Table 1: Summary of notations.

Variable	Meaning
n	Number of training samples
T	Number of testing samples
d	Degrees of freedom (one HGP for each)
α	$\lvert\psi_j(t)\rvert \leq \alpha$ (cf. Assumption 4.2)
κ	$k(t,t') \leq \kappa$ (cf. Assumption 4.1)
ν	$\nu = \max_{1\leq j\leq d}\lVert \frac{1}{\sqrt{n}}\mathbf{y}_j \rVert$
$\Sigma_{\text{noise}} \in \mathbb{R}^{n\times n}$	Diagonal time-varying noise variance at training points
$R = \Sigma_{\text{noise}}/n$	Normalized noise variance at training points
$\sigma^2_{\text{noise},t^*} \in \mathbb{R}$	Noise variance at testing point t^*
$r^2_{t^*} = \sigma^2_{\text{noise},t^*}/n$	Normalized noise variance at testing point t^*
$0 < \gamma < 1$	$R_{ii} > \gamma, 1 \leq i \leq n$
$S : \mathcal{H} \to \mathbb{R}^n$	Sampling operator with normalization $n^{-1/2}$
$L = SS^* = K/n \in \mathbb{R}^{n\times n}$	Normalized Gram matrix with exact kernel
$L_R \in \mathbb{R}^{n\times n}$	$L_R = L + R$
$\psi_j(t) = \psi(\omega_j, t)$	Element of approximate feat. vector
$\tilde{\phi}(t) = m^{-1/2}[\psi_1(t), \ldots, \psi_m(t)]^T$	Approximate feat. vector
$S_m : \mathbb{R}^m \to \mathbb{R}^n$	Sampling operator with normalization $n^{-1/2}$
$L_m = S_m S_m^* = \tilde{K}/n \in \mathbb{R}^{n\times n}$	Normalized Gram matrix with RF kernel
$L_{m,R} \in \mathbb{R}^{n\times n}$	$L_{m,R} = L_m + R$

A HGP Posterior Equations Revisited

In this section, we will rewrite the exact and approximated HGP posterior equations from Section 2 in terms of standard linear operators used in RKHS theory. For a linear operator A, we denote its adjoint by A^*. Let $\mathcal{H}$ be the RKHS associated to the kernel of interest. In order to retrieve a suitable expression, we denote $S : \mathcal{H} \to \mathbb{R}^n$ the sampling operator defined as $Sf := \frac{1}{\sqrt{n}}[f(t_1), \ldots f(t_n)]^T$. Moreover, the adjoint of the sampling operator is defined as $S^* : \mathbb{R}^n \to \mathcal{H} : S^*\mathbf{a} = \frac{1}{\sqrt{n}}\sum_{i=1}^n a_i k(t_i, \cdot)$, a_i being the i-th entry of $\mathbf{a}$. Now, let $L : \mathbb{R}^n \to \mathbb{R}^n, L := SS^*$. Note that $K = nL$. Let $R = \frac{1}{n}\Sigma_{\text{noise}}$, let $r_{t^*} = \frac{1}{n}\sigma^2_{\text{noise},t^*}$. Lastly, let $\langle\cdot,\cdot\rangle_{\mathbb{R}^n}$ denote the inner product of n-dimensional vectors. With this notation, let us consider a single DOF of the trajectory to be processed. The posterior mean of the associated exact HGP from Equation (1) is

$$\mu_{\text{post}}(t^*) = \left\langle (L+R)^{-1}Sk(t^*,\cdot), \frac{1}{\sqrt{n}}\mathbf{y} \right\rangle_{\mathbb{R}^n}. \tag{7}$$

Moreover, the posterior variance from Equation (2) is given by

$$\sigma^2_{\text{post}}(t^*) = k(t^*,t^*) + nr_{t^*} - \langle Sk(t^*,\cdot), (L+R)^{-1}Sk(t^*,\cdot)\rangle_{\mathbb{R}^n}. \tag{8}$$

Considering RFs, we can define the operator $S_m : \mathbb{R}^m \to \mathbb{R}^n, S_m := \frac{1}{\sqrt{n}}[\tilde{\phi}(t_1) \ \ldots, \tilde{\phi}(t_n)]^T$, and $L_m : \mathbb{R}^n \to \mathbb{R}^n, L_m := S_m S_m^*$. With this notation, let us consider a single DOF of the trajectory to be processed. The RF-based posterior mean of the associated HGP from Equation (4) can be rewritten as

$$\tilde{\mu}_{\text{post}}(t^*) = \left\langle (L_m+R)^{-1}S_m\tilde{\phi}(t^*), \frac{1}{\sqrt{n}}\mathbf{y} \right\rangle_{\mathbb{R}^n}. \tag{9}$$

On the other hand, the RF-based posterior variance from Equation (5) is given by

$$\tilde{\sigma}^2_{\text{post}}(t^*) = \tilde{k}(t^*,t^*) + nr_{t^*} - \langle S_m\tilde{\phi}(t^*), (L_m+R)^{-1}S_m\tilde{\phi}(t^*)\rangle_{\mathbb{R}^n}. \tag{10}$$

A summary of the main operators and constants that will appear in the proofs can be found in Table 1.

