Semi-Myopic Measurement Selection for Optimization under Uncertainty

David Tolpin and Solomon Eyal Shimony

November 29, 2009

Abstract

The following sequential decision problem is considered: given a set of items of unknown utility, an item of as high a utility as possible must be selected ("the selection problem"). Measurements (possibly noisy) of item features prior to selection are allowed, at known costs. The goal is to optimize the overall sequential decision process of measurements and selection.

Value of information (VOI) is a well-known scheme for selecting measurements, but the intractability of the problem typically leads to using myopic VOI estimates. In the selection problem, myopic VOI frequently badly underestimates the value of information, leading to inferior measurement policies. In this paper, the strict myopic assumption is relaxed into a scheme termed semi-myopic, providing a spectrum of methods that can improve the performance of measurement policies. In particular, the efficiently computable method of "blinkered" VOI is proposed, and theoretical bounds for important special cases are examined. Empirical evaluation of "blinkered" VOI in the selection problem with normally distributed item values shows that is performs much better than pure myopic VOI.

Chapter 1

Introduction

Optimization under uncertainty is a domain with numerous important applications [21], [17]. Problems of optimization under uncertainty are intractable in general, and thus special cases are of interest. In this paper, the following selection problem is examined: given a set of items of unknown utility (but a distribution of which is known), an item with as high a utility as possible must be selected. Measurements (possibly noisy) of item features are allowed prior to selection, at known costs. The objective is to optimize the overall decision process of measurement and selection. Even with the severe restrictions imposed by the above setting, this decision problem is intractable [20], and yet it is important to be able to find at least an approximate solution due its importance for potential applications such as sensor network planning and oil exploration.

Other settings where this problem is applicable are: considering which time-consuming deliberation steps to perform (metareasoning [22]) before selecting an action, locating a point of high temperature using a limited number of measurements (with dependencies between locations as in [14]), and the problem of finding a good set of parameters for setting up an industrial imaging system. The latter problem is actually the original motivation for this research, and is discussed in Section 6.5.

A widely adopted scheme for selecting measurements, also called sensing actions in some contexts (or deliberation steps, in the context of metareasoning), is based on value of information (VOI) [22]. Computing value of information is intractable in general, thus both researchers and practitioners often use various forms of myopic VOI estimates [22] coupled with greedy search (see Section 4). Even when not based on solid theoretical guarantees, such estimates lead to practically efficient solutions in many cases.

However, in a selection problem involving items with real-valued utilities (the main focus of this paper), coupled with the capability of the system to perform more than one measurement for each item, the myopic VOI estimate can be shown to badly underestimate the value of information (Section 4.4). This can lead to inferior measurement policies, due to the fact that in many cases either no measurement is seen to have a VOI estimate greater than its cost, or measurements with high true VOI are ignored, due to the myopic approximation. The goal is to find a scheme that, while still efficiently computable, can overcome these limitations of myopic VOI. The framework of semi-myopic VOI is proposed (Section 5), which includes the myopic VOI as the simplest special case, but also much more general schemes, including exhaustive subset selection at the other extreme. Within this framework the "blinkered" VOI estimate is introduced, as one that is efficiently computable and yet performs much better than myopic VOI for the selection problem. In indicative special cases we get theoretical bounds that show the benefit of the blinkered estimate (Section 5.2). Empirical results on artificial and real-world data further support using the blinkered scheme (Section 6).

Chapter 2

Background

2.1 Optimization under Uncertainty

Problems of optimization under uncertainty are characterized by the necessity of making a decision without being able to predict its exact effect. Such problems appear in many different areas and are both conceptually and computationally challenging. Optimization under uncertainty shares many basic notions with other branches of optimization, but still differs from them in several formulation, modeling, and solution issues.

The notion of optimization examined in this paper is that of making a single, "best" choice from a *feasible set*, according to a given *objective function*. Optimization problems can be either *discrete* or *continuous*.

The **Uncertainty** considered here is the lack of exact knowledge about the values the objective function takes for all or some members of the feasible set. The "best" choice thus depends on the *beliefs* about the the objective function values, which are represented by a multi-variate probability distribution. The agent can obtain information about the objective function through **observations**, which can can be either *exact* or *inexact*. An inexact observation deviates from the true value, and the observation distribution, conditional on the true value, may also be unknown.

Optimization under uncertainty is used to solve problems in diverse domains. A few examples are design of experiments, industrial design, medical diagnostics, and environmental control. Problems in these domains can be formulated as optimization of a function under constraints, when some of the data is unavailable or inexact.

Over the years, various uncertainty modeling philosophies have been developed [24]. The range of approaches to optimization under uncertainty includes, among others, stochastic and fuzzy programming. The Bayesian framework [4], which is assumed in this paper, is an extension of stochastic programming in which the uncertainty model incorporates prior beliefs and is then updated based on observed data using a Bayesian step.

Consider an optimization problem with uncertain coefficients:

$$\begin{aligned} \text{min:} \quad & f(x, \omega) \\ \text{s.t.:} \quad & x \in X \end{aligned} \tag{2.1}$$

where a value x from feasible set X must be found, and the parameter ω is uncertain. A typical stochastic optimization approach is to assume a distribution, conditional on x, for the uncertain parameter ω and optimize the expected value of the function:

$$\begin{aligned} \text{min:} \quad & \mathbb{E}_\omega\left[f(x,\omega)\right] = \int\limits_\Omega f(x,\omega)p(\omega|x)d\omega \\ \text{s.t.:} \quad & x \in X \end{aligned} \tag{2.2}$$

where $p(\omega|x)$ is the conditional probability density function of the uncertain parameter ω. In the Bayesian framework, the distribution for ω is not known in advance, but is gradually refined by combining prior beliefs with observed data into a posterior distribution for ω. Realizations of the Bayesian framework differ in the way in which data about ω are obtained.

2.2 Selection problems as POMDPs

In the selection problems considered in this paper, observations are achieved by making measurements, each at a known cost. The cost in some cases depends on the previous measurement, as defined formally in Section 3. The entire process, of selecting which measurments to perform (the selection possibly depending on observations obtained due to previous measurements) and then finally settling on a choice of item is a sequential stochastic decision process. The selection problem can be stated in terms of the well-known partially observable Markov decision process (POMDP) model.

A POMDP [13] is defined over a state space S, a set A of possible actions, and a set O of possible observations. In addition, a state transition model τ maps from state-action pairs to a distribution over states (that is, a probability of landing in a state s' given that an action a is executed in state s). An observation distribution σ from pre-action states, actions, and post action states to O is the stochastic observation model. A reward (or alternately, cost) function R from pre-action states, actions, and post-action states to real numbers is used to model the utility function. Finally, the initial state may be unknown, in which case an initial belief b_0 is given, as a "prior" distribution over S.

A policy in a POMDP is a mapping from belief states (probability distributions over S) to the set of actions, fully specifying what actions are to be done given each possible belief state. An initial belief state and a policy (together with the transition and observation distributions) define a Markov process, which can be used to stochastically generate sequences of (belief, action, observation) tuples, (and likewise a sequence of (state, action, reward) tuples). One is typically interested in finding a policy such that the expected sum of rewards over such sequences is maximized. There are various formulations of POMDP, the one that fits the selection problem is called an indefinite horizon POMDP [8], or a terminal-state POMDP: actions are executed until some action moves the system into a terminal state.

Although there are several different variants of the selection problem (more precisely defined in Section 3), they can all be stated as a special case of POMDP as follows. The state space S consists of all possible values of the items, the identity of the last item measured, and a special additional terminal state $\perp$. The action space A consists of all possible measurement actions and all possible selection actions. The transition distribution is actually a rather degenerate deterministic function, as the values of the items do not change, and the identity of the last measured item is defined by the last action. The initial belief is defined by the prior distribution over the item values. The observation distribution depends only on the (measurement) action and the value of the measured item. Finally, the reward function is defined as minus the cost of the measurement (for a measurement action), or the utility of the selected item (for a selection action). A selection action moves the system into the terminal state, $\perp$.

Defining selection problems as POMDPs as done above sheds light on the type of optimization problem we are trying to solve - the fact that we are (ideally) trying to find an optimal policy. Unfortunately, despite the fact that POMDP is a well-studied model with numerous (usually approximate) solution algorithms available [5][25], POMDP is a highly intractable problem [5], with complexity typically exponential in the number of states. Since in this case the number of states is itself exponential in the number of items, traditional POMDP solution methods are highly impractical (especially as in this paper the focus is on items with utility values taken from the reals). On the other hand, POMDPs defined by selection problems are a special case, due to the degenerate transition and observation distributions, so it may be possible to apply specialized techniques to such problems. Indeed, approximate optimization techniques developed for meta-reasoning are applicable to the selection problem.

2.3 Limited Rationality

Finding an approximate solution to the selection problem has been attempted in past research under the guise of *Rational metareasoning* [22]. In this setting, an agent can perform base-level actions from a known set $\{A_i\}$. Before committing to each action, the agent deliberates by performing a sequence of meta-level "deliberation" actions from a set $\{S_j\}$. At any given time there is a base-level action, A_α, that appears (in a decision-theoretic sense) best to the agent. The goal of subsequent meta-level actions is to improve the choice of A_α.

The current best action A_α is one which maximizes the agent's expected utility:

$$\mathbb{E}[U(A_i)] = \sum_k P(W_k)U(A_i, W_k) \tag{2.3}$$

$$\alpha = \arg\max_i \mathbb{E}[U(A_i)] \tag{2.4}$$

where $\{W_k\}$ is the set of possible world states, and $P(W_k)$ is the probability that the agent is currently in state W_k.

A meta-level action affects the choice of the base-level action A_α. The *value* of a meta-level action is measured by the resulting increase in the utility of A_α. Since neither the outcomes of meta-level actions nor the true utility of A_α are known in advance, a meta-level action is selected according to its expected influence on the expected utility of A_α. The *value of information* of a meta-level action S_j is the expected difference between the expected utility of S_j and the expected utility of the current A_α.

$$V(S_j) = \mathbb{E}(\mathbb{E}(U(S_j)) - \mathbb{E}(U(A_\alpha))) \tag{2.5}$$

Under certain assumptions, it is possible to capture the dependence of utility on time in a separate notion of *time cost* C. Then, the utility of an action A_i taken after a meta-level action S_j is the utility of A_i taken now less the cost of time for performing S_j:

$$U(A_i, S_j) = U(A_i) - C(A_i, S_j) \tag{2.6}$$

It is customary to call the current utility of a future base-level action its *intrinsic utility*. The separation into intrinsic utility and time cost allows to estimate the utility of a base-level action in a time-independent manner, and then refine the true utility estimate under varying the time pressure represented by C.

When the time cost C depends only on the meta-level action S_j, [2.5] can be rewritten with the cost and the intrinsic value of information of S_j as separate terms.

$$\begin{aligned} V(S_j) &= \mathbb{E}\left(\mathbb{E}(U(A_\alpha^j, S_J)) - \mathbb{E}(U(A_\alpha))\right) \\ &= \mathbb{E}\left(\mathbb{E}(U(A_\alpha^j)) - \mathbb{E}(U(A_\alpha))\right) - C(S_j) \\ &= \Lambda(S_j) - C(S_j) \end{aligned} \tag{2.7}$$

where *intrinsic value of information*

$$\Lambda(S_j) = \mathbb{E}\left(\mathbb{E}(U(A_\alpha^j)) - \mathbb{E}(U(A_\alpha))\right) \tag{2.8}$$

is the expected difference between the intrinsic expected utilities of the new and the old selected base-level action, computed after the meta-level action was taken.

Perfect meta-level rationality cannot be achieved; the simplest approximation of the utility of a meta-level action is based on the following assumptions [22]:

Meta-greedy assumption: Only sequences consisting of a single meta-level action are considered; the meta-level policy commits to the best single action and therefore is greedy.

Single-step assumption: The utility of a meta-level action is determined only by its immediate effect on the choice of a base-level action.

