Journal of Machine Learning Research 12 (2011) 2931-2974 Submitted 11/09; Revised 3/11; Published 10/11

Hierarchical Knowledge Gradient for Sequential Sampling

Martijn R.K. Mes M.R.K.MES@UTWENTE.NL
Department of Operational Methods for Production and Logistics
University of Twente
Enschede, The Netherlands

Warren B. Powell POWELL@PRINCETON.EDU
Department of Operations Research and Financial Engineering
Princeton University
Princeton, NJ 08544, USA

Peter I. Frazier PF98@CORNELL.EDU
Department of Operations Research and Information Engineering
Cornell University
Ithaca, NY 14853, USA

Editor: Ronald Parr

Abstract

We propose a sequential sampling policy for noisy discrete global optimization and ranking and selection, in which we aim to efficiently explore a finite set of alternatives before selecting an alternative as best when exploration stops. Each alternative may be characterized by a multi-dimensional vector of categorical and numerical attributes and has independent normal rewards. We use a Bayesian probability model for the unknown reward of each alternative and follow a fully sequential sampling policy called the knowledge-gradient policy. This policy myopically optimizes the expected increment in the value of sampling information in each time period. We propose a hierarchical aggregation technique that uses the common features shared by alternatives to learn about many alternatives from even a single measurement. This approach greatly reduces the measurement effort required, but it requires some prior knowledge on the smoothness of the function in the form of an aggregation function and computational issues limit the number of alternatives that can be easily considered to the thousands. We prove that our policy is consistent, finding a globally optimal alternative when given enough measurements, and show through simulations that it performs competitively with or significantly better than other policies.

Keywords: sequential experimental design, ranking and selection, adaptive learning, hierarchical statistics, Bayesian statistics

1. Introduction

We address the problem of maximizing an unknown function θ_x where $x = (x_1, \ldots, x_D)$, $x \in \mathcal{X}$, is a discrete multi-dimensional vector of categorical and numerical attributes. We have the ability to sequentially choose a set of measurements to estimate θ_x, after which we choose the value of x with the largest estimated value of θ_x. Our challenge is to design a measurement policy which produces fast convergence to the optimal solution, evaluated using the expected objective function after a specified number of iterations. Many applications in this setting involve measurements that are time consuming and/or expensive. This problem is equivalent to the ranking and selection (R&S)

problem, where the difference is that the number of alternatives $|\mathcal{X}|$ is extremely large relative to the measurement budget.

We do not make any explicit structural assumptions about θ_x, but we do assume that we are given an ordered set $\mathcal{G}$ and a family of aggregation functions $G^g : \mathcal{X} \to \mathcal{X}^g$, $g \in \mathcal{G}$, each of which maps $\mathcal{X}$ to a region $\mathcal{X}^g$, which is successively smaller than the original set of alternatives. After each observation $\hat{y}_x^n = \theta_x + \varepsilon^n$, we update a family of statistical estimates of θ at each level of aggregation. After n observations, we obtain a family of estimates $\mu_x^{g,n}$ of the function at different levels of aggregation, and we form an estimate μ_x^n of θ_x using

$$\mu_x^n = \sum_{g \in \mathcal{G}} w_x^{g,n} \mu_x^{g,n}, \tag{1}$$

where the weights $w_x^{g,n}$ sum to one over all the levels of aggregation for each point x. The estimates $\mu_x^{g,n}$ at more aggregate levels have lower statistical variance since they are based upon more observations, but exhibit aggregation bias. The estimates $\mu_x^{g,n}$ at more disaggregate levels will exhibit greater variance but lower bias. We design our weights to strike a balance between variance and bias.

Our goal is to create a measurement policy π that leads us to find the alternative x that maximizes θ_x. This problem arises in a wide range of problems in stochastic search including (i) which settings of several parameters of a simulated system has the largest mean performance, (ii) which combination of chemical compounds in a drug would be the most effective to fight a particular disease, and (iii) which set of features to include in a product would maximize profits. We also consider problems where x is a multi-dimensional set of continuous parameters.

A number of measurement policies have been proposed for the ranking and selection problem when the number of alternatives is not too large, and where our beliefs about the value of each alternative are independent. We build on the work of Frazier et al. (2009) which proposes a policy, the knowledge-gradient policy for correlated beliefs, that exploits correlations in the belief structure, but where these correlations are assumed known.

This paper makes the following contributions. First, we propose a version of the knowledge gradient policy that exploits aggregation structure and similarity between alternatives, without requiring that we specify an explicit covariance matrix for our belief. Instead, we develop a belief structure based on the weighted estimates given in (1). We estimate the weights using a Bayesian model adapted from frequentist estimates proposed by George et al. (2008). In addition to eliminating the difficulty of specifying an a priori covariance matrix, this avoids the computational challenge of working with large covariance matrices. Second, we show that a learning policy based on this method is optimal in the limit, that is, eventually it always discovers the best alternative. Our method requires that a family of aggregation functions be provided, but otherwise does not make any specific assumptions about the structure of the function or set of alternatives.

The remainder of this paper is structured as follows. In Section 2 we give a brief overview of the relevant literature. In Section 3, we present our model, the aggregation techniques we use, and the Bayesian updating approach. We present our measurement policy in Section 4 and a proof of convergence of this policy in Section 5. We present numerical experiments in Section 6 and 7. We close with conclusions, remarks on generalizations, and directions for further research in Section 8.

2. Literature

There is by now a substantial literature on the general problem of finding the maximum of an unknown function where we depend on noisy measurements to guide our search. Spall (2003) provides a thorough review of the literature that traces its roots to stochastic approximation methods first introduced by Robbins and Monro (1951). This literature considers problems with vector-valued decisions, but its techniques require many measurements to find maxima precisely, which is a problem when measurements are expensive.

Our problem originates from the ranking and selection (R&S) literature, which begins with Bechhofer (1954). In the R&S problem, we have a collection of alternatives whose value we can learn through sampling, and from which we would like to select the one with the largest value. This problem has been studied extensively since its origin, with much of this work reviewed by Bechhofer et al. (1995), more recent work reviewed in Kim and Nelson (2006), and research continuing actively today. The R&S problem has also been recently and independently considered within computer science (Even-Dar et al., 2002; Madani et al., 2004; Bubeck et al., 2009b).

There is also a related literature on online learning and multi-armed bandits, in which an algorithm is faced with a collection of noisy options of unknown value, and has the opportunity to engage these options sequentially. In the online learning literature, an algorithm is measured according to the *cumulative* value of the options engaged, while in the problem that we consider an algorithm is measured according to its ability to select the best at the end of experimentation. Rather than value, researchers often consider the regret, which is the loss compared to optimal sequence of decisions in hindsight. Cumulative value or regret is appropriate in settings such as dynamic pricing of a good sold online (learning while doing), while terminal value or regret is appropriate in settings such as optimizing a transportation network in simulation before building it in the real world (learn then do). Strong theoretical bounds on cumulative and average regret have been developed in the online setting (see, e.g., Auer et al., 2002; Flaxman et al., 2005; Abernethy et al., 2008).

General-purpose online-to-batch conversion techniques have been developed, starting with Littlestone (1989), for transforming online-learning methods with bounds on cumulative regret into methods with bounds on terminal regret (for a summary and literature review see Shalev-Shwartz, 2007, Appendix B). While these techniques are easy to apply and immediately produce methods with theoretical bounds on the rate at which terminal regret converges to zero, methods created in this way may not have the best achievable bounds on terminal regret: Bubeck et al. (2009b) shows that improving the upper bound on the cumulative regret of an online learning method causes a corresponding lower bound on the terminal regret to get worse. This is indicative of a larger difference between what is required in the two types of problems. Furthermore, as as example of the difference between cumulative and terminal performance, Bubeck et al. (2009b) notes that with finitely many unrelated arms, achieving optimal cumulative regret requires sampling suboptimal arms no more than a logarithmic number of times, while achieving optimal terminal regret requires sampling every arm a linear number of times.

Despite the difference between cumulative and terminal value, a number of methods have been developed that are often applied to both online learning and R&S problems in practice, as well as to more complex problems in reinforcement learning and Markov decision processes. These heuristics include Boltzmann exploration, interval estimation, upper confidence bound policies, and hybrid exploration-exploitation policies such as epsilon-greedy. See Powell and Frazier (2008) for

a review of these. Other policies include the Explicit Explore or Exploit (E^3) algorithm of Kearns and Singh (2002) and R-MAX of Brafman and Tennenholtz (2003).

Researchers from the online learning and multi-armed bandit communities have also directly considered R&S and other related problems in which one is concerned with terminal rather than cumulative value (Even-Dar et al., 2002, 2003; Madani et al., 2004; Mnih et al., 2008; Bubeck et al., 2009b). Most work that directly considers terminal value assumes no a-priori relationship between alternatives. One exception is Srinivas et al. (2010), which considers a problem with a Gaussian process prior on the alternatives, and uses a standard online-to-batch conversion to obtain bounds on terminal regret. We are aware of no work in the online learning community, however, whether considering cumulative value or terminal value, that considers the type of hierarchical aggregation structures that we consider here. A number of researchers have considered other types of dependence between alternatives, such as online convex and linear optimization (Flaxman et al., 2005; Kleinberg, 2005; Abernethy et al., 2008; Bartlett et al., 2008), general metric spaces with a Lipschitz or locally-Lipschitz condition (Kleinberg et al., 2008; Bubeck et al., 2009a), and Gaussian process priors (Grünewälder et al., 2010; Srinivas et al., 2010).

A related line of research has focused on finding the alternative which, if measured, will have the greatest impact on the final solution. This idea was originally introduced in Mockus (1975) for a one-dimensional continuous domain with a Wiener process prior, and in Gupta and Miescke (1996) in the context of the independent normal R&S problem as also considered in this paper. The latter policy was further analyzed in Frazier et al. (2008) under the name knowledge-gradient (KG) policy, where it was shown that the policy is myopically optimal (by construction) and asymptotically optimal. An extension of the KG policy when the variance is unknown is presented in Chick et al. (2010) under the name $\mathcal{LL}_1$, referring to the one-step linear loss, an alternative name when we are minimizing expected opportunity cost. A closely related idea is given in Chick and Inoue (2001) where samples are allocated to maximize an approximation to the expected value of information. Related search methods have also been developed within the simulation-optimization community, which faces the problem of determining the best of a set of parameters, where evaluating a set of parameters involves running what is often an expensive simulation. One class of methods evolved under the name optimal computing budget allocation (OCBA) (Chen et al., 1996; He et al., 2007).

The work in ranking and selection using ideas of expected incremental value is similar to work on Bayesian global optimization of continuous functions. In Bayesian global optimization, one would place a Bayesian prior belief on the unknown function θ. Generally the assumption is that unknown function θ is a realization from a Gaussian process. Wiener process priors, a special case of the Gaussian process prior, were common in early work on Bayesian global optimization, being used by techniques introduced in Kushner (1964) and Mockus (1975). Surveys of Bayesian global optimization may be found in Sasena (2002); Lizotte (2008) and Brochu et al. (2009).

While algorithms for Bayesian global optimization usually assume noise-free function evaluations (e.g., the EGO algorithm of Jones et al., 1998), some algorithms allow measurement noise (Huang et al., 2006; Frazier et al., 2009; Villemonteix et al., 2009). We compare the performance of HKG against two of these: Sequential Kriging Optimization (SKO) from Huang et al. (2006) and the knowledge-gradient policy for correlated normal beliefs (KGCB) from Frazier et al. (2009). The latter policy is an extension of the knowledge-gradient algorithm in the presence of correlated beliefs, where measuring one alternative updates our belief about other alternatives. This method was shown to significantly outperform methods which ignore this covariance structure, but the algorithm requires the covariance matrix to be known. The policies SKO and KGCB are further explained in

Section 6. Like the consistency results that we provide for HKG, consistency results are known for some algorithms: consistency of EGO is shown in Vazquez and Bect (2010), and lower bounds on the convergence rate of an algorithm called GP-UCB are shown in Srinivas et al. (2010).

An approach that is common in optimization of continuous functions, and which accounts for dependencies, is to fit a continuous function through the observations. In the area of Bayesian global optimization, this is usually done using Gaussian process priors. In other approaches, like the Response Surface Methodology (RSM) (Barton and Meckesheimer, 2006) one normally would fit a linear regression model or polynomials. An exception can be found in Brochu et al. (2009) where an algorithm is presented that uses random forests instead, which is reminiscent of the hierarchical prior that we employ in this paper. When we are dealing with nominal categorical dimensions, fitting a continuous function is less appropriate as we will show in this paper. Moreover, the presence of categorical dimensions might give a good indication for the aggregation function to be used. The inclusion of categorical variables in Bayesian global optimization methods, via both random forests and Gaussian processes, as well as a performance comparison between these two, is addressed in Hutter (2009).

There is a separate literature on aggregation and the use of mixtures of estimates. Aggregation, of course, has a long history as a method of simplifying models (see Rogers et al., 1991). Bertsekas and Castanon (1989) describes adaptive aggregation techniques in the context of dynamic programming, while Bertsekas and Tsitsiklis (1996) provides a good presentation of state aggregation methods used in value iteration. In the machine learning community, there is an extensive literature on the use of weighted mixtures of estimates, which is the approach that we use. We refer the reader to LeBlanc and Tibshirani (1996); Yang (2001) and Hastie et al. (2001). In our work, we use a particular weighting scheme proposed by George et al. (2008) due to its ability to easily handle state dependent weights, which typically involves estimation of many thousands of weights since we have a weight for each alternative at each level of aggregation.

3. Model

We consider a finite set $\mathcal{X}$ of distinct alternatives where each alternative $x \in \mathcal{X}$ might be a multi-dimensional vector $x = (x_1, \ldots, x_D)$. Each alternative $x \in \mathcal{X}$ is characterized by an independent normal sampling distribution with unknown mean θ_x and known variance $\lambda_x > 0$. We use M to denote the number of alternatives $|\mathcal{X}|$ and use θ to denote the column vector consisting of all θ_x, $x \in \mathcal{X}$.

Consider a sequence of N sampling decisions, $x^0, x^1, \ldots, x^{N-1}$. The sampling decision x^n selects an alternative to sample at time n from the set $\mathcal{X}$. The sampling error $\varepsilon_x^{n+1} \sim \mathcal{N}(0, \lambda_x)$ is independent conditioned on $x^n = x$, and the resulting sample observation is $\hat{y}_x^{n+1} = \theta_x + \varepsilon_x^{n+1}$. Conditioned on θ and $x^n = x$, the sample has conditional distribution

$$\hat{y}_x^{n+1} \sim \mathcal{N}(\theta_x, \lambda_x).$$

Because decisions are made sequentially, x^n is only allowed to depend on the outcomes of the sampling decisions $x^0, x^1, \ldots, x^{n-1}$. In the remainder of this paper, a random variable indexed by n means it is measurable with respect to $\mathcal{F}^n$ which is the sigma-algebra generated by $x^0, \hat{y}^1_{x^0}, x^1, \ldots, x^{n-1}, \hat{y}^n_{x^{n-1}}$.

In this paper, we derive a method based on Bayesian principles which offers a way of formalizing a priori beliefs and of combining them with the available observations to perform statistical

inference. In this Bayesian approach we begin with a prior distribution on the unknown values θ_x, $x \in \mathcal{X}$, and then use Bayes' rule to recursively to derive the posterior distribution at time $n+1$ from the posterior at time n and the observed data. Let μ^n be our estimate of θ after n measurements. This estimate will either be the Bayes estimate, which is the posterior mean $\mathbb{E}[\theta \mid \mathcal{F}^n]$, or an approximation to this posterior mean as we will use later on. Later, in Sections 3.1 and 3.2, we describe the specific prior and posterior that we use in greater detail. Under most sampling models and prior distributions, including the one we treat here, we may intuitively understand the learning that occurs from sampling as progressive concentration of the posterior distribution on θ, and as the tendency of μ^n, the mean of this posterior distribution, to move toward θ as n increases.

After taking N measurements, we make an implementation decision, which we assume is given by the alternative x^N that has the highest expected reward under the posterior, that is, $x^N \in \arg\max_{x \in \mathcal{X}} \mu_x^N$. Although we could consider policies making implementation decisions in other ways, this implementation decision is optimal when μ^N is the exact posterior mean and when performance is evaluated by the expected value under the prior of the true value of the implemented alternative. Our goal is to choose a sampling policy that maximizes the expected value of the implementation decision x^N. Therefore we define Π to be the set of sampling policies that satisfies the requirement $x^n \in \mathcal{F}^n$ and introduce $\pi \in \Pi$ as a policy that produces a sequence of decisions $\left(x^0, \ldots, x^{N-1}\right)$. We further write $\mathbb{E}^\pi$ to indicate the expectation with respect to the prior over both the noisy outcomes and the truth θ when the sampling policy is fixed to π. Our objective function can now be written as

$$\sup_{\pi \in \Pi} \mathbb{E}^\pi \left[\max_{x \in \mathcal{X}} \mathbb{E}[\theta_x \mid \mathcal{F}^N] \right].$$

If μ^N is the exact posterior mean, rather than an approximation, this can be written as

$$\sup_{\pi \in \Pi} \mathbb{E}^\pi \left[\max_{x \in \mathcal{X}} \mu_x^N \right].$$

As an aid to the reader, the notation defined throughout the next subsections is summarized in Table 1.

