Uniform Convergence of a Multilevel Energy-based Quantization Scheme

Maria Emelianenko [1] and Qiang Du [1]

Pennsylvania State University, University Park, PA 16803
emeliane@math.psu.edu and qdu@math.psu.edu

Abstract

Quantization has diverse applications in many areas of science and engineering. In this paper, we present a new nonlinear multilevel algorithm that accelerates existing numerical approaches for finding optimal quantizers. Both theoretical framework for the convergence analysis and results of computational experiments are provided.

1 Introduction

A vector quantizer maps N-dimensional vectors in the domain $\Omega \subset \mathbb{R}^N$ into a finite set of vectors $\{\mathbf{z}_i\}_{i=1}^k$. Each vector $\mathbf{z}_i$ is called a code vector or a *codeword*, and the set of all the codewords is called a codebook. A special quantization scheme is given by the Voronoi tessellation which associates with each codeword $\mathbf{z}_i$, also called a *generator*, a nearest neighbor region that is called a Voronoi region $\{V_i\}_{i=1}^k$. That is, for each i, V_i consists of all points in the domain Ω that are closer to $\mathbf{z}_i$ than to all the other generating points, and a Voronoi tessellation refers to the tessellation of a given domain by the Voronoi regions $\{V_i\}_{i=1}^k$ associated with a set of given generating points $\{\mathbf{z}_i\}_{i=1}^k \subset \Omega$.

For a given density function ρ defined on Ω, we may define the centroids, or mass centers, of regions $\{V_i\}_{i=1}^k$ by

$$\mathbf{z}_i^* = \Big(\int_{V_i} \mathbf{y}\rho(\mathbf{y})\,d\mathbf{y}\Big)\Big(\int_{V_i} \rho(\mathbf{y})\,d\mathbf{y}\Big)^{-1}. \tag{1}$$

Then, an *optimal quantization* may be constructed through a *centroidal Voronoi tessellation* (CVT) for which the generators of the Voronoi tessellation themselves are the centroids of their respective Voronoi regions, in other words, $\mathbf{z}_i = \mathbf{z}_i^*$ for all i. Besides providing an optimal least square vector quantizer design in electrical engineering applications [6],[7],[16], the concept

of CVT has other diverse applications in many areas of science and engineering, such as image and data analysis, resource optimization, sensor networks, and numerical partial differential equations [1],[2],[8],[9],[12],[14]. We refer to [1] for a more comprehensive review of the mathematical theory and diverse applications of CVTs.

In the seminal work of Lloyd on the least square quantization [13], one of the algorithms proposed for computing optimal quantizers is an iterative algorithm consisting of the following simple steps: starting from an initial quantization (a Voronoi tessellation corresponding to an old set of generators), a new set of generators is defined by the mass centers of the Voronoi regions. This process is continued until certain stopping criterion is met.

Given a set of points $\{\mathbf{z}_i\}_{i=1}^k$ and a tessellation $\{V_i\}_{i=1}^k$ of the domain, we may define the *energy functional* or the *distortion value* for the pair $(\{\mathbf{z}_i\}_{i=1}^k, \{V_i\}_{i=1}^k)$ by:

$$\mathcal{H}\Big(\{\mathbf{z}_i\}_{i=1}^k, \{V_i\}_{i=1}^k\Big) = \sum_{i=1}^k \int_{V_i} \rho(\mathbf{y})|\mathbf{y}-\mathbf{z}_i|^2\, d\mathbf{y}\,.$$

The minimizer of $\mathcal{H}$, that is, the optimal quantizer, necessarily forms a CVT which illustrates the optimization property of the CVT [1]. The terms optimal quantizer and CVT are thus to be used interchangeably in the sequel. It is also easy to see that the Lloyd algorithm is an energy descent iteration, which gives strong indications to its practical convergence.

Lloyd's algorithms sparked enormous research efforts in later years [7],[8],[9] and their variants have been proposed and studied in many contexts for different applications. For modern applications of the CVT concept in large scale scientific and engineering problems such as data communication and mesh generation, efficient algorithms for computing the CVTs play crucial roles. In this short paper, we first discuss some convergence theory recently derived in [3] on the Lloyd's algorithm to motivate our ongoing work. Then we outline a new multilevel approach to the optimal quantization problem introduced recently in [4],[5] which can be used to accelerate the convergence of the Lloyd's algorithm. We discuss the idea of a dynamic nonlinear preconditioner and also give a convergence theorem as well as some numerical results.

