Appendix

A.1 Proof of Theorem 1

Theorem 1 (Backward consistency of U-SGD with sample average)**.** *If the feature representation is tabular, the vectors* $\mathbf{u}$ *and* $\boldsymbol{\theta}$ *are initially set to zero, and* $0 \le \eta < 1$*, then U-SGD defined by (5)-(7) degenerates to the recency-weighted average estimator defined by (3) and (4), in the sense that each component of the parameter vector* $\boldsymbol{\theta}_{t+1}$ *of U-SGD becomes the recency-weighted average estimator of the corresponding input.*

Proof. Consider that t samples have been observed and among them t_x samples correspond to input x. Hence, $\sum_{x\in\mathcal{X}} t_x = t$. Let $Y_{x,k}$ denote the kth output corresponding to input x. Then the recency-weighted average estimator of $v(x)$ given overall data up to t can be equivalently redefined in the following way:

$$
\begin{aligned}
\tilde{V}_{t_x+1} &\doteq \frac{\sum_{k=1}^{t_x}(1-\eta)^{t_x-k}Y_{x,k}}{\sum_{k=1}^{t_x}(1-\eta)^{t_x-k}} \\
&= \tilde{V}_{t_x} + \frac{1}{\tilde{U}_{t_x+1}}\left(Y_{x,t_x} - \tilde{V}_{t_x}\right); \quad \tilde{V}_1 = 0, \\
\tilde{U}_{t_x+1} &\doteq (1-\eta)\tilde{U}_{t_x} + 1; \qquad\qquad \tilde{U}_1 = 0.
\end{aligned}
$$

Consider that the ith feature corresponds to input x. Then it is equivalent to prove that $[\boldsymbol{\theta}_{t+1}]_i = \tilde{V}_{t_x+1}$, where $[\cdot]_i$ denotes the ith component of a vector.

We prove by induction. First we show that $[\mathbf{u}_{t+1}]_i = \tilde{U}_{t_x+1}$. By assumption, $[\mathbf{u}_1]_i = \tilde{U}_1 = 0$. Now, consider that $[\mathbf{u}_t]_i = \tilde{U}_{(t-1)_x+1}$. Then the ith component of $\mathbf{u}_{t+1}$ can be written as

$$[\mathbf{u}_{t+1}]_i = (1-\eta[\boldsymbol{\phi}_t]_i^2)[\mathbf{u}_t]_i + [\boldsymbol{\phi}_t]_i^2.$$

If the tth input is not x, then $t_x = (t-1)_x$ and $[\boldsymbol{\phi}_t]_i = 0$. Hence

$$[\mathbf{u}_{t+1}]_i = (1-0)\tilde{U}_{(t-1)_x+1} + 0 = \tilde{U}_{(t-1)_x+1} = \tilde{U}_{t_x+1}.$$

On the other hand, if the tth input is x, then $t_x = (t-1)_x + 1$ and $[\boldsymbol{\phi}_t]_i = 1$. Hence,

$$[\mathbf{u}_{t+1}]_i = (1-\eta)\tilde{U}_{(t-1)_x+1} + 1 = (1-\eta)\tilde{U}_{t_x} + 1 = \tilde{U}_{t_x+1}.$$

Hence, $[\boldsymbol{\alpha}_{t+1}]_i = \frac{1}{\tilde{U}_{t_x+1}}$, if $t_x > 0$, or $[\boldsymbol{\alpha}_{t+1}]_i = 0$, otherwise.

Now, by assumption, $[\boldsymbol{\theta}_1]_i = \tilde{V}_1 = 0$. Consider $[\boldsymbol{\theta}_t]_i = \tilde{V}_{(t-1)_x+1}$ and $t_x > 0$. Then the ith component of $\boldsymbol{\theta}_{t+1}$ can be written as

$$
\begin{aligned}
[\boldsymbol{\theta}_{t+1}]_i &= [\boldsymbol{\theta}_t]_i + [\boldsymbol{\alpha}_{t+1}]_i\left(Y_t - \boldsymbol{\theta}_t^\top\boldsymbol{\phi}_t\right)[\boldsymbol{\phi}_t]_i \\
&= \tilde{V}_{(t-1)_x+1} + \frac{1}{\tilde{U}_{t_x+1}}\left(Y_t - \boldsymbol{\theta}_t^\top\boldsymbol{\phi}_t\right)[\boldsymbol{\phi}_t]_i.
\end{aligned}
$$

If the tth input is not x, then $[\boldsymbol{\theta}_{t+1}]_i = \tilde{V}_{(t-1)_x+1} + 0 = \tilde{V}_{t_x+1}$.

On the other hand, if the tth input is x, then $Y_t = Y_{x,t_x}$ and

$$
\begin{aligned}
[\boldsymbol{\theta}_{t+1}]_i &= \tilde{V}_{(t-1)_x+1} + \frac{1}{\tilde{U}_{t_x+1}}\left(Y_{x,t_x} - \tilde{V}_{(t-1)_x+1}\right) \\
&= \tilde{V}_{t_x} + \frac{1}{\tilde{U}_{t_x+1}}\left(Y_{x,t_x} - \tilde{V}_{t_x}\right) = \tilde{V}_{t_x+1}.
\end{aligned}
$$

The only case that is left is when $t_x = 0$. In this case, the tth input cannot be x, and $\tilde{V}_{t_x+1} = \tilde{V}_{(t-1)_x+1} = \cdots = \tilde{V}_1 = 0$. Then

$$
\begin{aligned}
[\boldsymbol{\theta}_{t+1}]_i &= [\boldsymbol{\theta}_t]_i + [\boldsymbol{\alpha}_{t+1}]_i\left(Y_t - \boldsymbol{\theta}_t^\top\boldsymbol{\phi}_t\right)[\boldsymbol{\phi}_t]_i \\
&= \tilde{V}_{(t-1)_x+1} + 0\cdot\left(Y_t - \boldsymbol{\theta}_t^\top\boldsymbol{\phi}_t\right)\cdot 0 \\
&= 0 = \tilde{V}_{t_x+1}. \quad \square
\end{aligned}
$$

A.2 Proof of Theorem 2

Theorem 2 (Backward consistency of WIS-SGD-1 with WIS)**.** *If the feature representation is tabular, the vectors* $\mathbf{u}$ *and* $\boldsymbol{\theta}$ *are initially set to zero, and* $0 \leq \eta < 1$*, then WIS-SGD-1 defined by (10)-(12) degenerates to recency-weighted WIS defined by (8) and (9) with* $Y_k \doteq G_k^{t+1}$ *and* $W_k \doteq \rho_k^{t+1}$*, in the sense that each component of the parameter vector of WIS-SGD-1* $\boldsymbol{\theta}_{t+1}^{t+1}$ *becomes the recency-weighted WIS estimator of the corresponding input.*

Proof. The proof is similar to that of Theorem 1.