3.1 Model Specification

In this section we describe our statistical model, beginning first by describing the aggregation structure upon which it relies, and then describing our Bayesian prior on the sampling means θ_x. Later, in Section 3.2, we describe the Bayesian inference procedure. Throughout these sections we make the following assumptions: (i) we assume independent beliefs across different levels of aggregation and (ii) we have two quantities which we assume are fixed parameters of our model whereas we estimate them using the empirical Bayes approach. Even though these are serious approximations, we show that posterior inference from the prior results in the same estimators as presented in George et al. (2008) derived using frequestist methods.

Aggregation is performed using a set of aggregation functions $G^g : \mathcal{X} \to \mathcal{X}^g$, where $\mathcal{X}^g$ represents the g^{th} level of aggregation of the original set $\mathcal{X}$. We denote the set of all aggregation levels by $\mathcal{G} = \{0, 1, \ldots, G\}$, with $g = 0$ being the lowest aggregation level (which might be the finest discretization of a continuous set of alternatives), $g = G$ being the highest aggregation level, and $G = |\mathcal{G}| - 1$.

The aggregation functions G^g are typically problem specific and involve a certain amount of domain knowledge, but it is possible to define generic forms of aggregation. For example, numeric

Variable	Description
G	highest aggregation level
$G^g(x)$	aggregated alternative of alternative x at level g
$\mathcal{G}$	set of all aggregation levels
$\mathcal{G}(x,x')$	set of aggregation levels that alternatives x and x' have in common
$\mathcal{X}$	set of alternatives
$\mathcal{X}^g$	set of aggregated alternatives $G^g(x)$ at the g^{th} aggregation level
$\mathcal{X}^g(x)$	set of alternatives sharing aggregated alternative $G^g(x)$ at aggregation level g
N	maximum number of measurements
M	number of alternatives, that is, $M = \|\mathcal{X}\|$
θ_x	unknown true sampling mean of alternative x
θ_x^g	unknown true sampling mean of aggregated alternative $G^g(x)$
λ_x	measurement variance of alternative x
x^n	n^{th} measurement decision
$\hat{y}_x^n$	n^{th} sample observation of alternative x
ε_x^n	measurement error of the sample observation $\hat{y}_x^n$
μ_x^n	estimate of θ_x after n measurements
$\mu_x^{g,n}$	estimate of aggregated alternative $G^g(x)$ on aggregation level g after n measurements
$w_x^{g,n}$	contribution (weight) of the aggregate estimate $\mu_x^{g,n}$ to the overall estimate μ_x^n of θ_x
$m_x^{g,n}$	number of measurements from the aggregated alternative $G^g(x)$
β_x^n	precision of μ_x^n, with $\beta_x^n = 1/(\sigma_x^n)^2$,
$\beta_x^{g,n}$	precision of $\mu_x^{g,n}$, with $\beta_x^{g,n} = 1/(\sigma_x^{g,n})^2$
$\beta_x^{g,n,\varepsilon}$	measurement precision from alternatives $x' \in \mathcal{X}^g(x)$, with $\beta_x^{g,n,\varepsilon} = 1/(\sigma_x^{g,n,\varepsilon})^2$
$\delta_x^{g,n}$	estimate of the aggregation bias
$\tilde{g}_x^n$	lowest level g for which $m_x^{g,n} > 0$.
$\nu_x^{g,n}$	variance of $\theta_x^g - \theta_x$
$\underline{\delta}$	lower bound on $\delta_x^{g,n}$

Table 1: Notation used in this paper.

data can be defined over a range, allowing us to define a series of aggregations which divide this range by a factor of two at each additional level of aggregation. For vector valued data, we can aggregate by simply ignoring dimensions, although it helps if we are told in advance which dimensions are likely to be the most important.

Using aggregation, we create a sequence of sets $\{\mathcal{X}^g, g = 0,1,\ldots,G\}$, where each set has fewer alternatives than the previous set, and where $\mathcal{X}^0$ equals the original set $\mathcal{X}$. We introduce the following notation and illustrate its value using the example of Figure 1:

$\mathcal{G}(x,x')$ Set of all aggregation levels that the alternatives x and x' have in common, with $\mathcal{G}(x,x') \subseteq \mathcal{G}$. In the example we have $\mathcal{G}(2,3) = \{1,2\}$.

$\mathcal{X}^g(x)$ Set of all alternatives that share the same aggregated alternative $G^g(x)$ at the g^{th} aggregation level, with $\mathcal{X}^g(x) \subseteq \mathcal{X}$. In the example we have $\mathcal{X}^1(4) = \{4,5,6\}$.

$g=2$	13								
$g=1$	10			11			12		
$g=0$	1	2	3	4	5	6	7	8	9

Figure 1: Example with nine alternatives and three aggregation levels.

Given this aggregation structure, we now define our Bayesian model. Define latent variables θ_x^g, where $g \in \mathcal{G}$ and $x \in \mathcal{X}$. These variables satisfy $\theta_x^g = \theta_{x'}^g$ when $G^g(x) = G^g(x')$. Also, $\theta_x^0 = \theta_x$ for all $x \in \mathcal{X}$. We have a belief about these θ_x^g, and the posterior mean of the belief about θ_x^g is $\mu_x^{g,n}$. We see that, roughly speaking, θ_x^g is the best estimate of θ_x that we can make from aggregation level g, given perfect knowledge of this aggregation level, and that $\mu_x^{g,n}$ may be understood to be an estimator of the value of θ_x^g for a particular alternative x at a particular aggregation level g.

We begin with a normal prior on θ_x that is independent across different values of x, given by

$$\theta_x \sim \mathcal{N}(\mu_x^0, (\sigma_x^0)^2).$$

The way in which θ_x^g relates to θ_x is formalized by the probabilistic model

$$\theta_x^g \sim \mathcal{N}(\theta_x, \nu_x^g),$$

where ν_x^g is the variance of $\theta_x^g - \theta_x$ under our prior belief.

The values $\theta_x^g - \theta_x$ are independent across different values of g, and between values of x that differ at aggregation level g, that is, that have different values of $\mathcal{G}^g(x)$. The value ν_x^g is currently a fixed parameter of the model. In practice this parameter is unknown, and while we could place a prior on it (e.g., inverse gamma), we later employ an empirical Bayes approach instead, first estimating it from data and then using the estimated value as if it were given a priori.

When we measure alternative $x^n = x$ at time n, we observe a value $\hat{y}_x^{n+1}$. In reality, this observation has distribution $\mathcal{N}(\theta_x, \lambda_x)$. But in our model, we make the following approximation. We suppose that we observe a value $\hat{y}_x^{g,n+1}$ for each aggregation level $g \in \mathcal{G}$. These values are independent and satisfy

$$\hat{y}_x^{g,n+1} \sim \mathcal{N}(\theta_x^g, 1/\beta_x^{g,n,\varepsilon}),$$

where again $\beta_x^{g,n,\varepsilon}$ is, for the moment, a fixed known parameter, but later will be estimated from data and used as if it were known a priori. In practice we set $\hat{y}_x^{g,n+1} = \hat{y}_x^{n+1}$. It is only a modeling assumption that breaks this equality and assumes independence in its place. This approximation allows us to recover the estimators derived using other techniques in George et al. (2008).

This probabilistic model for $\hat{y}_x^{g,n+1}$ in terms of θ_x^g induces a posterior on θ_x^g, whose calculation is discussed in the next section. This model is summarized in Figure 2.

3.2 Bayesian Inference

We now derive expressions for the posterior belief on the quantities of interest within the model. We begin by deriving an expression for the posterior belief on θ_x^g for a given g.

We define $\mu_x^{g,n}$, $(\sigma_x^{g,n})^2$, and $\beta_x^{g,n} = (\sigma_x^{g,n})^{-2}$ to be the mean, variance, and precision of the belief that we would have about θ_x^g if we had a noninformative prior on θ_x^g and then observed $\hat{y}_{x^{m-1}}^{g,m}$ for *only* those $m < n$ satisfying $G^g(x^m) = G^g(x)$ and *only* for the given value of g. These are the observations from level g pertinent to alternative x. The quantities $\mu_x^{g,n}$ and $\beta_x^{g,n}$ can be obtained recursively

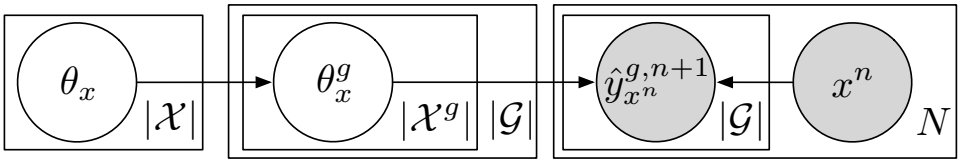

Figure 2: Probabilistic graphical model used by HKG. The dependence of x^n upon the past induced because HKG chooses its measurements adaptively is not pictured.

by considering two cases. When $G^g(x^n) \neq G^g(x)$, we let $\mu_x^{g,n+1} = \mu_x^{g,n}$ and $\beta_x^{g,n+1} = \beta_x^{g,n}$. When $G^g(x^n) = G^g(x)$ we let

$$\mu_x^{g,n+1} = \left[\beta_x^{g,n}\mu_x^{g,n} + \beta_x^{g,n,\varepsilon}\hat{y}_x^{n+1}\right]/\beta_x^{g,n+1}, \tag{2}$$

$$\beta_x^{g,n+1} = \beta_x^{g,n} + \beta_x^{g,n,\varepsilon}, \tag{3}$$

where $\beta_x^{g,0} = 0$ and $\mu_x^{g,0} = 0$.

Using these quantities, we may obtain an expression for the posterior belief on θ_x. We define μ_x^n, $(\sigma_x^n)^2$ and $\beta_x^n = (\sigma_x^n)^{-2}$ to be the mean, variance, and precision of this posterior belief. By Proposition 4 (Appendix B), the posterior mean and precision are

$$\mu_x^n = \frac{1}{\beta_x^n}\left[\beta_x^0\mu_x^0 + \sum_{g\in\mathcal{G}}\left((\sigma_x^{g,n})^2 + \nu_x^g\right)^{-1}\mu_x^{g,n}\right], \tag{4}$$

$$\beta_x^n = \beta_x^0 + \sum_{g\in\mathcal{G}}\left((\sigma_x^{g,n})^2 + \nu_x^g\right)^{-1}. \tag{5}$$

We generally work with a noninformative prior on θ_x in which $\beta_x^0 = 0$. In this case, the posterior variance is given by

$$(\sigma_x^n)^2 = \left(\sum_{g\in\mathcal{G}}\left((\sigma_x^{g,n})^2 + \nu_x^g\right)^{-1}\right)^{-1}, \tag{6}$$

and the posterior mean μ_x^n is given by the weighted linear combination

$$\mu_x^n = \sum_{g\in\mathcal{G}} w_x^{g,n}\mu_x^{g,n}, \tag{7}$$

where the weights $w_x^{g,n}$ are

$$w_x^{g,n} = \left((\sigma_x^{g,n})^2 + \nu_x^g\right)^{-1}\left(\sum_{g'\in\mathcal{G}}\left(\left(\sigma_x^{g',n}\right)^2 + \nu_x^{g'}\right)^{-1}\right)^{-1}. \tag{8}$$

Now, we assumed that we knew ν_x^g and $\beta_x^{g,n,\varepsilon}$ as part of our model, while in practice we do not. We follow the empirical Bayes approach, and estimate these quantities, and then plug in the estimates as if we knew these values a priori. The resulting estimator μ_x^n of θ_x will be identical to the estimator of θ_x derived using frequentist techniques in George et al. (2008).

First, we estimate ν_x^g. Our estimate will be $(\delta_x^{g,n})^2$, where $\delta_x^{g,n}$ is an estimate of the aggregation bias that we define here. At the unaggregated level ($g=0$), the aggregation bias is clearly 0, so we set $\delta_x^{0,n}=0$. If we have measured alternative x and $g>0$, then we set $\delta_x^{g,n}=\max(|\mu_x^{g,n}-\mu_x^{0,n}|,\underline{\delta})$, where $\underline{\delta}\geq 0$ is a constant parameter of the inference method. When $\underline{\delta}>0$, estimates of the aggregation bias are prevented from falling below some minimum threshold, which prevents the algorithm from placing too much weight on a frequently measured aggregate level when estimating the value of an infrequently measured disaggregate level. The convergence proof assumes $\underline{\delta}>0$, although in practice we find that the algorithm works well even when $\underline{\delta}=0$.

To generalize this estimate to include situations when we have not measured alternative x, we introduce a base level $\tilde{g}_x^n$ for each alternative x, being the lowest level g for which $m_x^{g,n}>0$. We then define $\delta_x^{g,n}$ as

$$\delta_x^{g,n}=\begin{cases}0 & \text{if } g=0 \text{ or } g<\tilde{g}_x^n,\\ \max(|\mu_x^{\tilde{g}_x^n}-\mu_x^{g,n}|,\underline{\delta}) & \text{if } g>0 \text{ and } g\geq\tilde{g}_x^n.\end{cases} \tag{9}$$

In addition, we set $w_x^{g,n}=0$ for all $g<\tilde{g}_x^n$.

Second, we estimate $\beta_x^{g,n,\varepsilon}$ using $\beta_x^{g,n,\varepsilon}=\left(\sigma_x^{g,n,\varepsilon}\right)^{-2}$ where $(\sigma_x^{g,n,\varepsilon})^2$ is the group variance (also called the population variance). The group variance $(\sigma_x^{0,n,\varepsilon})^2$ at the disaggregate ($g=0$) level equals λ_x, and we may use analysis of variance (see, e.g., Snijders and Bosker, 1999) to compute the group variance at $g>0$. The group variance over a number of subgroups equals the variance within each subgroup plus the variance between the subgroups. The variance within each subgroup is a weighted average of the variance $\lambda_{x'}$ of measurements of each alternative $x'\in\mathcal{X}^g(x)$. The variance between subgroups is given by the sum of squared deviations of the disaggregate estimates and the aggregate estimates of each alternative. The sum of these variances gives the group variance as

$$(\sigma_x^{g,n,\varepsilon})^2=\frac{1}{m_x^{g,n}}\left(\sum_{x'\in\mathcal{X}^g(x)}m_{x'}^{0,n}\lambda_{x'}+\sum_{x'\in\mathcal{X}^g(x)}m_{x'}^{0,n}\left(\mu_{x'}^{0,n}-\mu_x^{g,n}\right)^2\right),$$

where $m_x^{g,n}$ is the number of measurements from the aggregated alternative $G^g(x)$ at the g^{th} aggregation level, that is, the total number of measurements from alternatives in the set $\mathcal{X}^g(x)$, after n measurements. For $g=0$ we have $\left(\sigma_x^{g,n,\varepsilon}\right)^2=\lambda_x$.

In the computation of $\left(\sigma_x^{g,n,\varepsilon}\right)^2$, the numbers $m_{x'}^{0,n}$ can be regarded as weights: the sum of the bias and measurement variance of the alternative we measured the most contributes the most to the group variance $\left(\sigma_x^{g,n,\varepsilon}\right)^2$. This is because observations of this alternative also have the biggest impact on the aggregate estimate $\mu_x^{g,n}$. The problem, however, is that we are going to use the group variances $\left(\sigma_x^{g,n,\varepsilon}\right)^2$ to get an idea about the range of possible values of $\hat{y}_{x'}^{n+1}$ for all $x'\in\mathcal{X}^g(x)$. By including the number of measurements $m_{x'}^{0,n}$, this estimate of the range will heavily depend on the measurement policy. We propose to put equal weight on each alternative by setting $m_x^{g,n}=|\mathcal{X}^g(x)|$ (so $m_x^{0,n}=1$). The group variance $\left(\sigma_x^{g,n,\varepsilon}\right)^2$ is then given by

$$(\sigma_x^{g,n,\varepsilon})^2=\frac{1}{|\mathcal{X}^g(x)|}\left(\sum_{x'\in\mathcal{X}^g(x)}\lambda_{x'}+\left(\mu_{x'}^{0,n}-\mu_x^{g,n}\right)^2\right). \tag{10}$$

A summary of the Bayesian inference procedure can be found in Appendix A. Given this method of inference, we formally present in the next section the HKG policy for choosing the measurements x^n.

