Bolt: Theoretical Analysis

Davis W. Blalock, John V. Guttag

February 18 2017

Note: this document assumes you have already read the Bolt paper.

1 Quantization Error

1.1 Definitions

Let Q be the distribution of query subvectors $\mathbf{q}_m$ for lookup table m, X be the distribution of database subvectors $\mathbf{x}_m$ for this table, and Y be the scalar-valued distribution of distances within that table. I.e.:

$$p(Y = y) \triangleq \int_{Q,X} p(\mathbf{q}_m, \mathbf{x}_m) I\{d_m(\mathbf{q}_m, \mathbf{x}_m) = y\} \tag{1}$$

Recall that we seek to learn a quantization function $\beta_m : \mathbb{R} \to \{0, \ldots, 255\}$ of the form:

$$\beta_m(y) = \max(0, \min(255, \lfloor ay - b \rfloor)) \tag{2}$$

that minimizes the loss:

$$E_Y[(\hat{y} - y)^2] \tag{3}$$

where $\hat{y} \triangleq a(\beta_m(y) + b)$ is the *reconstruction* of y.

In the paper, we propose setting $b = F^{-1}(\alpha)$ and $a = 255/(F^{-1}(1-\alpha) - b)$ for some suitable α. F^{-1} is the inverse CDF of Y, estimated empirically on a training set. The value of α is optimized using a simple grid search.

To analyze the performance of β_m from a theoretical perspective, let us define the following:

- Let $|\hat{y} - y|$ be the *quantization error*.
- Let $B = 255$ be the number of quantization bins.
- Let $b_{min} \triangleq F^{-1}(\alpha)$ be the smallest value that can be quantized without clipping.
- Let $b_{max} \triangleq F^{-1}(1-\alpha)$ be the largest value that can be quantized without clipping.

- Let $\Delta \triangleq \frac{b_{max}-b_{min}}{B}$ be the width of each quantization bin.

Using these quantities, the quantization error for a given y value can be decomposed into:

$$|y-\hat{y}| = \begin{cases} b_{min}-y & \text{if } y \leq b_{min} \\ \Delta c(y) & \text{if } b_{min} < y \leq b_{max} \\ y-b_{max} & \text{if } y > b_{max} \end{cases} \tag{4}$$

where $c(y) = (y-\hat{y})/\Delta$ returns a value in $[0,1)$ indicating where $\hat{y}$ lies within its quantization bin. These three cases represent y being clipped at b_{min}, being rounded down to the nearest bin boundary, or being clipped at b_{max}, respectively.

It will also be helpful to define the following properties.

Definition 1.1. *A random variable X is (l,h)-exponential if and only if:*

$$l < E[X] < h \tag{5}$$

$$p(X < \gamma) < \frac{1}{\sigma_X} e^{-(E[X]-\gamma)/\sigma_X},\ \gamma \leq l \tag{6}$$

$$p(X > \gamma) < \frac{1}{\sigma_X} e^{-(\gamma-E[X])/\sigma_X},\ \gamma \geq h \tag{7}$$

where σ_X is the standard deviation of X.

In words, X is (l,h)-exponential if its tails are bounded by exponential distributions. For appropriate l and h, this definition includes distributions including the Laplace, Exponential, Gaussian, and all subgaussian distributions.

1.2 Guarantees

Lemma 1.1. *Let $p(Y < b_{min}) = 0$ and $p(Y > b_{max}) = 0$. Then $|\hat{y}-y| < \Delta$.*

Proof. The error $|\hat{y}-y| > \varepsilon$ can be decomposed according to (4). By assumption, the first and last terms in this decomposition, wherein Y clips, have probability 0. This leaves only:

$$|y-\hat{y}| = \Delta c(y) \tag{8}$$

where $0 \leq c(y) < 1$. For any value of $c(y)$, $|y-\hat{y}| < \Delta$. Intuitively, this means that if the distribution isn't clipped, the worst quantization error is the width of a quantization bin.

□

Theorem 1.1 (Two-tailed generalization bound)**.** *Let Y be (b_{min}, b_{max})-exponential. Then:*

$$p(|y-\hat{y}| > \varepsilon) < \frac{1}{\sigma_Y}\left(e^{-(b_{max}-E[Y])/\sigma_Y} + e^{-(E[Y]-b_{min})/\sigma_Y}\right) e^{-\varepsilon/\sigma_Y} \tag{9}$$

for all $\varepsilon > \Delta$.

