Appendix A

Theoretical Results: Proofs

The proofs of a number of theoretical results from Chapter 3 are presented in this Appendix.

A.1 Theorem 1

In order to prove Theorem 1, this section first establishes a series of supporting lemmas.

Lemma 6. Suppose that at some round \( t \), there exists a bound, \( C \), uniform over \( D \), on the conditional mutual information with respect to \( f(x) \) which will be acquired by the set of actions initiated since the last observation at round \( fb[t] \), where this bound is of form

\[
I(f(x): y_{fb[t]+1:t-1} | y_{1:fb[t]}) \leq C, \quad \forall x \in D,
\]

for some constant \( C > 0 \). Choose

\[
\beta_t = \exp(2C) \alpha_{fb[t]} \quad \text{(A.2)}
\]

where Equation (3.6) relates sequential confidence intervals \( C_{seq}^{fb[t]+1}(x) \) with the parameter \( \alpha_{fb[t]+1} \) and Equation (3.8) relates batch confidence intervals \( C_{t}^{batch}(x) \) with the parameter \( \beta_1 \). Then, conditional on the event that for all \( x \in D \), \( f(x) \in C_{seq}^{fb[t]+1}(x) \), it holds that \( f(x) \in C_{t}^{batch}(x) \) for all \( x \in D \) and all \( t' \) such that \( fb[t]+1 \leq t' \leq t \).

Proof. Noting that the confidence intervals \( C_{seq}^{fb[t]+1}(x) \) and \( C_{t}^{batch}(x) \) are both centered on \( \mu_{fb[t]}(x) \),

\[
C_{fb[t]+1}(x) \subseteq C_{t}^{batch}(x) \quad \forall x \in D \quad \iff \quad \alpha_{fb[t]}^{1/2} \sigma_{fb[t]}(x) \leq \beta_t^{1/2} \sigma_{t-1}(x) \quad \forall x \in D.
\]

By the definition of the conditional mutual information with respect to \( f(x) \), and by employing
Equation (A.1), Equation (3.14) follows. Choosing $\beta_t$ as in Equation (A.2), it follows that

$$\alpha_{fb[t]}^{1/2} \sigma_{fb[t]}(x) = \beta_t^{1/2} \exp(-C) \cdot \sigma_{fb[t]}(x) \leq \beta_t^{1/2} \sigma_{t-1}(x),$$

where the inequality is based on Equation (3.14), thus implying $C_{fb[t]+1}^{seq}(x) \subseteq C_t^{\text{batch}}(x)$ $\forall x \in D$. In turn, if $f(x) \in C_{fb[t]+1}^{seq}(x)$, then $f(x) \in C_t^{\text{batch}}(x)$. Further, since Equation (A.2) relates $\beta_t$ to $\alpha_{fb[t]}$, then $\beta_t = \beta_{t'}$ for all $t' \in \{fb[t] + 1, \ldots, t\}$. Since $\sigma_t$ is non-increasing, $C_t^{\text{batch}}(x) \supseteq C_t^{\text{batch}}(x)$ for all such $t'$, completing the proof.

With a bound $C$ on the conditional mutual information gain with respect to $f(x)$ for any $x \in D$, as in Equation (A.1), Lemma 6 links the confidence intervals and GP-BUCB decision rule at time $t$ with the GP posterior after observation $fb[t]$. Lemma 7 extends this link to all $t \geq 1$ and all $x \in D$, given a high-probability guarantee of confidence interval correctness at the beginning of all batches. This step is required for the regret bound of Theorem 1.

**Lemma 7.** Suppose there exist a constant $C > 0$, a sequence of actions $\{x_1, \ldots, x_{t-1}\}$, and a feedback mapping $fb[t]$ such that for all $x \in D$

$$C \geq I(f(x); y_{fb[t]+1:t-1} | y_{1:fb[t]}), \forall t \geq 1.$$

Then, if $\beta_t = \exp(2C)\alpha_{fb[t]}$, $\forall t \geq 1$,

$$P(f(x) \in C_{fb[t]+1}^{seq}(x) \forall x \in D, \forall t \geq 1) \geq 1 - \delta \implies P(f(x) \in C_t^{\text{batch}}(x) \forall x \in D, \forall t \geq 1) \geq 1 - \delta.$$

**Proof.** If $\beta_t$ is chosen as specified, then for any $t$ and $\tau$ such that $\tau = fb[t]$ and $f(x) \in C_{\tau+1}^{seq}(x)$, Lemma 6 implies that $f(x) \in C_t^{\text{batch}}(x)$ . If there exists a set $S = \{\tau_1, \tau_2, \ldots\}$ such that $fb[t] \in S$ for all $t \geq 1$, and additionally $f(x) \in C_{\tau+1}^{seq}(x)$ for all $x \in D$ and $\tau \in S$, then $f(x) \in C_t^{\text{batch}}(x)$ for all $x \in D$ and all $t \geq 1$. The event $f(x) \in C_{fb[t]+1}^{seq}(x), \forall x \in D, \forall t \geq 1$ satisfies these conditions directly. The lemma follows because for logical propositions $A$ and $B$, $[A \implies B] \implies [P(A) \leq P(B)]$ and thus if $P(A) \geq 1 - \delta \implies P(B) \geq 1 - \delta$.