Fast matrix inversion By defintion, the operators L_m and S_m are matrices. The inversion of the matrix $L_m + R$ appearing in Equations (9) and (10) can be performed by means of Woodbury identity [29], as follows:

$$L_{m,R}^{-1} = (S_m S_m^* + R)^{-1} \tag{11}$$
$$= R^{-1} - R^{-1} S_m (I + S_m^* R^{-1} S_m)^{-1} S_m^* R^{-1}. \tag{12}$$

The latter expression involves inverting an $m \times m$ matrix, which boosts the speed of the HGP posterior calculation if $m \ll n$.

B Proofs of the Main Results

In this appendix, we report the proofs of the two main theoretical results of our paper, along with some technical propositions that will be extensively used. In the following, we denote by A_R the operator $A + R$, with R diagonal positive definite matrix, and by A_γ the operator $A + \gamma I$. Moreover, in the remainder, $\|\cdot\|$ denotes the operator norm, while $\|\cdot\|_2$ denotes the Euclidean norm of a vector.

B.1 Useful Propositions

In this part, we report three propositions that will be useful in the proofs.

Proposition B.1 (Proposition 8 of [25])**.** *Let $\mathcal{H}$ be a separable Hilbert space, A, B be two bounded self-adjoint positive linear operators on $\mathcal{H}$, and $\lambda > 0$. Then*

$$\|A_\lambda^{-1/2} B^{1/2}\| \leq \|A_\lambda^{-1/2} B_\lambda^{1/2}\| \leq \frac{1}{(1-\beta)^{1/2}}, \tag{13}$$

where

$$\beta = \lambda_{max}\left[B_\lambda^{-1/2}(B - A)B_\lambda^{-1/2}\right]. \tag{14}$$

Proposition B.2. *Let $S_m : \mathbb{R}^m \to \mathbb{R}^n, S_m \coloneqq \frac{1}{\sqrt{n}}[\tilde{\boldsymbol{\phi}}(t_1)\ \ldots, \tilde{\boldsymbol{\phi}}(t_n)]^T$, and assume that the entries of the RF vectors are bounded, that is, $|\psi_j(t)| \leq \alpha, \forall j \in \{1, \ldots, m\}$. Then,*

$$\|S_m\| \leq \alpha. \tag{15}$$

Proof. The result follows from the definition of operator norm:

$$\|S_m\| = \sup_{\mathbf{a} \in \mathbb{R}^n, \|\mathbf{a}\|_2 \leq 1} \|S_m \mathbf{a}\|_2 \tag{16}$$
$$= \sup_{\mathbf{a} \in \mathbb{R}^n, \|\mathbf{a}\|_2 \leq 1} \frac{1}{\sqrt{n}} \sqrt{\langle \tilde{\boldsymbol{\phi}}(t_1), \mathbf{a}\rangle_2^2 + \cdots + \langle \tilde{\boldsymbol{\phi}}(t_n), \mathbf{a}\rangle_2^2} \tag{17}$$
$$\leq \frac{1}{\sqrt{n}} \sqrt{n\alpha^2} = \alpha, \tag{18}$$

as reported in the statement. □

Proposition B.3. *Let A be a bounded positive semi-definite operator, and let $A_R \coloneqq A + R$, with R diagonal positive definite and $A_\gamma = A + \gamma I$. Lastly, assume all entries in R are greater or equal to γ. Then,*

$$\|A_R^{-1/2} A_\gamma^{1/2}\| \leq 1. \tag{19}$$

Proof. Noting that $A_R - A_\gamma = (R - \gamma I) \succcurlyeq 0$ by hypothesis, it holds $A_\gamma \preccurlyeq A_R$ and thus

$$\|A_R^{-1/2} A_\gamma^{1/2}\|^2 = \|A_R^{-1/2} A_\gamma A_R^{-1/2}\| \leq \|I\| = 1. \tag{20}$$

□

B.2 Proof of Theorem 4.4 (Deviation of Approximate Posterior Mean)

We report here the proof of Theorem 4.4. We start by considering a single DOF, and generalize to a d-valued GP at the end of this section. We begin by proving a lemma that will be used to retrieve the main result.

Lemma B.4. *Let* $m \geq 8\left(\frac{1}{3}+\frac{\alpha^2}{\gamma}\right)\log(\frac{8\alpha^2}{\gamma\delta})$*, and* $\delta = (0,1]$*. Then, the following bound holds, with probability at least* $1-\delta$*,*

$$\|(L_{m,R}^{-1} - L_R^{-1})Sk(t^*,\cdot)\|_2 \leq \frac{\sqrt{2}\kappa}{\sqrt{\gamma}}\left[\frac{2\log\frac{8\kappa^2}{\gamma\delta}(1+\alpha^2/\gamma)}{3m} + \sqrt{\frac{2\log\frac{8\kappa^2}{\gamma\delta}\alpha^2}{\gamma m}}\right]. \tag{21}$$

