APPENDIX

Proof of Lemma 1

Suppose that at each discrete time step k, the proposed algorithm selects $u_k \in \widetilde{\mathcal{C}}(x(t_k))$. Consider the time interval $[t_k, t_{k+1}]$ where $t_{k+1} - t_k = \Delta T$ and suppose that we use a fixed control input $u(t) \equiv u_k$ on $[t_k, t_{k+1}]$ and $x(t_k) \in \mathcal{S}$. Then $x(t) \in \mathcal{S}$ for all $t \in [t_k, t_{k+1}]$.

Proof: To show this holds, if suffices to show that that u_k remains in $\mathcal{C}(x(t))$ for all $t \in [t_k, t_{k+1}]$.

By assumption we have that $x(t_k) \in \mathcal{S}$ which is equivalent to $h(x(t_k)) \geq 0$. Also by assumption we have that $u(t_k) = u_k \in \widetilde{\mathcal{C}}(x(t_k)) \subseteq \mathcal{C}(x(t_k))$. Let us start by showing that for any $\delta \in [0, \Delta T]$, the control input is in $\mathcal{C}(x(t_k + \delta))$. Using the Lipschitz continuity of the Lie derivatives, the solution to the dynamics, and the bound on the control inputs, we have that

$$\begin{aligned} &L_f h(x(t_k + \delta)) - L_f h(x(t_k)) + (L_g h(x(t_k + \delta)) - L_g h(x(t_k)))\, u_k \\ &\geq -\mathfrak{L}_h \mathfrak{L}_x (\mathfrak{L}_f + \mathfrak{L}_g \|u_k\|) \cdot \delta \\ &\geq -\mathfrak{L}_h \mathfrak{L}_x (\mathfrak{L}_f + \mathfrak{L}_g B_u) \cdot \delta. \end{aligned}$$

Rearranging, we have that

$$L_f h(x(t_k + \delta)) + L_g h(x(t_k + \delta)) u_k \geq L_f h(x(t_k)) + L_g h(x(t_k)) - \mathfrak{L}_h \mathfrak{L}_x (\mathfrak{L}_f + \mathfrak{L}_g B_u) \cdot \delta.$$

Adding and subtracting $\mathfrak{L}_{\alpha \circ h} \mathfrak{L}_x \delta$ on the right hand side of the inequality, we have that

$$\begin{aligned} L_f h(x(t_k + \delta)) + L_g h(x(t_k + \delta)) u_k &\geq L_f h(x(t_k)) + L_g h(x(t_k)) - \mathfrak{L}_h \mathfrak{L}_x (\mathfrak{L}_f + \mathfrak{L}_g B_u + \mathfrak{L}_{\mathcal{K} \circ h}) \cdot \delta + \mathfrak{L}_{\mathcal{K} \circ h} \mathfrak{L}_x \delta \\ &\geq -\mathcal{K}(h(x(t_k)) + \mathfrak{L}_{\mathcal{K} \circ h} \mathfrak{L}_x \delta. \end{aligned}$$

Since $\mathcal{K} \circ h$ is Lipschitz continuous, we also have that

$$|\mathcal{K}(h(x(t_k + \delta)) - \mathcal{K}(h(x(t_k))| \leq \mathfrak{L}_{\mathcal{K} \circ h} \mathfrak{L}_x \delta \;\Leftrightarrow\; -\mathfrak{L}_{\mathcal{K} \circ h} \mathfrak{L}_x \delta \leq \mathcal{K}(h(x(t_k + \delta)) - \mathcal{K}(h(x(t_k)) \leq \mathfrak{L}_{\mathcal{K} \circ h} \mathfrak{L}_x \delta.$$

