**A APPENDIX**

**A.1 PROOF OF THEOREM 2**

We begin by first proving Theorem 2 since the additional assumption of realizability makes it an easier read. For further ease of exposition, instead of directly working with \( g(\cdot) \), we translate the function to remove any constants not dependent on the variable. We write,

\[
l(S) := \|\mu_\pi\|_k^2 - g(S) = z^\top K^{-1} z.
\]

Some auxiliary Lemmas are proved later in this section. We use \( Z(S_j) := \sum_j w_j \phi(x_j) \). Further, note that the Assumption 2, when applied for \( h(\cdot) \), ensures that for any iterates considered in this proof we have that

\[
-\frac{m_\omega}{2} \|Z(S_i) - Z(S_j)\|_k^2 \\
\geq l(S_i) - l(S_j) - \langle \nabla l(S_i), Z(S_i) - Z(S_j) \rangle_k \\
\geq -\frac{M_\Omega}{2} \|Z(S_i) - Z(S_j)\|_k^2.
\]

**Proof.** Say \((i - 1)\) steps of the Algorithm 1 have been performed to select the set \( S \). Let \( w \in \mathbb{R}^{l - 1} \) be the corresponding weight vector. Let \( h(S, u) := \|\mu_\pi\|_k^2 - \|\mu_\pi - \sum_j u_j \phi(x_j)\|_k^2 \), so that \( l(S) = \min_u h(S, u) \) (as per Lemma 1). Set weight vector \( u \in \mathbb{R}^l \) as follows. For \( j \in [0, i - 1], u_j = w_i \). Set \( u_i = \alpha \), where \( \alpha \) is an arbitrary scalar.

From weight optimality proved in Lemma 1

\[
l(S \cup \{x_i\}) - l(S) \geq h(S \cup \{x_i\}, u) - l(S),
\]

for an arbitrary \( \alpha \in \mathbb{R} \). From Assumption 2 (smoothness),

\[
l(S \cup \{x_i\}) - l(S) \geq \alpha \langle \nabla l(S), \phi(x_i) \rangle_k - \alpha^2 M_\Omega.
\]

Let \( \gamma_S \) be the optimum value of the solution of the inner LMO problem. Since \( x_i \) is the optimizing atom,

\[
l(S \cup \{x_i\}) - l(S) \geq \alpha \gamma_S - \alpha^2 \frac{M_\Omega}{2}.
\]

Let \( S^*_\perp \) be the set obtained by orthogonalizing \( S^*_\perp \) with respect to \( S \) using the Gram-Schmidt procedure. Putting in \( \alpha = \frac{\gamma_S}{M_\Omega} \), we get,

\[
l(S \cup \{x_i\}) - l(S) \geq \frac{1}{2M_\Omega} \gamma_S \\
\geq \frac{1}{2r M_\Omega} \sum_{x_j \in S^*_\perp} \langle \phi(x_j), \nabla l(S) \rangle_k^2 \\
\geq \frac{m_\omega}{r M_\Omega} (l(S \cup S^*_\perp) - l(S)) \\
\geq \frac{m_\omega}{r M_\Omega} (l(S^*_\perp) - l(S)) \\
= \frac{m_\omega}{r M_\Omega} \left( \|\mu_\pi\|_k^2 - l(S) \right).
\]

The second inequality is true because \( \gamma_S = \langle \nabla l(S), x_i \rangle_k \) is the optimum value of the inner LMO problem in the \( i \)th iteration. The third inequality follows from Lemma 2. The fourth inequality is true because of monotonicity of \( l(\cdot) \), and the final equality is true because of Assumption 1 (realizability).

Let \( C := \frac{m_\omega}{r M_\Omega} \). We have \( l(S \cup \{x_i\}) - l(S) = g(S) - g(S \cup \{x_i\}) \geq C g(S) \iff g(S \cup \{x_i\}) \leq (1 - C) g(S) \). The result now follows.

\[\square\]

**A.2 PROOF OF THEOREM 3**

**Proof.** We proceed as in the proof of Theorem 2 but by replacing \( S^*_\perp \) with \( T_r \). From 2,

\[
l(S \cup \{x_i\}) - l(S) \geq \frac{m_\omega}{r M_\Omega} (l(T_r) - l(S))
\]

Adding and subtracting \( l(T_r) \) on the LHS and rearranging,
Thus after $k$ iterations,

$$l(T_r) - l(S_k) \leq (1 - \frac{m_\omega}{rM_\Omega})^k (l(T_r) - l(\emptyset)).$$

Rearranging,

$$l(S_k) \geq \left(1 - \frac{m_\omega}{rM_\Omega}\right)^k l(T_r) \geq \left(1 - \exp\left(-\frac{km_\omega}{rM_\Omega}\right)\right) l(T_r).$$

With $k = (r \frac{M_\Omega}{m_\omega} \log \frac{1}{\epsilon})$, we get,

$$l(S_k) \geq (1 - \epsilon)l(T_r).$$

The result now follows.