Consider that data is available up to time $t+1$, among which state s was visited on t_s steps. Let $G_{s,k}^{t+1}$ denote the kth flat truncated return originated from state s and $\rho_{s,k}^{t+1}$ its corresponding importance-sampling ratio. Then the recency-weighted WIS estimator of $v(s)$ given overall data up to $t+1$ can be equivalently redefined in the following way:

$$\begin{aligned}
\bar{V}_{t_s+1}^{t+1} &\doteq \bar{V}_{t_s}^{t+1} + \frac{\rho_{s,t_s}^{t+1}}{\bar{U}_{t_s+1}^{t+1}}\left(G_{s,t_s}^{t+1} - \bar{V}_{t_s}^{t+1}\right); \quad && \bar{V}_0^{t+1} = 0, \\
\bar{U}_{t_s+1}^{t+1} &\doteq (1-\eta)\bar{U}_{t_s}^{t+1} + \rho_{s,t_s}^{t+1}; && \bar{U}_0^{t+1} = 0.
\end{aligned}$$

Consider that the ith feature corresponds to input s. Then it is equivalent to prove that $\left[\boldsymbol{\theta}_{t+1}^{t+1}\right]_i = \bar{V}_{t_s+1}^{t+1}$, where $[\cdot]_i$ denotes the ith component of a vector. By abuse of notation, we drop all the $t+1$ from superscripts, as it is redundant in this proof.

We prove by induction. First we show that $[\mathbf{u}_{t+1}]_i = \bar{U}_{t_s+1}$. By assumption, $[\mathbf{u}_0]_i = \bar{U}_0 = 0$. Considering $[\mathbf{u}_t]_i = \bar{U}_{(t-1)_s+1}$. Then the ith component of $\mathbf{u}_{t+1}$ can be written as

$$[\mathbf{u}_{t+1}]_i = (1 - \eta[\boldsymbol{\phi}_t]_i^2)[\mathbf{u}_t]_i + \rho_t[\boldsymbol{\phi}_t]_i^2.$$

If the state at time t is not s, then $t_s = (t-1)_s$ and $[\boldsymbol{\phi}_t]_i = 0$. Hence

$$[\mathbf{u}_{t+1}]_i = (1-0)\bar{U}_{(t-1)_s+1} + 0 = \bar{U}_{(t-1)_s+1} = \bar{U}_{t_s+1}.$$

On the other hand, if the state at time t is s, then $t_s = (t-1)_s + 1$, $[\boldsymbol{\phi}_t]_i = 1$ and $\rho_t = \rho_{s,t_s}^{t+1}$. Hence,

$$\begin{aligned}
[\mathbf{u}_{t+1}]_i &= (1-\eta)\bar{U}_{(t-1)_s+1} + \rho_{s,t_s}^{t+1} \\
&= (1-\eta)\bar{U}_{t_s} + \rho_{s,t_s}^{t+1} = \bar{U}_{t_s+1}.
\end{aligned}$$

Hence, $[\boldsymbol{\alpha}_{t+1}]_i = \frac{1}{\bar{U}_{t_s+1}}$, if $t_s > 0$, or $[\boldsymbol{\alpha}_{t+1}]_i = 0$, otherwise.

Now, by assumption, $[\boldsymbol{\theta}_0]_i = \bar{V}_0 = 0$. Considering $[\boldsymbol{\theta}_t]_i = \bar{V}_{(t-1)_s+1}$ and $t_s > 0$, the ith component of $\boldsymbol{\theta}_{t+1}$ can be written as

$$\begin{aligned}
[\boldsymbol{\theta}_{t+1}]_i &= [\boldsymbol{\theta}_t]_i + [\boldsymbol{\alpha}_{t+1}]_i \rho_t \left(G_t - \boldsymbol{\phi}_t^\top \boldsymbol{\theta}_t\right)[\boldsymbol{\phi}_t]_i \\
&= \bar{V}_{(t-1)_s+1} + \frac{\rho_t}{\bar{U}_{t_s+1}}\left(G_t - \boldsymbol{\phi}_t^\top \boldsymbol{\theta}_t\right)[\boldsymbol{\phi}_t]_i.
\end{aligned}$$

If the state at time t is not s, then $[\boldsymbol{\theta}_{t+1}]_i = \bar{V}_{(t-1)_s+1} + 0 = \bar{V}_{t_s+1}$.

If the state at time t is not s, then $\rho_t = \rho_{s,t_s}$, $G_t = G_{s,t_s}$ and

$$\begin{aligned}
[\boldsymbol{\theta}_{t+1}]_i &= \bar{V}_{(t-1)_s+1} + \frac{\rho_{s,t_s}}{\bar{U}_{t_s+1}}\left(G_{s,t_s} - \bar{V}_{(t-1)_s+1}\right) \\
&= \bar{V}_{t_s} + \frac{\rho_{s,t_s}}{\bar{U}_{t_s+1}}\left(G_{s,t_s} - \bar{V}_{t_s}\right) = \bar{V}_{t_s+1}.
\end{aligned}$$

The only case that is left is when $t_s = 0$. In this case, the the state at time t cannot be s, and $\bar{V}_{t_s+1} = \bar{V}_{(t-1)_s+1} = \cdots = \bar{V}_0 = 0$. Then

$$\begin{aligned}
[\boldsymbol{\theta}_{t+1}]_i &= [\boldsymbol{\theta}_t]_i + [\boldsymbol{\alpha}_{t+1}]_i \rho_t \left(G_t - \boldsymbol{\theta}_t^\top \boldsymbol{\phi}_t\right)[\boldsymbol{\phi}_t]_i \\
&= \bar{V}_{(t-1)_s+1} + 0 \cdot \rho_t \left(G_t - \boldsymbol{\theta}_t^\top \boldsymbol{\phi}_t\right) \cdot 0 \\
&= 0 = \bar{V}_{t_s+1}. \quad \square
\end{aligned}$$