Proof. Using the decomposition in (4), we have:

$$\begin{aligned} p(|y-\hat{y}|>\varepsilon) = & \, p(c(y)\Delta>\varepsilon)p(b_{min}<y\leq b_{max}) \\ & + p(b_{min}-y>\varepsilon) \\ & + p(y-b_{max}>\varepsilon) \end{aligned} \tag{10}$$

The first term corresponds to y being truncated within a bin, and the latter two correspond to y clipping. Since $0 \leq c(y) < 1$ and $p(b_{min} < y \leq b_{max}) \leq 1$, the first term can be bounded as:

$$p(c(y)\Delta>\varepsilon)p(b_{min}<y\leq b_{max}) < I\{\varepsilon<\Delta\} \tag{11}$$

where $I\{\cdot\}$ is the indicator function. The latter two terms can be bounded using the fact that Y is (b_{min}, b_{max})-exponential:

$$p(b_{min}-y>\varepsilon) = p(y<b_{min}-\varepsilon) < \frac{1}{\sigma_Y} e^{-(E[Y]-b_{min}-\varepsilon)/\sigma_Y} \tag{12}$$

$$p(y-b_{max}>\varepsilon) = p(y>b_{max}+\varepsilon) < \frac{1}{\sigma_Y} e^{-(b_{max}+\varepsilon-E[Y])/\sigma_Y} \tag{13}$$

Combining (11)-(13), we have:

$$p(|y-\hat{y}|>\varepsilon) < I\{\varepsilon<\Delta\} + \frac{1}{\sigma_Y}\left(e^{-(b_{max}+\varepsilon-E[Y])/\sigma_Y} + e^{-(E[Y]-b_{min}-\varepsilon)/\sigma_Y}\right) \tag{14}$$

When $\varepsilon \geq \Delta$, the first term is zero and we obtain (9). □

For ease of understanding, it is helpful to consider the case wherein b_{min} and b_{max} are symmetric about the mean. When this holds, the bound of Theorem 1.1 simplifies to the more concise expression of Lemma 1.2. This shows that the error probability decays exponentially with the number of standard deviations b_{min} and b_{max} are from the mean, as well as the size of ε (normalized by the standard deviation).

Lemma 1.2 (Symmetric generalization bound)**.** *Let z be any scalar such that Y is $(E[y]-z\sigma_Y, E[y]+z\sigma_Y)$-exponential. Then:*

$$p(|y-\hat{y}|>\varepsilon) < \frac{1}{\sigma_Y} 2e^{-(z+\varepsilon/\sigma_Y)} \tag{15}$$

where $\varepsilon > \Delta = 2z\sigma_Y/B$.

Proof. This follows immediately from Theorem 1.1 using $b_{min} = E[y] - z\sigma_Y$, $b_{max} = E[y] + z\sigma_Y$. □

The bound of 1.1 is effective when Y is roughly symmetric (as is the case when quantizing dot products, and sometimes the case when quantizing L_p distances), but less so when Y is heavily skewed (as is often the case when

quantizing L_p distances). In the presence of severe skewness, $E[Y]$ is close to either b_{min} or b_{max}, and so one of the two exponentials in parentheses approaches 1. Theorem 1.3 describes a tighter bound for the case of right skew and a hard lower limit of 0, since this is often the distribution observed for L_p distances. The corresponding bound for left skew and a hard upper limit is trivial so we omit it. Note that this theorem is useful only if $b_{min} \approx \Delta$, but this is commonly the case when the L_p distances are highly skewed.

Lemma 1.3 (One-tailed generalization bound)**.** *Let Y be (b_{min}, b_{max})-exponential, with $p(Y < 0) = 0$. Then:*

$$p(|y - \hat{y}| > \varepsilon) < \frac{1}{\sigma_Y} e^{-(b_{max} + \varepsilon - E[Y])/\sigma_Y} \tag{16}$$

for all $\varepsilon > \max(\Delta, b_{min})$.

Proof. Using (17) with the fact that $\varepsilon > b_{min}$, we have:

$$\begin{aligned} p(|y - \hat{y}| > \varepsilon) = p(c(y)\Delta > \varepsilon) p(b_{min} < y \leq b_{max}) \\ + p(y - b_{max} > \varepsilon) \end{aligned} \tag{17}$$

Again applying the bounds from (11) and (13), we obtain (16). □

2 Dot Product Error

In this section, we bound the error in Bolt's approximate dot products. We also introduce a useful closed-form approximation that helps to explain the high performance of product quantization-based algorithms in general.