4. Measurement Decision

Our goal is to maximize the expected reward $\mu^N_{x^N}$ of the implementation decision $x^N = \arg\max_{x\in\mathcal{X}} \mu^N_x$. During the sequence of N sampling decisions, $x^0, x^1, \ldots, x^{N-1}$ we gain information that increases our expected final reward $\mu^N_{x^N}$. We may formulate an equivalent problem in which the reward is given in pieces over time, but the total reward given is identical. Then the reward we gain in a single time unit might be regarded as an increase in knowledge. The knowledge-gradient policy maximizes this single period reward. In Section 4.1 we provide a brief general introduction of the knowledge-gradient policy. In Section 4.2 we summarize the knowledge-gradient policy for independent and correlated multivariate normal beliefs as introduced in Frazier et al. (2008, 2009). Then, in Section 4.3, we adapt this policy to our hierarchical setting. We end with an illustration of how the hierarchical knowledge gradient policy chooses its measurements (Section 4.4).

4.1 The Knowledge-Gradient Policy

The knowledge-gradient policy was first introduced in Gupta and Miescke (1996) under the name $(R_1, \ldots, R_1)$, further analyzed in Frazier et al. (2008), and extended in Frazier et al. (2009) to cope with correlated beliefs. The idea works as follows. Let S^n be the knowledge state at time n. In Frazier et al. (2008, 2009) this is given by $S^n = (\mu^n, \Sigma^n)$, where the posterior on θ is $\mathcal{N}(\mu^n, \Sigma^n)$. If we were to stop measuring now, our final expected reward would be $\max_{x\in\mathcal{X}} \mu^n_x$. Now, suppose we were allowed to make one more measurement x^n. Then, the observation $\hat{y}^{n+1}_{x^n}$ would result in an updated knowledge state S^{n+1} which might result in a higher expected reward $\max_{x\in\mathcal{X}} \mu^{n+1}_x$ at the next time unit. The expected incremental value due to measurement x is given by

$$\upsilon^{KG}_x(S^n) = \mathbb{E}\left[\max_{x'\in\mathcal{X}} \mu^{n+1}_{x'} | S^n, x^n = x\right] - \max_{x'\in\mathcal{X}} \mu^n_{x'}. \tag{11}$$

The knowledge-gradient policy π^{KG} chooses its sampling decisions to maximize this expected incremental value. That is, it chooses x^n as

$$x^n = \arg\max_{x\in\mathcal{X}} \upsilon^{KG}_x(S^n).$$

4.2 Knowledge Gradient For Independent And Correlated Beliefs

In Frazier et al. (2008) it is shown that when all components of θ are independent under the prior and under all subsequent posteriors, the knowledge gradient (11) can be written

$$\upsilon^{KG}_x(S^n) = \tilde{\sigma}_x(\Sigma^n, x)\, f\left(\frac{-|\mu^n_x - \max_{x'\neq x} \mu^n_{x'}|}{\tilde{\sigma}_x(\Sigma^n, x)}\right),$$

where $\tilde{\sigma}_x(\Sigma^n, x) = Var\left(\mu^{n+1}_x | S^n, x^n = x\right) = \Sigma^n_{xx}/\sqrt{\lambda_x + \Sigma^n_{xx}}$, with Σ^n_{xx} the variance of our estimate μ^n_x, and where $f(z) = \varphi(z) + z\Phi(z)$ where $\varphi(z)$ and $\Phi(z)$ are, respectively, the normal density and cumulative distribution functions.

In the case of correlated beliefs, an observation $\hat{y}^{n+1}_x$ of alternative x may change our estimate $\mu^n_{x'}$ of alternatives $x' \neq x$. The knowledge gradient (11) can be written as

$$\upsilon^{KG,n}_x(S^n) = \mathbb{E}\left[\max_{x'\in\mathcal{X}} \mu^n_{x'} + \tilde{\sigma}_{x'}(\Sigma^n, x) Z | S^n, x^n = x\right] - \max_{x'\in\mathcal{X}} \mu^n_{x'}, \tag{12}$$

where Z is a standard normal random variable and $\tilde{\sigma}_{x'}(\Sigma^n, x) = \Sigma^n_{x'x}/\sqrt{\lambda_x + \Sigma^n_{xx}}$ with $\Sigma^n_{x'x}$ the covariance between $\mu^n_{x'}$ and μ^n_x.

Solving (12) involves the computation of the expectation over a piecewise linear convex function, which is given as the maximum of affine functions $\mu^n_{x'} + \tilde{\sigma}_{x'}(\Sigma^n, x) Z$. To do this, Frazier et al. (2009) provides an algorithm (Algorithm 2) which solves $h(a,b) = \mathbb{E}[\max_i a_i + b_i Z] - \max_i a_i$ as a generic function of any vectors a and b. In Frazier et al. (2009), the vectors a and b are given by the elements $\mu^n_{x'}$ and $\tilde{\sigma}_{x'}(\Sigma^n, x)$ for all $x' \in \mathcal{X}$ respectively, and the index i corresponds to a particular x'. The algorithm works as follows. First it sorts the sequence of pairs (a_i, b_i) such that the b_i are in nondecreasing order and ties in b are broken by removing the pair (a_i, b_i) when $b_i = b_{i+1}$ and $a_i \leq a_{i+1}$. Next, all pairs (a_i, b_i) that are dominated by the other pairs, that is, $a_i + b_i Z \leq \max_{j \neq i} a_j + b_j Z$ for all values of Z, are removed. Throughout the paper, we use $\tilde{a}$ and $\tilde{b}$ to denote the vectors that result from sorting a and b by b_i followed by the dropping of the unnecessary elements, producing a smaller $\tilde{M}$. The knowledge gradient can now be computed using

$$\upsilon^{KG}_x = \sum_{i=1,\ldots,\tilde{M}} (\tilde{b}_{i+1} - \tilde{b}_i) f\left(-\left|\frac{\tilde{a}_i - \tilde{a}_{i+1}}{\tilde{b}_{i+1} - \tilde{b}_i}\right|\right).$$

Note that the knowledge gradient algorithm for correlated beliefs requires that the covariance matrix Σ^0 be provided as an input. These correlations are typically attributed to physical relationships among the alternatives.

4.3 Hierarchical Knowledge Gradient

We start by generalizing the definition (11) of the knowledge-gradient in the following way

$$\upsilon^{KG}_x(S^n) = \mathbb{E}\left[\max_{x' \in \mathcal{X}} \mu^{n+1}_{x'} | S^n, x^n = x\right] - \max_{x' \in \mathcal{X}} \mathbb{E}\left[\mu^{n+1}_{x'} | S^n, x^n = x\right], \tag{13}$$

where the knowledge state is given by $S^n = \left\{\mu^{g,n}_x, \beta^{g,n}_x : x \in \mathcal{X}, g \in \mathcal{G}\right\}$.

When using the Bayesian updating equations from the original knowledge-gradient policy, the estimates μ^n_x form a martingale, in which case the conditional expectation of $\mu^{n+1}_{x'}$ given S^n is $\mu^n_{x'}$, and (13) is equivalent to the original definition (11). Because of approximations used in the updating equations derived in Section 3, μ^n_x is not a martingale in our case, and the term subtracted in (13) ensures the non-negativity of the KG factor.

Before working out the knowledge gradient (13), we first focus on the aggregate estimate $\mu^{g,n+1}_x$. We rewrite the updating Equation (2) as

$$\begin{aligned}
\mu^{g,n+1}_x &= \left[\beta^{g,n}_x \mu^{g,n}_x + \beta^{g,n,\varepsilon}_x \hat{y}^{n+1}_x\right] / \beta^{g,n+1}_x \\
&= \mu^{g,n}_x + \frac{\beta^{g,n,\varepsilon}_x}{\beta^{g,n}_x + \beta^{g,n,\varepsilon}_x}\left(\hat{y}^{n+1}_x - \mu^{g,n}_x\right) \\
&= \mu^{g,n}_x + \frac{\beta^{g,n,\varepsilon}_x}{\beta^{g,n}_x + \beta^{g,n,\varepsilon}_x}\left(\hat{y}^{n+1}_x - \mu^n_x\right) + \frac{\beta^{g,n,\varepsilon}_x}{\beta^{g,n}_x + \beta^{g,n,\varepsilon}_x}\left(\mu^n_x - \mu^{g,n}_x\right).
\end{aligned}$$

Now, the new estimate is given by the sum of (i) the old estimate, (ii) the deviation of $\hat{y}^{n+1}_x$ from the weighted estimate μ^n_x times the relative increase in precision, and (iii) the deviation of the estimate $\mu^{g,n}_x$ from the weighted estimate μ^n_x times the relative increase in precision. This means

that even if we observe precisely what we expected $\left(\hat{y}_x^{n+1} = \mu_x^n\right)$, the aggregate estimate $\mu_x^{g,n+1}$ still shrinks towards our current weighted estimate μ_x^n. However, the more observations we have, the less shrinking will occur because the precision of our belief on $\mu_x^{g,n}$ will be higher.

The conditional distribution of $\hat{y}_x^{n+1}$ is $\mathcal{N}\left(\mu_x^n, (\sigma_x^n)^2 + \lambda_x\right)$ where the variance of $\hat{y}_x^{n+1}$ is given by the measurement noise λ_x of the current measurement plus the variance $(\sigma_x^n)^2$ of our belief given by (6). So, $Z = \left(\hat{y}_x^{n+1} - \mu_x^n\right) / \sqrt{(\sigma_x^n)^2 + \lambda_x}$ is a standard normal. Now we can write

$$\mu_x^{g,n+1} = \mu_x^{g,n} + \frac{\beta_x^{g,n,\varepsilon}}{\beta_x^{g,n} + \beta_x^{g,n,\varepsilon}} \left(\mu_x^n - \mu_x^{g,n}\right) + \tilde{\sigma}_x^{g,n} Z, \tag{14}$$

where

$$\tilde{\sigma}_x^{g,n} = \frac{\beta_x^{g,n,\varepsilon} \sqrt{(\sigma_x^n)^2 + \lambda_x}}{\beta_x^{g,n} + \beta_x^{g,n,\varepsilon}}. \tag{15}$$

We are interested in the effect of decision x on the weighted estimates $\left\{\mu_{x'}^{n+1},\ \forall x' \in \mathcal{X}\right\}$. The problem here is that the values $\mu_{x'}^n$ for all alternatives $x' \in \mathcal{X}$ are updated whenever they share at least one aggregation level with alternative x, which is to say for all x' for which $\mathcal{G}(x', x)$ is not empty. To cope with this, we break our expression (7) for the weighted estimate $\mu_{x'}^{n+1}$ into two parts

$$\mu_{x'}^{n+1} = \sum_{g \notin \mathcal{G}(x',x)} w_{x'}^{g,n+1} \mu_{x'}^{g,n+1} + \sum_{g \in \mathcal{G}(x',x)} w_{x'}^{g,n+1} \mu_x^{g,n+1}.$$

After substitution of (14) and some rearrangement of terms we get

$$\begin{aligned} \mu_{x'}^{n+1} \quad = \quad & \sum_{g \in \mathcal{G}} w_{x'}^{g,n+1} \mu_{x'}^{g,n} + \sum_{g \in \mathcal{G}(x',x)} w_{x'}^{g,n+1} \frac{\beta_x^{g,n,\varepsilon}}{\beta_x^{g,n} + \beta_x^{g,n,\varepsilon}} \left(\mu_x^n - \mu_x^{g,n}\right) \\ & + Z \sum_{g \in \mathcal{G}(x',x)} w_{x'}^{g,n+1} \tilde{\sigma}_x^{g,n}. \end{aligned} \tag{16}$$

Because the weights $w_{x'}^{g,n+1}$ depend on the unknown observation $\hat{y}_{x'}^{n+1}$, we use an estimate $\bar{w}_{x'}^{g,n}(x)$ of the updated weights given we are going to sample x. Note that we use the superscript n instead of $n+1$ to denote its $\mathcal{F}^n$ measurability.

To compute $\bar{w}_{x'}^{g,n}(x)$, we use the updated precision $\beta_x^{g,n+1}$ due to sampling x in the weights (8). However, we use the current biases $\delta_x^{g,n}$ because the updated bias $\delta_x^{g,n+1}$ depends on the $\mu_x^{g,n+1}$ which we aim to estimate. The predictive weights $\bar{w}_{x'}^{g,n}(x)$ are

$$\bar{w}_{x'}^{g,n}(x) = \frac{\left(\left(\beta_{x'}^{g,n} + I_{x',x}^g \beta_{x'}^{g,n,\varepsilon}\right)^{-1} + \left(\delta_{x'}^{g,n}\right)^2\right)^{-1}}{\sum_{g' \in \mathcal{G}} \left(\left(\beta_{x'}^{g',n} + I_{x',x}^{g'} \beta_{x'}^{g',n,\varepsilon}\right)^{-1} + \left(\delta_{x'}^{g',n}\right)^2\right)^{-1}}, \tag{17}$$

where

$$I_{x',x}^g = \begin{cases} 1 & \text{if } g \in \mathcal{G}(x', x) \\ 0 & \text{otherwise} \end{cases}.$$

After combining (13) with (16) and (17), we get the following knowledge gradient

$$\upsilon_x^{KG}(S^n) = \mathbb{E}\left[\max_{x' \in \mathcal{X}} a_{x'}^n(x) + b_{x'}^n(x) Z | S^n\right] - \max_{x' \in \mathcal{X}} a_{x'}^n(x), \tag{18}$$

where

$$a^n_{x'}(x) = \sum_{g\in\mathcal{G}} \bar{w}^{g,n}_{x'}(x)\mu^{g,n}_{x'} + \sum_{g\in\mathcal{G}(x',x)} \bar{w}^{g,n}_{x'}(x)\frac{\beta^{g,n,\varepsilon}_x}{\beta^{g,n}_x+\beta^{g,n,\varepsilon}_x}\left(\mu^n_x-\mu^{g,n}_x\right), \tag{19}$$

$$b^n_{x'}(x) = \sum_{g\in\mathcal{G}(x',x)} \bar{w}^{g,n}_{x'}(x)\tilde{\sigma}^{g,n}_x. \tag{20}$$

Note that these equations for the knowledge gradient are quite different from those presented in Frazier et al. (2008) for the knowledge gradient for independent beliefs. However, it can be shown that without aggregation levels they coincide (if $G=0$, then $a^n_{x'}(x)=\mu^{0,n}_{x'}=\mu^n_{x'}$ and $b^n_{x'}(x)=\tilde{\sigma}^{0,n}_x$).

Following the approach of Frazier et al. (2009), which was briefly described in Section 4.2, we define $a^n(x)$ as the vector $\left\{a^n_{x'}(x),\ \forall x'\in\mathcal{X}\right\}$ and $b^n(x)$ as the vector $\left\{b^n_{x'}(x),\ \forall x'\in\mathcal{X}\right\}$. From this we derive the adjusted vectors $\tilde{a}^n(x)$ and $\tilde{b}^n(x)$. The knowledge gradient (18) can now be computed using

$$\upsilon^{KG,n}_x = \sum_{i=1,\ldots,\tilde{M}-1}\left(\tilde{b}^n_{i+1}(x)-\tilde{b}^n_i(x)\right) f\left(-\left|\frac{\tilde{a}^n_i(x)-\tilde{a}^n_{i+1}(x)}{\tilde{b}^n_{i+1}(x)-\tilde{b}^n_i(x)}\right|\right), \tag{21}$$

where $\tilde{a}^n_i(x)$ and $\tilde{b}^n_i(x)$ follow from (19) and (20), after the sort and merge operation as described in Section 4.2.

The form of (21) is quite similar to that of the expression in Frazier et al. (2009) for the correlated knowledge-gradient policy, and the computational complexities of the resulting policies are the same. Thus, like the correlated knowledge-gradient policy, the complexity of the hierarchical knowledge-gradient policy is $O\left(M^2\log M\right)$. An algorithm outline for the hierarchical knowledge-gradient measurement decision can be found in Appendix A.

