We thank the three anonymous reviewers for their insightful and largely positive comments!

Responses to Reviewer 1 Thanks again for your thoughtful comments.

Lemmas 3.1 and 3.2. Regarding Lemma 3.1, the novelty is the existence of the function $q(x)$. If $q(x) = 1$, this is well known in the literature, as the reviewer pointed out. We will make this point clear in the revision. Regarding Lemma 3.2, while we do not see this as a key contribution of the paper, it is not clear to us whether it follows from the results of Smola and Schoelkopf. This is because they consider greedy selection of a subset of training data, while we consider selection of points from the entire domain. If you know any existing work that explicitly states Lemmas 3.2, could you please let us know when you update your review? We will cite it and remove our proof in Appendix.

Lemma 4.1. Thank you for the suggestion, which we follow in the revision.

Proof of Proposition 4.2. Thank you very much for pointing this out. We agree that our argument of deriving (v) was flawed. We make the following correction, which we believe makes Proposition 4.2 still valid. (We will correct the proof of Prop. C.2 in a similar way.) We start from (iii), which can be stated as that there exists $n_0 \in \mathbb{N}$ such that $k_{X_n}(x,x) \leq \exp(-(c_1/c_2)n^{1/d})$ holds for all $n \geq n_0$. Now, define $c_4 > 0$ as a constant such that $c_4 \exp(-(c_1/c_2)n_0^{1/d}) = c_3$, and let $c_5 := \max(c_4, 1)$. Then, for $n < n_0$ we have $c_5 \exp(-(c_1/c_2)n^{1/d}) \geq c_4 \exp(-(c_1/c_2)n^{1/d}) \geq c_3 \geq k_{X_n}(x,x)$, where the last inequality follows from (iv). For $n \geq n_0$, we have $c_5 \exp(-(c_1/c_2)n^{1/d}) \geq \exp(-(c_1/c_2)n^{1/d}) \geq k_{X_n}(x,x)$. Therefore we conclude that $k_{X_n}(x,x) \leq c_5 \exp(-(c_1/c_2)n^{1/d})$ holds for all $n \in \mathbb{N}$ and $x \in \Omega$.

Responses to Reviewer 2 Thanks again for your insightful comments.

The estimator in line 120. This is a very good point, and thanks for pointing it out. We agree that the current presentation is confusing, and will make a correction. The quadrature estimator suggested in [8] (and used for WSABI-M [13]) can be described as $\int \mathbb{E}_{\acute{g}} T(\acute{g}(x))\pi(x)$, where $\acute{g} \sim \mathcal{GP}(m_{g,X_n}, k_{X_n})$ is the posterior GP. On the other hand, the estimator in line 120 is $\int T(m_{g,X_n}(x))\pi(x)$ and used by WSABI-L [13]. As we describe below, these two estimators are *both consistent with the same convergence rates*, and all theoretical guarantees obtained in the paper are applicable to the estimator $\int \mathbb{E}_{\acute{g}} T(\acute{g}(x))\pi(x)$ as well. Intuitively, this is because the posterior $\acute{g} \sim \mathcal{GP}(m_{g,X_n}, k_{X_n})$ contracts around the posterior mean m_{g,X_n} as n increases, and $\mathbb{E}_{\acute{g}} T(\acute{g})$ and $T(m_{g,X_n})$ get similar. We note however that for a finite n, it is not clear which estimator is "better," and this is an interesting topic for future research.

We sketch here that for the estimator $\int \mathbb{E}_{\acute{g}} T(\acute{g}(x))\pi(x)$, the essentially same upper bound as Proposition 2.1 holds, under an additional condition that $\mathbb{E}_{\acute{g}}(T'(|g(x)|+|\acute{g}(x)|))^2 < C$ holds for all $x \in \Omega$ and $n \in \mathbb{N}$ for some $C > 0$ (which can be shown to be satisfied for transformations T mentioned in our paper). By Taylor's theorem, there exists $\alpha_{x,X_n,\acute{g}} \in [0,1]$ such that for $y_{x,X_n,\acute{g}} := g(x) + \alpha_{x,X_n,\acute{g}}(\acute{g}(x) - g(x))$ we have $T(\acute{g}(x)) = T(g(x)) + T'(y_{x,X_n,\acute{g}})(\acute{g}(x) - g(x))$. Therefore $(\mathbb{E}_{\acute{g}}[T(\acute{g}(x))] - T(g(x)))^2 = (\mathbb{E}_{\acute{g}}[T'(y_{x,X_n,\acute{g}})(\acute{g}(x) - g(x))])^2 \leq \mathbb{E}_{\acute{g}}[(T'(y_{x,X_n,\acute{g}}))^2]\mathbb{E}_{\acute{g}}[(\acute{g}(x) - g(x))^2] \leq C\mathbb{E}_{\acute{g}}[(\acute{g}(x) - g(x))^2]$, where the last inequality follows from $|y_{x,X_n,\acute{g}}| \leq |g(x)| + |\acute{g}(x)|$ and the above assumption. Moreover, $\mathbb{E}_{\acute{g}}[(\acute{g}(x) - g(x))^2] \leq 2\mathbb{E}_{\acute{g}}[(\acute{g}(x) - m_{g,X_n}(x))^2] + 2(m_{g,X_n}(x) - g(x))^2 \leq 2k_{X_n}(x,x) + 2\|\tilde{g}\|^2_{\mathcal{H}_k} k_{X_n}(x,x)$, where the last inequality follows from Eq. (10). Thus, $|T(g(x)) - \mathbb{E}_{\acute{g}}[T(\acute{g}(x))]| \leq \sqrt{2C(1 + \|\tilde{g}\|^2_{\mathcal{H}_k})}\sqrt{k_{X_n}(x,x)}$. Following the argument in line 436, the essentially same bound as Prop. 2.1 can be obtained (with a different constant).

The generic form of acquisition functions (Eq. 4). We defined Eq. (4) so that the class of acquisition functions to which our convergence guarantees are applicable becomes as large as possible. One positive side of this generality is that it enables practitioners to design a new acquisition that results in a consistent algorithm; its derivation can be quite different from those of the existing acquisition functions (which are often done by approximating an intractable posterior covariance function of the transformed integrand). In this sense, we think that every special case of (4) does not have to have an interpretation as an approximate posterior covariance. We nevertheless agree that Eq. 4 is rather abstract. We will include a discussion about the roles of the components in Eq. (4).

Responses to Reviewer 3 Thanks again for encouraging comments and insightful suggestions.

BQ v.s. MC. We will make the tone of the comparison more neutral, including a discussion of the dependence of the dimensionality. We hope that you agree, however, that the existence of a convergence guarantee for MCMC (which didn't exist for adaptive BQ so far) is a key contributor to MCMC's popularity.

Ergodicity / detailed balance and the relation to weak adaptivity. We will modify the presentation so that the intuitive discussion on the connection to the detailed balance and ergodicity becomes minimal. We will also add an intuitive explanation about the weak adaptivity condition immediately after it is introduced.

Proof map / Appendix. We will provide a diagram in Appendix that describes the relationships between the various auxiliary results and how they yield the main results. We will also add a high level overview of the proof plan and techniques used.