Chapter 5

Toward Human Studies

5.1 Organization

The ultimate goal of this line of research is to develop a new methodology for automating human SCI therapy. This chapter lays out a roadmap of issues which must be resolved before full-scale human applications can be realized and provides suggestions for extending the approach of Chapter 4 to human experiments and therapy. Section 5.2 provides an overview of the existing human SCI therapy experiments to which this work could be applied. The experimental framework for the pilot studies executed so far is laid out in Section 5.3. Discussions of several specific technical issues and the solutions employed follow in Section 5.4. Section 5.5 gives results of the pilot experiments and discusses the significance of these results for the continued development of this approach. Finally, a number of extensions to the current techniques are discussed in Section 5.6.

5.2 Prior Human Experiments

The following discussion is based upon the work of our collaborators at the Frazier Rehab Institute and University of Lousiville in Louisville, Kentucky (see Harkema et al., 2011). Three major different types of experiments under multi-electrode stimulation are undertaken at Frazier Rehab: supine, standing, and stepping experiments. These experiments are carried out in SCI humans with paraplegia, each implanted with a RestoreAdvanced neurostimulator (Medtronic, Minneapolis, MN) connected to a Specify 5-6-5 electrode array (Medtronic) positioned over the lumbar enlargement of the spinal cord. This dual-component device was originally developed for chronic pain therapy, but has been adapted to this application.

The supine experiments involve the participant lying in a supine position as the electrical stimulus is changed to any of a variety of configurations in the stimulus space; the free parameters include the combination of electrodes used as cathodes and anodes, and the voltage, frequency, and pulse-width of the stimulus. This experiment is intended to "map out" the spinal cord's motor pools with

respect to the active electrodes. In many ways, this experiment could be treated as analogous to a multi-muscle, multi-electrode version of the current rat experiments, trying to find regions of the stimulus space which have large activations of particular muscles, or which produce a pattern of muscle activation matching a specified target. For this reason, these experiments might be a natural extension of the work of Chapter 4. However, the supine experiments are uncomfortable and boring for the participants; thus, only a limited amount of data from them has been collected, and relatively few opportunities to apply a machine learning algorithm to the supine experiments in closed-loop would be available.

Standing and stepping experiments present a substantially different sort of challenge than the supine experiments; in particular, the maintenance of a complex motor behavior is a much more challenging problem than the creation of a particular pattern of muscle activation, or the maximization of a particular muscle's activity. While it may be possible to describe stable standing as a particular, relatively constant pattern of muscle activation, it may also be that the body's muscular responses to perturbation are actually more important than those in equilibrium. Extracting this sort of information from EMG in an automated fashion may be very difficult. Stepping is a complex, cyclic activity, which, as in animals, is difficult to grade effectively, particularly at short timescales. Motion capture may be a useful means of assessing stepping performance, but motion capture is very time-consuming to manually process and so an automated motion capture analysis system may be necessary. As with standing, it may be that what is truly desirable in stepping is not simply the basic pattern of motor activity, but the ability to respond to perturbation while stepping. Fortunately, the basic stepping kinematic and motor pattern is fairly distinctive and responses to perturbations are relatively small, such that a performance measure could likely be created which responded to grossly correct or incorrect stepping, rather than needing to focus on the fine details of responses to rare and random events.

5.3 Pilot Applications of GP-BUCB to Human SCI Therapy: Introduction

Our pilot experiments focused on the use of GP-BUCB and variants of that algorithm in the context of stand training. While standing is a rhythmless motor behavior, with difficult-to-quantify success, human standing under epidural stimulation is somewhat understood and stand training is beneficial for the patients. As an additional point in favor of standing as an experimental target, stand training is conducted independently by the patients at home as an exercise, suggesting that monitoring and optimizing standing could fit into the daily routine of SCI patients undergoing EES therapy.

During an experimental session, the participant stands with assistance from the stimulator. For lateral stability and safety, the participant stands in the middle of a U-shaped apparatus, a

stand frame, which can also be used by the participant to assist the sit-to-stand transition. One important use of the standing sessions in the clinic is the modification of stimulus parameters to improve standing performance. These parameters, including the set of active electrodes and their assigned polarities, the master stimulus voltage (the RestoreAdvanced either applies a voltage of $-v$ or v to each active electrode, where v is the master voltage), and the stimulus frequency, may be varied during the experiment. Since the effect of changing the stimulus is nearly immediate, observations of the patient's responses can potentially provide feedback on the resulting performance. Previously, this feedback process has involved human expert experimenters interacting with the patient, observing the patient visually, and monitoring the EMG activity in the patient's leg muscles. Using these observations, the clinicians and scientists carefully and gradually modify the stimulus to aid standing, while also maintaining safety and the efficacy of the session as exercise.

One major constraint during this optimization process, in part due to the limitations of the repurposed Medtronic hardware, is that the stimulus must be temporarily stopped when the pattern of active stimulus electrodes is changed. This means that during the interim the participant must either support him- or herself (typically with the arms alone) or must sit. This transition may also be disruptive to the neurological state of the spinal cord. Radical transitions may cause collapse from standing, and, even for more gradual transitions, the spinal cord requires on the order of seconds to one minute to acclimate to a new set of stimulus parameters. This potential for disruption is the reason the stimulus parameters are changed slowly. Given these constraints, the pilot experiments have so far studied the problem of exploring and exploiting over the space of master voltage and stimulus frequency, with a fixed combination of active stimulus electrodes during this process and only slow changes in voltage and frequency.

A description of the pilot experiments so far performed follows. In Section 5.4, the necessary mathematical infrastructure is described, including performance measures (Section 5.4.1), some required extensions to the GP-BUCB algorithm (Section 5.4.2), and the novel covariance functions which were created for this problem (Section 5.4.3). Preliminary results and a discussion of these are presented in Section 5.5.

5.4 Mathematical Methods

As in Chapter 4, it is necessary to make modifications to the existing theoretical and mathematical framework in order to produce an experimentally useful implementation. In particular, formulating a useful measure of standing performance is a very challenging problem; the methods used in the preliminary experiments are described in Section 5.4.1, but even these methods have not proven entirely satisfactory. Some alternatives are discussed both in Section 5.4.1 and in Section 5.6.2. The problem of making decisions in a rigorous fashion using these alternative performance measures is

also not trivial; this problem is examined and solutions used in the preliminary experiments are described in Section 5.4.2. A further complication, described and addressed in Section 5.4.3, is the need to specify problem-specific covariance functions. All code implementing the algorithm was implemented in MATLAB (The MathWorks, Inc., Natick, MA).

