

Any-SSR: How Recursive Least Squares Works in Continual Learning of Large Language Models

Supplementary Material

A. Proof of Equation (17)

Proof. According to Eq. (15):

$$\hat{W}_{k+1}^r = R_{k+1}Q_{k+1} = R_{k+1}Q_k + R_{k+1}\tilde{h}_{k+1}^\top Y_{k+1}. \quad (26)$$

Noticed that $R_{k+1}Q_{k+1}$ can be written as:

$$\begin{aligned} R_{k+1}Q_k &= R_kQ_k - R_k\tilde{h}_{k+1}^\top K\tilde{h}_{k+1}R_kQ_k \\ &= (I - R_k\tilde{h}_{k+1}^\top K\tilde{h}_{k+1})\hat{W}_k, \end{aligned}$$

Where $K = (I + \tilde{h}_{k+1}R_k\tilde{h}_k^\top)^{-1}$ and $K \in \mathbb{R}^{n \times n}$.
Since

$$KK^{-1} = K(I + \tilde{h}_{k+1}R_k\tilde{h}_{k+1}^\top) = I, \quad (27)$$

we have

$$K = I - K\tilde{h}_{k+1}R_k\tilde{h}_{k+1}^\top. \quad (28)$$

Therefore

$$R_{k+1}Q_k = (I - R_{k+1}\tilde{h}_{k+1}^\top\tilde{h}_{k+1})\hat{W}_k^R. \quad (29)$$

Instantiate Eq. (29) into Eq. (15), we have

$$\hat{W}_{k+1}^r = (I - R_{k+1}\tilde{h}_{k+1}^\top\tilde{h}_{k+1})\hat{W}_k^R + R_{k+1}\tilde{h}_{k+1}^\top Y_{k+1}. \quad (30)$$

which completes the proof.

B. Additional Model Validation

We have supplemented experiments on TRACE with GLM-4-9B and Qwen3-8B using Any-SSR. Results remain consistent and confirms ASR’s generalizability across architectures:

Model	OP (%)	BWT (%)
GLM-4-9B	54.6	-0.3
Qwen3-8B	55.1	-0.1