A. Additional experimental results on Branin-Hoo function

In this section, we empirically evaluate the performance of our PO-GP-UCB algorithm using the dataset sampled from Branin-Hoo benchmark function[8]. The original inputs for this experiment are 2-dimensional vectors arranged into a uniform grid and discretized into a 31×31 input domain (i.e., $d = 2$ and $n = 961$). The function to maximize is sampled from the negation of Branin-Hoo function. The original output measurements are log-transformed to remove skewness and extremity in order to stabilize the GP covariance structure. The GP hyperparameters are learned using maximum likelihood estimation (Rasmussen & Williams, 2006). Similarly to the real-world loan applications dataset in Section 4, the original inputs are preprocessed to form an isotropic covariance function[4]. All results are averaged over 50 random runs, each of which uses a different set of initializations for BO. We set the GP-UCB parameter $\delta_{ucb} = 0.05$ (Theorem 3) and normalize the inputs to have a maximal norm of 25. We set the parameter $r = 10$ (Algorithm 1), DP parameter $\delta = 10^{-3}$ (Definition 2) and the GP-UCB parameter $T = 50$ for this experiment.

Fig. 3 shows the performances of PO-GP-UCB with different values of ϵ and that of non-private GP-UCB. The results are consistent with the previous experiments. Smaller values of ϵ (tighter privacy guarantees) generally lead to larger simple regret; PO-GP-UCB with the largest value of $\epsilon = \exp(2.3)$ satisfying the condition $\sigma_{min}(\mathcal{X}) \geq \omega$ incurs only $0.004\sigma_y$ more simple regret than non-private GP-UCB after 50 iterations; PO-GP-UCB with some values of ϵ in the single-digit range satisfying the condition $\sigma_{min}(\mathcal{X}) < \omega$ exhibits small difference in simple regret compared with non-private GP-UCB after 50 iterations: $\epsilon = \exp(2.0)$ and $\epsilon = \exp(1.8)$ result in $0.023\sigma_y$ and $0.051\sigma_y$ more simple regret respectively.

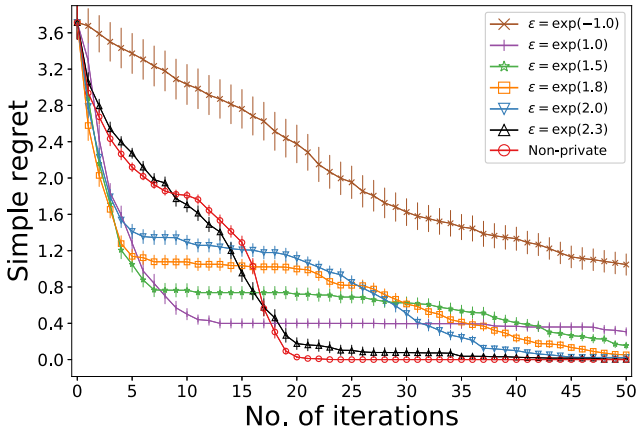

Figure 3. Simple regrets achieved by tested BO algorithms (with fixed $r = 10$ and different values of ϵ) vs. the number of iterations for the Branin-Hoo function dataset.

Similarly to the experiments in the main text, we investigate the impact of varying the value of the random projection parameter r on the performance of PO-GP-UCB. We consider 3 different values of DP parameter ϵ: $\epsilon = \exp(2.3)$, $\epsilon = \exp(2.5)$ and $\epsilon = \exp(2.7)$. We fix the value of ϵ and vary the value of r. The largest value of r satisfying the condition $\sigma_{min}(\mathcal{X}) \geq \omega$ is $r = 10$ for $\epsilon = \exp(2.3)$, $r = 15$ for $\epsilon = \exp(2.5)$ and $r = 20$ for $\epsilon = \exp(2.7)$. Tables 10, 11 and 12 reveal that the largest values of r satisfying the condition $\sigma_{min}(\mathcal{X}) \geq \omega$ lead to the smallest simple regret after 50 iterations. Decreasing the value of r increases the simple regret, which agrees with our analysis in Section 3.4 (i.e., smaller r results in worse regret upper bound). Increasing r such that the condition $\sigma_{min}(\mathcal{X}) < \omega$ is satisfied, on the other hand, also results in larger simple regret, which is again consistent with the analysis in Remark 2 stating that the regret upper bound becomes looser in this scenario. These observations are consisted with those for a synthetic GP dataset, a real-world loan applications dataset and a real-world property price dataset in the main text.

[8] https://www.sfu.ca/~ssurjano/branin.html.

Table 10. Simple regrets achieved by PO-GP-UCB with fixed $\epsilon = \exp(2.3)$ and different values of r after 50 iterations for the Branin-Hoo function dataset. The largest value of r satisfying the condition $\sigma_{min}(\mathcal{X}) \geq \omega$ is $r = 10$.

r	3	6	8	10	15	20
S_{50}	0.53	0.184	0.038	**0.0**	0.005	0.024

Table 11. Simple regrets achieved by PO-GP-UCB with fixed $\epsilon = \exp(2.5)$ and different values of r after 50 iterations for the Branin-Hoo function dataset. The largest value of r satisfying the condition $\sigma_{min}(\mathcal{X}) \geq \omega$ is $r = 15$.

r	3	9	12	15	20	30
S_{50}	0.259	0.001	**0.0**	**0.0**	0.014	0.026

Table 12. Simple regrets achieved by PO-GP-UCB with fixed $\epsilon = \exp(2.7)$ and different values of r after 50 iterations for the Branin-Hoo function dataset. The largest value of r satisfying the condition $\sigma_{min}(\mathcal{X}) \geq \omega$ is $r = 20$.

r	5	10	15	20	30	50
S_{50}	0.152	**0.0**	**0.0**	**0.0**	0.005	0.073

B. Proofs and derivations

B.1. Proof of Lemma 1

Theorem 4. *[Johnson-Lindenstrauss lemma (Johnson & Lindenstrauss, 1984)] Let $\nu \in (0, 1/2)$, $r \in \mathbb{N}$ and $d \in \mathbb{N}$ be given. Let M' be a $r \times d$ matrix whose entries are i.i.d. samples from $\mathcal{N}(0, 1)$. Then for any vector $y \in \mathbb{R}^d$*

$$P\Big((1-\nu)\|y\|^2 \leq r^{-1}\|M'y\|^2 \leq (1+\nu)\|y\|^2\Big) \geq 1 - 2\exp(-\nu^2 r/8).$$

Proof of lemma. Fix $x, x' \in \mathcal{X}$. It follows from Theorem 4 by setting vector $y = (x - x')^\top$ and $r \times d$ matrix $M' = M^\top$ that

$$\begin{aligned} &1 - 2\exp(-\nu^2 r/8) \\ &\leq P\Big((1-\nu)\|(x-x')^\top\|^2 \leq r^{-1}\|M^\top(x-x')^\top\|^2 \leq (1+\nu)\|(x-x')^\top\|^2\Big) \\ &= P\Big((1-\nu)\|x-x'\|^2 \leq r^{-1}\|xM - x'M\|^2 \leq (1+\nu)\|x-x'\|^2\Big). \end{aligned} \tag{2}$$

Since there are no more than $n^2/2$ pairs of inputs $x, x' \in \mathcal{X}$, applying the union bound to (2) gives that the probability of

$$(1-\nu)\|x-x'\|^2 \leq r^{-1}\|xM - x'M\|^2 \leq (1+\nu)\|x-x'\|^2$$

for all $x, x' \in \mathcal{X}$ is at least $1 - n^2\exp(-\nu^2 r/8)$.

To guarantee that the probability of $(1-\nu)\|x-x'\|^2 \leq r^{-1}\|Mx - Mx'\|^2 \leq (1+\nu)\|x-x'\|^2$ for all $x, x' \in \mathcal{X}$ is at least $1 - \mu$, the value of r has to satisfy the following inequality:

$$1 - n^2\exp(-\nu^2 r/8) \geq 1 - \mu,$$

which is equivalent to $r \geq 8\log(n^2/\mu)/\nu^2$.

B.2. Privacy guarantee of Algorithm 1

B.2.1. Comparison between Algorithm 1 and Algorithm 3 of Blocki et al. (2012)

There are several important differences between our Algorithm 1 and the work of Blocki et al. (2012). Firstly, Algorithm 3 of Blocki et al. (2012) outputs a DP estimate $r^{-1}\tilde{\mathcal{X}}^\top M^\top M\tilde{\mathcal{X}}$ (in the notations of Algorithm 1) of the covariance matrix $r^{-1}\mathcal{X}^\top\mathcal{X}$, while our Algorithm 1 outputs a DP transformation $r^{-1/2}\mathcal{X}M$ (or $r^{-1/2}\tilde{\mathcal{X}}M$) of the original dataset $\mathcal{X}$. However, the authors of Blocki et al. (2012) prove the privacy guarantee (see Theorem 4.1, p. 13 of their paper) by showing that releasing $\tilde{\mathcal{X}}^\top M^\top$ (using matrix M of size $r \times n$) preserves DP and then apply the post-processing property of DP to reconstruct $r^{-1}\tilde{\mathcal{X}}^\top M^\top M\tilde{\mathcal{X}}$. This observation allows us to modify their proof for our Algorithm 1. Additionally, matrix

$\tilde{\mathcal{X}}^\top M^\top$ (in the notations of Algorithm 1) in the proof of Blocki et al. (2012) has size $d \times r$, while matrices $r^{-1/2}\mathcal{X}M$ and $r^{-1/2}\tilde{\mathcal{X}}M$ returned by our Algorithm 1 have size $n \times r$, which requires us to modify the proof of Blocki et al. (2012). These modifications are discussed in Section B.2.2 below.

Secondly, Algorithm 3 of Blocki et al. (2012) does not have the "if/else" condition (line 6 of Algorithm 1) and always increases the singular values as in line 9 of Algorithm 1, since the authors are able to offset the bias introduced to the estimate of covariance of the dataset along a given dimension by increasing the singular values. Specifically, they do it by subtracting ω^2 from the computed estimate (see Algorithm 4 in Blocki et al. (2012)). For our case, however, the distances between the original inputs from the dataset $\mathcal{X}$ are no longer approximately the same as the distances between their images from the dataset $\mathcal{Z}$ when $\sigma_{min}(\mathcal{X}) < \omega$ (i.e., the "else" clause, line 8 of Algorithm 1), as shown in Theorem 2. Therefore, the case of $\sigma_{min}(\mathcal{X}) < \omega$ results in a slightly different regret bound (see Theorem 3 and Remark 2) and requires us to introduce the "if/else" condition into Algorithm 1. Introducing such an "if/else" condition, however, does not affect the proof of Theorem 4.1 of Blocki et al. (2012) and our proof: the "if" clause (line 6 of Algorithm 1) is stated in the Corollary (see p. 17 of Blocki et al. (2012)), while the "else" clause (line 8 of Algorithm 1) is proved in Theorem 4.1 of Blocki et al. (2012).

