SyncBN Implementation

Forward:

$$y_i = \gamma \cdot \frac{x_i - \mu}{\sigma} + \beta$$

$$\mu = \frac{\sum x_i}{N} = \frac{S}{N}$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{\sum x_i^2}{N} - \mu^2 = \frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N}\right)^2$$

$$\sigma = \sqrt{\frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N}\right)^2} = \sqrt{\frac{Q}{N} - \frac{S^2}{N^2} + \epsilon} = \frac{\Phi^{\frac{1}{2}}}{N}$$

$$y_i = \gamma \cdot \frac{x_i - \frac{S}{N}}{\sqrt{\frac{Q}{N} - \frac{S^2}{N^2} + \epsilon}} + \beta = \frac{\gamma(Nx_i - S)}{\sqrt{NQ - S^2 + N^2\epsilon}} = \gamma \cdot \Phi^{-\frac{1}{2}}(Nx_i - S) + \beta$$

Where $\Phi = NQ - S^2 + N^2\epsilon$

Backward:

$$\frac{d_\ell}{d_{x_i}} = \frac{d_\ell}{d_{y_i}} \cdot \frac{\partial_{y_i}}{\partial_{x_i}} + \frac{d_\ell}{d_Q} \cdot \frac{d_Q}{d_{x_i}} + \frac{d_\ell}{d_S} \cdot \frac{d_S}{d_{x_i}}$$

Where $\frac{\partial_{y_i}}{\partial_{x_i}} = \frac{\gamma}{\sigma}$, $\frac{d_\ell}{d_Q} = \sum \frac{d_\ell}{d_{y_i}} \cdot \frac{d_{y_i}}{d_Q}$ and $\frac{d_\ell}{d_S} = \sum \frac{d_\ell}{d_{y_i}} \cdot \frac{d_{y_i}}{d_S}$

$$\frac{d_{y_i}}{d_Q} = \gamma \cdot (Nx_i - S) \cdot \left(-\frac{1}{2}N\right) \cdot (NQ - S^2 + N^2\epsilon)^{-1.5}$$
$$= -0.5N \cdot \gamma(Nx_i - S)(NQ - S^2 + N^2\epsilon)^{-1.5}$$
$$= -0.5N \cdot \gamma \cdot \Phi^{-1.5} \cdot (\mathrm{Nx_i} - \mathrm{S})$$
$$= -0.5N\Phi^{-1}(y_i - \beta)$$

$$\frac{d_{y_i}}{d_S} = \frac{-\gamma}{\sqrt{NQ - S^2}} + \gamma(Nx_i - S) \cdot \left(-\frac{1}{2}\right)(N \cdot Q - S^2 + N^2\epsilon)^{-1.5} \cdot (-2S)$$
$$= -\frac{\gamma}{\sqrt{NQ - S^2}} + S \cdot \gamma(Nx_i - S)(NQ - S^2 + N^2\epsilon)^{-1.5}$$
$$= -\gamma\Phi^{-0.5} + S \cdot \gamma \cdot \Phi^{-1.5} \cdot (Nx_i - S)$$
$$= -\gamma\Phi^{-0.5} + S \cdot \Phi^{-1}(y_i - \beta)$$

Forward:

$$y_i = \gamma \cdot (Es - Ex^2 + \epsilon)^{-\frac{1}{2}}(x - Ex) + \beta$$

Where $Ex = \frac{\sum x_i}{N}$, $Es = \frac{\sum x_i^2}{N}$

Backward:

$$\frac{d\ell}{d_{x_i}} = \frac{d_\ell}{d_{y_i}} \cdot \gamma \cdot (Es - Ex^2 + \epsilon)^{-\frac{1}{2}} + \frac{d\ell}{d_{Ex}} \cdot \frac{1}{N} + \frac{d_\ell}{d_{Es}} \cdot \frac{2}{N} \cdot x_i$$

$$\frac{d\ell}{d_{Ex}} = \sum \frac{d\ell}{d_{y_i}} \cdot \left(-\gamma \cdot (Es - Ex^2 + \epsilon)^{-\frac{1}{2}} - \frac{1}{2}(y_i - \beta)(Es - Ex^2 + \epsilon)^{-1} \cdot (-2Ex)\right)$$

$$= -\frac{\gamma}{\sigma} \cdot \sum \frac{d\ell}{d_{y_i}} + \frac{Ex}{\sigma^2} \sum \frac{d\ell}{d_{y_i}} (y_i - \beta)$$

$$\frac{d\ell}{d_{Es}} = \sum \frac{d\ell}{d_{y_i}} \cdot \frac{\partial_{y_i}}{\partial_{Es}} = \sum \frac{d\ell}{d_{y_i}} \cdot \left(-\frac{1}{2}\right)(y_i - \beta)(Es - Ex^2 + \epsilon)^{-1}$$

$$= -\frac{1}{2} \cdot \frac{1}{\sigma^2} \sum \frac{d\ell}{d_{y_i}} \cdot (y_i - \beta)$$

$$\frac{d_\ell}{d_\gamma} = \sum \frac{d_\ell}{d_{y_i}} \cdot \frac{y_i - \beta}{\gamma}$$

$$\frac{d_\ell}{d_\beta} = \sum \frac{d_\ell}{d_{y_i}}$$