5.4.1 Performance Measures

As schematically shown in Figure 1.3, one of the key components of the EES system is the choice of what sensor information and methods are used to quantify performance of the stimulus, both in the execution of the desired motor task and in terms of any other variables of interest, e.g., the patient's comfort. This decision is crucial. Clearly, if the reward function optimized by the algorithm is not reflective of actual performance, the algorithm cannot optimize performance; at best, it will optimize this putative performance measure. The difficulty lies in choosing an appropriate performance measure. In Chapter 4, it is assumed that an appropriate performance measure is the peak-to-peak amplitude of activation of the left tibialis anterior muscle, which has the advantage of being a scalar function. However, this function might be maximized by stimuli which are uncomfortable or are otherwise unacceptable in humans, or which are simply not useful therapeutically. A number of possible alternatives are discussed in the following sections.

5.4.1.1 Subjective Ratings

One reasonable way to optimize the performance of a stimulation system is to simply ask the user for his or her opinion; if the user's ratings are both repeatable and reflective of actual therapeutic utility, than this method requires relatively little infrastructure, as well as giving patients a way to control their own therapy. User feedback could perhaps be beneficial in terms of reducing the number of clinical visits required, reducing patient frustration, and increasing the patients' sense of agency in their own recovery. It is quite plausible that patients could be trained in the use of a grading scale or rubric, as simple as integers from zero to ten, which would yield repeatable results; similar methods have been used for gait analysis in animals (e.g., Basso et al., 1995), and a human user could much more effectively assess some important aspects of the stimulus response, such as discomfort, than could a fully autonomous system. Under such a system, the object being regressed upon as a function of stimulus $\boldsymbol{x}$ is $f(\boldsymbol{x})$, the user's likely rating for that stimulus. In the preliminary experiments described here, a very simple zero to ten system was used with both subjects.

Several difficulties with such ratings are apparent. First, a rating system would require the aforementioned patient training, and so would likely not be useful without the user acquiring a substantial degree of experience with the stimulator and their personal, subjective experience of the system's effects. This characteristic would render it quite difficult to use such a simple grading system during the initial training period, already the most challenging period for autonomous search

of the stimulus parameter space. Secondly, it would likely be very difficult to develop a system under which the quantitative grades assigned by the user corresponded directly with utility. Certainly, it might quite plausibly be possible to develop a system which was ordinal, i.e., the nominal ordering of the categories in the rubric was correct in terms of utility, but it would likely be substantially more challenging to develop a system under which, for example, the difference in utility between an 8 and a 7 was the same as that between a 4 and a 3. Some discussion of the difficulties inherent in this non-equivalence of the rating and utility are discussed in Section 5.4.2.1. Further, if different aspects of the stimulus response are to be graded, e.g., performance and comfort, it might be difficult to determine how to balance these aspects against one another. A more precise and expressive rating system would also likely require a greater time for the user to assess and respond to any given stimulus, slowing the rate of testing. Without this detail, however, it might be very hard to use ratings to diagnose necessary changes to improve the stimulus, a major disadvantage as compared to muscle-activity-based systems. Additionally, if the stimulus applied was uncomfortable or dangerous, a system without its own sensing capabilities could harm the user without giving him or her the time or ability to respond appropriately. This would require careful design of safety systems. However, these difficulties aside, a performance measure based on subjective ratings is a reasonable approach for inexpensive and simple home use, particularly if it could be incorporated as part of a two-tier system for clinical and home adjustment of stimulus parameters. Medtronic's line of neurostimulators for chronic pain managment, for example, includes just such a two-tier system, though this system is for direct control of stimulus parameters, rather than user feedback; they manufacture both a complex and capable clinical control device, the N'Vision, and the MyStim, a simpler, less capable controller given to patients for home use. A simple, handheld unit like the MyStim would be an excellent interface device for a rating-based, closed-loop system.

5.4.1.2 Grading Vector-valued EMG

Performance must necessarily be quantified as a scalar in order for optimization of the performance to be a meaningful concept. Because the idea of standing performance itself is somewhat difficult to define, attempting to measure and work with performance directly is difficult, thus leading to the idea of using a holistic surrogate such as a user rating, discussed previously. Another alternative is to define performance as a known function of measurable physical quantities. Indeed, it is plausible that these measured variables, e.g., EMG signals, can be chosen in such a fashion as to have distinct patterns which result in useful, high-level behavior, such as naturalistic standing.

Mathematically, such a system must regress upon a vector-valued object, i.e., $\mathbf{f}(\boldsymbol{x}) \in \mathbb{R}^n$, such that there are n (possibly linked) observations which arise as the result of a single input $\boldsymbol{x} \in D$. In the case of interest, $\mathbf{f}(\boldsymbol{x})$ is vector of extracted features of several muscles' responses to a single EES

stimulus $\boldsymbol{x}$. From such a vector-valued object, one fairly natural way define the reward is

$$r(\mathbf{f}(\boldsymbol{x})) = -\sqrt{(\mathbf{f}(\boldsymbol{x}) - \mathbf{t})^T W (\mathbf{f}(\boldsymbol{x}) - \mathbf{t})}, \tag{5.1}$$

where W is a symmetric, positive definite penalty matrix, such that $r(\boldsymbol{x})$ is the negation of a weighted 2-norm of $(\mathbf{f}(\boldsymbol{x}) - \mathbf{t})$, the difference between $\mathbf{f}(\boldsymbol{x})$ and a target response $\mathbf{t}$ in $\mathbb{R}^n$. Careful selection of the feature representation, W, and $\mathbf{t}$ should allow relatively precise tailoring of the reward function to reflect more or less acceptable deviations from a particular motor behavior. In order to formulate a GP-UCB-like algorithm which uses such a reward function, it is necessary to create a regression model using Gaussian processes that captures the variation of $\mathbf{f}(\boldsymbol{x})$ with respect to $\boldsymbol{x}$. If these features are chosen appropriately, it may be relatively easy to incorporate expert knowledge, e.g., by the selection of prior means and covariance functions.