**A.3 AUXILIARY LEMMAS**

The following Lemma proves that the weights $w_i$ in $g(S)$ obtained using the posterior inference are an optimum choice that minimize the distance to $\mu_\pi$ in the RKHS over any set of weights [Khanna et al., 2019].

**Lemma 1.** The residual $\mu_\pi - \sum_j w_j \phi(x_j)$ is orthogonal to $x_i \in S \forall i$. In other words, for any set of samples $S$, $g(S) = \min_w \| \mu_\pi - \sum_j w_j \phi(x_j) \|_k$.

**Proof.** Recall that $w_i = \sum_j [K^{-1}]_{ij} z_j$, and $z_i = \int k(x, x_i) d\pi(x)$. For an arbitrary index $i$,

$$\langle \mu_\pi - \sum_j w_j \phi(x_j), \phi(x_i) \rangle_k = \int k(x, x_i) d\pi(x) - \langle \sum_j w_j \phi(x_j), \phi(x_i) \rangle_k = z_i - \langle \sum_j w_j \phi(x_j), \phi(x_i) \rangle_k = z_i - \sum_j w_j k(x_j, x_i) = z_i - \sum_t [K^{-1}]_{tj} z_t k(x_j, x_i) = z_i - z_i = z_i.$$

where the last equality follows by noting that $\sum_t K_{jt}[K^{-1}]_{tj}$ is inner product of row $i$ of $K$ and row $t$ of $K^{-1}$, which is 1 if $t = i$ and 0 otherwise. This completes the proof.

**Lemma 2.** For any set of chosen samples $S_1, S_2$, let $P$ be the operator of projection onto span($S_1 \cup S_2$). Then, $l(S_1 \cup S_2) - l(S_1) \leq \frac{P(\nabla l(S_1))}{2m_\omega}$.

**Proof.** Observe that

$$0 \leq l(S_1 \cup S_2) - l(S_1) \leq \langle \nabla l(S_1), Z(S_1 \cup S_2) - Z(S_1) \rangle_k - m_\omega \|Z(S_1 \cup S_2) - Z(S_1)\|_k^2 \leq \max_{X \in \text{span}(S_1 \cup S_2)} \langle \nabla l(S_1), X - Z(S_1) \rangle_k - \frac{m_\omega}{2} \|X - Z(S_1)\|_k^2 = \max_{X} \langle P(\nabla l(S_1)), X - Z(S_1) \rangle_k - \frac{m_\omega}{2} \|X - Z(S_1)\|_k^2.$$

Solving the argmax problem on the RHS for $X$, we get the required result.

**A.4 PROOF OF THEOREM 3**

We next present some notation and few lemmas that lead up to the main result of this section (Theorem 3). The domain of candidate atoms $X$ is split into $\{X_j, j \in [s]\}$ over $s$ machines, where machine $j$ runs WKH on $X_j$. Let $G_j$ be the $k$-sized solution returned by running Algorithm 1 on $X_j$, i.e., $G_j = \text{WKH}(X_j, k)$. Note that each $X_j$ induces a partition onto the optimal $r$-sized solution $S^*_r$ as follows ($r = 1$ for this theorem):

$$T_j := \{x \in S^*_1 : x \notin \text{WKH}(X_j \cup x, k)\},$$

$$T^*_j := \{x \in S^*_1 : x \in \text{WKH}(X_j \cup x, k)\}.$$

In other words, $T_j = S^*_1$ if the $j^{th}$ machine running WKH on $X_j \cup S^*_1$ will not select it as among its output, and it is empty otherwise, since $S^*_1$ is a singleton. We re-use the definition of $l(\cdot)$ used in Appendix A.1.