Proof. In order to bound the term of interest, we can use the fact that, for any invertible matrices A and B, $A^{-1} - B^{-1} = A^{-1}(I - AB^{-1}) = A^{-1}(B-A)B^{-1}$, and Proposition B.3, as follows:

$$\begin{aligned}
&\|(L_{m,R}^{-1} - L_R^{-1})Sk(t^*,\cdot)\|_2 && (22)\\
&=\|L_{m,R}^{-1}(L_R - L_{m,R})L_R^{-1}Sk(t^*,\cdot)\|_2 && (23)\\
&=\|L_{m,R}^{-1/2}L_{m,R}^{-1/2}L_{m,\gamma}^{1/2}L_{m,\gamma}^{-1/2}L_\gamma^{1/2}L_\gamma^{-1/2}(L-L_m)L_R^{-1}Sk(t^*,\cdot)\|_2 && (24)\\
&\leq\frac{1}{\sqrt{\gamma}}\|L_{m,R}^{-1/2}L_{m,\gamma}^{1/2}\|\|L_{m,\gamma}^{-1/2}L_\gamma^{1/2}\|\|L_\gamma^{-1/2}(L-L_m)L_R^{-1}Sk(t^*,\cdot)\|_2 && (25)\\
&\leq\frac{\kappa}{\sqrt{\gamma}}\|L_{m,\gamma}^{-1/2}L_\gamma^{1/2}\|\|L_\gamma^{-1/2}(L-L_m)L_\gamma^{-1/2}\|\|L_R^{-1/2}S\|. && (26)
\end{aligned}$$

We can now proceed to bound each of the three factors. To start off, let us consider $\|L_R^{-1/2}S\|$. This term can be bounded by using the *polar decomposition* of the bounded linear operator S, as follows. Let $S = (SS^*)^{1/2}U$, where U is a partial isometry. By Proposition B.3, the definition of polar decomposition, and by considering that $L \preccurlyeq L_\gamma$ by definition,

$$\begin{aligned}
\|L_R^{-1/2}S\| &= \|L_R^{-1/2}(SS^*)^{1/2}U\| && (27)\\
&\leq \|L_R^{-1/2}L^{1/2}\|\|U\| && (28)\\
&\leq \|L_R^{-1/2}L_\gamma^{1/2}\|\|U\| && (29)\\
&\leq 1. && (30)
\end{aligned}$$

Now, we can move on to bound $\|L_\gamma^{-1/2}(L-L_m)L_\gamma^{-1/2}\|$. To do so, we can observe that, by definition,

$$\begin{aligned}
L_m &= S_m S_m^* && (31)\\
&= \frac{1}{n}\frac{1}{m}\sum_{i=1}^m \begin{bmatrix}\psi_i(t_1)\\ \dots \\ \psi_i(t_n)\end{bmatrix} \otimes \begin{bmatrix}\psi_i(t_1)\\ \dots \\ \psi_i(t_n)\end{bmatrix}. && (32)
\end{aligned}$$

Moreover, due to linearity of expectation,

$$\mathbb{E}_\omega[L_m] = L. \tag{33}$$

We can therefore apply Proposition C.4, with $p = m$, $Q = L$, and $Q_p = L_m$. Note that $\operatorname{Tr} L$ is the trace of the normalized Gram matrix $\frac{1}{n}K$ and hence is smaller or equal to κ^2 under Assumption 4.1. Lastly, the value of the constant $\mathcal{F}_\infty(\gamma)$ in Proposition C.4 can be computed as follows:

$$\left\langle \frac{1}{\sqrt{n}}\begin{bmatrix}\psi_i(t_1)\\ \dots \\ \psi_i(t_n)\end{bmatrix}, \frac{1}{\sqrt{n}}L_\gamma^{-1}\begin{bmatrix}\psi_i(t_1)\\ \dots \\ \psi_i(t_n)\end{bmatrix}\right\rangle_{\mathbb{R}^n} \leq \frac{\alpha^2}{\gamma}. \tag{34}$$

Thus, we obtain, with probability at least $1-\delta$,

$$\|L_\gamma^{-1/2}(L-L_m)L_\gamma^{-1/2}\| \leq \frac{2\log\frac{8\kappa^2}{\gamma\delta}(1+\alpha^2/\gamma)}{3m} + \sqrt{\frac{2\log\frac{8\kappa^2}{\gamma\delta}\alpha^2}{\gamma m}}. \tag{35}$$

To conclude the proof, we can bound $\|L_{m,\gamma}^{-1/2} L_\gamma^{1/2}\|$. By Proposition B.1, we have that

$$\|L_{m,\gamma}^{-1/2} L_\gamma^{1/2}\| \leq \frac{1}{(1-\beta)^{1/2}}, \quad \text{where} \quad \beta = \lambda_{\max}\left[L_\gamma^{-1/2}(L - L_m)L_\gamma^{-1/2}\right]. \tag{36}$$