B.2.2. Proof of Theorem 1

Fix two neighboring datasets $\mathcal{X}$ and $\mathcal{X}'$. Let $E \triangleq \mathcal{X}' - \mathcal{X}$, such that E is a rank 1 matrix. Without loss of generality, we assume that in the definition of neighboring datasets (Definition 1) $\|x_{(i^*)} - x'_{(i^*)}\| = 1$. Then we can write E as the outer product $E = e_{i^*}v^\top$ where e_{i^*} is the indicator vector of row i^* and v is the vector of norm 1. Then the singular values of E are exactly $\{1, 0, \ldots, 0\}$ (see Blocki et al. (2012), p. 14).

Similar to Theorem 4.1 of Blocki et al. (2012), the proof is composed of two stages. For the first stage we work under the premise that both and $\mathcal{X}$ and $\mathcal{X}'$ have singular values no less than ω (the "if" clause, line 6 of Algorithm 1). For the second stage we denote $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{X}}'$ as the respective matrices from "else" clause (line 8 of Algorithm 1) and show what adaptations are needed to make the proof follow through.

We prove the theorem for the scaled output of the "if" clause of Algorithm 1 $\mathcal{X}M$ (the post-processing property of DP can be applied after that to reconstruct $r^{-1/2}\mathcal{X}M$). $\mathcal{X}M$ is composed of r columns each is an i.i.d. sample from $\mathcal{X}Y$ where $Y \sim \mathcal{N}(0, I_{d\times d})$. The following lemma is similar to Claim 4.3 of Blocki et al. (2012)(p. 14):

Lemma 2. *Let $\epsilon > 0$, $\delta \in (0,1)$, $r \in \mathbb{N}$, $d \in \mathbb{N}$, two neighboring datasets $\mathcal{X}$ and $\mathcal{X}'$ and Y sampled from $\mathcal{N}(0, I_{d\times d})$ be given. Fix $\epsilon_0 \triangleq \epsilon/\sqrt{4r\log(2/\delta)}$ and $\delta_0 \triangleq \delta/(2r)$. Denote*

$$S \triangleq \{\xi \in \mathbb{R}^n : \exp(-\epsilon_0)PDF_{\mathcal{X}'Y}(\xi) \le PDF_{\mathcal{X}Y}(\xi) \le \exp(\epsilon_0)PDF_{\mathcal{X}'Y}(\xi)\}$$

where PDF is the probability density function. Then $P(S) \ge 1 - \delta_0$.

Proof. Similar to the proof of Claim 4.3 of Blocki et al. (2012), first we formally define the PDF of the two distributions. We apply the fact that $\mathcal{X}Y$ and $\mathcal{X}'Y$ are linear transformations of $\mathcal{N}(0, I_{d\times d})$.

$$\text{PDF}_{\mathcal{X}Y}(\xi) = \frac{1}{\sqrt{(2\pi)^n \det(\mathcal{X}\mathcal{X}^\top)}} \exp\Big(-\frac{1}{2}\xi^\top(\mathcal{X}\mathcal{X}^\top)^{-1}\xi\Big)$$
$$\text{PDF}_{\mathcal{X}'Y}(\xi) = \frac{1}{\sqrt{(2\pi)^n \det(\mathcal{X}'\mathcal{X}'^\top)}} \exp\Big(-\frac{1}{2}\xi^\top(\mathcal{X}'\mathcal{X}'^\top)^{-1}\xi\Big).$$

If the matrix $\mathcal{X}\mathcal{X}^\top$ (all the reasoning here is exactly the same for $\mathcal{X}'\mathcal{X}'^\top$) is not full-rank, the SVD allows us to use similar notation to denote the generalizations of the inverse and of the determinant: The Moore-Penrose inverse of any square matrix M is $M^\dagger \triangleq V\Sigma^{-1}U^\top$ where $M = U\Sigma V^\top$ is the SVD of matrix M, and the pseudo-determinant of M is $\widetilde{det}(M) \triangleq \Pi_{i=1}^{rank(M)}\sigma_i(M)$ where $\sigma_i(M)$ are the singular values of matrix M. Furthermore, if $\mathcal{X}\mathcal{X}^\top$ has non-trivial kernel space (i.e., is not invertible) then $\text{PDF}_{\mathcal{X}Y}$ in the equation above is technically undefined. However, if we restrict ourselves only to the subspace $\mathcal{V} = (\text{Ker}(\mathcal{X}\mathcal{X}^\top))^\perp$, then $\text{PDF}^{\mathcal{V}}_{\mathcal{X}Y}$ is defined over $\mathcal{V}$ and $\text{PDF}^{\mathcal{V}}_{\mathcal{X}Y}(\xi) \triangleq \frac{1}{\sqrt{(2\pi)^{rank(\mathcal{X}\mathcal{X}^\top)}\widetilde{\det}(\mathcal{X}\mathcal{X}^\top)}} \exp\Big(-\frac{1}{2}\xi^\top(\mathcal{X}\mathcal{X}^\top)^\dagger\xi\Big)$
From now on, we omit the superscript from the PDF and refer to the above function as the PDF of $\mathcal{X}Y$. See p. 4–5 of Blocki et al. (2012) for more details.

Similar to the proof of Claim 4.3 of Blocki et al. (2012), first we show that

$$\exp(-\epsilon_0/2) \leq \sqrt{\frac{\det(\mathcal{X}'\mathcal{X}'^\top)}{\det(\mathcal{X}\mathcal{X}^\top)}} \leq \exp(\epsilon_0/2).$$

The proof copies the derivation of eq. 4 in Blocki et al. (2012) (p. 15) with replacing A to $\mathcal{X}^\top$, A' to $\mathcal{X}'^\top$, x to ξ and swapping n and d where necessary.

Next we prove an analogue of eq. 5 of Claim 4.3 of Blocki et al. (2012):

$$P_\xi\left(\frac{1}{2}|\xi^\top\big((\mathcal{X}\mathcal{X}^\top)^{-1} - (\mathcal{X}'\mathcal{X}'^\top)^{-1}\big)\xi| \geq \epsilon_0/2\right) \leq \delta_0. \tag{3}$$

To do this:

$$\begin{aligned}
&\xi^\top\big((\mathcal{X}\mathcal{X}^\top)^{-1} - (\mathcal{X}'\mathcal{X}'^\top)^{-1}\big)\xi \\
&= \xi^\top\big((\mathcal{X}\mathcal{X}^\top)^{-1} - (\mathcal{X}'\mathcal{X}'^\top)^{-1}\mathcal{X}\mathcal{X}^\top(\mathcal{X}\mathcal{X}^\top)^{-1}\big)\xi \\
&= \xi^\top\big((\mathcal{X}\mathcal{X}^\top)^{-1} - (\mathcal{X}'\mathcal{X}'^\top)^{-1}(\mathcal{X}'-E)(\mathcal{X}'-E)^\top(\mathcal{X}\mathcal{X}^\top)^{-1}\big)\xi \\
&= \xi^\top\big((\mathcal{X}\mathcal{X}^\top)^{-1} - (\mathcal{X}'\mathcal{X}'^\top)^{-1}(\mathcal{X}'\mathcal{X}'^\top - E\mathcal{X}'^\top - \mathcal{X}'E^\top + EE^\top)(\mathcal{X}\mathcal{X}^\top)^{-1}\big)\xi \\
&= \xi^\top\big((\mathcal{X}\mathcal{X}^\top)^{-1} - (\mathcal{X}\mathcal{X}^\top)^{-1} - (\mathcal{X}'\mathcal{X}'^\top)^{-1}(-E\mathcal{X}'^\top - \mathcal{X}'E^\top + EE^\top)(\mathcal{X}\mathcal{X}^\top)^{-1}\big)\xi \\
&= \xi^\top(\mathcal{X}'\mathcal{X}'^\top)^{-1}(E\mathcal{X}'^\top + \mathcal{X}'E^\top - EE^\top)(\mathcal{X}\mathcal{X}^\top)^{-1}\xi \\
&= \xi^\top(\mathcal{X}'\mathcal{X}'^\top)^{-1}(E\mathcal{X}^\top + \mathcal{X}'E^\top)(\mathcal{X}\mathcal{X}^\top)^{-1}\xi
\end{aligned} \tag{4}$$

where the second and the last equalities are due to $E = \mathcal{X}' - \mathcal{X}$. The expression in the last line of (4) is very similar to the one in the derivation of eq. 5 in Blocki et al. (2012) (p. 15). The difference is that in order for the proof to go through, we need to multiply $(\mathcal{X}'\mathcal{X}'^\top)^{-1}$ by $\mathcal{X}\mathcal{X}^\top(\mathcal{X}\mathcal{X}^\top)^{-1}$ in the second line of (4), while the original proof of Blocki et al. (2012) multiplies $(\mathcal{X}^\top\mathcal{X})^{-1}$ by $\mathcal{X}'^\top\mathcal{X}'(\mathcal{X}'^\top\mathcal{X}')^{-1}$ (in our notations), see eq. in the bottom of p. 15 of Blocki et al. (2012).

Now denoting singular value decompositions of $\mathcal{X} = U\Sigma V^\top$ and $\mathcal{X}' = U'\Lambda V'^\top$, and the fact that $E = e_{i^*}v^\top$, we continue (4):

$$\begin{aligned}
&\xi^\top(\mathcal{X}'\mathcal{X}'^\top)^{-1}(E\mathcal{X}^\top + \mathcal{X}'E^\top)(\mathcal{X}\mathcal{X}^\top)^{-1}\xi \\
&= \xi^\top(\mathcal{X}'\mathcal{X}'^\top)^{-1}E\mathcal{X}^\top(\mathcal{X}\mathcal{X}^\top)^{-1}\xi + \xi^\top(\mathcal{X}'\mathcal{X}'^\top)^{-1}\mathcal{X}'E^\top(\mathcal{X}\mathcal{X}^\top)^{-1}\xi \\
&= \xi^\top(U'\Lambda V'^\top V'\Lambda U'^\top)^{-1}(e_{i^*}\cdot v^\top V\Sigma U^\top)(U\Sigma V^\top V\Sigma U^\top)^{-1}\xi \\
&+\xi^\top(U'\Lambda V'^\top V'\Lambda U'^\top)^{-1}(U'\Lambda V'^\top v\cdot e_{i^*}^\top)(U\Sigma V^\top V\Sigma U^\top)^{-1}\xi \\
&= \xi^\top U'\Lambda^{-2}U'^\top e_{i^*}\cdot v^\top V\Sigma^{-1}U^\top\xi + \xi^\top U'\Lambda^{-1}V'^\top v\cdot e_{i^*}^\top U\Sigma^{-2}U^\top\xi
\end{aligned} \tag{5}$$

where the last equality is due to the properties of singular value decomposition.

