathbf{Y})} \rho(\mathbf{y})\, \mathrm{d}\mathbf{y} \\ &\leq \sum_{i=1}^{k} \frac{\epsilon}{k} \\ &\leq \epsilon\end{aligned}$$

and thus the result holds. □

REMARK 3.2. *In the case that $\rho \in L^q(\Omega)$ for some $q > 1$, the Hölder inequality again can be used to find a δ_ϵ in the above lemma which does not depend on the constant $\eta_{\epsilon/k}$ such as*

$$\delta = \delta_\epsilon := \min\left(\frac{d_{\mathbf{Z}}}{2},\ \frac{2^{d-1}\alpha_{\mathbf{Z}}}{3S_d R_\Omega^{d-1}}\left(\frac{\epsilon}{k\|\rho\|_{L^q(\Omega)}}\right)^{q/(q-1)}\right).$$

3.2. Non-degeneracy of limit points. Now, let us present our first important result.

THEOREM 3.5. *Let Assumption 3.1 be satisfied. Given $n \in \mathbb{N}$ and any initial point $\mathbf{Z}^0 \in \hat{\Omega}^n$. Let $\{\mathbf{Z}^i\}_{i=0}^\infty$ be the iterates of Lloyd algorithm starting with $\mathbf{Z}^0$. Then any limit point $\mathbf{Z}$ of these iterates is is non-degenerate (i.e., $\mathbf{Z} \in \hat{\Omega}^n$).*

Proof. Let $\mathbf{Z} \in \overline{\Omega}^n$ be any limit point of these iterates $\{\mathbf{Z}^i\}_{i=0}^\infty$. We will argue by contradiction that $\mathbf{Z} \in \hat{\Omega}^n$. Let us first assume that $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_n)$ is degenerate, i.e., in the sense that several of its components are equal. Without loss of generality, assume that $\mathbf{Z}_1 = \cdots = \mathbf{Z}_r$ for some $1 < r \leq n$.

Since $\mathbf{Z}^i$'s are Lloyd iterates, by the continuity of $\mathcal{G}$ on $\overline{\Omega}^n$ and the monotone decrease of $\mathcal{G}$ along $\mathbf{Z}^i$'s due to (2.2), we know

$$\mathcal{G}(\mathbf{Z}^i) \searrow_{i\to+\infty} \mathcal{G}(\mathbf{Z}), \tag{3.15}$$

i.e., $\mathcal{G}(\mathbf{Z}^i)$ monotonically decreases and converges to $\mathcal{G}(\mathbf{Z})$. Let $\tilde{\mathbf{Z}} \in \hat{\Omega}^k$ ($k = n-r+1 < n$) be the set of points in $\mathbf{Z}$ without duplication, i.e.,

$$\tilde{\mathbf{Z}} = (\tilde{\mathbf{Z}}_1, \cdots, \tilde{\mathbf{Z}}_k) = (\mathbf{Z}_1, \mathbf{Z}_{r+1}, \cdots, \mathbf{Z}_n).$$

It is obvious that $\mathcal{G}(\tilde{\mathbf{Z}}) = \mathcal{G}(\mathbf{Z})$. From $\{\mathbf{Z}^i\}_{i=0}^\infty$, select a subsequence $\{\mathbf{Z}^{i_j}\}_{j=0}^\infty$ such that $\mathbf{Z}^{i_j} \to \mathbf{Z}$ as $j \to +\infty$.

Now define

$$\alpha = \frac{M(V_1(\tilde{\mathbf{Z}}))}{2r}, \quad \text{and} \quad \beta = \frac{\eta_{\alpha/4}^2}{64R_\Omega^{2d-1}}, \quad \text{and} \quad \gamma = \frac{\alpha\beta^2}{4}.$$

Obviously we have $\alpha, \beta, \gamma > 0$ since $V_1(\tilde{\mathbf{Z}}))$ is a convex set with positive measure.