The value of information estimate and the greedy algorithm which follow from these assumptions are known as *myopic* [22]. The myopic value of information estimate MVI need not be restricted to single atomic actions; multiple atomic actions performed in a batch, without intermediate decisions, can be considered a single meta-level action. For a batch of n meta-level actions $S_{j_1} \dots S_{j_n}$, the myopic estimate is:

$$\begin{aligned} MVI_{j_{1\dots n}} &= \mathbb{E}\left(\mathbb{E}(U(A_\alpha^{j_{1\dots n}})) - \mathbb{E}(U(A_\alpha))\right) \\ &\quad - \sum_{k=1}^{n} C(S_{j_k}) \end{aligned} \tag{2.9}$$

where expectation $\mathbb{E}$ is computed according to the immediate posterior belief distributions. When all of the actions in the batch are identical, notation MVI_j^n will be used, denoting the myopic estimate of a batch of n actions S_j.

Theoretical bounds of the myopic estimate are proved for some restricted cases. In many applications, experiments show that the assumptions work well. When the assumptions turn out to be unjustified, better approximations can often be designed by extending or replacing either the meta-greediness or the single-step assumption.

Chapter 3

Problem Statement

The selection problem addressed in this paper is formally defined as follows.
Given:

- A set of items $S = \{s_1, s_2, \ldots\}$.
- A set of N_f item features $Z = \{z_1, z_2, \ldots, z_{N_f}\}$. (Each feature z_i has a domain $\mathcal{D}(z_i)$.)
- A joint distribution B_0 over the features of the items in S. That is, a joint distribution over the random variables $\{z_1(s_1), z_2(s_1), \ldots, z_1(s_2), z_2(s_2), \ldots\}$.
- A set of measurement types $M = \{(c^m, p^m)_k \,|\, k \in 1..N_m\}$, with potentially different intrinsic measurement cost c^m and observation distribution p^m, conditional on the true feature values, for each measurement type.
- A known utility function $u(\mathbf{z})\colon \mathbb{R}^{N_f} \to \mathbb{R}$ on features.
- A "movement cost" function $d(s_i, s_j)\colon S \times S \to \mathbb{R}$ on pairs of items.
- A measurement budget C.

Find a policy of measurement decisions and final selection (in the POMDP sense of a mapping from belief states into actions) that maximize the expectation of the following reward:

$$\begin{aligned} \text{max:} \quad & W = u(\mathbf{z}(s_\alpha)) - \sum_{i=1}^{N_q} \left(c^m_{k_i} + d\left(s_{i-1}, s_i\right)\right) \\ \text{s.t.:} \quad & \sum_{i=1}^{N_q} \left(c^m_{k_i} + d\left(s_{i-1}, s_i\right)\right) \leq C \end{aligned} \tag{3.1}$$

where $Q = \{q_i = (k_i, s_i) \,|\, i \in 1..N_q\}$ is the performed measurement sequence, where k_i is the type, and s_i is the location of measurement q_i, and s_α is the selected item. The expectation is taken according to the distribution B_0, which acts as the initial belief in the POMDP sense, updated according to information gathered by the measurements, if any.

There are several possible settings of the problem depending on the way information obtained from measurements is treated. This paper focuses on the *online measurement selection* [29] problem: a next measurement is selected after the outcomes of all preceding measurements are known. An alternative problem is *offline measurement subsequence selection*, in which the policy is just a predetermined sequence of measurements, without taking outcomes of individual measurements into account.

The above problem statement is still too general to handle directly. In typical settings of this selection problem, additional structure is usually imposed as follows.

Parameters: The set of items S can be indexed according to a set of parameters I. The parameter values of an item $s \in S$ are denoted by $I(s)$. In the simplest case I is just an integer index. Alternatively, I can be a set of coordinates, and the items S are indexed by points on a multi-dimensional grid. The latter case provides an approach to approximate optimization of an unknown continuous function.

Distance metric: When I is a set of coordinates, one common set of cost functions is where d is a function only of a distance metric over the item coordinates, i.e. $d(s_1, s_2) = f(||I(s_1), I(s_2)||)$ for some function f and some metric $||\cdot||$. This would be the type of cost encurred by having to physically move sensing equipment from location to location.

Distribution: The joint distribution over the feature space is also structured in many cases, with variables assumed independent except where indicated by a structured probability model such as a Bayes network, a Markov network, or a Markov random field. For example, when I is a multi-dimentional grid, the dependency between features of the items is typically modeled as a Markov random field with a neighborhood determined according to the grid—only (features of) items s' that are neighbors on the grid of s are in the Markov neighborhood of (features of) s.

The following examples can be viewed as instances of the selection problem:

Basic case: In this variant there is a finite number of items, and only one real-valued feature. The utility function is a simple monotonically increasing function of the feature values, such as the identity function. The features (of the different items) are mutually independent, the movement cost is zero, and the budget is infinite. This simple setting is still hard to solve, and resembles the well-known multi-armed bandit problem (see Section 7). Closed-form solutions are possible for very specific sub-cases, such as the example in Section 4.4.

Water reservoir monitoring: A water reservoir is monitored for contamination sources. Water probes can be taken in a number of predefined spots, and the goal is to predict the location and intensity of a contamination source based on analysis of water quality. The contamination must be identified quickly, before it distributes too far or affects the consumers. Here, the *features* are concentrations of possible contaminants, the *utility function* is a function of the concentrations; the *measurement cost* is determined by the time required to perform the analysis of a probe, and, in case of a large natural reservoir, the time and cost to move the probe physically between locations.

SVM parameter optimization: Classification quality of a support vector machine depends on one or more parameters (see Section 6.4 for a case study). While there are heuristics for selecting good parameter values for particular kernel types and data sets, several combinations of parameters must be tried before a good one can be chosen. The classifier accuracy is the only *feature*, and the *utility function* is a monotonicically increasing function of accuracy; the classifier parameters define *coordinates* of the items in a multi-dimensional grid, with distance-based *probabilistic dependencies* between the items. The *measurement cost* is the time required to train and validate the classifier for a combination of parameters. Movement cost for this problem is usually negligible.

Metrology machine setup: A focus and a filter color must be chosen for optimal setup of metrology equipment (see Section 6.5). For each combination of the focus and the color, the *item coordinates*, a number of *features* can be observed inexactly, and the *utility function* depends in a rather complicated way on the feature values. Both the measurements and the equipment movements take time and determine the *measurement* and *movement costs*, and the setup must be completed within a given *time budget*.

Chapter 4

Greedy Myopic Optimization

As the selection problem's state space is typically well beyond the capability of POMDP optimization techniques, we represent measurement policies in implicit form, as an adaption of the meta-resoning greedy myopic approach to optimization under uncertainty. The same simplifying assumptions used in meta-reasoning are used to achieve search efficiency (obviously at the cost of suboptimality) for the selection problem.

4.1 The Metareasoning Approach

In the framework of the rational metareasoning approach (Section 2.3), a measurement corresponds to a meta-level action, and the final selection is a base-level action. Thus, a measurement with the highest net value of information is performed at each step. The algorithm terminates when no measurement has positive apparent net value of information, at which point the item with the highest expected utility (given the current information) is selected and returned. The joint feature distribution is initialized to a *prior belief*, and then new evidence is incorporated into the belief through Bayesian inference [19].

The expected utility is computed according to (2.3), with integration instead of summation for the continuous case:

$$\mathbb{E}[U_i] = \iint\limits_{\mathbb{R}^{N_f}} u(\mathbf{z}) p_i(\mathbf{z}) d\mathbf{z} \tag{4.1}$$

where p_i is the current belief about feature values of item s_i.

The net value of information V is the difference between the intrinsic value, Λ, and the measurement cost. For a finite set of items of size N_s, there are $N_m \cdot N_s$ different measurements at each step, each determined by the measurement type and the measured item. Assuming that the measurements are arbitrarily indexed, the value of the jth measurement is:

$$V_j = \Lambda_j - c_j = \Lambda_j - d(s, s_j) - c_j^m \tag{4.2}$$

where s is the last measured item, and c_j is the net measurement cost—the sum of the intrinsic measurement cost and the movement cost. Under the myopic simplifying assumptions (Section 2.3), the intrinsic value of information is estimated myopically: as the expected immediate effect of

1: $budget \leftarrow C$
2: Initialize beliefs to B_0
3: **loop**
4:     **for all** items s_i **do**
5:         Compute $\mathbb{E}(U_i)$
6:     **for all** measurements m_j **do**
7:         **if** $c_j \leq budget$ **then**
8:             Compute V_j
9:         **else**
10:             $V_j \leftarrow 0$
11:     $j_{max} \leftarrow \arg\max_j V_j$
12:     **if** $V_{j_{max}} > 0$ **then**
13:         Perform measurement $m_{j_{max}}$
14:         Update beliefs
15:         $budget \leftarrow budget - c_{j_{max}}$
16:     **else**
17:         **break**
18: $\alpha \leftarrow \arg\max \mathbb{E}(U_i)$
19: **return** s_α

Figure 4.1: Greedy Myopic Algorithm

a single measurement. Following (2.8), the intrinsic value of information is determined by the following equation:

$$\begin{aligned}\Lambda_j &= \mathbb{E}(\mathbb{E}[U_{\alpha_j}] - \mathbb{E}[U_\alpha]) \\ &= \iint\limits_{\mathbb{R}^{2N_f}} (\mathbb{E}[U_{\alpha_j}] - \mathbb{E}[U_\alpha]) p_j^m(\mathbf{z}_m|\mathbf{z}) p_j(\mathbf{z}) d\mathbf{z}_m d\mathbf{z}\end{aligned} \tag{4.3}$$

where s_α is the item with the highest expected utility before the measurement, s_{α_j} — after the jth measurement, and the expected utilities are computed for the updated beliefs.

4.2 Algorithm Description

The greedy myopic algorithm (see Figure 4.1) maintains a persistent data structure which holds beliefs about feature values of the items. The beliefs are initialized to the prior beliefs (line 2), and then updated according to measurement outcomes (line 14). The main loop (lines 3–17) continues as long as there are measurements that fit within the budget (line 7) with positive value (line 12). Otherwise, the loop terminates (line 17), and the algorithm returns the parameter vector of an item with the maximum expected utility (line 18).

The remaining budget is held in variable *budget*, which is initialized to the total budget C, and decreased by the cost of each performed measurement. Thus, the algorithm is guaranteed to terminate when the costs of all measurements are positive and bounded away from zero.

4.3 Theoretical Bounds

Online measurement selection is NP^{PP}-hard [15], thus the need for approximation algorithms arises. The greedy myopic algorithm is one feasible approach; however, in the general case the algorithm can behave arbitrarily badly. For example, if at least two measurements are required to achieve a positive value, the algorithm will terminate without making any measurements (see below). On the other hand, if the myopic value of information estimate of a measurement is submodular, the algorithm is nearly optimal [16].

The myopic value of information estimate in the optimization problem can be shown to approach submodularity in some cases, for example, when the measurements are exact and the utility function is concave. Mostly, however, the estimate is nonsubmodular and no non-trivial performance guarantees can be provided.

4.4 Shortcomings of the Myopic Estimate

The simplifying assumptions behind the myopic estimate are related to the notion of *non-increasing returns*: an implicit hypothesis that the intrinsic value grows more slowly than the cost. When the hypothesis is correct, the assumptions should work well; otherwise, the myopic algorithm either gets stuck or makes measurements which gain little useful information.

