1 Natural gradient - alternate derivation

$$
\begin{aligned}
de &= \sum_{i,j} \frac{\partial e}{\partial \mathrm{w}_{ij}} d\mathrm{w}_{ij} \\
&= \sum_{i,j} \varphi_i \mathrm{x}_j d\mathrm{w}_{ij} + \sum_{i,j} \left(\underline{\mathbf{w}}^{-1}\right)_{ji} d\mathrm{w}_{ij} \\
&= \underline{\varphi}^T d\underline{\mathbf{w}}\underline{\mathbf{x}} + \mathrm{Tr}\left(d\underline{\mathbf{w}} \cdot \underline{\mathbf{w}}^{-1}\right) \\
&= \underline{\varphi}^T \underbrace{\left(d\underline{\mathbf{w}} \cdot \underline{\mathbf{w}}^{-1}\right) \overbrace{\widehat{\underline{\mathbf{s}}}}^{\widehat{\underline{\mathbf{s}}}=\underline{\mathbf{w}}\underline{\mathbf{x}}} + \mathrm{Tr}\left(d\underline{\mathbf{w}} \cdot \underline{\mathbf{w}}^{-1}\right)}_{d\underline{\mathbf{Z}}}
\end{aligned}
\tag{1}
$$

with

$$d\underline{\mathbf{Z}} = d\underline{\mathbf{w}} \cdot \underline{\mathbf{w}}^{-1} \tag{2}$$

we obtain

$$de = \underline{\varphi}^T \widehat{\underline{\mathbf{s}}} d\underline{\mathbf{Z}} + \mathrm{Tr}(d\underline{\mathbf{Z}}) \tag{3}$$

gradient ascent learning

$$
\begin{aligned}
\Delta \mathrm{Z}_{ij} &= \eta \frac{\partial e}{\partial \mathrm{z}_{ij}} \\
&= \eta\left(\varphi_i \widehat{s}_j + \delta_{ij}\right) \\
&= \eta\left(\varphi_i \sum_k \mathrm{w}_{jk} \mathrm{x}_k + \delta_{ij}\right) \\
&= \sum_k \Delta \mathrm{w}_{jk} \left(\underline{\mathbf{w}}^{-1}\right)_{kj}
\end{aligned}
\tag{4}
$$

$$\Delta \mathrm{w}_{il} = \eta\left(\mathrm{w}_{il} + \varphi_i \sum_{k,j} \mathrm{w}_{jl} \mathrm{w}_{jk} \mathrm{x}_k\right) \tag{5}$$