4.4 Remarks

Before presenting the convergence proofs and numerical results, we first provide the intuition behind the hierarchical knowledge gradient (HKG) policy. As illustrated in Powell and Frazier (2008), the independent KG policy prefers to measure alternatives with a high mean and/or with a low precision. As an illustration, consider Figure 3, where we use an aggregation structure given by a perfect binary tree (see Section 6.3) with 128 alternatives at the disaggregate level. At aggregation level 5, there are four aggregated alternatives. As a result, the first four measurements are chosen such that we have one observation for each of these alternatives. The fifth measurement will be either in an unexplored region one aggregation level lower (aggregation level 4 consisting of eight aggregated alternatives) or at an already explored region that has a high weighted estimate. In this case, HKG chooses to sample from the unexplored region $48 < x \leq 64$ since it has a high weighted estimate and a low precision. The same holds for the sixth measurements which would be either from one of the three remaining unexplored aggregated alternatives from level 4, or from an already explored alternative with high weighted mean. In this case, HKG chooses to sample from the region $32 < x \leq 40$, which corresponds with an unexplored alternative at the aggregation level 3. The last panel shows the results after 20 measurements. From this we see HKG concentrates its measurements around the optimum and we have a good fit in this area.

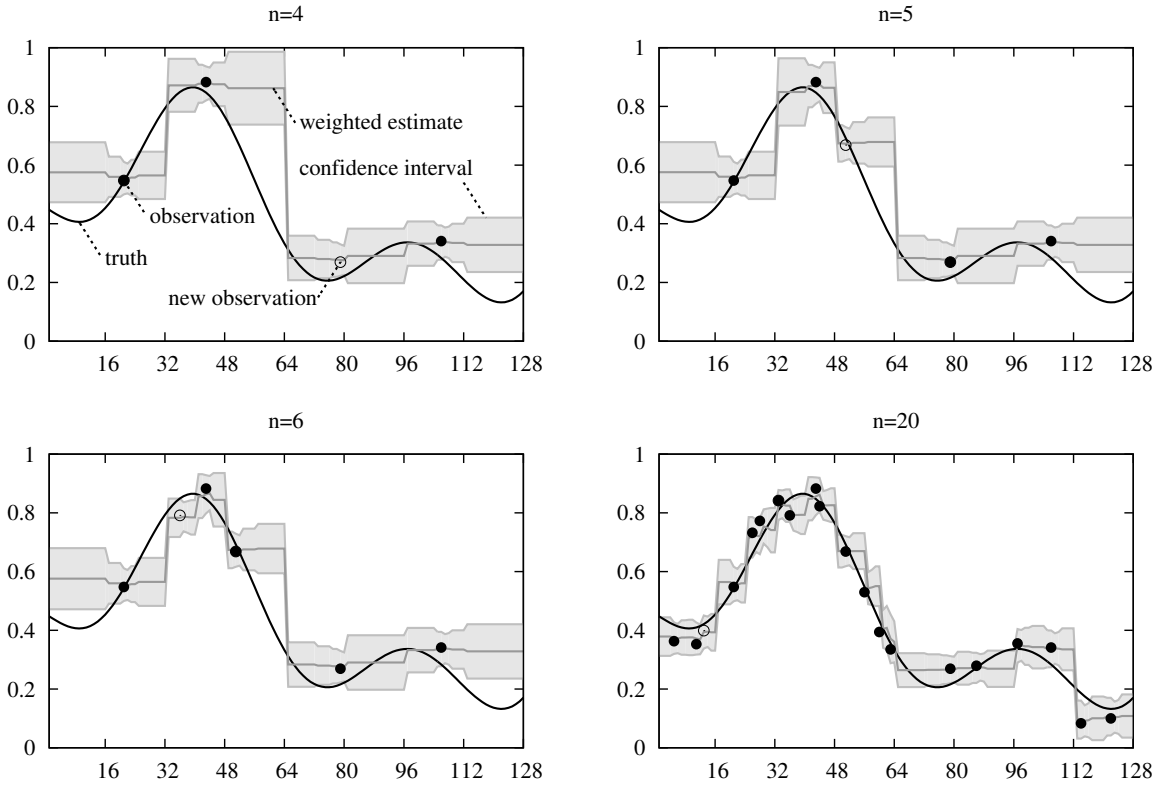

Figure 3: Illustration of the way HKG chooses its measurements.

5. Convergence Results

In this section, we show that the HKG policy measures each alternative infinitely often (Theorem 1). This implies that the HKG policy learns the true values of every alternative as $n \to \infty$ (Corollary 2) and eventually finds a globally optimal alternative (Corollary 3). This final corollary is the main theoretical result of this paper. The proofs of these results depend on lemmas found in Appendix C.

Although the posterior inference and the derivation of the HKG policy assumed that samples from alternative x were normal random variables with known variance λ_x, the theoretical results in this section allow general sampling distributions. We assume only that samples from any fixed alternative x are independent and identically distributed (iid) with finite variance, and that $\underline{\delta} > 0$. These distributions may, of course, differ across x. Thus, even if the true sampling distributions do not meet the assumptions made in deriving the HKG policy, we still enjoy convergence to a globally optimal alternative. We continue to define θ_x to be the true mean of the sampling distribution from alternative x, but the true variance of this distribution can differ from λ_x.

Theorem 1 *Assume that samples from any fixed alternative x are iid with finite variance, and that $\underline{\delta} > 0$. Then, the HKG policy measures each alternative infinitely often (i.e., $\lim_{n\to\infty} m_x^{0,n} = \infty$ for each $x \in \mathcal{X}$) almost surely.*

Proof Consider what happens as the number of measurements n we make under the HKG policy goes to infinity. Let $\mathcal{X}_\infty$ be the set of all alternatives measured infinitely often under our HKG policy,

and note that this is a random set. Suppose for contradiction that $\mathcal{X}_\infty \neq \mathcal{X}$ with positive probability, that is, there is an alternative that we measure only a finite number of times. Let N_1 be the last time we measure an alternative outside of $\mathcal{X}_\infty$. We compare the KG values $\upsilon_x^{KG,n}$ of those alternatives within $\mathcal{X}_\infty$ to those outside $\mathcal{X}_\infty$.

Let $x \in \mathcal{X}_\infty$. We show that $\lim_{n\to\infty} \upsilon_x^{KG,n} = 0$. Since f is an increasing non-negative function, and $\tilde{b}_{i+1}^n(x) - \tilde{b}_i^n(x) \geq 0$ by the assumed ordering of the alternatives, we have the bounds

$$0 \leq \upsilon_x^{KG,n} \leq \sum_{i=1,\ldots,\tilde{M}-1} \left(\tilde{b}_{i+1}^n(x) - \tilde{b}_i^n(x)\right) f(0).$$

Taking limits, $\lim_{n\to\infty} \upsilon_x^{KG,n} = 0$ follows from $\lim_{n\to\infty} \tilde{b}_i^n(x) = 0$ for $i = 1,\ldots,\tilde{M}$, which follows in turn from $\lim_{n\to\infty} b_{x'}^n(x) = 0\ \forall x' \in \mathcal{X}$ as shown in Lemma 8.

Next, let $x \notin \mathcal{X}_\infty$. We show that $\liminf_{n\to\infty} \upsilon_x^{KG,n} > 0$. Let $U = \sup_{n,i} |a_i^n(x)|$, which is almost surely finite by Lemma 7. Let $x' \in \mathcal{X}_\infty$. At least one such alternative x' must exist since we allocate an infinite number of measurements and $\mathcal{X}$ is finite. Lemma 9 shows

$$\upsilon_x^{KG,n} \geq \frac{1}{2} |b_{x'}^n(x) - b_x^n(x)|\, f\left(\frac{-4U}{|b_{x'}^n(x) - b_x^n(x)|}\right).$$

From Lemma 8, we know that $\liminf_{n\to\infty} b_x^n(x) > 0$ and $\lim_{n\to\infty} b_{x'}^n(x) = 0$. Thus, $b^* = \liminf_{n\to\infty} |b_x^n(x) - b_{x'}^n(x)| > 0$. Taking the limit inferior of the bound on $\upsilon_x^{KG,n}$ and noting the continuity and monotonicity of f, we obtain

$$\liminf_{n\to\infty} \upsilon_x^{KG,n} \geq \frac{1}{2} b^* f\left(\frac{-4U}{b^*}\right) > 0.$$

Finally, since $\lim_{n\to\infty} \upsilon_x^{KG,n} = 0$ for all $x \in \mathcal{X}_\infty$ and $\liminf_{n\to\infty} \upsilon_{x'}^{KG,n} > 0$ for all $x' \notin \mathcal{X}_\infty$, each $x' \notin \mathcal{X}_\infty$ has an $n > N_1$ such that $\upsilon_{x'}^{KG,n} > \upsilon_x^{KG,n}\ \forall x \in \mathcal{X}_\infty$. Hence we choose to measure an alternative outside $\mathcal{X}_\infty$ at a time $n > N_1$. This contradicts the definition of N_1 as the last time we measured outside $\mathcal{X}_\infty$, contradicting the supposition that $\mathcal{X}_\infty \neq \mathcal{X}$. Hence we may conclude that $\mathcal{X}_\infty = \mathcal{X}$, meaning we measure each alternative infinitely often. ■

Corollary 2 *Assume that samples from any fixed alternative x are iid with finite variance, and that $\underline{\delta} > 0$. Then, under the HKG policy, $\lim_{n\to\infty} \mu_x^n = \theta_x$ almost surely for each $x \in \mathcal{X}$.*

Proof Fix x. We first consider $\mu_x^{0,n}$, which can be written as

$$\mu_x^{0,n} = \frac{\beta_x^{0,0}\mu_x^{0,0} + m_x^{0,n}(\lambda_x)^{-1}\bar{y}_x^n}{\beta_x^{0,0} + m_x^{0,n}(\lambda_x)^{-1}},$$

where $\bar{y}_x^n$ is the average of all observations of alternative x by time n. As $n \to \infty$, $m_x^{0,n} \to \infty$ by Theorem 1. Thus, $\lim_{n\to\infty} \mu_x^{0,n} = \lim_{n\to\infty} \bar{y}_x^n$, which is equal to θ_x almost surely by the law of large numbers.

We now consider the weights $w_x^{g,n}$. For $g \neq 0$, (8) shows

$$w_x^{g,n} \leq \frac{\left((\sigma_x^{g,n})^2 + (\delta_x^{g,n})^2\right)^{-1}}{(\sigma_x^{0,n})^{-2} + \left((\sigma_x^{g,n})^2 + (\delta_x^{g,n})^2\right)^{-1}}.$$

When n is large enough that we have measured at least one alternative in $\mathcal{X}^g(x)$, then $\delta_x^{g,n} \geq \underline{\delta}$, implying $\left((\sigma_x^{g,n})^2 + (\delta_x^{g,n})^2\right)^{-1} \leq \underline{\delta}^{-2}$ and $w_x^{g,n} \leq \underline{\delta}^{-2}/((\sigma_x^{0,n})^{-2} + \underline{\delta}^{-2})$. As $n \to \infty$, $m_x^{0,n} \to \infty$ by Theorem 1 and $(\sigma_x^{0,n})^{-2} = \beta^{0,0} + m_x^{0,n}(\lambda_x)^{-1} \to \infty$. This implies that $\lim_{n\to\infty} w_x^{g,n} = 0$. Also observe that $w_x^{0,n} = 1 - \sum_{g\neq 0} w_x^{g,n}$ implies $\lim_{n\to\infty} w_x^{0,n} = 1$.

These limits for the weights, the almost sure finiteness of $\sup_n |\mu_x^{g,n}|$ for each g from Lemma 7, and the definition (7) of μ_x^n together imply $\lim_{n\to\infty} \mu_x^n = \lim_{n\to\infty} \mu_x^{0,n}$, which equals θ_x as shown above. ■

Finally, Corollary 3 below states that the HKG policy eventually finds a globally optimal alternative. This is the main result of this section. In this result, keep in mind that $\hat{x}^n = \arg\max_x \mu_x^N$ is the alternative one would estimate to be best at time N, given all the measurements collected by HKG. It is this estimate that converges to the globally optimal alternative, and not the HKG measurements themselves.

Corollary 3 *For each n, let $\hat{x}^n \in \arg\max_x \mu_x^n$. Assume that samples from any fixed alternative x are iid with finite variance, and that $\underline{\delta} > 0$. Then, under the HKG policy, there exists an almost surely finite random variable N' such that $\hat{x}^n \in \arg\max_x \theta_x$ for all $n > N'$.*

Proof Let $\theta^* = \max_x \theta_x$ and $\varepsilon = \min\{\theta^* - \theta_x : x \in \mathcal{X}, \theta^* > \theta_x\}$, where $\varepsilon = \infty$ if $\theta_x = \theta^*$ for all x. Corollary 2 states that $\lim_{n\to\infty} \mu_x^n = \theta_x$ almost surely for all x, which implies the existence of an almost surely finite random variable N' with $\max_x |\mu_x^n - \theta_x| < \varepsilon/2$ for all $n > N'$. On the event $\{\varepsilon = \infty\}$ we may take $N' = 0$. Fix $n > N'$, let $x^* \in \arg\max_x \theta_x$, and let $x' \notin \arg\max_x \theta_x$. Then $\mu_{x^*}^n - \mu_{x'}^n = (\theta_{x^*} - \theta_{x'}) + (-\theta_{x^*} + \mu_{x^*}^n) + (\theta_{x'} - \mu_{x'}^n) > \theta_{x^*} - \theta_{x'} - \varepsilon \geq 0$. This implies that $\hat{x}^n \in \arg\max_x \theta_x$. ■

6. Numerical Experiments

To evaluate the hierarchical knowledge-gradient policy, we perform a number of experiments. Our objective is to find the strengths and weaknesses of the HKG policy. To this end, we compare HKG with some well-known competing policies and study the sensitivity of these policies to various problem settings such as the dimensionality and smoothness of the function, and the measurement noise.

6.1 Competing Policies

We compare the Hierarchical Knowledge Gradient (HKG) algorithm against several ranking and selection policies: the Interval Estimation (IE) rule from Kaelbling (1993), the Upper Confidence Bound (UCB) decision rule from Auer et al. (2002), the Independent Knowledge Gradient (IKG) policy from Frazier et al. (2008), Boltzmann exploration (BOLTZ), and pure exploration (EXPL).

In addition, we compare with the Knowledge Gradient policy for correlated beliefs (KGCB) from Frazier et al. (2009) and, from the field of Bayesian global optimization, we select the Sequential Kriging Optimization (SKO) policy from Huang et al. (2006). SKO is an extension of the well known Efficient Global Optimization (EGO) policy (Jones et al., 1998) to the case with noisy measurements.

We also consider an hybrid version of the HKG algorithm (HHKG) in which we only exploit the similarity between alternatives in the updating equations and not in the measurement decision. As a result, this policy uses the measurement decision of IKG and the updating equations of HKG. The possible advantage of this hybrid policy is that it is able to cope with similarity between alternatives without the computational complexity of HKG.

Since several of the policies require choosing one or more parameters, we provide a brief description of the implementation of these policies in Appendix D. For those policies that require it, we perform tuning using all one-dimensional test functions (see Section 6.2). For the Bayesian approaches, we always start with a non-informative prior.

6.2 Test Functions

To evaluate the policies numerically, we use various test functions with the goal of finding the highest point of each function. Measuring the functions is done with normally distributed noise with variance λ. The functions are chosen from commonly used test functions for similar procedures.

6.2.1 ONE-DIMENSIONAL FUNCTIONS

First we test our approach on one-dimensional functions. In this case, the alternatives x simply represent a single value, which we express by i or j. As test functions we use a Gaussian process with zero mean and power exponential covariance function

$$Cov(i,j) = \sigma^2 \exp\left\{-\left(\frac{|i-j|}{(M-1)\rho}\right)^{\eta}\right\},$$

which results in a stationary process with variance σ^2 and a length scale ρ.

Higher values of ρ result in fewer peaks in the domain and higher values of η result in smoother functions. Here we fix $\eta = 2$ and vary $\rho \in 0.05, 0.1, 0.2, 0.5$. The choice of σ^2 determines the vertical scale of the function. Here we fix $\sigma^2 = 0.5$ and we vary the measurement variance λ.

To generate a truth θ_i, we take a random draw from the Gaussian process (see, e.g., Rasmussen and Williams, 2006) evaluated at the discretized points $i = 1, .., 128$. Figure 4 shows one test function for each value of ρ.