2 Convergence properties of the Lloyd's iteration

Even with their great success in practice, only limited rigorous results on the convergence properties of Lloyd's iteration have been obtained and many important computational issues remain to be explored [1]. Some important characterizations of convergence for the Lloyd's scheme have been obtained recently in [3]. The results stated below demonstrate the global convergence properties of the Lloyd iteration and its relationship to the critical points of the energy functional.

Theorem 2.1 *Any limit point of Lloyd algorithm is a fixed point of the Lloyd map, and this determines a stationary point of* $\mathcal{H}$ *. The set of limit points share the same distortion value* $\mathcal{H}$ *for a given iteration.*

Theorem 2.2 *If the iterations in the Lloyd algorithm stay in a compact set where the Lloyd map* T *is continuous, then the algorithm is globally convergent to a critical point of* $\mathcal{H}$ *.*

We refer to [3] for the proofs and further discussions of related results.

Beyond the study on the global convergence, the characterization of the convergence rate is often also important in practice. For instance, one may inquire if a geometric convergence rate can be established. This is indeed verified in [1] for the constant density function and later in [3] under strong type of log-concavity conditions, where the established geometric convergence rate r is shown to be of the order of $1 - ck^{-2}$, so that the Lloyd method slows down for large values of k, the total number of generators. Even in the one-dimensional case, both our theoretical estimates and the experiments indicate that the convergence of Lloyd iterations is at most linear.

3 The new energy-based nonlinear multilevel algorithm

The evidence of slow convergence of the Lloyd iteration and its descent properties motivated our search for a Lloyd iteration based numerical scheme with superior convergence properties.

One possible approach to the problem of speeding up convergence for Lloyd's method is to use a domain or space decomposition (or multigrid) strategies ([4],[5],[10],[11]). There are many ways one could implement such an algorithm in the context of CVTs. However, the problem of constructing a CVT is nonlinear in nature and hence cannot be analyzed using standard linear multigrid theory. Without using any type of linearization techniques, we hope to overcome the difficulties of the nonlinearity by essentially relying on the energy minimization.

3.1 Description of the algorithm

Our motivation in using the energy minimization approach was the optimality property of the CVTs mentioned above. The optimality property implies that at the optimal quantizer $\nabla\mathcal{H} = 0$.

Since the energy functional is in general non-convex, we use a dynamic nonlinear preconditioner to relate our problem to a convex optimization problem. More precisely, denote $R = \mathtt{diag}\{R_i^{-1}\}, i = 1, \ldots, k+1$ where $R_i = \int_{V_i} \rho(\mathbf{y})\, d\mathbf{y}$ are the masses of the corresponding Voronoi cells. We arrive at an equivalent formulation of the minimization problem: $R\nabla\mathcal{H} = 0$, or $\min ||R\nabla\mathcal{H}||^2$. This *preconditioning* makes the energy functional convex in a

large neighborhood of the minimizer and therefore the new formulation has advantages over the original problem. Hence if we define the set of iteration points $\mathbf{W}$ by

$$\mathbf{W} = \{(w_i)|_{i=0}^{k+1}| \quad 0 = w_0 \le w_i \le w_{i+1} \le w_{k+1} = 1, \ \forall 0 \le i \le k\} \ ,$$

we can base our new multilevel algorithm on the following nonlinear optimization problem

$$\min_{\mathbf{Z}\in\mathbf{W}} \tilde{\mathcal{H}}(\mathbf{Z}), \text{ where } \tilde{\mathcal{H}}(\mathbf{Z} = \{\mathbf{z}_i\}_{i=0}^{k+1}) = ||R\nabla\mathcal{H}(\{\mathbf{z}_i\}_{i=1}^{k}, \{V_i\}_{i=1}^{k})||^2 \qquad (2)$$

where $\{V_i\}_{i=1}^{k}$ is the Voronoi tessellation corresponding to the generators $\{\mathbf{z}_i\}_{i=1}^{k}$. For simplicity, consider the CVT on the one-dimensional unit interval $[0,1]$. Let S_k be the space of continuous piecewise linear functions with respect to the uniform partition of mesh size $1/(k+1)$ with a hierarchical basis $\{\{\psi_j^i\}_{j=1}^{n_i}\}_{i=1}^{H}$, where H is the number of levels. Let $\bar{\psi}_j^i = \{\psi_j^i(\frac{m}{k+1})\}_{m=0}^{k+1} \in \mathbb{R}^{k+2}$ and set $\mathbf{W}_i = \text{span}\{\bar{\psi}_j^i\}_{j=1}^{n_i}$. Using the above notations, we design a multilevel successive subspace correction algorithm as follows:

Algorithm 3.1 (Successive correction $V(\nu_1,\nu_2)$ scheme)

Input:
Ω, *the domain of interest;* ρ, *a probability distribution on* Ω;
k, *number of generators;*
$\mathbf{Z} = \{z_i\}_{i=0}^{k+1} \in \mathbf{W}$, *the ends plus the initial set of generators.*
Output:
$\mathbf{Z} = \{z_i\}_{i=0}^{k+1}$, *the ends plus the set of generators for CVT* $\{V_i\}_{i=1}^{k}$.
Method:
1. For i=H:-1:2
Repeat ν_1 *times:*
given $\mathbf{Z}$, *find* $\mathbf{Z} = \mathbf{Z} + \alpha_j^0\bar{\psi}_j^i \in \mathbf{W}$ *sequentially for* $1 \le j \le n_i$
such that $\tilde{\mathcal{H}}(\mathbf{Z} + \alpha_j^0\bar{\psi}_j^i) = \min_{\alpha_j} \tilde{\mathcal{H}}(\mathbf{Z} + \alpha_j\bar{\psi}_j^i)$
endfor
2. $\mathbf{Z} \leftarrow$ *CoarseGridSolve(*$\mathbf{Z}$*)*
3. For i=2:1:H
Repeat ν_2 *times:*
given $\mathbf{Z}$, *find* $\mathbf{Z} = \mathbf{Z} + \alpha_j^0\bar{\psi}_j^i \in \mathbf{W}$ *sequentially for* $1 \le j \le n_i$
such that $\tilde{\mathcal{H}}(\mathbf{Z} + \alpha_j^0\bar{\psi}_j^i) = \min_{\alpha_j} \tilde{\mathcal{H}}(\mathbf{Z} + \alpha_j\bar{\psi}_j^i)$
endfor
4. Repeat the procedure 1 to 3 until some stopping criterion is met.

Each step of the procedure outlined above involves solving a system of nonlinear equations which plays the role of relaxation. Solution at current iterate is updated after each nonlinear solve by the Gauss-Seidel type procedure, hence the resulting scheme is successive in nature. The more general algorithm and convergence results are given in our subsequent works.

It is worth noting that in the 1-dimensional case the set of basis functions

$$Q_i = [\bar{\psi}_1^i, \ldots, \bar{\psi}_{n_i}^i]^T \in R^{n_i \times k}$$

used at each iteration can be pre-generated using the recursive procedure: $Q_1 = I_{k\times k}$ and $Q_s = (\Pi_{i=1}^s P_i)Q_1$ where P_i is the basis transformation from space $\mathbf{W}_{i+1}$ to $\mathbf{W}_i$ which plays a role of a restriction operator.

Supply $\mathbf{W}$ with the following norm:

$$||y||_{1,\mathbf{W}}^2 = \frac{1}{k}\sum_{i=1}^{k+1}(y_i - y_{i-1})^2$$

Theorem 3.1 *Algorithm 3.1 converges uniformly in* $\mathbf{W}$ *for any density of the type* $\rho(x) = 1 + \epsilon g(x)$ *, where* $g(x)$ *is smooth and* ϵ *is small. Moreover,* $d_n = \tilde{\mathcal{H}}(u_n) - \tilde{\mathcal{H}}(u)$ *satisfies*

$$d_n \le r d_{n-1}, \quad r \in (0,1)$$

for some constant $r = \frac{C}{1+C}$ *, where* $C = C_1^2 C_2^2 L/K^3$ *, independent of the number of generators or the number of layers.*

To prove this result, we follow the framework of [15]. The key steps include demonstrating that for all densities of the given type there exist constants $K > 0, L > 0, p \ge q > 1$ such that

$$K||w - v||_{1,\mathbf{W}}^2 \le (\tilde{\mathcal{H}}'(w) - \tilde{\mathcal{H}}'(v), w - v) \le L||w - v||_{1,\mathbf{W}}^2, \forall w, v \in \mathbf{W}.$$

Moreover, the space decomposition (e.g. for the hierarchical basis) satisfies: 1) for any $v \in \mathbf{W}$, there exist $v_i \in \mathbf{W}_i$ such that

$$\sum_{i=1}^{H} v_i = v\ , \quad (\sum_{i=1}^{H} ||v_i||_{1,\mathbf{W}}^2)^{1/2} \le C_1 ||v||_{1,\mathbf{W}}\ ;$$

2) for any $w_{ij} \in \mathbf{W}, u_i \in \mathbf{W}_i, v_j \in \mathbf{W}_j$, we have

$$\sum_{i,j=1}^{H} (\tilde{\mathcal{H}}'(w_{ij} + u_i) - \tilde{\mathcal{H}}'(w_{ij}), v_j) \le C_2 (\sum_{i=1}^{H} ||u_i||_{1,\mathbf{W}_i}^2)^{1/2} (\sum_{j=1}^{H} ||v_j||_{1,\mathbf{W}_j}^2)^{1/2}\ .$$

The complete proof is given in [4] and is omitted here.

