A Proofs

A.1 Stability verification

Lemma 1. *Using Assumptions 1 and 2, let $\mathcal{X}_\tau$ be a discretization of $\mathcal{X}$ such that $\|\mathbf{x} - [\mathbf{x}]_\tau\|_1 \leq \tau$ for all $\mathbf{x} \in \mathcal{X}$. Then, for all $\mathbf{x} \in \mathcal{X}$, we have with probability at least $1-\delta$ that*

$$\left|v(\mu_{n-1}([\mathbf{z}]_\tau)) - v([\mathbf{x}]_\tau) - \big(v(f(\mathbf{z})) - v(\mathbf{x})\big)\right| \leq L_v \beta_n \sigma_{n-1}([\mathbf{z}]_\tau) + (L_v L_f (L_\pi + 1) + L_v)\tau, \quad (8)$$

where $\mathbf{z} = (\mathbf{x}, \pi(\mathbf{x}))$ and $[\mathbf{z}]_\tau = ([\mathbf{x}]_\tau, \pi([\mathbf{x}]_\tau))$.

Proof. Let $\mathbf{z} = (\mathbf{x}, \pi(\mathbf{x}))$, $[\mathbf{z}]_\tau = ([\mathbf{x}]_\tau, \pi([\mathbf{x}]_\tau))$, and $\mu = \mu_{n-1}$, $\sigma = \sigma_{n-1}$. Then we have that

$$\begin{aligned}
&\left|v(\mu([\mathbf{z}]_\tau)) - v([\mathbf{x}]_\tau) - \big(v(f(\mathbf{z})) - v(\mathbf{x})\big)\right|,\\
&= \left|v(\mu([\mathbf{z}]_\tau)) - v([\mathbf{x}]_\tau) - v(f(\mathbf{z})) + v(\mathbf{x})\right|,\\
&= \left|v(\mu([\mathbf{z}]_\tau)) - v(f([\mathbf{z}]_\tau)) + v(f([\mathbf{z}]_\tau)) - v(f(\mathbf{z})) + v(\mathbf{x}) - v([\mathbf{x}]_\tau)\right|,\\
&\leq \left|v(\mu([\mathbf{z}]_\tau)) - v(f([\mathbf{z}]_\tau))\right| + \left|v(f([\mathbf{z}]_\tau)) - v(f(\mathbf{z}))\right| + \left|v(\mathbf{x}) - v([\mathbf{x}]_\tau)\right|,\\
&\leq L_v \|\mu([\mathbf{z}]_\tau) - f([\mathbf{z}]_\tau)\|_1 + L_v \|f([\mathbf{z}]_\tau) - f(\mathbf{z})\|_1 + L_v \|\mathbf{x} - [\mathbf{x}]_\tau\|_1,\\
&\leq L_v \beta_n \sigma([\mathbf{z}]_\tau) + L_v L_f \|[\mathbf{z}]_\tau - \mathbf{z}\|_1 + L_v \|\mathbf{x} - [\mathbf{x}]_\tau\|_1,
\end{aligned}$$

where the last three inequalities follow from Assumptions 1 and 2 to last inequality follows from Lemma 3. The result holds with probability at least $1-\delta$. By definition of the discretization and the policy class Π_L we have on each grid cell that

$$\begin{aligned}
\|\mathbf{z} - [\mathbf{z}]_\tau\|_1 &= \|\mathbf{x} - [\mathbf{x}]_\tau\|_1 + \|\pi(\mathbf{x}) - \pi([\mathbf{x}]_\tau)\|_1,\\
&\leq \tau + L_\pi \|\mathbf{x} - [\mathbf{x}]_\tau\|_1,\\
&\leq (L_\pi + 1)\tau,
\end{aligned}$$

where the equality in the first step follows from the definition of the 1-norm. Plugging this into the previous bound yields

$$\left|v(\mu([\mathbf{z}]_\tau)) - v([\mathbf{x}]_\tau) - \big(v(f(\mathbf{z})) - v(\mathbf{x})\big)\right| \leq L_v \beta_n \sigma([\mathbf{z}]_\tau) + \left(L_v L_f (1 + L_\pi) + L_v\right)\tau,$$

which completes the proof. □

Lemma 2. *$v(f(\mathbf{x}, \mathbf{u})) \in \mathcal{Q}_n$ holds for all $\mathbf{x} \in \mathcal{X}$, $\mathbf{u} \in \mathcal{U}$, and $n > 0$ with probability at least $(1-\delta)$.*

Proof. The proof is analogous to Lemma 1 and follows from Assumptions 1 and 2. □

Corollary 1. *$v(f(\mathbf{x}, \mathbf{u})) \in \mathcal{C}_n$ holds for all $\mathbf{x} \in \mathcal{X}$, $\mathbf{u} \in \mathcal{U}$, and $n > 0$ with probability at least $(1-\delta)$.*

Proof. Direct consequence of the fact that Lemma 2 holds jointly for all $n > 0$ with probability at least $1-\delta$. □

Lemma 1 show that the decrease on the Lyapunov function on the discrete grid $\mathcal{X}_\tau$ is close to that on the continuous domain $\mathcal{X}$. Given these confidence intervals, we can now establish the region of attraction using Theorem 1:

Theorem 2. *Under Assumptions 1 and 2 with $L_{\Delta v} := L_v L_f (L_\pi + 1) + L_v$, let $\mathcal{X}_\tau$ be a discretization of $\mathcal{X}$ such that $\|\mathbf{x} - [\mathbf{x}]_\tau\|_1 \leq \tau$ for all $\mathbf{x} \in \mathcal{X}$. If, for all $\mathbf{x} \in \mathcal{V}(c) \cap \mathcal{X}_\tau$ with $c > 0$, $\mathbf{u} = \pi(\mathbf{x})$, and for some $n \geq 0$ it holds that $u_n(\mathbf{x}, \mathbf{u}) < v(\mathbf{x}) - L_{\Delta v}\tau$, then $v(f(\mathbf{x}, \pi(\mathbf{x}))) < v(\mathbf{x})$ holds for all $\mathbf{x} \in \mathcal{V}(c)$ with probability at least $(1-\delta)$ and $\mathcal{V}(c)$ is a region of attraction for (1) under policy π.*

Proof. Using Lemma 1 it holds that $v(f(\mathbf{x}, \pi(\mathbf{x})) - v(\mathbf{x}) < 0$ for all continuous states $\mathbf{x} \in \mathcal{V}(c)$ with probability at least $1-\delta$, since all discrete states $\mathbf{x}_\tau \in \mathcal{V}(c) \cap \mathcal{X}$ fulfill Theorem 2. Thus we can use Theorem 1 to conclude that $\mathcal{V}(c)$ is a region of attraction for (1). □