The high-probability confidence intervals are next be related to the instantaneous regret and thence to the cumulative regret.

**Lemma 8.** Conditional on the event $f(x) \in C_t^{\text{batch}}(x), \forall x \in D, \forall t \geq 1$, and given that actions $x_t, \forall t \geq 1$ are selected using Equation (3.7), it holds that

$$R_T \leq \sqrt{TC_1\beta_T\gamma_T},$$
where \( C_1 = \frac{8}{\log(1 + \sigma^{-2} n)} \), \( \gamma_T \) is defined in Equation (3.3), and \( \beta_t \) is defined in Equation (A.2).

**Proof.** Given \( f(x) \in C_t^{\text{batch}}(x), \forall x \in D, \forall t \geq 1 \), using the GP-BUCB decision rule, Equation (3.7), Lemma 5.2 in Srinivas et al. (2010), and our assumptions about \( C_t^{\text{batch}}(x) \), it follows that the instantaneous regret \( r_t \) is bounded as \( r_t \leq 2\beta_t^{1/2} \sigma_t^{-1}(x_t), \forall t \geq 1 \). From Lemma 5.3 and 5.4 in Srinivas et al. (2010) it follows that \( \sum_{t=1}^{T} r_t^2 \leq C_1 \beta_T \gamma_T \). The claim then follows as a consequence of the Cauchy Schwarz inequality, since \( R_T^2 \leq T \sum_{t=1}^{T} r_t^2 \).

**Proof of Theorem 1.** Taken together, Lemmas 6 through 8, a bound \( C \) satisfying Equation (A.1), and a high-probability guarantee that some set of sequential confidence intervals always contain the values of \( f \) allow us to construct a batch algorithm with high-probability regret bounds. Srinivas et al. (2010) develop choices of the exploration-exploitation tradeoff parameter \( \alpha_t \) such that guarantees of the form \( P(f(x) \in C_{t}^{\text{seq}}(x) \forall x \in D, \forall t \geq 1) \geq 1 - \delta \) are realizable for arbitrarily small \( \delta \).

Employing Lemma 8, as well as Lemmas 5.1 and 5.8 of Srinivas et al. (2010) (for assumptions 1 and 2) and Theorem 6 of Srinivas et al. (2010) (for assumption 3), Theorem 1 follows as an immediate corollary.

### A.2 Theorem 4: Initialization Set Size Bounds

Thorough initialization of GP-BUCB can drive down the constant \( C \), which bounds the information which can be hallucinated in the course of post-initialization batches and also governs the asymptotic scaling of the regret bound with batch size \( B \). First, we connect the information which can be gained in post-initialization batches with the amount of information being gained in the initialization, through Lemma 3, the formal statement of which is in Section 3.3.5, and the proof of which is presented here.

**Proof of Lemma 3.** Since the initialization procedure is greedy and information gain is submodular (See Section 3.2.3), the information gain from adding the last element of the initialization set, \( I(f; x_{\text{init}}^T \mid D_{\text{init}}^{T-1}) \), must be the smallest marginal information gain in the initialization process, and thus is no greater than the mean of the marginal gains, i.e,

\[
I(f; x_{\text{init}}^T \mid D_{\text{init}}^{T-1}) \leq I(f; D_{\text{init}}^{T-1})/T_{\text{init}}.
\]

Further, again because information gain is submodular and the initialization set was constructed greedily, no subsequent decision can yield information gain greater than \( I(f; x_{\text{init}}^T \mid D_{\text{init}}^{T-1}) \), and thus \( \gamma_{B-1}^{\text{init}} \leq (B-1) \cdot I(f; x_{\text{init}}^T \mid D_{\text{init}}^{T-1}) \). Combining these two inequalities with the definition of \( \gamma_{\text{init}} \) yields the result.
A.2.1 Initialization Set Size: Linear Kernel

For the linear kernel, there exists a logarithmic bound on the maximum information gain of a set of queries, precisely, \( \exists \eta \geq 0 : \gamma_t \leq \eta d \log (t + 1) \) (Srinivas et al., 2010). We attempt to initialize GP-BUCB with a set \( D_{init} \) of size \( T_{init} \), where, motivated by this bound and the form of Inequality (3.16), we assume \( T_{init} \) is of the form

\[
T_{init} = k\eta d(B - 1) \log B. \tag{A.3}
\]

We must show that there exists a \( k \) of finite size for which an initialization set of size \( T_{init} \) as in Equation (A.3) implies that any subsequent set \( S, |S| = B - 1 \), produces a conditional information gain with respect to \( f \) of no more than \( C \). This requires showing that the inequality \( \frac{B - 1}{T_{init}} \gamma_{T_{init}} \leq C \) holds for this choice of \( k \) and thus \( T_{init} \). Since we consider non-trivial batches, i.e., \( B - 1 \geq 1 \), if \( k \) is sufficiently large such that \( k\eta d(B - 1) \geq 1 \),

\[
\log (\log (B) + 1/((k\eta d(B - 1)))) \leq \log (\log (B) + 1) \leq \log B.
\]

