Rapid Near-Optimal VQ Design with a Deterministic Data Net

Michelle Effros[1] and Leonard J. Schulman[2]
California Institute of Technology · Pasadena, California 91125

We present a new algorithm for fixed-rate vector quantizer (VQ) design. A fixed-rate VQ of dimension d and rate $(\log K)/d$ represents each vector in $\mathbb{R}^d$ by one of K possible codewords. Given a pdf $p(\mathbf{x})$ on $\mathbb{R}^d$ and an integer $K \geq 1$, an optimal VQ $\{\mu_1^\star, \ldots, \mu_K^\star\} \subset \mathbb{R}^d$ achieves expected distortion $\Delta = \min_{\{\mu_1,\ldots,\mu_K\}\subset\mathbb{R}^d} \int_{\mathbb{R}^d} p(\mathbf{x}) \min_{k\in\{1,\ldots,K\}} ||\mathbf{x}-\mu_k||^2 d\mathbf{x}$.

Let $p(\mathbf{x})$ denote the empirical distribution of an n-point training set. Optimal VQ design is NP-hard even for $K = 2$ [1]. Iterative design finds a locally optimal VQ [4]. For any $\varepsilon > 0$, an ε-approximation algorithm guarantees a VQ with expected distortion $D \leq (1+\varepsilon)\Delta$. The best prior ε-approximation algorithms are a deterministic $O(\varepsilon^{-2K^2d} n \log^K n)$-time ε-approximation [5] and a randomized $O(\exp(\varepsilon^{-8}K^3(\ln K)(\ln \frac{1}{\varepsilon} + \ln K)) n \log^K n)$-time ε-approximation [2]. (The problem is called "K-clustering" and "K-median" in those papers.)

We present a new deterministic ε-approximation algorithm running in time quasilinear in n. The algorithm, which also serves as an approximation algorithm for the d-dimensional fixed-rate operational distortion-rate function, extends to a variety of network VQ problems.

Define a *data net* to be a set $\mathcal{Z} \subset \mathbb{R}^d$, regions $\{A_z\}_{z\in\mathcal{Z}}$, and mapping $\zeta : \mathbb{R}^d \to \mathcal{Z}$ such that

1. $\int_{A_z} p(x)||x-z||^2 d\mathbf{x} \leq \varepsilon\Delta/K$ for all $z \in \mathcal{Z}$
2. For any $\mu \in R^d$ and $x \notin A_{\zeta(\mu)}$,
$$||x-\zeta(\mu)||^2 \leq (1+\varepsilon)||x-\mu||^2.$$

Our deterministic algorithm designs a data net of size $M = c^d K^2(K^2+\varepsilon^{-2})\varepsilon^{-d-1}$ for some constant c, within time $Mn\log\log n$ and then returns the best codebook $\{\mu_1,\ldots,\mu_K\} \subset \mathcal{Z}$.

Theorem 1 *The above deterministic algorithm finds a rate-$(\log K)/d$ fixed-rate VQ with distortion $D \leq (1+\varepsilon)\Delta$ within time*
$$Mn\log\log n + M^{K+1}n.$$

Subsequent to codebook design, individual encodings can be peformed in time $\log(K/\varepsilon)$ using [3].

We can improve Theorem 1's dependence on K at the expense of d. The algorithm can also perform efficient VQ design for simply characterized continuous distributions.

The algorithm generalizes to give ε-approximation algorithms for many network VQ design problems. A few examples follow. Many more examples (including all combintations of listed examples) follow similarly.

Multiresolution VQ (MRVQ): A transmitter describes source X to L receivers. Receiver ℓ gets only the first $\sum_{i=1}^{\ell} \log K_i$ bits of the binary description.

Multiple description VQ (MDVQ): A transmitter sends L packets describing source X. Any subset of packets may be lost in transmission.

Side information VQ (SIVQ): A transmitter describes source X to a receiver that has access to side information unavailable to the encoder.

Broadcast VQ (BCVQ): A transmitter describes multiple sources to a family of decoders. Each source is intended for a distinct subset of the receivers, and each component of the description is received by a distinct subset of the receivers.

Joint source-channel VQ (JSCVQ): A transmitter describes a single source to a single receiver. The description may be corrupted during transmission.

Remote source VQ (RSVQ): An encoder observes a noisy copy of the true source and describes it to the decoder. The decoder reconstructs the true source as accurately as possible.

In each example, let S be the number of sources ($S = 1$ in all but the BCVQ), K be the maximal number of codewords that can be distinguished by any single decoder, T be the total number of d-dimensional codewords in the code, and Δ be the optimal performance (an expected distortion over some distribution on the reconstructions at different receivers).

Theorems 2–7 *In each scenario above, designing a data net of size $M = c^d K^2(K^2+\varepsilon^{-2})\varepsilon^{-d-1}$ for each source, and using the best T codewords from those data nets gives a code with performance $D \leq (1+\varepsilon)\Delta$ in time*
$$SMn\log\log n + (SM)^{T+1}n.$$

References

[1] P. Drineas, A. Frieze, R. Kannan, S. Vempala, and V. Vinay. Clustering in large graphs and matrices. In *Proc. 10th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 1999.

[2] W. Fernandez de la Vega, M. Karpinski, C. Kenyon, and Y. Rabani. Approximation schemes for clustering problems. In *Proc. 35'th Ann. Symp. on Theory of Computing (STOC)*, 2003.

[3] S. Har-Peled. A replacement for Voronoi diagrams in near linear size. In *Proc. 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2001.

[4] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design. *IEEE Trans. Comm.*, 28(1):84–95, 1980.

[5] J. Matousek. On approximate geometric k-clustering. *Discrete & Computational Geometry*, 24:61–84, 2000.

[1]M. Effros (effros@caltech.edu) is partially supported by NSF CCR-0220039 and Caltech's Lee Center for Advanced Networking.

[2]L. J. Schulman (schulman@caltech.edu) is partially supported by NSF CCR-0049092, the Charles Lee Powell Foundation, and a visiting membership at MSRI.