2.1 Definitions and preliminaries

Definition 2.1 (Codebook)**.** *A (B, J)-codebook C is an ordered collection of 2^B vectors $\mathbf{c} \in \mathbb{R}^J$. Each vector is referred to as a "centroid" or "codeword." The notation $\mathbf{c}_i$ denotes the ith centroid in the codebook.*

Definition 2.2 (Codelist)**.** *A (K, B, J)-codelist $\mathcal{C}$ is an ordered collection of k $(B, J/K)$-codebooks. Because zero-padding is trivial and does not affect any relevant measure of accuracy, we assume that J is a multiple of K. The notation $\mathbf{c}_{ij}$ denotes the ith centroid in the jth codebook. A codelist can be thought of (and stored) as a rank-3 tensor whose columns are codebooks, treated as row-major 2D arrrays.*

Definition 2.3 (Subvectors of a vector)**.** *Let $\mathbf{x} \in \mathbb{R}^J$ be a vector, let $K > 0$ be an integer, and let $L = J/K$. Then $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}$ are the subvectors of $\mathbf{x}$, where $\mathbf{x}^{(k)} \in \mathbb{R}^L \triangleq x_{(k-1)L+1}, \ldots, x_L$. As with codelists, J is assumed to be a multiple of K.*

Definition 2.4 (Encoding of a vector)**.** *Let $\mathbf{x} \in \mathbb{R}^J$ be a vector with subvectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}$ and let $\mathcal{C}$ be a (K, B, J)-codelist. Then the encoding of $\mathbf{x}$ is the sequence of integers $a_1, \ldots, a_K$, $0 < a_k \leq 2^B$ where $a_k \triangleq \arg\min_i ||\mathbf{x}^{(k)} - \mathbf{c}_{ik}||^2$.*

Definition 2.5 (Reconstruction)**.** *Let $a_1, \ldots, a_K$, $0 < a_k \leq 2^B$ be the encoding of some vector $\mathbf{x}$, and let $\mathcal{C}$ be a (K, B, J)-codelist. Then the concatenation of the vectors $c_{a_1 1}, c_{a_2 2}, \ldots, c_{a_K K}$ is the reconstruction of $\mathbf{x}$, denoted $\hat{\mathbf{x}}$.*

Definition 2.6 (Residuals)**.** *Let $\hat{\mathbf{x}}$ be the reconstruction of $\mathbf{x}$. Then $\mathbf{r} \triangleq \mathbf{x} - \hat{\mathbf{x}}$ is the residual vector for $\mathbf{x}$.*

Apart from these definitions, it is also necessary to establish several geometric properties of random (encoded) vectors in high-dimensional spaces.

Lemma 2.1 (Dot product bias ([1]))**.** *Let $\hat{\mathbf{x}}$ be the reconstruction of $\mathbf{x}$ using codelist $\mathcal{C}$, and suppose that the centroids of all codebooks within $\mathcal{C}$ were learned using k-means. Then $E[\mathbf{q}^\top \mathbf{x} - \mathbf{q}^\top \hat{\mathbf{x}}] = 0$.*

Lemma 2.2 (Euclidean distance bias ([4, 1]))**.** *Let $a_1, \ldots, a_K$, $0 < a_k \leq 2^B$ be the encoding of some vector $\mathbf{x}$ using codelist $\mathcal{C}$ and $\hat{\mathbf{x}}$ be the reconstruction of $\mathbf{x}$. Further suppose that the centroids of all codebooks within $\mathcal{C}$ were learned using k-means. Then:*

$$E[\|\mathbf{q} - \mathbf{x}\|^2 - \|\mathbf{q} - \hat{\mathbf{x}}\|^2] = \sum_{k=1}^{K} \mathrm{MSE}(a_k, k) \tag{18}$$

where $\mathrm{MSE}(a_k, k)$ is the expected squared Euclidean distance between centroid $\mathbf{c}_{a_k k}$ and the subvectors assigned to it by k-means. I.e.,

$$\mathrm{MSE}(a_k, k) \triangleq E_X[\|\mathbf{c}_{a_k k} - \mathbf{x}^{(k)}\|^2],\ a_k = \arg\min_i \|\mathbf{c}_{ik} - \mathbf{x}^{(k)}\|^2 \tag{19}$$

Lemma 2.3 (Area of a hyperspherical cap (Li. 2011 [5]))**.** *Suppose that a hyperphere in $\mathbb{R}^J$ with radius r is cut into two caps by a hyperplane, with the angle $\theta, 0 \leq \theta \leq \frac{\pi}{2}$ defining the radius of the smaller cap. Then the area of the smaller cap is given by*

$$A_J(r) = \frac{1}{2} A_J^s(r) I_{\sin^2(\theta)} \left(\frac{J-1}{2}, \frac{1}{2} \right) \tag{20}$$

where $A_J^s(r)$ is the area of the hyperpsphere and $I_x(a, b)$ denotes the regularized incomplete beta function (i.e., the CDF of a $\mathrm{Beta}(a, b)$ distribution).

