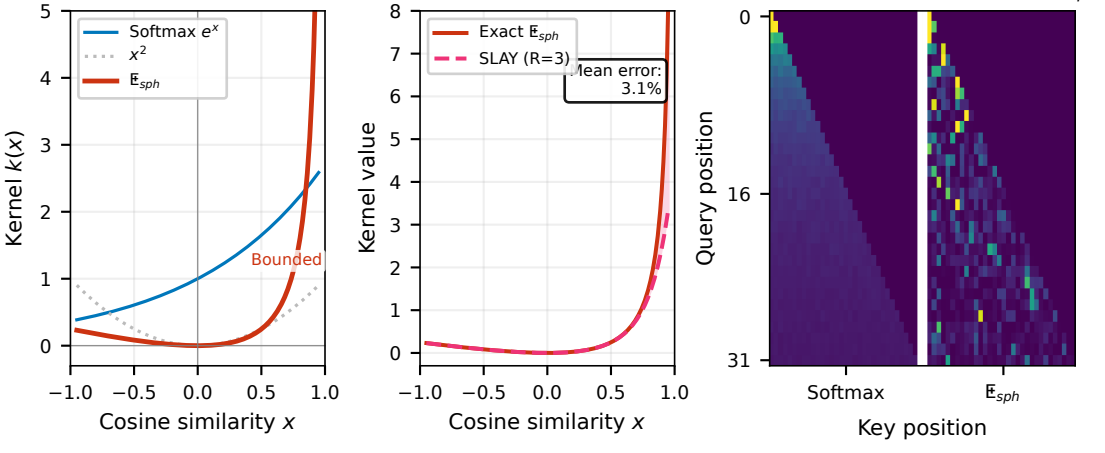
(a) Kernel functions
Softmax e^x
x^2
$\mathbb{E}_{sph}$
Bounded
Kernel $k(x)$
Cosine similarity x
(b) SLAY approximation
Exact $\mathbb{E}_{sph}$
SLAY (R=3)
Mean error: 3.1%
Kernel value
Cosine similarity x
(c) Attention: Softmax vs $\mathbb{E}_{sph}$
Query position
Softmax
$\mathbb{E}_{sph}$
Key position