Consider parametrized families $\lambda(k,\sigma,\omega)$ and $Q(i,u;\theta,\omega)$. Consider

$$\begin{aligned}
\theta_{n+1} \;=\; & \theta_n \;-\; a(n)\Big(\nabla_\theta Q(X_{n+1},v_n;\theta_n,\omega_n) \;-\; \nabla_\theta f(Q(\theta,\omega_n))\Big|_{\theta=\theta_n} \\
& -\; \nabla_\theta Q(X_n,U_n;\theta_n,\omega_n)\;\Big)\times \\
& \overline{\Big((1-U_n)(r(X_n,0)+\lambda(X_n,\sigma_n,\omega_n))+U_n r_n(X_n,1)+\max_{v\in\{0,1\}} Q(X_{n+1},v;\theta_n,\omega_n)} \\
& \overline{-\; f(Q(\theta_n,\omega_n))-Q(X_n,U_n;\theta_n,\omega_n)\Big)}+a(n)\xi_{n+1}, \\
\sigma_{n+1} = \sigma_n - b(n) & \overline{\Big(Q(X_n,1;\theta_n,\omega_n)-r(X_n,0)+\; f(Q(\theta_n,\omega_n))} \\
& \overline{-\max_{v\in\{0,1\}} Q(X_{n+1},v;\theta_n,\omega_n)-\lambda(X_n,\sigma_n,\omega_n)\Big)}\times \\
& \Big(-\nabla_\sigma\lambda(X_n,\sigma_n,\omega_n)\Big), \\
\omega_{n+1} \;=\; & \omega_n - b(n)\Big(\nabla_\omega Q(X_{n+1},v_n;\theta_n,\omega_n) \;-\; \nabla_\omega f(Q(\theta_n,\omega))\Big|_{\omega=\omega_n} \\
& -\; \nabla_\omega Q(X_n,U_n;\theta_n,\omega_n)\;\Big)\times \\
& \overline{\Big((1-U_n)(r(X_n,0)+\lambda(X_n,\sigma_n,\omega_n))+U_n r_n(X_n,1)+\max_{v\in\{0,1\}} Q(X_{n+1},v;\theta_n,\omega_n)} \\
& \overline{-\; f(Q(\theta_n,\omega_n))-Q(X_n,U_n;\theta_n,\sigma_n,\omega_n)\Big)}+a(n)\xi_{n+1} \\
& -\; b(n)\overline{\Big(Q(X_n,1;\theta_n,\omega_n)-r(X_n,0)+\; f(Q(\theta_n,\omega_n))} \\
& \overline{-\max_{v\in\{0,1\}} Q(X_{n+1},v;\theta_n,\omega_n)-\lambda(X_n,\sigma_n,\omega_n)\Big)}\times \\
& \Big(\nabla_\omega Q(X_{n+1},v_n;\theta_n,\omega_n) \;-\; \nabla_\omega f(Q(\theta_n,\omega))\Big|_{\omega=\omega_n} \\
& -\; \nabla_\omega Q(X_n,U_n;\theta_n,\omega_n)-\nabla_\omega\lambda(X_n,\sigma_n,\omega_n)\Big),
\end{aligned}$$

The θ_n iteration is the SGD for the mean square error

$$\begin{aligned}
\mathcal{E}_1 := E\Big[\Big\|&(1-U_n)(r(X_n,0)+\lambda(X_n,\sigma_n,\omega_n))+U_n r_n(X_n,1) \\
&+\max_{v\in\{0,1\}} Q(X_{n+1},v;\theta_n,\omega_n)-\; f(Q(\theta_n,\omega_n))-Q(X_n,U_n;\theta_n,\omega_n)\Big\|^2\Big].
\end{aligned}$$

The σ_n iteration is the SGD to minimize the mean square error

$$\begin{aligned}
\mathcal{E}_2 := E\Big[\Big\|&Q(X_n,1;\theta_n,\omega_n)-r(X_n,0)+\; f(Q(\theta_n,\omega_n)) \\
&-\max_{v\in\{0,1\}} Q(X_{n+1},v;\theta_n,\omega_n)-\lambda(X_n,\sigma_n,\omega_n)\Big\|^2\Big].
\end{aligned}$$