According to Equation (33), we can apply Proposition C.4 and see that with probability at least $1-\delta$

$$\beta \leq \frac{2\log\frac{8\kappa^2}{\gamma\delta}}{3m} + \sqrt{\frac{2\log\frac{8\kappa^2}{\gamma\delta}\alpha^2}{\gamma m}} \leq 0.5 \tag{37}$$

provided that $m \geq 8\left(\frac{1}{3} + \frac{\alpha^2}{\gamma}\right)\log(\frac{8\alpha^2}{\gamma\delta})$. $\square$

Proof of Theorem 4.4 In order to retrieve the main concentration result, we can consider the following decomposition of the error on the posterior mean. By Cauchy-Schwarz inequality and Equations (7) and (9),

$$\begin{aligned}
|\tilde{\mu}_{\text{post}}(t^*) - \mu_{\text{post}}(t^*)| &= \left|\left\langle L_{m,R}^{-1} S_m \tilde{\boldsymbol{\phi}}(t^*) - L_R^{-1} Sk(t^*,\cdot), \frac{1}{\sqrt{n}}\mathbf{y}\right\rangle_{\mathbb{R}^n}\right| && (38)\\
&\leq \nu\|L_{m,R}^{-1} S_m \tilde{\boldsymbol{\phi}}(t^*) - L_{m,R}^{-1} Sk(t^*,\cdot) + L_{m,R}^{-1} Sk(t^*,\cdot) - L_R^{-1} Sk(t^*,\cdot)\|_2 && (39)\\
&\leq \nu\|L_{m,R}^{-1}(S_m \tilde{\boldsymbol{\phi}}(t^*) - Sk(t^*,\cdot))\|_2 + \nu\|(L_{m,R}^{-1} - L_R^{-1})Sk(t^*,\cdot)\|_2 && (40)\\
&\leq \nu/\gamma\|S_m \tilde{\boldsymbol{\phi}}(t^*) - Sk(t^*,\cdot)\|_2 + \nu\|(L_{m,R}^{-1} - L_R^{-1})Sk(t^*,\cdot)\|_2. && (41)
\end{aligned}$$

Now, we can upper bound the two norms appearing in the expression above. The first addend can be directly bounded by applying Corollary C.3. The second addend in Equation (41) can be bounded by Lemma B.4. Hence, we obtain the following bound with probability at least $1-\delta$:

$$\begin{aligned}
|\tilde{\mu}_{\text{post}}(t^*) - \mu_{\text{post}}(t^*)| &\leq \nu/\gamma\|S_m \tilde{\boldsymbol{\phi}}(t^*) - Sk(t^*,\cdot)\|_2 + \nu\|(L_{m,R}^{-1} - L_R^{-1})Sk(t^*,\cdot)\|_2 && (42)\\
&\leq \sqrt{\frac{2\nu^2\alpha^4\log\frac{2Tn}{\delta}}{m\gamma^2}} + \frac{\sqrt{2}\kappa\nu}{\sqrt{\gamma}}\left[\frac{2\log\frac{8\kappa^2}{\gamma\delta}(1+\alpha^2/\gamma)}{3m} + \sqrt{\frac{2\log\frac{8\kappa^2}{\gamma\delta}\alpha^2}{\gamma m}}\right]. && (43)
\end{aligned}$$

The final result for the vector-valued GP can be obtained by applying a union bound.

B.3 Proof of Theorem 4.5 (Deviation of Approximate Posterior Variance)

In this section, we prove our result related to the concentration of the approximate posterior variance. Again, we begin by stating some lemmas that will be used in the proof.

Lemma B.5. *Let $\delta = (0,1]$. Then, the following bound holds, with probability at least $1-\delta$,*

$$|\langle Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*), L_R^{-1}Sk(t^*,\cdot)\rangle_{\mathbb{R}^n}| \leq \sqrt{\frac{2\kappa^2\alpha^4\log\frac{2Tn}{\delta}}{\gamma m}}. \tag{44}$$

Proof. By Cauchy-Schwarz,

$$|\langle Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*), L_R^{-1}Sk(t^*,\cdot)\rangle_{\mathbb{R}^n}| \leq \|Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*)\|_2\|L_R^{-1}Sk(t^*,\cdot)\|_2. \tag{45}$$

By using the polar decomposition of S, for a suitable partial isometry operator U, and according to Propositions B.1 and B.3

$$\|L_R^{-1}Sk(t^*,\cdot)\|_2 \leq \|L_R^{-1}(SS^*)^{1/2}U\|\|k(t^*,\cdot)\|_{\mathcal{H}} \tag{46}$$
$$\leq \kappa\|L_R^{-1}L^{1/2}\| \tag{47}$$
$$\leq \kappa\|L_R^{-1/2}\|\|L_R^{-1/2}L^{1/2}\| \tag{48}$$
$$\leq \frac{\kappa}{\sqrt{\gamma}}\|L_R^{-1/2}L_\gamma^{1/2}\|\|L_\gamma^{-1/2}L^{1/2}\| \tag{49}$$
$$\leq \frac{\kappa}{\sqrt{\gamma}}\|L_\gamma^{-1/2}L_\gamma^{1/2}\| \tag{50}$$
$$\leq \frac{\kappa}{\sqrt{\gamma}}. \tag{51}$$