One of the most important problems associated with such a model for $\mathbf{f}(\boldsymbol{x})$ is the problem of appropriately interlinking the individual GPs corresponding to the different entries in the vector $\mathbf{f}$. Anatomically, the nerves, interneurons, and motor pools responsible for activity in different muscles are located in close proximity to one another, such that nearby structures may be brought to threshold at roughly the same amplitude of stimulation. Thus, different variables of interest can be expected to co-vary due to their co-dependence on the same physical phenomena. Capturing this structure is important, and can lead to greater efficiency, particularly in the case in which sensors or channels of information are intermittently unavailable. One can think of the identity of the muscle or feature as a piece of side information supplied to the model. Recently, Krause and Ong (2011) approached this problem of using side information from the GP-bandit perspective. In their setting, in each round, the algorithm is presented with a *context* m_t from the set of possible contexts M and must then choose an action $\boldsymbol{x}_t$ to take in that round to maximize reward (where the reward depends on the context and the action). Krause and Ong formulate an algorithm, CGP-UCB, which uses covariance functions on the context space M to enable regression on a *contextual Gaussian process* over $D \times M$. Decisions in the SCI therapy setting yield a reward which is assumed in this section to be dependent on all of the contexts simultaneously, thus making employment of the complete CGP-UCB algorithm inappropriate, due to its assumption that a single context is in effect at any time. However, the core contextual GP model upon which CGP-UCB is based holds substantial promise for the multi-muscle or multi-characteristic problem. Using this model requires covariance functions which are capable of expressing the covariance of $f(\boldsymbol{x}, m)$ and $f(\boldsymbol{x}', m')$. Some ideas for anatomically appropriate covariance functions are discussed in Section 5.4.3.

In the preliminary experiments, some trials were also carried out using a reward function based on a GP model of an EMG feature vector over the allowed stimulus space. This necessitated formulating a decision rule which used the reward function and the GP model together to trade off exploration

and exploitation. The chosen decision rule is discussed in Section 5.4.2.2.

5.4.2 Algorithmic Extensions

Some important extensions to the theory described by Srinivas et al. (2010) and Chapter 3 should be considered in light of the requirements for human application. As in the case of GP-BUCB, it is useful to consider alternate decision rules, as well as the circumstances under which probabilistic guarantees regarding the convergence of algorithms using a GP posterior and these decision rules can be demonstrated. Another important extension is the creation of an algorithm for selecting smoothly varying paths of stimuli, rather than the unconstrained transitions in the conventional GP-BUCB and GP-AUCB algorithms.

5.4.2.1 Divorcing Reward from the Function Regressed Upon

In GP-UCB, GP-BUCB, and GP-AUCB, the response function f that is actually measured and regressed upon is the same as the reward r; thus the GP-UCB decision rule

$$\boldsymbol{x}_t = \underset{\boldsymbol{x} \in D}{\operatorname{argmax}} [\mu_{t-1}(\boldsymbol{x}) + \alpha_t^{1/2} \sigma_{t-1}(\boldsymbol{x})] \tag{5.2}$$

is sensible, because it trades off the expected reward $\mathbb{E}[r(f(\boldsymbol{x}))] = \mathbb{E}[f(\boldsymbol{x})] = \mu_{t-1}(\boldsymbol{x})$ with a measure of the information to be gained by making the corresponding observation. It is not clear, however, that the reward is always the correct object upon which to regress; in particular, it may be that the reward is a quantity which is difficult to measure (e.g., standing performance) or even not directly observable, while there may be one or more objects which are relatively easy to measure (e.g., evoked potential amplitude, used in Chapter 4, and subjective ratings, proposed in Section 5.4.1.1), and which are strongly related to the reward. In addition, it may be that the reward function does not easily lend itself to being modeled as a GP, perhaps because covariance functions are very hard to specify or the reward is erratic, whereas the available surrogates are more easily modeled. As discussed in Sections 4.1 and 5.4.1.1, it is important that these easily measured and modeled surrogates are chosen such that when they are maximized, reward is also maximized. If more is known about the relationship between the surrogate and the true reward, perhaps an algorithm which understands this relationship can exploit this knowledge to perform better than if it simply considered the reward and surrogate as equal.

One particularly interesting case, and one for which theoretical results are likely obtainable with a modest degree of effort, involves cases in which f is a scalar function, modeled as drawn from a GP, and the mapping from $f(\boldsymbol{x})$ to the utility r is a scalar function g, such that $r(\boldsymbol{x}) = g(f(\boldsymbol{x}))$. If g is a finite, Lipschitz continuous, non-decreasing function of $f(\boldsymbol{x}) \in \mathbb{R}$, with Lipschitz constant k, it follows immediately that the set $\boldsymbol{X}^* \in D$ of maximizers of f also is a subset of the set of

maximizers of $r = g(f)$ over D. Further, it follows from Theorem 1 that if the algorithm is simply run conventionally, ignoring $g(f)$ and treating f as the reward, the regret with respect to $r = g(f)$ is no more than k times the regret of the algorithm with respect to $r = f$, i.e., the regret increases by at most a factor of k. If $g(f)$ is given to the algorithm, it is likely that clever design could produce an algorithm which exploits knowledge of g to do considerably better than this. While g is unknown in the subjective rating setting arising from the human experiments, a number of intriguing forms for g exist. One potentially fruitful choice for theoretical analysis is

$$g(f) = \tanh(k(f - c)),$$

or similarly, the related logistic function,

$$g(f) = 1/(1 + \exp(4k(f - c))),$$

which both saturate to both the left and right, implementing a gentle classification between success and failure at a threshold c. Other reasonable choices include a hinge loss

$$g(f) = \min[0, k(f - c)]$$

or gain

$$g(f) = \max[0, k(f - c)],$$

or piecewise continuous functions of a variety of forms. There are many other interesting possibilities, but exploitation of a properly chosen g might give quite a bit of expressiveness and flexibility to the family of UCB-based algorithms.

5.4.2.2 Making Decisions Using Vector-Valued Functions

Section 5.4.1.2 introduces the idea of modeling many characteristics of the EMG activity of multiple muscles using a contextual GP, as well as the idea of using a reward function of the form of Equation (5.1), a weighted Euclidean norm. Making decisions using this combination of reward function and contextual GP model is examined in some detail in Appendix D.1. In summary, one plausible decision rule for batch or delay selection in this case is

$$\boldsymbol{x}_t = \underset{\boldsymbol{x} \in D}{\operatorname{argmax}} \left[-\sqrt{(\mu_{\mathrm{fb}[t]}(\boldsymbol{x}) - \mathbf{t})^T W (\mu_{\mathrm{fb}[t]}(\boldsymbol{x}) - \mathbf{t})} + \beta_t^{1/2} \sqrt{\operatorname{trace}(W \Sigma_{t-1}(\boldsymbol{x}))} \right]. \tag{5.3}$$

This decision rule is related to trading off the expected reward and a quantity similar to the standard deviation of a scalar function. Another useful characteristic is that the uncertainty term (the second term inside the brackets) is non-increasing as observations are added to the decision set, meaning

it can be calculated lazily. Finally, the scalar case of this decision rule reduces to the GP-BUCB decision rule, Equation 3.7, for $t \to \infty$. A version of this decision rule has been implemented in the suite of code prepared for the pilot experiments.