Next, we consider non-stationary covariance functions. We choose to use the Gibbs covariance function (Gibbs, 1997) as it has a similar structure to the exponential covariance function but is non-stationary. The Gibbs covariance function is given by

$$Cov(i,j) = \sigma^2 \sqrt{\frac{2l(i)l(j)}{l(i)^2 + l(j)^2}} \exp\left(-\frac{(i-j)^2}{l(i)^2 + l(j)^2}\right),$$

where $l(i)$ is an arbitrary positive function in i. In our experiments we use a horizontally shifted periodic sine curve for $l(i)$,

$$l(i) = 1 + 10\left(1 + \sin\left(2\pi\left(\frac{i}{128} + u\right)\right)\right),$$

where u is a random number from [0,1] that shifts the curve horizontally across the x-axis. The function $l(i)$ is chosen so that, roughly speaking, the resulting function has one full period, that is, one area with relatively low correlations and one area with relatively high correlations. The area

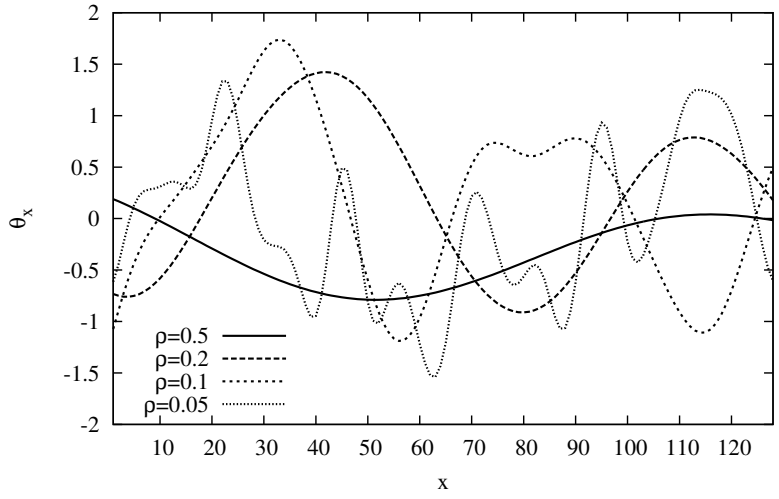

Figure 4: Illustration of one-dimensional test functions.

with low correlations visually resembles the case of having a stationary function with $\rho = 0.05$, whereas the area with high correlations visually resembles the case of having a stationary function with $\rho = 0.5$.

The policies KGCB, SKO and HKG all assume the presence of correlations in function values. To test the robustness of these policies in the absence of any correlation, we consider one last one-dimensional test function. This function has an independent truth generated by $\theta_i = U[0,1], i = 1,...,128$.

6.2.2 TWO-DIMENSIONAL FUNCTIONS

Next, we consider two-dimensional test functions. First, we consider the Six-hump camel back (Branin, 1972) given by

$$f(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4.$$

Different domains have been proposed for this function. Here we consider the domain $x \in [-1.6, 2.4] \times [-0.8, 1.2]$ as also used in Huang et al. (2006) and Frazier et al. (2009), and a slightly bigger domain $x \in [-2, 3] \times [-1, 1.5]$. The extended part of this domain contains only values far from the optimum. Hence, the extension does not change the value and location of the optimum.

The second function we consider is the Tilted Branin (Huang et al., 2006) given by

$$f(x) = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10 + \frac{1}{2}x_1,$$

with $x \in [-5, 10] \times [0, 15]$.

The Six-hump camel back and Tilted Branin function are relatively smooth functions in the sense that a Gaussian process can be fitted to the truth relatively well. Obviously, KGCB and SKO benefit from this. To also study more messy functions, we shuffle these functions by placing a 2×2 grid onto the domain and exchange the function values from the lower left quadrant with those from the upper right quadrant.

With the exception of SKO, all policies considered in this paper require problems with a finite number of alternatives. Therefore, we discretize the set of alternatives and use an 32×32 equispaced grid on $\mathbb{R}^2$. We choose this level of discretization because, although our method is theoretically capable of handling any finite number of alternatives, computational issues limit the possible number to the order of thousands. This limit also holds for KGCB, which has the same computational complexity as HKG. For SKO we still use the continuous functions which should give this policy some advantage.

6.2.3 CASE EXAMPLE

To give an idea about the type of practical problems for which HKG can be used, we consider a transportation application (see Simao et al., 2009). Here we must decide where to send a driver described by three attributes: (i) the location to which we are sending him, (ii) his home location (called his domicile) and (iii) to which of six fleets he belongs. The "fleet" is a categorical attribute that describes whether the driver works regionally or nationally and whether he works as a single driver or in a team. The spatial attributes (driver location and domicile) are divided into 100 regions (by the company). However, to reduce computation time, we aggregate these regions into 25 regions. Our problem is to find which of the $25 \times 25 \times 6 = 3750$ is best.

To allow replicability of this experiment, we describe the underlying truth using an adaption of a known function which resembles some of the characteristics of the transportation application. For this purpose we use the Six-hump camel back function, on the smaller domain, as presented earlier. We let x_1 be the location and x_2 be the driver domicile, which are both discretized into 25 pieces to represent regions. To include the dependence on capacity type, we use the following transformation

$$g(x_1, x_2, x_3) = p_1(x_3) - p_2(x_3)(|x_1 - 2x_2|) - f(x_1, x_2),$$

where x_3 denotes the capacity type. We use $p_2(x_3)$ to describe the dependence of capacity type on the distance between the location of the driver and his domicile.

We consider the following capacity types: CAN for Canadian drivers that only serve Canadian loads, WR for western drivers that only serve western loads, US_S for United States (US) solo drivers, US_T for US team drivers, US_IS for US independent contractor solo drivers, and US_IT for US independent contractor team drivers. The parameter values are shown in Table 2. To cope with the fact that some drivers (CAN and WR) cannot travel to certain locations, we set the value to zero for the combinations $\{x_3 = \text{CAN} \wedge x_1 < 1.8\}$ and $\{x_3 = \text{WR} \wedge x_1 > -0.8\}$. The maximum of $g(x_1, x_2, x_3)$ is attained at $g(0, 0, \text{US\_S})$ with value 6.5.

x_3	CAN	WR	US_S	US_T	US_IS	US_IT
$p_1(x_3)$	7.5	7.5	6.5	5.0	2.0	0.0
$p_2(x_3)$	0.5	0.5	2.0	0.0	2.0	0.0

Table 2: Parameter settings.

To provide an indication of the resulting function, we show $\max_{x_3} g(x_1, x_2, x_3)$ in Figure 5. This function has similar properties to the Six-hump camel back, except for the presence of discontinuities due to the capacity types CAN and WR, and a twist at $x_1 = x_2$.

An overview of all test functions can be found in Table 3. Here σ denotes the standard deviation of the function measured over the given discretization.

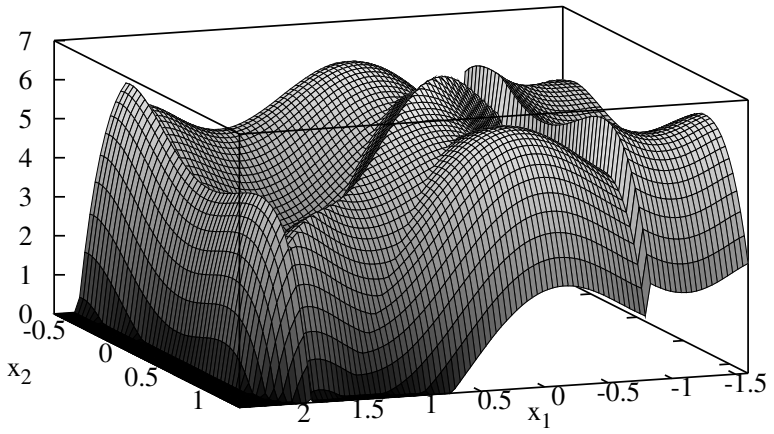

Figure 5: $max_{x_3} g(x_1, x_2, x_3)$.

Type	Function name	σ	Description
One-dimensional	GP1R005	0.32	stationary GP with $\rho = 0.05$
	GP1R01	0.49	stationary GP with $\rho = 0.1$
	GP1R02	0.57	stationary GP with $\rho = 0.2$
	GP1R05	0.67	stationary GP with $\rho = 0.5$
	NSGP	0.71	non-stationary GP
	IT	0.29	independent truth
Two-dimensional	SHCB-DS	2.87	Six-hump camel back on small domain
	SHCB-DL	18.83	Six-hump camel back on large domain
	TBRANIN	51.34	Tilted Branin
	SHCB-DS-SH	2.87	shuffled SHCB-DS
	SHCB-DB-SH	18.83	shuffled SHCB-DL
	TBRANIN-SH	51.34	shuffled TBRANIN
Case example	TA	3.43	transportation application

Table 3: Overview of test functions.

6.3 Experimental Settings

We consider the following experimental factors: the measurement variance λ, the measurement budget N, and for the HKG policy the aggregation structure. Given these factors, together with the nine policies from Section 6.1 and the 15 test functions from Section 6.2, a full factorial design is not an option. Instead, we limit the number of combinations as explained in this section.

As mentioned in the introduction, our interest is primarily in problems where M is larger than the measurement budget N. However, for these problems it would not make sense to compare with the tested versions of IE, UCB and BOLTZ since, in the absence of an informed prior, these methods typically choose one measurement of each of the M alternatives before measuring any alternative a second time. Although we do not do so here, one could consider versions of these policies with informative priors (e.g., the GP-UCB policy of Srinivas et al. (2010), which uses UCB with a Gaussian process prior), which would perform better on problems with M much larger than

N. To obtain meaningful results for the tested versions of IE, UCB and BOLTZ, we start with an experiment with a relatively large measurement budget and relatively large measurement noise. We use all one-dimensional test functions with $N = 500$ and $\sqrt{\lambda} \in \{0.5, 1\}$. We omit the policy HHKG, which will be considered later.

In the remaining experiments we omit the policies IE, UCB, and BOLTZ that use non-informative priors because they would significantly underperform the other policies. This is especially true with the multi-dimensional problems where the number of alternatives after discretization is much bigger then the measurement budget. We start with testing the remaining policies, together with the hybrid policy HHKG, on all one-dimensional test functions using $\sqrt{\lambda} \in \{0.1, 0.5, 1\}$ and $N = 200$. Next, we use the non-stationary function to study (i) the sensitivity of all policies on the value of λ, using $\sqrt{\lambda} \in \{0.1, 0.5, 1, 1.5, 2, 2.5\}$ and (ii) the sensitivity of HKG on the aggregation structure. For the latter, we consider two values for $\sqrt{\lambda}$, namely 0.5 and 1, and five different aggregation structures as presented at the end of this subsection.

For the stationary one-dimensional setting, we generate 10 random functions for each value of ρ. For the non-stationary setting and the random truth setting, we generate 25 random functions each. This gives a total of 90 different functions. We use 50 replications for each experiment and each generated function.

For the multi-dimensional functions we only consider the policies KGCB, SKO, HKG, and HHKG. For the two-dimensional functions we use $N = 200$. For the transportation application we use $N = 500$ and also present the results for intermediate values of n. We set the values for λ by taking into account the standard deviation σ of the functions (see Table 3). For the Six-hump camel back we use $\sqrt{\lambda} \in \{1, 2, 4\}$, for the Tilted Branin we use $\sqrt{\lambda} \in \{2, 4, 8\}$, and for the case example we use $\sqrt{\lambda} \in \{1, 2\}$. For the multi-dimensional functions we use 100 replications.

During the replications we keep track of the opportunity costs, which we define as $OC(n) = (\max_i \theta_i) - \theta_{i^*}$, with $i^* \in \arg\max_x \mu_x^n$, that is, the difference between the true maximum and the value of the best alternative found by the algorithm after n measurements. Our key performance indicator is the mean opportunity costs $\mathbb{E}[OC(n)]$ measured over all replications of one or more experiments. For clarity of exposition, we also group experiments and introduce a set GP1 containing the 40 stationary one-dimensional test functions and a set NS0 containing the 50 non-stationary and independent truth functions. When presenting the $\mathbb{E}[OC(n)]$ in tabular form, we bold and underline the lowest value, and we also bold those values that are not significantly different from the lowest one (using Welch's t test at the 0.05 level).

We end this section with an explanation of the aggregation functions used by HKG. Our default aggregation structure is given by a binary tree, that is, $|\mathcal{X}^g(x)| = 2^g$ for all $x \in \mathcal{X}^g$ and $g \in \mathcal{G}$. As a result, we have 8 $(\ln(128)/\ln(2) + 1)$ aggregation levels for the one-dimensional problems and 6 $(\ln(32)/\ln(2) + 1)$ for the two-dimensional problems. For the experiment with varying aggregation functions, we introduce a variable ω to denote the number of alternatives $G^g(x), g < G$ that should be aggregated in a single alternative $G^{g+1}(x)$ one aggregation level higher. At the end of the domain this might not be possible, for example, if we have an odd number of (aggregated) alternatives. In this case, we use the maximum number possible. We consider the values $\omega \in \{2, 4, 8, 16\}$, where $\omega = 2$ resembles the original situation of using a binary tree. To evaluate the impact of having a difference in the size of aggregated sets, we introduce a fifth aggregation structure where ω alternately takes values 2 and 4.

For the transportation application, we consider five levels of aggregation. At aggregation level 0, we have 25 regions for location and domicile, and 6 capacity types, producing 3750 attribute vectors.

At aggregation level 1, we represent the driver domicile as one of 5 areas. At aggregation level 2, we ignore the driver domicile; at aggregation level 3, we ignore capacity type; and at aggregation level 4, we represent location as one of 5 areas.

An overview of all experiments can be found in Table 4.

Experiment	Number of runs
One-dimensional long	$90 \times 8 \times 2 \times 1 \times 50 = 72,000$
One-dimensional normal	$90 \times 6 \times 3 \times 1 \times 50 = 81,000$
One-dimensional varying λ	$25 \times 6 \times 6 \times 1 \times 50 = 45,000$
One-dimensional varying ω	$25 \times 1 \times 2 \times 5 \times 50 = 12,500$
Two-dimensional	$6 \times 3 \times 3 \times 1 \times 100 = 27,000$
Transportation application	$2 \times 3 \times 2 \times 1 \times 100 = 6000$

Table 4: Overview of experiments. The number of runs is given by #functions $\times$ #policies $\times$ #λ's $\times$ #ω's $\times$ #replications. The total number of experiments, defined by the number of unique combinations of function, policy, λ, and ω, is 2696.

7. Numerical Results

In this section we present the results of the experiments described in Section 6. We demonstrate that HKG performs best when measured by the average performance across all problems. In particular, it outperforms others on functions for which the use of an aggregation function seems to be a natural choice, but it also performs well on problems for which the other policies are specifically designed. In the following subsections we present the policies, the test functions, and the experimental design.

7.1 One-dimensional Functions

In our first experiment, we focus on the comparison with R&S policies using a relatively large measurement budget. A complete overview of the results, for $n = 500$ and an intermediate value $n = 250$, can be found in Appendix E. To illustrate the sensitivity of the performance of these policies to the number of measurements n, we also provide a graphical illustration in Figure 6. To keep these figures readable, we omit the policies UCB and IKG since their performance is close to that of IE (see Appendix E).

As expected, the R&S policies perform well with many measurements. IE generally performs best, closely followed by UCB. BOLTZ only performs well for few measurements ($n \leq M$) after which it underperforms the other policies with the exception of EXPL, which spends an unnecessary portion of its measurements on less attractive alternatives.

With increasing n, IE eventually outperforms at least one of the advanced policies (KGCB, SKO, and HKG). However, it seems that the number of measurements required for IE to outperform KGCB and HKG increases with increasing measurement variance λ. We further see, from Appendix E, that IE outperforms IKG on most instances. However, keep in mind that we tuned IE using exactly the functions on which we test while IKG does not require any form of tuning. The qualitative change in the performance of IE at $n = 128$ samples is due to the fact that the version of IE against

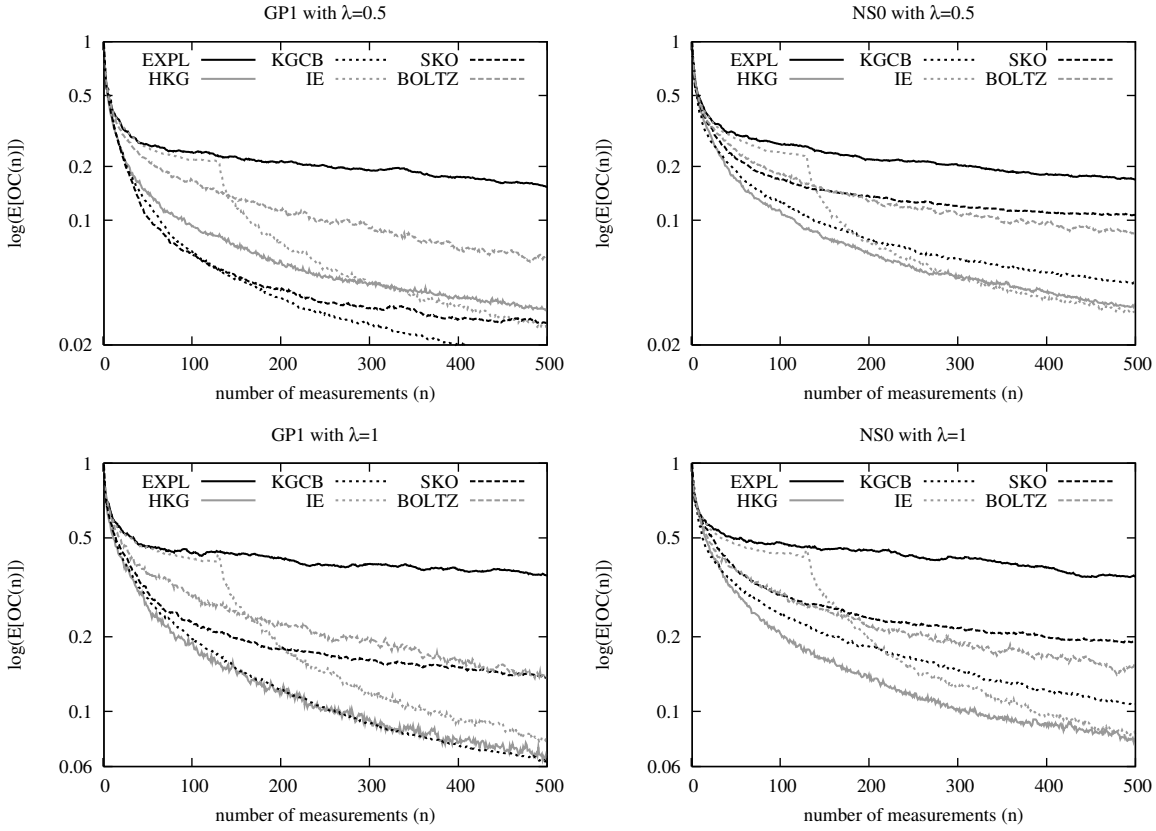

Figure 6: Results for the one-dimensional long experiments.

which we compare uses a non-informative prior, which causes it to measure each alternative exactly once before it can use the IE logic to decide where to allocate future samples.