Lemma 2.4 (Minimum angle between random vectors)**.** *Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^J$ be vectors such that $\frac{\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \mathbf{x}$ is sampled uniformly from the surface of the unit hypersphere S^{J-1}, and let $\theta \triangleq \arccos(\frac{\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|})$ be the angle between $\mathbf{x}$ and $\mathbf{y}$. Then for $0 \leq a \leq \frac{\pi}{2}$,*

$$p(|\theta| \geq a) = I_{\sin^2(a)} \left(\frac{J-1}{2}, \frac{1}{2} \right) \tag{21}$$

Proof. Since the angle between $\mathbf{x}$ and $\mathbf{y}$ is independent of their norms, assume without loss of generality that $\mathbf{x}$ and $\mathbf{y}$ have been scaled such that $\|x\| = \|y\| = 1$. For a given $\mathbf{x}$, the set of $\mathbf{y}$ vectors such that $\cos(\theta) \geq a, \theta \leq \frac{\pi}{2}$ is exactly the set of vectors comprising a hyperspherical cap of S^{J-1} with radius defined by a. Because the projection onto $\mathbf{x}$ of $\mathbf{y}$ has probability mass uniformly distributed across S^{J-1}, the probability that $\mathbf{y}$ lies within this cap is equal to the area of the cap divided by the area of the hypersphere. Using Lemma 2.3, this ratio is given by:

$$\frac{1}{2} I_{\sin^2(a)} \left(\frac{J-1}{2}, \frac{1}{2} \right) \tag{22}$$

By symmetry, this is also $p(\theta < -a)$. Summing the probabilities of these two events yields (21).

□

Lemma 2.5 (Gaussian approximation to angle between random vectors)**.** *Let* $\mathbf{x}$, $\mathbf{q}$, *and* θ *be defined as in Lemma 2.4. Then*

$$p(cos(\theta) > a) \approx \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(-cos(\theta) \sqrt{\frac{J}{2}} \right) \tag{23}$$

Proof. Using the identity $I_z(\alpha, \alpha) = \frac{1}{2} I_{4z(1-z)}(\alpha, \frac{1}{2})$ [2, Eq. 8.17.6], we can rewrite (22) as:

$$I_\phi(\frac{J-1}{2}, \frac{J-1}{2}) \tag{24}$$

where $\phi = \frac{1}{2}(1 - \cos(\theta))$. Recall that a Beta$(\alpha, \beta)$ distribution can be approximated by a normal distribution with:

$$\mu = \frac{\alpha}{\alpha + \beta} \tag{25}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)} \tag{26}$$

Using $\alpha = \beta = \frac{J-1}{2}$, this yeilds

$$\mu = \frac{1}{2} \tag{27}$$

$$\sigma^2 = \frac{\left(\frac{J-1}{2}\right)^2}{4\left(\frac{J-1}{2}\right)^2 \left(1 + 2\frac{J-1}{2}\right)} = \frac{1}{4J} \tag{28}$$

Further recall that the CDF of a normal distribution with a given mean μ and variance σ^2 is given by

$$\Phi(a) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{a - \mu}{\sigma\sqrt{2}} \right) \tag{29}$$

Substituting (27) and (28) into (29), we obtain

$$I_\phi(\frac{J-1}{2}, \frac{J-1}{2}) \approx \text{erf}((\phi - \frac{1}{2})\sqrt{2J}) \tag{30}$$

Finally, substituting $\frac{1}{2}(1 - \cos(\theta))$ for ϕ yields (23). □

Lemma 2.6 (Gaussian PDF approximation)**.** *Let* $\mathbf{x}$*,* $\mathbf{q}$*, and* θ *be defined as in Lemma 2.4. Then*

$$cos(\theta) \sim \mathcal{N}(0, J^{-1}) \tag{31}$$

Proof. Writing (23) in the form of (29) gives

$$\mu = 0 \tag{32}$$

$$\sigma^2 = \frac{1}{J} \tag{33}$$

Because (23) is the CDF of a Gaussian random variable with this μ and σ^2, the PDF is given by $\mathcal{N}(0, J^{-1})$. □

2.2 Guarantees

We now prove several bounds on the errors caused by product quantization using an arbitrary number of subvectors. We begin with no distributional assumptions, and then prove increasingly tight bounds as more assumptions are added.