Theorem 3. *Let $\mathcal{R}_{\pi_n}$ be the true region of attraction of* (1) *under the policy π_n. For any $\delta \in (0,1)$, we have with probability at least $(1-\delta)$ that $\mathcal{V}(c_n) \subseteq \mathcal{R}_{\pi_n}$ for all $n > 0$.*

Proof. Following the definition of $\mathcal{D}_n$ in (2), it is clear from the constraint in the optimization problem (3) that for all $\mathbf{x} \in \mathcal{D}_n$ it holds that $(\mathbf{x}, \pi_n(\mathbf{x})) \in \mathcal{D}_n$ or, equivalently that $u_n(\mathbf{x}, \pi(\mathbf{x})) - v(\mathbf{x}) < -L_{\Delta v}\tau$, see (2). The result $\mathcal{V}(c_n) \subseteq \mathcal{R}_{\pi_n}$ then follows from Theorem 2. □

Note that the initialization of the confidence intervals $\mathcal{Q}_0$ ensures that the decrease condition is always fulfilled for the initial policy.

A.2 Gaussian process model

One particular assumption that satisfies both the Lipschitz continuity and allows us to use GPs as a model of the dynamics is that the model errors $g(\mathbf{x}, \mathbf{u})$ live in some reproducing kernel Hilbert space (RKHS, [40]) corresponding to a differentiable kernel k and have RKHS norm smaller than B_g [35]. In our theoretical analysis, we use this assumption to prove exploration guarantees.

A $\mathcal{GP}(\mu(\mathbf{z}), k(\mathbf{z}, \mathbf{z}'))$ is a distribution over well-behaved, smooth functions $f\colon \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ (see Remark 1 for the vector-case, $\mathbb{R}^q$) that is parameterized by a mean function μ and a covariance function (kernel) k, which encodes assumptions about the functions [16]. In our case, the mean is given by the prior model h, while the kernel corresponds to the one in the RKHS. Given noisy measurements of the dynamics, $\hat{f}(\mathbf{z}) = f(\mathbf{z}) + \epsilon$ with $\mathbf{z} = (\mathbf{x}, \mathbf{u})$ at locations $A_n = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$, corrupted by independent, Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$ (we relax the Gaussian noise assumption in our analysis), the posterior is a GP distribution again with mean, $\mu_n(\mathbf{z}) = \mathbf{k}_n(\mathbf{z})^{\mathrm{T}}(\mathbf{K}_n + \sigma^2\mathbf{I})^{-1}\mathbf{y}_n$, covariance $k_n(\mathbf{z}, \mathbf{z}') = k(\mathbf{z}, \mathbf{z}') - \mathbf{k}_n(\mathbf{z})^{\mathrm{T}}(\mathbf{K}_n + \sigma^2\mathbf{I})^{-1}\mathbf{k}_n(\mathbf{z}')$, and variance $\sigma_n^2(\mathbf{z}) = k_n(\mathbf{z}, \mathbf{z})$. The vector $\mathbf{y}_n = [\hat{f}(\mathbf{z}_1) - h(\mathbf{z}_1), \dots, \hat{f}(\mathbf{z}_n) - h(\mathbf{z}_n)]^{\mathrm{T}}$ contains observed, noisy deviations from the mean, $\mathbf{k}_n(\mathbf{z}) = [k(\mathbf{z}, \mathbf{z}_1), \dots, k(\mathbf{z}, \mathbf{z}_n)]$ contains the covariances between the test input $\mathbf{z}$ and the data points in $\mathcal{D}_n$, $\mathbf{K}_n \in \mathbb{R}^{n \times n}$ has entries $[\mathbf{K}_n]_{(i,j)} = k(\mathbf{z}_i, \mathbf{z}_j)$, and $\mathbf{I}$ is the identity matrix.

Remark 1. *In the case of multiple output dimensions ($q > 1$), we consider a function with one-dimensional output $f'(\mathbf{x}, \mathbf{u}, i)\colon \mathcal{X} \times \mathcal{U} \times \mathcal{I} \to \mathbb{R}$, with the output dimension indexed by $i \in \mathcal{I} = \{1, \dots, q\}$. This allows us to use the standard definitions of the RKHS norm and GP model. In this case, we define the GP posterior distribution as $\mu_n(\mathbf{z}) = [\mu_n(\mathbf{z}, 1), \dots, \mu_n(\mathbf{z}, q)]^{\mathrm{T}}$ and $\sigma_n(\mathbf{z}) = \sum_{1 \leq i \leq q} \sigma_n(\mathbf{z}, i)$, where the unusual definition of the standard deviation is used in Lemma 3.*

Given the previous assumptions, it follows from [28, Lemma 2] that the dynamics in (1) are Lipschitz continuous with Lipschitz constant $L_f = L_h + L_g$, where L_g depends on the properties (smoothness) of the kernel.

Moreover, we can construct high-probability confidence intervals on the dynamics in (1) that fulfill Assumption 2 using the GP model.

Lemma 3. *([35, Theorem 6]) Assume σ-sub-Gaussian noise and that the model error $g(\cdot)$ in* (1) *has RKHS norm bounded by B_g. Choose $\beta_n = B_g + 4\sigma\sqrt{\gamma_n + 1 + \ln(1/\delta)}$. Then, with probability at least $1-\delta$, $\delta \in (0,1)$, for all $n \geq 1$, $\mathbf{x} \in \mathcal{X}$, and $\mathbf{u} \in \mathcal{U}$ it holds that $\|f(\mathbf{x}, \mathbf{u}) - \mu_{n-1}(\mathbf{x}, \mathbf{u})\|_1 \leq \beta_n \sigma_{n-1}(\mathbf{x}, \mathbf{u})$.*

Proof. From [34, Theorem 2] it follows that $|f(\mathbf{x}, \mathbf{u}, i) - \mu_{n-1}(\mathbf{x}, \mathbf{u}, i)| \leq \beta_n \sigma_n(\mathbf{x}, \mathbf{u}, i)$ holds with probability at least $1-\delta$ for all $1 \leq i \leq q$. Following Remark 1, we can model the multi-output function as a single-output function over an extended parameter space. Thus the result directly transfers by definition of the one norm and our definition of σ_n for multiple output dimensions in Remark 1. Note that by iteration n we have obtained nq measurements in the information capacity γ_n. □

That is, the true dynamics are contained within the GP posterior confidence intervals with high probability. The bound depends on the information capacity,

$$\gamma_n = \max_{A \subset \mathcal{X} \times \mathcal{U} \times \mathcal{I}\colon\, |A| = nq} \mathrm{I}(\mathbf{y}_A; \mathbf{f}_A), \tag{9}$$

which is the maximum mutual information that could be gained about the dynamics f from samples. The information capacity has a sublinear dependence on $n (\neq t)$ for many commonly used kernels such as the linear, squared exponential, and Matérn kernels and it can be efficiently and accurately approximated [35]. Note that we explicitly account for the q measurements that we get for each of the q states in (9).

