A Multilevel Energy-based Quantization Scheme

Maria Emelianenko [1] and Qiang Du [2]

[1] Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213. masha@cmu.edu
[2] Department of Mathematics, Pennsylvania State University, University Park, PA 16802. qdu@math.psu.edu

Summary. 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 methods for finding optimal quantizers. Both a theoretical framework for the convergence analysis and results of some 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, yet popular, quantization scheme is given by the Voronoi tessellation associated with some codewords $\{\mathbf{z}_i\}$, also called *generators*.

A Voronoi tessellation for the given generating points $\{\mathbf{z}_i\}_{i=1}^k \subset \Omega$ refers to the tessellation of a given domain Ω by the Voronoi regions $\{V_i\}_{i=1}^k$ where, for each i, the Voronoi region V_i consists of all points in Ω that are closer to $\mathbf{z}_i$ than to the other generating points. For a density function ρ defined on Ω, we define the centroids, or mass centers, of the 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 [10],[11],[20], 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],[12],[13],[16],[18]. 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 [17], 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 a 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 Lloyd's algorithm is an energy descent iteration, which gives strong indications to its practical convergence.

Lloyd's algorithm has [11],[12],[13], in recent years, sparked an enormous research effort and its variants have been proposed and studied in many contexts for different applications. Efficient algorithms for computing the CVTs play crucial roles for modern application of CVT in large scale scientific and engineering problems such as data communication and mesh generation. In this short paper, we first discuss some convergence theory recently derived in [3] for Lloyd's algorithm to motivate our ongoing work. Then we outline a new multilevel approach to the optimal quantization problem introduced recently in [4],[6] which can be used to accelerate the convergence of 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 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 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 the 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 a strong type of log-concavity conditions, where the established geometric convergence rate r is shown to be of the order of $1 - ck^{-2}$, therefore 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 the 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. A possible approach to speeding up the convergence of Lloyd's method is to use a domain or space decomposition (or multigrid) strategies ([4],[5],[6],[14],[15]). There are many ways of implementing 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, let $R = \mathtt{diag}\{R_i^{-1}\}, i = 1, \ldots, k+1$ 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, defining the set of iteration points $\mathbf{W}$ by

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

our new multilevel algorithm is then based on the following nonlinear optimization problem

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

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 on a uniform mesh with k sub-intervals and a hierarchical basis $\{\{\psi_j^i\}_{j=1}^{n_i}\}_{i=1}^{H}$ with H 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 $\mathbf{W}_i = \text{span}\{\bar{\psi}_j^i\}_{j=1}^{n_i}$. We now present our 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 generators.*
Output:
$\mathbf{Z} = \{z_i\}_{i=0}^{k+1}$, *the ends plus 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* $1 \leq j \leq n_i$
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* $1 \leq j \leq n_i$
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 steps 1 to 3 until some stopping criterion is met.

Each step of the above algorithm involves solving a system of nonlinear equations which plays the role of a relaxation. The solution at the current iterate is updated after each nonlinear solve by the Gauss-Seidel type procedure, hence the resulting scheme is sequential in nature. Here ν_1, ν_2 denote the number of Gauss-Seidel iterations used at each level. Although $\nu_{1,2} = 1$ is sufficient in theory, larger values need to be used in practice due to the numerical error in solving the nonlinear system. The values $\nu_{1,2} \leq 3$ usually suffice for the optimization to reach saturation. More general algorithms and convergence results will be given in future work. It is worth noting that in the one-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 the space $\mathbf{W}_{i+1}$ to $\mathbf{W}_i$ which plays a role of a restriction operator.