Now let us select a j^* large enough such that the following inequalities hold:

$$M(\cup_{m=1}^{r} V_m(\mathbf{Z}^{i_{j^*}})) > \frac{M(V_1(\tilde{\mathbf{Z}}))}{2}, \tag{3.16}$$

$$\mathcal{G}(\mathbf{Z}^{i_{j^*}}) - \mathcal{G}(\tilde{\mathbf{Z}}) < \frac{\gamma}{2}, \tag{3.17}$$

$$|\mathbf{Z}_m^{i_{j^*}} - \mathbf{Z}_m| < \frac{\beta}{4}, \quad m = 1, 2, \cdots, n. \tag{3.18}$$

Inequality (3.16) holds for large j^* by Lemma 3.4 using $\epsilon = \dfrac{M(V_1(\tilde{\mathbf{Z}}))}{2}$. Further, that also means that for some $m \in \{1, \cdots, r\}$, $M(V_m(\mathbf{Z}^{i_{j^*}})) > \alpha$. Without loss of generality, assume this m is 1, i.e., $M(V_1(\mathbf{Z}^{i_{j^*}})) > \alpha$. That gives us

$$\delta_{V_1(\mathbf{Z}^{i_{j^*}})} \geq \beta, \qquad \gamma_{V_1(\mathbf{Z}^{i_{j^*}})} \geq \gamma. \tag{3.19}$$

The inequality (3.18) and (3.19) imply

$$\mathrm{dist}(\mathbf{Z}_1^{i_{j^*}}, \partial V_1(\mathbf{Z}^{i_{j^*}})) \leq |\mathbf{Z}_1^{i_{j^*}} - \mathbf{Z}_1| + |\mathbf{Z}_2^{i_{j^*}} - \mathbf{Z}_1| < \frac{\beta}{2} \leq \frac{\delta_{V_1(\mathbf{Z}^{i_{j^*}})}}{2}.$$

Then by Lemma 3.3 and (3.19), we immediately obtain

$$\begin{aligned} \int_{V_1(\mathbf{Z}^{i_{j^*}})} & \rho(\mathbf{y}) |\mathbf{y} - \mathbf{c}_{V_1(\mathbf{Z}^{i_{j^*}})}|^2 \, \mathrm{d}\mathbf{y} \\ & \leq \int_{V_1(\mathbf{Z}^{i_{j^*}})} \rho(\mathbf{y}) |\mathbf{y} - \mathbf{Z}_1^{i_{j^*}}|^2 \, \mathrm{d}\mathbf{y} - \gamma_{V_1(\mathbf{Z}^{i_{j^*}})} \\ & \leq \int_{V_1(\mathbf{Z}^{i_{j^*}})} \rho(\mathbf{y}) |\mathbf{y} - \mathbf{Z}_1^{i_{j^*}}|^2 \, \mathrm{d}\mathbf{y} - \gamma. \end{aligned} \tag{3.20}$$

Now, let us estimate the quantization energy $\mathcal{G}(\mathbf{T}(\mathbf{Z}^{i_{j^*}}))$. Using the dual minimization property, we get

$$\begin{aligned} \mathcal{G}(\mathbf{T}(\mathbf{Z}^{i_{j^*}})) &= \mathcal{H}(\mathbf{T}(\mathbf{Z}^{i_{j^*}}), \mathbf{T}(\mathbf{Z}^{i_{j^*}})) \\ &\leq \mathcal{H}(\mathbf{Z}^{i_{j^*}}, \mathbf{T}(\mathbf{Z}^{i_{j^*}})) \\ &= \int_{V_1(\mathbf{Z}^{i_{j^*}})} \rho(\mathbf{y}) |\mathbf{y} - \mathbf{T}_1(\mathbf{Z}^{i_{j^*}})|^2 \, \mathrm{d}\mathbf{y} \\ &\qquad + \sum_{k=2}^{n} \int_{V_k(\mathbf{Z}^{i_{j^*}})} \rho(\mathbf{y}) |\mathbf{y} - \mathbf{T}_k(\mathbf{Z}^{i_{j^*}})|^2 \, \mathrm{d}\mathbf{y} \end{aligned}$$