So now, assume ξ is sampled from $\mathcal{X}'Y$ (the case of $\mathcal{X}Y$ is symmetric). That is, assume that we've sampled χ from $Y \sim \mathcal{N}(0, I_{d\times d})$ and we have $\xi = \mathcal{X}'\chi = U'\Lambda V'^\top\chi$ and equivalently $\xi = (\mathcal{X}+E)\chi = U\Sigma V^\top\chi + e_{i^*}v^\top\chi$. Plugging it into (5) gives:

$$\begin{aligned}
&|\xi^\top U'\Lambda^{-2}U'^\top e_{i^*}\cdot v^\top V\Sigma^{-1}U^\top\xi + \xi^\top U'\Lambda^{-1}V'^\top v\cdot e_{i^*}^\top U\Sigma^{-2}U^\top\xi| \\
&= |(U'\Lambda V'^\top\chi)^\top U'\Lambda^{-2}U'^\top e_{i^*}\cdot v^\top V\Sigma^{-1}U^\top(U\Sigma V^\top\chi + e_{i^*}v^\top\chi) \\
&+(U'\Lambda V'^\top\chi)^\top U'\Lambda^{-1}V'^\top v\cdot e_{i^*}^\top U\Sigma^{-2}U^\top(U\Sigma V^\top\chi + e_{i^*}v^\top\chi)| \\
&= |\chi^\top V'\Lambda U'^\top U'\Lambda^{-2}U'^\top e_{i^*}\cdot v^\top V\Sigma^{-1}U^\top(U\Sigma V^\top\chi + e_{i^*}v^\top\chi) \\
&+\chi^\top V'\Lambda U'^\top U'\Lambda^{-1}V'^\top v\cdot e_{i^*}^\top U\Sigma^{-2}U^\top(U\Sigma V^\top\chi + e_{i^*}v^\top\chi)| \\
&\leq term_1\cdot term_2 + term_3\cdot term_4
\end{aligned}$$

where for $i = 1,2,3,4$ we have $term_i = |vec_i\cdot\chi|$ and

$$\begin{aligned}
&vec_1 \\
&= (V'\Lambda U'^\top U'\Lambda^{-2}U'^\top e_{i^*})^\top \\
&= (V'\Lambda^{-1}U'^\top e_{i^*})^\top
\end{aligned}$$

so $\|vec_1\| \leq 1/\lambda_d$;

$$\begin{aligned}
&vec_2 \\
&= v^\top V\Sigma^{-1}U^\top(U\Sigma V^\top + e_{i^*}v^\top) \\
&= v^\top + v^\top V\Sigma^{-1}U^\top e_{i^*}v^\top
\end{aligned}$$

so $\|vec_2\| \leq 1 + 1/\sigma_d$;

$$\begin{aligned} &vec_3 \\ &= (V'\Lambda U'^\top U'\Lambda^{-1}V'^\top v)^\top \\ &= v^\top \end{aligned}$$

so $\|vec_3\| \leq 1$;

$$\begin{aligned} &vec_4 \\ &= e_{i^*}^\top U\Sigma^{-2}U^\top(U\Sigma V^\top + e_{i^*}v^\top) \\ &= e_{i^*}^\top U\Sigma^{-1}V^\top + e_{i^*}^\top U\Sigma^{-2}U^\top e_{i^*}v^\top \end{aligned}$$

so $\|vec_4\| \leq 1/\sigma_d + 1/\sigma_d^2$ where σ_d and λ_d are the smallest singular values of $\mathcal{X}$ and $\mathcal{X}'$, respectively. The remainder of the proof now follows the proof of Claim 4.3 of Blocki et al. (2012) with replacing A to $\mathcal{X}^\top$, A' to $\mathcal{X}'^\top$, x to ξ and swapping n and d where necessary. □

For the second stage we assume that "else" clause (line 8 of Algorithm 1) is applied and denote $\tilde{\mathcal{X}} \triangleq U\sqrt{\Sigma^2 + \omega^2 I_{n\times d}}V^\top$ and $\tilde{\mathcal{X}}' \triangleq U'\sqrt{\Lambda^2 + \omega^2 I_{n\times d}}V'^\top$. The theorem requires an analogue of Lemma 2 to hold, which depends on the following two conditions:

$$\exp(-\epsilon_0/2) \leq \sqrt{\frac{\det(\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)}{\det(\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)}} \leq \exp(\epsilon_0/2). \tag{6}$$

$$P_\xi\left(\frac{1}{2}|\xi^\top\big((\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1} - (\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)^{-1}\big)\xi| \geq \epsilon_0/2\right) \leq \delta_0. \tag{7}$$

Derivation of (6) copies the derivation of eq. 6 in Blocki et al. (2012) (p. 16). To derive (7), we start with an observation regarding $\mathcal{X}'\mathcal{X}'^\top$ and $\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top$:

$$\begin{aligned} &\mathcal{X}'\mathcal{X}'^\top = (\mathcal{X}+E)(\mathcal{X}+E)^\top = \mathcal{X}\mathcal{X}^\top + \mathcal{X}'E^\top + E\mathcal{X}^\top \\ &\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top = U(\Sigma^2+\omega^2 I)U^\top = U\Sigma^2U^\top + \omega^2 I = \mathcal{X}\mathcal{X}^\top + \omega^2 I \\ &\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top = U'(\Lambda^2+\omega^2 I)U'^\top = U'\Lambda^2U'^\top + \omega^2 I = \mathcal{X}'\mathcal{X}'^\top + \omega^2 I \\ &\implies \tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top - \tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top = \mathcal{X}'E^\top + E\mathcal{X}^\top. \end{aligned} \tag{8}$$

Now we can follow the same outline as in the proof of (3). Fix ξ, then

$$\begin{aligned} &\xi^\top\big((\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1} - (\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)^{-1}\big)\xi \\ &= \xi^\top\big((\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1} - (\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)^{-1}\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top(\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1}\big)\xi \\ &= \xi^\top\big((\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1} - (\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)^{-1}(\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top - \mathcal{X}'E^\top - E\mathcal{X}^\top)(\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1}\big)\xi \\ &= \xi^\top\big((\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1} - (\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1} - (\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)^{-1}(-\mathcal{X}'E^\top - E\mathcal{X}^\top)(\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1}\big)\xi \\ &= \xi^\top(\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)^{-1}(\mathcal{X}'E^\top + E\mathcal{X}^\top)(\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1}\xi \\ &= \xi^\top(\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)^{-1}(\mathcal{X}'E^\top - EE^\top + EE^\top + E\mathcal{X}^\top)(\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1}\xi \\ &= \xi^\top(\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)^{-1}((\mathcal{X}'-E)E^\top + E(\mathcal{X}^\top + E^\top))(\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1}\xi \\ &= \xi^\top(\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)^{-1}(\mathcal{X}'-E)v\cdot e_{i^*}^\top(\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1}\xi \\ &+\xi^\top(\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)^{-1}e_{i^*}\cdot v^\top(\mathcal{X}^\top + E^\top)(\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1}\xi \end{aligned} \tag{9}$$

where the second equality follows from (8) and the last equality follows from $E = e_{i^*}v^\top$. The expression in the last line of (9) is very similar to the one in the derivation of equation in Blocki et al. (2012) (p. 17, second equation array from the top). The difference is that in order for the proof to go trhough, we need to multiply $(\tilde{\mathcal{X}}'\tilde{\mathcal{X}}'^\top)^{-1}$ by $\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top(\tilde{\mathcal{X}}\tilde{\mathcal{X}}^\top)^{-1}$ in the second line of (9), while the original proof of Blocki et al. (2012) multiplies $(\tilde{\mathcal{X}}^\top\tilde{\mathcal{X}})^{-1}$ by $\tilde{\mathcal{X}}'^\top\tilde{\mathcal{X}}'(\tilde{\mathcal{X}}'^\top\tilde{\mathcal{X}}')^{-1}$ (in our notations), see second equation array from the top, p. 17 of Blocki et al. (2012). The remainder of the proof now follows the proof of Theorem 4.1 of Blocki et al. (2012) (p. 17).

B.3. Proof of Theorem 2

Proof. Fix $x, x' \in \mathcal{X}$ and their images $z, z' \in \mathcal{Z}$. If $\sigma_{min}(\mathcal{X}) \geq \omega$, according to Algorithm 1, $\mathcal{Z} = r^{-1/2}\mathcal{X}M$ (line 7) and

$$\begin{aligned}&\|z - z'\|^2\\ &= \|r^{-1/2}xM - r^{-1/2}x'M\|^2\\ &= r^{-1}\|xM - x'M\|^2\end{aligned}$$

and Lemma 1 can be immediately applied.

If $\sigma_{min}(\mathcal{X}) < \omega$, according to Algorithm 1, $\mathcal{Z} = r^{-1/2}\tilde{\mathcal{X}}M$ (line 10) and

$$\begin{aligned}&\|z - z'\|^2\\ &= \|r^{-1/2}\tilde{x}M - r^{-1/2}\tilde{x}'M\|^2\\ &= r^{-1}\|\tilde{x}M - \tilde{x}'M\|^2\\ &\leq (1+\nu)\|\tilde{x} - \tilde{x}'\|^2\\ &\leq (1+\nu)(1+\omega^2/\sigma^2_{min}(\mathcal{X}))\|x - x'\|^2\end{aligned}$$

where the first inequality follows from Lemma 1 and the second inequality follows from Lemma 7. Similarly,

$$\begin{aligned}&\|z - z'\|^2\\ &= \|r^{-1/2}\tilde{x}M - r^{-1/2}\tilde{x}'M\|^2\\ &= r^{-1}\|\tilde{x}M - \tilde{x}'M\|^2\\ &\geq (1-\nu)\|\tilde{x} - \tilde{x}'\|^2\\ &\geq (1-\nu)\|x - x'\|^2\end{aligned}$$

where the first inequality follows from Lemma 1 and the second inequality follows from Lemma 7. □

B.4. Bounding the covariance change

Theorem 5. *Let a dataset $\mathcal{X} \subset \mathbb{R}^d$ be given and $\sigma_{min}(\mathcal{X}) > 0$ be the smallest singular value of $\mathcal{X}$. Let $r \in \mathbb{N}$ be the input parameter of Algorithm 1, a dataset $\mathcal{Z} \subset \mathbb{R}^r$ be the output of Algorithm 1 and ω be defined in line 5 of Algorithm 1. Let $\mathrm{d} = diam(\mathcal{X})/l$ where $diam(\mathcal{X})$ is the diameter of the dataset $\mathcal{X}$. Let $\nu \in (0, 1/2)$, $\mu \in (0, 1)$ be given. If $\nu \leq 2/\mathrm{d}^2$ and $r \geq 8\log(n^2/\mu)/\nu^2$, then the probability of*

$$|k_{zz'} - k_{xx'}| \leq C \cdot k_{xx'}$$

for all $x, x' \in \mathcal{X}$ and their images under Algorithm 1 $z, z' \in \mathcal{Z}$ is at least $1 - \mu$ where

$$C \triangleq \begin{cases}\nu\mathrm{d}^2 & \text{if } \sigma_{min}(\mathcal{X}) \geq \omega,\\ \max\left(\nu\mathrm{d}^2, 1 - \exp\left(-0.5(\nu + \nu\omega^2/\sigma^2_{min}(\mathcal{X}) + \omega^2/\sigma^2_{min}(\mathcal{X}))\mathrm{d}^2\right)\right) & \text{otherwise.}\end{cases} \tag{10}$$

Remark 5. It immediately follows from Theorem 5 that the probability of $k_{zz'} \leq (1 + C) \cdot k_{xx'}$ for all $x, x' \in \mathcal{X}$ and their images $z, z' \in \mathcal{Z}$ is at least $1 - \mu$.