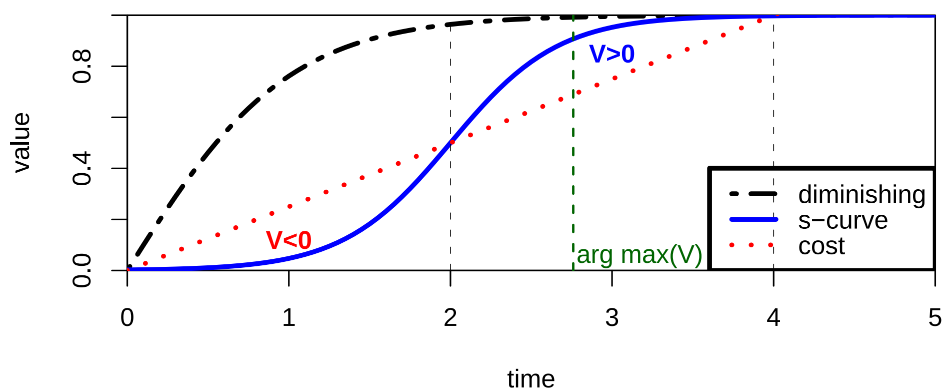

Figure 4.2: The s-curve

However, the *law of diminishing returns* [12] only holds asymptotically: while it is often true that starting with some point in time the returns never grow, until that point they can alternate between increases and decreases. Figure 4.2 presents the curve of diminishing returns (the dashed curve), the s-curve (the solid curve), and the areas with negative and positive values for the s-curved returns for a linear cost (the dotted straight line). Investments for the first two units of time do not pay off, and the maximum return is achieved at approximately 2.8.

Sigmoid-shaped returns were discovered in marketing [11]. As experimental results show [31], they are not uncommon in sensing and planning. In such cases, an approach that can deal with increasing returns must be used.

Even in simple cases the myopic estimate may behave poorly. Consider the following example:

- S is a set of two items , s_1 and s_2 with a scalar feature vector $\mathbf{z} = z$;

- the value of feature z of s_1 is known exactly, $z(s_1) = 1$;
- the prior belief about s_2 is a normal distribution $p_2(z) = N(z; 0, 1)$;
- the observation conditional variance is $\sigma_m^2 = 5$;
- the measurement cost is constant, and chosen so that the net value estimate of a two-observation step is zero: $c(j) = \frac{MVI_2^2}{2} \approx 0.00144$; the movement cost is zero.
- the utility is a step function:

$$u(z) = \begin{cases} 0 & \text{if } z < 1 \\ 0.5 & \text{if } z = 1 \\ 1 & \text{if } z > 1 \end{cases}$$

The plot in Figure 4.3 depicts the intrinsic value of information as a function of the number of observations in a single step. The straight line corresponds to the measurement cost.

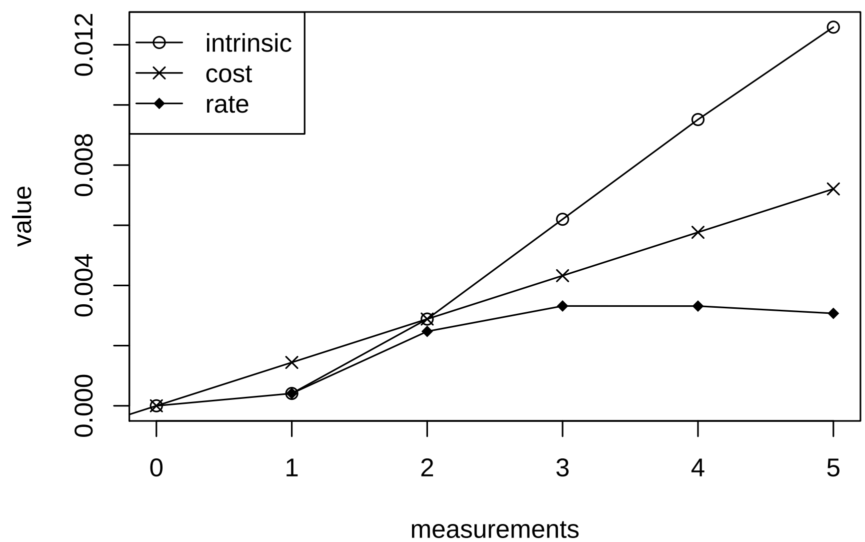

Figure 4.3: Intrinsic value and measurement cost

Under these conditions, the pure myopic algorithm (which considers only one measurement per step) will terminate without gathering evidence because the value estimate of the first step is negative, and will return item s_1 as best. However, observing s_2 several times in a row has a positive value, and the updated expected utility of s_2 can eventually become greater than $u(\mathbf{z}(s_1))$. Figure 4.3 also shows the intrinsic value growth rate as a function of the number of measurements: the rate increases up to a maximum at 3 measurements, and then goes down. Apparently, the myopic algorithm does not “see” as far as the initial increase.

Chapter 5

Semi-Myopic Optimization

As the pure myopic scheme is too short-sighted in many cases, some lookahead is needed, at increased computational cost. The proposed semi-myopic optimzation framework allows for a spectrum of schemes that allow a choice of such tradeoffs. The scheme is analyzed qualitatively, and theoretical bounds for two important extreme cases are presented.

5.1 Semi-Myopic Schemes

Keeping the complexity manageable (the number of possible sensing plans over continuous variables is uncountably infinite) while overcoming the limitations of the myopic algorithm is the basis for the *semi-myopic* framework. Let $\mathcal{M}$ be the set of all possible measurement actions, and $\mathcal{C}$ be a constraint over sets of measurements from $\mathcal{M}$. In the semi-myopic framework, all possible subsets (batches) B of measurements from $\mathcal{M}$ that obey the constraint $\mathcal{C}$ are considered, and for each such subset $\mathcal{B}$ a 'batch' value of information estimate is computed under the assumption that all measurements in $\mathcal{B}$ are made, followed by a decision (selection of an item). Then, the batch $\mathcal{B}^*$ with the best value estimate is chosen. Once the best batch $\mathcal{B}^*$ is chosen, there are still several options ([2] discusses a similar approach to discrete feature acquisition):

1. Actually do all the measurements in $\mathcal{B}^*$.
2. Attempt to optimize $\mathcal{B}^*$ into some form of conditional plan of measurements and the resulting observations.
3. Perform the best measurement in $\mathcal{B}^*$.

In all cases, after measurements are made, the selection is repeated, until no batch has a positive value, at which point the algorithm *terminates* and selects an item. Option 1, a combination of online and offline selection, misses the possibility to choose better measurements while executing a batch, and therefore is inferior to options 2 and 3. While limited efficient implementation for option 2 is possible, optimizing a conditional plan is intractable in general, recreating the optimal policy problem on a smaller scale. However, option 3, with a suitable constraint, remains tractable and offers an opportunity to make a better choice at each step.

for all batches b_j satisfying constraint $\mathcal{C}$ **do**
  if $cost(b_j) \geq budget$ **then**
    compute V_j^b
  else
    $V_j^b \leftarrow 0$
$j_{max} \leftarrow \arg\max_j V_j^b$
if $V_{j_{max}}^b > 0$ **then**
  for all measurements $m_k \in b_{j_{max}}$ **do**
    compute V_k
  $k_{max} \leftarrow \arg\max_k V_k$
  perform measurement $m_{k_{max}}$
  update beliefs
  $budget \leftarrow budget - c_{k_{max}}$

Figure 5.1: Semi-myopic Scheme

In the semi-myopic version, the measurement selection part (lines 6–15) of the algorithm in Figure 4.1 is replaced with the algorithm in Figure 5.1. Here, value of information is computed twice: first, value of information V_j^b of every batch b_j satisfying the constraint $\mathcal{C}$ is computed; then, if the maximum value of information of a batch is positive, value of information V_k of every single measurement m_k belonging to the batch $b_{j_{max}}$ is computed, and a measurement with the highest value of information (which can be negative) is performed.

The constraint $\mathcal{C}$ is crucial and defines a *scheme*. For the empty constraint — the *exhaustive* scheme — all possible measurement sets are considered; this scheme has an exponential computational cost, while still not guaranteeing optimality. At the other extreme is the constraint where only singleton sets are allowed. This extreme results in the greedy single-step assumption — the original myopic scheme. [2] proposes a constraint based on the *value of information lattice*, a data structure introduced in that paper. The induced number of subsets is still exponential in a general case, but for certain kinds of dependency structure the constraint gives significant improvement. Yet another — *omni-myopic* — constraint can be constructed along the lines of [9]: the measurements are ordered according to their myopic VOI estimates, and subsets of measurements with greatest VOI estimates are considered.

This paper investigates the constraint that restricts the batches to repeated identical measurements— the *blinkered* scheme. This scheme is tractable, corresponds to a common approach in which noisy experiments are repeated to increase confidence, and demonstrates improved algorithm efficiency in empirical evaluations.

5.2 Blinkered Estimate

As stated above, the blinkered scheme considers sets of independent identical measurements; this constitutes unlimited lookahead, but along a single “direction”, as if one “had one’s blinkers on”. Although this scheme has a computational overhead over the myopic one, the factor is only linear in the budget. [1] The "blinkered" value of information is defined as:

$$
\begin{aligned}
BVI_i &= \max_k MVI_i^k \\
\text{s.t.}: & \quad c_i = d_i + c_i^m k \le C
\end{aligned}
\tag{5.1}
$$

where, as before, d_i and c_i^m are the movement cost and the intrinsic measurement cost of the ith measurement, correspondingly.

Driven by this estimate, the blinkered scheme selects a single measurement of the item where *some* number of measurements gains the greatest value of information. A single step is expected to be just the first one in the right direction, rather than to actually achieve the value. Thus, the estimate relaxes the *single-step* assumption, while still underestimating the value of information.

For budget C and a single measurement cost bounded from below by c the time to compute the estimate T_{BVI} is: $T_{BVI} = O\left(T_{MVI}\frac{C}{c}\right)$. This time bound is still exponential in the representation, but can be made polynomial in one of the following ways:

- If MVI^k is a unimodal function of k, which can be shown for some forms of distributions and utility functions, the time is logarithmic in the budget C. Indeed, using the bisection method to find k for which $MVI^{k+1} - MVI^k$ changes sign yields the logarithmic time.
- Otherwise, BVI can be estimated as the maximum for powers of 2: $k = 2^{k'}$.

In the presence of inexact recurring measurements, the blinkered scheme plays a role similar to the role of the myopic scheme for exact measurements. It seems to be the simplest approximation that still works for a wide range of conditions under realistic assumptions.

5.3 Theoretical Bounds

Bounds on the blinkered scheme are established for two special cases. The first bound shows that the scheme does not stop measuring too early.

Theorem 1. *Let $S = \{s_1, s_2\}$, where the value of s_1 is known exactly. Let C_b be the remaining budget of measurements when the blinkered scheme terminates. Then the (exact) value of information of an optimal policy from this point onwards is at most C_b.*

The second bound is related to a common case of a finite budget and free measurements. The bound provides certain performance guarantees for the case when dependencies between items are sufficiently weak, and shows that the blinkered scheme chooses reasonable search directions under weaker assumptions than those for the myopic scheme.

Definition. *Measurements of two items s_1, s_2 are mutually subadditive if, given sets of measurements of each item Q_1 and Q_2, the intrinsic value of information of the union of the sets is not greater than the sum of the intrinsic values of each of the sets, i.e.:* $\Lambda(Q_1 \cup Q_2) \le \Lambda(Q_1) + \Lambda(Q_2)$

Theorem 2. *For a set of N_s items, the zero measurement cost $c = 0$, and a finite budget, if measurements of every two items are mutually subadditive, the value of information of the blinkered scheme is no more than a factor of N_s worse than the value of information of an optimal algorithm.*

[1] This complexity assumes either normal distributions, or some other distribution that has compact statistics. For general distributions, sets of observations may provide information beyond the mean and variance, and the resources required to compute value of information may even be exponential in the number of measurements.

Proofs for the bounds are provided in Appendix A. Both bounds are asymptotically tight, as can be shown by constructing appropriate problem instances.

