Supplementary Material

Vijay Anavangot and Animesh Kumar

October 21, 2019

1 Product LALM Quantizer

{sec:typestyl

The vector quantization of the source is challenging for high dimensions ($d \gg 1$). An alternate suboptimal scheme is the product quantizer (PQ). It simplifies the design procedure as it requires to design for only d scalar quantizers. Eventhough, the MSE performance is affected by the choice of PQ, we value the computational simplicity that the method offers. We introduce a product-LALM quantizer which performs vector quantization of multi-dimensional data. We show that the product-LALM quantizer deviates from the optimal k-means quanitizer with increasing dimension.

1.1 Analysis of MSE Performance

The product-LALM quantizer performs piecewise-linear approximation along the marginal densities. Using the independence accross dimensions, the product-LALM results in a product of piecewise-linear line-segments in every hyper-rectangular (higher dimensional rectangle) voronoi region. Then, the density approximation can be expressed as a hyperbolic function;

$$f_{\text{app}}(x_1, \ldots, x_d) = \prod_{l=1}^{d}(a_l x_l + b_l) \tag{1}$$ {eq:prod-dens

for $(x_1, \ldots, x_d) \in \mathcal{R}_k$, where $\mathcal{R}_k$ is the hyper-rectangular region bounded by the box inequalities $q_{L,k}^{(l)} \leq x_l \leq q_{R,k}^{(l)}$. The notation $q_{L,k}^{(l)}$ and $q_{R,k}^{(l)}$ represents the immediate left and right neighbors of q_k along the dimension x_l for $l = 1, \ldots, d$. The approximation of the true density by the product form in (1), can be expressed in terms of the Taylor series expansion of multivariate function $f_{\vec{X}}(\vec{x})$ [**?**]. Assume the density function is d-times differentiable.

$$\begin{aligned} f_{\vec{X}}(\vec{x}) &= \sum_{|\alpha| \leq d} \frac{D^{\alpha} f(\vec{q}_i)}{\alpha!}(\vec{x} - \vec{q}_i)^{\alpha} + \sum_{|\alpha| = d} h_{\alpha}(\vec{x})(\vec{x} - \vec{q})^{\alpha}, \\ &= f_{\text{app}}(x) + \mathcal{O}(\varepsilon_K^d), \end{aligned} \tag{2}$$ {eq:Taylor_ex

where $\varepsilon_K = \max_k \left\{ \max_l |q_{R,k}^{(l)} - q_{L,k}^{(l)}| \right\}$, is the maximum length of any hyper-rectangle edge. The Taylor series approximation will be extensively used to prove the main result that follows. We first glean some terminology that will make the exposition of the proof tractable. These are listed in the table below

Scheme	`prod-LALM`	`Subopt k-means`	`k-means`
Voronoi	Rectangular : $\mathcal{R}_k$	Rectangular : $\mathcal{H}_k$	Polytope : $\mathcal{V}_k$
Density	$f_{\text{app}}(\vec{x})$	True $f_{\vec{X}}(\vec{x})$	True $f_{\vec{X}}(\vec{x})$
MSE	$\text{MSE}_{\text{pd-LALM}}$	$\text{MSE}_{\text{subopt}}$	$\text{MSE}_{\text{kmeans}}$

Theorem 1 (*MSE performance of product-LALM*)**.** *Let the MSE of the product-LALM and the optimal k-means be represented at $MSE_{pd\text{-}LALM}$ and MSE_{kmeans} respectively. Then,*

$$MSE_{pd\text{-}LALM} \geq MSE_{kmeans} + \mathcal{O}(\varepsilon_K^{2d}) \tag{3}$$