Corollary 3.2 *For the* hat *basis, we may take* $C_1 = 1$ *and* $C_2 = 2L$ *, and when* $\rho(x) = 1$ *,* $C = 4$ *.*

It follows that for a suitable choice of decomposition in 1D the asymptotic convergence factor of our multilevel algorithm is independent of the size of the problem and the number of grid levels, which gives a significant speedup comparing to other methods, like the traditional Lloyd iteration. We justify this in the numerical examples that follow.

3.2 Numerical results

Below are the computational results obtained for the V(1,1) multigrid implementation of the new algorithm in comparison with the regular Gauss-Seidel performance. We plot the convergence factor $\rho \approx \dfrac{z_{n+1}-z_n}{z_n-z_{n-1}}$ for each V(1,1) cycle with respect to the total number of generators (grid points) involved.

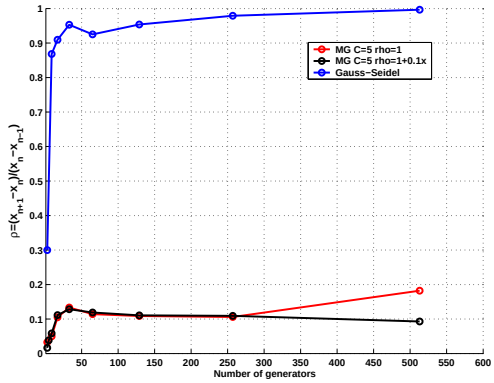

Fig. 1. Plot of the convergence factor over the number of generators for the multigrid method vs. regular Gauss-Seidel method for $\rho(x) = 1$ and $\rho(x) = 1 + 0.1x$

Figure 1 justifies the fact that the speed of convergence for the proposed scheme does not grow with the number of generators, while Table I shows the stabilization of the number of multigrid cycles $V(\nu_1, \nu_2, \mu)$ needed to reduce the error to $\epsilon = 10^{-12}$.

$k/V(\nu_1,\nu_2)$	V(1,0)	V(0,1)	V(1,1)	V(2,0)	V(0,2)	V(2,2)
3	7	8	6	6	7	4
5	11	11	8	8	8	6
9	13	14	9	9	9	7
17	18	18	12	12	12	8
33	21	20	13	12	13	8
65	21	22	12	12	12	8
129	21	21	12	12	12	8
257	20	23	12	12	13	7
513	20	22	12	11	13	7
1025	19	22	11	11	13	7

Table I. Number of $V(\nu_1, \nu_2)$ cycles needed to reduce the error to machine zero vs. the number of generators.

The geometric rate of energy and error reduction asserted by the Theorem 3.1 is confirmed by the experiments. Indeed, Figure 2 shows convergence history of a $V(1,1)$-cycle vs. total number of relaxations for the $k = 64$ case.

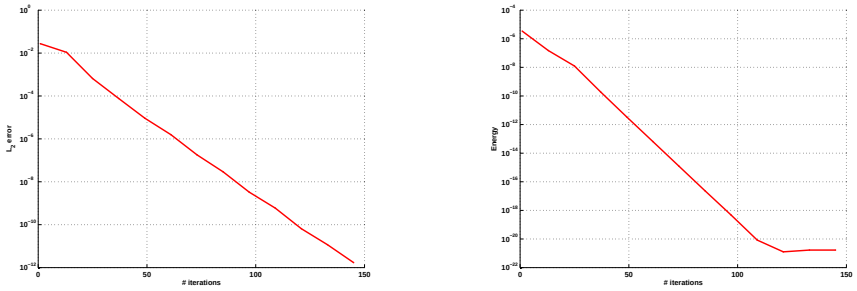

Fig. 2. (a) Convergence history for $k = 64$ generators (log-normal scale); (b)Energy reduction for $k = 64$ generators (log-normal scale)

The results for other nonlinear densities, though not shown here, also comply with the theoretical conclusions reached above (see [4]). Multidimensional extensions are discussed in [5] .