We begin with Lemma 2.7, which is not probabalistic.

Lemma 2.7 (Worst-case dot product error)**.** *Let* $\hat{\mathbf{x}}$ *be the reconstruction of* $\mathbf{x}$ *and let* $\mathbf{q} \in \mathbb{R}^J$ *be a vector. Then* $|\mathbf{q}^\top \mathbf{x} - \mathbf{q}^\top \hat{\mathbf{x}}| < \|q\| \cdot \|r\|$.

Proof. This follows immediately from application of the Cauchy-Schwarz inequality.

$$|\mathbf{q}^\top \mathbf{x} - \mathbf{q}^\top \hat{\mathbf{x}}| = |\mathbf{q}^\top \mathbf{x} - \mathbf{q}^\top (\mathbf{x} - \mathbf{r})| = |\mathbf{q}^\top \mathbf{r}| \leq \|q\| \cdot \|r\| \tag{34}$$

□

If we are willing to make the extremely pessimistic assumption that the cosine of the angle between $\mathbf{q}$ and $\mathbf{r}$ is uniformly distributed, a tighter bound (and indeed, an exact expression for the error probability) is possible (Theorem 2.2). This assumption is pessimistic because angles close to 0, which yield smaller errors, are much more probable in high dimensions.

Theorem 2.1 (Pessimistic dot product error)**.** *Let* θ *denote the angle between* $\mathbf{r}$ *and some vector* $\mathbf{q}$*, and suppose that* $\cos(\theta) \sim Unif(-1, 1)$*. Then*

$$p(|\mathbf{q}^\top \mathbf{x} - \mathbf{q}^\top \hat{\mathbf{x}}| > \varepsilon) = \max(0, 1 - \frac{\varepsilon}{\|\mathbf{q}\| \cdot \|\mathbf{r}\|}) \tag{35}$$

Proof. Simple algebra shows that

$$\mathbf{q}^\top \mathbf{x} - \mathbf{q}^\top \hat{\mathbf{x}} = \mathbf{q}^\top (\hat{\mathbf{x}} + \mathbf{r}) - \mathbf{q}^\top \hat{\mathbf{x}} = \mathbf{q}^\top \mathbf{r} = \|\mathbf{q}\| \cdot \|\mathbf{r}\| \cos(\theta) \tag{36}$$

Since $\cos(\theta) \sim Unif(-1, 1)$, we have that $|\cos(\theta)| \sim Unif(0, 1)$, and therefore

$$p(|\mathbf{q}^\top \mathbf{x} - \mathbf{q}^\top \hat{\mathbf{x}}| > \varepsilon) = p(|\|\mathbf{q}\| \cdot \|\mathbf{r}\| \cos(\theta)| > \varepsilon) \tag{37}$$

$$= p\left(|\cos(\theta)| > \frac{\varepsilon}{\|\mathbf{q}\| \cdot \|\mathbf{r}\|}\right) \tag{38}$$

$$= \max\left(0, 1 - \frac{\varepsilon}{\|\mathbf{q}\| \cdot \|\mathbf{r}\|}\right) \tag{39}$$

□

The assumption that the cosine similarity of vectors is uniform can be replaced with the slightly more optimistic assumption that the errors in quantizing each subvector are independent, yielding Theorem 2.2.

Theorem 2.2 (Dot product error with independent subspaces)**.** *Let* $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}$ *be the subvectors of* $\mathbf{x}$*, let* $\hat{\mathbf{x}}^{(1)}, \ldots, \hat{\mathbf{x}}^{(K)}$ *be the subvectors of* $\hat{\mathbf{x}}$*, and let* $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(K)}$ *be the subvectors of an arbitrary vector* $\mathbf{q} \in R^J$*. Further let* $\mathbf{r}^{(k)} \triangleq \mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)}$*, and assume that the values of* $\|\mathbf{r}^{(k)}\|$ *are independent for all* k*. Then:*

$$p(|\mathbf{q}^\top \mathbf{x} - \mathbf{q}^\top \hat{\mathbf{x}}| \geq \varepsilon) \leq 2 \exp\left(\frac{-\varepsilon^2}{4 \sum_{k=1}^{K} (\|\mathbf{q}^{(k)}\| \cdot \|\mathbf{r}^{(k)}\|))^2}\right) \tag{40}$$