A.3 Proof of Theorem 3

Theorem 3 (Online equivalence technique)**.** *Consider any forward view that updates toward an interim scalar target Y_k^t with*

$$\boldsymbol{\theta}_{k+1}^{t+1} \doteq \mathbf{F}_k \boldsymbol{\theta}_k^{t+1} + Y_k^{t+1}\mathbf{w}_k + \mathbf{x}_k, \quad 0 \le k < t+1,$$

where $\boldsymbol{\theta}_0^t \doteq \boldsymbol{\theta}_0$ for some initial $\boldsymbol{\theta}_0$, and both $\mathbf{F}_k \in \mathbb{R}^{n\times n}$ and $\mathbf{w}_k \in \mathbb{R}^n$ can be computed using data available at k. Assume that the temporal difference $Y_k^{t+1} - Y_k^t$ at k is related to the temporal difference at $k+1$ as follows:

$$Y_k^{t+1} - Y_k^t = d_{k+1}\left(Y_{k+1}^{t+1} - Y_{k+1}^t\right) + b_t g_k \prod_{j=k+1}^{t-1} c_j, 0 \le k < t,$$

where b_k, c_k, d_k and g_k are scalars that can be computed using data available at time k. Then the final weight $\boldsymbol{\theta}_{t+1} \doteq \boldsymbol{\theta}_{t+1}^{t+1}$ can be computed through the following backward-view update, with $\mathbf{e}_{-1} \doteq \mathbf{0}$, $\mathbf{d}_0 \doteq \mathbf{0}$, and $t \ge 0$:

$$\begin{aligned}
\mathbf{e}_t &\doteq \mathbf{w}_t + d_t \mathbf{F}_t \mathbf{e}_{t-1}, \\
\boldsymbol{\theta}_{t+1} &\doteq \mathbf{F}_t \boldsymbol{\theta}_t + (Y_t^{t+1} - Y_t^t)\mathbf{e}_t + Y_t^t \mathbf{w}_t + b_t \mathbf{F}_t \mathbf{d}_t + \mathbf{x}_t, \\
\mathbf{d}_{t+1} &\doteq c_t \mathbf{F}_t \mathbf{d}_t + g_t \mathbf{e}_t.
\end{aligned}$$

Proof. We can write the difference between two consecutive estimates as

$$\begin{aligned}
\boldsymbol{\theta}_{t+1}^{t+1} - \boldsymbol{\theta}_t^t &= \mathbf{F}_t \boldsymbol{\theta}_t^{t+1} - \boldsymbol{\theta}_t^t + Y_t^{t+1}\mathbf{w}_k + \mathbf{x}_t \\
&= \mathbf{F}_t\left(\boldsymbol{\theta}_t^{t+1} - \boldsymbol{\theta}_t^t\right) + Y_t^{t+1}\mathbf{w}_k + (\mathbf{F}_t - \mathbf{I})\boldsymbol{\theta}_t^t + \mathbf{x}_t.
\end{aligned}$$

Now let us expand $\boldsymbol{\theta}_t^{t+1} - \boldsymbol{\theta}_t^t$:

$$\begin{aligned}
\boldsymbol{\theta}_t^{t+1} - \boldsymbol{\theta}_t^t &= \mathbf{F}_{t-1}\boldsymbol{\theta}_{t-1}^{t+1} + Y_{t-1}^{t+1}\mathbf{w}_{t-1} + \mathbf{x}_{t-1} \\
&\quad - \mathbf{F}_{t-1}\boldsymbol{\theta}_{t-1}^t - Y_{t-1}^t \mathbf{w}_{t-1} - \mathbf{x}_{t-1} \\
&= \mathbf{F}_{t-1}\left(\boldsymbol{\theta}_{t-1}^{t+1} - \boldsymbol{\theta}_{t-1}^t\right) + \left(Y_{t-1}^{t+1} - Y_{t-1}^t\right)\mathbf{w}_{t-1} \\
&= \mathbf{F}_{t-1}\cdots\mathbf{F}_0(\boldsymbol{\theta}_0^{t+1} - \boldsymbol{\theta}_0^t) + \sum_{k=0}^{t-1}\mathbf{F}_{t-1}\cdots\mathbf{F}_{k+1}(Y_k^{t+1} - Y_k^t)\mathbf{w}_k \\
&= \sum_{k=0}^{t-1}\mathbf{F}_{t-1}\cdots\mathbf{F}_{k+1}(Y_k^{t+1} - Y_k^t)\mathbf{w}_k \\
&= \sum_{k=0}^{t-1}\mathbf{F}_{t-1}\cdots\mathbf{F}_{k+1}\left(d_{k+1}(Y_{k+1}^{t+1} - Y_{k+1}^t) + b_t g_k \prod_{j=k+1}^{t-1} c_j\right)\mathbf{w}_k \\
&= \sum_{k=0}^{t-1}\mathbf{F}_{t-1}\cdots\mathbf{F}_{k+1}\Bigg(d_{k+1}\Bigg(d_{k+2}(Y_{k+2}^{t+1} - Y_{k+2}^t) \\
&\quad + b_t g_{k+1}\prod_{j=k+2}^{t-1} c_j\Bigg) + b_t g_k \prod_{j=k+1}^{t-1} c_j\Bigg)\mathbf{w}_k \\
&= \sum_{k=0}^{t-1}\mathbf{F}_{t-1}\cdots\mathbf{F}_{k+1}\Bigg(d_{k+1}d_{k+2}(Y_{k+2}^{t+1} - Y_{k+2}^t) \\
&\quad + b_t g_{k+1} d_{k+1}\prod_{j=k+2}^{t-1} c_j + b_t g_k \prod_{j=k+1}^{t-1} c_j\Bigg)\mathbf{w}_k \\
&= \sum_{k=0}^{t-1}\mathbf{F}_{t-1}\cdots\mathbf{F}_{k+1}\Bigg(\prod_{j=k+1}^{t} d_j (Y_t^{t+1} - Y_t^t)
\end{aligned}$$

$$
\begin{aligned}
&\quad + b_t \sum_{n=k}^{t-1} g_n \prod_{i=k+1}^{n} d_i \prod_{j=n+1}^{t-1} c_j \Bigg) \mathbf{w}_k \\
&= d_t (Y_t^{t+1} - Y_t^t) \underbrace{\sum_{k=0}^{t-1} \mathbf{F}_{t-1} \cdots \mathbf{F}_{k+1} \prod_{j=k+1}^{t-1} d_j \mathbf{w}_k}_{\mathbf{e}_{t-1}} \\
&\quad + b_t \underbrace{\sum_{k=0}^{t-1} \mathbf{F}_{t-1} \cdots \mathbf{F}_{k+1} \sum_{n=k}^{t-1} g_n \prod_{i=k+1}^{n} d_i \prod_{j=n+1}^{t-1} c_j \mathbf{w}_k}_{\mathbf{d}_t} \\
&= (Y_t^{t+1} - Y_t^t) d_t \mathbf{e}_{t-1} + b_t \mathbf{d}_t.
\end{aligned}
$$