It is easy to see that

$$\sum_{k=2}^{n} \int_{V_k(\mathbf{Z}^{i_{j^*}})} \rho(\mathbf{y}) |\mathbf{y} - \mathbf{T}_k(\mathbf{Z}^{i_{j^*}})|^2 \, \mathrm{d}\mathbf{y} \leq \sum_{k=2}^{n} \int_{V_k(\mathbf{Z}^{i_{j^*}})} \rho(\mathbf{y}) |\mathbf{y} - \mathbf{Z}_k^{i_{j^*}}|^2 \, \mathrm{d}\mathbf{y}. \tag{3.21}$$

since $\mathbf{T}_k(\mathbf{Z}^{i_{j^*}})$ is the centroid of $V_k(\mathbf{Z}^{i_{j^*}})$. With (3.17), (3.20) and (3.21), we have

$$\begin{aligned}\mathcal{G}(\mathbf{T}(\mathbf{Z}^{i_{j^*}})) &\le \int_{V_1(\mathbf{Z}^{i_{j^*}})} \rho(\mathbf{y})|\mathbf{y}-\mathbf{Z}_1^{i_{j^*}}|^2 \,\mathrm{d}\mathbf{y} - \gamma \\ &\qquad + \sum_{k=2}^{n} \int_{V_k(\mathbf{Z}^{i_{j^*}})} \rho(\mathbf{y})|\mathbf{y}-\mathbf{Z}_k^{i_{j^*}}|^2 \,\mathrm{d}\mathbf{y} \\ &= \mathcal{G}(\mathbf{Z}^{i_{j^*}}) - \gamma \\ &< \mathcal{G}(\tilde{\mathbf{Z}}) + \frac{\gamma}{2} - \gamma \\ &= \mathcal{G}(\tilde{\mathbf{Z}}) - \frac{\gamma}{2}\end{aligned}$$

that contradicts with the fact (3.15). We conclude that $\mathbf{Z} \in \hat{\Omega}^n$, i.e., any limit point of the Lloyd iteration cannot be a degenerate point. □

THEOREM 3.6. *Let Assumption 3.1 be satisfied. Given $n \in \mathbb{N}$ and any initial point $\mathbf{Z}^0 \in \hat{\Omega}^n$. Let $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ be the iterates of Lloyd algorithm starting with $\mathbf{Z}^0$. Then there exists a constant $D_{\mathbf{Z}^0} > 0$ depending only on n, Ω, ρ and $\mathbf{Z}^0$ such that for any $i \ge 0$,*

$$|\mathbf{Z}_j^i - \mathbf{Z}_k^i| > D_{\mathbf{Z}^0}$$

for all $j, k \in \{1, \cdots, n\}$ and $j \ne k$.

Proof. We will use the contradiction argument again. Let us assume that the theorem is not true, then $\forall m \in \mathbb{N}$, there always exists an index i_m, some $j_m, k_m \in \{1, \cdots, n\}$ and $j_m \ne k_m$, such that

$$|\mathbf{Z}_{j_m}^{i_m} - \mathbf{Z}_{k_m}^{i_m}| < \frac{1}{m}.$$

Let us consider the subsequence $\{\mathbf{Z}^{i_m}\}_{m=1}^{\infty}$. Without loss of generality, assume the index set $I = \{i_m\}_{m=1}^{\infty}$ is a monotonically increasing sequence (otherwise we can do that by adding some restriction such as $i_{k+1} > i_k$ during choosing $\{\mathbf{Z}^{i_m}\}_{m=1}^{\infty}$). Notice that the set $JK = \{(j_m, k_m)\}_{m=1}^{\infty}$ has at most $n(n-1)$ different elements, so there exists a pair $\{(j^*, k^*)\} \in JK$ that appears infinitely many times along the increasing of m and let us record their appearances by the increasing sequence $\{m_1, m_2, \cdots\} \subset \mathbb{N}$. This means that

$$|\mathbf{Z}_{j^*}^{i_{m_p}} - \mathbf{Z}_{k^*}^{i_{m_p}}| < \frac{1}{m_p}, \quad \forall\, p \in \mathbb{N}. \tag{3.22}$$

Since $\{\mathbf{Z}^{i_{m_p}}\}$ are bounded, there always exists a convergent subsequence of $\{\mathbf{Z}^{i_{m_p}}\}$, and assume $\mathbf{Z}$ is the corresponding limit point. Due to (3.22), we must have $\mathbf{Z}_{j^*} = \mathbf{Z}_{k^*}$, that means $\mathbf{Z}$ is a degenerate point. However, $\mathbf{Z}$ is also a limit point of the original Lloyd iterates $\{\mathbf{Z}^i\}_{i=0}^{\infty}$, according to Theorem 3.5, $\mathbf{Z}$ could not be a degenerate point. We then get a contradiction and so that the theorem holds. □

REMARK 3.3. *In Theorem 3.6, we are only able to show existence of the lower bound of the distances between any pair of generators of all Lloyd iterates starting with $\mathbf{Z}^0$. For the one dimensional case, an explicit estimate was derived in [7] when the density function is strictly positive and smooth.*

A direct consequence of Theorem 3.6 is the following corollary.