Using Equation (A.3) and the bound for \( \gamma_{T_{init}} \), and following algebraic rearrangement, this inequality implies that if \( k\eta d(B - 1) \geq 1 \),

\[
\frac{B - 1}{T_{init}} \gamma_{T_{init}} \leq C \iff \frac{\log k}{k \log B} + \frac{\log \eta + \log d}{k \log B} + \frac{2}{k} \leq C.
\]

By noting that the maximum of \( \frac{\log k}{k} \) over \( k \in (0, \infty) \) is \( 1/e \) and choosing for convenience \( C = 2/e \), we obtain for \( k \geq 1/((\eta d(B - 1))) \):

\[
\frac{B - 1}{T_{init}} \gamma_{T_{init}} \leq \frac{2}{e} \iff \frac{1}{e \log B} + \frac{1}{k} \left( \frac{\log \eta + \log d + 2 \log B}{\log B} \right) \leq \frac{2}{e},
\]

or equivalently, choosing \( k \) to satisfy both constraints simultaneously,

\[
\frac{B - 1}{T_{init}} \gamma_{T_{init}} \leq \frac{2}{e} \iff k \geq \max \left[ \frac{1}{\eta d(B - 1)}, \frac{e(\log \eta + \log d + 2 \log B)}{2 \log (B) - 1} \right].
\]

Thus, for a linear kernel and such a \( k \), an initialization set \( D_{init} \) of size \( T_{init} \), where \( T_{init} \geq k\eta d(B - 1) \log (B) \), ensures that the hallucinated conditional information in any future batch of size \( B \) is \( \leq \frac{2}{e} \).
A.2.2 Initialization Set Size: Matérn Kernel

For the Matérn kernel, $\gamma_t \leq \nu t^\epsilon$, $\epsilon \in (0,1)$ for some $\nu > 0$ (Srinivas et al., 2010). Hence:

\[
\frac{(B-1)}{T_{\text{init}}} \gamma_{T_{\text{init}}} \leq C \iff \frac{\nu(B-1)(T_{\text{init}})^\epsilon}{T_{\text{init}}} \leq C \\
\iff \nu(B-1)(T_{\text{init}})^{\epsilon-1} \leq C \\
\iff T_{\text{init}} \geq \left( \frac{\nu(B-1)}{C} \right)^{1/(1-\epsilon)}.
\]

Thus, for a Matérn kernel, an initialization set $D_{\text{init}}$ of size $T_{\text{init}}$ implies that the conditional information gain of any future batch is $\leq C$. Choosing $C = 1$, we obtain the results presented in the corresponding row of Table 3.1.

A.2.3 Initialization Set Size: Squared Exponential (RBF) Kernel

For the RBF kernel, the information gain is bounded by an expression similar to that of the linear kernel, $\gamma_t \leq \eta (\log (t+1))^{d+1}$ (Srinivas et al., 2010). Again, motivated by Inequality (3.16), one reasonable choice for an initialization set size is $T_{\text{init}} = k\eta(B-1)(\log B)^{d+1}$. It is necessary to show that there exists a finite $k$ such that the conditional information gain of any post-initialization batch is $\leq C$. By a similar parallel argument to that for the linear kernel (Appendix A.2.1), and assuming that $B \geq 2$ and $k\eta(B-1) \geq 1$, it follows that

\[
\frac{B-1}{T_{\text{init}}} \gamma_{T_{\text{init}}} \leq C \\
\iff \log k + \log \eta + \log (B-1) \log [(\log B)^{d+1} + 1] \frac{1}{k^{1/(d+1)}(\log B)} \leq C^{1/(d+1)} \\
\iff \frac{\log k}{k^{1/(d+1)}(\log B)} + \frac{\log \eta}{k^{1/(d+1)}(\log B)} + \frac{(d+2) \log B}{k^{1/(d+1)}} \leq C^{1/(d+1)},
\]

where the last implication follows because for $a \geq 0, b \geq 1, (a^b + 1) \leq (a+1)^b$.

By noting that the maximum of $k^{-1/(d+1)} \log k$ over $k \in (0, \infty)$ is $(d+1)/e$ and choosing $C = (2(d+1)/e)^{d+1}$, we obtain for $k \geq 1/(\eta(B-1))$:

\[
\frac{B-1}{T_{\text{init}}} \gamma_{T_{\text{init}}} \leq \left( \frac{2d+2}{e} \right)^{d+1} \iff \frac{d+1}{e \log B} + \frac{1}{k^{1/(d+1)}} \left( \frac{\log \eta + (d+2) \log B}{\log B} \right) \leq \frac{2d+2}{e},
\]

or equivalently, incorporating the constraint $k \geq 1/(\eta(B-1))$ explicitly,

\[
\frac{B-1}{T_{\text{init}}} \gamma_{T_{\text{init}}} \leq \left( \frac{2d+2}{e} \right)^{d+1} \iff k \geq \max \left[ \frac{1}{\eta(B-1)}, \left( \frac{e(\log \eta + (d+2) \log B)}{(d+1)(2 \log (B-1))} \right)^{d+1} \right].
\]

Thus, for a Squared Exponential kernel and such a $k$, an initialization set $D_{\text{init}}$ of size $T_{\text{init}}$, ...
where $T_{\text{init}} \geq k\eta(B - 1)(\log(B))^{d+1}$, ensures that the hallucinated conditional information in any future batch of size $B$ is no more than $(2d+2)^{d+1}$.