The vectors $\mathbf{e}_t$ and $\mathbf{d}_t$ can be incrementally updated as follows:

$$
\begin{aligned}
\mathbf{e}_t &= \sum_{k=0}^{t} \mathbf{F}_t \cdots \mathbf{F}_{k+1} \prod_{j=k+1}^{t} d_j \mathbf{w}_k \\
&= \mathbf{w}_t + d_t \mathbf{F}_t \sum_{k=0}^{t-1} \mathbf{F}_{t-1} \cdots \mathbf{F}_{k+1} \prod_{j=k+1}^{t-1} d_j \mathbf{w}_k \\
&= \mathbf{w}_t + d_t \mathbf{F}_t \mathbf{e}_{t-1},
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{d}_t &= \sum_{k=0}^{t-1} \mathbf{F}_{t-1} \cdots \mathbf{F}_{k+1} \sum_{n=k}^{t-1} g_n \prod_{i=k+1}^{n} d_i \prod_{j=n+1}^{t-1} c_j \mathbf{w}_k \\
&= \sum_{k=0}^{t-1} \mathbf{F}_{t-1} \cdots \mathbf{F}_{k+1} \left(\sum_{n=k}^{t-2} g_n \prod_{i=k+1}^{n} d_i \prod_{j=n+1}^{t-1} c_j \mathbf{w}_k + g_{t-1} \prod_{j=k+1}^{t-1} d_j \mathbf{w}_k \right) \\
&= \sum_{k=0}^{t-1} \mathbf{F}_{t-1} \cdots \mathbf{F}_{k+1} \sum_{n=k}^{t-2} g_n \prod_{i=k+1}^{n} d_i \prod_{j=n+1}^{t-1} c_j \mathbf{w}_k + g_{t-1} \sum_{k=0}^{t-1} \mathbf{F}_{t-1} \cdots \mathbf{F}_{k+1} \prod_{j=k+1}^{t-1} d_j \mathbf{w}_k \\
&= c_{t-1} \mathbf{F}_{t-1} \sum_{k=0}^{t-2} \mathbf{F}_{t-1} \cdots \mathbf{F}_{k+1} \sum_{n=k}^{t-2} g_n \prod_{i=k+1}^{n} d_i \prod_{j=n+1}^{t-2} c_j \mathbf{w}_k + g_{t-1} \mathbf{e}_{t-1} \\
&= c_{t-1} \mathbf{F}_{t-1} \mathbf{d}_{t-1} + g_{t-1} \mathbf{e}_{t-1}.
\end{aligned}
$$

Then plugging back in

$$
\begin{aligned}
\boldsymbol{\theta}_{t+1}^{t+1} &= \boldsymbol{\theta}_t^t + \mathbf{F}_t \left(\boldsymbol{\theta}_t^{t+1} - \boldsymbol{\theta}_t^t \right) + Y_t^{t+1} \mathbf{w}_t + (\mathbf{F}_t - \mathbf{I}) \boldsymbol{\theta}_t^t + \mathbf{x}_t \\
&= \boldsymbol{\theta}_t^t + d_t \mathbf{F}_t \mathbf{e}_{t-1} (Y_t^{t+1} - Y_t^t) + b_t \mathbf{F}_t \mathbf{d}_t + Y_t^{t+1} \mathbf{w}_t + (\mathbf{F}_t - \mathbf{I}) \boldsymbol{\theta}_t^t + \mathbf{x}_t \\
&= \mathbf{F}_t \boldsymbol{\theta}_t^t + (\mathbf{e}_t - \mathbf{w}_t)(Y_t^{t+1} - Y_t^t) + Y_t^{t+1} \mathbf{w}_t + b_t \mathbf{F}_t \mathbf{d}_t + \mathbf{x}_t \\
&= \mathbf{F}_t \boldsymbol{\theta}_t^t + (Y_t^{t+1} - Y_t^t) \mathbf{e}_t + Y_t^t \mathbf{w}_t + b_t \mathbf{F}_t \mathbf{d}_t + \mathbf{x}_t. \quad \square
\end{aligned}
$$

A.4 Proof of Theorem 4

Theorem 4 (Generality of the new equivalence technique)**.** *The online equivalence technique by van Hasselt, Mahmood and Sutton (2014, Theorem 1) can be retrieved as a special case from the online equivalence technique given in Theorem 3.*

Proof. We describe the online equivalence technique by van Hasselt et al. (2014) in the following.

Consider any forward view that updates toward an interim scalar target Y_k^t with

$$\boldsymbol{\theta}_{k+1}^{t+1} = \boldsymbol{\theta}_k^{t+1} + \mu_k \left(Y_k^{t+1} - \boldsymbol{\phi}_k^\top \boldsymbol{\theta}_k^{t+1} \right) \boldsymbol{\phi}_k + \mathbf{x}_k, 0 \le k < t,$$

where $\boldsymbol{\theta}_0^t = \boldsymbol{\theta}_0$ for some initial $\boldsymbol{\theta}_0$. Assume that the temporal difference $Y_k^{t+1} - Y_k^t$ at k is related to the temporal difference at $k+1$ as follows:

$$Y_k^{t+1} - Y_k^t = d_{k+1}(Y_{k+1}^{t+1} - Y_{k+1}^t), 0 \le k < t,$$

where d_k is a scalar that can be computed using data available at time k. Then the final weight $\boldsymbol{\theta}_{t+1} \doteq \boldsymbol{\theta}_{t+1}^{t+1}$ can be computed through the following backward-view update, with $\mathbf{e}_{-1} = \mathbf{0}$ and $t \ge 0$:

$$\begin{aligned} \mathbf{e}_t &= \mu_t \boldsymbol{\phi}_t + d_t (\mathbf{I} - \mu_t \boldsymbol{\phi}_t \boldsymbol{\phi}_t^\top) \mathbf{e}_{t-1}, \\ \boldsymbol{\theta}_{t+1} &= \boldsymbol{\theta}_t + (Y_t^{t+1} - Y_t^t) \mathbf{e}_t + \mu_t (Y_t^t - \boldsymbol{\phi}_t^\top \boldsymbol{\theta}_t) \boldsymbol{\phi}_t + \mathbf{x}_t. \end{aligned}$$

The above equivalence technique can be obtained from Theorem 3 as a special case by substituting $\mathbf{F}_k = \mathbf{I} - \mu_k \boldsymbol{\phi}_k \boldsymbol{\phi}_k^\top$, $\mathbf{w}_k = \mu_k \boldsymbol{\phi}_k$ and $b_k = 0$. □

A.5 Proof of Theorem 5

Theorem 5 (Backward view update for $\boldsymbol{\alpha}_t$ of WIS-TD(λ))**.** *The step-size vector $\boldsymbol{\alpha}_t$ computed by the following backward-view update and the forward-view update defined by (18) – (20) are equal at each step t:*

$$\mathbf{u}_{t+1} \doteq (\mathbf{1} - \eta \boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{u}_t + \rho_t \boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t + (\rho_t - 1)\gamma_t \lambda_t (\mathbf{1} - \eta \boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{v}_t, \tag{22}$$

$$\mathbf{v}_{t+1} \doteq \gamma_t \lambda_t \rho_t (\mathbf{1} - \eta \boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{v}_t + \rho_t \boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t, \tag{23}$$

$$\boldsymbol{\alpha}_{t+1} \doteq \mathbf{1} \oslash \mathbf{u}_{t+1}. \tag{24}$$

Proof. First, note that the component-wise vector multiplication in (19) can be written equivalently as a matrix-vector multiplication in the following way:

$$(\mathbf{1} - \eta \boldsymbol{\phi}_k \circ \boldsymbol{\phi}_k) \circ \mathbf{u}_k^{t+1} = (\mathbf{I} - \eta \mathrm{Diag}\left(\boldsymbol{\phi}_k \circ \boldsymbol{\phi}_k\right)) \mathbf{u}_k^{t+1},$$

where $\mathrm{Diag}(\mathbf{v}) \in \mathbb{R}^{|\mathbf{v}| \times |\mathbf{v}|}$ is a diagonal matrix with the components of $\mathbf{v}$ in its diagonal.