5.4 Reusing Computations

On the one hand, limited rationality approach (Section 2.3) assumes that the time required to estimate value of information of a meta-level action is negligible. On the other hand, the greedy algorithm (Chapter 4) performs massive computations at each step:

- evaluates the expected utilitiy of every item;
- evaluates the value of information of every measurement;
- updates the beliefs by propagating the newly obtained evidence;

and the need to implement the computations efficiently is even more prominent when a semi-myopic value of information estimate is used, because more evaluations of value of information per step must be performed.

This section suggests improvements to the greedy algorithm that make the algorithm run faster and scale better to large numbers of items and evaluation kinds. Most of the improvements are implemented in the code used for empirical evaluation (Chapter 6) and proved useful in practice.

5.4.1 Expected Utilities

At each step, the greedy algorithm evaluates the expected utility of every item. The expected utility of an item depends only on beliefs about the item features. If the beliefs have not been affected by the last update, the computation from an earlier step can be re-used.

The beliefs are affected by an update when either the item has been directly observed, or when the beliefs about the item's features have been changed due to belief propagation after observing another item. Thus, to implement selective recomputation of expected utilities, the dependency model must be extended to provide information about a change in the belief at every item. With this modification, the utility updating loop (lines 4–6) of the algorithm in Figure 4.1 will take the following form:

for all items s_i **do**
  if s_i updated **then**
    compute $\mathbb{E}(U_i)$

This change is exact: the algorithm still recomputes all expected utilities which might have changed. However, the change considerably decreases the computation time, in particular, in the case of large dependency models with disconnected components.

5.4.2 Domain-specific Inference

The greedy algorithm spends a significant portion of the deliberation time updating beliefs. When the dependency model is large or complicated, a straightforward implementation of an inference algorithm is too slow. Better results are achieved when the structure of updates is exploited.

An iterative belief propagation algorithm is commonly used for inference in networks of general topology. The algorithm initializes messages between nodes to some values, and then traverses the network repeatedly until the computed beliefs converge or the iteration limit is reached.

Incremental Updating

Generally, the messages are initialized arbitrarily. Then, if the algorithm converges, the messages approach a fixed point. Good initial values cannot be chosen easily and are not worth bothering with: the algorithm rapidly approaches the proximity of a fixed point and then slowly, in the presence of loops,[2] achieves the required precision.

However, in the optimization problem, the evidence is added gradually, and the model drives the search but does not significantly affect the final result: at the end of the search, best and close to best nodes are known with high confidence. Therefore, the messages can be reused and the computation can be sped up:

- final message values in each iteration become initial values for the next iteration;
- since only a small amount of evidence is added at each iteration, most messages remain almost unchanged;
- relatively low precision is sufficient, therefore good initial values significantly decrease the computation time.

Messages are initialized once in the beginning, and then the model is updated incrementally.

Propagation Order

The convergence speed depends on the traversal order. The following example illustrates the dependence: in a one-dimensional Markov random field, $x_1, \ldots, x_n$ (Figure 5.2), the only obtained evidence is y_n for the rightmost node x_n.

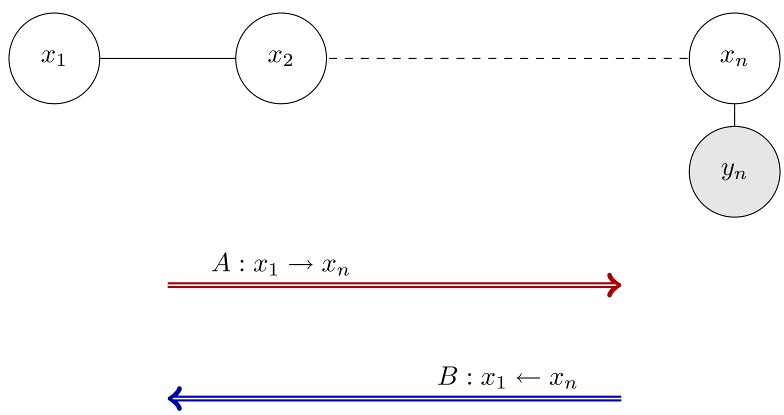

Figure 5.2: Propagation direction

When the propagation algorithm traverses the chain from left to right (A), it takes $O(n)$ iterations to propagate the evidence through the chain. However, if the propagation direction is from right to left (B), only $O(1)$ iterations are required.

[2] When the network graph is a polytree, the algorithm converges to an exact solution in a fixed number of iterations [6].

Every time the optimization algorithm updates beliefs, new observations are available for just a small number of variables.

- when a measurement is chosen by the algorithm, only a single node is measured;
- before a measurement is chosen, the algorithm simulates batches of measurements; but every batch should be reasonably small for the algorithm to remain tractable.

Like in the one-dimensional case, the propagation should start from the newly observed variables and recursively proceed to the neighbors. This order ensures that the convergence speed does not depend on the measurement location. Besides that, if the dependency model is disconnected—consists of multiple components with probabilistic dependencies only within the components but not between them (see Section 6.5 for a case study)—only affected components participate in message passing and the computation time is not wasted in the parts which were not updated.

Value of Information Estimates

The utility function depends on the parameter vector (see Section **??**). Thus, to compute value of information of a batch of measurements, a multidimensional integral must be evaluated. The integral is commonly estimated using a Monte Carlo method, which means a function must be evaluated multiple times with different arguments. Each such evaluation corresponds to a new evidence, and the dependency model must be updated (and reverted) multiple times per computation of value of information of each batch.

With large models and multiple measurements this amount of computation becomes prohibitive. Fewer iterations of belief propagation is a natural solution which can be justified in the following way:

- an error in value of information estimate results in a less efficient measurement (in the worst case), but does not affect the dependency model and the final choice;
- the propagation error does not accumulate: the evidence is added incrementally, and while the estimated influence of the batch is computed with a lower precision, the propagation messages are are initialized from precise inference of all the evidence up to date.

The approximation can be taken to extreme, and value of information can be computed without propagating beliefs at all: only the beliefs that are directly affected by a measurement are updated. The algorithm used for the empirical evaluation (Section 6) does not propagate beliefs while computing value of information. This approximation has two potential drawbacks:

- value of information of measurements in close vicinity of the α item can be overestimated;
- superadditivity of measurements caused by dependencies cannot be discovered.

However, these cases are infrequent, and in practice computation of value of information without belief propagation works well.

Chapter 6

Empirical Evaluation

The blinkered scheme is evaluated empirically both with artificial data and in real-world case studies.

6.1 Algorithm Implementation

UNCERTIMA, a toolkit for optimization under uncertainty, has been developed with the purpose to explore optimization approaches and deploy efficient optimization algorithms in applications. At the core of the toolkit is a generalized implementation of the greedy algorithm. The greedy algorithm is parameterized by the measurement selection scheme, the dependency model, and the optimization object.

- The myopic and blinkered schemes are available; other schemes, both based on value of information and fixed, used for burn-in, may be added in the future.
- The currently implemented dependency model is a stationary multi-dimensional rectangular lattice. The lattice parameterized by prior beliefs and covariances in each direction. The work is underway to add other topologies and non-stationary models.

The output of the optimization algorithm is the chosen combination of parameters, the expected utility and the measurement budget surplus. Additionally, the final beliefs and the log of all measurements are stored for analysis.

6.2 Simulated Random Inputs

In this set of experiments, a set of independent items with a single feature is given. The utility function is the step function (see the example in Section 4.4), and the initial beliefs are normal distributions $N(\mu = 0, \sigma^2 = 1)$. True feature value is randomly generated for each item from the initial belief distribution; then, for a range of measurement costs, budgets, and observation precisions, the simulation is run. Observation outcomes are generated randomly according to the exact values of the item features and the measurement model. The performance measure is the regret—the difference between the utility of the best item and the utility of the selected item less the measurement costs. Results of using the myopic scheme, the blinkered scheme and other semi-myopic schemes are compared.

6.2.1 Blinkered vs. Myopic

The first comparison is the difference in regret between the myopic and the blinkered scheme, done for 2 items (Table 6.1). Positive values in the cells indicate an improvement due to the blinkered estimate. Table 6.2 repeats the experiment for 4 items.

Table 6.1: Myopic vs. blinkered: 2 items, measurement budget 5

σ_o C	0.0005	0.0010	0.0015	0.0020
3.0	0.0147	0.0156	0.0199	0.2648
4.0	0.0619	0.2324	0.2978	0.2137
5.0	0.2526	0.2322	0.1729	0.1776
6.0	0.1975	0.1762	0.1466	0.0000

Table 6.2: Myopic vs. blinkered: 4 items, measurement budget 10

σ_o C	0.0005	0.0010	0.0015	0.0020
3.0	0.0113	-0.00459	-0.0024	0.4352
4.0	0.0374	0.43435	0.4184	0.3902
5.0	0.4060	0.40004	0.3534	0.3599
6.0	0.4082	0.37804	0.3337	0.0000

Averaged over 100 runs of the experiment, the difference is significantly positive for most combinations of the parameters. In the first experiment (Table 6.1), the mean regret of the myopic scheme compared to the blinkered scheme is 0.15 with standard deviation 0.1. In the second experiment (Table 6.2), the mean regret is 0.27 with standard deviation 0.19.

6.2.2 Other Semi-Myopic Estimates

Three semi-myopic schemes, introduced in Section 5.1, are compared: blinkered, omni-myopic, and exhaustive[1]. All schemes are run on a set of 5 items with a measurement budget of 10.

The results show that

- blinkered is significantly better than omni-myopic: the mean regret is 0.20 with standard deviation 0.16;
- exhaustive is only marginally better than blinkered, despite evaluating an exponential number of sets of measurements: the mean regret is 0.03 with standard deviation 0.10;
- omni-myopic is only marginally better than myopic: the mean regret is 0.02 with standard deviation 0.07.

6.2.3 Dependencies between Items

When linear dependencies between the items are added, e.g. when: $z_i = z_{i-1} + w$ with w being a random variable distributed as $N(0, \sigma_w^2)$, the VOI of series of observations of several items can

[1]The omni-myopic and the exhaustive scheme were added to the implementation for these experiments but are not available generally due to their intractability.

be greater than the sum of VOI of each observation. The experiment examines the influence of dependencies on the relative quality of the blinkered and the omni-myopic scheme.

Without dependencies, i.e. for $\frac{\sigma_o^2}{\sigma_w^2} = 0$, the blinkered scheme is significantly better. But as σ_w decreases, the omni-myomic estimate performs better. Figure 6.1 shows the difference between achieved utility of the blinkered and the myopic schemes with dependencies. The experiment is run on a set of 5 items, with the prior belief $N(0,1)$, measurement precision $\sigma_o^2 = 4$, measurement cost $C = 0.002$ and a budget of 10 measurements. The results are averaged over 100 runs of the experiment.

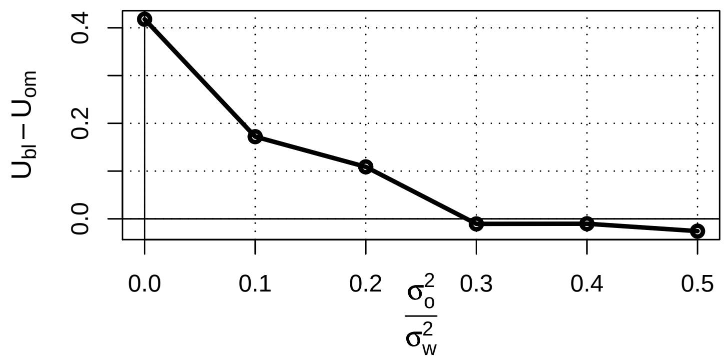

Figure 6.1: Influence of dependencies

In the absence of dependencies, the omni-myopic algorithm does not perform measurements and chooses an item at random, thus performing poorly. As the dependencies become stronger, the omni-myopic scheme collects evidence and eventually outperforms the blinkered scheme. In the experiment, the omni-myopic scheme begins to outperform the blinkered scheme when dependencies between the items are roughly half as strong as the measurement accuracy.