COROLLARY 3.7. *Let Assumption 3.1 be satisfied. Given $n \in \mathbb{N}$ and any initial point $\mathbf{Z}^0 \in \hat{\Omega}^n$. Let $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ be the iterates of Lloyd algorithm starting with $\mathbf{Z}^0$. Then* *the Lloyd map is continuous at any of the iterates* $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ *and* $m(V_j(\mathbf{Z}^i)) \geq (\frac{D_{\mathbf{Z}^0}}{2})^d$ *for any* $1 \leq j \leq n$ *and* $i \geq 0$.

3.3. Main result. Finally let's present our main result in the following theorem:

THEOREM 3.8. (Weak Global Convergence) *Let Assumption 3.1 be satisfied. Given* $n \in \mathbb{N}$ *and any initial point* $\mathbf{Z}^0 \in \hat{\Omega}^n$*. Let* $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ *be the iterates of Lloyd algorithm starting with* $\mathbf{Z}^0$*. Then*

(1) $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ *is weakly convergent (i.e.,* $\lim_{i\to+\infty} \nabla\mathcal{G}(\mathbf{Z}^i) = 0$*) and any limit point of* $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ *is also a non-degenerate critical point of the quantization energy* $\mathcal{G}$ *(and thus a CVT).*

(2) Moreover, it also holds that $\lim_{i\to+\infty} \|\mathbf{Z}^{i+1} - \mathbf{Z}^i\| = 0$.

Proof. Using the result of Corollary 3.7, we see that we can define a compact set (away from the degenerate points) such that for any initial point $\mathbf{Z}^0 \in \hat{\Omega}^n$, the Lloyd iterates (the images of the Lloyd maps) will stay in such a compact set. Thus, we can apply Theorems 2.1 and Corollary 3.7 to deduce the weak global convergence of the Lloyd iterations. The first part of the theorem is then complete with the help of Theorem 3.5.

From the Corollary 3.7, we also know that for any $1 \leq k \leq n$ and $i \geq 0$,

$$m(V_k(\mathbf{Z}^i)) \geq \left(\frac{D_{\mathbf{Z}^0}}{2}\right)^d > 0.$$

Together with Lemma 3.1 and the assumption that ρ is positive almost everywhere, we know there exists some constant $M^* > 0$ such that

$$M(V_k(\mathbf{Z}^i)) \geq M^* > 0. \tag{3.23}$$

Using the equation (2.1), we then get

$$\begin{aligned}
\|\nabla\mathcal{G}(\mathbf{Z}^i)\| &= \left(\sum_{k=1}^{n} \Big|\frac{\partial\mathcal{G}}{\partial\mathbf{z}_k}(\mathbf{Z}^i)\Big|^2\right)^{1/2} \\
&= \left(\sum_{k=1}^{n} |2M(V_k(\mathbf{Z}^i))(\mathbf{Z}_k^i - \mathbf{T}_k(\mathbf{Z}^i))|^2\right)^{1/2} \\
&\geq 2M^* \left(\sum_{k=1}^{n} |\mathbf{Z}_k^i - \mathbf{T}_k(\mathbf{Z}^i)|^2\right)^{1/2} \\
&= 2M^*\|\mathbf{Z}^i - \mathbf{Z}^{i+1}\|.
\end{aligned} \tag{3.24}$$

Consequently, we get

$$0 \leq \lim_{i\to+\infty} \|\mathbf{Z}^{i+1} - \mathbf{Z}^i\| \leq \frac{1}{2M^*} \lim_{i\to+\infty} \|\nabla\mathcal{G}(\mathbf{Z}^i)\| = 0$$

which deduces the second part of this theorem. ☐

Theorem 3.8 in fact tells us that the Lloyd method always terminates under general practical stopping criteria.