With respect to the more advanced policies, we see that HKG outperforms the others on the NS0 functions (non-stationary covariance and independent truth) and performs competitively on the stationary GPs in the case of relatively large λ. Obviously, KGCB and SKO are doing well on the latter case since the truths are drawn from a Gaussian process and these policies fit a Gaussian process to the evaluated function values. Apart from the given aggregation function, HKG does not assume any structure and therefore has a slower rate of convergence on these instances. Further, it is remarkable to see that SKO is only competitive on GP1 with $\lambda = 0.5$ but not with $\lambda = 1$. We return to this issue in the next experiment.

For a more detailed comparison between KGCB, SKO and HKG we now focus on smaller measurement budgets. A summary of the results can be found in Table 5. More detailed results in combination with a further analysis can be found in Appendix E. As mentioned before, we bold and underline the lowest value, and we also bold those values that are not significantly different from the lowest one.

On the GP1 functions with $\lambda \leq 0.5$, HKG is outperformed by KGCB and SKO. SKO does particularly well during the early measurements (n=50) after which it is outperformed by KGCB (n=200). On the GP1 functions with $\lambda = 1$, we see HKG becomes more competitive: in almost

Function	$\sqrt{\lambda}$	n	EXPL	IKG	KGCB	SKO	HKG	HHKG
GP1	0.1	50	0.090	0.081	0.010	**0.008**	0.034	0.078
		200	0.051	0.006	**0.002**	0.004	0.008	0.008
	0.5	50	0.265	0.252	0.123	**0.104**	0.141	0.175
		200	0.214	0.075	**0.037**	**0.041**	0.059	0.065
	1	50	0.460	0.441	0.286	0.302	**0.265**	0.305
		200	0.415	0.182	**0.122**	0.181	**0.121**	0.135
NS0	0.1	50	0.111	0.096	0.066	0.093	**0.051**	0.113
		200	0.043	**0.008**	0.017	0.060	**0.009**	0.014
	0.5	50	0.301	0.288	0.189	0.221	**0.170**	0.212
		200	0.219	0.086	0.078	0.136	**0.065**	0.081
	1	50	0.498	0.468	0.323	0.375	**0.306**	0.335
		200	0.446	0.213	0.183	0.238	**0.141**	0.163

Table 5: $\mathbb{E}[OC(n)]$ on the one-dimensional normal experiments.

all cases it outperforms SKO, and with a limited measurement budget (n=50) it also outperforms KGCB.

On the NS0 functions, we see that HKG always outperforms KGCB and SKO with the only exception being the independent truth (IT) function with $\lambda = 1$ and $n = 50$ (see Appendix E). We also see that SKO is always outperformed by KGCB. Especially in the case with low measurement noise ($\lambda = 0.1$) and a large number of measurements ($n = 200$), SKO performs relatively poorly. This is exactly the situation in which one would expect to obtain a good fit, but a fitted Gaussian process prior with zero correlation is of no use. With an increasing number of measurements, we see SKO is even outperformed by EXPL.

In general, HKG seems to be relatively robust in the sense that, whenever it is outperformed by other policies, it still performs well. This claim is also supported by the opportunity costs measured over all functions and values of λ found in Table 6 (note this is not a completely fair comparison since we have slightly more non-stationary functions, and the average opportunity costs over all policies is slightly higher in the non-stationary cases). Even though HKG seems to be quite competitive, HKG seems to have convergence problems in the low noise case ($\lambda = 0.1$). We analyze this issue further in Appendix E. The hybrid policy does not perform well, although it outperforms IKG on most problem instances.

	EXPL	IKG	KGCB	SKO	HKG	HHKG
$\mathbb{E}[OC(50)]$	0.289	0.273	0.169	0.189	**0.163**	0.205
$\mathbb{E}[OC(200)]$	0.232	0.096	0.075	0.114	**0.068**	0.078

Table 6: Aggregate results for the one-dimensional normal experiments.

In the next experiment we vary the measurement variance λ. Figure 7 shows the relative reduction in $\mathbb{E}[OC(50)]$ compared with the performance of EXPL. For clarity of exposition, we omitted the results for $n = 200$ and the performance of IKG. These results confirm our initial conclusions with respect to the measurement variance: increasing λ gives HKG a competitive advantage whereas the opposite holds for SKO. On the GP1R02 functions, HKG is outperformed by SKO and KGCB for $\lambda \leq 0.5$. With $\lambda > 0.5$, the performance of KGCB, HKG, and HHKG is close and they all

outperform SKO. On the NSGP functions, the ordering of policies seem to remain the same for all values of λ, with the exception that with $\lambda \geq 1$, SKO is outperformed by all policies. The difference between KGCB and HKG seems to decline with increasing λ.

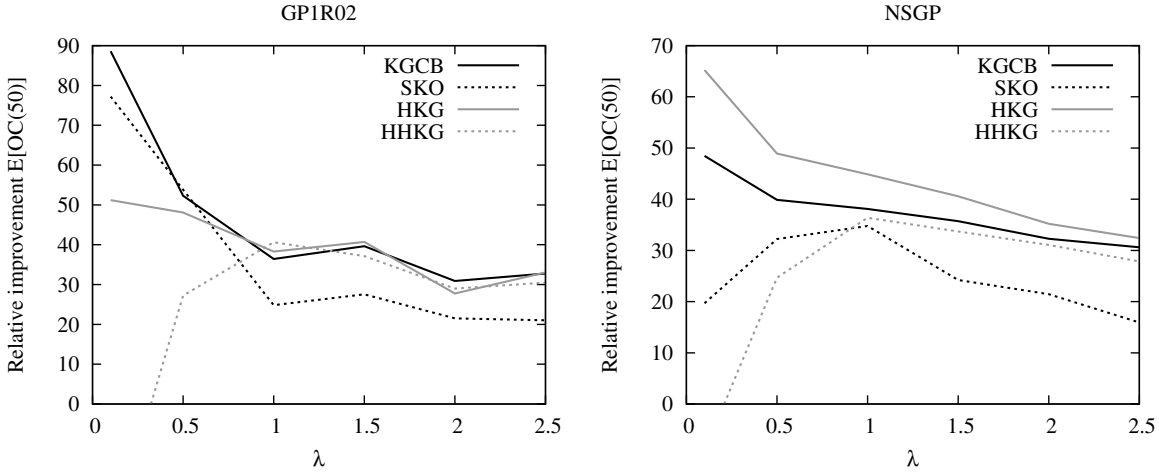

Figure 7: Sensitivity to the measurement noise.

As a final test with one-dimensional functions, we now vary the aggregation structure used by HKG. The results can be found in Figure 8. Obviously, HKG is sensitive to the choice of aggregation structure. The aggregation function with $\omega = 16$ is so coarse that, even on the lowest aggregation level, there exists aggregate alternatives that have local maxima as well as local minima in their aggregated set. We also see that the performance under the $\omega = 2/4$ structure is close to that of $\omega = 4$, which indicates that having some symmetry in the aggregation function is preferable. When comparing the two figures, we see that the impact of the aggregation function decreases with increasing λ. The reason for this is that with higher λ, more weight is given to the more aggregate levels. As a result, the benefit of having more precise lower aggregation levels decreases.

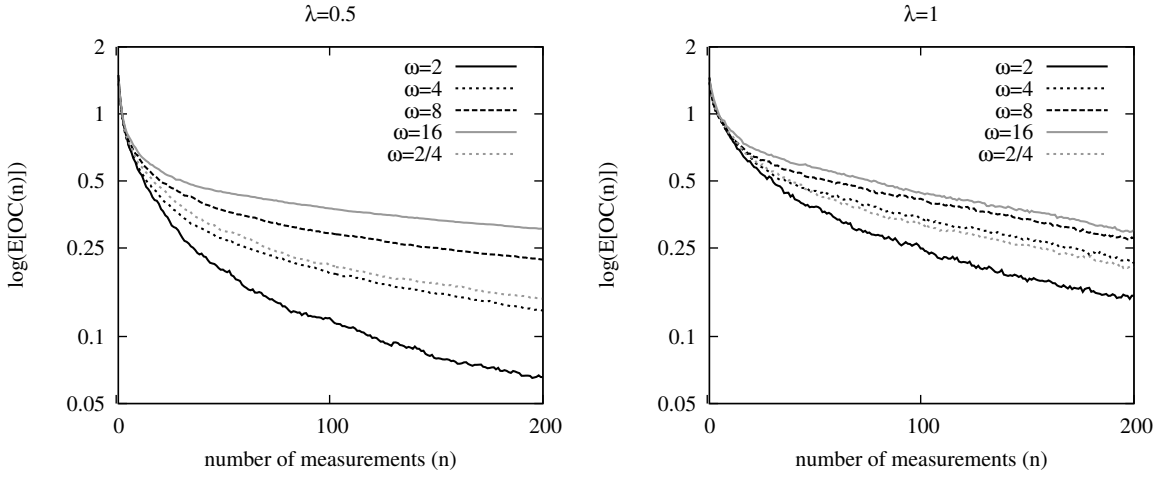

Figure 8: Sensitivity of HKG to the aggregation function.

7.2 Two-dimensional Functions

An overview of results for the two-dimensional functions can be found in Table 7. From these results we draw the following conclusions:

1. On the standard test functions, SHCB-DS and TBRANIN, HKG is outperformed by KGCB and SKO. However, with increasing λ, HKG still outperforms SKO.

2. In case of the Six-hump camel back function, just extending the domain a bit (where the extended part of the domain only contains points with large opportunity costs) has a major impact on the results. With the exception of one outcome (KGCB with $\lambda = 1$), the opportunity costs increase for all policies. This makes sense because there are simply more alternatives with higher opportunity costs. For KGCB and SKO, these extreme values also play a role in fitting the Gaussian process prior. As a result, we have a less reliable fit at the area of interest, something especially SKO suffers from. Obviously, also HKG 'loses' measurements on these extreme values. However, their influence on the fit (via the aggregation function) is limited since HKG automatically puts a low weight on them. As a result, HKG outperforms the other policies in almost all cases.

3. Shuffling the Six-hump camel back has a similar influence to extending the domain. In all cases, HKG outperforms KGCB and SKO. Shuffling the TBRANIN has an especially large impact on the performance of KGCB and SKO. However, not all performance differences with the shuffled TBRANIN are significant due to relatively large variances, especially in the case of $n = 50$.

7.3 Example Case

The results for the transportation application can be found in Figure 9. As mentioned in Section 6, the first two dimensions of this problem are described by the Six-hump camel back function on the small domain. This function is also considered in Huang et al. (2006) and Frazier et al. (2009) where the policies SKO and KGCB respectively are introduced. Compared to HKG, these policies perform relatively well on this standard test function. It is interesting to see that the addition of a third, categorical, dimension changes the situation.

As can be seen from Figure 9, HKG outperforms SKO and KGCB for both values of λ and almost all intermediate values of n. Measured at $n = 100$ and $n = 200$, the differences between HKG and both KGCB and SKO are significant (again using the 0.05 level). The hybrid policy HHKG is doing remarkably well; the differences with HKG at $n = 200$ are not significant, which is partly due to the fact that the variances with HHKG are higher. The performance of HHKG is especially remarkable since this policy requires only a fraction of the computation time of the others. Given, the large number of measurements and alternatives, the running times of KGCB, SKO, and HKG take multiple hours per replication whereas HHKG requires around 10 seconds.

8. Conclusions

We have presented an efficient learning strategy to optimize an arbitrary function that depends on a multi-dimensional vector with numerical and categorical attributes. We do not attempt to fit a

		$\mathbb{E}[OC(50)]$				$\mathbb{E}[OC(100)]$			
Function	$\sqrt{\lambda}$	KGCB	SKO	HKG	HHKG	KGCB	SKO	HKG	HHKG
SHCB-DS	1	**0.28**	**0.35**	**0.37**	0.55	**0.18**	0.30	0.29	0.33
	2	**0.56**	**0.53**	0.72	0.84	**0.38**	**0.41**	**0.48**	0.54
	4	**0.95**	**1.17**	**1.19**	**1.08**	**0.72**	**0.89**	**0.92**	**0.78**
SHCB-DB	1	**0.53**	0.70	**0.57**	**0.58**	**0.12**	0.53	0.41	0.35
	2	1.03	1.11	**0.73**	0.92	0.83	0.95	**0.46**	0.64
	4	1.55	1.50	**1.21**	**1.34**	1.33	1.42	**0.89**	**1.05**
SHCB-DS-SF	1	0.60	0.63	**0.32**	0.51	0.35	0.41	**0.20**	0.31
	2	0.90	0.95	**0.67**	**0.81**	0.69	0.86	**0.42**	**0.51**
	4	**1.17**	1.44	**1.13**	**1.22**	**1.05**	1.23	**0.86**	**0.89**
SHCB-DB_SF	1	1.19	0.75	**0.48**	**0.65**	0.60	0.81	**0.29**	**0.38**
	2	1.66	1.23	**0.69**	0.99	1.08	1.07	**0.48**	0.64
	4	1.85	1.41	**1.00**	**1.14**	1.36	1.43	**0.74**	**0.86**
TBRANIN	2	**0.16**	0.30	2.33	3.30	**0.08**	0.23	0.79	1.57
	4	**0.67**	1.21	2.40	4.12	**0.33**	0.85	1.16	2.27
	8	**3.64**	**2.88**	**3.81**	4.99	**1.29**	2.03	2.12	2.80
TBRANIN-SF	2	21.85	**1.42**	2.18	3.76	7.59	1.42	**0.82**	1.68
	4	10.61	**2.84**	**2.57**	4.55	3.17	1.99	**1.25**	2.22
	8	7.63	**5.01**	**4.07**	**4.50**	6.47	3.46	**2.33**	**2.48**

Table 7: Results for the 2-dimensional test functions.

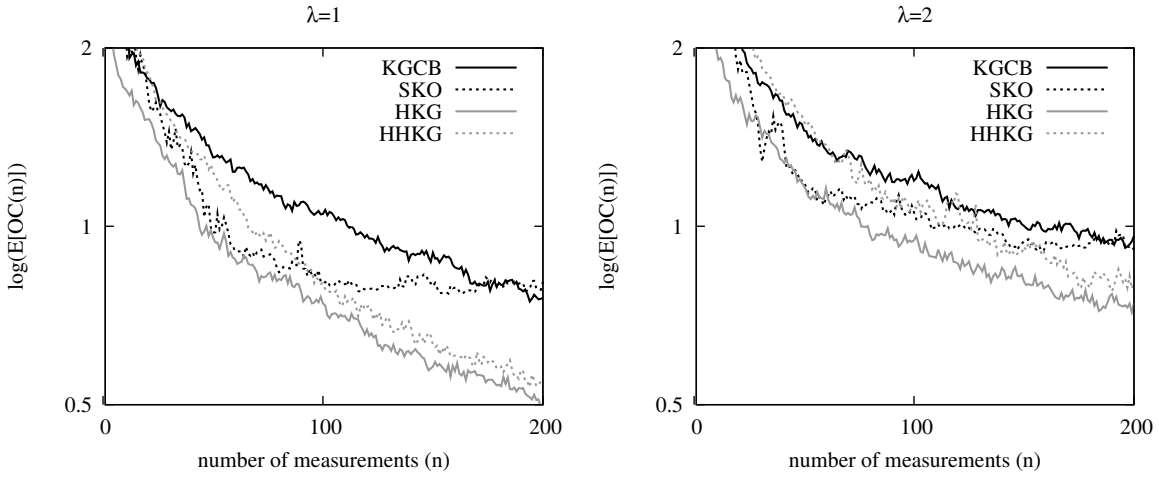

Figure 9: Results for the transportation application.

function to this surface, but we do require a family of aggregation functions. We produce estimates of the value of the function using a Bayesian adaptation of the hierarchical estimation procedure suggested by George et al. (2008). We then present an adaptation of the knowledge-gradient procedure of Frazier et al. (2009) for problems with correlated beliefs. That method requires the use of a known covariance matrix, while in our strategy, we compute covariances from our statistical model.

