A PROOF OF MAGNITUDE REDUCTION TRANSFORMATION

Let the original query and key matrices be $Q, K \in \mathbb{R}^{B \times d}$. We introduce a pair of linear transformations:

$$Q' = QR^{-T}, \qquad K' = KR,$$

where $R \in \mathbb{R}^{d \times d}$ is an invertible transformation. This transformation preserves the attention scores:

$$Q'K'^{\top} = (QR^{-T})(KR)^{\top} = QR^{-T}R^{\top}K^{\top} = QK^{\top}.$$

Let the second-moment (or Hessian) matrices of Q and K be:

$$X = \mathbb{E}[Q^{\top}Q], \qquad Y = \mathbb{E}[K^{\top}K],$$

which capture the energy distributions of Q and K along each feature direction.

After applying the reparameterization R, the average squared magnitudes of the transformed queries and keys become:

$$\mathbb{E}[\|Q'\|^2] = \mathrm{tr}(R^{-1}XR^{-T}), \qquad \mathbb{E}[\|K'\|^2] = \mathrm{tr}(R^{\top}YR).$$

To jointly minimize their overall energy (and hence the quantization error), we minimize the product of these quantities:

$$\min_{R} \ \mathrm{tr}(R^{-1}XR^{-T}) \cdot \mathrm{tr}(R^{\top}YR). \tag{1}$$

We solve (1) analytically using singular value decomposition (SVD). Let

$$X^{1/2}Y^{1/2} = USV^{\top}$$

be the SVD of the symmetric positive semi-definite matrix product $X^{1/2}Y^{1/2}$, where U, V are orthogonal and $S = \mathrm{diag}(s_i)$ contains the singular values.

Define the optimal transformation as:

$$R = Y^{-1/2}VS^{1/2}, \qquad R^{-1} = S^{-1/2}V^{\top}Y^{1/2}.$$

such that it makes the traces equal.

$$\begin{aligned}
\mathrm{tr}(R^{-1}XR^{-T}) &= \mathrm{tr}\Big(S^{-1/2}V^{\top}Y^{1/2}XY^{1/2}VS^{-1/2}\Big) \\
&= \mathrm{tr}(S), \\
\mathrm{tr}(R^{\top}YR) &= \mathrm{tr}\Big(S^{1/2}V^{\top}Y^{-1/2}YY^{-1/2}VS^{1/2}\Big) \\
&= \mathrm{tr}(S).
\end{aligned}$$

Hence, the minimized joint energy becomes:

$$\mathrm{tr}(R^{-1}XR^{-T}) \cdot \mathrm{tr}(R^{\top}YR) = \left(\mathrm{tr}(S)\right)^2.$$

This is the minimum achievable value of (1), and the corresponding R aligns the spectral bases of X and Y optimally.