5.4.2.3 Choosing Paths

One important observation from early feasibility testing in the human model is that it is important to choose stimuli which form smooth sample paths, i.e., to only command gradual changes in the stimulus. This constraint is necessary because the spinal cord's neurological and functional state has a memory and abrupt changes in the stimulus result in a disruption of behavior. This suggests that, particularly in terms of frequency and voltage, it is important to plan paths through the space of candidate stimuli. From an algorithmic perspective, a problem immediately presents itself; this sort of multi-step search is likely exponentially complex in the length B of the path, since such a path search becomes a search over leaves of a tree with depth B and a branching factor of greater than 1 (typically 5 in these pilot experiments). This exponential complexity in constructing batches is the very reason the GP-BUCB and GP-AUCB algorithms use a greedy decision rule to sequentially select individual stimuli and thereby construct a batch, rather than attempting to choose one out of all possible batches. One potential (though not entirely satisfactory) solution to this problem is to require that all valid paths must follow a particular set of construction rules, and further, to restrict the construction rules such that there is at most one valid path of length B or less to any $\boldsymbol{x} \in D$, given the current state, $\boldsymbol{x}_{t-1}$; let this set of legal paths be designated L. In this case, since $|L| \leq |D|$, there are again at most $|D|$ entities among which the algorithm must choose at any given time t. Following this formulation, two suggested decision rules, both with discounted and undiscounted versions, are presented in Appendix D.3. Among these four forms, the decision rule

$$\boldsymbol{X}_t = \underset{\boldsymbol{X} \in L}{\operatorname{argmax}} \left[\sum_{\tau=t}^{t+B-1} [\lambda^{\tau-t}(\mu_{\mathrm{fb}[t]}(\boldsymbol{x}_\tau) + \beta_{\mathrm{fb}[t]}^{1/2} \sigma_{\tau-1}(\boldsymbol{x}_\tau))] \right]$$

has been implemented for the human experiments, where any path $\boldsymbol{X} \in L$ is the ordered sequence $\boldsymbol{X} = \{\boldsymbol{x}_t, \ldots, \boldsymbol{x}_{t+B-1}\}$, $\boldsymbol{X}_t \in L$ is the selected path, and λ is a discounting rate corresponding to the probability of a failure occurring at each step, i.e., the assumed likelihood that the path must be stopped due to that step's stimulus being unacceptable. This decision rule is intended to be used to search the space of voltage and frequency, given a fixed set of active electrodes, where subjective ratings quantify the reward. Alternatively, the decision rule

$$\boldsymbol{X}_t = \underset{\boldsymbol{X} \in L}{\operatorname{argmax}} \left[\sum_{\tau=t}^{t+B-1} [\lambda^{\tau-t}(-\sqrt{(\mu_{\mathrm{fb}[t]}(\boldsymbol{x}) - \mathbf{t})^T W (\mu_{\mathrm{fb}[t]}(\boldsymbol{x}) - \mathbf{t})} + \beta_t^{1/2} \sqrt{\operatorname{trace}(W \Sigma_{t-1}(\boldsymbol{x}))})] \right]$$

has been implemented for path-based planning using the vector-valued EMG feature representation discussed in Sections 5.4.1.2 and 5.4.2.2. A second, more fundamental problem also occurs if decisions must constitute paths; it may be possible to construct a combination of D, the construction rules, and B, such that no paths or sequence of paths can ever connect particular parts of D. Even if this does not happen, it could occur that poorly-responding zones prevent the algorithm from moving between two regions of good performance, even if both high-performing zones have been previously explored. In such situations, there is a non-zero probability that the algorithm cannot find the optimum. This is, unfortunately, a consequence of requiring legal paths, rather than allowing the algorithm to "jump," as do GP-BUCB and GP-AUCB. One partial work-around might be to allow paths to transition through an "off" state, which neighbors many other states, but this makes decision-making very complex and further divorces the algorithm from rigorous theory.

5.4.3 Novel Covariance Functions

The multipolar array configurations needed for the current and future human experiments require new, carefully structured covariance functions. The parameterization of the input space used in Chapter 4 for bipolar electrode configurations is based upon the four spatial coordinates of the single cathode and single anode used. For any two different bipolar pairs $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^4$, the covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ compares the (rescaled) Euclidean proximity of the cathodes and anodes and computes the covariance between the muscle responses to these stimuli on the basis of this proximity. The logical extension of such a covariance function to $n \geq 2$ active electrodes in a given stimulus would be to again describe the spatial locations of the active electrodes in $\mathbb{R}^{2n}$; however, this approach has many deficiencies. With several electrodes of the same polarity, exchangeability issues arise, i.e., the set of active electrodes (A1+, A2+, C9-) is actually the same as (A2+, A1+, C9-), though these two descriptions might naturally correspond to different points in $\mathbb{R}^6$. Thus, the covariance function would have to recognize the equivalence of different descriptions of the same configuration, a problem not present in the bipolar case. Worse, if the algorithm is allowed to use multi-electrode configurations with different numbers of active electrodes, it must be able to compare these objects. Indeed, a configuration with two cathodes and one anode might be functionally very similar to the bipolar configuration in which one of these two cathodes is off (i.e., is neither a cathode nor an anode). In such a case, the covariance function certainly should express the commonality between their respective responses, even though these two objects would exist in $\mathbb{R}^6$ and $\mathbb{R}^4$ respectively. Clearly, a covariance function which operates directly on vector quantities of arbitrary size is not appropriate. The covariance function should therefore act on some representation of the electrode configuration which is independent of the number of cathodes or anodes.