Therefore, by adding and subtracting $\mathcal{K}(h(x(t_k + \delta))$ on the right hand side of the above inequality, we deduce that

$$\begin{aligned} L_f h(x(t_k + \delta)) + L_g h(x(t_k + \delta)) u_k &\geq \mathcal{K}(h(x(t_k + \delta)) - \mathcal{K}(h(x(t_k)) + \mathfrak{L}_{\mathcal{K} \circ h} \mathfrak{L}_x \delta - \mathcal{K}(h(x(t_k + \delta)) \\ &\geq -\mathcal{K}(h(x(t_k + \delta)). \end{aligned}$$

This shows for any $\delta \in [0, \Delta T]$, that we have $u_k \in \mathcal{C}(x(t_k + \delta))$ and therefore $x(t_k + \delta) \in \mathcal{S}$ as a consequence by the standard continuous time CBF arguments. ■

Proof of Lemma 3

Consider a loss function $\mathcal{L}(\lambda)$ (see (14)). Given a risk threshold $\alpha \in (0, 1)$ and confidence level $\gamma \in (0, 1)$, if we compute $\hat{\lambda}$ using non-exchangeable CRC to satisfy $\mathbb{E}[\mathcal{L}(\hat{\lambda})] \leq \alpha + \beta$ (where β accounts for non-exchangeability), then $\epsilon = \frac{\alpha+\beta}{\gamma}$ gives,

$$P(|\mathcal{B}_k(x_k, u_k) - \hat{\mathcal{B}}_k(x_k, u_k)| \leq \hat{\lambda}_k + \epsilon) \geq 1 - \gamma. \tag{A.1}$$

Proof: First, observe that by the definition of our loss function $L(\hat{\lambda})$ in equation (14):

$$L(\hat{\lambda}) = \max\left(0, |\mathcal{B}_k - \hat{\mathcal{B}}_k| - \hat{\lambda}\right) \tag{A.2}$$

For any $\epsilon > 0$, if the barrier prediction error exceeds $\hat{\lambda} + \epsilon$, then the loss must be greater than ϵ:

$$|\mathcal{B}_k - \hat{\mathcal{B}}_k| > \hat{\lambda} + \epsilon \implies L(\hat{\lambda}) > \epsilon \tag{A.3}$$

This implication allows us to bound the probability of large prediction errors:

$$P(|\mathcal{B}_k - \hat{\mathcal{B}}_k| > \hat{\lambda} + \epsilon) \leq P(L(\hat{\lambda}) > \epsilon) \tag{A.4}$$

By Markov's concentration inequality, for any non-negative random variable X and $a > 0$:

$$P(X > a) \leq \frac{\mathbb{E}[X]}{a} \tag{A.5}$$

Applying Markov's inequality to our loss function and using our non-exchangeable CRC guarantee that $\mathbb{E}[L(\hat{\lambda})] \leq \alpha + \beta$:

$$P(L(\hat{\lambda}) > \epsilon) \leq \frac{\mathbb{E}[L(\hat{\lambda})]}{\epsilon} \leq \frac{\alpha + \beta}{\epsilon} \tag{A.6}$$

Setting $\epsilon = \frac{\alpha+\beta}{\gamma}$, where α is the user-specified risk threshold, β is the total variation penalty term, and γ is the user-specified confidence level, we obtain:

$$P\left(L(\hat{\lambda}) > \frac{\alpha + \beta}{\gamma}\right) \leq \gamma \tag{A.7}$$

Taking the complement of this probability:

$$P\left(L(\hat{\lambda}) \leq \frac{\alpha + \beta}{\gamma}\right) \geq 1 - \gamma \tag{A.8}$$

Since $L(\hat{\lambda}) = \max\left(0, |\mathcal{B}_k - \hat{\mathcal{B}}_k| - \hat{\lambda}\right)$, we have:

$$P\left(|\mathcal{B}_k - \hat{\mathcal{B}}_k| - \hat{\lambda} \leq \frac{\alpha+\beta}{\gamma}\right) \geq 1-\gamma \tag{A.9}$$