Remark 2. *The GP model assumes Gaussian noise, while Lemma 3 considers σ-sub-Gaussian noise. Moreover, we consider functions with bounded RKHS norm, rather than samples from a GP. Lemma 3 thus states that even though we make different assumptions than the model, the confidence intervals are conservative enough to capture the true function with high probability.*

A.3 Safe exploration

Remark 3. *In the following we assume that $\mathcal{D}_n$ and $\mathcal{S}_n$ are defined as in* (4) *and* (5)*.*

Baseline As a baseline, we consider a class of algorithms that know about the Lipschitz continuity properties of v, f, and π. In addition, we can learn about $v(f(\mathbf{x},\mathbf{u}))$ up to some arbitrary statistical accuracy ϵ by visiting state $\mathbf{x}$ and obtaining a measurement for the next state after applying action $\mathbf{u}$, but face the safety restrictions defined in Sec. 2. Suppose we are given a set $\mathcal{S}$ of state-action pairs about which we can learn safely. Specifically, this means that we have a policy such that, for any state-action pair $(\mathbf{x},\mathbf{u})$ in $\mathcal{S}$, if we apply action $\mathbf{u}$ in state $\mathbf{x}$ and then apply actions according to the policy, the state converges to the origin. Such a set can be constructed using the initial policy π_0 from Sec. 2 as $\mathcal{S}_0 = \{(\mathbf{x},\pi_0(\mathbf{x})) \,|\, \mathbf{x} \in \mathcal{S}_0^x\}$.

The goal of the algorithm is to expand this set of states that we can learn about safely. Thus, we need to estimate the region of attraction by certifying that state-action pairs achieve the $-L_{\Delta v}\tau$ decrease condition in Theorem 2 by learning about state-action pairs in $\mathcal{S}$. We can then generalize the gained knowledge to unseen states by exploiting the Lipschitz continuity,

$$R^{\mathrm{dec}}(\mathcal{S}) = \mathcal{S}_0 \cup \big\{ \mathbf{z} \in \mathcal{X}_\tau \times \mathcal{U}_\tau \,|\, \exists (\mathbf{x},\mathbf{u}) \in \mathcal{S} \colon v(f(\mathbf{x},\mathbf{u})) - v(\mathbf{x}) + \epsilon + L_{\Delta v} \| \mathbf{z} - (\mathbf{x},\mathbf{u}) \|_1 < -L_{\Delta v}\tau \big\}, \tag{10}$$

where we use that we can learn $v(f(\mathbf{x},\mathbf{u}))$ up to ϵ accuracy within $\mathcal{S}$. We specifically include $\mathcal{S}_0$ in this set, to allow for initial policies that are safe, but does not meet the strict decrease requirements of Theorem 2. Given that all states in $R^{\mathrm{dec}}(\mathcal{S})$ fulfill the requirements of Theorem 2, we can estimate the corresponding region of attraction by committing to a control policy $\pi \in \Pi_L$ and estimating the largest safe level set of the Lyapunov function. With $\mathcal{D} = R^{\mathrm{dec}}(\mathcal{S})$, the operator

$$R^{\mathrm{lev}}(\mathcal{D}) = \mathcal{V}\big(\operatorname{argmax} c, \quad \text{such that } \exists \pi \in \Pi_L \colon \forall \mathbf{x} \in \mathcal{V}(c) \cap \mathcal{X}_\tau,\, (\mathbf{x},\pi(\mathbf{x})) \in \mathcal{D} \big) \tag{11}$$

encodes this operation. It optimizes over safe policies $\pi \in \Pi_L$ to determine the largest level set, such that all state-action pairs $(\mathbf{x},\pi(\mathbf{x}))$ at discrete states $\mathbf{x}$ in the level set $\mathcal{V}(c) \cap \mathcal{X}_\tau$ fulfill the decrease condition of Theorem 2. As a result, $R^{\mathrm{lev}}(R^{\mathrm{dec}}(\mathcal{S}))$ is an estimate of the largest region of attraction given the ϵ-accurate knowledge about state-action pairs in $\mathcal{S}$. Based on this increased region of attraction, there are more states that we can safely learn about. Specifically, we again use the Lipschitz constant and statistical accuracy ϵ to determine all states that map back into the region of attraction,

$$R_\epsilon(\mathcal{S}) = \mathcal{S} \cup \big\{ \mathbf{z}' \in R_\tau^{\mathrm{lev}}(R^{\mathrm{dec}}(\mathcal{S})) \times \mathcal{U}_\tau \,|\, \exists \mathbf{z} \in \mathcal{S} \colon v(f(\mathbf{z})) + \epsilon + L_v L_f \| \mathbf{z} - \mathbf{z}' \|_1 \leq \max_{\mathbf{x} \in R^{\mathrm{lev}}(R^{\mathrm{dec}}(\mathcal{S}))} v(\mathbf{x}) \big\}, \tag{12}$$

where $R_\tau^{\mathrm{lev}}(\mathcal{D}) = R^{\mathrm{lev}}(\mathcal{D}) \cap \mathcal{X}_\tau$. Thus, $R_\epsilon(\mathcal{S}) \supseteq \mathcal{S}$ contains state-action pairs that we can visit to learn about the system. Repeatedly applying this operator leads the largest set of state-action pairs that any safe algorithm with the same knowledge and restricted to policies in Π_L could hope to reach. Specifically, let $R_\epsilon^0(\mathcal{S}) = \mathcal{S}$ and $R_\epsilon^{i+1}(\mathcal{S}) = R_\epsilon(R_\epsilon^i(\mathcal{S}))$. Then $\overline{R}_\epsilon(\mathcal{S}) = \lim_{i \to \infty} R_\epsilon^i(\mathcal{S})$ is the set of all state-action pars on the discrete grid that any algorithm could hope to classify as safe without leaving this safe set. Moreover, $R^{\mathrm{lev}}(\overline{R}_\epsilon(\mathcal{S}))$ is the largest corresponding region of attraction that any algorithm can classify as safe for the given Lyapunov function.