4. Conclusions and Open Questions. In this paper, we prove that any limit point of the Lloyd iteration is non-degenerate provided that Ω is a convex and bounded set and ρ belongs to $L^1(\Omega)$ and is positive almost everywhere in any dimensional spaces. It follows that, for every initial point set, the Lloyd iteration always approaches the set of non-degenerate critical points of the so-called quantization energy (which is exactly the set of CVTs). These results go beyond those presented in earlier papers; they enlarge the class of densities in the one-dimensional case and provide a stronger meaning to the statement of global convergence of the Lloyd iteration by eliminating the possibility of degeneracy. Although the proofs in their generality hold for all L^1 density functions, we have been able to provide a more constructive characterization for the constants δ_V and δ_ϵ in the case of $L^q, q > 1$, which possess their own merit from practical point of view.

There are still some open questions. Throughout our proof, the convexity of Ω is required in many places. The first question is whether this condition is really necessary or can Ω be a nonconvex set with some constraints. Another question is under which conditions weak convergence transforms into a strong single limit-point convergence. Also, it remains to explore whether the proof can be extended to replace the density function with a more general measure: ideally, it would be extended to a set of measures more general than those considered in [24]. Finally, it is also very meaningful and interesting to study the convergence properties of the Lloyd algorithm when the CVT is defined under some other error weighting functions, such as the very commonly used l^1 and l^∞ that arise in many engineering applications. The geometric proof provided here clearly provides room for such generalizations, however, the detailed analysis would complicate current presentation hence it is left to a separate treatment. These questions and generalizations are the focus of current research and will be addressed in subsequent publications.

5. Acknowledgment. The authors are grateful to Noel Walkington of Carnegie Mellon University, Qiang Du of Penn State University and Max Gunzburger of Florida State University for many stimulating discussions. Special thanks are extended to the anonymous referees for providing valuable suggestions that improved the quality of the presentation and for bringing several relevant references to the authors' attention. M. Emelianenko also acknowledges the support provided by the Center for Nonlinear Analysis for the part of the work that was done at Carnegie Mellon University.

Appendix A. Proof of Lemma 3.1.

Proof. Suppose that the claim is false. Then there exists a $\epsilon_0 > 0$ such that $\forall k \in \mathbb{N}$, $\exists A_k \subset \Omega$ with $m(A_k) \leq (1/2)^{k+1}$, it holds that $\int_{A_k} \rho(\mathbf{y})\ \mathrm{d}\mathbf{y} > \epsilon_0$. Let $B_k = \cup_{i=k}^{\infty} A_i$, clearly $m(B_k) \leq \sum_{k=1}^{\infty} m(A_i) \leq (1/2)^k$ and $\int_{B_k} \rho(\mathbf{y})\ \mathrm{d}\mathbf{y} > \epsilon_0$ for all $k \in \mathbb{N}$. Since $\{\int_{B_k} \rho(\mathbf{y})\ \mathrm{d}\mathbf{y}\}_{k=1}^{\infty}$ is a positive nonincreasing sequence, it is easy to see that

$$\lim_{k\to+\infty} \int_{B_k} \rho(\mathbf{y})\ \mathrm{d}\mathbf{y} \geq \epsilon_0 > 0.$$

On the other hand, notice that $B_1 \supset B_2 \supset \cdots$, by the Dominated Convergence Theorem in measure theory, we have

$$\begin{aligned}\lim_{k\to+\infty} \int_{B^*} \rho(\mathbf{y})\ \mathrm{d}\mathbf{y} &= \lim_{k\to+\infty} \int_{\Omega} (\mathbf{1}_{B_k}(\mathbf{y})\rho(\mathbf{y}))\ \mathrm{d}\mathbf{y} \\ &= \int_{\Omega} \lim_{k\to+\infty} (\mathbf{1}_{B_k}(\mathbf{y})\rho(\mathbf{y}))\ \mathrm{d}\mathbf{y} \\ &= 0\end{aligned}$$

that gives us a contradiction, so this lemma holds. □

REFERENCES

[1] E.F. Abaya and G. Wise, Convergence of vector quantizers with application to optimal quantization, em SIAM J. Appl. Math. **44**, 1984, pp.183-189.