### A.3 GP-AUCB: Finite Batch Size

In the absence of an explicitly specified maximum batch size, it is interesting to consider the scaling of batch sizes produced by GP-AUCB for large $T$. We are concerned with the case where actions are chosen when much is known about the structure of the reward function: many actions could be selected with little “danger” of choosing poorly, but also little information gain. In such a case, a great deal of regret could be accumulated between observations if the posterior mean fails to correctly order the available actions in $D$ with respect to their reward.

The set of size $T$ which gains the least information with respect to $f$, conditioned on observations $y(S)$, is one which queries $x^* = \arg\min_{x \in D} \sigma^2_n(x|y(S))$ $T$ times. These samples gain information $1/2 \log(1 + T\sigma_n^{-2}\sigma_S^2(x_*))$, where $\sigma^2(x|y_S)) = \sigma_S^2(x)$ is the posterior variance, conditioned on the observations $y_S$. Using this observation, if a batch is terminated when a threshold $C$ for hallucinated conditional information with respect to $f$ is exceeded, as in the GP-AUCB algorithm, the maximum possible length of a batch, $B_{\text{max}}$, can be bounded as follows:

$$C \geq 1/2 \log(1 + (B_{\text{max}} - 1)\sigma_n^{-2}\sigma_S^2(x_*)) = \log\left(\frac{\sigma^2_n}{\sigma_S^2(x_*)}\right) + 1 \geq B_{\text{max}}. \quad (A.4)$$

Thus, if there does not exist any $x \in D$ such that $\sigma^2(x) = 0$, which is the case under the GP model for any finite noise, this upper limit on $B_{\text{max}}$ is finite for any finite $C$ and any previous sampling history; the batch sizes of GP-AUCB do not diverge to infinity in a finite number of rounds.

Bounding the rate at which the batch length $B_{\text{max}}$ can grow is of interest, however. Consider cases where time is indexed by action number $t$ or by batch number $N$. In the case of iteration number, by rearrangement of Equation (3.14) and using Inequalities (3.11) and (3.13), we have

$$\sigma_t^2(x) \geq \sigma_0^2(x) \exp(-2I(f; y_{1:t-1})) \geq \sigma_0^2(x) \exp(-2\gamma_{t-1}) \forall t \in \mathbb{N}.$$ 

At time $t$, using this result and Inequality (A.4), the maximum length of the batch which can be constructed under GP-AUCB (or any sampling procedure such that the batch terminates when the information gain threshold $C$ is exceeded) is bounded as

$$B_{\text{max}} \leq \left[\frac{\sigma_n^2}{\min_{x \in D}(\sigma_0^2(x))}\exp(2C) - 1\right]\left[\exp(2\gamma_{t-1})\right] + 1.$$ 

This bound is $O(\exp(tC))$, since $\gamma_t$ is no more than linear in $t$. 
A similar bounding result may be obtained for the $N$th batch. After $N - 1$ batches, the posterior variance of $f(x)$, $\sigma^2_{N-1}(x)$, may be bounded as follows, for any $x \in D$, via Equation (3.14) and Inequalities (3.11) and (3.13):

$$\sigma^2_{N-1}(x) \geq \sigma^2_0(x) \exp(-2(N - 1)C_B) \forall t \in \mathbb{N}.$$

Here, $C_B$ is an upper bound on the information which is obtained when the observations corresponding to the batch are made. $C_B$ is greater than $C$, since the batch terminates only when the information which would be hallucinated in order to select the next action exceeds the threshold $C$. One useful bound is $C_B \leq C + 1/2 \log (1 + \sigma_n^{-2} \max_{x \in D} \sigma^2_0(x))$, which follows because the termination condition is checked every round and mutual information is submodular. Using Equation (A.4), the length of the $N$th batch is thus bounded as

$$B_{\max} \leq \left[ \frac{\sigma^2_n}{\min_{x \in D} (\sigma^2_0(x))} [\exp(2C) - 1][\exp (2(N - 1)C_B)] \right] + 1,$$

which is $O(\exp(NC))$, but is bounded for finite batch number.
Appendix B

Tabulated Computational Results

B.1 Tables of Results from Experiments

These experiments are described in detail in Section 3.6. Tables of numerical results are presented here; these include the regret (or elapsed time) with the standard error. Each table presents the results of each data set and algorithm combination tested for a particular experimental setting, averaged over 200 runs. Minimum regret of zero indicates that the optimal set was visited by every run.

