

---

# Supplementary Material

---

**Armin Lederer**  
 Technical University of Munich  
 armin.lederer@tum.de

**Jonas Umlauf**  
 Technical University of Munich  
 jonas.umlauft@tum.de

**Sandra Hirche**  
 Technical University of Munich  
 hirche@tum.de

## 1 Proof of Theorem 3.1

*Proof of Theorem 3.1.* We first prove the Lipschitz constant of the posterior mean  $\nu_N(\mathbf{x})$  and the modulus of continuity of the standard deviation  $\sigma_N(\mathbf{x})$ , before we derive the bound of the regression error. The norm of the difference between the posterior mean  $\nu_N(\mathbf{x})$  evaluated at two different points is given by

$$\|\nu_N(\mathbf{x}) - \nu_N(\mathbf{x}')\| = \|(\mathbf{k}(\mathbf{x}, \mathbf{X}_N) - \mathbf{k}(\mathbf{x}', \mathbf{X}_N)) \boldsymbol{\alpha}\|$$

with

$$\boldsymbol{\alpha} = (\mathbf{K}(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 \mathbf{I}_N)^{-1} \mathbf{y}_N. \quad (1)$$

Due to the Cauchy-Schwarz inequality and the Lipschitz continuity of the kernel we obtain

$$\|\nu_N(\mathbf{x}) - \nu_N(\mathbf{x}')\| \leq L_k \sqrt{N} \|\boldsymbol{\alpha}\| \|\mathbf{x} - \mathbf{x}'\|,$$

which proves Lipschitz continuity of the mean  $\nu_N(\mathbf{x})$ . In order to calculate a modulus of continuity for the posterior standard deviation  $\sigma_N(\mathbf{x})$  observe that the difference of the variance at two points  $\mathbf{x}, \mathbf{x}' \in \mathbb{X}$  can be expressed as

$$|\sigma_N^2(\mathbf{x}) - \sigma_N^2(\mathbf{x}')| = |\sigma_N(\mathbf{x}) - \sigma_N(\mathbf{x}')| |\sigma_N(\mathbf{x}) + \sigma_N(\mathbf{x}')|. \quad (2)$$

Since the standard deviation is positive semidefinite we have

$$|\sigma_N(\mathbf{x}) + \sigma_N(\mathbf{x}')| \geq |\sigma_N(\mathbf{x}) - \sigma_N(\mathbf{x}')| \quad (3)$$

and hence, we obtain

$$|\sigma_N^2(\mathbf{x}) - \sigma_N^2(\mathbf{x}')| \geq |\sigma_N(\mathbf{x}) - \sigma_N(\mathbf{x}')|^2. \quad (4)$$

Therefore, it is sufficient to bound the difference of the variance at two points  $\mathbf{x}, \mathbf{x}' \in \mathbb{X}$  and take the square root of the resulting expression. Due to the Cauchy-Schwarz inequality and Lipschitz continuity of  $k(\cdot, \cdot)$  the absolute value of the difference of the variance can be bounded by

$$\begin{aligned} |\sigma_N^2(\mathbf{x}) - \sigma_N^2(\mathbf{x}')| &\leq 2L_k \|\mathbf{x} - \mathbf{x}'\| \\ &+ \|\mathbf{k}(\mathbf{x}, \mathbf{X}_N) - \mathbf{k}(\mathbf{x}', \mathbf{X}_N)\| \|(\mathbf{K}(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 \mathbf{I}_N)^{-1}\| \|\mathbf{k}(\mathbf{X}_N, \mathbf{x}) + \mathbf{k}(\mathbf{X}_N, \mathbf{x}')\|. \end{aligned} \quad (5)$$