Proofs In the following we implicitly assume that the assumptions of Lemma 3 hold and that β_n is defined as specified within Lemma 3. Moreover, for ease of notation we assume that $\mathcal{S}_0^x$ is a level set of the Lyapunov function $v(\cdot)$.

Lemma 4. $\mathcal{V}(c_n) = R^{\mathrm{lev}}(\mathcal{D}_n)$ *and* $c_n = \max_{\mathbf{x} \in R^{\mathrm{lev}}(\mathcal{D}_n)} v(\mathbf{x})$

Algorithm 2 Theoretical algorithm

Input: Initial safe policy $\mathcal{S}_0$, dynamics model $\mathcal{GP}(\mu(\mathbf{z}), k(\mathbf{z}, \mathbf{z}'))$

for all $n = 1, \dots$ **do**

$\quad \mathcal{D}_n = \bigcup_{(\mathbf{x},\mathbf{u}) \in \mathcal{S}_{n-1}} \left\{ \mathbf{z}' \in \mathcal{X}_\tau \times \mathcal{U}_\tau \,|\, u_n(\mathbf{x}, \mathbf{u}) - v(\mathbf{x}) + L_{\Delta v} \|\mathbf{z}' - (\mathbf{x}, \mathbf{u})\|_1 < -L_{\Delta v}\tau \right\}$,

$\quad \pi_n, c_n = \operatorname{argmax}_{\pi \in \Pi_L, c \in \mathbb{R}_{>0}} c,$ such that for all $\mathbf{x} \in \mathcal{V}(c) \cap \mathcal{X}_\tau \colon (\mathbf{x}, \pi(\mathbf{x})) \in \mathcal{D}_n$

$\quad \mathcal{S}_n = \bigcup_{\mathbf{z} \in \mathcal{S}_{n-1}} \left\{ \mathbf{z}' \in \mathcal{V}(c_n) \cap \mathcal{X}_\tau \times \mathcal{U}_\tau \,|\, u_n(\mathbf{z}) + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1 \leq c_n \right\}$

$\quad\quad = \bigcup_{\mathbf{z} \in \mathcal{S}_{n-1}} \left\{ \mathbf{z}' \in R_\tau^{\mathrm{lev}}(\mathcal{D}_n) \times \mathcal{U}_\tau \,|\, u_n(\mathbf{z}) + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1 \leq \max_{\mathbf{x} \in R^{\mathrm{lev}}(\mathcal{D}_n)} v(\mathbf{x}) \right\}$

$\quad (\mathbf{x}_n, \mathbf{u}_n) = \operatorname{argmax}_{(\mathbf{x},\mathbf{u}) \in \mathcal{S}_n} u_n(\mathbf{x}, \mathbf{u}) - l_n(\mathbf{x}, \mathbf{u})$

$\quad \mathcal{S}_n = \{ \mathbf{z} \in \mathcal{V}(c_n) \times \mathcal{U}_\tau \,|\, u_n(\mathbf{z}) \leq c_n \}$

$\quad$ Update GP with measurements $f(\mathbf{x}_n, \mathbf{u}_n) + \epsilon_n$

Proof. Directly by definition, compare (3) and (11). ☐

Remark 4. *Lemma 4 allows us to write the proofs entirely in terms of operators, rather than having to deal with explicit policies. In the following and in Algorithm 2 we replace $\mathcal{V}(c_n)$ and c_n according to Lemma 4. This moves the definitions closer to the baseline and makes for an easier comparison.*

We roughly follow the proof strategy in [18], but deal with the additional complexity of having safe sets that are defined in a more difficult way (indirectly through the policy). This is non-trivial and the safe sets are carefully designed in order to ensure that the algorithm works for general nonlinear systems.

We start by listing some fundamental properties of the sets that we defined below.

Lemma 5. *It holds for all $n \geq 1$ that*

(i) $\forall \mathbf{z} \in \mathcal{X}_\tau \times \mathcal{U}_\tau,\ u_{n+1}(\mathbf{z}) \leq u_n(\mathbf{z})$

(ii) $\forall \mathbf{z} \in \mathcal{X}_\tau \times \mathcal{U}_\tau,\ l_{n+1}(\mathbf{z}) \geq l_n(\mathbf{z})$

(iii) $\mathcal{S} \subseteq \mathcal{R} \implies R^{\mathrm{lev}}(\mathcal{S}) \subseteq R^{\mathrm{lev}}(\mathcal{R})$

(iv) $\mathcal{S} \subseteq \mathcal{R} \implies R^{\mathrm{dec}}(\mathcal{S}) \subseteq R^{\mathrm{dec}}(\mathcal{R})$

(v) $\mathcal{S} \subseteq \mathcal{R} \implies R_\epsilon(\mathcal{S}) \subseteq R_\epsilon(\mathcal{R})$

(vi) $\mathcal{S} \subseteq \mathcal{R} \implies \overline{R}_\epsilon(\mathcal{S}) \subseteq \overline{R}_\epsilon(\mathcal{R})$

(vii) $\mathcal{S}_n \supseteq \mathcal{S}_{n-1} \implies \mathcal{D}_{n+1} \supseteq \mathcal{D}_n$

(viii) $\mathcal{D}_1 \supseteq \mathcal{S}_0$

(ix) $\mathcal{S}_n \supseteq \mathcal{S}_{n-1}$

(x) $\mathcal{D}_n \supseteq \mathcal{D}_{n-1}$

Proof. (i) and (ii) follow directly form the definition of $\mathcal{C}_n$.

(iii) Let $\pi \in \Pi_L$ be a policy such that for some $c > 0$ it holds for all $\mathbf{x} \in \mathcal{V}(c) \cap \mathcal{X}_\tau$ that $(\mathbf{x}, \pi(\mathbf{x})) \in \mathcal{S}$. Then we have that $(\mathbf{x}, \pi(\mathbf{x})) \in \mathcal{R}$, since $\mathcal{S} \subseteq \mathcal{R}$. Thus it follows that with

$$c_s = \operatorname{argmax} c \quad \text{s.t. } \exists \pi \in \Pi_L \colon \forall \mathbf{x} \in \mathcal{V}(c) \cap \mathcal{X}_\tau,\ (\mathbf{x}, \pi(\mathbf{x})) \in \mathcal{S} \tag{13}$$

and

$$c_r = \operatorname{argmax} c \quad \text{s.t. } \exists \pi \in \Pi_L \colon \forall \mathbf{x} \in \mathcal{V}(c) \cap \mathcal{X}_\tau,\ (\mathbf{x}, \pi(\mathbf{x})) \in \mathcal{R} \tag{14}$$

we have that $c_r \geq c_s$. This implies $\mathcal{V}(c_r) \supseteq \mathcal{V}(c_s)$. The result follows.