To conclude the proof, we can observe that, according to Corollary C.3,

$$\|Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*)\|_2 \leq \sqrt{\frac{2\alpha^4 \log \frac{2Tn}{\delta}}{m}}. \tag{52}$$

□

Lemma B.6. *Let* $\delta = (0,1]$. *Then, the following bound holds, with probability at least* $1-\delta$*,*

$$|\langle S_m\tilde{\boldsymbol{\phi}}(t^*), L_R^{-1}(Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*))\rangle_{\mathbb{R}^n}| \leq \frac{\alpha^2}{\gamma}\sqrt{\frac{2\alpha^4 \log \frac{2Tn}{\delta}}{m}}. \tag{53}$$

Proof. By Cauchy-Schwarz inequality and Proposition B.2,

$$|\langle S_m\tilde{\boldsymbol{\phi}}(t^*), L_R^{-1}(Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*))\rangle_{\mathbb{R}^n}| \leq \|S_m\tilde{\boldsymbol{\phi}}(t^*)\|_2\|L_R^{-1}(Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*))\|_2 \tag{54}$$
$$\leq \|S_m\|\|\tilde{\boldsymbol{\phi}}(t^*)\|_2\|L_R^{-1}(Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*))\|_2 \tag{55}$$
$$\leq \frac{\alpha^2}{\gamma}\|Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*)\|_2. \tag{56}$$

Now, we can again observe that, according to Corollary C.3,

$$\|Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*)\|_2 \leq \sqrt{\frac{2\alpha^4 \log \frac{2Tn}{\delta}}{m}}, \tag{57}$$

which concludes the proof. □

Lemma B.7. *Let* $m \geq 8\left(\frac{1}{3} + \frac{\alpha^2}{\gamma}\right)\log(\frac{8\alpha^2}{\gamma\delta})$*, and* $\delta = (0,1]$. *Then, the following bound holds, with probability at least* $1-\delta$*,*

$$|\langle S_m\tilde{\boldsymbol{\phi}}(t^*), (L_R^{-1} - L_{m,R}^{-1})S_m\tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^n}| \leq \frac{\alpha^3\sqrt{2}}{\sqrt{\gamma}}\left[\frac{2\log\frac{8\kappa^2}{\gamma\delta}(1+\alpha^2/\gamma)}{3m} + \sqrt{\frac{2\log\frac{8\kappa^2}{\gamma\delta}\alpha^2}{\gamma m}}\right]. \tag{58}$$

Proof. Firstly, we can observe that, by Cauchy-Schwarz inequality, Propositions B.2 and B.3, the polar decomposition of S_m, and the fact that $L_m \preccurlyeq L_{m,\gamma}$ by definition, we have that

$$
\begin{aligned}
|\langle S_m\tilde{\boldsymbol{\phi}}(t^*),&(L_R^{-1} - L_{m,R}^{-1})S_m\tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^n}| \\
&= |\langle S_m\tilde{\boldsymbol{\phi}}(t^*), L_{m,R}^{-1}(L - L_m)L_R^{-1}S_m\tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^n}| && (59)\\
&= |\langle L_{m,R}^{-1/2}S_m\tilde{\boldsymbol{\phi}}(t^*), L_{m,R}^{-1/2}(L - L_m)L_R^{-1}S_m\tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^n}| && (60)\\
&\le \|L_{m,R}^{-1/2}S_m\tilde{\boldsymbol{\phi}}(t^*)\|_2\|L_{m,R}^{-1/2}(L - L_m)L_R^{-1}S_m\tilde{\boldsymbol{\phi}}(t^*)\|_2 && (61)\\
&\le \alpha^3\|L_{m,R}^{-1/2}(S_mS_m^*)^{1/2}U\|\|L_{m,R}^{-1/2}L_{m,\gamma}^{1/2}L_{m,\gamma}^{-1/2}(L - L_m)L_R^{-1}S_m\tilde{\boldsymbol{\phi}}(t^*)\|_2 && (62)\\
&\le \alpha^3\|L_{m,R}^{-1/2}L_{m,\gamma}^{1/2}\|\|L_{m,\gamma}^{-1/2}L_\gamma^{1/2}\|\|L_\gamma^{-1/2}(L - L_m)L_\gamma^{-1/2}\|\|L_\gamma^{1/2}L_R^{-1/2}\|\|L_R^{-1/2}\| && (63)\\
&\le \frac{\alpha^3}{\sqrt{\gamma}}\|L_{m,\gamma}^{-1/2}L_\gamma^{1/2}\|\|L_\gamma^{-1/2}(L - L_m)L_\gamma^{-1/2}\|. && (64)
\end{aligned}
$$