On the one hand, we have

$$\|\mathbf{k}(\mathbf{x}, \mathbf{X}_N) - \mathbf{k}(\mathbf{x}', \mathbf{X}_N)\| \leq \sqrt{N} L_k \|\mathbf{x} - \mathbf{x}'\| \quad (6)$$

due to Lipschitz continuity of  $k(\mathbf{x}, \mathbf{x}')$ . On the other hand we have

$$\|\mathbf{k}(\mathbf{x}, \mathbf{X}_N) + \mathbf{k}(\mathbf{x}', \mathbf{X}_N)\| \leq 2\sqrt{N} \max_{\mathbf{x}, \mathbf{x}' \in \mathbb{X}} k(\mathbf{x}, \mathbf{x}'). \quad (7)$$

The modulus of continuity  $\omega_{\sigma_N}(\tau)$  follows from substituting (6) and (7) in (5) and taking the square root of the resulting expression. Finally, we prove the probabilistic uniform error bound by exploiting the fact that for every grid  $\mathbb{X}_\tau$  with  $|\mathbb{X}_\tau|$  grid points and

$$\max_{\mathbf{x} \in \mathbb{X}} \min_{\mathbf{x}' \in \mathbb{X}_\tau} \|\mathbf{x} - \mathbf{x}'\| \leq \tau \quad (8)$$

it holds with probability of at least  $1 - |\mathbb{X}_\tau|e^{-\beta(\tau)/2}$  that [1]

$$|f(\mathbf{x}) - \nu_N(\mathbf{x})| \leq \sqrt{\beta(\tau)}\sigma_N(\mathbf{x}) \quad \forall \mathbf{x} \in \mathbb{X}_\tau. \quad (9)$$

Choose  $\beta(\tau) = 2 \log \left( \frac{|\mathbb{X}_\tau|}{\delta} \right)$ , then

$$|f(\mathbf{x}) - \nu_N(\mathbf{x})| \leq \sqrt{\beta(\tau)}\sigma_N(\mathbf{x}) \quad \forall \mathbf{x} \in \mathbb{X}_\tau \quad (10)$$

holds with probability of at least  $1 - \delta$ . Due to continuity of  $f(\mathbf{x})$ ,  $\nu_N(\mathbf{x})$  and  $\sigma_N(\mathbf{x})$  we obtain

$$\min_{\mathbf{x}' \in \mathbb{X}_\tau} |f(\mathbf{x}) - f(\mathbf{x}')| \leq \tau L_f \quad \forall \mathbf{x} \in \mathbb{X} \quad (11)$$

$$\min_{\mathbf{x}' \in \mathbb{X}_\tau} |\nu_N(\mathbf{x}) - \nu_N(\mathbf{x}')| \leq \tau L_{\nu_N} \quad \forall \mathbf{x} \in \mathbb{X} \quad (12)$$

$$\min_{\mathbf{x}' \in \mathbb{X}_\tau} |\sigma_N(\mathbf{x}) - \sigma_N(\mathbf{x}')| \leq \omega_{\sigma_N}(\tau) \quad \forall \mathbf{x} \in \mathbb{X}. \quad (13)$$

Moreover, the minimum number of grid points satisfying (8) is given by the covering number  $M(\tau, \mathbb{X})$ . Hence, we obtain

$$P \left( |f(\mathbf{x}) - \nu_N(\mathbf{x})| \leq \sqrt{\beta(\tau)}\sigma_N(\mathbf{x}) + \gamma(\tau), \forall \mathbf{x} \in \mathbb{X} \right) \geq 1 - \delta, \quad (14)$$

where

$$\beta(\tau) = 2 \log \left( \frac{M(\tau, \mathbb{X})}{\delta} \right) \quad (15)$$

$$\gamma(\tau) = (L_f + L_{\nu_N})\tau + \sqrt{\beta(\tau)}\omega_{\sigma_N}(\tau). \quad (16)$$

□

## 2 Proof of Theorem 3.2

In order to proof Theorem 3.2, several auxiliary results are necessary, which are derived in the following. The first lemma concerns the expected supremum of a Gaussian process.