Let us now state the following convergence theorem [4]:

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*

$$d_n \leq r d_{n-1},$$

for some constant $r = \frac{C}{1+C}$ *, where* C *is a constant independent of the number of generators or the number of levels.*

A proof of this result can be derived based on the framework of [19]. Supply $\mathbf{W}$ with the norm $||y||_{1,\mathbf{W}}^2 = \frac{1}{k}\sum_{i=1}^{k+1}(y_i - y_{i-1})^2$, the key steps of the proof include demonstrating the continuity and local convexity of the functional $\tilde{\mathcal{H}}$ with respect to the norm $||\cdot||_{1,\mathbf{W}}$, and establishing a strengthened Cauchy-Schwartz inequality with respect to the space decomposition $\mathbf{W} = \bigoplus_{i=1}^{H} \bar{\mathbf{W}}_i$ where $\bar{\mathbf{W}}_i = \mathbf{W}_i/\mathbf{W}_{i-1}$ for $i > 1$ and $\bar{\mathbf{W}}_1 = \mathbf{W}_1$. The complete proof is given in [4] and is omitted here. 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 in comparison to other methods, like the traditional Lloyd iteration. Moreover, we have

Corollary 3.2 *For the* hat *basis and the constant density function, we may take* $C = 4$ *and thus* $r = 0.8$ *.*

We note that the estimated convergence rate of $r = 0.8$ is merely an upper bound, and the actually convergence rate is much smaller in practice. We justify the above theoretical results in the numerical examples that follow.

3.2 Numerical results

For the V(1,1) multigrid implementation of the new algorithm, we compared our algorithm with the regular Gauss-Seidel performance. We plotted 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 k , the total number of generators (grid points) taken for $\rho(x) = 1$.

Figure 1 justifies the fact that the speed of convergence for the proposed scheme does not grow with the number of generators, while Table 1 shows the stabilization of the number of multigrid cycles $V(\nu_1, \nu_2)$ needed to reduce the error to $\epsilon = 10^{-12}$ in the $\rho(x) = 1$ case. The difference in the number of iterations required for $V(1,1)$ and $V(2,2)$ comes from the approximation error in solving the optimization problem at each level, which decreases quickly as the number of relaxations grows.

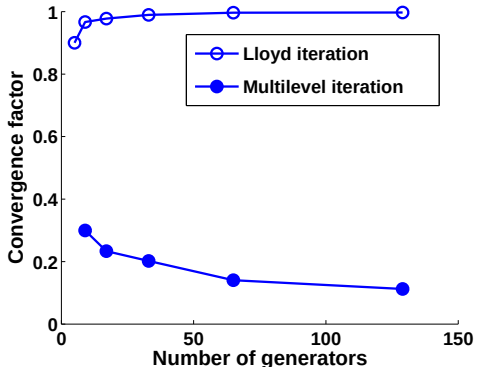

Fig. 1. Convergence factor ρ vs. k for the multigrid and Gauss-Seidel methods.

$k/V(\nu_1,\nu_2)$	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 1. Number of $V(\nu_1,\nu_2)$ cycles 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 the convergence history of the error (left) of a $V(1,1)$-cycle and the energy (right) vs. total number of relaxations for the $k=64$, $\rho(x)=1$ case (in log-normal scale).

The results for other nonlinear densities, though not shown here, are also consistent with the theoretical conclusions reached above (see [4]). Multidimensional extensions are discussed in [6].

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 [7] and an example of meshes generated by means of CVT [9].

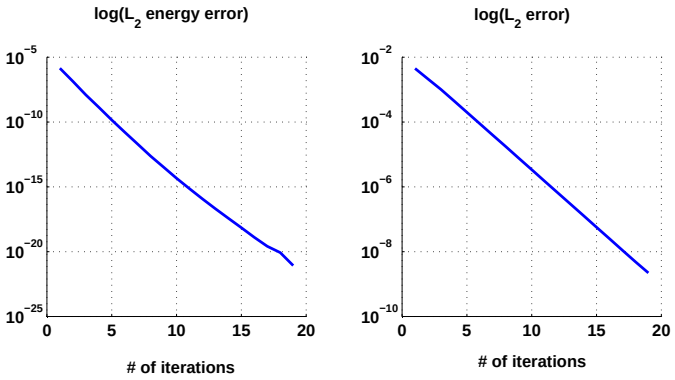

Fig. 2. Error and energy reduction of the $V(1,1)$-cycle.