4 Applications

CVTs have a rich field of applications in various areas of mathematics as well as engineering. Here we provide a couple of geometric examples to give a flavor of the kind of problems that benefit from the study of this concept.

Figure 3 shows tessellations of the sphere for different density functions and an example of mesh generated by means of CVT. The point distributions

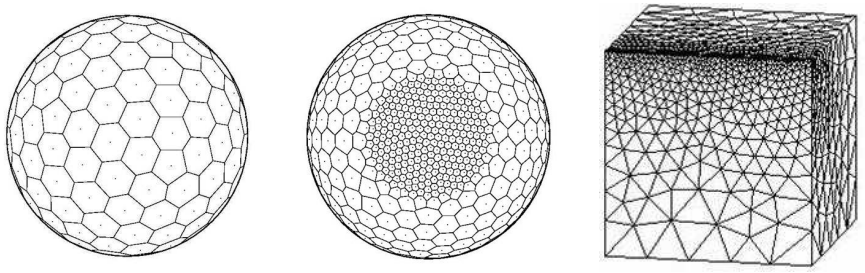

Fig. 3. (a),(b) Examples of CVTs for a sphere; (c) CVT-based mesh for the cube

generated via CVT can be used for vector quantization, optimal resource allocation, image compression, mesh generation and in many other applications [1] . In many of these applications, the efficiency of the numerical scheme plays a crucial role, so possible new approaches in accelerating existing numerical methods such as the multilevel approach discussed here are very important.

5 Conclusion

We introduced a new energy-based multilevel method for the optimal quantization where a dynamic nonlinear preconditioning is adopted to take advantage of a nonlinear convex optimization setting. Uniform convergence of the method with respect to the grid size and the number of grid levels and significant speedup comparing to Lloyd's method were demonstrated. More works are under way for the analysis of the multilevel scheme in higher dimensions.

References

1. Q. Du, V. Faber and M. Gunzburger: Centroidal Voronoi tessellations: applications and algorithms. SIAM Review, **41**, 637–676 (1999)
2. Q. Du and M. Gunzburger: Grid Generation and Optimization Based on Centroidal Voronoi Tessellations, to appear
3. Q. Du, M. Emelianenko and L. Ju: Convergence Properties of the Lloyd Algorithm for Computing the Centroidal Voronoi Tessellations, submitted to SIAM Num. Anal., (2004)
4. Q. Du and M. Emelianenko: Nonlinear optimization-based Multilevel Quantization Scheme, preprint, (2004)
5. Q. Du, M. Emelianenko and L. Zikatanov: An Energy-based Multilevel Quantization Scheme in Multidimension, in preparation, (2005)
6. A. Gersho: Asymptotically optimal block quantization. IEEE Trans. Inform. Theory, **25**, 373–380 (1979)
7. R. Gray and D. Neuhoff: Quantization. IEEE Trans. Inform. Theory, **44**, 2325–2383 (1998)
8. L. Ju, Q. Du, M. Gunzburger: Probabilistic methods for centroidal Voronoi tessellations and their parallel implementations. Para. Comp., **28**, 1477-1500 (2002)
9. L. Ju, M. Gunzburger and Q. Du: Meshfree, Probabilistic Determination of Points, Support Spheres, and Connectivities for Meshless Computing. Computer Methods in Applied Mechanics and Engineering, **191**, 1349-1366 (2002)
10. Y. Koren, I. Yavneh and A. Spira: A Multigrid Approach to the 1-D Quantization Problem, preprint, (2003)
11. Y. Koren, I. Yavneh: Adaptive Multiscale Redistribution for Vector Quantization, accepted to SIAM J. Sci. Comp. (2004)
12. Y. Linde, A. Buzo, and R. Gray: An algorithms for vector quantizer design. IEEE Trans. Comm., **28**, 84–95 (1980)
13. S. Lloyd: Least square quantization in PCM. IEEE Trans. Infor. Theory, **28**, 129–137 (1982)
14. J. MacQueen: Some methods for classification and analysis of multivariate observations. Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability, I, Ed. by L. LeCam & J. Neyman, 281–297 (1967)
15. X.-C. Tai and J. Xu: Global Convergence of Subspace Correction Methods for Some Convex Optimization Problems. Math. Comp., **71**, 105-124 (2002).
16. A. Trushkin: On the design of an optimal quantizer. IEEE Trans. Infor. Theory, **39**, 1180–1194 (1993)