The hierarchical knowledge-gradient (HKG) algorithm shares the inherent steepest ascent property of the knowledge gradient algorithm, which chooses samples that produce the greatest single-sample improvement in our ability to maximize the function. We also prove that the algorithm is guaranteed to produce the optimal solution in the many-sample limit, since the HKG algorithm measures every alternative infinitely often.

We close with experimental results on a class of one and two dimensional scalar functions and a multi-attribute problem drawn from a transportation application. In these experiments, HKG performs better than all competing policies tested, when measured by average performance across all problems. In particular, it outperforms the other policies on functions for which the use of an aggregation function seems to be a natural choice (e.g., those with categorical dimensions), but it also performs well on problems for which the other policies are specifically designed.

The limitation of the HKG policy is that it requires a given aggregation structure, which means that we depend on having some insight into the problem. When this is the case, the ability to capture this knowledge in an aggregation structure is actually a strength, since we can capture the most important features in the highest levels of aggregation. If we do not have this insight, designing the aggregation functions imposes an additional modeling burden.

We mention two other limitations that give rise to further research. First, we observe convergence problems for HKG in the case of low measurement variance where HKG tends to become to confident about values of alternatives never measured before. We describe this issue in more detail in Appendix E. Second, the HKG policy requires enumerating all possible choices before determining the next measurement. This is appropriate for applications where we need to make good choices with a small number of measurements, typically far smaller than the set of alternatives. However, this limits our approach to handling perhaps thousands of choices, but not millions. A solution here would be to create a limited set of choices for the next measurement. As a starting point we might create this set by running HKG on a higher aggregation level which has fewer elements. Preliminary experiments have shown that this method can drastically reduce computation time without harming the performance too much. Future research could further explore such computational improvements.

We mention one final direction for future research. While we have presented a proof of convergence for the HKG policy, there are no theoretical results currently available that bound the rate at which it converges. Future research could derive such bounds, or could create new techniques appropriate for problems with hierarchical aggregation structures that have bounds on their convergence rates. One approach for creating such techniques would be to begin with an online learning technique with bounds on cumulative regret, and then to use a batch-to-online conversion technique to derive a procedure with a bound on the rate at which its terminal regret converges to zero.

Appendix A.

The overall sampling and updating procedure used for HKG is shown in Algorithm 1 and an outline for the HKG measurement decision is shown in Algorithm 2.

Algorithm 1 Sampling and updating procedure.

Require: Inputs $(G^g)\forall g \in \mathcal{G}$, $(\lambda_x)\forall x \in \mathcal{X}$, and $\underline{\delta}$

Initialize $(\mu_x^0, \beta_x^0, \tilde{g}_x^0)\forall x \in \mathcal{X}$, $(\mu_x^{g,0}, \beta_x^{g,0}, \delta_x^{g,0}, \beta_x^{g,0,\varepsilon})\forall g \in \mathcal{G}, x \in \mathcal{X}$
for $n = 1$ to N **do**
  Use Algorithm 2 to get measurement decision x^*
  Measure x^* and observe $\hat{y}_{x^*}^n$
  Compute $\tilde{g}_x^n \ \forall x \in \mathcal{X}$
  Compute $\mu_x^{g,n}$, $\beta_x^{g,n}$, and $\delta_x^{g,n}$ $\forall g \in \mathcal{G}, x \in \mathcal{X}$ using (2), (3), and (9)
  Compute $w_x^{g,n}$ with $(\sigma_x^{g,n})^2 = 1/\beta_x^{g,n}$ $\forall g \in \mathcal{G}, x \in \mathcal{X}$ using (8)
  Compute $\beta_x^{g,n,\varepsilon} = (\sigma_x^{g,n,\varepsilon})^{-2}$ $\forall g \in \mathcal{G}, x \in \mathcal{X}$ using (10)
  Compute μ_x^n and β_x^n with $(\sigma_x^{g,n})^2 = 1/\beta_x^{g,n}$ $\forall x \in \mathcal{X}$ using (4) and (5)
end for
return $x^N \in \arg\max_{x \in \mathcal{X}} \mu_x^N$

Algorithm 2 Hierarchical knowledge-gradient measurement decision.

Require: Inputs $(G^g)\forall g \in \mathcal{G}$, $(\lambda_x, \mu_x^n, \beta_x^n)\forall x \in \mathcal{X}$, $(\mu_x^{g,n}, \beta_x^{g,n}, \delta_x^{g,n}, \beta_x^{g,n,\varepsilon})\forall g \in \mathcal{G}, x \in \mathcal{X}$

for $x = 1$ to M **do**
  Compute $\tilde{\sigma}_x^{g,n}$ $\forall g \in \mathcal{G}$ using (15) with $(\sigma_x^n)^2 = 1/\beta_x^n$
  for $x' = 1$ to M **do**
    Compute $\bar{w}_{x'}^{g,n}(x)$ $\forall g \in \mathcal{G}$ using (17)
    Compute $a_{x'}^n(x)$ and $b_{x'}^n(x)$ using (19) and (20)
  end for
  Sort the sequence of pairs $(a_i^n(x), b_i^n(x))_{i=1}^M$ so that the $b_i^n(x)$ are in non-decreasing order and ties are broken so that $a_i^n(x) < a_{i+1}^n(x)$ if $b_i^n(x) = b_{i+1}^n(x)$.
  for $i = 1$ to $M-1$ **do**
    if $b_i^n(x) = b_{i+1}^n(x)$ **then**
      Remove entry i from the sequence $(a_i^n(x), b_i^n(x))_{i=1}^M$
    end if
  end for
  Use Algorithm 1 from Frazier et al. (2009) to compute $\tilde{a}_i^n(x)$ and $\tilde{b}_i^n(x)$
  Compute $\upsilon_x^{KG,n}$ using (21)
  if $x = 1$ or $\upsilon_x^{KG,n} \geq \upsilon^*$ **then**
    $\upsilon^* = \upsilon_x^{KG,n}$, $x^* = x$
  end if
end for
return x^*

Appendix B.

Proposition 4 *The posterior belief on θ_x given observations up to time n for all aggregation levels is normally distributed with mean and precision*

$$\begin{aligned}\mu_x^n &= \frac{1}{\beta_x^n}\left[\beta_x^0\mu_x^0 + \sum_{g\in\mathcal{G}}\left((\sigma_x^{g,n})^2+\nu_x^g\right)^{-1}\mu_x^{g,n}\right],\\ \beta_x^n &= \beta_x^0 + \sum_{g\in\mathcal{G}}\left((\sigma_x^{g,n})^2+\nu_x^g\right)^{-1}.\end{aligned}$$

Proof Let $Y_x^{g,n} = \left\{\hat{y}_{x^{m-1}}^{g,m} : m \le n, G^g(x) = G^g(x^{m-1})\right\}$. This is the set of observations from level g pertinent to alternative x.

Let H be a generic subset of $\mathcal{G}$. We show by induction on the size of the set H that the posterior on θ_x given $Y_x^{g,n}$ for all $g \in H$ is normal with mean and precision

$$\begin{aligned}\mu_x^{H,n} &= \frac{1}{\beta_x^{H,n}}\left[\beta_x^0\mu_x^0 + \sum_{g\in H}\left((\sigma_x^{g,n})^2+\nu_x^g\right)^{-1}\mu_x^{g,n}\right],\\ \beta_x^{H,n} &= \beta_x^0 + \sum_{g\in H}\left((\sigma_x^{g,n})^2+\nu_x^g\right)^{-1}.\end{aligned}$$

Having shown this statement for all H, the proposition follows by taking $H = \mathcal{G}$.

For the base case, when the size of H is 0, we have $H = \emptyset$ and the posterior on θ is the same as the prior. In this case the induction statement holds because $\mu_x^{H,n} = \mu_x^0$ and $\beta_x^{H,n} = \beta_x^0$.

Now suppose the induction statement holds for all H of a size m and consider a set H' with $m+1$ elements. Choose $g \in H'$ and let $H = H' \setminus \{g\}$. Then the induction statement holds for H because it has size m. Let $\mathbb{P}_H$ denote the prior conditioned on $Y_x^{g',n}$ for $g' \in H$, and define $\mathbb{P}_{H'}$ similarly. We show that the induction statement holds for H' by considering two cases: $Y_x^{g,n}$ empty and non-empty.

If $Y_x^{g,n}$ is empty, then the distribution of θ_x is the same under both $\mathbb{P}_H$ and $\mathbb{P}_{H'}$. Additionally, from the fact that $\sigma_x^{g,n} = \infty$ it follows that $\mu_x^{H,n} = \mu_x^{H',n}$ and $\beta_x^{H,n} = \beta_x^{H',n}$. Thus, the induction statement holds for H'.

Now consider the case that $Y_x^{g,n}$ is non-empty. Let φ be the normal density, and let y denote the observed value of $Y_x^{g,n}$. Then, by the definitions of H and H', and by Bayes rule,

$$\mathbb{P}_{H'}\{\theta_x \in du\} = \mathbb{P}_H\{\theta_x \in du \mid Y_x^{g,n} = y\} \propto \mathbb{P}_H\{Y_x^{g,n} \in dy \mid \theta_x = u\}\,\mathbb{P}_H\{\theta_x \in du\}.$$

The second term may be rewritten using the induction statement as $\mathbb{P}_H\{\theta_x \in du\} = \varphi\left((u-\mu_x^{H,n})/\sigma_x^{H,n}\right)$. The first term may be rewritten by first noting that $Y_x^{g,n}$ is independent of $Y_x^{g',n}$ for $g' \in H$, and then conditioning on θ_x^g. This provides

$$\begin{aligned}\mathbb{P}_H\{Y_x^{g,n} \in dy \mid \theta_x = u\} &= \mathbb{P}\{Y_x^{g,n} \in dy \mid \theta_x = u\}\\ &= \int_{\mathbb{R}} \mathbb{P}\{Y_x^{g,n} \in dy \mid \theta_x^g = v\}\,\mathbb{P}\{\theta_x^g = v \mid \theta_x = u\}\,dv\\ &\propto \int_{\mathbb{R}} \varphi\left(\frac{\mu_x^{g,n}-v}{\sigma_x^{g,n}}\right)\varphi\left(\frac{v-u}{\sqrt{\nu_x^g}}\right)dv\\ &\propto \varphi\left(\frac{\mu_x^{g,n}-u}{\sqrt{(\sigma_x^{g,n})^2+\nu_x^g}}\right).\end{aligned}$$

In the third line, we use the fact that $\mathbb{P}_H\left\{Y_x^{g,n} \in dy \mid \theta_x^g = v\right\}$ is proportional (with respect to u) to $\varphi\left((\mu_x^{g,n} - v)/\sigma_x^{g,n}\right)$, which may be shown by induction on n from the recursive definitions for $\mu_x^{g,n}$ and $\beta_x^{g,n}$.

Using this, we write

$$\mathbb{P}_{H'}\{\theta_x \in du\} \propto \varphi\left(\frac{u-\mu_x^{g,n}}{\sqrt{(\sigma_x^{g,n})^2+\nu_x^g}}\right)\varphi\left(\frac{u-\mu_x^{H,n}}{\sigma_x^{H,n}}\right) \propto \varphi\left(\frac{u-\mu_x^{H',n}}{\sigma_x^{H',n}}\right),$$

which follows from an algebraic manipulation that involves completing the square.

This shows that the posterior is normally distributed with mean $\mu_x^{H',n}$ and variance $(\sigma_x^{H',n})^2$, showing the induction statement. ■

Appendix C.

This appendix contains all the lemmas required in the proofs of Theorem 1 and Corollaries 2 and 3.

Lemma 5 *If $z_1, z_2, \ldots$ is a sequence of non-negative real numbers bounded above by a constant $a < \infty$, and $s_n = \sum_{k \le n} z_k$, then $\sum_n (z_n/s_n)^2 \mathbf{1}_{\{s_n > 0\}}$ is finite.*

Proof Let $n_0 = \inf\{n \ge 0 : s_n > 0\}$, and, for each integer k, let $n_k = \inf\{n \ge 0 : s_n > ka\}$. Then, noting that $s_n = 0$ for all $n < n_0$ and that $s_n > 0$ for all $n \ge n_0$, we have

$$\sum_n (z_n/s_n)^2 \mathbf{1}_{\{s_n>0\}} = \left[\sum_{n_0 \le n < n_1} (z_n/s_n)^2\right] + \sum_{k=1}^{\infty}\left[\sum_{n_k \le n < n_{k+1}} (z_n/s_n)^2\right].$$

We show that this sum is finite by showing that the two terms are both finite. The first term may be bounded by

$$\sum_{n_0 \le n < n_1} (z_n/s_n)^2 \le \sum_{n_0 \le n < n_1} (z_n/z_{n_0})^2 \le \left(\sum_{n_0 \le n < n_1} z_n/z_{n_0}\right)^2 \le (a/z_{n_0})^2 < \infty.$$

The second term may be bounded by

$$\begin{aligned}
\sum_{k=1}^{\infty}\sum_{n=n_k}^{n_{k+1}-1} (z_n/s_n)^2 \le \sum_{k=1}^{\infty}\sum_{n=n_k}^{n_{k+1}-1} (z_n/ka)^2 &\le \sum_{k=1}^{\infty}\left(\sum_{n=n_k}^{n_{k+1}-1} z_n/ka\right)^2 \\
= \sum_{k=1}^{\infty}\left(\frac{s_{n_{k+1}-1} - s_{n_k} + z_{n_k}}{ka}\right)^2 &\le \sum_{k=1}^{\infty}\left(\frac{(k+1)a - ka + a}{ka}\right)^2 \\
= \sum_{k=1}^{\infty} (2/k)^2 = \frac{2}{3}\pi^2 &< \infty.
\end{aligned}$$

■

Lemma 6 *Assume that samples from any fixed alternative x are iid with finite variance. Fix $g \in \mathcal{G}$ and $x \in \mathcal{X}$ and let*

$$\bar{y}_x^n = \left[\sum_{m<n} \beta_x^{g,m,\varepsilon} \hat{y}_x^{m+1} \mathbf{1}_{\{x^m=x\}}\right] \Big/ \left[\sum_{m<n} \beta_x^{g,m,\varepsilon} \mathbf{1}_{\{x^m=x\}}\right]$$

for all those n for which the denominator is strictly positive, and let $\bar{y}_x^n = 0$ for those n for which the denominator is zero. Then, $\sup_n |\bar{y}_x^n|$ is finite almost surely.

Proof Let $\alpha^n = \left[\beta_x^{g,n,\varepsilon} \mathbf{1}_{\{x^n=x\}}\right] / \left[\sum_{m\le n} \beta_x^{g,m,\varepsilon} \mathbf{1}_{\{x^m=x\}}\right]$, so that

$$\bar{y}_x^{n+1} = (1-\alpha^n)\bar{y}_x^n + \alpha^n \hat{y}_x^{n+1}.$$

Let v_x be the variance of samples from alternative x, which is assumed finite. Let $M^n = (\bar{y}_x^n - \theta_x)^2 + \sum_{m=n}^{\infty} \mathbf{1}_{\{x^m=x\}} v_x (\alpha^m)^2$, and note that Lemma 5 and the upper bound $(\min_{x'} \lambda_{x'})^{-1}$ on $\beta_x^{g,m,\varepsilon}$ together imply that M^0 is finite. We will show that M^n is a supermartingale with respect to the filtration generated by $(\hat{y}_x^n)_{n=1}^{\infty}$. In this proof, we write $\mathbb{E}^n$ to indicate $\mathbb{E}[\cdot \mid \mathcal{F}^n]$, the conditional expectation taken with respect to $\mathcal{F}^n$.