**Lemma 2.1.** *Consider a Gaussian process with a continuously differentiable covariance function  $k(\cdot, \cdot)$  and let  $L_k$  denote its Lipschitz constant on the set  $\mathbb{X}$  with maximum extension  $r = \max_{\mathbf{x}, \mathbf{x}' \in \mathbb{X}} \|\mathbf{x} - \mathbf{x}'\|$ . Then, the expected supremum of a sample function  $f(\mathbf{x})$  of this Gaussian process satisfies*

$$E \left[ \sup_{\mathbf{x} \in \mathbb{X}} f(\mathbf{x}) \right] \leq 12\sqrt{6d} \max \left\{ \max_{\mathbf{x} \in \mathbb{X}} \sqrt{k(\mathbf{x}, \mathbf{x})}, \sqrt{rL_k} \right\}. \quad (17)$$

*Proof.* We prove this lemma by making use of the metric entropy criterion for the sample continuity of some version of a Gaussian process [2]. This criterion allows to bound the expected supremum of a sample function  $f(\mathbf{x})$  by

$$E \left[ \sup_{\mathbf{x} \in \mathbb{X}} f(\mathbf{x}) \right] \leq \int_0^{\max_{\mathbf{x} \in \mathbb{X}} \sqrt{k(\mathbf{x}, \mathbf{x})}} \sqrt{\log(N(\varrho, \mathbb{X}))} d\varrho, \quad (18)$$

where  $N(\varrho, \mathbb{X})$  is the  $\varrho$ -packing number of  $\mathbb{X}$  with respect to the covariance pseudo-metric

$$d_k(\mathbf{x}, \mathbf{x}') = \sqrt{k(\mathbf{x}, \mathbf{x}) + k(\mathbf{x}', \mathbf{x}') - 2k(\mathbf{x}, \mathbf{x}')}. \quad (19)$$

Instead of bounding the  $\varrho$ -packing number, we bound the  $\varrho/2$ -covering number, which is known to be an upper bound. The covering number can be easily bounded by transforming the problem of

covering  $\mathbb{X}$  with respect to the pseudo-metric  $d_k(\cdot, \cdot)$  into a coverage problem in the original metric of  $\mathbb{X}$ . For this reason, define

$$\psi(\varrho') = \sup_{\substack{\mathbf{x}, \mathbf{x}' \in \mathbb{X} \\ \|\mathbf{x} - \mathbf{x}'\|_\infty \leq \varrho'}} d_k(\mathbf{x}, \mathbf{x}'), \quad (20)$$

which is continuous due to the continuity of the covariance kernel  $k(\cdot, \cdot)$ . Consider the inverse function

$$\psi^{-1}(\varrho) = \inf \{ \varrho' > 0 : \psi(\varrho') > \varrho \}. \quad (21)$$

Continuity of  $\psi(\cdot)$  implies  $\varrho = \psi(\psi^{-1}(\varrho))$ . In particular, this means that we can guarantee  $d_k(\mathbf{x}, \mathbf{x}') \leq \frac{\varrho}{2}$  if  $\|\mathbf{x} - \mathbf{x}'\| \leq \psi^{-1}(\frac{\varrho}{2})$ . Due to this relationship it is sufficient to construct a uniform grid with grid constant  $2\psi^{-1}(\frac{\varrho}{2})$  in order to obtain a  $\varrho/2$ -covering net of  $\mathbb{X}$ . Furthermore, the cardinality of this grid is an upper bound for the  $\varrho/2$ -covering number, i.e.

$$M(\varrho/2, \mathbb{X}) \leq \left\lceil \frac{r}{2\psi^{-1}(\frac{\varrho}{2})} \right\rceil^d. \quad (22)$$

Therefore, it follows that

$$N(\varrho, \mathbb{X}) \leq \left\lceil \frac{r}{2\psi^{-1}(\frac{\varrho}{2})} \right\rceil^d. \quad (23)$$

