Let N be the dimension of the vector. For one sample:

$$\boldsymbol{y}^{(\boldsymbol{n})} = \begin{bmatrix} y_1^{(n)} & y_2^{(n)} & \dots & y_{N-1}^{(n)} & y_N^{(n)} \end{bmatrix} \ (n \in \mathbb{N})$$

Let $\boldsymbol{x}$ be the original vector, and let $\boldsymbol{x} + Sublayer(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{y}^{(\boldsymbol{n})}$ be

$$\boldsymbol{z}^{(\boldsymbol{n})} = \begin{bmatrix} y_1^{(n)} + x_1 & y_2^{(n)} + x_2 & \dots & y_{N-1}^{(n)} + x_{N-1} & y_N^{(n)} + x_N \end{bmatrix}$$
$$= \begin{bmatrix} z_1^{(n)} & z_2^{(n)} & \dots & z_{N-1}^{(n)} & z_N^{(n)} \end{bmatrix}$$
$$\boldsymbol{y}^{(\boldsymbol{n+1})} = \frac{\gamma}{\sigma}\left(\boldsymbol{z}^{(\boldsymbol{n})} - \mu\right) + \beta$$

Let C be the loss function and assuming we already found

$$\frac{\partial C}{\partial \boldsymbol{y}^{(\boldsymbol{n+1})}} = \begin{bmatrix} \frac{\partial C}{\partial y_1^{(n+1)}} & \frac{\partial C}{\partial y_2^{(n+1)}} & \dots & \frac{\partial C}{\partial y_{N-1}^{(n+1)}} & \frac{\partial C}{\partial y_N^{(n+1)}} \end{bmatrix}$$

then:

$$\frac{\partial C}{\partial \gamma} = \frac{\partial C}{\partial \boldsymbol{y}^{(\boldsymbol{n+1})}} \frac{\partial \boldsymbol{y}^{(\boldsymbol{n+1})}}{\partial \gamma} = \frac{\partial C}{\partial \boldsymbol{y}^{(\boldsymbol{n+1})}} \frac{\boldsymbol{z}^{(\boldsymbol{n})} - \mu}{\sigma} = \frac{1}{N} \sum \frac{\partial C}{\partial y_k^{(n+1)}} \frac{z_k^{(n)} - \mu}{\sigma}$$

$$\frac{\partial C}{\partial \beta} = \frac{\partial C}{\partial \boldsymbol{y}^{(\boldsymbol{n+1})}} \frac{\partial \boldsymbol{y}^{(\boldsymbol{n+1})}}{\partial \beta} = \frac{\partial C}{\partial \boldsymbol{y}^{(\boldsymbol{n+1})}} = \frac{1}{N} \sum \frac{\partial C}{\partial y_k^{(n+1)}}$$

$$\frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}} = \frac{\partial C}{\partial \boldsymbol{y}^{(\boldsymbol{n+1})}} \frac{\partial \boldsymbol{y}^{(\boldsymbol{n+1})}}{\partial \boldsymbol{z}^{(\boldsymbol{n})}}$$

We have

$$\mu = \frac{\sum z_k^{(n)}}{N} \Longrightarrow \frac{\partial \mu}{\partial z_k^{(n)}} = \frac{1}{N}$$

$$V = Var\left(\boldsymbol{z}^{(\boldsymbol{n})}\right) = \frac{\sum \left(z_k^{(n)}\right)^2}{N} - \mu^2 \Longrightarrow \frac{\partial V}{\partial z_k^{(n)}} = \frac{2 z_k^{(n)}}{N} - \frac{2\mu}{N} = \frac{2}{N}\left(z_k^{(n)} - \mu\right)$$

$$V = \sigma^2 \Longrightarrow \frac{\partial V}{\partial z_k^{(n)}} = 2\sigma \frac{\partial \sigma}{\partial z_k^{(n)}} \Leftrightarrow \frac{\partial \sigma}{\partial z_k^{(n)}} = \frac{\left(z_k^{(n)} - \mu\right)}{N\sigma}$$

Combining the above, for 2 natural numbers k, l we have two cases:

Case 1: $k \neq l$, then we have:

$$\frac{\partial y_k^{(n+1)}}{\partial z_l^{(n)}} = \gamma \frac{-\frac{1}{N}\sigma - \frac{\left(z_l^{(n)} - \mu\right)}{N\sigma}\left(z_k^{(n)} - \mu\right)}{\sigma^2} = \frac{-\gamma}{N\sigma}\left(1 + \frac{\left(z_l^{(n)} - \mu\right)\left(z_k^{(n)} - \mu\right)}{\sigma^2}\right)$$

Case 2: $k = l$, then we have:

$$\frac{\partial y_k^{(n+1)}}{\partial z_k^{(n)}} = \gamma \frac{\left(1 - \frac{1}{N}\right)\sigma - \frac{\left(z_k^{(n)} - \mu\right)^2}{N\sigma}}{\sigma^2} = \frac{\gamma}{N\sigma}\left(N - 1 - \frac{\left(z_k^{(n)} - \mu\right)^2}{\sigma^2}\right)$$