In Theorem 3, we substitute $\boldsymbol{\theta}_k^{t+1} = \mathbf{u}_k^{t+1}$, $\mathbf{F}_k = (\mathbf{I} - \eta \mathrm{Diag}\left(\boldsymbol{\phi}_k \circ \boldsymbol{\phi}_k\right))$, $\mathbf{x}_k = \mathbf{0}$, $\mathbf{w}_k = \boldsymbol{\phi}_k \circ \boldsymbol{\phi}_k$ and $Y_k^{t+1} = \tilde{\rho}_k^{t+1}$.

Now, $\tilde{\rho}_k^{t+1}$ can be recursively in t written as follows

$$\begin{aligned} \tilde{\rho}_k^{t+1} &= \rho_k \sum_{i=k+1}^{t} C_k^{i-1}(1 - \gamma_i \lambda_i) + \rho_k C_k^t \\ &= \rho_k \sum_{i=k+1}^{t-1} C_k^{i-1}(1 - \gamma_i \lambda_i) + \rho_k C_k^{t-1}(1 - \gamma_t \lambda_t) + \rho_k C_k^t \\ &= \rho_k \sum_{i=k+1}^{t-1} C_k^{i-1}(1 - \gamma_i \lambda_i) + \rho_k C_k^{t-1} + \rho_k C_k^{t-1} \rho_t \gamma_t \lambda_t - \rho_k C_k^{t-1} \gamma_t \lambda_t \\ &= \tilde{\rho}_k^t + (\rho_t - 1)\gamma_t \lambda_t \rho_k C_k^{t-1}. \end{aligned}$$

Hence, it proves that

$$Y_k^{t+1} - Y_k^t = d_{k+1}\left(Y_{k+1}^{t+1} - Y_{k+1}^t\right) + b_t g_k \prod_{j=k+1}^{t-1} c_j, 0 \le k < t,$$

with $d_i = 0$, $b_i = (\rho_i - 1)\gamma_i\lambda_i$, $g_i = \rho_i$ and $c_i = \gamma_i\lambda_i\rho_i, \forall i$.

Inserting these substitutes in Theorem 3 yields us the backward-view defined by (22) – (24). □

A.6 Proof of Theorem 6

Theorem 6 (Backward view update for $\boldsymbol{\theta}_t^t$ of WIS-TD(λ))**.** *The parameter vector $\boldsymbol{\theta}_t$ computed by the following backward-view update and the parameter vector $\boldsymbol{\theta}_t^t$ computed by the forward-view update defined by (17) and (21) are equal at every time step t:*

$$\mathbf{e}_t \doteq \rho_t \boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t + \gamma_t\lambda_t\rho_t\left(\mathbf{e}_{t-1} - \rho_t\left(\boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t\right)\boldsymbol{\phi}_t^\top \mathbf{e}_{t-1}\right), \tag{25}$$

$$\begin{aligned}\boldsymbol{\theta}_{t+1} \doteq\ & \boldsymbol{\theta}_t + \boldsymbol{\alpha}_{t+1} \circ \rho_t\left(\boldsymbol{\theta}_{t-1}^\top\boldsymbol{\phi}_t - \boldsymbol{\theta}_t^\top\boldsymbol{\phi}_t\right)\boldsymbol{\phi}_t + (R_{t+1} + \gamma_{t+1}\boldsymbol{\theta}_t^\top\boldsymbol{\phi}_{t+1} - \boldsymbol{\theta}_{t-1}^\top\boldsymbol{\phi}_t)\mathbf{e}_t \\ & + (\rho_t - 1)\gamma_t\lambda_t\left(\mathbf{d}_t - \rho_t\left(\boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t\right)\boldsymbol{\phi}_t^\top\mathbf{d}_t\right),\end{aligned} \tag{26}$$

$$\mathbf{d}_{t+1} \doteq \gamma_t\lambda_t\rho_t\left(\mathbf{d}_t - \rho_t\left(\boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t\right)\boldsymbol{\phi}_t^\top\mathbf{d}_t\right) + \left(R_{t+1} + \boldsymbol{\theta}_t^\top\boldsymbol{\phi}_{t+1} - \boldsymbol{\theta}_{t-1}^\top\boldsymbol{\phi}_t\right)\mathbf{e}_t. \tag{27}$$

Proof. First, we redefine (21) for convenience:

$$\boldsymbol{\theta}_{k+1}^{t+1} \doteq \boldsymbol{\theta}_k^{t+1} + \boldsymbol{\alpha}_{k+1} \circ \rho_k\left(\zeta_{k,t+1}^\rho - \boldsymbol{\phi}_k^\top\boldsymbol{\theta}_k^{t+1}\right)\boldsymbol{\phi}_k, \tag{28}$$

where $G_{k,t+1}^\rho = \rho_k\zeta_{k,t+1}^\rho$. Hence, $\zeta_{k,t+1}^\rho$ can be given by:

$$\begin{aligned}\zeta_{k,t+1}^\rho \doteq\ & C_k^t\Big((1-\gamma_{t+1})G_k^{t+1} + \gamma_{t+1}\left(G_k^{t+1} + \boldsymbol{\phi}_{t+1}^\top\boldsymbol{\theta}_t\right)\Big) + \sum_{i=k+1}^{t} C_k^{i-1}\Big((1-\gamma_i)G_k^i + \gamma_i(1-\lambda_i)\left(G_k^i + \boldsymbol{\phi}_i^\top\boldsymbol{\theta}_{i-1}\right)\Big) \\ & - \left(C_k^t + \sum_{i=k+1}^{t} C_k^{i-1}(1-\gamma_i\lambda_i) - 1\right)\boldsymbol{\phi}_k^\top\boldsymbol{\theta}_{k-1}.\end{aligned}$$