Now, we can bound the two factors. According to Propositions B.1 and C.4, with probability at least $1-\delta$, for $\delta \in (0,1]$ and $m \ge 8\left(\frac{1}{3} + \frac{\alpha^2}{\gamma}\right)\log(\frac{8\alpha^2}{\gamma\delta})$, we have that

$$
\|L_{m,\gamma}^{-1/2}L_\gamma^{1/2}\|\|L_\gamma^{-1/2}(L_\gamma - L_{m,\gamma})L_\gamma^{-1/2}\| \le \sqrt{2}\left[\frac{2\log\frac{8\kappa^2}{\gamma\delta}(1+\alpha^2/\gamma)}{3m} + \sqrt{\frac{2\log\frac{8\kappa^2}{\gamma\delta}\alpha^2}{\gamma m}}\right], \quad (65)
$$

concluding the proof. □

Proof of Theorem 4.5 We are now ready to prove Theorem 4.5. According to Equations (8) and (10), and similarly to what we did for the posterior mean, we can decompose the error on the variance of a single DOF as follows:

$$
\begin{aligned}
|\sigma^2_{\text{post}}(t^*) - \sigma^2_{\text{post}}(t^*)| =&|k(t^*,t^*) - \langle Sk(t^*,\cdot), L_R^{-1}Sk(t^*,\cdot)\rangle_{\mathbb{R}^n} \\
&- \tilde{k}(t^*,t^*) + \langle S_m\tilde{\boldsymbol{\phi}}(t^*), L_{m,R}^{-1}S_m\tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^n}| && (66)\\
\le&|k(t^*,t^*) - \tilde{k}(t^*,t^*)| \\
&+ |\langle Sk(t^*,\cdot), L_R^{-1}Sk(t^*,\cdot)\rangle_{\mathbb{R}^n} - \langle S_m\tilde{\boldsymbol{\phi}}(t^*), L_{m,R}^{-1}S_m\tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^n}| && (67)\\
\le&|k(t^*,t^*) - \tilde{k}(t^*,t^*)| \\
&+ |\langle Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*), L_R^{-1}Sk(t^*,\cdot)\rangle_{\mathbb{R}^n}| \\
&+ |\langle S_m\tilde{\boldsymbol{\phi}}(t^*), L_R^{-1}Sk(t^*,\cdot) - L_{m,R}^{-1}S_m\tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^n}| && (68)\\
\le&|k(t^*,t^*) - \tilde{k}(t^*,t^*)| \\
&+ |\langle Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*), L_R^{-1}Sk(t^*,\cdot)\rangle_{\mathbb{R}^n}| \\
&+ |\langle S_m\tilde{\boldsymbol{\phi}}(t^*), L_R^{-1}(Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*))\rangle_{\mathbb{R}^n}| \\
&+ |\langle S_m\tilde{\boldsymbol{\phi}}(t^*), (L_R^{-1} - L_{m,R}^{-1})S_m\tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^n}|. && (69)
\end{aligned}
$$

Now, we can upper bound the four addends appearing in the decomposition above. The first addend can by directly bounded by Corollary C.2. The second addend of the decomposition in Equation (69) can be bounded by Lemma B.5. The third addend in Equation (69) can be bounded by Lemma B.6. The last addend in Equation (69) can be bounded by Lemma B.7. In this way, we retrieve the result of Theorem 4.5, obtaining the following bound holding with probability at least $1-\delta$. Having defined

$$C := \sqrt{\frac{2\alpha^4 \log \frac{2T}{\delta}}{m}} + \sqrt{\frac{2\kappa^2\alpha^4 \log \frac{2Tn}{\delta}}{\gamma m}} + \frac{\alpha^2}{\gamma}\sqrt{\frac{2\alpha^4 \log \frac{2Tn}{\delta}}{m}} + \frac{\alpha^3\sqrt{2}}{\sqrt{\gamma}}\left[\frac{2\log \frac{8\kappa^2}{\gamma\delta}(1+\alpha^2/\gamma)}{3m} + \sqrt{\frac{2\log \frac{8\kappa^2}{\gamma\delta}\alpha^2}{\gamma m}}\right]:$$

$$\begin{aligned}
|\sigma^2_{\text{post}}(t^*) - \sigma^2_{\text{post}}(t^*)| \leq & |\langle k(t^*,\cdot), k(t^*,\cdot)\rangle_{\mathcal{H}} - \langle \tilde{\boldsymbol{\phi}}(t^*), \tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^m}| \\
& + \langle Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*), L_R^{-1}Sk(t^*,\cdot)\rangle_{\mathbb{R}^n}| \\
& + |\langle S_m\tilde{\boldsymbol{\phi}}(t^*), L_R^{-1}(Sk(t^*,\cdot) - S_m\tilde{\boldsymbol{\phi}}(t^*))\rangle_{\mathbb{R}^n}| \\
& + |\langle S_m\tilde{\boldsymbol{\phi}}(t^*), (L_R^{-1} - L_{m,R}^{-1})S_m\tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^n}| \qquad (70)\\
\leq & \; C. \qquad (71)
\end{aligned}$$

The final result for the vector-valued GP can be obtained by applying a union bound.

C Concentration Results

We first provide a few lemmas for the concentration of the approximate kernel functions that derive from Hoeffding inequality, and then a lemma for the concentration of random operators that derives from Bernstein inequality. Again, we denote by A_γ the operator $A + \gamma I$. $\|\cdot\|$ denotes the operator norm, while $\|\cdot\|_2$ denotes the Euclidean norm of a vector.