So finally, we have:

$$\frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}} = \begin{bmatrix} \frac{\partial C}{\partial z_1^{(n)}} & \frac{\partial C}{\partial z_2^{(n)}} & \cdots & \frac{\partial C}{\partial z_{N-1}^{(n)}} & \frac{\partial C}{\partial z_N^{(n)}} \end{bmatrix} = \frac{\gamma}{N\sigma}\left(\frac{\partial C}{\partial \boldsymbol{y}^{(\boldsymbol{n}+\mathbf{1})}} \cdot \boldsymbol{\mathcal{J}}\right)$$

where "$\cdot$" is the dot product, $\boldsymbol{\mathcal{J}}$ is the Jacobian matrix with value at the i-th row and j-th column be:

$$\boldsymbol{\mathcal{J}}_{i,j} = \begin{cases} -1 - \frac{\left(z_i^{(n)} - \mu\right)\left(z_j^{(n)} - \mu\right)}{\sigma^2} & if\ i \neq j \\ N - 1 - \frac{\left(z_i^{(n)} - \mu\right)^2}{\sigma^2} & if\ i = j \end{cases}$$

Note that if we store the values

$$w_i^{(n)} = \frac{z_i^{(n)} - \mu}{\sigma} \rightarrow \boldsymbol{w}^{(\boldsymbol{n})}$$

then the expressions will come out cleanly as:

$$\frac{\partial C}{\partial \gamma} = \frac{1}{N}\sum \frac{\partial C}{\partial y_k^{(n+1)}} w_i^{(n)}$$

$$\frac{\partial C}{\partial \beta} = \frac{1}{N}\sum \frac{\partial C}{\partial y_k^{(n+1)}}$$

$$\boldsymbol{\mathcal{J}} = \left(NI_N - 1_N - \boldsymbol{w}^{(\boldsymbol{n})} \otimes \boldsymbol{w}^{(\boldsymbol{n})}\right)$$

$$\frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}} = \frac{\gamma}{N\sigma}\left(\frac{\partial C}{\partial \boldsymbol{y}^{(\boldsymbol{n}+\mathbf{1})}} \cdot \boldsymbol{\mathcal{J}}\right)$$

Where I_N is the identity matrix, 1_N is the $N \times N$ matrix filled with 1s and $\otimes$ is the outer product.

Note that in reality however, this Jacobian matrix is simple enough to manually calculate so we don't need the third formula.

Now, having calculated $\boldsymbol{z}^{(\boldsymbol{n})} = \boldsymbol{x} + \boldsymbol{y}^{(\boldsymbol{n})}$, we want to calculate the gradient of $\boldsymbol{x}$ and $\boldsymbol{y}^{(\boldsymbol{n})}$ separately. Notice that:

$$\frac{\partial C}{\partial \boldsymbol{y}^{(\boldsymbol{n})}} = \frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}} \frac{\partial \boldsymbol{z}^{(\boldsymbol{n})}}{\partial \boldsymbol{y}^{(\boldsymbol{n})}} = \frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}} \left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}^{(\boldsymbol{n})}} + I_N \right) = \frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}}$$

since $\boldsymbol{x}$ is independent with respect to $\boldsymbol{y}^{(\boldsymbol{n})}$. More precisely, $\boldsymbol{y}^{(\boldsymbol{n})}$ is derived from $\boldsymbol{x}$ and therefore is a function of $\boldsymbol{x}$. However, this is only a one-way relation since there were no steps in which we update the values of $\boldsymbol{x}$ based on $\boldsymbol{y}^{(\boldsymbol{n})}$, so any changes on $\boldsymbol{y}^{(\boldsymbol{n})}$ does not affect $\boldsymbol{x}$ whatsoever, which makes $\boldsymbol{x}$ essentially independent from $\boldsymbol{y}^{(\boldsymbol{n})}$.

Next, we have:

$$\frac{\partial C}{\partial \boldsymbol{x}} = \frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}} \frac{\partial \boldsymbol{z}^{(\boldsymbol{n})}}{\partial \boldsymbol{x}} = \frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}} \left(I_N + \frac{\partial \boldsymbol{y}^{(\boldsymbol{n})}}{\partial \boldsymbol{x}} \right) = \frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}} + \frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}} \frac{\partial \boldsymbol{y}^{(\boldsymbol{n})}}{\partial \boldsymbol{x}}$$

Intuitively, since $\boldsymbol{z}^{(\boldsymbol{n})} = \boldsymbol{x} + \boldsymbol{y}^{(\boldsymbol{n})}$, any change on x affect z by two ways:

1. Directly onto $\boldsymbol{z}^{(\boldsymbol{n})}$ (any change on $\boldsymbol{x}$ yields the same effect on $\boldsymbol{z}^{(\boldsymbol{n})}$).

2. The change propagates through the layers, in which $\boldsymbol{y}^{(\boldsymbol{n})}$ acts as an intermediate to change $\boldsymbol{z}^{(\boldsymbol{n})}$, so any change on $\boldsymbol{x}$ affects $\boldsymbol{y}^{(\boldsymbol{n})}$, which then yields the same effect on $\boldsymbol{z}^{(\boldsymbol{n})}$.

So having calculated $\frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}}$, we can first add it into the gradient of $\boldsymbol{x}$, then we input it into the previous layer to calculate $\frac{\partial C}{\partial \boldsymbol{z}^{(\boldsymbol{n})}} \frac{\partial \boldsymbol{y}^{(\boldsymbol{n})}}{\partial \boldsymbol{x}}$ using the conventional back propagation procedure.