

Iterative Refinement in the Continuous Space for Non-Autoregressive Neural Machine Translation: Appendix

Anonymous EMNLP submission

A Training objective

$$\begin{aligned} & \min_{\psi} \left\| \left(-E_{\psi}(\bar{\mathbf{z}}) + E_{\psi}(\mathbf{z}) \right) - \left(\tau(\bar{\mathbf{z}}) - \tau(\mathbf{z}) \right) \right\|^2 \\ & \approx \min_{\psi} \left\| \left((\bar{\mathbf{z}} - \mathbf{z})^{\top} \cdot \nabla_{\mathbf{z}}(-E_{\psi}(\mathbf{z})) \right) \right. \\ & \quad \left. - \left((\bar{\mathbf{z}} - \mathbf{z})^{\top} \cdot \nabla_{\mathbf{z}}\tau(\mathbf{z}) \right) \right\|^2 \end{aligned} \quad (1)$$

$$\begin{aligned} & \approx \min_{\psi} \left\| (\bar{\mathbf{z}} - \mathbf{z})^{\top} \left(\nabla_{\mathbf{z}}(-E_{\psi}(\mathbf{z})) - \nabla_{\mathbf{z}}\tau(\mathbf{z}) \right) \right\|^2 \\ & \approx \min_{\psi} \left(\nabla_{\mathbf{z}}(-E_{\psi}(\mathbf{z})) - \nabla_{\mathbf{z}}\tau(\mathbf{z}) \right)^{\top} (\bar{\mathbf{z}} - \mathbf{z}) \\ & \quad (\bar{\mathbf{z}} - \mathbf{z})^{\top} \left(\nabla_{\mathbf{z}}(-E_{\psi}(\mathbf{z})) - \nabla_{\mathbf{z}}\tau(\mathbf{z}) \right) \\ & \approx \min_{\psi} \left(\nabla_{\mathbf{z}}(-E_{\psi}(\mathbf{z})) - \nabla_{\mathbf{z}}\tau(\mathbf{z}) \right)^{\top} \|\bar{\mathbf{z}} - \mathbf{z}\|^2 \\ & \quad \left(\nabla_{\mathbf{z}}(-E_{\psi}(\mathbf{z})) - \nabla_{\mathbf{z}}\tau(\mathbf{z}) \right) \end{aligned} \quad (2)$$

$$\begin{aligned} & \approx \min_{\psi} \left\| \nabla_{\mathbf{z}}(-E_{\psi}(\mathbf{z})) - \nabla_{\mathbf{z}}\tau(\mathbf{z}) \right\|^2 \\ & \approx \min_{\psi} \left\| \nabla_{\mathbf{z}}(-E_{\psi}(\mathbf{z})) \right\|^2 + \left\| \nabla_{\mathbf{z}}\tau(\mathbf{z}) \right\|^2 \\ & \quad - 2 \left(\nabla_{\mathbf{z}}(-E_{\psi}(\mathbf{z}))^{\top} \cdot \nabla_{\mathbf{z}}\tau(\mathbf{z}) \right) \end{aligned} \quad (3)$$

$$\approx \min_{\psi} \left\| \nabla_{\mathbf{z}}E_{\psi}(\mathbf{z}) \right\|^2 + 2 \left(\nabla_{\mathbf{z}}E_{\psi}(\mathbf{z})^{\top} \cdot \nabla_{\mathbf{z}}\tau(\mathbf{z}) \right)$$

Eq. 1 follows from linear approximation, as

$$\begin{aligned} -\left(E_{\psi}(\bar{\mathbf{z}}) - E_{\psi}(\mathbf{z})\right) & \approx (\bar{\mathbf{z}} - \mathbf{z})^{\top} \cdot \nabla_{\mathbf{z}}(-E_{\psi}(\mathbf{z})) \\ \tau(\bar{\mathbf{z}}) - \tau(\mathbf{z}) & \approx (\bar{\mathbf{z}} - \mathbf{z})^{\top} \cdot \nabla_{\mathbf{z}}\tau(\mathbf{z}) \end{aligned}$$

$\|\bar{\mathbf{z}} - \mathbf{z}\|^2$ in Eq. 2 can be eliminated, as dividing the objective with a positive constant does not change the solution. The second term in Eq. 3 is also a constant with respect to ψ , hence can be ignored.