(iv) Let $\mathbf{z} \in R^{\mathrm{dec}}(\mathcal{S})$. Then there exists $(\mathbf{x}, \mathbf{u}) \in \mathcal{S}$ such that $v(f(\mathbf{x}, \mathbf{u})) - v(\mathbf{x}) + \epsilon + L_{\Delta v} \|\mathbf{z} - (\mathbf{x}, \mathbf{u})\|_1 < -L_{\Delta v}\tau$. Since $\mathcal{S} \subseteq \mathcal{R}$ we have that $(\mathbf{x}, \mathbf{u}) \in \mathcal{R}$ as well and thus $\mathbf{z} \in R^{\mathrm{dec}}(\mathcal{R})$.

(v) $\mathcal{S} \subseteq \mathcal{R} \implies R^{\text{lev}}(R^{\text{dec}}(\mathcal{S})) \subseteq R^{\text{lev}}(R^{\text{dec}}(\mathcal{R}))$ due to (iii) and (iv). Since $\mathbf{z}' \in R_\epsilon(\mathcal{S})$, there must exist an $\mathbf{z} \in \mathcal{S}$ such that $v(f(\mathbf{z})) + \epsilon + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1 \leq \max_{\mathbf{x} \in R^{\text{lev}}(R^{\text{dec}}(\mathcal{S}))} v(\mathbf{x})$. Since $\mathcal{S} \subseteq \mathcal{R}$ it follows that $\mathbf{z} \in \mathcal{R}$. Moreover,

$$\max_{\mathbf{x} \in R^{\text{lev}}(R^{\text{dec}}(\mathcal{S}))} v(\mathbf{x}) \leq \max_{\mathbf{x} \in R^{\text{lev}}(R^{\text{dec}}(\mathcal{R}))} v(\mathbf{x}) \tag{15}$$

follows from $R^{\text{lev}}(R^{\text{dec}}(\mathcal{S})) \subseteq R^{\text{lev}}(R^{\text{dec}}(\mathcal{R}))$, so that we conclude that $\mathbf{z}' \in R_\epsilon(\mathcal{R})$.

(vi) This follows directly by repeatedly applying the result of (v).

(vii) Let $\mathbf{z}' \in \mathcal{D}_n$. Then $\exists (\mathbf{x}, \mathbf{u}) \in \mathcal{S}_{n-1} \colon u_n(\mathbf{x}, \mathbf{u}) - v(\mathbf{x}) + L_{\Delta v} \|\mathbf{z}' - (\mathbf{x}, \mathbf{u})\|_1 < -L_{\Delta v} \tau$. Since $\mathcal{S}_n \supseteq \mathcal{S}_{n-1}$ it follows that $(\mathbf{x}, \mathbf{u}) \in \mathcal{S}_n$ as well. Moreover, we have

$$\begin{aligned} &u_{n+1}(\mathbf{x}, \mathbf{u}) - v(\mathbf{x}) + L_{\Delta v} \|\mathbf{z}' - (\mathbf{x}, \mathbf{u})\|_1 \\ &\leq u_n(\mathbf{x}, \mathbf{u}) - v(\mathbf{x}) + L_{\Delta v} \|\mathbf{z}' - (\mathbf{x}, \mathbf{u})\|_1 < -L_{\Delta v} \tau \end{aligned}$$

since u_{n+1} is non-increasing, see (i). Thus $\mathbf{z}' \in \mathcal{D}_{n+1}$.

(viii) By definition of $\mathcal{C}_0$ we have for all $(\mathbf{x}, \mathbf{u}) \in \mathcal{S}_0$ that $u_0(\mathbf{x}, \mathbf{u}) < v(\mathbf{x}) - L_{\Delta v} \tau$. Now we have that

$$\begin{aligned} &u_1(\mathbf{x}, \mathbf{u}) - v(\mathbf{x}) + L_{\Delta v} \|(\mathbf{x}, \mathbf{u}) - (\mathbf{x}, \mathbf{u})\|_1, \\ &= u_1(\mathbf{x}, \mathbf{u}) - v(\mathbf{x}), \\ &\leq u_0(\mathbf{x}, \mathbf{u}) - v(\mathbf{x}), \quad \text{by Lemma 5 (i)} \\ &< \, -L_{\Delta v} \tau, \end{aligned}$$

which implies that $(\mathbf{x}, \mathbf{u}) \in \mathcal{D}_1$.

(ix) Proof by induction. We consider the base case, $\mathbf{z} \in \mathcal{S}_0$, which implies that $\mathbf{z} \in \mathcal{D}_1$ by (viii). Moreover, since $\mathcal{S}_0^x$ is a level set of the Lyapunov function v by assumption, we have that $R^{\text{lev}}(\mathcal{S}_0) = \mathcal{S}_0^x$. The previous statements together with (iii) imply that $\mathbf{z} \in R_\tau^{\text{lev}}(\mathcal{D}_1) \times \mathcal{U}_\tau$, since $\mathcal{D}_1 \supseteq \mathcal{S}_0$ by (viii). Now, we have that

$$u_1(\mathbf{z}) + L_v L_f \|\mathbf{z} - \mathbf{z}\|_1 = u_1(\mathbf{z}) \overset{(i)}{\leq} u_0(\mathbf{z}).$$

Moreover, by definition of $\mathcal{C}_0$, we have for all $(\mathbf{x}, \mathbf{u}) \in \mathcal{S}_0$ that

$$u_0(\mathbf{x}, \mathbf{u}) < v(\mathbf{x}) - L_{\Delta v} \tau < v(\mathbf{x}).$$

As a consequence,

$$\begin{aligned} u_0(\mathbf{x}, \mathbf{u}) &\leq \max_{(\mathbf{x}, \mathbf{u}) \in \mathcal{S}_0} v(\mathbf{x}), &\quad (16) \\ &= \max_{\mathbf{x} \in R^{\text{lev}}(\mathcal{S}_0)} v(\mathbf{x}), &\quad (17) \\ &\leq \max_{\mathbf{x} \in R^{\text{lev}}(\mathcal{D}_1)} v(\mathbf{x}), &\quad (18) \end{aligned}$$

where the last inequality follows from (iii) and (viii). Thus we have $\mathbf{z} \in \mathcal{S}_1$.