C.1 Approximation of the Kernel Function

Note that if a *uniform* convergence of the RF-HGP posterior is seeked w.r.t. the domain of the function modelled with the HGP, our proofs could be adapted by replacing the following Lemma C.1 with a uniform convergence result. For instance, in the case of RFFs, such a result can be found in [17, Claim 1].

Lemma C.1. *Let* $\delta = (0, 1]$. *Then, for any* (t_1, t_2) *it holds with probability at least* $1 - \delta$*, it holds*

$$\left|\tilde{\boldsymbol{\phi}}(t_1)^T\tilde{\boldsymbol{\phi}}(t_2) - k(t_1,t_2)\right| \leq \sqrt{\frac{2\alpha^4 \log \frac{2}{\delta}}{m}}, \quad \forall t_1, t_2 \in \mathcal{X}. \tag{72}$$

Proof. To upper bound the quantity of interest, we can use Hoeffding's inequality for bounded random variables. Let $A_j(t_1,t_2) := \psi_j(t_1)\psi_j(t_2) - \mathbb{E}_\omega\psi(\omega,t_1)\psi(\omega,t_2)$. Since $-\alpha^2 \leq \psi_j(t_1)\psi_j(t_2) \leq \alpha^2$ according to Assumption 4.2, by Hoeffding inequality, we have that

$$\Pr\left\{\frac{1}{m}\left|\sum_{j=1}^m A_j(t_1,t_2)\right| \geq \frac{t}{m}\right\} \leq 2e^{-\frac{2t^2}{4m\alpha^4}}. \tag{73}$$

Therefore, by setting the above upper bound smaller than δ, for $\delta \in (0, 1]$, we get that with probability at least $1 - \delta$

$$\left|\tilde{\boldsymbol{\phi}}(t_1)^T\tilde{\boldsymbol{\phi}}(t_2) - k(t_1,t_2)\right| = \frac{1}{m}\left|\sum_{j=1}^m A_j(t_1,t_2)\right| \leq \sqrt{\frac{2\alpha^4 \log \frac{2}{\delta}}{m}}. \tag{74}$$

□

Corollary C.2. *Let* $\delta = (0, 1]$. *Then with probability at least* $1 - \delta$*, it holds*

$$\left|\tilde{\boldsymbol{\phi}}(t^*)^T\tilde{\boldsymbol{\phi}}(t^*) - k(t^*,t^*)\right| \leq \sqrt{\frac{2\alpha^4 \log \frac{2|\mathcal{T}|}{\delta}}{m}}, \quad \forall t^* \in \mathcal{T}. \tag{75}$$

Proof. We apply Lemma C.1 on each element of $\mathcal{T}$ with $\delta' := \delta/T$. The claimed result then follows using a union bound. □

Corollary C.3. *Let $\delta = (0, 1]$. Then with probability at least $1 - \delta$,*

$$\|S_m \tilde{\phi}(t^*) - Sk(t^*, \cdot)\|_2 \leq \sqrt{\frac{2\alpha^4 \log \frac{2Tn}{\delta}}{m}}, \quad \forall t^* \in \mathcal{T}. \tag{76}$$

Proof. It holds

$$\|S_m \tilde{\phi}(t^*) - Sk(t^*, \cdot)\|_2^2 = \frac{1}{n} \sum_{i=1}^{n} \left[\tilde{\phi}(t_i)^T \tilde{\phi}(t^*) - k(t_i, t^*) \right]^2 \tag{77}$$

The result thus follows from applying nT times Lemma C.1 on the pairs $((t_i, t^*))_{1 \leq i \leq n, t^* \in \mathcal{T}}$ with probability $\delta' := \delta/(nd)$ and using a union bound. □

C.2 Concentration of the Kernel matrix

The following result derives from the Bernstein inequality for sums of random operators on separable Hilbert spaces in operator norm.

Proposition C.4 (Proposition 6 and Remark 10 of [25]). *Let $\mathbf{v}_1, ..., \mathbf{v}_p$ with $p \geq 1$, be independent and identically distributed random vectors on a separable Hilbert spaces $\mathcal{H}$ such that $Q = \mathbb{E}\mathbf{v} \otimes \mathbf{v}$ is trace-class, and for any $\lambda > 0$ there exists a constant $\mathcal{F}_\infty(\lambda) < \infty$ such that $\langle \mathbf{v}, (Q + \lambda I)^{-1} \mathbf{v} \rangle \leq \mathcal{F}_\infty(\lambda)$ almost everywhere. Let $Q_p = \frac{1}{p} \sum_{i=1}^{p} \mathbf{v}_i \otimes \mathbf{v}_i$ and take $0 < \lambda \leq \|Q\|$. Then for any $\delta \geq 0$, the following holds with probability at least $1 - \delta$:*

$$\|Q_\lambda^{-1/2}(Q - Q_p)Q_\lambda^{-1/2}\| \leq \frac{2w(1 + \mathcal{F}_\infty(\lambda))}{3p} + \sqrt{\frac{2w\mathcal{F}_\infty(\lambda)}{p}} \tag{78}$$

where $w = \log \frac{8 \operatorname{Tr} Q}{\lambda \delta}$. Moreover, with the same probability,

$$\lambda_{max} \left[Q_\lambda^{-1/2}(Q - Q_p)Q_\lambda^{-1/2} \right] \leq \frac{2w}{3p} + \sqrt{\frac{2w\mathcal{F}_\infty(\lambda)}{p}}. \tag{79}$$