One natural way to create the required unified representation would be to calculate the electric field or voltage distribution created by each stimulus configuration, and then compare key charac-

teristics of the fields generated any pair of stimuli. This is reasonable because these fields can be calculated from first principles (given good measurements of the electrical properties of the tissue) and the effects of epidural electrostimulation arise via these fields. Particularly if computational speed is an issue, very simple (i.e., closed-form algebraic) calculations may be appropriate. Fortunately, electricity is well-understood and follows simple physical laws. As an example, given a spherical electrode of radius ρ with a fixed surface voltage V, the voltage $V(r)$ at distance $r \geq \rho$ from the center of the electrode in a homogeneous medium is

$$V(r) = V\frac{\rho}{r}, \tag{5.4}$$

assuming that the infinite distance boundary condition is ground, i.e., $V(r) \to 0$ as $r \to \infty$. If several electrodes of different positions and voltages are desired (i.e., the chosen stimulus $\boldsymbol{x}$ includes more than one active electrode), an approximate solution can be obtained by summing the corresponding voltage functions, yielding a voltage value corresponding to the influence of all of the electrodes together. Given a set of target locations believed to be representative of the voltage distribution's influence on the spinal cord, the vector of voltage values $\mathbf{V}(\boldsymbol{x})$ corresponding to a stimulus $\boldsymbol{x}$ can be calculated. Since these locations are fixed, $\mathbf{V}(\boldsymbol{x})$ has the same size and meaning, regardless of how many electrodes are active in stimulus $\boldsymbol{x}$. To calculate a covariance function for $\boldsymbol{x}$ and $\boldsymbol{x}'$, it then remains to compare the vectors $\mathbf{V}(\boldsymbol{x})$ and $\mathbf{V}(\boldsymbol{x}')$ in terms of their similarity with respect to the responses of interest. One plausible way to do this is to weight some elements of $\mathbf{V}$ more heavily than others, using anatomical intuition. Using information from Sharrard (1955), as well as Sharrard (1964) and Harkema et al. (2011), it is possible to localize any of several motor pools of interest within the human spinal cord with respect to the electrode. These localizations can also be verified with respect to the array by use of the data from the supine stimulation experiments described in Section 5.2.

If the model must consider EMG from several muscles, the set of which is designated M, it is necessary to construct a covariance function which enables predicting these responses, treating the individual muscle or feature as a context. One way to do this is to create a diagonal weight matrix W_m for each muscle $m \in M$. Then, if the ith entry in $\mathbf{V}$ is at a location where electrical stimulus might plausibly exert substantial effects on the spinal inputs and circuits associated with muscle m, e.g., near the corresponding motor pool, $[W_m]_{i,i}$ should be large; conversely, if the ith entry in $\mathbf{V}$ can be expected to not be particularly influential on muscle m, the weight $[W_m]_{i,i}$ should be small. One way to do this simply is to choose the weights as the height values of a Gaussian bump centered on the location of the motor pool associated with muscle m. The re-weighted vectors $W_m\mathbf{V}(\boldsymbol{x})$ and

$W_{m'}\mathbf{V}(\boldsymbol{x}')$ can then be fed into a covariance function

$$k(W_m\mathbf{V}(\boldsymbol{x}), W_{m'}\mathbf{V}(\boldsymbol{x}'))$$

which computes the covariance of the responses to $\boldsymbol{x}$ in muscle m and the responses to $\boldsymbol{x}'$ in muscle m'. Appropriate covariance functions might be linear,

$$k_{lin}(W_m\mathbf{V}(\boldsymbol{x}), W_{m'}\mathbf{V}(\boldsymbol{x}')) = (W_m\mathbf{V}(\boldsymbol{x})) \cdot (W_{m'}\mathbf{V}(\boldsymbol{x}')), \tag{5.5}$$