Rearranging:

$$P\left(|\mathcal{B}_k - \hat{\mathcal{B}}_k| \leq \hat{\lambda} + \frac{\alpha+\beta}{\gamma}\right) \geq 1-\gamma \tag{A.10}$$

Thus, with $\epsilon = \frac{\alpha+\beta}{\gamma}$, we have proven that:

$$P(|\mathcal{B}_k - \hat{\mathcal{B}}_k| \leq \hat{\lambda} + \epsilon) \geq 1-\gamma \tag{A.11}$$

This completes the proof. ■

Proof of Theorem 1

Consider the human-robot system (7) with barrier certificates defined in (12). Given a confidence level $\gamma \in (0,1)$ and risk threshold $\alpha \in (0,1)$, if we have the non-exchangeable CRC guarantee such that $\hat{\lambda}$ satisfies $\mathbb{E}[\mathcal{L}(\hat{\lambda})] \leq \alpha + \beta$ and set $\epsilon = \frac{\alpha+\beta}{\gamma}$, then the prediction set defining the safe set of control inputs under uncertainty,

$$C_\lambda(x_k) = \{u_{\mathtt{R}} \in \mathcal{U}_{\mathtt{R}} \mid \hat{\mathcal{B}}_k(x_k, u_k) - (\hat{\lambda}_k + \epsilon) \geq 0\}, \tag{A.12}$$

ensures that $P(h(x_{k+1}) \geq 0) \geq 1-\gamma$ holds.

Proof: From Lemma 3, we know that with $\epsilon = \frac{\alpha+\beta}{\gamma}$:

$$P(|\mathcal{B}_k - \hat{\mathcal{B}}_k| \leq \hat{\lambda} + \epsilon) \geq 1-\gamma \tag{A.13}$$

For any robot control action $u_{\mathtt{R}} \in \mathcal{C}_{\hat{\lambda}}$, by definition of our prediction set in (13):

$$\hat{\mathcal{B}}_k - (\hat{\lambda} + \epsilon) \geq 0 \tag{A.14}$$

When the barrier prediction error is bounded (which occurs with probability at least $1-\gamma$), we have:

$$|\mathcal{B}_k - \hat{\mathcal{B}}_k| \leq \hat{\lambda} + \epsilon \tag{A.15}$$

This inequality can be written as a two-sided bound:

$$\hat{\lambda} + \epsilon \geq \mathcal{B}_k - \hat{\mathcal{B}}_k \geq -(\hat{\lambda} + \epsilon) \tag{A.16}$$

Rearranging inequalities:

$$\hat{\mathcal{B}}_k + (\hat{\lambda} + \epsilon) \geq \mathcal{B}_k \geq \hat{\mathcal{B}}_k - (\hat{\lambda} + \epsilon) \tag{A.17}$$

Combining with our prediction set constraint:

$$\mathcal{B}_k \geq \hat{\mathcal{B}}_k - (\hat{\lambda} + \epsilon) \geq 0 \tag{A.18}$$

By the properties of barrier certificates and Lemma 2, we know that:

$$\mathcal{B}_k \geq 0 \implies h(x_{\mathtt{R},k+1}, x_{\mathtt{H},k+1}) \geq 0 \tag{A.19}$$

Therefore, we have chain of probabilities:

$$P(h(x_{\mathtt{R},k+1}, x_{\mathtt{H},k+1}) \geq 0) \geq P(\mathcal{B}_k \geq 0) \tag{A.20}$$

$$\geq P(|\mathcal{B}_k - \hat{\mathcal{B}}_k| \leq \hat{\lambda} + \epsilon) \tag{A.21}$$

$$\geq 1-\gamma \tag{A.22}$$

This establishes our desired probabilistic safety guarantee:

$$P(h(x_{\mathtt{R},k+1}, x_{\mathtt{H},k+1}) \geq 0) \geq 1-\gamma \tag{A.23}$$

This completes the proof. ■