6.3 Optimization Benchmarks

A popular way to compare optimization algorithms is to apply them to real-valued functions with multiple minima and maxima. Such functions are difficult for optimization and reflect the challenges which optimization algorithms meet in real world problems. Two functions were used for the study, the Ackley function [1] and a modified version of the Himmelblau function [28]. The functions are scaled into a cube with coordinate range $[-1, 1]$ along each edge.

The two-argument form of the Ackley function is used. The function is defined by the expression (6.1):

$$\begin{aligned} A(x,y) &= 20 \cdot \exp\left(-0.2\sqrt{\frac{x^2+y^2}{2}}\right) \\ &\quad + \exp\left(\frac{cos(2\pi x) + cos(2\pi y)}{2}\right) \end{aligned} \tag{6.1}$$

6.3.1 Optimization Problems

Two optimization problems are considered, with a sigmoid utility function $\tanh(2z)$. In the former problem the movements are free; in the latter one the movement cost is proportional to the Manhattan distance. In both problems, the measurements are noisy with $\sigma_e^2 = 0.5$ and and there are uniform dependencies with $\sigma_w^2 = 0.5$ between neighbor nodes in both directions of the coordinate grid with a step of 0.2 along each axis.

6.3.2 Experiment and Representation of Results

For each problem, the myopic and the blinkered scheme were run 64 times. The chosen locations were recorded, as shown graphically (see Figures 6.2, 6.4). In each figure, the plots in the upper and the lower rows depict the selected locations[2] and the histogram of rewards (the sum of the utility and the budget surplus) for the myopic and the blinkered scheme, correspondingly. The extent to which the locations are grouped around the maxima, and the utility distribution histograms provide for easy visual comparison of the results. The accompanying textual description includes some quantitative performance measures.

6.3.3 Free Movements

The case of free movements is relatively easy for the information-based optimization, and both schemes behave comparably (Figure 6.2). close to the global maximum of the Ackley function but fail to identify the drops in the utility in the saddles $(\pm 0.2, 0)$, $(0, \pm 0.2)$. For the Himmelblau function, most selected locations are in high utility areas, but since the proximity of the global maximum is rather flat, the algorithm terminates without achieving the exact maximum: further measurements are not worth their cost. $(\pm 0.2, 0)$, $(0, \pm 0.2)$. Still, for the Ackley function, the myopic scheme produced two outliers: $(0.6, 0.6)$ and $(-1, -1)$. The mean reward is 0.92 for the blinkered, 0.81 for the myopic scheme. Thus, while on average in the case of free movements both schemes behave comparably for the given parameters, the myopic scheme is more likely to select a location with a low utility.

For the Himmelblau function, the utility histogram for the myopic scheme has a longer left tail, with rewards as low as ≈ 0.6, while the lowest reward for the blinkered scheme is ≈ 0.9. The myopic scheme also chose a few locations with low true utilities: $(-0.2, -0.8)$, $(0.0, -0.6)$. The difference in the mean reward is less prominent: 1.25 for the blinkered, 1.23 for the myopic scheme.

Thus, while on average in the case of free movements both schemes behave comparably for the given parameters, the myopic scheme is more likely to select a location with a low utility.

6.3.4 Manhattan Distance Movement Costs

When the movement cost is non-zero, the myopic scheme, which cannot take into account amortization of the movement cost over multiple measurements, is essentially inferior to the blinkered scheme. Indeed, the experiment shows better performance of the blinkered scheme (Figure 6.4) on both functions.

For the Ackley function, the mean reward is 0.59 for the blinkered, 0.40 for the myopic scheme. The locations selected by the myopic scheme are spread over a wide area and are less focused around

[2] The coordinates of selection locations are slightly randomly moved in the plots to show that a particular location was selected multiple times.

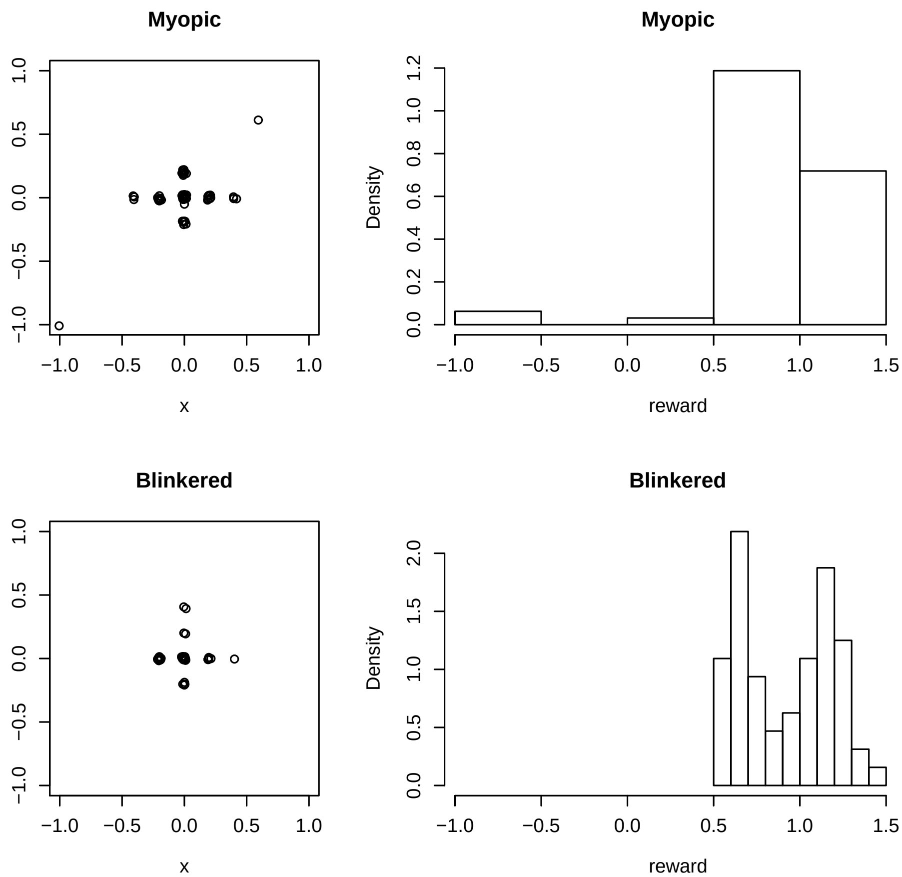

Figure 6.2: The Ackley function, free movements.

the true maximum $(0, 0)$; 9 locations among selected by the myopic scheme have a negative utility, compared to 3 for the blinkered scheme.

The difference in performance for the Himmelblau function is less obvious from the location plot due to the flatter shape of the function, but the reward distribution histogram shows more locations with lower utility. The mean reward is 1.20 for the blinkered, 1.10 for the myopic scheme.

6.4 Case Study: Optimization of SVM Parameters

Support Vector Machine (SVM) is a popular family of supervized learning methods for classification and regression [3]. SVM-based classifiers are efficient and robust. However, most kernel functions are parameterized, and the parameters must be properly chosen for each particular kind of data in order to achieve good classification accuracy.

An SVM classifier based on the *radial basis function* has two parameters: C and γ. While there are heuristics for selecting an initial approximation for a good parameter combination, an efficient algorithm for determining their values is not known and several different combinations

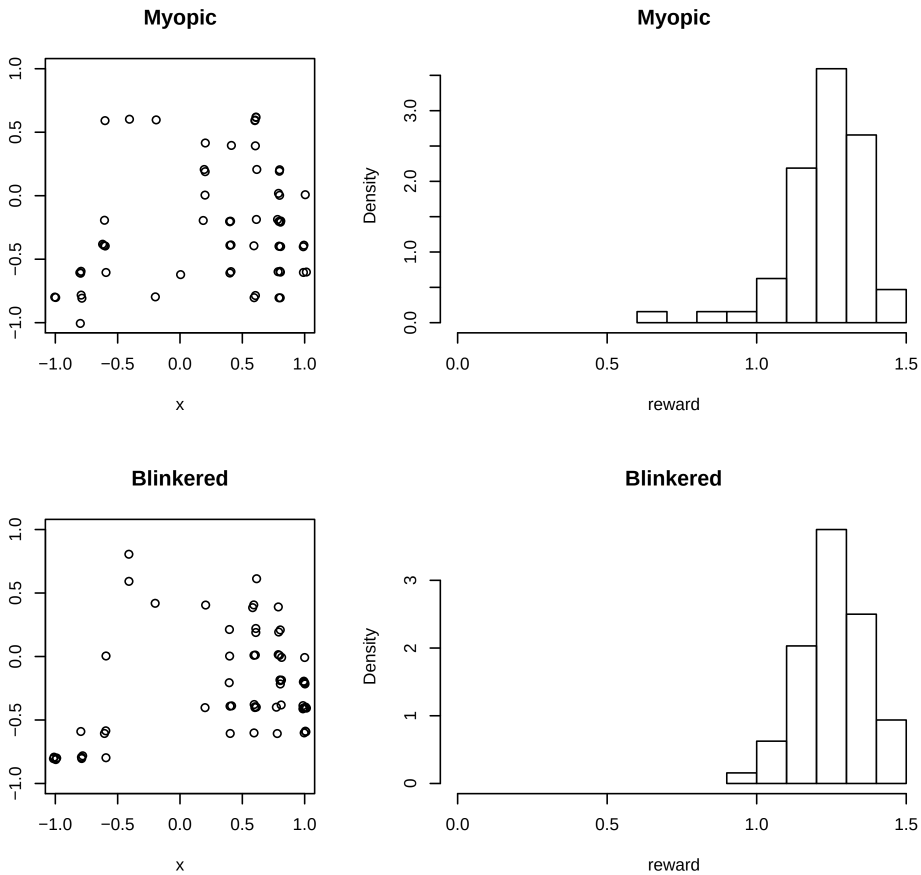

Figure 6.3: The Himmbelblau function, free movements.

must be tried. The parameter search space is usually exponential, such that $C_i = \alpha^i$ and $\gamma_j = \beta^j$, and dependency of the classification accuracy on the parameters has a shape similar to shown in Figure 6.6. A point in the upper flat region, preferably away from the edges, corresponds to a good parameter combination. Points lying in the lower flat region yield the **baseline** accuracy, which is no better than a random class label.

A trial for a combination of parameters determines estimated accuracy of the classifier through cross-validation. One possibility is the *complete data* approach— the complete training data set is used for each trial, and the accuracy is determined with high confidence, that is, an exact, or an almost exact, measurement is performed. However, the time required to estimate the accuracy is roughly quadratic in the size of the data set, and with larger sets the computation time can become prohibitive.