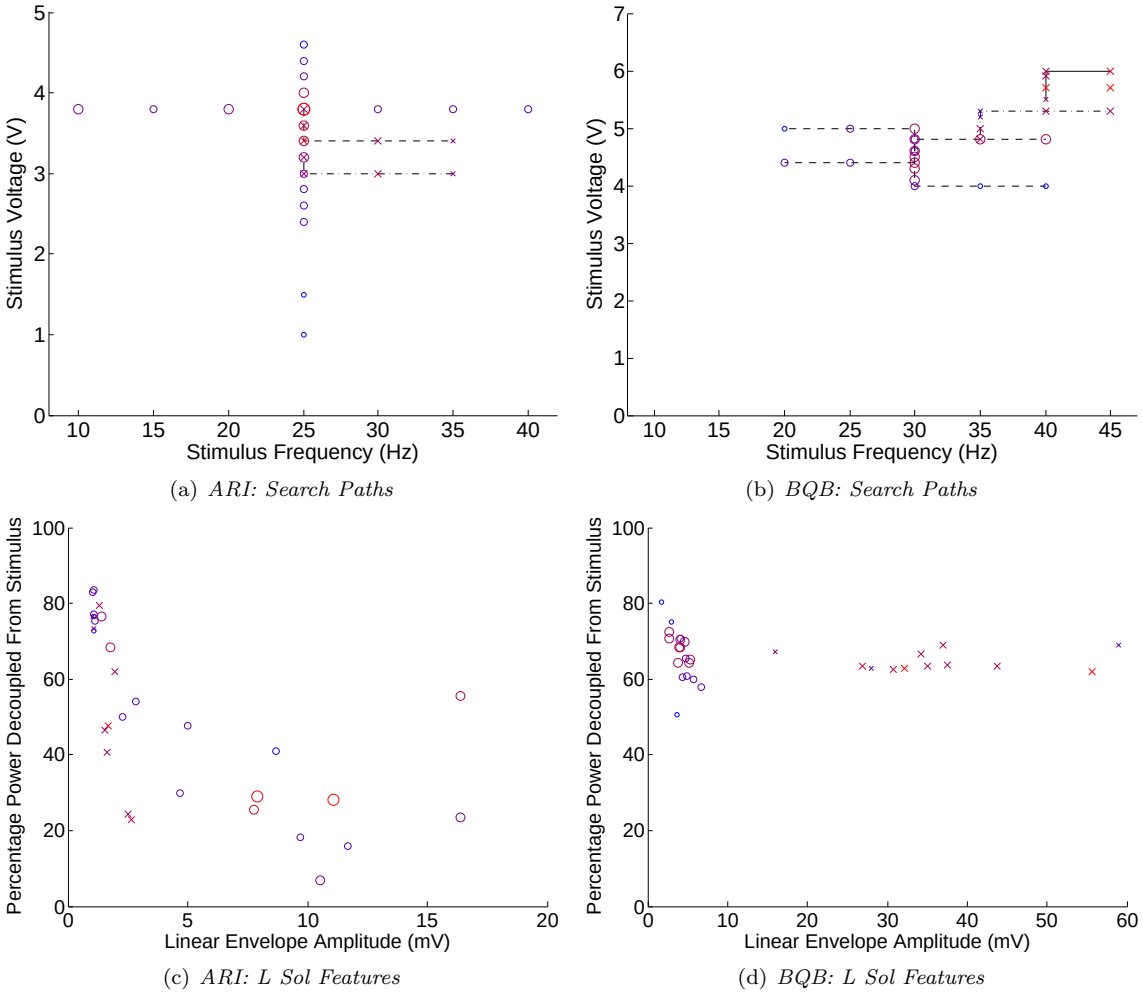
or perhaps could take the form

$$k_{lin,0} = \max[0, (W_m\mathbf{V}(\boldsymbol{x})) \cdot (W_{m'}\mathbf{V}(\boldsymbol{x}'))],$$

where this second form would avoid having large, negative covariances between polarity-flipped configurations. Alternatively, rather than focusing on individual muscles, as in the vector-valued EMG case, it might be more appropriate to consider a holistic view of activity in general, as in the subjective rating case; it would then be reasonable to attempt to correspondingly use knowledge of motor pool locations to weight elements of $\mathbf{V}$, via the choice of a combined weight matrix $W_{combined}$, constructed as a linear combination of individual matrices W_m. The covariance function has been employed using both the subjective rating and vector-valued EMG modes, and results are described in the next section.

5.5 Preliminary Results and Discussion

Some fairly simple preliminary experiments have been conducted with two different human patients. In the first patient, designated ARI (who has no voluntary motor control or sensation below the level of his injury), data from two sessions were used. In the first, the human experts performed a standard procedure, a voltage and frequency sweep, using a configuration of active electrodes known to produce good standing for ARI (3+ 4+ 8+ 14+ 15+ // 9- 10-, in which cathodes form a horseshoe shape around two anodes at the caudal end of the array). This experimental session consisted of a gradual (0.2 V increment after reaching threshold) increase in voltage from 0 V to 4.6 V, with the stimulus frequency fixed at 25 Hz. After this voltage sweep, the voltage was fixed at 3.8 V, and the frequency of stimulation was increased in 5 Hz steps from 10 Hz to 40 Hz. During this process, the participant was instructed to provide a rating from zero to ten of how well he was standing, one such rating at each set of stimulus parameters. Subjective ratings from these two sweeps are presented in Figure 5.1(a). On a subsequent experimental day, the algorithm was given the data from the previous session and given a quadratic mean function chosen such that the maximum subjective



(a) *ARI: Search Paths*

(b) *BQB: Search Paths*

(c) *ARI: L Sol Features*

(d) *BQB: L Sol Features*

Figure 5.1: Search paths, ratings, and EMG features from experiments with participants ARI and BQB. (a): Voltage and frequency sweeps executed by human experts (circle) and the algorithm using subjective ratings (dashed, x) with participant ARI. Subjective ratings given by the participant are shown as color and size of the marker, where red and large markers are high subjective ratings and blue and small markers are low subjective ratings. Participant ARI was instructed to give ratings from 0 (least favored) to 10 (most favored) and gave ratings covering this whole range. (b): Voltage and frequency sweeps conducted by the algorithm with participant BQB, where the first session was executed using subjective ratings (circle, dashed) and the second was executed using EMG-based grading (x, with individual paths respectively shown as dash-dot and solid). The starting state for the first session lies at the rough center of the four separate, diverging paths executed on that day; these paths give good coverage of the region around that state. Participant BQB was instructed to give ratings from 0 to 10, with 5 as a "baseline," and gave no rating lower than 3 and no rating higher than 6. (c): Features calculated from participant ARI's EMG activity in a representative muscle, the left soleus, in response to each stimulus. The same scheme for labeling trials and subjective ratings is maintained from (a). The crescent shape shown in this feature space seems to correspond with a continuum from quiescence (upper left) to activation driven by the stimulus (lower right). (d): Features calculated from BQB's EMG activity under the stimuli shown in (b), again maintaining the labeling of trials and subjective ratings used in that figure. Qualitiatively different behavior appears to have occurred in the left soleus during these two sessions; in the first session (circles), the muscle was quiescent, while in the second (x) it was active, quite strongly in some cases.

rating was at 3.7 V, 25 Hz, near the previous day's maximum. The algorithm then continued this search over voltage and frequency, making two individual five-step paths. The first path started from 3.8 V and 25 Hz. The second started from 3.4 V and 25 Hz, an intermediate point on the first path, a value selected by the human experts overseeing the experiment. These experimental paths were successfully completed, yielding subjective standing quality ratings which were generally similar to those given in the previous session.

Two sessions were also conducted with the second patient, designated BQB (who has some sensation, but no motor control below the level of his lesion). Again, a known effective combination of active stimulus electrodes (2+ 3+ 4+ 9+ 13+ 14+ 15+ // 7- 8- 10-, consisting of two outer rows of cathodes and one midline cathode, combined with midline anodes, and all positioned at the caudal end of the array) was selected, with voltage and frequency values (4.5 V, 30 Hz) known to produce effective standing chosen as the starting point for the first session. The algorithm then was asked to choose paths. The first path started from this known point. After the completion of the first path, the stimulus was set to 4.4 V and 30Hz, which was used as the starting state for the remaining three paths of the first session. All four paths in this first session were selected using the subjective rating method. During the second session, the algorithm selected actions based on hand-selected EMG targets and weights. These targets and weights were chosen such that the highest-rated stimuli also tended to have the highest EMG-based grades when the reward was calculated using Equation (5.1), substituting the observations actually made for the vector $\mathbf{f}$. The algorithm was then given the EMG observations from an earlier session involving quiet standing using the same combination of active electrodes and one combination of voltage and frequency, as well as the EMG data from the first session, which had been guided by subjective ratings. Since it was making decisions using only EMG, the algorithm was now unable to directly use the subjective ratings given during the rating-based session. Within this second, EMG-based session, the algorithm selected two paths, the first beginning from the best stimuli found during the fourth path of the previous session. The second path executed during this session used the EMG observed during the first path to make decisions about what stimuli to apply during the second. This session thus provided an opportunity to test the efficiency of the data handling and processing necessary to select a subsequent path based on EMG observations acquired earlier in the same session. Using the current manual, two computer process, it may be possible to execute three paths during one hour-long session, but more paths will likely require an integrated data acquisition and processing system.

