

The diagram illustrates the decomposition of trainable step sizes into \mathbf{v} scaling and \mathbf{x} scaling components for two models, Γ^+ and Λ^+ .

Top Model (Γ^+):

$$\Gamma^+[\mathbf{v}_k; \zeta_{\mathbf{v}}] \equiv \mathbf{v}_k \odot \exp\left(\frac{\epsilon_{\mathbf{v}}^k}{2} s_{\mathbf{v}}^k(\zeta_{\mathbf{v}_k})\right) - \frac{\epsilon_{\mathbf{v}}^k}{2} \left[\partial_x S(x_k) \odot \exp\left(\epsilon_{\mathbf{v}}^k q_{\mathbf{v}}^k(\zeta_{\mathbf{v}_k})\right) \right] + t_{\mathbf{v}}^k(\zeta_{\mathbf{v}_k})$$

Bottom Model (Λ^+):

$$\Lambda^+[\bar{\mathbf{x}}_k; \zeta_{\bar{\mathbf{x}}_k}] \equiv \bar{\mathbf{x}}_k \odot \exp\left(\epsilon_{\bar{\mathbf{x}}}^k s_{\bar{\mathbf{x}}}^k(\zeta_{\bar{\mathbf{x}}_k})\right) + \epsilon_{\bar{\mathbf{x}}}^k \left[v'_k \odot \exp\left(\epsilon_{\bar{\mathbf{x}}}^k q_{\bar{\mathbf{x}}}^k(\zeta_{\bar{\mathbf{x}}_k})\right) \right] + t_{\bar{\mathbf{x}}}^k(\zeta_{\bar{\mathbf{x}}_k})$$

Annotations:

- trainable step sizes:** Indicated by a dashed line connecting the $\epsilon_{\mathbf{v}}^k$ and $\epsilon_{\bar{\mathbf{x}}}^k$ terms.
- \mathbf{v} scaling:** Associated with the $\epsilon_{\mathbf{v}}^k$ term in the top model.
- force scaling:** Associated with the $\partial_x S(x_k)$ term in the top model.
- translation:** Associated with the $t_{\mathbf{v}}^k(\zeta_{\mathbf{v}_k})$ term in the top model.
- \mathbf{x} scaling:** Associated with the $\epsilon_{\bar{\mathbf{x}}}^k$ term in the bottom model.
- \mathbf{v} scaling:** Associated with the $\epsilon_{\bar{\mathbf{x}}}^k$ term in the bottom model.
- translation:** Associated with the $t_{\bar{\mathbf{x}}}^k(\zeta_{\bar{\mathbf{x}}_k})$ term in the bottom model.