An alternative, *incremental data* approach is to perform the trials on smaller subsets of the data set, such that estimating the accuracy on each subset is imprecise but fast, and the trials

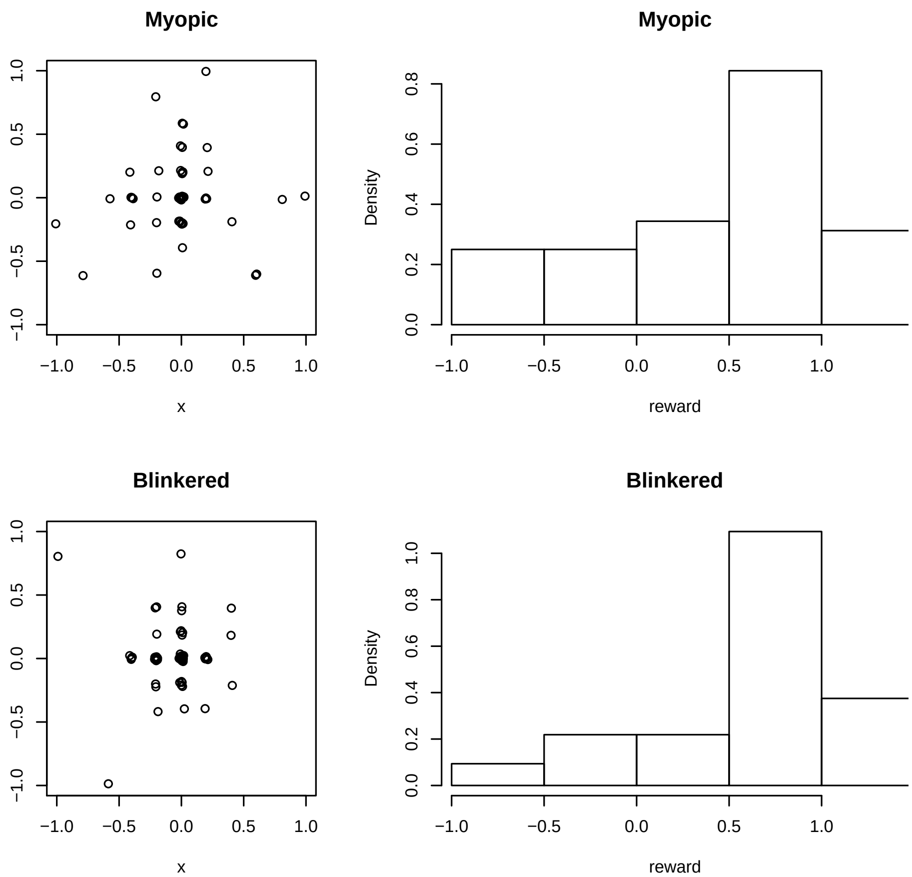

Figure 6.4: The Ackley function, manhattan movement cost.

are repeated multiple times in some of the locations to increase the confidence, each time with a different subset. The latter approach is more time-efficient, and is based on recurring imprecise measurements. This case study compares the myopic scheme for the complete data approach with both the myopic and the blinkered scheme for the incremental data approach. It is anticipated that the best results are achieved through combining the incremental data approach with the blinkered scheme.

6.4.1 Data Sets

Two data sets are used for the case study: USPS [10] and SVMGUIDE2 [27]. While the data sets have a different number of features and specimens, the most essential difference is in the number of classes: 10 and 3 correspondingly. Due to the difference in the number of classes, the accuracy for a poor combination of parameters, which can be estimated as $\frac{1}{number-of-classes}$, is different, as well as the steepness of the slope between the good and the bad regions.

USPS — 10 classes, 2007 specimens, 256 features;

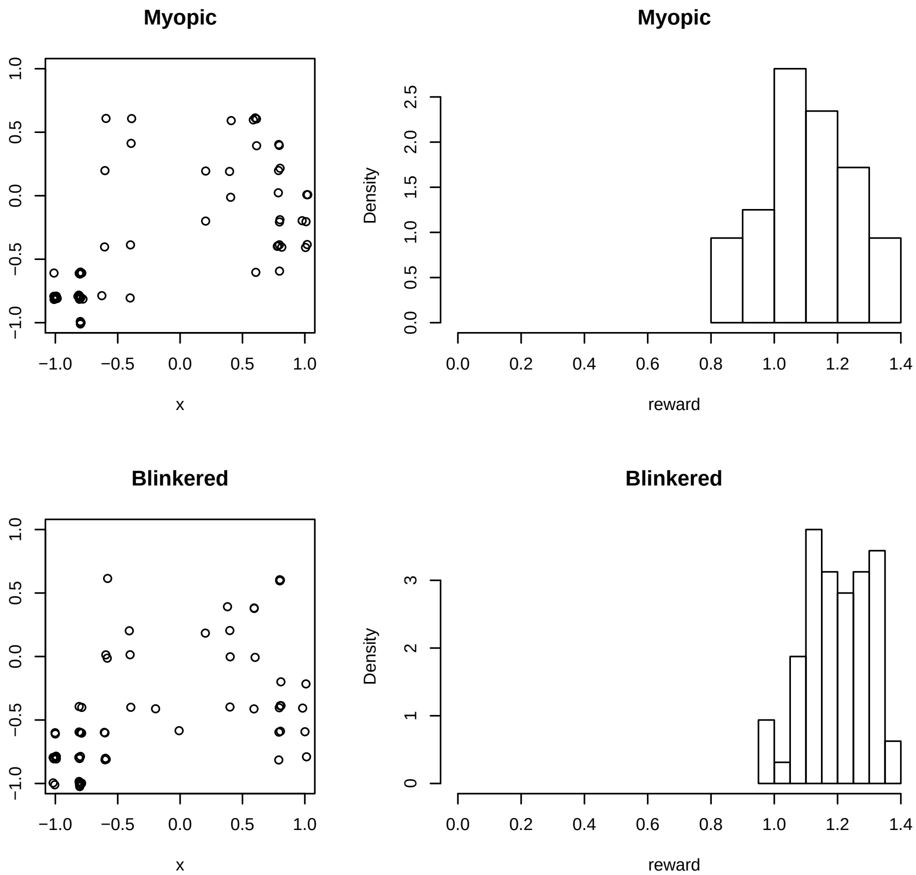

Figure 6.5: The Himmelblau function, manhattan movement cost.

SVMGUIDE2 — 3 classes, 391 specimens, 20 features.

6.4.2 Optimization Problems

Two optimization problems are defined, one for each of the approaches. The problem definitions are the same for both data sets, and the parameters are estimated from the data. For the incremental data approach, the data sets are divided into 8 equal parts. It is essential that the same problem definitions used for both data sets; the same algorithm should be applicable to different problem instances without much tuning. The utility function in both problems is $tanh(4(x-0.5))$, the $\log C$ and $\log \gamma$ axes are scaled for uniformity to ranges [1..21] and there are uniform dependencies along both axes with $\sigma_w^2 = 0.4$. The difference is in the accuracy (0.01 for complete, 0.25 for incremental trials) and the measurement cost (0.025 for complete, 0.01 for incremental trials). The movement cost is zero in both problems.

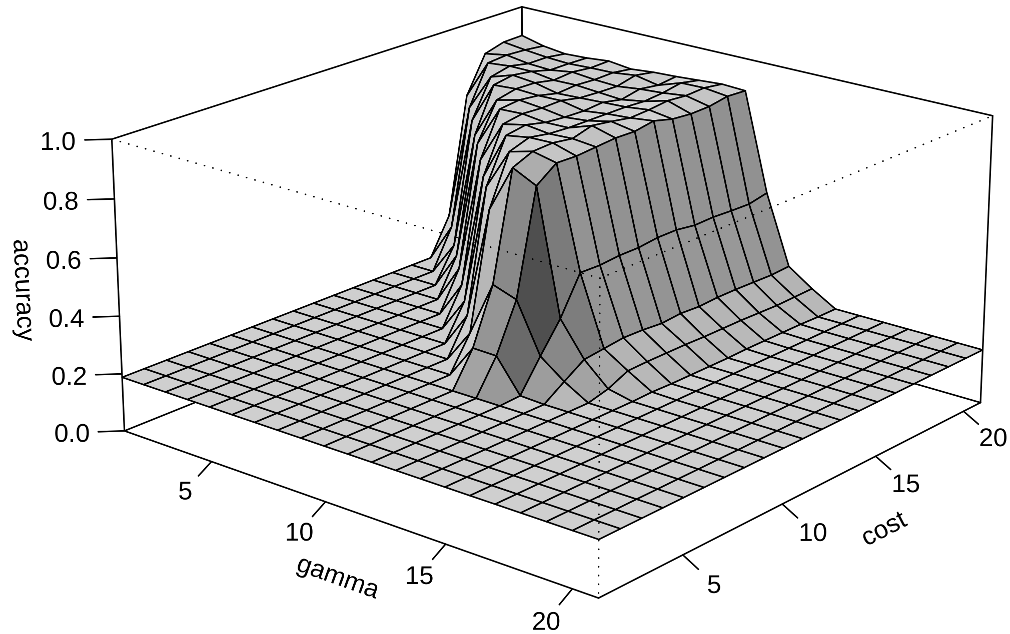

Figure 6.6: SVM classification accuracy as function of $\log C, \log \gamma$.

6.4.3 Experiment and Representation of Results

The experiment design is similar to that of the function optimization in Section 6.3.2. The myopic scheme is run for the complete approach, and both the myopic and the blinkered schemes are run on each data set for the incremental data approach, 64 times in each case. Results for the SVMGUIDE2 data set, turned out to be more difficult for optimization, are presented in Figure 6.7. In the figure, the first row corresponds to the complete trials, the second and the third row—to the myopic and the blinkered scheme for the incremental trials. The plots in the left column show the combinations of parameters chosen by the algorithm on each run (slightly randomly moved, as before, to show multiply selected combinations), and the plots in the right column depict distributions of the net utility of the result.

6.4.4 Results and Conclusions

The complete data approach exhausted the budget in most runs, and in several cases the selected parameter combinations yielded the **baseline** accuracy. The myopic scheme for the incremental data approach selected an optimal combination in most runs, but there still were quite a few outliers. The blinkered scheme, however, succeeded in always selecting an optimal combination. The mean reward is 0.88 for the complete data approach, 1.10 for the myopic, 1.12 for the blinkered scheme.

6.5 Case Study: Metrology Machine Setup

Another case study considers a complicated optimization problem arising in the field of metrology equipment setup: two parameters, the focus and the filter color must be chosen to optimize confidence of outcomes of subsequent metrology of a wafer for electronic chip production.

For each focus and color filter, the measurements can be performed in several locations (sites), and for two orientations of the wafer. The utility function (see the problem definition below) depends on 6 compound features; and each of the features is an aggregate expression over the sites.

The dependency model consists of 5 (the number of filter colors) unconnected three-dimensional lattices (an axis along the foci and two axes in the wafer plane, with measurement sites in the nodes).

The movement cost is a step function of the parameter combinations, as it is the fact that a parameter is changed that makes the difference, rather than the magnitude. Parameter changes differ in cost: changing the wafer orientation is more expensive than changing the filter.

6.5.1 Optimization Problem

To save the computation time for simulations, only each third focus (1, 4, 7, ... 31) and only 3 of 9 sites are considered. Even with this simplifications, there are 330 possible measurements and 1980 optimization features in the dependency model at each point in time.

6.5.2 Experiment and Representation of Results

Simulation data for two wafers, A and B, was used. For each of the wafers, the myopic and the blinkered scheme were run 10 times; results for wafer A are presented in the plots (Figures 6.8, 6.9)[3]. In the plots, the horizontal axis corresponds to the focus, the color of each line denotes the filter color, and the vertical axis corresponds to the intrinsic utility of the combination. The selected combinations are marked by circles.

6.5.3 Results and Conclusions

The experiment results for wafer A are presented in Figures 6.8, 6.9, the results for the other wafer are similar. the blinkered scheme selected parameters with an almost optimal utility in all but one case, while the myopic scheme exhausted the budget or failed to select a next measurement before gathering sufficient evidence in about one third of the runs, selecting parameter combinations with a low utility.

Both the myopic and the blinkered scheme used up the measurement budget in most runs. Therefore, the blinkered scheme was advantageous in this problem not just because this scheme prevents premature termination, but also because better measurement locations are chosen and the evidence is gathered more efficiently.

[3]As before, the coordinates were slightly randomized to show multiple choices in the same location.