Proof. We note that at iteration $i = 0$, the voronoi regions of product-LALM and suboptimal k-means are matched, that is $\mathcal{H}_k^{(0)} = \mathcal{R}_k^{(0)}$. Note that the suboptimal k-means picks its rectangular voronoi region by having a fixed set of quantization levels as its neighbors (however this may not correspond to the voronoi region of the k-means). Using the optimality condition of Lloyd algorithm in each of the voronoi regions, we have

$$\int_{\mathcal{H}_k^{(0)}} (\vec{y} - \vec{q}_k^{\,(1)}) f_{\vec{X}}(\vec{y}) d\vec{y} = 0 \;\&\; \int_{\mathcal{R}_k^{(0)}} (\vec{y} - \vec{q}_{kA}^{\,(1)}) f_{\text{app}}(\vec{y}) d\vec{y} = 0.$$

We denote the quantization level of product-LALM by $\vec{q}_{kA}$ to distinguish it from the suboptimal kmeans. Subtracting the two integrals and rearranging the terms, we obtain the bound on the distance,

$$\|\vec{q}_k^{\,(1)} - \vec{q}_{kA}^{\,(1)}\| = \mathcal{O}(\varepsilon_K^d)$$

After iteration $i = 1$, the recomputed voronoi regions of the two schemes would be $\mathcal{H}_k^{(1)}$ and $\mathcal{R}_k^{(1)}$ respectively. The relation between the two are denoted using the shorthand notation,

$$\mathcal{H}_k^{(1)} = \mathcal{R}_k^{(1)} \pm \mathcal{O}(\varepsilon_K^d) \times \ldots \times \mathcal{O}(\varepsilon_K^d).$$

The above notation signifies that the voronoi (hyper)rectangle of the suboptimal k-means is within $\mathcal{O}(\varepsilon_K^d)$ of the product-LALM region. This change in the voronoi region would result in an additional factor in the distance between the quantization levels. That is, at iteration $i = 2$, (using the Lloyd optimality)

$$\begin{aligned} \|\vec{q}_k^{\,(2)} - q_{kA}^{(2)}\| &= \mathcal{O}(\varepsilon_K^d) \pm \mathcal{O}(\varepsilon^{d^2}), \\ &= (1+\gamma)\mathcal{O}(\varepsilon_K^d) \end{aligned}$$

In the above relation, the first term is as result of the density approximation and the second term is due to the voronoi regions. The factor $\gamma < 1$, due to the assumption that the density $f_{\vec{X}}(\vec{x})$ has a bounded gradient (and also since the data is bounded). Using the above fact inductively, we show that

$$\|\vec{q}_k^{\,(N)} - \vec{q}_{kA}^{\,(N)}\| = (1 + \gamma + \gamma^2 + \ldots + \gamma^{N-1})\mathcal{O}(\varepsilon_K^d).$$

As $N \to \infty$, we have the relation $\|\vec{q}_k - \vec{q}_{kA}\| = \frac{1}{1-\gamma}\mathcal{O}(\varepsilon_K^d)$. Using this distance relation, we can simplify the mean square error expression of the product-LALM scheme.

$$\begin{aligned} \text{MSE}_{\text{pd-LALM}} &= \sum_k \int_{\mathcal{R}_k} \|\vec{q}_{kA} - \vec{y}\|^2 f_{\text{app}}(\vec{y}) d\vec{y} \\ &= \sum_k \int_{\mathcal{R}_k} \|\vec{q}_{kA} - \vec{q}_k + \vec{q}_k - \vec{y}\|^2 f_{\text{app}}(\vec{y}) d\vec{y} \\ &= \sum_k \int_{\mathcal{R}_k} \left[\|\vec{q}_{kA} - \vec{q}_k\|^2 + \|\vec{q}_k - \vec{y}\|^2 + 2(\vec{q}_{kA} - \vec{q}_k)^T(\vec{q}_k - \vec{y})\right] f_{\text{app}}(\vec{y}) d\vec{y} \\ &= \frac{1}{1-\gamma}\mathcal{O}(\varepsilon_K^{2d}) + \text{MSE}_{\text{subopt}} + \mathcal{O}(\varepsilon_K^{3d}) \end{aligned} \tag{4}$$