Before moving to the proof of the main theorem, we prove two prerequisites. Recall $G_j$ is the set of iterates selected by machine $j$. In this mini-result, we lower bound the expected improvement in the loss at the aggregator machine.

**Lemma 3.** For the aggregator machine that runs WKH over $\cup_j G_j$ (step 6 of Algorithm 2), we have, at selection of next sample point $x_i$ after having selected $S$, $\exists$ machine $j$ such that

$$E[l(S \cup \{x_i\}) - l(S)] \geq \frac{m_\omega}{M_\Omega} E \left(l(T^*_j) - l(S)\right).$$
Proof. The expectation is over the random split of $\mathcal{X}$ into $\mathcal{X}_j$ for $j \in [s]$. We denote $T^c_j$ to be the complement of $T_j$. Then, we have that

$$
\mathbb{E}[l(S \cup \{x_i\}) - l(S)]
\geq \mathbb{E}\left[\frac{1}{2M_\Omega} \mathcal{G}_j\right]
\geq \frac{1}{2M_\Omega} \sum_{x \in \mathcal{S}_j^c} \mathbb{P}(x \in \bigcup_j \mathcal{G}_j) \mathbb{E}(\phi(x), \nabla l(S))^2_k
= \frac{1}{2sM_\Omega} \sum_{b=1}^s \sum_{x \in T_b^c} \mathbb{E}(\phi(x), \nabla l(S))^2_k
\geq \frac{m_\omega}{sM_\Omega} \sum_{b=1}^s \mathbb{E}(l(S \cup T_b^c) - l(S))
\geq \frac{m_\omega}{M_\Omega} \min_b \mathbb{E}(l(T_b^c) - l(S)).
$$

The equality in step 3 above is because of Lemma \ref{lemma:local_lower_bound}. \hfill \Box

In the following lemma, we lower bound the greedy improvement in the loss on each machine.

Lemma 4. For any individual worker machine $j$ running local WKH, if $S$ is the set of $(i-1)$ iterates already chosen, then at the selection of next sample point $x_i$, $l(S \cup \{x_i\}) \geq (l(T_j) - l(S))$.

Proof. We proceed as in proof of Theorem \ref{theorem:main_theorem} in Appendix A. From (1), we have,

$$
l(S \cup \{x\}) - l(S) \geq \frac{1}{2M_\Omega} \mathcal{G}_j
\geq \frac{1}{2M_\Omega} \sum_{x \in T_j} \mathbb{E}(\phi(x), \nabla l(S))^2_k
\geq \frac{m_\omega}{M_\Omega} (l(S \cup T_j) - l(S))
\geq \frac{m_\omega}{M_\Omega} (l(T_j) - l(S)).
$$

We are now ready to prove Theorem \ref{theorem:main_theorem}.\hfill \Box

Proof of Theorem \ref{theorem:main_theorem}. If, for a random split of $\mathcal{X}$, for any $j \in [s]$, $T_j = \mathcal{S}_j^c$, then the given rate follows from Lemma \ref{lemma:local_lower_bound} after following the straightforward steps covered in proof of Theorem \ref{theorem:main_theorem} for proving the rate from the given condition on $l(\cdot)$. On the other hand, if none of $j \in [s]$, $T_j = \mathcal{S}_j^c$, then $\forall j \in [s]$, $T_j = \emptyset \implies T_j^c = \mathcal{S}_j^c$. In this case, the given rate follows from Lemma \ref{lemma:local_lower_bound}.

Finally, here is the statement and proof of an auxiliary lemma that was used above.

Lemma 5. For any $x \in \mathcal{X}$, $\mathbb{P}(x \in \bigcup_j \mathcal{G}_j) = \frac{1}{s} \sum_j \mathbb{P}(x \in T_j^c)$.

Proof. We have

$$
\mathbb{P}(x \in \bigcup_j \mathcal{G}_j)
= \sum_j \mathbb{P}(x \in \mathcal{X}_j \cap x \in WKH(\mathcal{X}_j, k))
= \sum_j \mathbb{P}(x \in \mathcal{X}_j)\mathbb{P}(x \in WKH(\mathcal{X}_j, k)|x \in \mathcal{X}_j)
= \sum_j \mathbb{P}(x \in \mathcal{X}_j)\mathbb{P}(x \in T_j^c)
= \frac{1}{s} \mathbb{P}(x \in T_j^c).
$$