Moreover, for any $s \in (0, 1]$, if $\|\mathbf{v}_i\| \leq \alpha$, we have that, with probability at least $1 - \delta$,

$$\lambda_{max} \left[Q_\lambda^{-1/2}(Q - Q_p)Q_\lambda^{-1/2} \right] \leq s. \tag{80}$$

provided that $p \geq \frac{2}{t^2} \left[\frac{2t}{3} + \mathcal{F}_\infty(\gamma) \right] \log \frac{8\alpha^2}{\lambda \delta}$ and $\lambda \leq \|Q\|$.

D Efficient Matrix Inversion and Online Updates

In this section, we show how the expression of the posterior mean and variance can easily be updated when adding new samples to the dataset.

We recall that the operators S_m and $L_{m,R}$ are matrices and are defined in Appendix A. As discussed in Appendix A, the inversion of $L_{m,R}$, involved the posteriors of Equations (9) and (10), can be simplified by applying Woodbury identity [29], as follows:

$$\begin{aligned} L_{m,R}^{-1} &= (S_m S_m^* + R)^{-1} & (81) \\ &= R^{-1} - R^{-1} S_m (I + S_m^* R^{-1} S_m)^{-1} S_m^* R^{-1}. & (82) \end{aligned}$$

The posterior mean of the HGP in Equation (9) becomes:

$$\begin{aligned} \tilde{\mu}_{\text{post}}(t^*) &= \langle \left[R^{-1} - R^{-1} S_m (I + S_m^* R^{-1} S_m)^{-1} S_m^* R^{-1} \right] S_m \tilde{\phi}(t^*), \frac{1}{\sqrt{n}} \mathbf{y} \rangle_{\mathbb{R}^n} & (83) \\ &= \tilde{\phi}(t^*)^T \left[I - S_m^* R^{-1} S_m (I + S_m^* R^{-1} S_m)^{-1} \right] S_m^* R^{-1} \frac{1}{\sqrt{n}} \mathbf{y} & (84) \\ &= \tilde{\phi}(t^*)^T \left[I - B(I + B)^{-1} \right] A & (85) \end{aligned}$$

where $A := \frac{1}{\sqrt{n}} S_m^* R^{-1}\mathbf{y} \in \mathbb{R}^m$ and $B := S_m^* R^{-1} S_m \in \mathbb{R}^{m\times m}$. Moreover, the only term in the expression of the posterior variance of Equation (10)

$$\tilde{\sigma}^2_{\text{post}}(t^*) = \langle \tilde{\boldsymbol{\phi}}(t^*), \tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^m} + n r_{t^*} - \langle S_m \tilde{\boldsymbol{\phi}}(t^*), (L_m + R)^{-1} S_m \tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^n} \tag{86}$$

which varies with n is

$$\langle S_m \tilde{\boldsymbol{\phi}}(t^*), (L_m + R)^{-1} S_m \tilde{\boldsymbol{\phi}}(t^*)\rangle_{\mathbb{R}^n} \tag{87}$$

$$= \tilde{\boldsymbol{\phi}}^T(t^*) \left[S_m^* R^{-1} S_m - S_m^* R^{-1} S_m (I + S_m^* R^{-1} S_m)^{-1} S_m^* R^{-1} S_m \right] \tilde{\boldsymbol{\phi}}(t^*), \tag{88}$$

$$= \tilde{\boldsymbol{\phi}}^T(t^*) \left[B - B(I+B)^{-1} B \right] \tilde{\boldsymbol{\phi}}(t^*). \tag{89}$$

When a new human demonstration is gathered, the training set is enlarged by adding n_{new} training points. This means that the matrix S_m is updated by adding n_{new} rows (and renormalized), containing the RF embeddings of the new training points. The same happens to vector $\mathbf{y}$ and to the diagonal matrix R, which is enlarged by adding n_{new} rows and columns. This means that matrices A and B support online updates. In particular, after initializing A and B to the null matrix, having collected the new embeddings in $S_{m,new} \in \mathbb{R}^{n_{new}\times m}$ (with normalization $n_{new}^{-1/2}$) and the new noise variance values in $R_{new} \in \mathbb{R}^{n_{new}\times n_{new}}$ (with normalization n_{new}^{-1}), the updates are as follows:

$$A \leftarrow A + \frac{1}{\sqrt{n_{new}}} S^*_{m,new} R^{-1}_{new} \mathbf{y}_{new} \tag{90}$$

$$B \leftarrow B + S^*_{m,new} R^{-1}_{new} S_{m,new}. \tag{91}$$

Having computed the updates, the matrices appearing in the posterior mean and variance can be computed in constant time w.r.t. the current size of the training set during the data streaming.