loss
func-
tion
$\mathbf{W}$

$$\mathcal{E}(\mathbf{W}+\Delta\mathbf{W}) = \mathcal{E}(\mathbf{W})+\mathbf{g}\mathcal{E}^T\Delta\mathbf{W}+\frac{1}{2}\Delta\mathbf{W}^T\mathbf{H}\Delta\mathbf{W},$$
(1)

$\mathbf{H}$
Hes-
sian
$\mathbf{g}$
??
$$\Delta\mathbf{W}^* = \mathbf{H}^{-1}\mathbf{g},$$
(2)

Con-
ju-
gate
Gra-
di-
ent
$$\Delta\mathbf{W} = \alpha\mathbf{d},$$
(3)

α
$\mathbf{d}$
$\mathbf{W}\in$
$R^n\mathcal{E}(\mathbf{W}),$
$\mathcal{E}$
$\mathbf{W}$
??
$\{W_k\}$
k

$$\mathbf{W}_{k+1} = \mathbf{W}_k+\alpha\mathbf{d}_k, k = 0, 1, ...,$$
(4)

α_k
$\mathbf{d}_k$
??
$\mathbf{d}_k$
$$\mathbf{d}_k^T\mathbf{H}\mathbf{d}_{tk-1} = 0,$$
(5)

$\mathbf{d}_{k-1}$
$\mathbf{g}$
$\mathbf{d}_{k-1}$
$$\mathbf{d}_k = \{-\mathbf{g}_0, if$$

;
$-$
$\mathbf{g}_k+$
$\beta_k\mathbf{d}_{k-1}, otherwise,$
(6)

β_k
$\mathbf{d}_k$
$\mathbf{d}_{k-1}$
??
$k =$
0
$\mathbf{W}_0$
$k >$
1
??

$$\mathbf{d}_k^+ = \{-\mathbf{g}_0, if$$

;
$-$
$(1+$
$\beta_k\frac{\mathbf{g}_k^T\mathbf{d}_k}{\|\mathbf{g}_k\|})\mathbf{g}_k+$
$\beta_k\mathbf{d}_{k-1}^+, otherwise.$
(7)

β
MHS^+
??
$$\beta^+ = max\{\beta, 0\}.$$
(8)

β
β
??
d_k

$$\beta_k^{PR} = \frac{\mathbf{g}_k^T(\mathbf{g}_k - \mathbf{g}_{k-1})}{\|\mathbf{g}_{k-1}\|^2}, \beta_k^{FR} = \frac{\|\mathbf{g}_k\|^2}{\|\mathbf{g}_{k-1}\|^2}, \beta_k^{HS} = \frac{\mathbf{g}_k^T(\mathbf{g}_k - \mathbf{g}_{k-1})}{(\mathbf{g}_k - \mathbf{g}_{k-1}^T\mathbf{d}_{k-1})}.$$
(9)