Table B.1: Average (AR) and Minimum regret (MR) for fixed batch size $B = 5$.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Algorithm</th>
<th>AR, Query 100</th>
<th>MR, Query 100</th>
<th>AR, Query 200</th>
<th>MR, Query 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>GP</td>
<td>0.0187 ± 0.0010</td>
<td>0.0180 ± 0.0009</td>
<td>0.0181 ± 0.0009</td>
<td>0.0179 ± 0.0008</td>
</tr>
<tr>
<td>GP-BUCB</td>
<td>0.0702 ± 0.0009</td>
<td>0.0698 ± 0.0008</td>
<td>0.0700 ± 0.0008</td>
<td>0.0698 ± 0.0008</td>
<td></td>
</tr>
<tr>
<td>SM-MEI</td>
<td>0.0590 ± 0.0011</td>
<td>0.0580 ± 0.0010</td>
<td>0.0592 ± 0.0010</td>
<td>0.0581 ± 0.0009</td>
<td></td>
</tr>
<tr>
<td>SE-GP</td>
<td>GP</td>
<td>0.0342 ± 0.0013</td>
<td>0.0335 ± 0.0012</td>
<td>0.0344 ± 0.0013</td>
<td>0.0337 ± 0.0012</td>
</tr>
<tr>
<td>GP-BUCB</td>
<td>0.0661 ± 0.0014</td>
<td>0.0656 ± 0.0013</td>
<td>0.0662 ± 0.0014</td>
<td>0.0657 ± 0.0013</td>
<td></td>
</tr>
<tr>
<td>SM-MEI</td>
<td>0.0557 ± 0.0013</td>
<td>0.0552 ± 0.0012</td>
<td>0.0560 ± 0.0013</td>
<td>0.0556 ± 0.0012</td>
<td></td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>GP-BUCB</td>
<td>0.0001 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
</tr>
<tr>
<td>GP-BUCB</td>
<td>0.0001 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td></td>
</tr>
<tr>
<td>SM-MEI</td>
<td>0.0001 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td></td>
</tr>
</tbody>
</table>

Table B.2: Average (AR) and Minimum regret (MR) for fixed delay length $B = 5$.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Algorithm</th>
<th>AR, Query 100</th>
<th>MR, Query 100</th>
<th>AR, Query 200</th>
<th>MR, Query 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>GP</td>
<td>0.0187 ± 0.0010</td>
<td>0.0180 ± 0.0009</td>
<td>0.0181 ± 0.0009</td>
<td>0.0179 ± 0.0008</td>
</tr>
<tr>
<td>GP-BUCB</td>
<td>0.0702 ± 0.0009</td>
<td>0.0698 ± 0.0008</td>
<td>0.0700 ± 0.0008</td>
<td>0.0698 ± 0.0008</td>
<td></td>
</tr>
<tr>
<td>SM-MEI</td>
<td>0.0590 ± 0.0011</td>
<td>0.0580 ± 0.0010</td>
<td>0.0592 ± 0.0010</td>
<td>0.0581 ± 0.0009</td>
<td></td>
</tr>
<tr>
<td>SE-GP</td>
<td>GP</td>
<td>0.0342 ± 0.0013</td>
<td>0.0335 ± 0.0012</td>
<td>0.0344 ± 0.0013</td>
<td>0.0337 ± 0.0012</td>
</tr>
<tr>
<td>GP-BUCB</td>
<td>0.0661 ± 0.0014</td>
<td>0.0656 ± 0.0013</td>
<td>0.0662 ± 0.0014</td>
<td>0.0657 ± 0.0013</td>
<td></td>
</tr>
<tr>
<td>SM-MEI</td>
<td>0.0557 ± 0.0013</td>
<td>0.0552 ± 0.0012</td>
<td>0.0560 ± 0.0013</td>
<td>0.0556 ± 0.0012</td>
<td></td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>GP-BUCB</td>
<td>0.0001 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
</tr>
<tr>
<td>GP-BUCB</td>
<td>0.0001 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td></td>
</tr>
<tr>
<td>SM-MEI</td>
<td>0.0001 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td></td>
</tr>
<tr>
<td>Data Set</td>
<td>Algorithm</td>
<td>AR, Query 100</td>
<td>MR, Query 100</td>
<td>AR, Query 200</td>
<td>MR, Query 200</td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>SCIC</td>
<td>GP-BUCB, B = 5</td>
<td>0.1405 ± 0.0033</td>
<td>0.0177 ± 0.0013</td>
<td>0.1405 ± 0.0013</td>
<td>0.0177 ± 0.0013</td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>GP-BUCB, B = 10</td>
<td>0.0558 ± 0.0004</td>
<td>0.0000 ± 0.0000</td>
<td>0.0558 ± 0.0004</td>
<td>0.0000 ± 0.0000</td>
</tr>
<tr>
<td>Conklin</td>
<td>GP-BUCB, B = 5</td>
<td>0.2010 ± 0.0013</td>
<td>0.0000 ± 0.0000</td>
<td>0.2010 ± 0.0013</td>
<td>0.0000 ± 0.0000</td>
</tr>
<tr>
<td>Vaccine, Dengue</td>
<td>GP-BUCB, B = 5</td>
<td>0.0413 ± 0.0046</td>
<td>0.0040 ± 0.0014</td>
<td>0.0413 ± 0.0046</td>
<td>0.0040 ± 0.0014</td>
</tr>
<tr>
<td>MC</td>
<td>GP-BUCB, B = 5</td>
<td>0.2348 ± 0.0104</td>
<td>0.0062 ± 0.0026</td>
<td>0.2348 ± 0.0104</td>
<td>0.0062 ± 0.0026</td>
</tr>
</tbody>
</table>