For the induction step, assume that for $n \geq 2$ we have $\mathbf{z}' \in \mathcal{S}_n$ with $\mathcal{S}_n \supseteq \mathcal{S}_{n-1}$. Now since $\mathbf{z}' \in \mathcal{S}_n$ we must have that $\mathbf{z}' \in R_\tau^{\text{lev}}(\mathcal{D}_n) \times \mathcal{U}_\tau$. This implies that $\mathbf{z}' \in R_\tau^{\text{lev}}(\mathcal{D}_{n+1}) \times \mathcal{U}_\tau$, due to Lemma 5 (iii) and (vii) together with the induction assumption of $\mathcal{S}_n \supseteq \mathcal{S}_{n-1}$. Moreover, there must exist a $\mathbf{z} \in \mathcal{S}_{n-1} \subseteq \mathcal{S}_n$ such that

$$\begin{aligned} u_{n+1}(\mathbf{z}) + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1, &\leq u_n(\mathbf{z}) + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1, &\quad (19) \\ &\leq \max_{\mathbf{x} \in R^{\text{lev}}(\mathcal{D}_n)} v(\mathbf{x}), &\quad (20) \\ &\leq \max_{\mathbf{x} \in R^{\text{lev}}(\mathcal{D}_{n+1})} v(\mathbf{x}), &\quad (21) \end{aligned}$$

which in turn implies $\mathbf{z} \in \mathcal{S}_{n+1}$. The last inequality follows from Lemma 5 (iii) and (vii) together with the induction assumption that $\mathcal{S}_n \supseteq \mathcal{S}_{n-1}$.

(x) This is a direct consequence of (vii), (viii), and (ix).

□

Given these set properties, we first consider what happens if the safe set $\mathcal{S}_n$ does not expand after collecting data points. We use these results later to conclude that the safe set must either expand or that the maximum level set is reached. We denote by

$$\mathbf{z}_n = (\mathbf{x}_n, \mathbf{u}_n)$$

the data point the is sampled according to (6).

Lemma 6. *For any $n_1 \geq n_0 \geq 1$, if $\mathcal{S}_{n_1} = \mathcal{S}_{n_0}$, then for any n such that $n_0 \leq n < n_1$, it holds that*

$$2\beta_n \sigma_n(\mathbf{z}_n) \leq \sqrt{\frac{C_1 q \beta_n^2 \gamma_n}{n - n_0}}, \tag{22}$$

where $C_1 = 8/\log(1 + \sigma^{-2})$.

Proof. We modify the results for $q = 1$ by [35] to this lemma, but use the different definition for β_n from [34]. Even though the goal of [35, Lemma 5.4] is different from ours, we can still apply their reasoning to bound the amplitude of the confidence interval of the dynamics. In particular, in [35, Lemma 5.4], we have $r_n = 2\beta_n \sigma_{n-1}(\mathbf{z}_n)$ with $\mathbf{z}_n = (\mathbf{x}_n, \mathbf{u}_n)$ according to Lemma 3. Then

$$\begin{aligned}
r_n^2 &= 4\beta_n^2 \sigma_{n-1}^2(\mathbf{z}_n), && (23) \\
&= 4\beta_n^2 \left(\sum_{i=1}^{q} \sigma_{n-1}(\mathbf{z}_n, i) \right)^2, && (24) \\
&\leq 4\beta_n^2 q \sum_{i=1}^{q} \sigma_{n-1}^2(\mathbf{z}_n, i) \qquad \text{(Jensen's ineq.)}, && (25) \\
&\leq 4\beta_n^2 q \sigma^2 C_2 \sum_{i=1}^{q} \log(1 + \sigma^{-2} \sigma_{n-1}^2(\mathbf{z}_n, i)), && (26)
\end{aligned}$$

where $C_2 = \sigma^{-2}/\log(1 + \sigma^{-2})$. The result then follows analogously to [35, Lemma 5.4] by noting that

$$\sum_{j=1}^{n} r_j^2 \leq C_1 \beta_n^2 q \gamma_n \qquad \forall n \geq 1 \tag{27}$$

according to the definition of γ_n in this paper and using the Cauchy-Schwarz inequality. □

The previous result allows us to bound the width of the confidence intervals:

Corollary 2. *For any $n_1 \geq n_0 \geq 1$, if $\mathcal{S}_{n_1} = \mathcal{S}_{n_0}$, then for any n such that $n_0 \leq n < n_1$, it holds that*

$$u_n(\mathbf{z}_n) - l_n(\mathbf{z}_n) \leq L_v \sqrt{\frac{C_1 q \beta_n^2 \gamma_n}{n - n_0}}, \tag{28}$$

where $C_1 = 8/\log(1 + \sigma^{-2})$.

Proof. Direct consequence of Lemma 6 together with the definition of $\mathcal{C}$ and $\mathcal{Q}$. □

Corollary 3. *For any $n \geq 1$ with C_1 as defined in Lemma 6, let N_n be the smallest integer satisfying $\frac{N_n}{\beta_{n+N_n}^2 \gamma n + N_n} \geq \frac{C_1 L_v^2 q}{\epsilon^2}$ and $\mathcal{S}_{n+N_n} = \mathcal{S}_{N_n}$, then, for any $\mathbf{z} \in \mathcal{S}_{n+N_n}$ it holds that*

$$u_n(\mathbf{z}) - l_n(\mathbf{z}) \leq \epsilon. \tag{29}$$

Proof. The result trivially follows from substituting N_n in the bound in Corollary 2. □

Lemma 7. *For any $n \geq 1$, if $\overline{R}_\epsilon(\mathcal{S}_0) \setminus \mathcal{S}_n \neq \emptyset$, then $R_\epsilon(\mathcal{S}_n) \setminus \mathcal{S}_n \neq \emptyset$.*

Proof. As in [18, Lemma 6]. Assume, to the contrary, that $R_\epsilon(\mathcal{S}_n) \setminus \mathcal{S}_n = \emptyset$. By definition $R_\epsilon(\mathcal{S}_n) \supseteq \mathcal{S}_n$, therefore $R_\epsilon(\mathcal{S}_n) = \mathcal{S}_n$. Iteratively applying R_ϵ to both sides, we get in the limit $\overline{R}_\epsilon(\mathcal{S}_n) = \mathcal{S}_n$. But then, by Lemma 5,(vi) and (ix), we get

$$\overline{R}_\epsilon(\mathcal{S}_0) \subseteq \overline{R}_\epsilon(\mathcal{S}_n) = \mathcal{S}_n, \tag{30}$$

which contradicts the assumption that $\overline{R}_\epsilon(\mathcal{S}_0) \setminus \mathcal{S}_n \neq \emptyset$. □