Due to the Lipschitz continuity of the covariance function, we can bound  $\psi(\cdot)$  by

$$\psi(\varrho') \leq \sqrt{2L_k \varrho'}. \quad (24)$$

Hence, the inverse function satisfies

$$\psi^{-1}\left(\frac{\varrho}{2}\right) \geq \left(\frac{\varrho}{2\sqrt{2L_k}}\right)^2 \quad (25)$$

and consequently

$$N(\varrho, \mathbb{X}) \leq \left(1 + \frac{4rL_k}{\varrho^2}\right)^d \quad (26)$$

holds, where the ceil operator is resolved through the addition of 1. Substituting this expression in the metric entropy bound (18) yields

$$E \left[ \sup_{\mathbf{x} \in \mathbb{X}} f(\mathbf{x}) \right] \leq 12\sqrt{d} \int_0^{\max_{\mathbf{x} \in \mathbb{X}} \sqrt{k(\mathbf{x}, \mathbf{x})}} \sqrt{\log \left( 1 + \frac{4rL_k}{\varrho^2} \right)} d\varrho. \quad (27)$$

As shown in [3] this integral can be bounded by

$$\int_0^{\max_{\mathbf{x} \in \mathbb{X}} \sqrt{k(\mathbf{x}, \mathbf{x})}} \sqrt{\log \left( 1 + \frac{4rL_k}{\varrho^2} \right)} d\varrho \leq \sqrt{6} \max \left\{ \max_{\mathbf{x} \in \mathbb{X}} \sqrt{k(\mathbf{x}, \mathbf{x})}, \sqrt{rL_k} \right\} \quad (28)$$

which proves the lemma.  $\square$

Based on the expected supremum of Gaussian process it is possible to derive a high probability bound for the supremum of a sample function.

**Lemma 2.2.** *Consider a Gaussian process with a continuously differentiable covariance function  $k(\cdot, \cdot)$  and let  $L_k$  denote its Lipschitz constant on the set  $\mathbb{X}$  with maximum extension  $r = \max_{\mathbf{x}, \mathbf{x}' \in \mathbb{X}} \|\mathbf{x} - \mathbf{x}'\|$ . Then, with probability of at least  $1 - \delta_L$  the supremum of a sample function  $f(\mathbf{x})$  of this Gaussian process is bounded by*

$$\sup_{\mathbf{x} \in \mathbb{X}} f(\mathbf{x}) \leq \sqrt{2 \log \left( \frac{1}{\delta_L} \right)} \max_{\mathbf{x} \in \mathbb{X}} \sqrt{k(\mathbf{x}, \mathbf{x})} + 12\sqrt{6d} \max \left\{ \max_{\mathbf{x} \in \mathbb{X}} \sqrt{k(\mathbf{x}, \mathbf{x})}, \sqrt{rL_k} \right\}. \quad (29)$$

*Proof.* We prove this lemma by exploiting the wide theory of concentration inequalities to derive a bound for the supremum of the sample function  $f(\mathbf{x})$ . We apply the Borell-TIS inequality [4]

$$P\left(\sup_{\mathbf{x} \in \mathbb{X}} f(\mathbf{x}) - E\left[\sup_{\mathbf{x} \in \mathbb{X}} f(\mathbf{x})\right] \geq c\right) \leq \exp\left(-\frac{c^2}{2 \max_{\mathbf{x} \in \mathbb{X}} k(\mathbf{x}, \mathbf{x})}\right). \quad (30)$$

Due to Lemma 2.1 we have

$$E\left[\sup_{\mathbf{x} \in \mathbb{X}} f(\mathbf{x})\right] \leq 12\sqrt{6d} \max\left\{\max_{\mathbf{x} \in \mathbb{X}} \sqrt{k(\mathbf{x}, \mathbf{x})}, \sqrt{rL_k}\right\}. \quad (31)$$

The lemma follows from substituting (31) in (30) and choosing  $c = \sqrt{2 \log\left(\frac{1}{\delta_L}\right)} \max_{\mathbf{x} \in \mathbb{X}} \sqrt{k(\mathbf{x}, \mathbf{x})}$ .  $\square$

Finally, we exploit the fact that the derivative of a sample function is a sample function from another Gaussian process to prove the high probability Lipschitz constant in Theorem 3.2.