Proof.

$$\begin{aligned}&k_{zz'} - k_{xx'}\\ &= \sigma_y^2 \exp\left(-0.5\|z - z'\|^2/l^2\right) - \sigma_y^2 \exp\left(-0.5\|x - x'\|^2/l^2\right)\\ &\leq \sigma_y^2 \exp\left(-0.5(1-\nu)\|x - x'\|^2/l^2\right) - \sigma_y^2 \exp\left(-0.5\|x - x'\|^2/l^2\right)\\ &= k_{xx'}\left(\exp\left(0.5\nu\|x - x'\|^2/l^2\right) - 1\right)\\ &\leq k_{xx'}\left(2 \cdot \left(0.5\nu\|x - x'\|^2/l^2\right)\right)\\ &\leq k_{xx'} \cdot \nu\mathrm{d}^2\end{aligned}$$

where the first inequality follows from Theorem 2 (since the condition $(1-\nu)\|x - x'\|^2 \leq \|z - z'\|^2$ holds in both cases $\sigma_{min}(\mathcal{X}) \geq \omega$ and otherwise), and the second inequality follows from the identity $\exp c \leq 1 + 2c$ for $c \in (0, 1)$ by setting $c = 0.5\nu\|x - x'\|^2/l^2$ since $\nu \leq 2/\mathrm{d}^2$ and

$$\begin{aligned}&0.5\nu\|x - x'\|^2/l^2\\ &\leq 0.5\nu\ (\mathrm{diam}(\mathcal{X}))^2/l^2\\ &\leq 0.5 \cdot 2/\mathrm{d}^2 \cdot (\mathrm{diam}(\mathcal{X}))^2/l^2\\ &= 1.\end{aligned} \tag{11}$$

If $\sigma_{min}(\mathcal{X}) \geq \omega$,

$$\begin{aligned}
&k_{xx'} - k_{zz'} \\
&= \sigma_y^2 \exp\left(-0.5\|x - x'\|^2/l^2\right) - \sigma_y^2 \exp\left(-0.5\|z - z'\|^2/l^2\right) \\
&\leq \sigma_y^2 \exp\left(-0.5\|x - x'\|^2/l^2\right) - \sigma_y^2 \exp\left(-0.5(1+\nu)\|x - x'\|^2/l^2\right) \\
&= k_{xx'}\left(1 - \exp\left(-0.5\nu\|x - x'\|^2/l^2\right)\right) \\
&= k_{xx'}\left(\exp\left(0.5\nu\|x - x'\|^2/l^2\right) - 1\right)\exp\left(-0.5\nu\|x - x'\|^2/l^2\right) \\
&\leq k_{xx'}\left(\exp\left(0.5\nu\|x - x'\|^2/l^2\right) - 1\right) \\
&\leq k_{xx'}\left(2 \cdot \left(0.5\nu\|x - x'\|^2/l^2\right)\right) \\
&\leq k_{xx'} \cdot \nu \mathrm{d}^2
\end{aligned}$$

where the first inequality follows from Theorem 2, since if $\sigma_{min}(\mathcal{X}) \geq \omega$, $C' = 1$ in the statement of Theorem 2, the second inequality follows from $0.5\nu\|x - x'\|^2/l^2 \geq 0$ and the third inequality follows from the identity $\exp c \leq 1 + 2c$ for $c \in (0, 1)$ by setting $c = 0.5\nu\|x - x'\|^2/l^2$ and (11).

Similarly, if $\sigma_{min}(\mathcal{X}) < \omega$,

$$\begin{aligned}
&k_{xx'} - k_{zz'} \\
&= \sigma_y^2 \exp\left(-0.5\|x - x'\|^2/l^2\right) - \sigma_y^2 \exp\left(-0.5\|z - z'\|^2/l^2\right) \\
&\leq \sigma_y^2 \exp\left(-0.5\|x - x'\|^2/l^2\right) - \sigma_y^2 \exp\left(-0.5(1+\nu)(1+\omega^2/\sigma_{min}^2(\mathcal{X}))\|x - x'\|^2/l^2\right) \\
&= k_{xx'}\left(1 - \exp\left(-0.5(\nu + \nu\omega^2/\sigma_{min}^2(\mathcal{X}) + \omega^2/\sigma_{min}^2(\mathcal{X}))\|x - x'\|^2/l^2\right)\right) \\
&\leq k_{xx'}\left(1 - \exp\left(-0.5(\nu + \nu\omega^2/\sigma_{min}^2(\mathcal{X}) + \omega^2/\sigma_{min}^2(\mathcal{X}))\mathrm{d}^2\right)\right)
\end{aligned}$$

where the first inequality follows from Theorem 2, since if $\sigma_{min}(\mathcal{X}) < \omega$, $C' = 1 + \omega^2/\sigma_{min}^2(\mathcal{X})$ in the statement of Theorem 2.

□

B.5. Proof of Theorem 3

First we recall and introduce a few notations which we will use throughout this section. Let $\mathcal{X} \subset \mathbb{R}^d$ be a dataset and its image under Algorithm 1 be a dataset $\mathcal{Z} \subset \mathbb{R}^r$, $\mathcal{Z}_{t-1} \triangleq \{z_1, \ldots, z_{t-1}\}$ be a set of transformed inputs selected by Algorithm 2 run on transformed dataset $\mathcal{Z}$ after $t-1$ iterations and the preimage of $\mathcal{Z}_{t-1}$ under Algorithm 1 be a set $\mathcal{X}_{t-1} \triangleq \{x_1, \ldots, x_{t-1}\}$. Let $z \in \mathcal{Z}$ be an (unobserved) transformed input and $x \in \mathcal{X}$ be its preimage under Algorithm 1. Let f be a latent function sampled from a GP. Define

$$\begin{aligned}
&\tilde{f}(z) \triangleq f(x) \\
&\alpha_t(x, \mathcal{X}_{t-1}) \triangleq \mu_t(x) + \beta_t^{1/2}\sigma_t(x) \\
&\alpha_t(z, \mathcal{Z}_{t-1}) \triangleq \tilde{\mu}_t(z) + \beta_t^{1/2}\tilde{\sigma}_t(z) \\
&z_t \triangleq \operatorname*{argmax}_{z \in \mathcal{Z}} \alpha_t(z, \mathcal{Z}_{t-1}).
\end{aligned} \tag{12}$$

That is, $\tilde{f}$ is the latent function f defined over the transformed dataset $\mathcal{Z}$, $\alpha_t(z, \mathcal{Z}_{t-1})$ is the function maximized by Algorithm 2 at iteration t, $\alpha_t(x, \mathcal{X}_{t-1})$ is the function maximized by GP-UCB algorithm run on the original dataset, z_t is the transformed input selected by Algorithm 2 at iteration t and x_t is the preimage of z_t under Algorithm 1.

Lemma 3. *Let $\delta' \in (0, 1)$ be given and $\beta_t \triangleq 2\log(nt^2\pi^2/6\delta')$. Then*

$$|f(x) - \mu_t(x)| \leq \beta_t^{1/2}\sigma_t(x) \;\; \forall x \in \mathcal{X} \;\; \forall t \in \mathbb{N}$$

holds with probability at least $1 - \delta'$.

Proof. Lemma 3 above corresponds to Lemma 5.1 in Srinivas et al. (2010); see its proof therein. □

Lemma 4. *Let $\delta' \in (0, 1)$ be given and $\beta_t \triangleq 2\log(nt^2\pi^2/6\delta')$. Then the probability of*

$$\tilde{f}(z^*) - \tilde{f}(z_t) \leq 2\max_{x,z} |\alpha_t(z, \mathcal{Z}_{t-1}) - \alpha_t(x, \mathcal{X}_{t-1})| + 2\beta_t^{1/2}\sigma_t(x_t)$$

for all $t \in \mathbb{N}$ is at least $1 - \delta'$ where z^* is the maximizer of $\tilde{f}$ and $x \in \mathcal{X}$ is the preimage of $z \in \mathcal{Z}$ under Algorithm 1.

Proof.

$$\begin{aligned}
&\tilde{f}(z^*) - \tilde{f}(z_t)\\
&= f(x^*) - f(x_t)\\
&\leq \alpha_t(x^*, \mathcal{X}_{t-1}) - f(x_t)\\
&= \alpha_t(x^*, \mathcal{X}_{t-1}) - \alpha_t(z^*, \mathcal{Z}_{t-1}) + \alpha_t(z^*, \mathcal{Z}_{t-1}) - f(x_t)\\
&\leq \alpha_t(x^*, \mathcal{X}_{t-1}) - \alpha_t(z^*, \mathcal{Z}_{t-1}) + \alpha_t(z_t, \mathcal{Z}_{t-1}) - f(x_t)\\
&= \alpha_t(x^*, \mathcal{X}_{t-1}) - \alpha_t(z^*, \mathcal{Z}_{t-1}) + \alpha_t(z_t, \mathcal{Z}_{t-1}) - \alpha_t(x_t, \mathcal{X}_{t-1}) + \alpha_t(x_t, \mathcal{X}_{t-1}) - f(x_t)\\
&\leq 2\max_{x,z} |\alpha_t(z, \mathcal{Z}_{t-1}) - \alpha_t(x, \mathcal{X}_{t-1})| + \alpha_t(x_t, \mathcal{X}_{t-1}) - f(x_t)\\
&\leq 2\max_{x,z} |\alpha_t(z, \mathcal{Z}_{t-1}) - \alpha_t(x, \mathcal{X}_{t-1})| + 2\beta_t^{1/2}\sigma_t(x_t)
\end{aligned}$$

where the first equality is due to (12) and x^* is the maximizer of f, the first and the last inequalities are due to Lemma 3 and the second inequality is due to the choice of z_t in (12). □

Lemma 4 resembles Lemma 5.2 of Srinivas et al. (2010) with an added term $2\max_{x,z} |\alpha_t(z, \mathcal{Z}_{t-1}) - \alpha_t(x, \mathcal{X}_{t-1})|$. It suggests that in order to bound regret $\tilde{f}(z^*) - \tilde{f}(z_t)$ incurred by Algorithm 2 at iteration t, we need to bound $|\alpha_t(z, \mathcal{Z}_{t-1}) - \alpha_t(x, \mathcal{X}_{t-1})|$. Using the diagonal dominance assumption (Definition 3), we do it in the following two lemmas:

Lemma 5. *Let $C > 0$ be given. If for all $x, x' \in \mathcal{X}$ and their images under Algorithm 1 $z, z' \in \mathcal{Z}$ holds $|k_{zz'} - k_{xx'}| \leq C \cdot k_{xx'}$, for all $t = 1, \ldots, T$ matrix $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$ is diagonally dominant, then for every unobserved transformed input $z \in \mathcal{Z}$ and its preimage under Algorithm 1 $x \in \mathcal{X}$*

$$|\tilde{\sigma}_t^2(z) - \sigma_t^2(x)| \leq C_1/\sqrt{|\mathcal{X}_{t-1}|}$$

where

$$C_1 \triangleq C\sigma_y\sqrt{2\sigma_y^2 + \sigma_n^2}\Big(\sqrt{2}(1+C)^2\sigma_y^2/\sigma_n^2 + (2+C)C\Big).$$

Proof.

$$\begin{aligned}
&|\tilde{\sigma}_t^2(z) - \sigma_t^2(x)|\\
&= |\big(k_{zz} - K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{Z}_{t-1}z}\big) - \big(k_{xx} - K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{X}_{t-1}x}\big)|\\
&= |K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{Z}_{t-1}z} - K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{X}_{t-1}x}|\\
&\leq |K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{Z}_{t-1}z} - K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{Z}_{t-1}z}|\\
&+ |K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{Z}_{t-1}z} - K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{Z}_{t-1}z}|\\
&+ |K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{Z}_{t-1}z} - K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{X}_{t-1}x}|\\
&\leq (1+C)^2\|K_{x\mathcal{X}_{t-1}}\| \cdot \sigma_y^2/\sigma_n^2 \cdot \sqrt{2}C/\sqrt{|\mathcal{X}_{t-1}|} + (2+C)C \cdot \|K_{x\mathcal{X}_{t-1}}\|/\sqrt{|\mathcal{X}_{t-1}|}\\
&= C\|K_{x\mathcal{X}_{t-1}}\|/\sqrt{|\mathcal{X}_{t-1}|}\Big(\sqrt{2}(1+C)^2\sigma_y^2/\sigma_n^2 + (2+C)C\Big)\\
&\leq C\sigma_y\sqrt{2\sigma_y^2 + \sigma_n^2}/\sqrt{|\mathcal{X}_{t-1}|}\Big(\sqrt{2}(1+C)^2\sigma_y^2/\sigma_n^2 + (2+C)C\Big)
\end{aligned} \tag{13}$$

where the first equality is due to (1), the second equality is due to $k_{xx} = k_{zz} = \sigma_y^2$ for every x and z, the first inequality is due to triangle inequality, the second inequality is due to

$$\begin{aligned}
&|K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{Z}_{t-1}z} - K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}K_{\mathcal{Z}_{t-1}z}|\\
&= |K_{z\mathcal{Z}_{t-1}}\big((K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1} - (K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\big)K_{\mathcal{Z}_{t-1}z}|\\
&\leq \|K_{z\mathcal{Z}_{t-1}}\|^2 \cdot \|(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1} - (K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2\\
&\leq (1+C)^2\|K_{x\mathcal{X}_{t-1}}\|^2 \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1} - (K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2\\
&\leq (1+C)^2\|K_{x\mathcal{X}_{t-1}}\|^2 \cdot \|(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})\|_2 \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2\\
&\leq (1+C)^2\|K_{x\mathcal{X}_{t-1}}\|^2 \cdot \|(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}\|_2 \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2\\
&\leq (1+C)^2\|K_{x\mathcal{X}_{t-1}}\|^2 \cdot 1/\sigma_n^2 \cdot \|K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}\|_2 \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2\\
&\leq (1+C)^2\|K_{x\mathcal{X}_{t-1}}\|^2 \cdot 1/\sigma_n^2 \cdot \sqrt{2}C\sigma_y^2/\sqrt{|\mathcal{X}_{t-1}|} \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2\\
&\leq (1+C)^2\|K_{x\mathcal{X}_{t-1}}\|^2 \cdot 1/\sigma_n^2 \cdot \sqrt{2}C\sigma_y^2/\sqrt{|\mathcal{X}_{t-1}|} \cdot 1/(\sqrt{|\mathcal{X}_{t-1}|}\|K_{x\mathcal{X}_{t-1}}\|)\\
&= (1+C)^2\|K_{x\mathcal{X}_{t-1}}\| \cdot \sigma_y^2/\sigma_n^2 \cdot \sqrt{2}C/|\mathcal{X}_{t-1}|\\
&\leq (1+C)^2\|K_{x\mathcal{X}_{t-1}}\| \cdot \sigma_y^2/\sigma_n^2 \cdot \sqrt{2}C/\sqrt{|\mathcal{X}_{t-1}|}
\end{aligned}$$

where the first inequality is due to property of quadratic forms $|v^\top A v| \leq \|v\|^2 \cdot \|A\|_2$ for any vector v (see Theorem 2.11, Section II.2.2 in Stewart & Sun (1990)), the second inequality follows from the statement of the lemma and Remark 5 to Theorem 5, the third inequality follows from Theorem 2.5 (see Section III.2.2 in Stewart & Sun (1990)), the fourth inequality is due to the submultiplicativity of the spectral norm (see Section II.2.2, p. 69 in Stewart & Sun (1990)), the fifth inequality follows from Lemma 8, the sixth inequality follows from Lemma 9, the second last inequality follows from Lemma 10 and the last inequality follows from $|\mathcal{X}_{t-1}| \geq 1$;

and

$$
\begin{aligned}
&|K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1} K_{\mathcal{Z}_{t-1}z} - K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1} K_{\mathcal{Z}_{t-1}z}| \\
&+ |K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1} K_{\mathcal{Z}_{t-1}z} - K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1} K_{\mathcal{X}_{t-1}x}| \\
&= |(K_{z\mathcal{Z}_{t-1}} - K_{x\mathcal{X}_{t-1}})(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1} K_{\mathcal{Z}_{t-1}z}| \\
&+ |K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}(K_{\mathcal{Z}_{t-1}z} - K_{\mathcal{X}_{t-1}x})| \\
&\leq \|K_{z\mathcal{Z}_{t-1}} - K_{x\mathcal{X}_{t-1}}\| \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|K_{\mathcal{Z}_{t-1}z}\| \\
&+ \|K_{x\mathcal{X}_{t-1}}\| \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|K_{\mathcal{Z}_{t-1}z} - K_{\mathcal{X}_{t-1}x}\| \\
&\leq (1 + 1 + C) \cdot \|K_{z\mathcal{Z}_{t-1}} - K_{x\mathcal{X}_{t-1}}\| \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|K_{\mathcal{X}_{t-1}x}\| \\
&\leq (2 + C) \cdot C\|K_{x\mathcal{X}_{t-1}}\| \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|K_{x\mathcal{X}_{t-1}}\| \\
&\leq (2 + C) \cdot C\|K_{x\mathcal{X}_{t-1}}\| \cdot 1/(\sqrt{|\mathcal{X}_{t-1}|}\|K_{x\mathcal{X}_{t-1}}\|) \cdot \|K_{x\mathcal{X}_{t-1}}\| \\
&= (2 + C)C \cdot \|K_{x\mathcal{X}_{t-1}}\|/\sqrt{|\mathcal{X}_{t-1}|}
\end{aligned}
$$

where the first inequality is due to property of bilinear forms $|u^\top A v| \leq \|u\| \cdot \|A\|_2 \cdot \|v\|$ for any vectors u, v (see Theorem 2.11, Section II.2.2 in Stewart & Sun (1990)), the second and the third inequalities follow from the statement of the lemma and Remark 5 to Theorem 5 and the last inequality follows from Lemma 10.

The last inequality in (13) follows from

$$
\begin{aligned}
&\|K_{x\mathcal{X}_{t-1}}\|^2 \\
&= \|K_{x\mathcal{X}_{t-1}}\|^2 \cdot \psi_{max}^{-1}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I) \cdot \psi_{max}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I) \\
&= \|K_{x\mathcal{X}_{t-1}}\|^2 \cdot \psi_{min}((K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}) \cdot \psi_{max}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I) \\
&= \|K_{x\mathcal{X}_{t-1}}\|^2 \cdot \psi_{min}((K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}) \cdot \|K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I\|_2 \\
&= \|K_{x\mathcal{X}_{t-1}}\|^2 \cdot \psi_{min}((K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}) \cdot (\|K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}\|_2 + \sigma_n^2) \\
&\leq \|K_{x\mathcal{X}_{t-1}}\|^2 \cdot \psi_{min}((K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}) \cdot (2\sigma_y^2 + \sigma_n^2) \\
&\leq K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1} K_{\mathcal{X}_{t-1}x} \cdot (2\sigma_y^2 + \sigma_n^2) \\
&\leq k_{xx} \cdot (2\sigma_y^2 + \sigma_n^2) \\
&= \sigma_y^2(2\sigma_y^2 + \sigma_n^2)
\end{aligned}
$$

where $\psi_{max}(\cdot)$ and $\psi_{min}(\cdot)$ denote the largest and the smallest eigenvalues of a matrix, respectively, the first fourth equalities are properties of eigenvalues, the first inequality is due to Lemma 11, the second inequality follows from Lemma 12, the third inequality follows from the fact that conditioning does not increase variance and the last equality is due to $k_{xx} = \sigma_y^2$. □

Lemma 6. *Let $C > 0$ be given. If for all $x, x' \in \mathcal{X}$ and their images under Algorithm 1 $z, z' \in \mathcal{Z}$ holds $|k_{zz'} - k_{xx'}| \leq C \cdot k_{xx'}$, for all $t = 1, \ldots, T$ matrix $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$ is diagonally dominant and $|y_t| \leq L$, then for every unobserved transformed input $z \in \mathcal{Z}$ and its preimage under Algorithm 1 $x \in \mathcal{X}$*

$$|\tilde{\mu}_t(z) - \mu_t(x)| \leq CL + C_2/\sqrt{|\mathcal{X}_{t-1}|}$$

where

$$C_2 = \sqrt{2}(1 + C) \cdot C\sigma_y^2/\sigma_n^2 \cdot L.$$

Proof.

