Optimal vector Quantization: from signal processing to clustering and numerical probability

Gilles Pagès

Université P.M. Curie - Paris 6

Optimal vector quantization has been originally introduced in Signal processing as a discretization method of random signals, leading to an optimal trade-off between the speed of transmission and the quality of the transmitted signal. In machine learning, similar methods applied to a dataset are the historical core of unsupervised classification methods known as "clustering". In both case it appears as an optimal way to produce a set of weighted prototypes (or codebook) which makes up a kind of skeleton of a dataset, a signal and more generally, from a mathematical point of view, of a probability distribution.

Quantization has encountered in recent years a renewed interest in various application fields like automatic classification, learning algorithms, optimal stopping and stochastic control, Backward SDEs and more generally numerical probability. In all these various applications, practical implementation of such clustering/quantization methods more or less rely on two procedures (and their countless variants): the Competitive Learning Vector Quantization ($CLVQ$) which appears as a stochastic gradient descent derived from the so-called *distortion* potential and the (randomized) Lloyd's procedure (also known as k-means algorithm, nuées dynamiques) which is but a fixed point search procedure. Batch version of those procedures can also be implemented when dealing with a dataset (or more generally a discrete distribution).

In a more formal form, if μ is probability distribution on an Euclidean space $\mathbb{R}^d$, the optimal quantization problem at level N boils down to exhibiting an N-tuple $(x_1^*, \ldots, x_N^*)$, solution to

$$\operatorname{argmin}_{(x_1,\ldots,x_N)\in(\mathbb{R}^d)^N} \int_{\mathbb{R}^d} \min_{1\leq i\leq N} |x_i - \xi|^2 \mu(d\xi)$$

and its distribution *i.e.* the weights $(\mu(C(x_i^*))_{1 \le i \le N}$ where $(C(x_i^*)$ is a (Borel) partition of $\mathbb{R}^d$ satisfying

$$C(x_i^*) \subset \left\{ \xi \in \mathbb{R}^d : |x_i^* - \xi| \le \min_{1 \le j \le N} |x_j^* - \xi| \right\}.$$

To produce an unsupervised classification (or clustering) of a (large) dataset $(\xi_k)_{1 \le k \le n}$, one considers its empirical measure

$$\mu = \frac{1}{n} \sum_{k=1}^{n} \delta_{\xi_k}$$

whereas in numerical probability $\mu = \mathcal{L}(X)$ where X is an $\mathbb{R}^d$-valued *simulable* random vector. In both situations, $CLVQ$ and Lloyd's procedures rely on massive sampling of the distribution μ.

As for clustering, the classification into N *clusters* is produced by the partition of the dataset induced by the Voronoi cells $C(x_i^*)$, $i = 1, \dots, N$ of the optimal quantizer.

In this second case, which is of interest for solving non linear problems like Optimal stopping problems (variational inequalities in terms of PDEs) or Stochastic control problems (HJB equations) in medium dimensions, the idea is to produce a quantization tree optimally fitting the dynamics of (a time discretization) of the underlying structure process.

We will explore (briefly) this vast panorama with a focus on the algorithmic aspects where few theoretical results coexist with many heuristics in a burgeoning literature. We will present few simulations in two dimensions.

References

[1] M. DUFLO (1996) (1997). *Algorithmes stochastiques*, coll. Mathématiques & Applications, **23**, Springer-Verlag, Berlin, 319p.

[2] A. GERSHO AND R.M. GRAY (1988). Special issue on Quantization, **I-II** (A. Gersho and R.M. Gray eds.), *IEEE Trans. Inform. Theory*, **28**.

[3] A. GERSHO AND R.M. GRAY (1992). *Vector Quantization and Signal Compression*. Kluwer, Boston.

[4] S. GRAF, H. LUSCHGY (2000). *Foundations of quantization for probability distributions*, LNM 1730, Springer, Berlin, 230p.

[5] H.J. KUSHNER, G.G. YIN (2003). Stochastic approximation and recursive algorithms and applications, 2nd edition, *Applications of Mathematics, Stochastic Modelling and Applied Probability*, **35**, Springer-Verlag, New York.

[6] G. Pagès (2013). P Introduction to vector quantization and its applications for numerics. CEMRACS 2013modelling and simulation of complex systems: stochastic and deterministic approaches, *ESAIM Proc. & Surveys*, **48**:29-79, EDP Sci., Les Ulis.

[7] G. Pagès, J. Printems (2009). Optimal quantization for finance: from random vectors to stochastic processes, chapter from *Mathematical Modeling and Numerical Methods in Finance* (special volume) (A. Bensoussan, Q. Zhang guest éds.), coll. *Handbook of Numerical Analysis* (P.G. Ciarlet Editor), North Holland, 595-649.

[8] G. Pagès, B. Wilbertz (2011). Optimal Delaunay et Voronoi quantization methods for pricing American options, pré-pub PMA, 2011, *Numerical methods in Finance* (R. Carmona, P. Hu, P. Del Moral, .N. Oudjane éds.), Springer, New York, 171-217.