In Theorem 3, we substitute $\mathbf{F}_k = \mathbf{I} - \rho_k(\boldsymbol{\alpha}_{k+1} \circ \boldsymbol{\phi}_k)\boldsymbol{\phi}_k^\top$, $\mathbf{w}_k = \rho_k\boldsymbol{\alpha}_{k+1} \circ \boldsymbol{\phi}_k$, $Y_k^{t+1} = \zeta_{k,t+1}^\rho$ and $\mathbf{x}_k = 0, \forall k$, to get (28). Now, the next step is to establish a recursive relation for ζ^ρ both in k and t. For that, we use the following identities:

$$\begin{aligned}G_k^{k+1} &= R_{k+1}, \\ G_k^{t+1} &= \sum_{i=k}^{t} R_{i+1} = R_{k+1} + G_{k+1}^{t+1}.\end{aligned}$$

First we establish the recurrence relation in k:

$$\begin{aligned}\zeta_{k,t+1}^\rho =\ & C_k^t\Big((1-\gamma_{t+1})G_k^{t+1} + \gamma_{t+1}\left(G_k^{t+1} + \boldsymbol{\phi}_{t+1}^\top\boldsymbol{\theta}_t\right)\Big) + \sum_{i=k+1}^{t} C_k^{i-1}\Big((1-\gamma_i)G_k^i + \gamma_i(1-\lambda_i)\left(G_k^i + \boldsymbol{\phi}_i^\top\boldsymbol{\theta}_{i-1}\right)\Big) \\ & - \left(C_k^t + \sum_{i=k+1}^{t} C_k^{i-1}(1-\gamma_i\lambda_i) - 1\right)\boldsymbol{\phi}_k^\top\boldsymbol{\theta}_{k-1} \\ =\ & C_k^t\Big((1-\gamma_{t+1})\left(R_{k+1} + G_{k+1}^{t+1}\right) + \gamma_{t+1}\left(R_{k+1} + G_{k+1}^{t+1} + \boldsymbol{\phi}_{t+1}^\top\boldsymbol{\theta}_t\right)\Big) \\ & + \Big((1-\gamma_{k+1})G_k^{k+1} + \gamma_{k+1}(1-\lambda_{k+1})\left(G_k^{k+1} + \boldsymbol{\phi}_{k+1}^\top\boldsymbol{\theta}_k\right)\Big) \\ & + \sum_{i=k+2}^{t} C_k^{i-1}\Big((1-\gamma_i)\left(R_{k+1} + G_{k+1}^i\right) + \gamma_i(1-\lambda_i)\left(R_{k+1} + G_{k+1}^i + \boldsymbol{\phi}_i^\top\boldsymbol{\theta}_{i-1}\right)\Big)\end{aligned}$$

$$
\begin{aligned}
&- \left(C_k^t + \sum_{i=k+1}^{t} C_k^{i-1}(1-\gamma_i\lambda_i) - 1 \right) \boldsymbol{\phi}_k^\top \boldsymbol{\theta}_{k-1} \\
&= \rho_{k+1}\gamma_{k+1}\lambda_{k+1} C_{k+1}^t \left((1-\gamma_{t+1}) G_{k+1}^{t+1} + \gamma_{t+1}(G_{k+1}^{t+1} + \boldsymbol{\phi}_{t+1}^\top \boldsymbol{\theta}_t) \right) \\
&+ \rho_{k+1}\gamma_{k+1}\lambda_{k+1} \sum_{i=k+2}^{t} C_{k+1}^{i-1} \Big((1-\gamma_i) G_{k+1}^i + \gamma_i (1-\lambda_i) \left(G_{k+1}^i + \boldsymbol{\phi}_i^\top \boldsymbol{\theta}_{i-1} \right) \Big) \\
&- \rho_{k+1}\gamma_{k+1}\lambda_{k+1} \left(C_{k+1}^t + \sum_{i=k+2}^{t} C_{k+1}^{i-1}(1-\gamma_i\lambda_i) - 1 \right) \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k \\
&+ \left(C_k^t + \sum_{i=k+2}^{t} C_k^{i-1}(1-\gamma_i\lambda_i) - \rho_{k+1}\gamma_{k+1}\lambda_{k+1} \right) \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k \\
&+ C_k^t R_{k+1} + (1-\gamma_{k+1}\lambda_{k+1}) R_{k+1} + \gamma_{k+1}(1-\lambda_{k+1}) \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k \\
&+ R_{k+1} \sum_{i=k+2}^{t} C_k^{i-1}(1-\gamma_i\lambda_i) \\
&- \left(C_k^t + \sum_{i=k+1}^{t} C_k^{i-1}(1-\gamma_i\lambda_i) - 1 \right) \boldsymbol{\phi}_k^\top \boldsymbol{\theta}_{k-1} \\
&= \rho_{k+1}\gamma_{k+1}\lambda_{k+1} \zeta_{k+1,t+1}^\rho + \left(C_k^t + \sum_{i=k+1}^{t} C_k^{i-1}(1-\gamma_i\lambda_i) - 1 \right) \left(R_{k+1} + \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k - \boldsymbol{\phi}_k^\top \boldsymbol{\theta}_{k-1} \right) \\
&+ R_{k+1} + \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k - \rho_{k+1}\gamma_{k+1}\lambda_{k+1} \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k + \gamma_{k+1}(1-\lambda_{k+1}) \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k - (1-\gamma_{k+1}\lambda_{k+1}) \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k \\
&= \rho_{k+1}\gamma_{k+1}\lambda_{k+1} \zeta_{k+1,t+1}^\rho + \left(C_k^t + \sum_{i=k+1}^{t} C_k^{i-1}(1-\gamma_i\lambda_i) - 1 \right) \left(R_{k+1} + \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k - \boldsymbol{\phi}_k^\top \boldsymbol{\theta}_{k-1} \right) \\
&+ R_{k+1} + \gamma_{k+1} \left(1 - \rho_{k+1}\lambda_{k+1} \right) \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k .
\end{aligned}
$$

