Contraction of the log distance function

The one step TD-update for transition (s_1, s_2, a) with the log-distance approach is

$$D^*(s_1, s_g, a) \leftarrow \log\left(1 + \min_b [e^{D(s_2,s_g,b)}]\right)$$

To show this update is a contraction in the sup-norm, we need to show that for some $\gamma \in [0, 1)$:

$$||D_1^* - D_2^*||_\infty < \gamma ||D_1 - D_2||_\infty$$

For the below derivation, we assume that the log distances D_1 and D_2 are bounded on $[1, M]$, where M is the maximum log-distance allowed (this bound is necessary for the contraction property). Note: the following derivation is for a deterministic environment. The extension to stochastic transitions is straightforward.

$$\|D_1^*(s_1, s_g, a) - D_2^*(s_1, s_g, a)\| =$$

$$\begin{aligned}
&= \max_{(s_1,s_g,a)} \left\| \log(1 + \min_b [e^{D_1(s_2,s_g,b)]}) - \log(1 + \min_b [e^{D_2(s_2,s_g,b})]) \right\| && (1)\\
&= \max_{(s_1,s_g,a)} \left\| \log \frac{1 + \min_b [e^{D_1(s_2,s_g,b)]}}{1 + \min_b [e^{D_2(s_2,s_g,b})]} \right\| && (2)\\
&< \max_{(s_1,s_g,a)} \gamma \left\| \log \frac{\min_b [e^{D_1(s_2,s_g,b)]}}{\min_b [e^{D_2(s_2,s_g,b})]} \right\| && (3)\\
&= \gamma \max_{(s_1,s_g,a)} \left\| \log \min_b [e^{D_1(s_2,s_g,b)}] - \log \min_b [e^{D_2(s_2,s_g,b})] \right\| && (4)\\
&= \gamma \max_{(s_1,s_g,a)} \left\| \min_b [D_1(s_2, s_g, b)] - \min_b [D_2(s_2, s_g, b)] \right\| && (5)\\
&\leq \gamma \max_{(s_1,s_g,a)} \max_b \|[D_1(s_2, s_g, b)] - [D_2(s_2, s_g, b)]\| && (6)\\
&= \gamma \max_{(s_2,s_g,b)} \|[D_1(s_2, s_g, b)] - [D_2(s_2, s_g, b)]\| && (7)\\
&= \gamma \|D_1 - D_2\|_\infty && (8)
\end{aligned}$$

The strict inequality in step (3) comes from the fact that $|\log \frac{1+A}{1+B}| < |\log \frac{A}{B}|$, and since the distance functions D_1 and D_2 are bounded on $[1, M]$, there must exist a $\gamma \in [0, 1)$ such that this inequality is satisfied for all values of D_1 and D_2. The inequality in step (6) comes from the fact that $\max_x |f(x) - g(x)| \geq |\min_x f(x) - \min_x g(x)|$. To get the last line, we notice that maximizing over s_1 and a is the same as maximizing over s_2 (assuming every state in the environment is reachable by some other state, which is a fine assumption - if this isn't the case, just redefine the environment as only reachable states, since those are the only ones we care about for reinforcement learning anyway). Notice that with the log formulation, we don't even need to set a discount factor (γ) for contraction, we get one inherently from the update in log-space! Although it is worth noting that the effective γ increases as we increase the maximum log-distance M.