$$
\begin{aligned}
&|\tilde{\mu}_t(z) - \mu_t(x)| \\
&= |K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}\mathbf{y}_{t-1} - K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\mathbf{y}_{t-1}| \\
&\leq |K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\mathbf{y}_{t-1} - K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\mathbf{y}_{t-1}| \\
&+ |K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}\mathbf{y}_{t-1} - K_{z\mathcal{Z}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\mathbf{y}_{t-1}| \\
&= |(K_{z\mathcal{Z}_{t-1}} - K_{x\mathcal{X}_{t-1}})(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\mathbf{y}_{t-1}| \\
&+ |K_{z\mathcal{Z}_{t-1}}\big((K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1} - (K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\big)\mathbf{y}_{t-1}| \\
&\leq C \cdot L + C_2/\sqrt{|\mathcal{X}_{t-1}|}
\end{aligned}
$$

where the first equality is due to (1), the first inequality is due to triangle inequality and the second inequality follows from

$$\begin{aligned}
&|(K_{z\mathcal{Z}_{t-1}} - K_{x\mathcal{X}_{t-1}})(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\mathbf{y}_{t-1}| \\
&\leq \|K_{z\mathcal{Z}_{t-1}} - K_{x\mathcal{X}_{t-1}}\| \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|\mathbf{y}_{t-1}\| \\
&\leq C\|K_{x\mathcal{X}_{t-1}}\| \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|\mathbf{y}_{t-1}\| \\
&\leq C\|K_{x\mathcal{X}_{t-1}}\| \cdot 1/(\sqrt{|\mathcal{X}_{t-1}|}\|K_{x\mathcal{X}_{t-1}}\|) \cdot \|\mathbf{y}_{t-1}\| \\
&\leq C \cdot L
\end{aligned}$$

where the first inequality is due to property of bilinear forms $|u^\top A v| \leq \|u\| \cdot \|A\|_2 \cdot \|v\|$ for any vectors u, v (see Theorem 2.11, Section II.2.2 in Stewart & Sun (1990)), the second inequality follows from the statement of the lemma, the third inequality follows from Lemma 10 and the last inequality follows from the condition $|y_t| \leq L$ for all $t = 1, \ldots, T$;

and

$$\begin{aligned}
&|K_{z\mathcal{Z}_{t-1}}\big((K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1} - (K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\big)\mathbf{y}_{t-1}| \\
&\leq \|K_{z\mathcal{Z}_{t-1}}\| \cdot \|(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1} - (K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|\mathbf{y}_{t-1}\| \\
&\leq \|K_{z\mathcal{Z}_{t-1}}\| \cdot \|(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|\mathbf{y}_{t-1}\| \\
&\leq \|K_{z\mathcal{Z}_{t-1}}\| \cdot \|(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}\|_2 \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|\mathbf{y}_{t-1}\| \\
&\leq \|K_{z\mathcal{Z}_{t-1}}\| \cdot 1/\sigma_n^2 \cdot \|K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}\|_2 \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|\mathbf{y}_{t-1}\| \\
&\leq \|K_{z\mathcal{Z}_{t-1}}\| \cdot 1/\sigma_n^2 \cdot \sqrt{2}C\sigma_y^2/\sqrt{|\mathcal{X}_{t-1}|} \cdot \|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \cdot \|\mathbf{y}_{t-1}\| \\
&\leq \|K_{z\mathcal{Z}_{t-1}}\| \cdot 1/\sigma_n^2 \cdot \sqrt{2}C\sigma_y^2/\sqrt{|\mathcal{X}_{t-1}|} \cdot 1/(\sqrt{|\mathcal{X}_{t-1}|}\|K_{x\mathcal{X}_{t-1}}\|) \cdot \|\mathbf{y}_{t-1}\| \\
&\leq (1+C)\|K_{x\mathcal{X}_{t-1}}\| \cdot 1/\sigma_n^2 \cdot \sqrt{2}C\sigma_y^2/\sqrt{|\mathcal{X}_{t-1}|} \cdot 1/(\sqrt{|\mathcal{X}_{t-1}|}\|K_{x\mathcal{X}_{t-1}}\|) \cdot \|\mathbf{y}_{t-1}\| \\
&\leq \sqrt{2}(1+C) \cdot C\sigma_y^2/\sigma_n^2 \cdot L/\sqrt{|\mathcal{X}_{t-1}|} \\
&= C_2/\sqrt{|\mathcal{X}_{t-1}|}
\end{aligned}$$

where the first inequality is due to property of bilinear forms $|u^\top A v| \leq \|u\| \cdot \|A\|_2 \cdot \|v\|$ for any vectors u, v (see Theorem 2.11, Section II.2.2 in Stewart & Sun (1990)), the second inequality follows from Theorem 2.5 (see Section III.2.2 in Stewart & Sun (1990)), the third inequality is due to the submultiplicativity of the spectral norm (see Section II.2.2, p. 69 in Stewart & Sun (1990)) the fourth inequality follows from Lemma 8, the fifth inequality follows from Lemma 9, the third last inequality follows from Lemma 10, the second last inequality follows from the statement of the lemma and Remark 5 to Theorem 5 and the last inequality follows from the condition $|y_t| \leq L$ for all $t = 1, \ldots, T$. □

Proof of the theorem. By Lemma 4 for $\delta' = \delta_{ucb}/2$ and $\beta_t = 2\log(nt^2\pi^2/3\delta_{ucb})$ for all $t \in \mathbb{N}$:

$$\begin{aligned}
&r_t \\
&= f(x^*) - f(x_t) \\
&= \tilde{f}(z^*) - \tilde{f}(z_t) \\
&\leq 2\max_{x,z} |\alpha_t(z, \mathcal{Z}_{t-1}) - \alpha_t(x, \mathcal{X}_{t-1})| + 2\beta_t^{1/2}\sigma_t(x_t) \\
&\leq 2\max_{x,z} |\tilde{\mu}_t(z) - \mu_t(x)| + 2\beta_t^{1/2}\max_{x,z} |\tilde{\sigma}_t^2(z) - \sigma_t^2(x)| + 2\beta_t^{1/2}\sigma_t(x_t)
\end{aligned} \tag{14}$$

with probability at least $1 - \delta_{ucb}/2$ where the second equality follows from (12), the first inequality follows from Lemma 4 and the second inequality follows from triangle inequality. Suppose $\nu \in (0, \min(1/2, 2/\mathrm{d}^2))$, $\mu \in (0, 1)$ are given (we will set the exact values of μ, ν later) and the input parameter of Algorithm 1 $r \geq 8\log(n^2/\mu)/\nu^2$. By Theorem 5 for all $x, x' \in \mathcal{X}$ and their images under Algorithm 1 $z, z' \in \mathcal{Z}$ holds $|k_{zz'} - k_{xx'}| \leq C \cdot k_{xx'}$ with probability at least $1 - \mu$. Let $\mu = \delta_{ucb}/2$. Then we can apply Lemma 5 and Lemma 6 to (14). Using the union bound we obtain that for all $t = 1, \ldots, T$

$$\begin{aligned}
&r_t \\
&\leq 2\max_{x,z} |\tilde{\mu}_t(z) - \mu_t(x)| + 2\beta_t^{1/2}\max_{x,z} |\tilde{\sigma}_t^2(z) - \sigma_t^2(x)| + 2\beta_t^{1/2}\sigma_t(x_t) \\
&\leq 2(CL + C_2/\sqrt{|\mathcal{X}_{t-1}|}) + 2C_1\beta_t^{1/2}/\sqrt{|\mathcal{X}_{t-1}|} + 2\beta_t^{1/2}\sigma_t(x_t)
\end{aligned} \tag{15}$$

with probability at least $1 - \delta_{ucb}$ where C_1 and C_2 are defined in Lemma 5 and Lemma 6, respectively. Summing over

$t = 1, \ldots, T$:

$$\begin{aligned}
&\sum_{t=1}^{T} r_t^2 \\
&\leq 4\sum_{t=1}^{T} \left(CL + C_2/\sqrt{|\mathcal{X}_{t-1}|} + C_1\beta_t^{1/2}/\sqrt{|\mathcal{X}_{t-1}|} + \beta_t^{1/2}\sigma_t(x_t)\right)^2 \\
&\leq 12\sum_{t=1}^{T} \left(C^2L^2 + (C_2 + C_1\beta_t^{1/2})^2/|\mathcal{X}_{t-1}| + \beta_t\sigma_t^2(x_t)\right) \\
&= 12C^2L^2T + 12\sum_{t=1}^{T}(C_2 + C_1\beta_t^{1/2})^2|\mathcal{X}_{t-1}| + 12\sum_{t=1}^{T}\beta_t\sigma_t^2(x_t) \\
&\leq 12C^2L^2T + 24(C_2 + C_1\beta_T^{1/2})^2\log T + 12\beta_T\sum_{t=1}^{T}\sigma_t^2(x_t) \\
&\leq 12C^2L^2T + 24(C_2 + C_1\beta_T^{1/2})^2\log T + 12\beta_T/\log(1+\sigma_n^{-2})\sum_{t=1}^{T}\log(1+\sigma_n^{-2}\sigma_t^2(x_t)) \\
&\leq 12C^2L^2T + 24(C_2 + C_1\beta_T^{1/2})^2\log T + 24\beta_T/\log(1+\sigma_n^{-2})\cdot\gamma_T
\end{aligned} \tag{16}$$

where the first inequality follows from (15), the second inequality follows from identity $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, the third inequality follows from $\sum_{t=1}^{T} 1/|\mathcal{X}_{t-1}| \leq \sum_{t=1}^{T} 1/t \leq 2\log T$ and the fact that β_t is nondecreasing, the fourth inequality corresponds to an intermediate step of Lemma 5.4 in Srinivas et al. (2010) and the last step follows from Lemma 5.3 and Lemma 5.4 in Srinivas et al. (2010) where $\gamma_T \triangleq \max_{\mathcal{X}_T \subset \mathcal{X}} \mathbb{I}[\mathbf{f}_{\mathcal{X}}; \mathbf{y}_{t-1}] = \mathcal{O}\big((\log T)^{d+1}\big)$ and $\mathbf{f}_{\mathcal{X}} \triangleq (f(x))^{\top}_{x\in\mathcal{X}}$ (see Theorem 5 in Srinivas et al. (2010)). Therefore,

$$\begin{aligned}
&S_T^2 \\
&\leq R_T^2/T^2 \\
&\leq \sum_{t=1}^{T} r_t^2/T \\
&\leq 12C^2L^2 + 24(C_2 + C_1\beta_T^{1/2})^2\log T/T + 24\beta_T/\log(1+\sigma_n^{-2})\gamma_T/T
\end{aligned} \tag{17}$$

where the second inequality follows from Cauchy-Schwarz inequality and the last inequality follows from (16). If $\sigma_{min}(\mathcal{X}) \geq \omega$ then, according to Theorem 5, $C = \nu\mathrm{d}^2$. To guarantee that $12C^2L^2 \leq \epsilon_{ucb}^2$ and to satisfy the premise of Lemma 1 (i.e. $\nu \leq 1/2$) and Theorem 5 (i.e. $\nu \leq 2/\mathrm{d}^2$), we need to set the value of $\nu = \min(\varepsilon_{ucb}/(2\sqrt{3}\mathrm{d}^2L), 2/\mathrm{d}^2, 1/2)$.