Then the recurrence in t can be established by subtracting $\zeta_{k,t}^\rho$ from $\zeta_{k,t+1}^\rho$:

$$
\begin{aligned}
\zeta_{k,t+1}^\rho - \zeta_{k,t}^\rho &\doteq \rho_{k+1}\gamma_{k+1}\lambda_{k+1} \zeta_{k+1,t+1}^\rho + \left(C_k^t + \sum_{i=k+1}^{t} C_k^{i-1}(1-\gamma_i\lambda_i) - 1 \right) \left(R_{k+1} + \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k - \boldsymbol{\phi}_k^\top \boldsymbol{\theta}_{k-1} \right) \\
&\quad + R_{k+1} + \gamma_{k+1} \left(1 - \rho_{k+1}\lambda_{k+1} \right) \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k \\
&- \rho_{k+1}\gamma_{k+1}\lambda_{k+1} \zeta_{k+1,t}^\rho - \left(C_k^{t-1} + \sum_{i=k+1}^{t-1} C_k^{i-1}(1-\gamma_i\lambda_i) - 1 \right) \left(R_{k+1} + \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k - \boldsymbol{\phi}_k^\top \boldsymbol{\theta}_{k-1} \right) \\
&\quad - R_{k+1} + \gamma_{k+1} \left(1 - \rho_{k+1}\lambda_{k+1} \right) \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k \\
&= \rho_{k+1}\gamma_{k+1}\lambda_{k+1} \left(\zeta_{k+1,t+1}^\rho - \zeta_{k+1,t}^\rho \right) \\
&\quad + \left(C_k^t - C_k^{t-1} + C_k^{t-1}(1-\gamma_t\lambda_t) \right) \left(R_{k+1} + \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k - \boldsymbol{\phi}_k^\top \boldsymbol{\theta}_{k-1} \right) \\
&= \rho_{k+1}\gamma_{k+1}\lambda_{k+1} \left(\zeta_{k+1,t+1}^\rho - \zeta_{k+1,t}^\rho \right) + (\rho_t - 1)\gamma_t\lambda_t C_k^{t-1} \left(R_{k+1} + \boldsymbol{\phi}_{k+1}^\top \boldsymbol{\theta}_k - \boldsymbol{\phi}_k^\top \boldsymbol{\theta}_{k-1} \right) .
\end{aligned}
$$

The above recurrence relation establishes

$$
Y_k^{t+1} - Y_k^t = d_{k+1} \left(Y_{k+1}^{t+1} - Y_{k+1}^t \right) + b_t g_k \prod_{j=k+1}^{t-1} c_j, 0 \leq k < t,
$$

with $d_i = \rho_i\gamma_i\lambda_i$, $b_i = (\rho_i - 1)\gamma_i\lambda_i$, $g_i = R_{i+1} + \boldsymbol{\phi}_{i+1}^\top \boldsymbol{\theta}_i - \boldsymbol{\phi}_i^\top \boldsymbol{\theta}_{i-1}$ and $c_i = \gamma_i\lambda_i\rho_i, \forall i$. Inserting these substitutes in Theorem 3 yields us the backward-view defined by (25) – (27). $\square$

A.7 Description of WIS-TD(λ), WIS-GTD(λ), WIS-TO-GTD(λ), U-TD(λ) and U-TO-TD(λ)

Algorithm 1 WIS-TD(λ)

Initialization:
Choose $\boldsymbol{\theta}_0, u_0 \geq 0, \eta \geq 0$
Set $\mathbf{u}_0 = u_0\mathbf{1}, \mathbf{v}_0 = \mathbf{0}, \mathbf{e}_{-1} = \mathbf{0}, \mathbf{d}_0 = \mathbf{0}$
for $t = 0, 1, \cdots$ **do**
receive $\boldsymbol{\phi}_t, \rho_t, \gamma_t, \lambda_t, R_{t+1}, \boldsymbol{\phi}_{t+1}, \gamma_{t+1}, \lambda_{t+1}$

$$\mathbf{u}_{t+1} = (\mathbf{1} - \eta\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{u}_t + \rho_t\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t + (\rho_t - 1)\gamma_t\lambda_t (\mathbf{1} - \eta\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{v}_t$$
$$\mathbf{v}_{t+1} = \gamma_t\lambda_t\rho_t (\mathbf{1} - \eta\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{v}_t + \rho_t\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t$$
$$\boldsymbol{\alpha}_{t+1} = \mathbf{1} \oslash \mathbf{u}_{t+1}$$
$$\mathbf{e}_t = \rho_t\boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t + \gamma_t\lambda_t\rho_t \left(\mathbf{e}_{t-1} - \rho_t (\boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t) \boldsymbol{\phi}_t^\top \mathbf{e}_{t-1}\right)$$
$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \boldsymbol{\alpha}_{t+1} \circ \rho_t \left(\boldsymbol{\theta}_{t-1}^\top \boldsymbol{\phi}_t - \boldsymbol{\theta}_t^\top \boldsymbol{\phi}_t\right) \boldsymbol{\phi}_t + (R_{t+1} + \gamma_{t+1}\boldsymbol{\theta}_t^\top \boldsymbol{\phi}_{t+1} - \boldsymbol{\theta}_{t-1}^\top \boldsymbol{\phi}_t)\mathbf{e}_t + (\rho_t - 1)\gamma_t\lambda_t \left(\mathbf{d}_t - \rho_t (\boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t) \boldsymbol{\phi}_t^\top \mathbf{d}_t\right)$$
$$\mathbf{d}_{t+1} = \gamma_t\lambda_t\rho_t \left(\mathbf{d}_t - \rho_t (\boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t) \boldsymbol{\phi}_t^\top \mathbf{d}_t\right) + \left(R_{t+1} + \boldsymbol{\theta}_t^\top \boldsymbol{\phi}_{t+1} - \boldsymbol{\theta}_{t-1}^\top \boldsymbol{\phi}_t\right) \mathbf{e}_t$$

end for

Algorithm 2 WIS-GTD(λ)

Initialization:
Choose $\boldsymbol{\theta}_0, \mathbf{w}_0, u_0 \geq 0, \eta \geq 0, \beta \geq 0$
Set $\mathbf{u}_0 = u_0\mathbf{1}, \mathbf{v}_0 = \mathbf{0}, \mathbf{e}_{-1} = \mathbf{0}$
for $t = 0, 1, \cdots$ **do**
receive $\boldsymbol{\phi}_t, \rho_t, \gamma_t, \lambda_t, R_{t+1}, \boldsymbol{\phi}_{t+1}, \gamma_{t+1}, \lambda_{t+1}$

$$\mathbf{u}_{t+1} = (\mathbf{1} - \eta\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{u}_t + \rho_t\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t + (\rho_t - 1)\gamma_t\lambda_t (\mathbf{1} - \eta\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{v}_t$$
$$\mathbf{v}_{t+1} = \gamma_t\lambda_t\rho_t (\mathbf{1} - \eta\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{v}_t + \rho_t\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t$$
$$\boldsymbol{\alpha}_{t+1} = \mathbf{1} \oslash \mathbf{u}_{t+1}$$
$$\mathbf{e}_t = \rho_t (\gamma_t\lambda_t\mathbf{e}_{t-1} + \boldsymbol{\phi}_t)$$
$$\delta_t = R_{t+1} + \gamma_{t+1}\boldsymbol{\theta}_t^\top \boldsymbol{\phi}_{t+1} - \boldsymbol{\theta}_t^\top \boldsymbol{\phi}_t$$
$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \boldsymbol{\alpha}_{t+1} \circ \delta_t\mathbf{e}_t - \boldsymbol{\alpha}_{t+1} \circ \gamma_{t+1}(1 - \lambda_{t+1})(\mathbf{e}_t^\top \mathbf{w}_t)\boldsymbol{\phi}_{t+1}$$
$$\mathbf{w}_{t+1} = \mathbf{w}_t + \beta \left[\delta_t\mathbf{e}_t - (\mathbf{w}_t^\top \boldsymbol{\phi}_t)\boldsymbol{\phi}_t\right]$$

end for

Algorithm 3 WIS-TO-GTD(λ)