Table B.3: Average (AR) and Minimum regret (MR) for batch sizes B = 5, 10, and 20, non-adaptive algorithms.
Table B.4: Average (AR) and Minimum regret (MR) for maximum adaptive batch sizes $B_{\text{max}} = 5$, 10, and 20.
Table B.5: Average (AR) and Minimum regret (MR) for delay lengths $B = 5, 10, \text{and } 20.$
<table>
<thead>
<tr>
<th>DATA SET</th>
<th>ALGORITHM</th>
<th>QUERY 40</th>
<th>QUERY 100</th>
<th>QUERY 200</th>
<th>QUERY 400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spinal Cord Therapy</td>
<td>GP-UCB</td>
<td>0.594</td>
<td>1.042</td>
<td>1.404</td>
<td>1.946</td>
</tr>
<tr>
<td></td>
<td>GP-BUCB</td>
<td>0.591</td>
<td>0.973</td>
<td>1.032</td>
<td>1.047</td>
</tr>
<tr>
<td></td>
<td>GP-AUCB Lazy</td>
<td>0.589</td>
<td>0.762</td>
<td>0.858</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>GP-AUCB Lazy</td>
<td>0.593</td>
<td>0.838</td>
<td>0.966</td>
<td>1.072</td>
</tr>
<tr>
<td></td>
<td>HBBO MEI</td>
<td>0.570</td>
<td>0.549</td>
<td>0.563</td>
<td>0.567</td>
</tr>
<tr>
<td></td>
<td>GP-AUCB Lazy</td>
<td>0.579</td>
<td>0.596</td>
<td>0.601</td>
<td>0.607</td>
</tr>
</tbody>
</table>

Table B.6: Mean wall-clock execution times and standard deviations of estimate (S).
Appendix C

Action-matched Animal Plots

During two of the animal experiment runs (animal 5, run 1 and animal 7) a substantial number of actions performed by the human experimenter were missed or dropped. An alternate view of these experiments is presented here, where “pass” actions are inserted for those missed by either the algorithm or the human experimenter. Without inserted passes, the same action indices for the human and algorithm do not correspond to the same point in time, and visual interpretation of the regret plots is difficult; after inserted passes, this synchrony is restored.
Figure C.1: Action-matched plots for animal 5, run 1 and animal 7. A missed action is treated as a “pass” to restore action synchrony between the human experimenter and the algorithm. These actions are shown with an “x” in the plots above. C.1(a) & (c): Animal 5, run 1. During this experimental run, the human experimenter missed a full day of experiments (P15). Compare these plots to Figures 4.8(b) and (d), which do not have the passes corresponding to the three missed batches on P15. (b) & (d): Animal 7. The human experimenter did not conduct a fourth and final batch on the second testing day (P12). Several actions were also missing from the third testing day, P13. These plots are action compensated versions of Figures 4.11(b) and (d).
Appendix D

Toward Human Studies: Mathematical Results

D.1 Decision-making with an Aggregated Objective

When trying to use several, possibly related GP models $f_1, \ldots, f_n$ to make a decision about a known function $r$ of those individual GPs, it is natural to attempt to apply a UCB-like approach to the problem. Unfortunately, unless $r$ is a linear combination of these individual GPs, $r(f)$ is not itself a Gaussian process, nor is the posterior over $r(f(x))$ Gaussian. This problem even arises if one common and natural formulation of a reward function, that of penalized deviation from a target $t$ via a weighted norm term, is used, e.g.,

$$r(f(x)) = -\sqrt{(f(x) - t)^TW(f(x) - t)},$$

(D.1)

where $W$ is a symmetric, positive definite penalty matrix, such that $r$ is $-1$ times a weighted 2-norm in $\mathbb{R}^n$. Such an objective function makes a great deal of sense in terms of convex optimization, and has a unique global maximum at $f(x) = t$. Further, for any $x$, the posterior over $f(x)$ is $f(x) \sim N(\mu(x), \Sigma(x))$, and as $\mu - t$ becomes very large, the distribution of the weighted squared norm begins to look like the corresponding marginalization of the posterior onto the unit vector in the direction $\mu - t$; this marginal distribution is a Gaussian. However, of the most interest in terms of active learning is the region near the optimum, where such an $f(f)$ is most strongly non-Gaussian.

Inspired by GP-UCB and GP-BUCB, it seems reasonable that it would be desirable to create a decision function of form

$$x_t = \arg\max_{x \in D} \left[ \mathbb{E}[r(f(x))|y_{1:s}] + \beta_t^{1/2} \sqrt{\text{Var} \left( r(f(x))|y_{1:s} \right)} \right],$$

(D.2)