The third term in the equation above captures the cross terms in the norm expansion. Using the fact that $\text{MSE}_{\text{subopt}} > \text{MSE}_{\text{kmeans}}$, we show that the $\text{MSE}_{\text{pd-LALM}}$ exceeds $\text{MSE}_{\text{kmeans}}$ by atleast a constant factor of ε_K^{2d}. □

2 Experiments

2.1 Scalar LALM

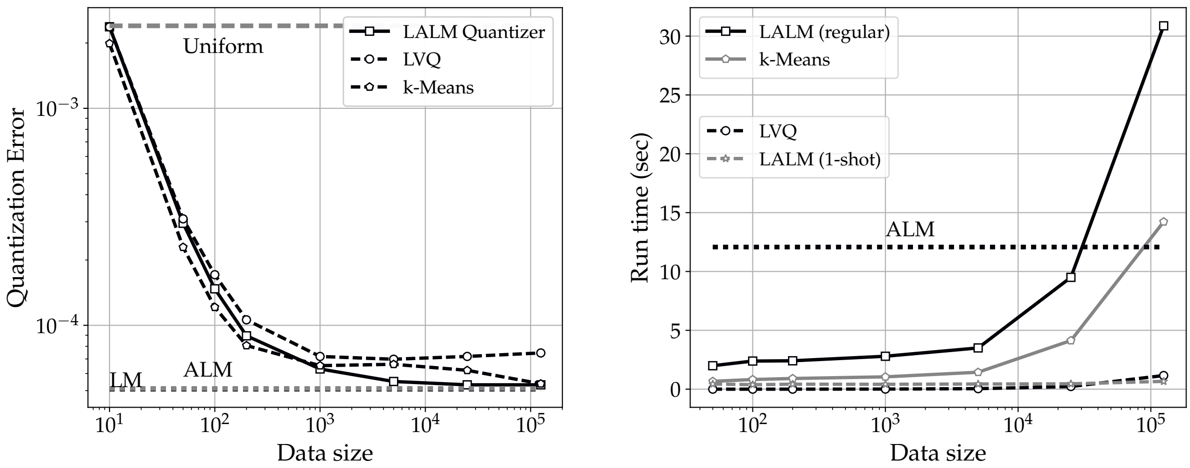

Figure 1: fig:scalar_LALM

N	pdt-LALM		equi-spaced		uniform(rand)		k-means++ init		LVQ		k-means	
33	0.0056	0.0062	0.0059	0.0061	0.0350	0.0383	-	-	0.0318	0.0367	-	-
58	0.0048	0.0047	0.0058	0.0060	0.0207	0.0216	-	-	0.0195	0.0214	-	-
100	0.0047	0.0046	0.0055	0.0062	0.0170	0.0156	-	-	0.0158	0.0150	-	-
190	0.0047	0.0047	0.0059	0.0059	0.0108	0.0105	-	-	0.0104	0.0104	-	-
330	0.0046	0.0045	0.0060	0.0063	0.0076	0.0076	-	-	0.0074	0.0074	-	-
580	0.0045	0.0044	0.0061	0.0060	0.0072	0.0071	0.0012	0.0039	0.0070	0.0070	0.0009	0.0039
1730	0.0045	0.0045	0.0061	0.0061	0.0075	0.0074	0.0028	0.0039	0.0073	0.0073	0.0018	0.0035
3000	0.0045	0.0045	0.0061	0.0061	0.0072	0.0072	0.0032	0.0037	0.0070	0.0071	0.0022	0.0033
5200	0.0045	0.0045	0.0061	0.0061	0.0073	0.0073	0.0033	0.0037	0.0071	0.0072	0.0025	0.0033
10000	0.0045	0.0045	0.0061	0.0061	0.0072	0.0072	0.0035	0.0037	0.0070	0.0070	0.0025	0.0031

Table 1: tbl:comp_bet_DIM3 Comparison of Beta(4,2) various schemes for $K = 343$ and $d = 3$. Each dimension has 7 levels