Since $\nu \leq 2/\mathrm{d}^2$ and hence $C = \nu\mathrm{d}^2 \leq 2$

$$\begin{aligned}
&C_1 \\
&= C\sigma_y\sqrt{2\sigma_y^2+\sigma_n^2}\left(\sqrt{2}(1+C)^2\sigma_y^2/\sigma_n^2 + (2+C)C\right) \\
&\leq 2\sigma_y\sqrt{2\sigma_y^2+\sigma_n^2}\left(\sqrt{2}(1+2)^2\sigma_y^2/\sigma_n^2 + (2+2)\cdot 2\right) \\
&= \mathcal{O}\left(\sigma_y\sqrt{\sigma_y^2+\sigma_n^2}(\sigma_y^2/\sigma_n^2+1)\right)
\end{aligned}$$

and

$$\begin{aligned}
&C_2 \\
&= \sqrt{2}(1+C)\cdot C\sigma_y^2/\sigma_n^2\cdot L \\
&\leq \sqrt{2}(1+2)\cdot 2\sigma_y^2/\sigma_n^2\cdot L \\
&= \mathcal{O}(\sigma_y^2/\sigma_n^2\cdot L)
\end{aligned}$$

where C_1 and C_2 are defined in Lemma 5 and Lemma 6, respectively.

Remark 6. If $\sigma_{min}(\mathcal{X}) < \omega$, a similar form of regret bound to that of (17) can be proven: According to Theorem 5, $C = \max(\nu\mathrm{d}^2, 1 - \exp\left(-0.5(\nu + \nu\omega^2/\sigma_{min}^2(\mathcal{X}) + \omega^2/\sigma_{min}^2(\mathcal{X}))\mathrm{d}^2\right))$ instead of $C = \nu\mathrm{d}^2$ and the entire proof of Theorem 3 can be directly copied to reach (17). In this case, however, the term $12C^2L^2$ in (17) cannot be set arbitrarily small. That is explained by the fact that when $\sigma_{min}(\mathcal{X}) < \omega$, Algorithm 1 increases the singular values of dataset $\mathcal{X}$ (see line 9) and the pairwise distances between the original inputs from $\mathcal{X}$ are no longer approximately the same as the distances between their respective transformed images (see Theorem 2) resulting in a looser regret bound.

B.6. Auxiliary results

Lemma 7. *Let a dataset $\mathcal{X} \subset \mathbb{R}^d$ be given. Let a dataset $\tilde{\mathcal{X}} \subset \mathbb{R}^d$ be defined in line 9 of Algorithm 1 (i.e., $\tilde{\mathcal{X}} = U\sqrt{\Sigma^2 + \omega^2 I_{n\times d}}V^\top$ where $\mathcal{X} = U\Sigma V^\top$ is the singular value decomposition of $\mathcal{X}$). Let $\sigma_{min}(\mathcal{X}) > 0$ be the smallest singular value of $\mathcal{X}$. Then for all $x, x' \in \mathcal{X}$ and their corresponding $\tilde{x}, \tilde{x}' \in \tilde{\mathcal{X}}$ (when viewing datasets $\mathcal{X}$ and $\tilde{\mathcal{X}}$ as matrices)*

$$\|x - x'\| \leq \|\tilde{x} - \tilde{x}'\| \leq \sqrt{1 + \omega^2/\sigma^2_{min}(\mathcal{X})}\|x - x'\|.$$

Proof. Denote the rows of U as $u_{(i)}$ so that

$$U = \begin{bmatrix} u_{(1)} \\ \vdots \\ u_{(n)} \end{bmatrix}.$$

For $i = 1, \ldots, n$ denote the input in the i-th row of the datset $\mathcal{X}$ ($\tilde{\mathcal{X}}$) viewed as matrix as $x_{(i)}$ ($\tilde{x}_{(i)}$). From the singular value decomposition, $x_{(i)} = u_{(i)}\Sigma V^\top$ and $\tilde{x}_{(i)} = u_{(i)}\sqrt{\Sigma^2 + I_{n\times d}\omega^2}V^\top$ Then for $i, j = 1, \ldots, n$

$$\begin{aligned}
&\|\tilde{x}_{(i)} - \tilde{x}_{(j)}\|^2 \\
&= \|(u_{(i)} - u_{(j)})\sqrt{\Sigma^2 + \omega^2 I_{n\times d}}V^\top\|^2 \\
&= (u_{(i)} - u_{(j)})\sqrt{\Sigma^2 + \omega^2 I_{n\times d}}V^\top V\sqrt{\Sigma^2 + \omega^2 I_{n\times d}}^\top (u_{(i)} - u_{(j)})^\top \\
&= (u_{(i)} - u_{(j)})\sqrt{\Sigma^2 + \omega^2 I_{n\times d}}\sqrt{\Sigma^2 + \omega^2 I_{n\times d}}^\top (u_{(i)} - u_{(j)})^\top \\
&= \sum_{k=1}^{\min(n,d)} (u_{(i)k} - u_{(j)k})^2(\sigma_k^2 + \omega^2) \\
&\leq \sum_{k=1}^{\min(n,d)} (u_{(i)k} - u_{(j)k})^2\sigma_k^2(1 + \omega^2/\sigma^2_{min}(\mathcal{X})) \\
&= (1 + \omega^2/\sigma^2_{min}(\mathcal{X}))(u_{(i)} - u_{(j)})\Sigma\Sigma^\top(u_{(i)} - u_{(j)})^\top \\
&= (1 + \omega^2/\sigma^2_{min}(\mathcal{X}))(u_{(i)} - u_{(j)})\Sigma V^\top V\Sigma^\top(u_{(i)} - u_{(j)})^\top \\
&= (1 + \omega^2/\sigma^2_{min}(\mathcal{X}))\|(u_{(i)} - u_{(j)})\Sigma V^\top\|^2 \\
&= (1 + \omega^2/\sigma^2_{min}(\mathcal{X}))\|x_{(i)} - x_{(j)}\|^2
\end{aligned} \tag{18}$$

where the second and the second last equalities follow from $\|v\|^2 = vv\top$ for any row vector v, the third and the third last equalities follow from orthonormality of matrix V, and the inequality follows from

$$\begin{aligned}
&\sigma_k^2 + \omega^2 \\
&= \sigma_k^2(1 + \omega^2/\sigma_k^2) \\
&\leq \sigma_k^2(1 + \omega^2/\sigma^2_{min}(\mathcal{X}))
\end{aligned}$$

where the inequality follows from $\sigma_k \geq \sigma_{min}(\mathcal{X})$ for every $k = 1, \ldots, \min(n, d)$.

Similarly,

$$\begin{aligned}
&\|\tilde{x}_{(i)} - \tilde{x}_{(j)}\|^2 \\
&= \sum_{k=1}^{\min(n,d)} (u_{(i)k} - u_{(j)k})^2(\sigma_k^2 + \omega^2) \\
&= \sum_{k=1}^{\min(n,d)} (u_{(i)k} - u_{(j)k})^2\sigma_k^2 + \omega^2 \sum_{k=1}^{\min(n,d)} (u_{(i)k} - u_{(j)k})^2 \\
&\geq \sum_{k=1}^{\min(n,d)} (u_{(i)k} - u_{(j)k})^2\sigma_k^2 \\
&= \|x_{(i)} - x_{(j)}\|^2
\end{aligned} \tag{19}$$

where the first and the last equalities follow from the fourth and the fifth equalities of (18), respectively. Since (18) and (19) both hold for all $i, j = 1, \ldots, n$, the lemma follows. □

Lemma 8. *In the notations of Section B.5, for all $t = 1, \dots, T$ holds $\|(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \leq 1/\sigma_n^2$.*

Proof. Since $(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}$ is positive definite, by definition of spectral norm for all $t = 1, \dots, T$ and $\mathcal{Z}_{t-1}$

$$\begin{aligned}
&\|(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \\
&= \psi_{max}((K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)^{-1}) \\
&= \frac{1}{\psi_{min}(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} + \sigma_n^2 I)} \\
&= \frac{1}{\psi_{min}(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}}) + \sigma_n^2} \\
&\leq 1/\sigma_n^2
\end{aligned}$$

where $\psi_{max}(\cdot)$ and $\psi_{min}(\cdot)$ denote the largest and the smallest eigenvalues of a matrix, respectively, the second and the third equalities are properties of eigenvalues and the inequality is due to the fact that matrix $K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}}$ is positive semidefinite. □

Lemma 9. *In the notations of Section B.5, if for all $x, x' \in \mathcal{X}$ and their images under Algorithm 1 $z, z' \in \mathcal{Z}$ holds $|k_{zz'} - k_{xx'}| \leq C \cdot k_{xx'}$, and for all $t = 1, \dots, T$ matrix $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$ is diagonally dominant (Definition 3), then*

$$\|K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}\|_2 \leq \sqrt{2}C\sigma_y^2/\sqrt{|\mathcal{X}_{t-1}|}.$$

Proof. Fix $t = 1, \dots, T$. For some $i = 1, \dots, t-1$:

$$\begin{aligned}
&\|K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}\|_2^2 \\
&= \psi_{max}\big((K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})^\top (K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})\big) \\
&= \psi_{max}\big((K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})^2\big) \\
&\leq \sum_{j,j\neq i} |[(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})^2]_{ij}| + [(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})^2]_{ii} \\
&\leq 2C^2\sigma_y^4/\big(\sqrt{|\mathcal{X}_{t-1}| - 1} + 1\big)^2 \\
&\leq 2C^2\sigma_y^4/|\mathcal{X}_{t-1}|
\end{aligned}$$

where $\psi_{max}(\cdot)$ denotes the largest eigenvalue of a matrix, the first equality is the definition of spectral norm, the second equality follows from the fact that matrices $K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}}$ and $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$ are symmetric, the first inequality is due to Gershgorin circle theorem, the last inequality follows from $\sqrt{|\mathcal{X}_{t-1}| - 1} + 1 \geq \sqrt{|\mathcal{X}_{t-1}|}$ and the second last inequality follows from

$$\begin{aligned}
&\sum_{j,j\neq i} |[(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})^2]_{ij}| \\
&= \sum_{j,j\neq i} |\sum_p [K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}]_{ip}[K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}]_{pj}| \\
&= \sum_{j,j\neq i} |\sum_p (k_{z_i z_p} - k_{x_i x_p})(k_{z_p z_j} - k_{x_p x_j})| \\
&= \sum_{j,j\neq i} |\sum_{p,p\neq j,i} (k_{z_i z_p} - k_{x_i x_p})(k_{z_p z_j} - k_{x_p x_j})| \\
&\leq \sum_{j,j\neq i} \sum_{p,p\neq j,i} |k_{z_i z_p} - k_{x_i x_p}| \cdot |k_{z_p z_j} - k_{x_p x_j}| \\
&\leq C^2 \sum_{j,j\neq i} \sum_{p,p\neq j} k_{x_i x_p} \cdot k_{x_p x_j} \\
&= C^2 \sum_{p,p\neq j,i} k_{x_i x_p} \sum_{j,j\neq i,p} k_{x_p x_j} \\
&\leq C^2 \sum_{p,p\neq j,i} k_{x_i x_p} k_{x_p x_p}/\big(\sqrt{|\mathcal{X}_{t-1}| - 1} + 1\big) \\
&= C^2\sigma_y^2/\big(\sqrt{|\mathcal{X}_{t-1}| - 1} + 1\big) \sum_{p,p\neq j,i} k_{x_i x_p} \\
&\leq C^2\sigma_y^2/\big(\sqrt{|\mathcal{X}_{t-1}| - 1} + 1\big) k_{x_i x_i}/\big(\sqrt{|\mathcal{X}_{t-1}| - 1} + 1\big) \\
&= C^2\sigma_y^4/\big(\sqrt{|\mathcal{X}_{t-1}| - 1} + 1\big)^2
\end{aligned}$$