which once again trades off exploitation, captured by the mean reward term on the left, with exploration, captured by the standard deviation term on the right. It is possible to calculate
\[
E[r(f(x))^2|y_{1:t}] = E[(f(x) - t)^TW(f(x) - t)|y_{1:t}] = (\mu_{0[t]}(x) - t)^TW(\mu_{0[t]}(x) - t) + \text{trace}(W\Sigma_{0[t]}).
\]
This leaves calculating the expected reward, \(E[r(f(x))|y_{1:t}]\). Unfortunately, despite its relation to the \(\chi\) distribution, I was unable to obtain a general expression for this expectation, and so I made recourse to bounding arguments. By use of Jensen’s inequality, which states that for a convex function \(h(x)\),
\[
E[h(x)] \geq h(E[x]),
\]
it is possible to derive an upper bound
\[
E[r(f(x))|y_{1:t}] \leq -\sqrt{(\mu_{0[t]}(x) - t)^TW(\mu_{0[t]}(x) - t)}
\]
via the concavity of \(r\) with respect to \(f\), as well as a lower bound
\[
E[r(f(x))|y_{1:t}] \geq \sqrt{(\mu_{0[t]}(x) - t)^TW(\mu_{0[t]}(x) - t) + \text{trace}(W\Sigma_{t-1})}
\]
by noting that \(-\sqrt{r}\) is convex with respect to \(r\) over \(r \in \mathbb{R}^+\). Using the definition of the variance in terms of the expectation of the square and the square of the expectation, and substituting in the upper bound on \(E[r(f(x))|y_{1:t}]\), it can be shown that
\[
\text{Var} \left( r(f(x))|y_{1:t} \right) = E[r(f(x))^2|y_{1:t}] - E[r(f(x))|y_{1:t}]^2 \leq \text{trace}(W\Sigma_{t-1}). \tag{D.4}
\]
By analogy to the GP-UCB and GP-BUCB decision rules (Equations 3.5 and 3.7), and using the upper bounds above to create a term capturing the reward and another term capturing the uncertainty, this suggests a decision rule of form
\[
x_t = \text{argmax}_{x \in D} \left[ -\sqrt{(\mu_{0[t]}(x) - t)^TW(\mu_{0[t]}(x) - t) + \beta_t^{1/2}\sqrt{\text{trace}(W\Sigma_{t-1}(x))}} \right] \tag{D.5}
\]
should have useful characteristics. Note that for the scalar case, with a very large target \(t\), the decision rule reduces to that of GP-BUCB. Further, as we learn more and more about the function near the optima, \(\Sigma_{t-1}(x)\) should decrease for these decisions, making the upper bound on \(E[r(f(x))|y_{1:t}]\) tighter. Additionally, for \(\mu_{t-1}(x) - t\) large, the decision rule can also be expected to closely bound the actual form in Equation D.2, allowing us to disregard these decisions as poor-performing. This leaves the poor cases as those in which \(\Sigma_{t-1}(x)\) is very large and \(\mu_{t-1}(x) - t\) is small; in this case, overestimating either the mean or variance should result in allocation of observations to these actions, driving down \(\Sigma_{t-1}(x)\) and resolving the issue through observation. This decision rule is also practical because it is very easy to calculate; if the posterior over the values \(f(x)\) is available, this decision rule simply requires some linear algebraic calculations. Further, it can be shown that, under the assumption that observations can only be added to the observation
set, but none can leave, the term \( \sqrt{\text{trace}(W \Sigma_t(x))} \) is non-increasing (see Appendix D.2). Because this term is non-increasing as observations are added to the observation set, the calculation of \( \sqrt{\text{trace}(W \Sigma_t(x))} \) can be done lazily, as can be done for the standard deviation in Section 3.5, enabling substantial computational savings.

### D.2 Proof Multi-Muscle Uncertainty Term is Non-Increasing

A quite useful characteristic of the GP-BUCB decision rule, Equation (3.7), is that the uncertainty term (i.e., the standard deviation) cannot increase as observations are added. For computational reasons, it is important to demonstrate that Equation (D.5) also has the same property. As a first step, we define the matrix root of the weight matrix \( W \) as a symmetric, positive definite matrix \( W^{1/2} = (W^{1/2})^T > 0 \), such that \((W^{1/2})^2 = W\). Such a matrix can be constructed by noting that \( W = V D V^T \), where \( V \in \mathbb{R}^{n \times n} \) is a matrix whose columns are the set of orthonormal right eigenvectors of \( W \) and \( D \) is the diagonal matrix of the corresponding eigenvalues of \( W \); choosing \( W^{1/2} = V D^{1/2} V^T \) produces the desired properties. We then consider the uncertainty of the algorithm's estimate of \( f(x) \) after steps \( t \) and \( t' \) of the algorithm, where \( t' > t \), the corresponding observation sets \( y_t \) and \( y_{t'} \), and the sets of past actions \( X_t \) and \( X_{t'} \). We may describe the posterior covariance function between stimuli \( x \) and \( x' \) and muscle indices \( i \) and \( j \) as

\[
 k_t((x, i), (x', j)) = k((x, i), (x', j)) y_t \quad \text{and write the posterior covariance matrix at time } t' \text{ for } f(x) \text{ as}
\]

\[
 \Sigma_t(x) = \Sigma_t(x) - k_t(K + \sigma_n^2 I)^{-1} k_t^T,
\]

where \( X_{t+1:t'} = X_{t'} - X_t \) is the set of observations occurring between times \( t \) and \( t' \), \( k_t = k_t(x, X_{t+1:t'}) \in \mathbb{R}^{n \times [(t' - t) \times n]} \) is the covariance between the observations associated with \( X_{t+1:t'} \), including all \( n \) channels, and \( K_t = K_t(X_{t+1:t'}, X_{t+1:t'}) \in \mathbb{R}^{[(t' - t) \times n] \times [(t' - t) \times n]} \) is the posterior covariance at time \( t \) between the noisy observations \( y_{t+1:t'} \) of \( f(X_{t+1:t'}) \) in \( X_{t+1:t'} \). Multiplying left and right by \( W^{1/2} \), and then using the linearity of the trace and its invariance to circular permutations of symmetric matrices, we obtain

\[
 \text{trace}(W \Sigma_t(x)) = \text{trace}(W \Sigma_t(x)) - \text{trace}(W^{1/2} k_t(K_t + \sigma_n^2 I)^{-1} k_t^T W^{1/2}). \tag{D.6}
\]