Initialization:
Choose $\boldsymbol{\theta}_0, \mathbf{w}_0, u_0 \geq 0, \eta \geq 0, \beta \geq 0$
Set $\mathbf{u}_0 = u_0\mathbf{1}, \mathbf{v}_0 = \mathbf{0}, \mathbf{e}_{-1} = \mathbf{e}_{-1}^\nabla = \mathbf{e}_{-1}^\mathbf{w} = \mathbf{0}, \rho' = 0$
for $t = 0, 1, \cdots$ **do**
receive $\boldsymbol{\phi}_t, \rho_t, \gamma_t, \lambda_t, R_{t+1}, \boldsymbol{\phi}_{t+1}, \gamma_{t+1}, \lambda_{t+1}$

$$\mathbf{u}_{t+1} = (\mathbf{1} - \eta\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{u}_t + \rho_t\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t + (\rho_t - 1)\gamma_t\lambda_t (\mathbf{1} - \eta\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{v}_t$$
$$\mathbf{v}_{t+1} = \gamma_t\lambda_t\rho_t (\mathbf{1} - \eta\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{v}_t + \rho_t\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t$$
$$\boldsymbol{\alpha}_{t+1} = \mathbf{1} \oslash \mathbf{u}_{t+1}$$
$$\mathbf{e}_t = \rho_t\boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t + \gamma_t\lambda_t\rho_t \left(\mathbf{e}_{t-1} - \rho_t (\boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t) \boldsymbol{\phi}_t^\top \mathbf{e}_{t-1}\right)$$
$$\mathbf{e}_t^\nabla = \rho_t (\gamma_t\lambda_t\mathbf{e}_{t-1} + \boldsymbol{\phi}_t)$$
$$\mathbf{e}_t^\mathbf{w} = \gamma_t\lambda_t\rho'\mathbf{e}_{t-1}^\mathbf{w} + \beta \left(1 - \gamma_t\lambda_t\rho'\boldsymbol{\phi}_t^\top \mathbf{e}_{t-1}^\mathbf{w}\right) \boldsymbol{\phi}_t$$
$$\delta_t = R_{t+1} + \gamma_{t+1}\boldsymbol{\theta}_t^\top \boldsymbol{\phi}_{t+1} - \boldsymbol{\theta}_t^\top \boldsymbol{\phi}_t$$
$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \delta_t\mathbf{e}_t + (\mathbf{e}_t - \boldsymbol{\alpha}_{t+1} \circ \rho_t\boldsymbol{\phi}_t)(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1})^\top \boldsymbol{\phi}_t - \boldsymbol{\alpha}_{t+1} \circ \gamma_{t+1}(1 - \lambda_{t+1})(\mathbf{w}_t^\top \mathbf{e}_t^\nabla)\boldsymbol{\phi}_{t+1}$$
$$\mathbf{w}_{t+1} = \mathbf{w}_t + \rho_t\delta_t\mathbf{e}_t^\mathbf{w} - \beta(\mathbf{w}_t^\top \boldsymbol{\phi}_t)\boldsymbol{\phi}_t$$
$$\rho' = \rho_t$$

end for

Algorithm 4 U-TD(λ)

Initialization:
Choose $\boldsymbol{\theta}_0, u_0 \geq 0, \eta \geq 0$
Set $\mathbf{u}_0 = u_0\mathbf{1}, , \mathbf{e}_{-1} = \mathbf{0}$
for $t = 0, 1, \cdots$ **do**
receive $\boldsymbol{\phi}_t, \gamma_t, \lambda_t, R_{t+1}, \boldsymbol{\phi}_{t+1}, \gamma_{t+1}, \lambda_{t+1}$

$$\mathbf{u}_{t+1} = (\mathbf{1} - \eta\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{u}_t + \boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t$$
$$\boldsymbol{\alpha}_{t+1} = \mathbf{1} \oslash \mathbf{u}_{t+1}$$
$$\mathbf{e}_t = \gamma_t\lambda_t\mathbf{e}_{t-1} + \boldsymbol{\phi}_t$$
$$\delta_t = R_{t+1} + \gamma_{t+1}\boldsymbol{\theta}_t^\top \boldsymbol{\phi}_{t+1} - \boldsymbol{\theta}_t^\top \boldsymbol{\phi}_t$$
$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \boldsymbol{\alpha}_{t+1} \circ \delta_t\mathbf{e}_t$$

end for

Algorithm 5 U-TO-TD(λ)

Initialization:
Choose $\boldsymbol{\theta}_0, u_0 \geq 0, \eta \geq 0$
Set $\mathbf{u}_0 = u_0\mathbf{1}, \mathbf{e}_{-1} = \mathbf{0}$
for $t = 0, 1, \cdots$ **do**
receive $\boldsymbol{\phi}_t, \gamma_t, \lambda_t, R_{t+1}, \boldsymbol{\phi}_{t+1}, \gamma_{t+1}, \lambda_{t+1}$

$$\mathbf{u}_{t+1} = (\mathbf{1} - \eta\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t) \circ \mathbf{u}_t + \boldsymbol{\phi}_t \circ \boldsymbol{\phi}_t$$
$$\boldsymbol{\alpha}_{t+1} = \mathbf{1} \oslash \mathbf{u}_{t+1}$$
$$\mathbf{e}_t = \boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t + \gamma_t\lambda_t \left(\mathbf{e}_{t-1} - (\boldsymbol{\alpha}_{t+1} \circ \boldsymbol{\phi}_t) \boldsymbol{\phi}_t^\top \mathbf{e}_{t-1}\right)$$
$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \boldsymbol{\alpha}_{t+1} \circ \left(\boldsymbol{\theta}_{t-1}^\top \boldsymbol{\phi}_t - \boldsymbol{\theta}_t^\top \boldsymbol{\phi}_t\right) \boldsymbol{\phi}_t + (R_{t+1} + \gamma_{t+1}\boldsymbol{\theta}_t^\top \boldsymbol{\phi}_{t+1} - \boldsymbol{\theta}_{t-1}^\top \boldsymbol{\phi}_t)\mathbf{e}_t$$

end for