N	pdt-LALM		equi-spaced		uniform(rand)		k-means++ init		LVQ		k-means	
330	-	-	-	-	-	-	-	-	-	-	-	-
580	0.0118	0.0118	0.0182	0.0183	0.0106	0.0105	0.0026	0.0072	0.0104	0.0104	0.0018	0.0072
1730	0.0118	0.0119	0.0183	0.0183	0.0098	0.0096	0.0053	0.0069	0.0096	0.0095	0.0035	0.0058
3000	0.0119	0.0117	0.0185	0.0181	0.0097	0.0095	0.0061	0.0071	0.0095	0.0094	0.0039	0.0056
5200	0.0117	0.0117	0.0183	0.0182	0.0100	0.0099	0.0066	0.0072	0.0098	0.0097	0.0042	0.0054
10000	0.0118	0.0118	0.0185	0.0183	0.0097	0.0099	0.0066	0.0069	0.0095	0.0097	0.0043	0.0051

Table 2: tbl:comp_arcsine_DIM3 Comparison of arcsine for $K = 343$ and $d = 3$. Each dimension has 7 levels

N	pdt-LALM		equi-spaced		uniform(rand)		k-means++ init		LVQ		k-means	
330	-	-	-	-	-	-	-	-	-	-	-	-
580	0.0085	0.0083	0.0136	0.0131	0.0075	0.0075	0.0003	0.0056	0.0074	0.0074	0.0006	0.0059
1730	0.0085	0.0083	0.0136	0.0134	0.0087	0.0085	0.0034	0.0053	0.0085	0.0084	0.0022	0.0046
3000	0.0082	0.0083	0.0133	0.0133	0.0069	0.0072	0.0042	0.0052	0.0068	0.0071	0.0026	0.00441
5200	0.0084	0.0085	0.0134	0.0136	0.0073	0.0075	0.0046	0.0051	0.0072	0.0074	0.0031	0.0042
10000	0.0084	0.0084	0.0135	0.0134	0.0076	0.0075	0.0048	0.0051	0.0075	0.0074	0.0032	0.0040

Table 3: tbl:comp_arcsine_DIM3 Comparison of arcsine for $K = 512$ and $d = 3$. Each dimension has 7 levels

d	pdt-LALM		equi-spaced		uniform(rand)		k-means++ init		LVQ		k-means	
4	0.0231	0.0232	0.0323	0.0324	0.0482	0.0483	0.0235	0.0237	0.0429	0.0431	0.0172	0.0184
6	0.0348	0.0347	0.0488	0.0481	0.0603	0.0602	0.0281	0.0302	0.0589	0.0588	0.0211	0.0281
8	0.0465	0.0465	0.0648	0.0648	0.0771	0.0771	0.0118	0.0379	0.0766	0.0766	0.0089	0.0384

tbl:dim_scaling_PERLEV3

Table 4: Comparison of Beta(4,2) for $k = 3$ and $N = 10000$. Each dimension has 3 levels

d	pdt-LALM		equi-spaced		uniform(rand)		k-means++ init		LVQ		k-means	
3	0.0033	0.0033	0.0045	0.0045	0.0055	0.0055	0.0027	0.0028	0.0054	0.0054	0.0019	0.0024
9	0.1123	0.1122	0.1606	0.1607	0.2055	0.2054	0.0923	0.0976	0.1884	0.1886	0.0703	0.0871
32	0.9569	0.9562	1.5295	1.530	2.1591	2.1569	0.8913	0.9386	1.0156	1.0141	0.7983	0.8349
64	2.1473	2.1470	3.4368	3.433	5.5844	5.587	2.3062	2.435	2.0295	2.0337	1.9642	1.9779
100	3.4794	3.4791	5.5795	5.5801	9.6999	9.6958	4.0499	4.2675	3.1771	3.1762	3.1316	3.1430

tbl:dim_scaling_LEV512

Table 5: Comparison of Beta(4,2) for $K = 512$ and $N = 10000$.