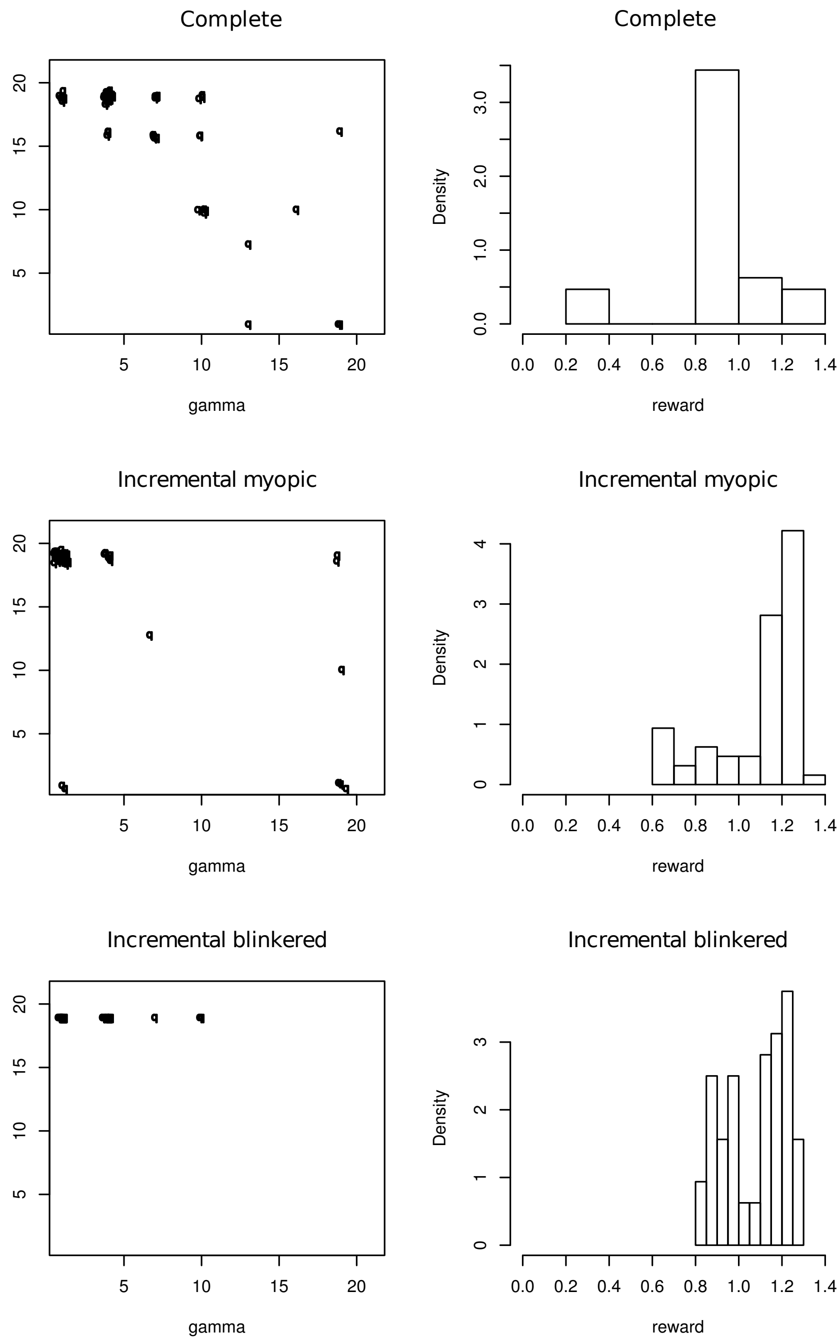

Figure 6.7: The SVMGUIDE2 data set.

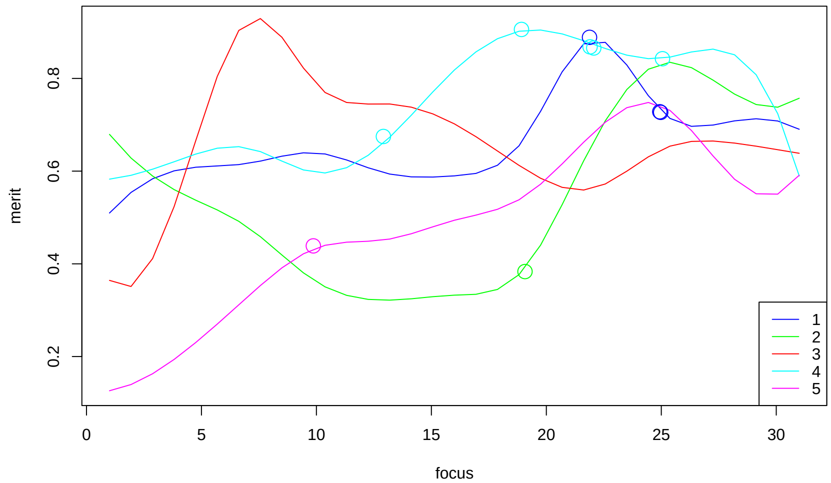

Figure 6.8: Myopic scheme.

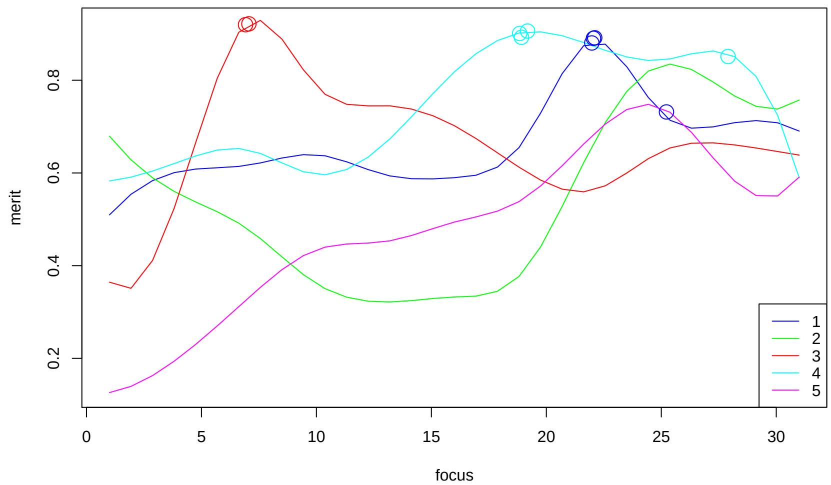

Figure 6.9: Blinkered scheme.

Chapter 7

Related Work

7.1 Limited rationality

Limited rationality, a model of deliberation based on value of utility revision and deliberation time cost was introduced in [22]. Notions of value of computation and its estimate were defined, as well as the class of meta-greedy algorithms and simplifying assumptions under which the algorithms are applicable. The theory of bounded optimality, on which the approach is based, is further developed in [23]. [30] employs limited rationality techniques to analyze any-time algorithms and proves optimality of myopic algorithm monitoring under assumptions about the class of value and time cost functions.

7.2 Greedy myopic algorithm

[16] consider a greedy myopic algorithm for observation selection based on value of observation. The paper shows that when values of measurements for different items are mutually submodular and the measurement cost is fixed, the algorithm is nearly optimal. The authors describe applications of the algorithm to various problems, such as fault diagnosis, robotic explorations, minimizing human attention and others. However, none of the problems is a problem of single-choice optimization. When a single item must be selected, as opposed to optimizing an aggregate function on a set, the utility function is usually not submodular; thus, there are no established performance guarantees of the greedy myopic algorithm in this case.

7.3 Exact measurements

Several researchers consider non-myopic on-line selection of measurements in the cases where the submodularity restriction does not hold. The proposed approaches are better than the greedy myopic algorithm, but are restricted to cases of exact measurements. Besides, the proposed improved algorithms often rely on a particular structure of dependencies between measurement outcomes, such as dependencies represented by trees or sparse acyclic graphs, and are evaluated on cases with a small number (less than a hundred) of different measurements available. In problems of optimization under uncertainty of continuous functions, thousands of different measurements can be required to achieve sufficient accuracy, and the dependency structure can be rather tight.

In [9], a case of discrete Bayesian networks with a single decision node is analyzed. The authors propose to consider subsequences of observations in the descending order of their myopic value estimates. If any such subsequence has non-negative value estimate, then the computation with the greatest myopic estimate is chosen. However, this approach always chooses a measurement for the myopically best item, and when applied to the selection problem either looks at sequences of measurements on a single item with the greatest myopic value estimate, or, if sequences with one measurement per item are considered, may fail to provide an improvement over the myopic estimate even for simple cases. Still, in may cases their scheme shows an improvement in performance. [18] describes and experimentally analyzes an algorithm for influence diagrams based on a non-myopic VOI estimate.

[2] addresses the problem of efficient feature value acquisition in the presence of varying acquisition costs. A new structure, *value of information lattice*, and an algorithm to build the structure are introduced in the paper. The paper compares different acquisition strategies: greedy acquisition, acquisition in sets and a mixed strategy, according to which the most valuable feature in the most valuable set is selected and shows how the mixed strategy can be implemented efficiently using the introduced structure. However, the method is still intractable in the general case, the features are assumed to be discrete, and the number of features to be relatively small (the largest set in the empirical evaluation consisted of 20 features).

7.4 Other research directions

Multi-armed bandits [26] bear similarity to the measurement selection problem, in particular, when the reward distribution is continuous and unknown. The Multi-armed bandit is a well known problem based on the analogy with traditional slot machines, but with more than one lever. When pulled, each lever provides a reward drawn from a distribution associated with that specific lever. The objective of the gambler is to maximize the collected reward sum through iterative pulls. Some of the algorithms, e.g. POKER (Price of Knowledge and Estimated Reward) [26] employ the notion of value of information. However, the multi-armed bandit problem is different from the single-choice optimization problem in that the reward accumulates in the process of the search, and a compromise must be found between exploration (looking for the best lever) and exploitation (pulling the lever that brings the highest benefit). Additionally, most solutions concentrate on particular features of the value function, such as linear dependence of reward from pushing a lever on the time left, and do not facilitate generalization. On the other hand, achievements in limited rationality techniques should be helpful in development of improved solutions in this domain.

[7] explores a methodology for selective sampling of non-stationary Gaussian processes. The authors show that by using the Bayesian approach to measure predictive uncertainty and guide sampling, an accurate model of the process can be discovered at a lower cost. The objective of the method is different though: to simulate a complex process over the whole field, rather than to optimize a combination of parameters. In the problem of optimization of uncertainty, an accurate model of areas with low utility is unnecessary, and measurements which do not affect the selection of the best combination of parameters are wasted.

Chapter 8

Summary

8.1 Conclusions

The semi-myopic scheme introduced in this work extended applicability of the greedy optimization algorithm to problems with recurring inexact evaluations of the function being optimized. A particular kind of the myopic scheme, the blinkered scheme, and the correspoding blinkered value of information estimate were shown to cope well with problems for which the myopic scheme fails to maintain robustness. Theoretical bounds were proved for important boundary cases, thus providing certain guarantees that the blinkered scheme selects evaluation parameters rationally and does not stop too early. As follows from both the theoretical analysis and the empirical evaluation, advantage of the blinkered estimate is most pronounced when a change in evaluation parameters decreases the final reward ("the movement has a cost"), or when the utility function is S-shaped.

The blinkered scheme can be implemented efficiently with only a small overhead. A reference implementation of the blinkered scheme showed good performance and consistent robustness on a number of challenging optimization problems. Careful design of data structures and deliberation logic resulted in an optimization algorithm capable of solving optimization problems with multi-argument utility functions, hundreds of nodes in the grid, and thousands of possible measurements at each point in time.

8.2 Future Research

Influence of dependencies on the properties of the blinkered estimate has been investigated only partially In particular, when, due to sufficiently strong dependencies, observations in different locations are not mutually subadditive, the blinkered estimate alone may not prevent premature termination of the measurement plan, and its combination with the approach proposed in [9] may be worthwhile.

During the algorithm analysis, several assumptions have been made about the shape of utility functions and belief distributions. Certain special cases, such as normally distributed beliefs and s-shaped utility functions, are frequently met in applications and may lead to stronger bounds and discovery of additional features of semi-myopic schemes.

An efficient implementation of the optimization algorithm demands ubiquitous reuse of computations. While some ideas were formulated and implemented, a lot of work remains to be done both in theoretical analysis and in empirical evaluation. The work should result in an algorithm that scales to large problems and can be deployed in real-time environments.

The myopic algorithm acts as though the time is unlimited, and takes the budget constraint into account only as a termination condition. This behavior is reasonable when the utility and the costs are such that in most runs the algorithm stops because it cannot find a measurement with a positive value of information. However, when the budget is tight, an algorithm that takes into account measurement value per cost unit is worth consideration.