Noting that \( h = W^{1/2} k_t y_{t+1:t'} \in \mathbb{R}^{n \times 1} \) is a linear combination of the \( n(t' - t) \) multivariate Gaussian observations, \( h \) also has a multivariate Gaussian distribution, such that its covariance matrix, \( W^{1/2} k_t(K_t + \sigma_n^2 I)^{-1} k_t^T W^{1/2} \), is positive semi-definite. Since the trace of this matrix is therefore non-negative, it follows from Equation (D.6) that \( \text{trace}(W \Sigma_t(x)) \leq \text{trace}(W \Sigma_t(x)) \) for \( t' > t \).
D.3 Path-Based Decision Rules

As discussed in Section 5.4.2.3, it may be desirable to plan for smooth paths of length no more than $B$ which travel from the present stimulus state through the decision set. This set of possible paths may be denoted $L$. While there are potentially exponentially many paths through the graph of possible stimuli, if there is a set of restrictions on path construction such that each end-point (i.e., $x \in D$) may be reached by at most one path, these restrictions imply $|L| \leq |D|$. One reasonable idea for selecting paths from $L$ is to extend the decision rule used by the GP-BUCB algorithm, resulting in the following equation:

$$X_t = \text{argmax}_{X \in L} \left[ \sum_{\tau = t}^{t+B-1} \left( \mu_{\theta_t}(x_{\tau}) + \frac{\beta_{\theta_t}^{1/2}}{\beta_{\theta_t}} \sigma_{\tau-1}(x_{\tau}) \right) \right], \quad (D.7)$$

where $X = \{x_t, \ldots x_{t+B-1}\}$. This construction follows the form of the GP-BUCB decision rule, and might even be amenable to the same confidence interval analysis, at least locally. However, this decision rule may philosophically differ from the GP-BUCB rule in that the quantity which corresponds to information gained no longer maps easily to the actual information gain $I(f; y(X)|y(X_{\theta_t}))$.

Motivated by the transformation between $\sigma_{t-1}(x_t)$ and $I(f; y(x_t) | y(X_{\theta_t}))$, i.e.,

$$\sigma_{t-1}(x_t) = \sigma_n \sqrt{-1 + \exp(I(f; y(x_t) | y(X_{\theta_t})))}, \quad (D.8)$$

it may be reasonable to apply the same transformation to $I(f; y(X)|y(X_{\theta_t}))$ to obtain a quantity $e(X)$ which corresponds to the information gain from the group of observations as follows:

$$e(X) = \sigma_n \sqrt{-1 + \exp(I(f; y(X)|y(X_{\theta_t})))}$$

$$= \sigma_n \sqrt{-1 + \prod_{\tau = t}^{t+B-1} (1 + \sigma_{\tau}^{-2} \sigma_{\tau-1}(x_{\tau}))} \quad (D.9)$$

where $X$ is again $X = \{x_t, \ldots x_{t+B-1}\}$. This yields a decision rule of the form

$$X_t = \text{argmax}_{X \in L} \left[ \sum_{\tau = t}^{t+B-1} \left[ \mu_{\theta_t}(x_{\tau}) + \frac{\beta_{\theta_t}^{1/2}}{\beta_{\theta_t}} e(X) \right] \right]. \quad (D.10)$$

In either or both of these cases, it might be appropriate to consider the possibility that the experiment might have to be stopped during the traversal of $X$ with some probability. Letting the uniform probability of failure of each transition be $1 - \lambda$, and assuming the reward and observation are
obtained even if the individual action is a failure, the discounted version of the first decision rule is

\[
X_t = \arg\max_{X \in L} \left[ \sum_{\tau=t}^{t+B-1} \left[ \lambda^{\tau-t} (\mu_{b[\tau]}(x_{\tau}) + \beta^{1/2} (X)) \right] \right].
\]  (D.11)

This decision rule has been implemented for a version of the human experimental code which is intended to search over the space of voltage and frequency parameters corresponding to a fixed set of active electrodes. Similarly, the discounted version of Equation (D.9), designated \( e_\lambda(X) \), is

\[
e_\lambda(X) = \sigma_n \left( -1 + \prod_{\tau=t}^{t+B-1} \left[ (1 + \sigma_n^2 \sigma_{\tau-1}(x_{\tau}))^{\lambda_{\tau-1}} \right] \right)
\]  (D.12)

and the corresponding decision rule becomes

\[
X_t = \arg\max_{X \in L} \left[ \sum_{\tau=t}^{t+B-1} \left[ \lambda^{\tau-t} \mu_{b[\tau]}(x_{\tau}) + \beta^{1/2} \right] \right].
\]  (D.13)

Either of these frameworks may make sense as a method of selecting paths through the stimulus space.
Appendix E

Code Availability

Code implementing the algorithms discussed in Chapter 3 is available at www.its.caltech.edu/~tadesaut/.