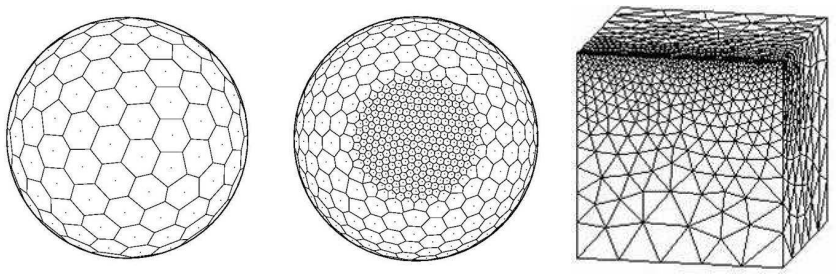

Fig. 3. Examples of CVTs for a sphere and a CVT-based mesh for a cube.

The point distributions 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

A new energy-based multilevel method is introduced for the optimal quantization which adopts dynamic nonlinear preconditioning to take advantage of a nonlinear convex optimization setting. The uniform convergence of the method with respect to the grid size and the number of grid levels and significant speedup compared to Lloyd's method are demonstrated. More work is under way for the multilevel scheme in higher dimensions.

Research support by the NSF through the grants DMR-0205232 and DMS-0409297 is gratefully acknowledged.

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, Applied Mathematics and Computation, **133**, 591–607 (2002)
3. Q. Du, M. Emelianenko and L. Ju: Convergence Properties of the Lloyd Algorithm for Computing the Centroidal Voronoi Tessellations, to appear in SIAM J. Numer. Anal., (2005)
4. Q. Du and M. Emelianenko: Uniform convergence of a nonlinear optimization-based multilevel quantization Scheme, preprint, (2005)
5. Q. Du and M. Emelianenko: Acceleration schemes for computing centroidal Voronoi tessellations to appear in special issue of Numerical Linear Algebra with Applications, (2006)
6. Q. Du, M. Emelianenko and L. Zikatanov: An Energy-based Multilevel Quantization Scheme in Multidimension, in preparation, (2005)
7. Q. Du, M. Gunzburger and L. Ju: Constrained centroidal Voronoi tessellations on general surfaces, SIAM J. Scientific Computing, **24**, 1488–1506 (2003)
8. Q. Du, M. Gunzburger, L. Ju and X. Wang: Centroidal Voronoi tessellation algorithms for image processing, to appear in J. Math. Imaging & Vision (2005)
9. Q. Du and D. Wang: Tetrahedral mesh generation and optimization based on centroidal Voronoi tessellations, IJNME, **56**, 1355–1373 (2002)
10. A. Gersho: Asymptotically optimal block quantization. IEEE Trans. Inform. Theory, **25**, 373–380 (1979)
11. R. Gray and D. Neuhoff: Quantization. IEEE Trans. Inform. Theory, **44**, 2325–2383 (1998)
12. L. Ju, Q. Du, M. Gunzburger: Probabilistic methods for centroidal Voronoi tessellations and their parallel implementations. Parallel Comput., **28**, 1477-1500 (2002)
13. L. Ju, M. Gunzburger and Q. Du: Meshfree, Probabilistic Determination of Points, Support Spheres, and Connectivities for Meshless Computing. Comput. Methods in Appl. Mech. and Engrg., **191**, 1349-1366 (2002)
14. Y. Koren, I. Yavneh and A. Spira: A Multigrid Approach to the 1-D Quantization Problem, preprint, (2003)
15. Y. Koren, I. Yavneh: Adaptive Multiscale Redistribution for Vector Quantization, accepted to SIAM J. Sci. Comp. (2004)
16. Y. Linde, A. Buzo, and R. Gray: An algorithms for vector quantizer design. IEEE Trans. Comm., **28**, 84–95 (1980)
17. S. Lloyd: Least square quantization in PCM. IEEE Trans. Infor. Theory, **28**, 129–137 (1982)
18. 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)
19. X.-C. Tai and J. Xu: Global Convergence of Subspace Correction Methods for Some Convex Optimization Problems. Math. Comp., **71**, 105-124 (2002).
20. A. Trushkin: On the design of an optimal quantizer. IEEE Trans. Infor. Theory, **39**, 1180–1194 (1993)