Lemma 8. *For any $n \geq 1$, if $\overline{R}_\epsilon(\mathcal{S}_0) \setminus \mathcal{S}_n \neq \emptyset$, then the following holds with probability at least $1-\delta$:*

$$\mathcal{S}_{n+N_n} \supset \mathcal{S}_n. \tag{31}$$

Proof. By Lemma 7, we have that $R_\epsilon(\mathcal{S}_n) \setminus \mathcal{S}_n \neq \emptyset$. By definition, this means that there exist $\mathbf{z} \in R_\epsilon(\mathcal{S}_n) \setminus \mathcal{S}_n$ and $\mathbf{z}' \in \mathcal{S}_n$ such that

$$v(f(\mathbf{z}')) + \epsilon + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1 \leq \max_{\mathbf{x} \in R^{\text{lev}}(R^{\text{dec}}(\mathcal{S}_n))} v(\mathbf{x}) \tag{32}$$

Now we assume, to the contrary, that $\mathcal{S}_{n+N_n} = \mathcal{S}_n$ (the safe set cannot decrease due to Lemma 5 (ix)). This implies that $\mathbf{z} \in \mathcal{X}_\tau \times \mathcal{U}_\tau \setminus \mathcal{S}_{n+N_n}$ and $\mathbf{z}' \in \mathcal{S}_{n+N_n} = \mathcal{S}_{n+N_n-1}$. Due to Corollary 2, it follows that

$$\begin{aligned}
& u_{n+N_n}(\mathbf{z}') + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1 && (33)\\
\leq\ & v(f(\mathbf{z}')) + \epsilon + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1 && (34)\\
\leq\ & \max_{\mathbf{x} \in R^{\text{lev}}(R^{\text{dec}}(\mathcal{S}_n))} v(\mathbf{x}) \qquad \text{by (32)} && (35)\\
=\ & \max_{\mathbf{x} \in R^{\text{lev}}(R^{\text{dec}}(\mathcal{S}_{n+N_n}))} v(\mathbf{x}) \qquad \text{by (iii), (iv) and (ix)} && (36)
\end{aligned}$$

Thus, to conclude that $\mathbf{z} \in \mathcal{S}_{n+N_n}$ according to (5), we need to show that $R^{\text{lev}}(\mathcal{D}_{n+N_n}) \supseteq R^{\text{lev}}(R^{\text{dec}}(\mathcal{S}_n))$. To this end, we use Lemma 5 (iii) and show that $\mathcal{D}_{n+N_n} \supseteq R^{\text{dec}}(\mathcal{S}_{n+N_n})$. Consider $(\mathbf{x}, \mathbf{u}) \in R^{\text{dec}}(\mathcal{S}_{n+N_n})$, we know that there exists a $(\mathbf{x}', \mathbf{u}') \in \mathcal{S}_{n+N_n} = \mathcal{S}_{n+N_n-1}$ such that

$$\begin{aligned}
-L_{\Delta v}\tau &> v(f(\mathbf{x}', \mathbf{u}')) - v(\mathbf{x}') + \epsilon + L_{\Delta v}\|(\mathbf{x}, \mathbf{u}) - (\mathbf{x}', \mathbf{u}')\|_1, && (37)\\
&\geq u_{n+N_n}(\mathbf{x}', \mathbf{u}') - v(\mathbf{x}') + L_{\Delta v}\|(\mathbf{x}, \mathbf{u}) - (\mathbf{x}', \mathbf{u}')\|_1, && (38)
\end{aligned}$$

where the second inequality follows from Corollary 2. This implies that $(\mathbf{x}, \mathbf{u}) \in \mathcal{D}_n$ and thus $\mathcal{D}_{n+N_n} \supseteq R^{\text{dec}}(\mathcal{S}_{n+N_n})$. This, in turn, implies that $\mathbf{z} \in \mathcal{S}_{n+N_n}$, which is a contradiction. □

Lemma 9. *For any $n \geq 0$, the following holds with probability at least $1-\delta$:*

$$\mathcal{S}_n \subseteq \overline{R}_0(\mathcal{S}_0). \tag{39}$$

Proof. Proof by induction. For the base case, $n = 0$, we have $\mathcal{S}_0 \subseteq \overline{R}_0(\mathcal{S}_0)$ by definition.

For the induction step, assume that for some $n \geq 1$, $\mathcal{S}_{n-1} \subseteq \overline{R}_0(\mathcal{S}_0)$. Let $\mathbf{z} \in \mathcal{S}_n$. Then, by definition, $\exists \mathbf{z}' \in \mathcal{S}_{n-1}$ such that

$$u_n(\mathbf{z}') + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1 \leq \max_{\mathbf{x} \in R^{\text{lev}}(\mathcal{D}_n)} v(\mathbf{x}), \tag{40}$$

which, by Corollary 1, implies that

$$v(f(\mathbf{z}')) + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1 \leq \max_{\mathbf{x} \in R^{\text{lev}}(\mathcal{D}_n)} v(\mathbf{x}) \tag{41}$$

Now since $\mathbf{z}' \in \overline{R}_0(\mathcal{S}_0)$ by the induction hypothesis, in order to conclude that $\mathbf{z} \in \overline{R}_0(\mathcal{S}_0)$ we need to show that $R^{\text{lev}}(\mathcal{D}_n) \subseteq R^{\text{lev}}(R^{\text{dec}}(\overline{R}(\mathcal{S}_0)))$.

Let $(\mathbf{x}, \mathbf{u}) \in \mathcal{D}_n$, then there exist $(\mathbf{x}', \mathbf{z}') \in \mathcal{S}_{n-1}$ such that

$$u_{n-1}(\mathbf{x}', \mathbf{u}') - v(\mathbf{x}') + L_{\Delta v}\|(\mathbf{x}, \mathbf{u}) - (\mathbf{x}', \mathbf{u}')\|_1 < -L_{\Delta v}\tau, \tag{42}$$

which, by Corollary 1, implies that

$$v(f(\mathbf{x}', \mathbf{u}')) - v(\mathbf{x}') + L_{\Delta v} \|(\mathbf{x}, \mathbf{u}) - (\mathbf{x}', \mathbf{u}')\|_1 < -L_{\Delta v} \tau, \tag{43}$$

which means that $(\mathbf{x}, \mathbf{u}) \in R^{\text{dec}}(\overline{R}_0(\mathcal{S}_0))$ since $\mathcal{S}_{n-1} \subseteq \overline{R}_0(\mathcal{S}_0)$ and therefore $(\mathbf{x}', \mathbf{u}') \in \overline{R}_0(\mathcal{S}_0)$ holds by the induction hypothesis. We use (iii) to conclude that $R^{\text{lev}}(\mathcal{D}_n) \subseteq R^{\text{lev}}(R^{\text{dec}}(\overline{R}(\mathcal{S}_0)))$, which concludes the proof. □