While no rigorous assertions regarding the search strategy's effectiveness can be made on the basis of the current experiments, several conclusions may be drawn from this experience. First, the algorithm and software implementation are capable of collaboratively planning experiments with human experts. In this procedure, given parameters set by the human experts, the algorithm proposed experiments, which were then approved and executed. This collaborative system allows the

combination of the rigorous, quantitative search and optimization capabilities of GP-BUCB and its derivatives with the human experts' ability to assess a variety of important diagnostic information unavailable to the algorithm via the subjective ratings or EMG. Second, wide variation is present in the subjective grades given by different patients, perhaps in part because the instructions given may not have been consistent, but also perhaps due to differences in the perceived effects of EES therapy for individual patients. This difference in perceptions may be related to the differences in individual injuries as well, particularly the degree of sensation preserved. From an algorithmic or procedural standpoint, this implies that the rating scale must be both carefully specified and also customized to each patient. Careful specification is important because ratings must be repeatable and consistent to be useful. Patient-specific customization is also important because patient perception is highly individual, yet each patient's ratings must have sufficient resolution to identify improvement if and when it occurs, even if this improvement is small. Third, the Euclidean distance metric used for reward under the EMG-based selection procedure may require modification or replacement. It is clear that in some features (e.g., linear envelope amplitude, which roughly translates to contractile activity), there are qualitative differences between regions of the feature space (e.g., a muscle being "on" vs. "off") which are not effectively captured by simple Euclidean distance from a target; Equation (5.1) instead implies linear and symmetrical degradation of reward as distance from the target increases. From the perspective of standing performance, it may be that it is very important that a particular muscle is indeed "on", but not terribly important how strongly it is contracting during the period recorded, making the distance metric a poor fit. Another type of reward function may be more suitable to these features than Equation (5.1), e.g., the soft classification reward functions suggested in Section 5.4.2.1. Third, as a simple matter of logistics, the experimental cycle must eventually be made faster than it currently is, such that more than two or three paths of EMG-based stimuli selections can be made; without greater throughput, it will be very hard for the algorithm to effectively search while independent of expert assistance.

Another important point concerns the EMG features selected. Ten muscles were used in the EMG-based selection of actions for patient BQB: each of the soleus, tibialis anterior, medial gastrocnemius, medial hamstring, and vastus lateralis, in each of the left and right lower limbs. For each muscle, two features were computed: the average amplitude of the linear enveloped signal (using a second-order Butterworth filter, 4 Hz cutoff, applied forward and backward using the MATLAB filtfilt command) and the percentage power decoupled from the stimulus frequency and its multiples. This second feature was computed by performing a Fast Fourier Transform of the EMG signal and then computing the ratio of the power in frequency ranges other than $\bigcup_{n=1}^{\infty}[(n-0.1)f,(n+0.1)f]$, where f is the stimulus frequency, to the power of the original signal; since the frequency content of a repeating signal with period f should lie in the suppressed bands, this ratio should give the proportion of the signal which is not driven by the stimulus. The suppressed bands are 20% of the

frequency spectrum, so this ratio's value is expected to be approximately 80% when the signal is white noise (which has a flat frequency spectrum and should be present when the muscle is quiescent) and substantially lower when the EMG is driven in synchrony with the stimulus. It was hypothesized that these features would give the ability to detect the level of activity of the muscle, as well as the degree to which this activity was being controlled by the spinal cord, rather than being driven by the stimulus, and that high performance would be observable as moderate-to-high linear envelope amplitudes and large amounts of stimulus decoupled power. Unfortunately, the experiments conducted so far have not borne out this hypothesis. In patient ARI, most muscles, as typified by Figure 5.1(c), did not show a co-occurrence of both high activity and high proportion of decoupled power, even under those stimuli which received very high subjective ratings. On the other hand, patient BQB did achieve EMG which had large values for both features, but this may be an artifact associated with a persistent and large amplitude tremor which occurs in both lower limbs during most of his standing bouts, rather than emergent, spinal-controlled activity. Neither of these observations implies that EMG with the hypothesized feature combination is infeasible for either patient, nor do these observations indicate that good standing would not appear in this region of the feature space; however, it does appear that (subjectively) good standing can occur outside of this region and that poor standing may in some cases produce EMG signals which do lie in this region. Both of these facts suggest that different (or at least additional) features may be necessary for low-error EMG-based recognition of good standing.

5.6 Extensions

The methods discussed in Section 5.4 have been used for the pilot experiments, but several substantial opportunities for improvement, alternatives to current approaches, or extensions to the theory suggest themselves.

5.6.1 Time Series Information and Coordination of Muscles

While it is reasonable to suggest using contextual GP models for modeling the activity of different muscles under the same stimulus, representing the aggregate activity of each muscle as a scalar or set of scalars may miss physiologically and behaviorally important information. In particular, properties such as coordination of muscles in response to disturbances require examining multiple EMG channels and how they interact with one another over time; muscle groups should activate in the proper anatomical pattern for coordination, e.g., flexors in the same leg activating together or extensors in the same leg activating together. These patterns are not describable by the time-averaged activity of each muscle; the crucial diagnostic information is instead carried in the relative times at which the muscles are active. This is not clearly something to which GPs are applicable; a

(contextual) GP predicts a (set of) scalar(s), rather than a time trajectory. A set of scalars could be used to parameterize these interactions over time, but it is not clear how this representation should be constructed, or how expert knowledge should be used to construct covariance functions which respect the structure of these functions over the stimulus space. Further, it is not clear how a representation of muscle activity in terms of these features should be used to determine the quality or performance of the high-level motor activity. These issues remain intriguing, but open.