where the third, the fifth and the last equalities follow from $k_{z_p z_p} = k_{x_p x_p} = \sigma_y^2$ for every p, the first inequality follows from triangle inequality, the second inequality follows from the statement of the lemma, the third and the last inequalities follow from the diagonal dominance property of $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$ (Definition 3); and

$$\begin{aligned}
&[(K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})^2]_{ii} \\
&= \sum_p [K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}]_{ip} [K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}]_{pi} \\
&= \sum_p [K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}} - K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}]_{ip}^2 \\
&= \sum_p (k_{z_i z_p} - k_{x_i x_p})^2 \\
&= \sum_{p, p\neq i} (k_{z_i z_p} - k_{x_i x_p})^2 \\
&\leq C^2 \sum_{p, p\neq i} k_{x_i x_p}^2 \\
&\leq C^2 \big(\sum_{p, p\neq i} k_{x_i x_p} \big)^2 \\
&\leq C^2 k_{x_i x_i}^2 / \big(\sqrt{|\mathcal{X}_{t-1}| - 1} + 1 \big)^2 \\
&= C^2 \sigma_y^4 / \big(\sqrt{|\mathcal{X}_{t-1}| - 1} + 1 \big)^2
\end{aligned}$$

where the second equality follows from the fact that $K_{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}}$ and $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$ are symmetric, the fourth and the last equalities follow from $k_{z_p z_p} = k_{x_p x_p} = \sigma_y^2$ for every p, the first inequality follows from the statement of the lemma and the last inequality follows from the diagonal dominance of $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$ (Definition 3). □

Lemma 10. *In the notations of Section B.5, if for all $t = 1, \ldots, T$ matrix $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$ is diagonally dominant (Definition 3), then for any unobserved original input $x \in \mathcal{X}$ at iteration t*

$$\|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}} + \sigma_n^2 I)^{-1}\|_2 \leq 1/(\sqrt{|\mathcal{X}_{t-1}|}\|K_{x\mathcal{X}_{t-1}}\|).$$

.

Proof. By applying Gershgorin circle theorem for $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$:

$$\begin{aligned}
&\psi_{min}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}) \\
&\geq \min_{x_i \in \mathcal{X}_{t-1}} \big(k_{x_i x_i} - R_{\mathcal{X}_{t-1}}(x_i) \big) \\
&= k_{xx} - \max_{x_i \in \mathcal{X}_{t-1}} R_{\mathcal{X}_{t-1}}(x_i) \\
&\geq (\sqrt{|\mathcal{X}_{t-1}|} + 1) \max_{x_i \in \mathcal{X}_{t-1} \cup \{x\}} R_{\mathcal{X}_{t-1}\cup\{x\}}(x_i) - \max_{x_i \in \mathcal{X}_{t-1}} R_{\mathcal{X}_{t-1}}(x_i)
\end{aligned}$$

where $\psi_{min}(\cdot)$ denotes the smallest eigenvalue of a matrix, $R_{\mathcal{X}_{t-1}}(x_i) \triangleq \sum_{x_j \in \mathcal{X}_{t-1}\setminus\{x_i\}} k_{x_i x_j}$, the first equality follows from the fact that $k_{xx} = \sigma_y^2 = k_{x_i x_i}$ for all x_i and x, and the second inequality holds because $K_{(\mathcal{X}_{t-1}\cup\{x\})(\mathcal{X}_{t-1}\cup\{x\})}$ is assumed to be diagonally dominant. On the other hand, since $x \notin \mathcal{X}_{t-1}$, $R_{\mathcal{X}_{t-1}\cup\{x\}}(x_i) = R_{\mathcal{X}_{t-1}}(x_i) + k_{x_i x}$ for all $x_i \in \mathcal{X}_{t-1}$, which immediately implies $\max_{x_i \in \mathcal{X}_{t-1}\cup\{x\}} R_{\mathcal{X}_{t-1}\cup\{x\}}(x_i) \geq \max_{x_i \in \mathcal{X}_{t-1}} R_{\mathcal{X}_{t-1}\cup\{x\}}(x_i) \geq \max_{x_i \in \mathcal{X}_{t-1}} R_{\mathcal{X}_{t-1}}(x_i)$. Plugging this into above inequality,

$$\begin{aligned}
&\psi_{min}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}) \\
&\geq (\sqrt{|\mathcal{X}_{t-1}|} + 1) \max_{x_i \in \mathcal{X}_{t-1} \cup \{x\}} R_{\mathcal{X}_{t-1}\cup\{x\}}(x_i) - \max_{x_i \in \mathcal{X}_{t-1}} R_{\mathcal{X}_{t-1}}(x_i) \\
&\geq \sqrt{|\mathcal{X}_{t-1}|} \max_{x_i \in \mathcal{X}_{t-1} \cup \{x\}} R_{\mathcal{X}_{t-1}\cup\{x\}}(x_i) \\
&\geq \sqrt{|\mathcal{X}_{t-1}|} R_{\mathcal{X}_{t-1}\cup\{x\}}(x).
\end{aligned}$$

Since $\|K_{x\mathcal{X}_{t-1}}\| = \sqrt{\sum_{x_i \in \mathcal{X}_{t-1}} k_{x_i x}^2} \leq \sum_{x_i \in \mathcal{X}_{t-1}} k_{x_i x} = R_{\mathcal{X}_{t-1}\cup\{x\}}(x)$, it follows that $\psi_{min}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}) \geq$

$\sqrt{|\mathcal{X}_{t-1}|}\|K_{x\mathcal{X}_{t-1}}\|$. Finally,

$$\begin{aligned}&\|(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}+\sigma_n^2 I)^{-1}\|_2\\&=1/(\psi_{min}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})+\sigma_n^2 I)\\&\leq 1/(\psi_{min}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}))\\&\leq 1/\big(\sqrt{|\mathcal{X}_{t-1}|}\|K_{x\mathcal{X}_{t-1}}\|\big).\end{aligned}$$

□

Lemma 11. *In the notations of Section B.5, if for all $t=1,\ldots,T$ matrix $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$ is diagonally dominant (Definition 3), then $\|K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}\|_2\leq 2\sigma_y^2$.*

Proof. Fix all $t=1,\ldots,T$. By applying Gershgorin circle theorem to matrix $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$, for some point $x_i\in\mathcal{X}_{t-1}$:

$$\begin{aligned}&|\psi_{max}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})-k_{x_ix_i}|\\&\leq\sum_{x_j\in\mathcal{X}_{t-1}\setminus x_i}k_{x_ix_j}\\&\leq k_{x_ix_i}/\big(\sqrt{|\mathcal{X}_{t-1}|-1}+1\big)\\&=\sigma_y^2/\big(\sqrt{|\mathcal{X}_{t-1}|-1}+1\big)\end{aligned}$$

where $\psi_{max}(\cdot)$ denotes the largest eigenvalue of a matrix, the second inequality is due to diagonal dominance property of matrix $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$ and the equality is due to $k_{x_ix_i}=\sigma_y^2$ for every x_i. Since $K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}$ is a symmetric, positive-semidefinite matrix, it follows that

$$\begin{aligned}&\|K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}\|_2\\&=\psi_{max}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}})\\&\leq\sigma_y^2/\big(\sqrt{|\mathcal{X}_{t-1}|-1}+1\big)+k_{x_ix_i}\\&\leq\sigma_y^2\big(1+1/(\sqrt{|\mathcal{X}_{t-1}|-1}+1)\big)\\&\leq 2\sigma_y^2.\end{aligned}$$

. □

Lemma 12. *In the notations of Section B.5, for all $t=1,\ldots,T$ and any unobserved input $x\in\mathcal{X}$ at iteration t $\|K_{x\mathcal{X}_{t-1}}\|^2\cdot\psi_{min}((K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}+\sigma_n^2 I)^{-1})\leq K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}+\sigma_n^2 I)^{-1}K_{\mathcal{X}_{t-1}x}$ where $\psi_{min}(\cdot)$ denotes the smallest eigenvalue of a matrix.*

Proof. Since $(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}+\sigma_n^2 I)^{-1}$ is a symmetric, positive-definite matrix, there exists an orthonormal basis comprising the eigenvectors $E\triangleq[e_1\ldots e_{|\mathcal{X}_{t-1}|}]$ ($e_i^\top e_i=1$ and $e_i^\top e_j=0$ for $i\neq j$) and their associated positive eigenvalues $\Psi^{-1}\triangleq\mathrm{Diag}[\psi_1^{-1},\ldots,\psi_{|\mathcal{X}_{t-1}|}^{-1}]$ such that $(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}+\sigma_n^2 I)^{-1}=E\Psi^{-1}E^\top$ (i.e., spectral theorem). Denote $\{p_i\}_{i=1}^{|\mathcal{X}_{t-1}|}$ as the set of coefficients when $K_{\mathcal{X}_{t-1}x}$ is projected on E. Then

$$\begin{aligned}&K_{x\mathcal{X}_{t-1}}(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}+\sigma_n^2 I)^{-1}K_{\mathcal{X}_{t-1}x}\\&=\left(\sum_{i=1}^{|\mathcal{X}_{t-1}|}p_ie_i^\top\right)(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}+\sigma_n^2 I)^{-1}\left(\sum_{i=1}^{|\mathcal{X}_{t-1}|}p_ie_i\right)\\&=\left(\sum_{i=1}^{|\mathcal{X}_{t-1}|}p_ie_i^\top\right)\left(\sum_{i=1}^{|\mathcal{X}_{t-1}|}p_i(K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}+\sigma_n^2 I)^{-1}e_i\right)\\&=\left(\sum_{i=1}^{|\mathcal{X}_{t-1}|}p_ie_i^\top\right)\left(\sum_{i=1}^{|\mathcal{X}_{t-1}|}p_i\psi_i^{-1}e_i\right)\\&=\sum_{i=1}^{|\mathcal{X}_{t-1}|}p_i^2\psi_i^{-1}\\&\geq\psi_{min}((K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}+\sigma_n^2 I)^{-1})\sum_{i=1}^{|\mathcal{X}_{t-1}|}p_i^2\\&=\psi_{min}((K_{\mathcal{X}_{t-1}\mathcal{X}_{t-1}}+\sigma_n^2 I)^{-1})\,\|K_{x\mathcal{X}_{t-1}}\|^2.\end{aligned}$$

□