Proof. The quantity $\mathbf{q}^\top \mathbf{x} - \mathbf{q}^\top \hat{\mathbf{x}}$ can be expressed as the sum

$$\sum_{k=1}^{K} \mathbf{q}^{(k)\top} (\mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)}) = \sum_{k=1}^{K} \mathbf{q}^{(k)\top} \mathbf{r}^{(k)} \tag{41}$$

Each element of this sum can be viewed as an independent random variable v_k. By Lemma 2.7, $-\|\mathbf{q}^{(k)}\| \cdot \|\mathbf{r}^{(k)}\| < v_k < \|\mathbf{q}^{(k)}\| \cdot \|\mathbf{r}^{(k)}\|$. The inequality (40) then follows from Hoeffding's inequality. □

This bound assumes the worst-case distribution of errors for each subvector. If we instead assume that the errors are random as defined in Lemma 2.4, it is possible to obtain not only a bound, but a closed-form expression for the probability of a given error.

Theorem 2.3 (Dot product error approximation)**.** *Let* $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}$*,* $\hat{\mathbf{x}}^{(1)}, \ldots, \hat{\mathbf{x}}^{(K)}$*,* $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(K)}$*,* $\mathbf{r}^{(1)}, \ldots, \mathbf{r}^{(K)}$ *be defined as in Theorem 2.2. Suppose that each* $(\mathbf{q}^{(k)}, \mathbf{r}^{(k)})$ *satisfy the conditions of Lemma 2.4 and the values of* $\mathbf{q}^{(k)\top} \mathbf{r}^{(k)}$ *are independent across all* k*.*

$$p(\mathbf{q}^\top \mathbf{x} - \mathbf{q}^\top \hat{\mathbf{x}}) \approx \mathcal{N}(0, \sigma^2) \tag{42}$$

where $\sigma^2 \triangleq J^{-1} \sum_{k=1}^{K} \|\mathbf{q}^{(k)}\| \cdot \|\mathbf{r}^{(k)}\|$.

Proof. Applying Lemma 2.6 to a given $(\mathbf{q}^{(k)}, \mathbf{r}^{(k)})$, we have the approximation:

$$cos(\theta_k) \sim \mathcal{N}(0, L^{-1}) \tag{43}$$

where $\theta_k \triangleq \frac{{\mathbf{q}^{(k)}}^\top \mathbf{r}^{(k)}}{\|\mathbf{q}^{(k)}\| \cdot \|\mathbf{r}^{(k)}\|}$. Recalling from (36) that $\mathbf{q}^\top \mathbf{x} - \mathbf{q}^\top \hat{\mathbf{x}} = \|\mathbf{q}\| \cdot \|\mathbf{r}\| \cos(\theta)$, this implies that

$${\mathbf{q}^{(k)}}^\top \mathbf{x}^{(k)} - \mathbf{q}^\top \hat{\mathbf{x}} \sim \mathcal{N}(0, \sigma_k^2) \tag{44}$$

where $\sigma_k^2 = \frac{\|\mathbf{q}^{(k)}\| \cdot \|\mathbf{r}^{(k)}\|}{L}$. Because the errors from each subspace are independent, one can sum their variances and divide by K to obtain (42). Observe that we have simplified $(KL)^{-1}$ to J^{-1}. □

This approximation is optimistic if the codebooks are trained from k-means using the Euclidean distance, since the residuals' directions are unlikely to be uniformly distributed on the unit hypersphere. However, if the centroids are trained under the Mahalanobis distance as in [3, 1], then this approximation may be pessimistic. This is because the latter approach tends to concentrate $cos(\theta)$ around 0 (by construction), which yields even smaller variances in each subspace.

It is also interesting to note that this error expression is independent of K. One might expect that having more independent estimators would be preferable, but the tendency of larger subvectors to concentrate $\cos(\theta)$ near 0 exactly cancels the resulting variance reduction, making the error depend only on the vector dimensionality J and residuals within each subspace.

3 Euclidean Distance Error

The guarantees in this section closely parallel those of the previous section, so we state them without comment.