5.6.2 Dynamical Systems Approaches: Cost Functions and LQG

The cost functions proposed so far may fail to strike the right balance of abstraction versus concreteness; the subjective rating system is high-level and could yield meaningful results, but is somewhat poorly defined conceptually, while the vector-valued EMG approach relies on the specification of low-level targets which may not be easy to choose so as to actually produce good standing. Another alternative, which offers both greater rigor and an intermediate level of abstraction, is to treat the human participant as a dynamical system, for which some states (motion capture, center of pressure, etc.) are observed directly and for which some CNS control outputs (EMG) may also be observed. In this case, a physical model of the human could be used to infer a simple parameterization of the composite controller (the spinal cord under the influence of the EES system); this becomes a system identification problem, classically treated in the literature of controls and dynamical systems (for an introduction to system identification, see the text by Ljung, 1999). If the EES system is assumed to modulate the parameters of the controller, it might be possible to model these parameters as functions of the stimulus. The controller parameters would then be the objects of interest for learning and regression, perhaps modeled as draws from a Gaussian process. Given a regression model for the controller parameters, a closed-loop therapy system requires the ability to predict the reward corresponding to any stimulus; depending on the controller model used, well-known methods like linear quadratic Gaussian control (LQG, see Athans, 1971, for an early and thorough review) could be used to assign a cost to any given set of controller parameters. In the LQG case, for any draw from the posterior over controller parameters, there is an analytical expectation for the expected controller cost with respect to disturbances; this cost is the time integral of the variance of a Gaussian (accounting for deviations of the state from the desired trajectory and control costs) summed with other Gaussian variances (accounting for terminal costs). With relatively few samples, it might be possible to build a useful notion of the distribution over the composite controller parameters of the cost of running such a system, such that stimulus decisions could be made by comparing these cost distributions. This quantification would require a stochastic procedure, such that the algorithm would no longer be as computationally efficient. However, such a perspective might be more reflective of useful performance than direct functions of the EMG or other sensors and more tractable than attempting to work with user ratings directly. Given access to clinical sensors and off-board

computing, such an approach might be reasonable.

5.6.3 Alternative Covariance Functions

A reasonable criticism of the covariance functions proposed in Section 5.4.3 is that, while they have the virtue of being very efficient computationally, this efficiency has been obtained at the price of both accuracy and precision in terms of the stimulating voltage distribution in and around the spinal cord. Certainly, the voltage function used is quite crude; the electrodes are not spheres, nor point sources, and the medium is not a single, homogeneous material. Additionally, capacitative and electrochemical effects may be significant, and the chosen weights and target regions of the spinal cord may not be appropriate. It is important to remember that the algorithmic purpose of the covariance function is to obtain a useful notion of how the responses to different electrode configurations co-vary, and how large the uncertainty about these responses may plausibly be, but this is only a means to an end. The covariance function does not need to perfectly capture the response function's shape; rather, it only needs to provide enough of a guide to enable efficient and intelligent experimental choices, the resulting data from which will drive the GP model.

Nevertheless, it remains reasonable to refine the proposed covariance function. Sophisticated finite element models of the spinal cord, epidural electrode array, and the surrounding volume have been created, which might be suitable for these purposes (see, e.g., Minassian et al., 2007). These simulations are typically quite computationally expensive. For this reason, creating a large, precomputed set of such simulations (corresponding to large sets of observations $\boldsymbol{X}_{t-1}$ or decision sets D) is likely infeasible. It might be possible to execute some subset of these simulations ahead of time, however, particularly if the decision set is small, perhaps growing slowly with time. Additional simulations could be performed on an online, as-needed basis, in a scheme somewhat similar to the lazy variance calculations discussed in Section 3.5. A hybrid method might also be reasonable; a sufficiently large "library" of full finite element calculations could be used to calculate corrections to the predictions of fast but crude models (the simple, homogeneous medium models discussed above, or perhaps linear combinations of very simple finite element simulations), giving the advantages speed, accuracy, and large decision or observation sets. It may also be reasonable to use prior knowledge of the relevant structures in the spinal cord and the mechanisms of neuronal excitation (see Minassian et al., 2007) or neuronal modeling software like NEURON (Hines and Carnevale, 2001) to choose appropriate functional weightings on the resulting simulation outputs to describe actual activation (and hopefully the degree of assistance toward the desired motor behavior).

One interesting note follows from the possible studies with NEURON; since the finite element and NEURON simulations are computationally expensive, a limited budget of these experiments is available, making this problem itself an appropriate application for the GP-UCB, GP-BUCB, and GP-AUCB algorithms. The algorithms would choose which simulations should be run, attempting

to find stimuli which yield favorable patterns of neural activation, as determined by finite element simulations and NEURON. The results of these experiments could be used (perhaps with smoothing or abstraction) to build priors on the responses in actual human patients, thus avoiding the need to relearn the anatomy of the human spinal cord for every individual patient, consequently freeing the algorithm's clinical action selections to learn patient-specific variations.

5.6.4 Expansions of the Decision Set

Somewhat related to the question of the decision set "moving" with respect to time, discussed in Section 4.4.1, it is reasonable to consider decision sets which expand with respect to time. Certainly, the proof of the cases of Theorem 1 which allow the decision set D to be continuous rest on notions of how to grow the decision set appropriately as the algorithm acquires more data, discussed by Srinivas et al. (2009). However, these methods of growing the finite decision set D_t over which the algorithm must make a choice in each round constitute increasingly fine discretizations of a continuous, true D. An interesting question is how, and when, to expand the decision set into new and unexplored territory. One reason to do this is clear; if a set of stimuli are known to be safe and productive, it would be reasonable to allow the algorithm to explore in a limited "sandbox" including this set and its immediate surroundings within the stimulus space. Subsequently, the sandbox should be allowed to grow in some fashion which respects notions of sensitivity and safety. In the case of experimental stimuli, some reasonable notions of neighborship may be created, e.g., moving an active cathode or anode by one interval on the array, inactivating a cathode or anode, activating a cathode or anode next to an already existing cathode or anode, etc. Again, it is not clear precisely how these notions of neighborship correspond to functional similarity or sensitivity, nor is it clear how to choose when to expand the decision set in a rigorous, algorithmic fashion. Further, while this may be necessary for practical reasons, this sort of expansion of the decision set in a fashion which is contiguous with respect to the domain may preclude finding the optimum in some cases. It may also (or instead) be appropriate to track which regions of the decision set are known to be safe and which are known to be dangerous; dangerous regions should be avoided, even if they would otherwise be worth exploring. This could perhaps be captured in parallel by employing a Gaussian process classifier (see Rasmussen and Williams, 2006, Chapter 3), using a similarly structured covariance function and operating over the same domain. An open question concerns how to formulate the decision rule in this case; certainly, it is important and useful to learn more about this classification of safe versus dangerous, but should this be incorporated into the decision rule, or should this information be acquired as made available by a more conventional decision rule? Additionally, if it is desirable to incorporate this model of safety and confidence therein into the decision rule, how should this be done? These questions will have to be answered if it is desirable to incorporate explicit models of safety into GP-BUCB-like algorithms.