Acknowledgments

The research is partially supported by the IMG4 consortium under the MAGNET program, funded by the Israel Ministry of Trade and Industry, by the Israel Science Foundation, and by the Lynne and William Frankel Center for Computer Sciences.

Appendix A

Proofs of Theorems

A.1 Proof of Theorem 1

Proof. The intrinsic value of information over the remaining budget when the blinkered scheme terminates Λ_b is at most the remaining budget C_b, since otherwise the scheme would not have terminated. There is only one kind of measurement, thus the intrinsic value of information Λ_o achieved by an optimal policy is at most equal to that of all measurements, $\Lambda_o \leq \Lambda_b \leq C_b$. Since measurement costs are positive, the value of information V_o of an optimal policy must therefore also be at most C_b. □

A.2 Proof of Theorem 2

Proof. Since the measurement cost is 0, the net value of information is the same as the intrinsic value, $V = \Lambda$. Value of information cannot decrease due to additional measurements, therefore the value of any sequence of measurements containing all of the measurements in an optimal sequence is at least as high as the value V_o of an optimal policy. Consider an exhaustive sequence containing N_q measurements of each of the N_s items, $N_s \cdot N_q$ measurements total, with value V_e. The exhaustive sequence contains all measurements made according to an optimal policy within the budget, thus $V_o \leq V_e$.

Let s_i be the item with the highest blinkered value for N_q measurements, denote its value by $V_{b\,i_{max}} = \max_i V_{b\,i}$. Since measurements of different items are mutually subadditive, $V_e \leq N_s V_{b\,i_{max}}$, and thus $V_{b\,i_{max}} \geq \frac{V_o}{N}_s$.

The blinkered scheme selects at every step a measurement from a sequence with value of information which is at least as high as the value of measurements of s_i for the rest of the budget. Thus, its value of information $V_b \geq V_{b\,i_{max}} \geq \frac{V_o}{N}_s$. □

Appendix B

Tightness of Bounds

B.1 Tightness of the Bound in Theorem 1

The bound is asymptotically tight. Consider the following problem instance:

- the utility of s_1 is exactly known an equal 0;
- according to the prior belief, s_2 can have utility $-\frac{h}{C_b}p$, 0 or $\frac{C_b}{p}$ with equal probability, for some $0 \le p \le 1$;
- the measurement cost function is defined as follows[1]:

$$\begin{aligned} c(x_0, x_2) &= 0 \\ c(x_2, x_2) &= C_b, \end{aligned}$$

  that is, at most two measurements can be made, the first measurement is free, the cost of the second measurement is C_b;
- the first measurement of s_2 either discovers with probability p that the utility of s_2 is $-\frac{C_b}{p}$ or $\frac{C_b}{p}$, or, with probability $1-p$ that the utility is exactly 0;
- the second measurement of s_2 determines the exact utility of the item.

The blinkered value of information estimate is 0: the intrinsic value of information of the first measurement is 0, and of the first and the second measurement is $\frac{C_b}{p}p = C_b$ — the utility of s_2 is always determined exactly, and is $\pm\frac{C_b}{p}$ with probability p and 0 with probability $1-p$. Thus, the blinkered scheme will not perform any measurements.

The optimal policy is to

1. perform the first measurement,
2. if the utility of s_2 is not zero, perform the second measurement.

[1] The second measurement is more expensive than the first one when, for example, in a physical experiment one has to wait in order to exclude inference between the measurements.

The probability of the second measurement is p and the expected cost of the policy is pC_b. The value of information of the optimal policy is thus:

$$V_o = p\frac{C_b}{p} - pC_b = (1-p)C_b$$

and approaches C_b when p approaches 0:

$$\lim_{p\to 0} V_o = C_b$$

The latter limit shows that the bound stated in Theorem 1 is asymptotically tight.

B.2 Tightness of the Bound in Theorem 2

The bound is asymptotically tight. Construct a problem instance as follows: M items, with value of information of jth item for i measurements $v_j(i)$:

$$v_1(i) = \sqrt[k]{\frac{i}{N}}$$

$$v_{j>1}(i) = \begin{cases} 0 & \text{if } i < \frac{N}{M} \\ \sqrt[k]{\frac{1}{M}} & \text{otherwise} \end{cases}$$

Value of information of a combination of the measurements is the sum of the values for each item:

$$v(i) = \sum_{j=1}^{M} v_j(i_j)$$

Here, the optimal policy is to measure each item $\frac{N}{M}$ times. The resulting value of information for N measurements is:

$$V_o = v_1\left(\frac{N}{M}\right) + (M-1)v_{j>1}\left(\frac{N}{M}\right) = M\sqrt[k]{\frac{1}{M}}$$

$$\lim_{k\to\infty} M\sqrt[k]{\frac{1}{M}} = M$$

But the blinkered scheme will always choose the first item with

$$V_b = v_1(N) = 1.$$

B.2.1 Comparative Example

In the following example the blinkered scheme selects better measurements than the myopic one. Consider a modification of the example in Section **??**, with an additional item s_3 and free measurements:

- S is a set of three items, s_1, s_2, and s_3;
- the value of s_1 is known exactly, $x_1 = 1$;
- the prior belief about s_2 is nomal $p_2(x) = \mathcal{N}(x; 0, 1)$;
- the observation variance of s_2 is $\sigma_{e2}^2 = 5$;
- the prior belief about s_3 is normal: $p_3(x) = \mathcal{N}(x; 0.0, 0.16)$;
- a measurement of the item returns the exact value: $\sigma_{e3}^2 = 0.0$;
- measurements are free: $c = 0$;
- the budget is two measurements;
- the utility is a step function:

$$u(x) = \begin{cases} 0 & \text{if } x < 1 \\ 0.5 & \text{if } x = 1 \\ 1 & \text{if } x > 1 \end{cases}$$

The myopic scheme measures s_3, and then s_2, since $MVI_3 \approx 0.0024 > MVI_2 \approx 0.0004$. However, the blinkered scheme first measures s_2, since $MVI_2^2 \approx 0.0028 > MVI_3^2 \approx 0.0024$, and then chooses between s_2 and s_3 depending on the obtained observation. In particular, if the posterior mean of s_2 is approximately in the range $[0.39, 1.6]$, s_2 is chosen, otherwise, s_3 is chosen.

It is easy to show that only these two scenarios are possible:

1. measure s_3, and then s_2;
2. measure s_2, and then either s_3 or s_2, depending on the value of each measurement;

and the latter policy is optimal. Thus, the blinkered scheme implements the optimal policy, while the myopic one misses cases where s_2 should be measured twice for the best expected result. Of course, one can construct an example where the blinkered scheme is far from optimal. Still, the empirical evaluations show that the blinkered scheme behaves significantly better than the myopic one, and gives good results in cases where the myopic scheme fails.

Bibliography

[1] David H. Ackley. *A connectionist machine for genetic hillclimbing.* Kluwer Academic Publishers, Norwell, MA, USA, 1987.

[2] Mustafa Bilgic and Lise Getoor. Voila: Efficient feature-value acquisition for classification. In *AAAI*, pages 1225–1230. AAAI Press, 2007.

[3] Christopher M. Bishop. *Pattern Recognition and Machine Learning.* Springer Science+Business Media, LLC, New York, NY, USA, 2007.

[4] C. C. Caramanis and S. Mannor. A bayesian approach to data-driven optimization under uncertainty. In *Allerton Conference on Communication, Control, and Computing*, 2007.

[5] Anthony R. Cassandra. *Exact and Approximate Algorithms for Partially Observable Markov Decision Processes.* PhD thesis, Brown University, Providence, RI, USA, 1998.

[6] Robert Cowell. Advanced inference in bayesian networks. In *Proceedings of the NATO Advanced Study Institute on Learning in graphical models*, pages 27–49, Norwell, MA, USA, 1998. Kluwer Academic Publishers.

[7] Robert B. Gramacy, Herbert K. H. Lee, and William G. Macready. Parameter space exploration with gaussian process trees. In *Proceedings of the International Conference on Machine L earning (pp. 353360). Omnipress & ACM Digital Library*, pages 353–360, 2004.

[8] Eric A. Hansen. Indefinite-horizon pomdps with action-based termination. In *AAAI*, pages 1237–1242, 2007.

[9] D. Heckerman, E. Horvitz, and B. Middleton. An approximate nonmyopic computation for value of information. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(3):292–298, 1993.

[10] J. J. Hull. A database for handwritten text recognition research. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(5):550–554, 1994.

[11] J. K. Johansson. Advertising and the s-curve: A new approach. *Journal of Marketing Research*, 16(3):346–354, August 1979.

[12] K. E. Johns and R. C. Fair. *Principles of Economics.* Prentice-Hall, 5th edition, 1999.

[13] Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134, 1998.

[14] Andreas Krause and Carlos Guestrin. Near-optimal nonmyopic value of information in graphical models. In *Proceedings of the 21th Annual Conference on Uncertainty in Artificial Intelligence (UAI-05)*, pages 324–33, Arlington, Virginia, 2005. AUAI Press.

[15] Andreas Krause and Carlos Guestrin. Optimal nonmyopic value of information in graphical models — efficient algorithms and theoretical limits. In *In Proc. of IJCAI*, pages 1339–1345, 2005.

[16] Andreas Krause and Carlos Guestrin. Near-optimal observation selection using submodular functions. In *AAAI*, pages 1650–1654, 2007.

[17] Andreas Krause, Jure Leskovec, Carlos Guestrin, Jeanne VanBriesen, and Christos Faloutsos. Efficient sensor placement optimization for securing large water distribution networks. *Journal of Water Resources Planning and Management*, 134(6):516–526, November 2008.

[18] Wenhui Liao and Qiang Ji. Efficient non-myopic value-of-information computation for influence diagrams. *Int. J. Approx. Reasoning*, 49(2):436–450, 2008.

[19] David J. C. MacKay. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, 2003. Available from http://www.inference.phy.cam.ac.uk/mackay/itila/.

[20] Yan Radovilsky and Solomon Eyal Shimony. Observation subset selection as local compilation of performance profiles. In *UAI*, pages 460–467, 2008.

[21] I. Rish, M. Brodie, Sheng Ma, N. Odintsova, A. Beygelzimer, G. Grabarnik, and K. Hernandez. Adaptive diagnosis in distributed systems. *Neural Networks, IEEE Transactions on*, 16(5):1088–1109, 2005.

[22] Stuart Russell and Eric Wefald. *Do the right thing: studies in limited rationality.* MIT Press, Cambridge, MA, USA, 1991.

[23] Stuart J. Russell and Devika Subramanian. Provably bounded-optimal agents. *Journal of Artificial Intelligence Research*, 2:575–609, 1995.

[24] Nikolaos V. Sahinidis. Optimization under uncertainty: State-of-the-art and opportunities. *Computers and Chemical Engineering*, 28:971–983, 2004.

[25] Guy Shani, Ronen I. Brafman, and Solomon Eyal Shimony. Prioritizing point-based pomdp solvers. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 38(6):1592–1605, 2008.

[26] Joannès Vermorel and Mehryar Mohri. Multi-armed bandit algorithms and empirical evaluation. In *ECML*, pages 437–448, 2005.

[27] Chih wei Hsu, Chih chung Chang, and Chih jen Lin. A practical guide to support vector classification. Technical report, 2003.

[28] Wikipedia. Himmelblau's function—Wikipedia, The Free Encyclopedia, 2009. [Online; accessed 9-July-2009].

[29] Alice X. Zheng, Irina Rish, and Alina Beygelzimer. Efficient test selection in active diagnosis via entropy approximation. In *UAI*, pages 675–682, 2005.

[30] Shlomo Zilberstein. *Operational Rationality through Compilation of Anytime Algorithms*. PhD thesis, University of California at Berkeley, Berkeley, CA, USA, 1993.

[31] Shlomo Zilberstein. Resource-bounded sensing and planning in autonomous systems. *Autonomous Robots*, 3:31–48, 1996.