Theorem 3.1 (Worst-case L_2 error)**.** *Let* $\hat{\mathbf{x}}$ *be the reconstruction of* $\mathbf{x}$ *and let* $\mathbf{q} \in \mathbb{R}^J$ *be a vector. Then* $|\|\mathbf{q} - \mathbf{x}\| - \|\mathbf{q} - \hat{\mathbf{x}}\|| < \|\mathbf{r}\|$.

Proof. This follows immediately from application of the triangle inequality.

$$\|\mathbf{q} - \mathbf{x}\| - \|\mathbf{r}\| \leq \|\mathbf{q} - \hat{\mathbf{x}}\| = \|\mathbf{q} - \mathbf{x} + \mathbf{r}\| \leq \|\mathbf{q} - \mathbf{x}\| + \|\mathbf{r}\| \tag{45}$$

and therefore

$$|\|\mathbf{q} - \mathbf{x}\| - \|\mathbf{q} - \hat{\mathbf{x}}\|| \leq \|\mathbf{r}\| \tag{46}$$

□

Theorem 3.2 (Pessimistic L_2 error)**.** *Let* θ *denote the angle between* $\mathbf{r}$ *and some vector* $\mathbf{q}$*, and suppose that* $\cos(\theta) \sim Unif(-1, 1)$*. Then*

$$p(|\|\mathbf{q} - \mathbf{x}\|^2 - \|\mathbf{q} - \hat{\mathbf{x}}\|^2| > \varepsilon) = \max\left(0, 1 - \frac{\|\mathbf{r}\|^2 - \varepsilon}{2\|\mathbf{r}\|\|\mathbf{q} - \mathbf{x}\|}\right) \tag{47}$$

Proof. Using the Law of Cosines, we have

$$\begin{aligned}\|\mathbf{q}-\mathbf{x}\|^2 &= \|\mathbf{q}-\hat{\mathbf{x}}\|^2 + \|\mathbf{x}-\hat{\mathbf{x}}\|^2 - 2\|\mathbf{q}-\hat{\mathbf{x}}\|\|\mathbf{x}-\hat{\mathbf{x}}\|\cos(\theta) \\ &= \|\mathbf{q}-\hat{\mathbf{x}}\|^2 + \|\mathbf{r}\|^2 - 2\|\mathbf{q}-\hat{\mathbf{x}}\|\|\mathbf{r}\|\cos(\theta) \end{aligned} \tag{48}$$

and therefore

$$\|\mathbf{q}-\mathbf{x}\|^2 - \|\mathbf{q}-\hat{\mathbf{x}}\|^2 = \|\mathbf{r}\|^2 - 2\|\mathbf{r}\|\|\mathbf{q}-\hat{\mathbf{x}}\|\cos(\theta) \tag{49}$$

This implies that

$$\begin{aligned} p(\|\mathbf{q}-\mathbf{x}\|^2 - \|\mathbf{q}-\hat{\mathbf{x}}\|^2 > \varepsilon) &= p(\|\mathbf{r}\|^2 - 2\|\mathbf{r}\|\|\mathbf{q}-\hat{\mathbf{x}}\|\cos(\theta) > \varepsilon) \\ &= p(\frac{\|\mathbf{r}\|^2 - \varepsilon}{2\|\mathbf{r}\|\|\mathbf{q}-\hat{\mathbf{x}}\|} > \cos(\theta)) \\ &= \frac{1}{2}\max\left(0, 1 - \frac{\|\mathbf{r}\|^2 - \varepsilon}{2\|\mathbf{r}\|\|\mathbf{q}-\hat{\mathbf{x}}\|}\right) \end{aligned} \tag{50}$$

Equation (47) follows by symmetry.

□

Theorem 3.3 (L_2 error with independent subspaces)**.** *Let $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}$ be the subvectors of $\mathbf{x}$, let $\hat{\mathbf{x}}^{(1)}, \ldots, \hat{\mathbf{x}}^{(K)}$ be the subvectors of $\hat{\mathbf{x}}$, and let $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(K)}$ be the subvectors of an arbitrary vector $\mathbf{q} \in R^J$. Further let $\mathbf{r}^{(k)} \triangleq \mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)}$, and assume that the values of $\|\mathbf{q}^{(k)} - \mathbf{x}^{(k)}\|^2 - \|\mathbf{q}^{(k)} - \hat{\mathbf{x}}^{(k)}\|^2$ are independent for all k.*

$$p(|\|\mathbf{q}-\mathbf{x}\|^2 - \|\mathbf{q}-\hat{\mathbf{x}}\|^2| > \varepsilon) \leq 2\exp\left(\frac{-\varepsilon^2}{4\sum_{k=1}^K \|\mathbf{r}^{(k)}\|^4}\right) \tag{51}$$