*Proof of Theorem 3.2.* Continuity of the sample function  $f(\mathbf{x})$  follows directly from [5, Theorem 5]. Furthermore, this theorem guarantees that the derivative functions  $\frac{\partial}{\partial x_i} f(\mathbf{x})$  are samples from derivative Gaussian processes with covariance functions

$$k_{\partial i}(\mathbf{x}, \mathbf{x}') = \frac{\partial^2}{\partial x_i \partial x'_i} k(\mathbf{x}, \mathbf{x}'). \quad (32)$$

Therefore, we can apply Lemma 2.2 to each of the derivative processes and obtain with probability of at least  $1 - \frac{\delta_L}{d}$

$$-L_{f_{\partial i}} \leq \sup_{\mathbf{x} \in \mathbb{X}} \frac{\partial}{\partial x_i} f(\mathbf{x}) \leq L_{f_{\partial i}}, \quad (33)$$

where

$$L_{f_{\partial i}} = \sqrt{2 \log\left(\frac{2d}{\delta_L}\right)} \max_{\mathbf{x} \in \mathbb{X}} \sqrt{k_{\partial i}(\mathbf{x}, \mathbf{x})} + 12\sqrt{6d} \max\left\{\max_{\mathbf{x} \in \mathbb{X}} \sqrt{k_{\partial i}(\mathbf{x}, \mathbf{x})}, \sqrt{rL_k^{\partial i}}\right\} \quad (34)$$

and  $L_k^{\partial i}$  is the Lipschitz constant of derivative kernel  $k_{\partial i}(\mathbf{x}, \mathbf{x}')$ . Applying the union bound over all partial derivative processes  $i = 1, \dots, d$  finally yields the result.  $\square$

### 3 Proof of Theorem 3.3

*Proof of Theorem 3.3.* Due to Theorem 3.1 with  $\beta_N(\tau) = 2 \log\left(\frac{M(\tau, \mathbb{X})\pi^2 N^2}{3\delta}\right)$  and the union bound over all  $N > 0$  it follows that

$$\sup_{\mathbf{x} \in \mathbb{X}} |f(\mathbf{x}) - \nu_N(\mathbf{x})| \leq \sqrt{\beta_N(\tau)} \sigma_N(\mathbf{x}) + \gamma_N(\tau) \quad \forall N > 0 \quad (35)$$

with probability of at least  $1 - \delta/2$ . A trivial bound for the covering number can be obtained by considering a uniform grid over the cube containing  $\mathbb{X}$ . This approach leads to

$$M(\tau, \mathbb{X}) \leq \left(1 + \frac{r}{\tau}\right)^d, \quad (36)$$

where  $r = \max_{\mathbf{x}, \mathbf{x}' \in \mathbb{X}} \|\mathbf{x} - \mathbf{x}'\|$ . Therefore, we have

$$\beta_N(\tau) \leq 2d \log\left(1 + \frac{r}{\tau}\right) + 4 \log(\pi N) - 2 \log(3\delta). \quad (37)$$

Furthermore, the Lipschitz constant  $L_{\nu_N}$  is bounded by

$$L_{\nu_N} \leq L_k \sqrt{N} \left\|(\mathbf{K}(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 \mathbf{I}_N)^{-1} \mathbf{y}_N\right\| \quad (38)$$

due to Theorem 3.1. Since the Gram matrix  $\mathbf{K}(\mathbf{X}_N, \mathbf{X}_N)$  is positive semidefinite and  $f(\cdot)$  is bounded by  $\bar{f}$ , we can bound  $\|(\mathbf{K}(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 \mathbf{I}_N)^{-1} \mathbf{y}_N\|$  by