[2] E.S. Barnes and N.J. Sloane, The optimal lattice quantizer in three dimensions, *SIAM J. Algebraic Discrete Methods* **4**, 1983, pp. 31-40.

[3] J.H. Conway and N.J. Sloane, *Sphere Packings, Lattices and Groups*, Springer-Verlag, New York, 1988.

[4] Q. Du, V. Faber, and M. Gunzburger, Centroidal Voronoi tessellations: Applications and algorithms, *SIAM Review* **41**, 1999, pp. 637-676.

[5] Q. Du, M. Gunzburger, and L. Ju, Constrained centroidal Voronoi tessellations on general surfaces, *SIAM J. Sci. Comput.* **24**, 2003, pp. 1488-1506.

[6] Q. Du and M. Emelianenko, Uniform convergence of a nonlinear energy-based multilevel quantization scheme via centroidal Voronoi tessellations, *SIAM J. Numer. Anal.*, to appear, 2007.

[7] Q. Du, M. Emelianenko, and L. Ju, Convergence of the Lloyd algorithm for computing centroidal Voronoi tessellations, *SIAM J. Numer. Anal.* **44**, 2006, pp. 102-119.

[8] Q. Du and D. Wang, Anisotropic centroidal Voronoi tessellations and their applications, *SIAM. J. Sci. Comput.* **26**, 2005, pp. 737-761.

[9] P. Fleischer, Sufficient conditions for achieving minimum distortion in a quantizer, *IEEE Int. Convention Record*, I, 1964, pp. 104-111.

[10] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*, Kluwer, Boston, 1992.

[11] R. Gray, J. Kieffer, and Y. Linde, Locally optimal block quantizer design, *Inform. Control* **45**, 1980, pp. 178-198.

[12] R. Gray and D. Neuhoff, Quantization, *IEEE Trans. Inform. Theory* **44**, 1998, pp. 2325-2383.

[13] P. M. Gruber, Optimum quantization and its applications, *Advances in Mathematics* **186**, 2004, pp. 456-497.

[14] L. Ju, Q. Du, and M. Gunzburger, Probabilistic methods for centroidal Voronoi tessellations and their parallel implementations, *Parallel Comput.* **28**, 2002, pp. 1477-1500.

[15] J. Kieffer, Uniqueness of locally optimal quantizer for log-concave density and convex error function, *IEEE Trans. Infor. Theory* **29**, 1983, pp. 42-47.

[16] T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu, An efficient k-means clustering algorithm: Analysis and implementation, *IEEE Trans. Pattern Anal. Machine Intell.* **24**, 2002, pp. 881-892.

[17] Y. Linde, A. Buzo and R. Gray, An algorithm for vector quantizer design, *IEEE Trans. Commun.* **28**, 1980, pp.84-95.

[18] S. Lloyd, Least squares quantization in PCM, *IEEE Trans. Infor. Theory* **28**, 1982, pp. 129-137.

[19] F. Lu and G. Wise, A further investigation of the Lloyd-Max algorithm for quantizer design, *Proceedings of Twenty-First Annual Allerton Conference on Communication, Control, and Computing*, University of Illinois, 1983, pp. 481-490.

[20] D.G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA, 1984.

[21] J.B. MacQueen, Some methods for classification and analysis of multivariate observations, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281-297.

[22] D. Newman, The hexagon theorem, *IEEE Trans. Inform. Theory* **28**, 1982, pp. 137-139.

[23] S.Z. Selim and M.A. Ismail, K-means-type algorithms: A generalized convergence theorem and characterization of local optimality, *IEEE Trans. Pattern Anal. Machine Intell.* **6**, 1984, pp. 81-87.

[24] J. Sabin and R. Gray, Global Convergence and Empirical Consistency of the Generalized Lloyd Algorithm, *IEEE Trans. on Inform. Theory* **32**, 1986, pp. 148-155.

[25] A.V. Trushkin, Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weghting functions, *IEEE Trans. Inform. Theory* **28**, 1982, pp. 187-198.

[26] X. Wu, On Convergence of Lloyd's Method I, *IEEE Trans. Inform. Theory* **38**, 1992, pp. 171-174.