Proof. The quantity $\|\mathbf{q}-\mathbf{x}\|^2 - \|\mathbf{q}-\hat{\mathbf{x}}\|^2$ can be expressed as the sum

$$\sum_{k=1}^K \|\mathbf{q}^{(k)} - \mathbf{x}^{(k)}\|^2 - \|\mathbf{q}^{(k)} - \hat{\mathbf{x}}^{(k)}\|^2 \tag{52}$$

By assumption, each element of this sum can be viewed as an independent random variable v_k. By Lemma 3.1, $-\|\mathbf{r}^{(k)}\|^2 \leq v_k \leq \|\mathbf{r}^{(k)}\|^2$. Assuming that one adds in the bias correction described in Lemma 2.2, one can apply Hoeffding's inequality to obtain (40). □

Theorem 3.4 (L_2 error approximation)**.** *Let $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}$, $\hat{\mathbf{x}}^{(1)}, \ldots, \hat{\mathbf{x}}^{(K)}$, $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(K)}$, $\mathbf{r}^{(1)}, \ldots, \mathbf{r}^{(K)}$ be defined as in Theorem 3.3. Suppose that each pair $(\mathbf{q}^{(k)} - \hat{\mathbf{x}}^{(k)}, \mathbf{r}^{(k)})$ satisfy the conditions of Lemma 2.4 and the values of $(\mathbf{q}^{(k)} - \hat{\mathbf{x}}^{(k)})^\top \mathbf{r}^{(k)}$ are independent across all k.*

$$p(\|\mathbf{q}-\mathbf{x}\|^2 - \|\mathbf{q}-\hat{\mathbf{x}}\|^2) \approx \mathcal{N}(\|r\|^2, \sigma^2) \tag{53}$$

where $\sigma^2 \triangleq 4\|\mathbf{r}\|^2\|\mathbf{q}-\mathbf{x}\|^2 J^{-1}$.

Proof. Applying Lemma 2.6 to a given $(\mathbf{q}^{(k)}, \mathbf{r}^{(k)})$, we have the approximation:

$$cos(\theta_k) \sim \mathcal{N}(0, L^{-1}) \tag{54}$$

where $\theta_k \triangleq \frac{\mathbf{q}^{(k)\top}\mathbf{r}^{(k)}}{\|\mathbf{q}^{(k)}\|\cdot\|\mathbf{r}^{(k)}\|}$. Further recall from (49) that

$$\|\mathbf{q}^{(k)} - \mathbf{x}^{(k)}\|^2 - \|\mathbf{q}^{(k)} - \hat{\mathbf{x}}^{(k)}\|^2 = \|\mathbf{r}^{(k)}\|^2 - 2\|\mathbf{r}^{(k)}\|\|\mathbf{q}^{(k)} - \hat{\mathbf{x}}^{(k)}\|\cos(\theta_k) \tag{55}$$

Combining (55) and (54) yields

$$\|\mathbf{q}^{(k)} - \mathbf{x}^{(k)}\|^2 - \|\mathbf{q}^{(k)} - \hat{\mathbf{x}}^{(k)}\|^2 \sim \mathcal{N}(\|\mathbf{r}^{(k)}\|^2, \sigma_k^2) \tag{56}$$

where $\sigma_k^2 \triangleq 4\|\mathbf{r}^{(k)}\|^2\|\mathbf{q}^{(k)} - \hat{\mathbf{x}}^{(k)}\|^2 L^{-1}$. Because the errors from each subspace are independent, one can sum their variances and divide by K to obtain (53). □

References

[1] Babenko, A., Arandjelovic, R., and Lempitsky, V. Pairwise quantization. *arXiv preprint arXiv:1606.01550* (2016).

[2] *NIST Digital Library of Mathematical Functions.* http://dlmf.nist.gov/, Release 1.0.14 of 2016-12-21. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

[3] Guo, R., Kumar, S., Choromanski, K., and Simcha, D. Quantization based fast inner product search. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics* (2016), pp. 482–490.

[4] Jegou, H., Douze, M., and Schmid, C. Product quantization for nearest neighbor search. *IEEE transactions on pattern analysis and machine intelligence 33*, 1 (2011), 117–128.

[5] Li, S. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics 4*, 1 (2011), 66–70.