$$\begin{aligned} \|(\mathbf{K}(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 \mathbf{I}_N)^{-1} \mathbf{y}_N\| &\leq \frac{\|\mathbf{y}_N\|}{\rho_{\min}(\mathbf{K}(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 \mathbf{I}_N)} \\ &\leq \frac{\sqrt{N} \bar{f} + \|\boldsymbol{\xi}_N\|}{\sigma_n^2}, \end{aligned} \quad (39)$$

where  $\boldsymbol{\xi}_N$  is a vector of  $N$  i.i.d. zero mean Gaussian random variables with variance  $\sigma_n^2$ . Therefore, it follows that  $\frac{\|\boldsymbol{\xi}_N\|^2}{\sigma_n^2} \sim \chi_N^2$ . Due to [6], with probability of at least  $1 - \exp(-\eta_N)$  we have

$$\|\boldsymbol{\xi}_N\|^2 \leq \left(2\sqrt{N\eta_N} + 2\eta_N + N\right) \sigma_n^2. \quad (40)$$

Setting  $\eta_N = \log(\frac{\pi^2 N^2}{3\delta})$  and applying the union bounds over all  $N > 0$  yields

$$\|(\mathbf{K}(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 \mathbf{I}_N)^{-1} \mathbf{y}_N\| \leq \frac{\sqrt{N} \bar{f} + \sqrt{2\sqrt{N\eta_N} + 2\eta_N + N} \sigma_n}{\sigma_n^2} \quad \forall N > 0 \quad (41)$$

with probability of at least  $1 - \delta/2$ . Hence, the Lipschitz constant of the posterior mean function  $\nu_N(\cdot)$  satisfies with probability of at least  $1 - \delta/2$

$$L_{\nu_N} \leq L_k \frac{N \bar{f} + \sqrt{N(2\sqrt{N\eta_N} + 2\eta_N + N)} \sigma_n}{\sigma_n^2} \quad \forall N > 0. \quad (42)$$

Since  $\eta_N$  grows logarithmically with the number of training samples  $N$ , it holds that  $L_{\nu_N} \in \mathcal{O}(N)$  with probability of at least  $1 - \delta/2$ . The modulus of continuity  $\omega_{\sigma_N}(\cdot)$  of the posterior standard deviation can be bounded by

$$\omega_{\sigma_N}(\tau) \leq \sqrt{2L_k \tau \left( \frac{N \max_{\tilde{\mathbf{x}}, \tilde{\mathbf{x}}' \in \mathbb{X}} k(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}')}{\sigma_n^2} + 1 \right)} \quad (43)$$

because  $\|(\mathbf{K}(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 \mathbf{I}_N)^{-1}\| \leq \frac{1}{\sigma_n^2}$ . Due to the union bound (35) holds with probability of at least  $1 - \delta$  with

$$\gamma_N(\tau) \leq \sqrt{2L_k \tau \beta(\tau) \left( \frac{N \max_{\tilde{\mathbf{x}}, \tilde{\mathbf{x}}' \in \mathbb{X}} k(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}')}{\sigma_n^2} + 1 \right)} + L_f \tau + L_k \frac{N \bar{f} + \sqrt{N(2\sqrt{N\eta_N} + 2\eta_N + N)} \sigma_n}{\sigma_n^2} \tau. \quad (44)$$

