(a) Case 1. $k = Q = x + \mathcal{E}$.

Score $k^T Q = (x + \mathcal{E})^T (x + \mathcal{E})$

$$= (x^T + \mathcal{E}^T)(x + \mathcal{E})$$

$$= (x^T x + x^T \mathcal{E} + \mathcal{E}^T x + \mathcal{E}^T \mathcal{E})$$

$$= (x^T x + \mathcal{E}^T \mathcal{E}) + x^T \mathcal{E} + \mathcal{E}^T x$$

terms meaning →

$x^T x$ and $\mathcal{E}^T \mathcal{E}$ → Captures correlations between different features for both feature and positional embedding space.

$\mathcal{E}^T x$ → this matrix calculates projection of each positional embedding vector in the feature space.

$x^T \mathcal{E}$ → this matrix calculates projection of each sample in embedding space.

Case 2 → $k = Q = \text{Cat}(x, \mathcal{E})$.

$$k = [\; x \;\vdots\; \mathcal{E} \;]$$

Sort of like additional samples in features space.

Clearly $k^T Q$ now will be the matrix which calculates covariance matrix but now with 2N samples.