A Bounding the sum of confidence set widths

We are interested in bounding $\min\{\tau \sum_{k=1}^{m} \sum_{i=1}^{\tau} \min\{\beta_k s_{t_k+i}, a_{t_k+i}), 1\}, T\}$ which we claim is $O(\tau S\sqrt{AT\log(SAT)}$ for $\beta_k(s,a) := \sqrt{\frac{14S\log(2SAmt_k)}{\max\{1, N_{t_k}(s,a)\}}}$.

Proof. In a manner similar to [4] we can say:

$$\sum_{k=1}^{m}\sum_{i=1}^{\tau}\sqrt{\frac{14S\log(2SAmt_k)}{\max\{1,N_{t_k}(s,a)\}}} \quad\leq\quad \sum_{k=1}^{m}\sum_{i=1}^{\tau}\mathbb{1}_{\{N_{t_k}\leq\tau\}} + \sum_{k=1}^{m}\sum_{i=1}^{\tau}\mathbb{1}_{\{N_{t_k}>\tau\}}\sqrt{\frac{14S\log(2SAmt_k)}{\max\{1,N_{t_k}(s,a)\}}}$$

Now, the consider the event $(s_t, a_t) = (s,a)$ and $(N_{t_k}(s,a) \leq \tau)$. This can happen fewer than 2τ times per state action pair. Therefore, $\sum_{k=1}^{m}\sum_{i=1}^{\tau}\mathbf{1}(N_{t_k}(s,a) \leq \tau) \leq 2\tau SA$.Now, suppose $N_{t_k}(s,a) > \tau$. Then for any $t \in \{t_k, .., t_{k+1}-1\}$, $N_t(s,a) + 1 \leq N_{t_k}(s,a) + \tau \leq 2N_{t_k}(s,a)$. Therefore:

$$\begin{aligned}
\sum_{k=1}^{m}\sum_{t=t_k}^{t_{k+1}-1}\sqrt{\frac{\mathbf{1}(N_{t_k}(s_t,a_t)>\tau)}{N_{t_k}(s_t,a_t)}} \quad &\leq\quad \sum_{k=1}^{m}\sum_{t=t_k}^{t_{k+1}-1}\sqrt{\frac{2}{N_t(s_t,a_t)+1}} = \sqrt{2}\sum_{t=1}^{T}(N_t(s_t,a_t)+1)^{-1/2} \\
&\leq\quad \sqrt{2}\sum_{s,a}\sum_{j=1}^{N_{T+1}(s,a)} j^{-1/2} \leq \sqrt{2}\sum_{s,a}\int_{x=0}^{N_{T+1}(s,a)} x^{-1/2}\,dx \\
&\leq\quad \sqrt{2SA\sum_{s,a}N_{T+1}(s,a)} = \sqrt{2SAT}
\end{aligned}$$

Note that since all rewards and transitions are absolutely constrained $\in [0,1]$ our regret

$$\begin{aligned}
\min\{\tau\sum_{k=1}^{m}\sum_{i=1}^{\tau}\min\{\beta_k(s_{t_k+i}, a_{t_k+i}), 1\}, T\} \quad &\leq\quad \min\{2\tau^2 SA + \tau\sqrt{28S^2AT\log(SAT)}, T\} \\
&\leq\quad \sqrt{2\tau^2 SAT} + \tau\sqrt{28S^2AT\log(SAT)} \leq \tau S\sqrt{30AT\log(SAT)}
\end{aligned}$$

Which is our required result. ☐