Lemma 10. *Let n^* be the smallest integer, such that $n^* \geq |\overline{R}_0(\mathcal{S}_0)| N_{n^*}$. Then, there exists $n_0 \leq n^*$ such that $\mathcal{S}_{n_0 + N_{n_0}} = \mathcal{S}_{n_0}$ holds with probability $1 - \delta$.*

Proof. By contradiction. Assume, to the contrary, that for all $n \leq n^*$, $\mathcal{S}_n \subset \mathcal{S}_{n+N_n}$. From Lemma 5 (ix) we know that $\mathcal{S}_n \subseteq \mathcal{S}_{n+N_n}$. Since N_n is increasing in n, we have that $N_n \leq N_{n^*}$. Thus, we must have

$$\mathcal{S}_0 \subset \mathcal{S}_{N_{n^*}} \subset \mathcal{S}_{2N_{n^*}} \cdots, \tag{44}$$

so that for any $0 \leq j \leq |\overline{R}_0(\mathcal{S}_0)|$, it holds that $|S_{jT_{n^*}}| > j$. In particular, for $j = |\overline{R}_0(\mathcal{S}_0)|$, we get

$$|\mathcal{S}_{jN_{n^*}}| > |\overline{R}_0(\mathcal{S}_0)|, \tag{45}$$

which contradicts $\mathcal{S}_{jN_{n^*}} \subseteq \overline{R}_0(\mathcal{S}_0)$ from Lemma 9. □

Corollary 4. *Let n^* be the smallest integer such that*

$$\frac{n^*}{\beta_{n^*} \gamma_{n^*}} \geq \frac{C_1 L_v^2 q |\overline{R}_0(\mathcal{S}_0)|}{\epsilon^2}, \tag{46}$$

then there exists a $n_0 \leq n^*$ such that $\mathcal{S}_{n_0 + N_{n_0}} = \mathcal{S}_{n_0}$.

Proof. A direct consequence of Lemma 10 and Corollary 3. □

A.4 Safety and policy adaptation

In the following, we denote the true region of attraction of (1) under a policy π by $\mathcal{R}_\pi$.

Lemma 11. $R^{\text{lev}}(\mathcal{D}_n) \subseteq \mathcal{R}_{\pi_n}$ *for all $n \geq 0$.*

Proof. By definition, we have for all $(\mathbf{x}, \mathbf{u}) \in \mathcal{D}_n$ that the exists $(\mathbf{x}', \mathbf{u}') \in \mathcal{S}_{n-1}$ such that

$$\begin{aligned} -L_{\Delta v} \tau &\geq u_n(\mathbf{x}', \mathbf{u}') - v(\mathbf{x}') + L_{\Delta v} \|(\mathbf{x}, \mathbf{u}) - (\mathbf{x}', \mathbf{u}')\|_1, \\ &\geq v(f(\mathbf{x}', \mathbf{u}')) - v(\mathbf{x}') + L_{\Delta v} \|(\mathbf{x}, \mathbf{u}) - (\mathbf{x}', \mathbf{u}')\|_1, \\ &\geq v(f(\mathbf{x}, \mathbf{u})) - v(\mathbf{x}), \end{aligned}$$

where the first inequality follows from Corollary 1 and the second one by Lipschitz continuity, see Lemma 1.

By definition of R^{lev} in (11), it follows that for all $\mathbf{x} \in R^{\text{lev}}(\mathcal{D}_n) \cap \mathcal{X}_\tau$ we have that $(\mathbf{x}, \pi_n(\mathbf{x})) \in \mathcal{D}_n$. Moreover, $R^{\text{lev}}(\mathcal{D}_n)$ is a level set of the Lyapunov function by definition. Thus the result follows from Theorem 2. □

Lemma 12. $f(\mathbf{x}, \mathbf{u}) \in \mathcal{R}_{\pi_n}$ $\forall (\mathbf{x}, \mathbf{u}) \in \mathcal{S}_n$.

Proof. This holds for $\mathcal{S}_0$ by definition. For $n \geq 1$, by defition, we have for all $\mathbf{z} \in \mathcal{S}_n$ there exists an $\mathbf{z}' \in \mathcal{S}_{n-1}$ such that

$$\begin{aligned} \max_{\mathbf{x} \in R^{\text{lev}}(\mathcal{D}_n)} v(\mathbf{x}) &\geq u_n(\mathbf{z}') + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1 \\ &\geq v(f(\mathbf{z}')) + L_v L_f \|\mathbf{z} - \mathbf{z}'\|_1 \\ &\geq v(f(\mathbf{z})) \end{aligned}$$

where the first inequality follows from Corollary 1 and the second one by Lipschitz continuity, see Lemma 1. Since $R^{\text{lev}}(\mathcal{D}_n) \subseteq \mathcal{R}_{\pi_n}$ by Lemma 11, we have that $f(\mathbf{z}) \in \mathcal{R}_{\pi_n}$. □

Theorem 4. *Assume σ-sub-Gaussian measurement noise and that the model error $g(\cdot)$ in* (1) *has RKHS norm smaller than B_g. Under the assumptions of Theorem 2, with $\beta_n = B_g + 4\sigma\sqrt{\gamma_n + 1 + \ln(1/\delta)}$, and with measurements collected according to* (6)*, let n^* be the smallest positive integer so that $\frac{n^*}{\beta_{n^*}^2 \gamma_{n^*}} \geq \frac{Cq(|\overline{R}(\mathcal{S}_0)|+1)}{L_v^2 \epsilon^2}$ where $C = 8/\log(1+\sigma^{-2})$. Let $\mathcal{R}_\pi$ be the true region of attraction of* (1) *under a policy π. For any $\epsilon > 0$, and $\delta \in (0,1)$, the following holds jointly with probability at least $(1-\delta)$ for all $n > 0$:*

(i) $\mathcal{V}(c_n) \subseteq \mathcal{R}_{\pi_n}$ *(ii)* $f(\mathbf{x}, \mathbf{u}) \in \mathcal{R}_{\pi_n}\ \forall (\mathbf{x}, \mathbf{u}) \in \mathcal{S}_n$. *(iii)* $\overline{R}_\epsilon(\mathcal{S}_0) \subseteq \mathcal{S}_n \subseteq \overline{R}_0(\mathcal{S}_0)$.

Proof. See Lemmas 11 and 12 for (i) and (ii), respectively. Part (iii) is a direct consequence of Corollary 4 and Lemma 9. □