This function must converge to 0 for  $N \rightarrow \infty$  in order to guarantee a vanishing regression error. This is only ensured if  $\tau(N)$  decreases faster than  $\mathcal{O}((N \log(N))^{-1})$ . Therefore, set  $\tau(N) \in \mathcal{O}(N^{-2})$  in order to guarantee

$$\lim_{N \rightarrow \infty} \gamma_N(\tau_N) = 0. \quad (45)$$

However, this choice of  $\tau(N)$  implies that  $\beta_N(\tau(N)) \in \mathcal{O}(\log(N))$  due to (37). Since there exists an  $\epsilon > 0$  such that  $\sigma_N(\mathbf{x}) \in \mathcal{O}(\log(N)^{-\frac{1}{2}-\epsilon})$ ,  $\forall \mathbf{x} \in \mathbb{X}$  by assumption, we have

$$\sqrt{\beta_N(\tau(N))} \sigma_N(\mathbf{x}) \in \mathcal{O}(\log(N)^{-\epsilon}) \quad \forall \mathbf{x} \in \mathbb{X}, \quad (46)$$

which concludes the proof.  $\square$

## 4 Proof of Theorem 4.1

Lyapunov theory provides the following statement [7].

**Lemma 4.1.** *A dynamical system  $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$  is globally ultimately bounded to a set  $\mathbb{B} \subset \mathbb{X}$ , containing the origin, if there exists a positive definite (so called Lyapunov) function,  $V : \mathbb{X} \rightarrow \mathbb{R}_{+,0}$ , for which  $\dot{V}(\mathbf{x}) < 0$ , for all  $\mathbf{x} \in \mathbb{X} \setminus \mathbb{B}$ .*

This allows to proof Theorem 4.1 as following.

*Proof of Theorem 4.1.* Consider the Lyapunov function  $V(\mathbf{x}) = \frac{1}{2}r^2$

$$\dot{V}(\mathbf{x}) = \frac{\partial V}{\partial r} \dot{r} = r \left( f(\mathbf{x}) - \hat{f}(\mathbf{x}) - \mathfrak{e}_c r \right) \leq |r| |f(\mathbf{x}) - \nu_N(\mathbf{x})| - \mathfrak{e}_c |r|^2 \leq 0 \quad \forall |r| > \frac{f(\mathbf{x}) - \nu_N(\mathbf{x})}{\mathfrak{e}_c}$$

Based on Theorem 3.1, the model error is bounded with high probability, which allows to conclude

$$P \left( \dot{V}(\mathbf{x}) < 0 \quad \forall \mathbf{x} \in \mathbb{X} \setminus \mathbb{B} \right) \geq 1 - \delta.$$

The global ultimate boundedness of the closed-loop system, is thereby shown according to Lemma 4.1.  $\square$

## 5 Report on Computational Complexity of the Numerical Evaluation

Simulations are performed in MATLAB 2019a on a i5-6200U CPU with 2.3GHz and 8GB RAM. The simulation in Sec. 5.1 took 77s and used 1 MB of workspace memory. The simulation in Sec. 5.2 took 39s and used 134 MB of workspace memory. The code is available as supplementary material.

## References

- [1] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, “Information-theoretic regret bounds for Gaussian process optimization in the bandit setting,” *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3250–3265, 2012.
- [2] R. M. Dudley, “The sizes of compact subsets of Hilbert space and continuity of Gaussian processes,” *Journal of Functional Analysis*, vol. 1, no. 3, pp. 290–330, 1967.
- [3] S. Grünewälder, J.-Y. Audibert, M. Opper, and J. Shawe-Taylor, “Regret Bounds for Gaussian Process Bandit Problems,” *Journal of Machine Learning Research*, vol. 9, pp. 273–280, 2010.
- [4] M. Talagrand, “Sharper Bounds for Gaussian and Empirical Processes,” *The Annals of Probability*, vol. 22, no. 1, pp. 28–76, 1994.
- [5] S. Ghosal and A. Roy, “Posterior consistency of Gaussian process prior for nonparametric binary regression,” *The Annals of Statistics*, vol. 34, no. 5, pp. 2413–2429, 2006.
- [6] B. Laurent and P. Massart, “Adaptive Estimation of a Quadratic Functional by Model Selection,” *The Annals of Statistics*, vol. 28, no. 5, pp. 1302–1338, 2000.
- [7] H. K. Khalil, *Nonlinear Systems; 3rd ed.* Upper Saddle River, NJ: Prentice-Hall, 2002.