Consider $\mathbb{E}^n[M^{n+1}]$. On the event $\{x^n \neq x\}$ (which is $\mathcal{F}^n$ measurable), we have $M^{n+1} = M^n$ and $\mathbb{E}^n\left[M^{n+1} - M^n\right] = 0$. On the event $\{x^n = x\}$ we compute $\mathbb{E}^n\left[M^{n+1} - M^n\right]$ by first computing

$$\begin{aligned} M^{n+1} - M^n &= (\bar{y}_x^{n+1} - \theta_x)^2 - (\bar{y}_x^n - \theta_x)^2 - v_x(\alpha^n)^2 \\ &= ((1-\alpha^n)\bar{y}_x^n + \alpha^n \hat{y}_x^{n+1} - \theta_x)^2 - (\bar{y}_x^n - \theta_x)^2 - v_x(\alpha^n)^2 \\ &= -(\alpha^n)^2(\bar{y}_x^n - \theta_x)^2 + 2\alpha^n(1-\alpha^n)(\bar{y}_x^n - \theta_x)(\hat{y}_x^{n+1} - \theta_x) \\ &\quad + (\alpha^n)^2\left[(\hat{y}_x^{n+1} - \theta_x)^2 - v_x\right]. \end{aligned}$$

Then, the $\mathcal{F}^n$ measurability of α^n and $\bar{y}_x^n$, together with the facts that $\mathbb{E}^n\left[\hat{y}_x^{n+1} - \theta_x\right] = 0$ and $\mathbb{E}^n\left[\left(\hat{y}_x^{n+1} - \theta_x\right)^2\right] = v_x$, imply

$$\mathbb{E}\left[M^{n+1} - M^n\right] = -(\alpha^n)^2 (\bar{y}_x^n - \theta_x)^2 \le 0.$$

Since $M^n \ge 0$ and $M^0 < \infty$, the integrability of M^n follows. Thus, $(M^n)_n$ is a supermartingale and has a finite limit almost surely. Then,

$$\lim_{n\to\infty} M^n = \lim_{n\to\infty} (\bar{y}_x^n - \theta_x)^2 + \sum_{m=n}^{\infty} \mathbf{1}_{\{x^m=x\}} v_x (\alpha^m)^2 = \lim_{n\to\infty} (\bar{y}_x^n - \theta_x)^2.$$

The almost sure existence of a finite limit for $(\hat{y}_x^n - \theta_x)^2$ implies the almost sure existence of a finite limit for $|\hat{y}_x^n - \theta_x|$ as well. Finally, the fact that a sequence with a limit has a finite supremum implies that $\sup_n |\bar{y}_x^n| \le \sup_n |\bar{y}_x^n - \theta_x| + |\theta_x| < \infty$ almost surely. ■

Lemma 7 *Assume that samples from any fixed alternative x are iid with finite variance. Let $x, x' \in \mathcal{X}$, $g \in \mathcal{G}$. Then $\sup_n |\mu_{x'}^{g,n}|$ and $\sup_n |a_{x'}^n(x)|$ are almost surely finite.*

Proof We first show $\sup_n |\mu_x^{g,n}| < \infty$ almost surely for fixed x and g. We write $\mu_x^{g,n}$ as

$$\mu_x^{g,n} = \frac{\beta_x^{g,0}\mu_x^{g,0} + \sum_{m<n}\beta_x^{g,m,\varepsilon}\mathbf{1}_{\{x^m \in \mathcal{X}^g(x)\}}\hat{y}_{x^m}^{m+1}}{\beta_x^{g,0} + \sum_{m<n}\beta_x^{g,m,\varepsilon}\mathbf{1}_{\{x^m \in \mathcal{X}^g(x)\}}} = p_0^n \mu_x^{g,0} + \sum_{x' \in \mathcal{X}^g(x)} p_{x'}^n \bar{y}_{x'}^n,$$

where the $\bar{y}_{x'}^n$ are as defined in Lemma 6 and the $p_{x'}^n$ are defined for $x' \in \mathcal{X}^g(x)$ by

$$p_0^n = \frac{\beta_x^{g,0}}{\beta_x^{g,0} + \sum_{m<n}\beta_x^{g,m,\varepsilon}\mathbf{1}_{\{x^m \in \mathcal{X}^g(x)\}}}, \qquad p_{x'}^n = \frac{\sum_{m<n}\beta_x^{g,m,\varepsilon}\mathbf{1}_{\{x^m = x'\}}}{\beta_x^{g,0} + \sum_{m<n}\beta_x^{g,m,\varepsilon}\mathbf{1}_{\{x^m \in \mathcal{X}^g(x)\}}}.$$

Note that p_0^n and each of the $p_{x'}^n$ are bounded uniformly between 0 and 1. We then have

$$\sup_n |\mu_x^{g,n}| \le \sup_n \left[|\mu_x^{g,0}| + \sum_{x' \in \mathcal{X}^g(x)} |\bar{y}_{x'}^n| \right] \le |\mu_x^{0,g}| + \sum_{x' \in \mathcal{X}^g(x)} \sup_n |\bar{y}_{x'}^n|.$$

By Lemma 6, $\sup_n |\bar{y}_{x'}^n|$ is almost surely finite, and hence so is $\sup_n |\mu_x^{g,n}|$.

We now turn our attention to $a_{x'}^n(x)$ for fixed x and x'. $a_{x'}^n(x)$ is a weighted linear combinations of the terms $\mu_{x'}^{g,n}$, $g \in \mathcal{G}$ (note that $\mu_{x'}^n$ is itself a linear combination of such terms), where the weights are uniformly bounded. This, together with the almost sure finiteness of $\sup_n |\mu_{x'}^{g,n}|$ for each g, implies that $\sup_n |a_{x'}^n(x)|$ is almost surely finite. ■

Lemma 8 *Assume that $\underline{\delta} > 0$ and samples from any fixed alternative x are iid with finite variance. Let $\mathcal{X}_\infty$ be the (random) set of alternatives measured infinitely often by HKG. Then, for each $x', x \in \mathcal{X}$, the following statements hold almost surely,*

- *If $x \in \mathcal{X}_\infty$ then $\lim_{n\to\infty} b_{x'}^n(x) = 0$ and $\lim_{n\to\infty} b_x^n(x') = 0$.*
- *If $x \notin \mathcal{X}_\infty$ then $\liminf_{n\to\infty} b_x^n(x) > 0$.*

Proof Let x' and x be any pair of alternatives.

First consider the case $x \in \mathcal{X}_\infty$. Let $g \in \mathcal{G}(x', x)$ and $B = \sup_n (\sigma_x^{g,n,\varepsilon})^2$. Lemma 7 and (10) imply that B is almost surely finite. Since $\beta_x^{g,n,\varepsilon} \ge 1/B$ for each n, we have $\beta_x^{g,n} \ge m_x^{g,n} B$. Then $x \in \mathcal{X}_\infty$ implies $\lim_{n\to\infty} m_x^{g,n} = \infty$ and $\lim_{n\to\infty} \beta_x^{g,n} = \infty$. Also, x and x' share aggregation level g, so $\beta_x^{g,n} = \beta_{x'}^{g,n}$ and $\lim_{n\to\infty} \beta_{x'}^{g,n} = \infty$. Then consider $\tilde{\sigma}_x^{g,n}$ for n large enough that we have measured alternative x at least once. From (10), $(\sigma_x^{g,n,\varepsilon})^2 \ge \lambda_x / |\mathcal{X}^g(x)|$, which gives a uniform upper bound $\beta_x^{g,n,\varepsilon} \le |\mathcal{X}^g(x)|/\lambda_x$. Also, the definition (6) implies $(\sigma_x^n)^2 \le (\sigma_x^{g,n})^2 \le 1/B$. This, the definition (15), and $\lim_{n\to\infty} \beta_x^{g,n} = \infty$ together imply $\lim_{n\to\infty} \tilde{\sigma}_x^{g,n} = 0$. The limit $\lim_{n\to\infty} \tilde{\sigma}_{x'}^{g,n} = 0$ follows similarly from the bounds $\beta_{x'}^{g,n,\varepsilon} \le |\mathcal{X}^g(x)|/\lambda_{x'}$ and $(\sigma_{x'}^n)^2 \le (\sigma_x^{n,g})^2 \le 1/B$, and $\lim_{n\to\infty} \beta_{x'}^{g,n} = \infty$. Hence, (20) and the boundedness of the weights $\bar{w}_{x'}^{g,n}$ and $\bar{w}_x^{g,n}$ imply $\lim_{n\to\infty} b_{x'}^n(x) = \lim_{n\to\infty} b_x^n(x') = 0$.

Now consider the case $x \notin \mathcal{X}_\infty$. We show that $\liminf_{n\to\infty} b_x^n(x) > 0$. From (20) and $0 \in \mathcal{G}(x,x)$,

$$b_x^n(x) \ge \bar{w}_x^{0,n}(x) \frac{(\lambda_x)^{-1}\sqrt{\left(\sum_{g' \in \mathcal{G}} \beta_x^{g',n}\right)^{-1} + \lambda_x}}{\beta_x^{0,n} + (\lambda_x)^{-1}}.$$

Because $x \notin \mathcal{X}_\infty$, there is some random time $N_1 < \infty$ after which we do not measure x, and $\beta_x^{0,n} \leq \beta_x^{N_1,0}$ for all n.

$$b_x^n(x) \geq \bar{w}_x^{0,n}(x) \frac{(\lambda_x)^{-1}\sqrt{\lambda_x}}{\beta_x^{0,N_1} + (\lambda_x)^{-1}},$$

where the weights are given by

$$\bar{w}_x^{0,n}(x) = \frac{\left(\beta_x^{0,n} + (\lambda_x)^{-1}\right)^{-1}}{\left(\beta_x^{0,n} + (\lambda_x)^{-1}\right)^{-1} + \sum_{g \in \mathcal{G}\setminus\{0\}} \psi_x^{g,n}},$$

with

$$\psi_x^{g,n} = \left((\beta_x^{g,n} + \beta_x^{g,n,\varepsilon})^{-1} + (\delta_x^{g,n})^2\right)^{-1}.$$

We now show $\limsup_n \psi_x^{g,n} < \infty$ for all $g \in \mathcal{G}\setminus\{0\}$. We consider two cases for g. In the first case, suppose that an alternative in $\mathcal{X}^g(x)$ is measured at least once. Then, for all n after this measurement, $m_x^{g,n} > 0$ and $\delta_x^{g,n} \geq \underline{\delta}$ (by (9)), implying $\psi_x^{g,n} \leq \underline{\delta}^{-2}$ and $\limsup_n \psi_x^{g,n} \leq \underline{\delta}^{-2} < \infty$. In the second case, suppose no alternative in $\mathcal{X}^g(x)$ is ever measured. Then, $\limsup_n \psi_x^{g,n} \leq \limsup_n \beta_x^{g,n} + \beta_x^{g,n,\varepsilon} < \infty$.

Finally, $\limsup_n \psi_x^{g,n} < \infty$ and $\left(\beta_x^{0,n} + (\lambda_x)^{-1}\right)^{-1} \geq \left(\beta_x^{0,N_1} + (\lambda_x)^{-1}\right)^{-1} > 0$ together imply $\liminf_{n\to\infty} \bar{w}_x^{0,n}(x) > 0$. This shows $\liminf_{n\to\infty} b_x^n(x) > 0$. ■

Lemma 9 *Let $a \in \mathbb{R}^d$ with $\max_i |a_i| \leq c$, $b \in \mathbb{R}^d$, and let Z be a standard normal random variable. If $x \neq x'$, then,*

$$\mathbb{E}\left[\max_i a_i + b_i Z\right] - \max_i a_i \geq \frac{|b_{x'} - b_x|}{2} f\left(\frac{-4c}{|b_{x'} - b_x|}\right),$$

where this expression is understood to be 0 if $b_{x'} = b_x$.

Proof Let $x^* \in \arg\max_i a_i$ and $a^* = \max_i a_i$. Then adding and subtracting $a_{x^*} + b_{x^*} Z = a^* + b_{x^*} Z$ and observing $\mathbb{E}[b_{x^*} Z] = 0$ provides

$$\begin{aligned}\mathbb{E}\left[\max_i a_i + b_i Z\right] - a^* &= \mathbb{E}\left[\left(\max_i (a_i - a^*) + (b_i - b_{x^*})Z\right) + a^* + b_{x^*} Z\right] - a^* \\ &= \mathbb{E}\left[\max_i (a_i - a^*) + (b_i - b_{x^*})Z\right].\end{aligned}$$

Let $j \in \arg\max_{i \in \{x,x'\}} |b_i - b^*|$. Then, by taking the maximum in the previous expression over only j and x^*, we obtain the lower bound

$$\begin{aligned}\mathbb{E}\left[\max_i a_i + b_i Z\right] - a^* &\geq \mathbb{E}\left[\max\left(0, a_j - a^* + (b_j - b_{x^*})Z\right)\right] \\ &\geq \mathbb{E}\left[\max\left(0, -2c + (b_j - b_{x^*})Z\right)\right] \\ &= |b_j - b_{x^*}| f\left(\frac{-2c}{|b_j - b_{x^*}|}\right) \geq \frac{|b_{x'} - b_x|}{2} f\left(\frac{-4c}{|b_{x'} - b_x|}\right).\end{aligned}$$

The second line follows from the bound $\max_i |a_i| \leq c$. The equality in the third line can be verified by evaluating the expectation analytically (see, e.g., Frazier et al., 2008), where the expression is taken to be 0 if $b_j = b_{x^*}$. The inequality in the third line then follows from $|b_j - b^*| \geq |b_x - b_{x'}|/2$ and from f being an increasing non-negative function. ■

Appendix D.

Here we provide a brief description of the implementation of the policies considered in our numerical experiments.

Interval estimation (IE) The IE decision rule by Kaelbling (1993) is given by

$$x^n = \arg\max_{x \in \mathcal{X}} \left(\mu_x^n + z_{\alpha/2} \cdot \sigma_x^n\right)$$

where $z_{\alpha/2}$ is a tunable parameter. Kaelbling (1993) suggests that values of 2, 2.5 or 3 often works best. The IE policy is quite sensitive to this parameter. For example, we observe that the following cases require higher values for $z_{\alpha/2}$: more volatile functions (low values for ρ, see Section 6.2), a higher measurement variance λ, and higher measurement budget N. To find a value that works reasonably well on most problem instances, we tested values between 0.5 and 4 with increments of .1 and found that $z_{\alpha/2} = 2.3$ works best on average. Since we assume the measurement noise is known, we use $\sigma_x^n = \sqrt{\frac{\lambda}{m_x^n}}$, where m_x^n is the number of times x has been measured up to and including time n.

UCB1-Normal (UCB1) The study by Auer et al. (2002) proposes different variations of the Upper Confidence Bound (UCB) decision rule originally proposed by Lai (1987). The UCB1-Normal policy is proposed for problems with Gaussian rewards and is given by

$$x^n = \arg\max_{x \in \mathcal{X}} \left(\mu_x^n + 4\sqrt{\frac{\lambda \log n}{N_x^n}}\right).$$

The original presentation of the policy uses a frequentist estimate of the measurement variance λ, which we replace by the known value. We improve the performance of UCB1 by treating the coefficient 4 as a tunable parameter. As with IE, we observe that the performance is quite sensitive to the value of this parameter. Using a setup similar to IE, we found that a value of 0.9 produced the best results on average.

Independent KG (IKG) This is the knowledge-gradient policy as presented in Section 4.1 of this paper.

Boltzmann exploration (BOLTZ) Boltzmann exploration chooses its measurements by

$$\mathbb{P}(x^n = x) = \left(\frac{e^{\mu_x^n / T^n}}{\sum_{x' \in \mathcal{X}} e^{\mu_{x'}^n / T^n}}\right),$$

where the policy is parameterized by a decreasing sequence of "temperature" coefficients $(T^n)_{n=0}^{N-1}$. We tune this temperature sequence within the set of exponentially decreasing sequences defined by $T^{n+1} = \gamma T^n$ for some constant $\gamma \in (0,1]$. The set of all such sequences is parameterized by γ and T^N. We tested combinations of $\gamma \in \{.1,.2,...,1\}$ and $T^N \in \{.1,.5,1,2\}$ and found that the combination $\gamma = 1$ and $T^N = .3$ produces the best results on average.

Pure exploration (EXPL) The pure exploration policy measures each alternative x with the same probability, that is, $\mathbb{P}(x^n = x) = 1/M$.

Sequential Kriging Optimization (SKO) This is a blackbox optimization method from Huang et al. (2006) that fits a Gaussian process onto the observed variables. The hyperparameters of the Gaussian process prior are estimated using an initial Latin hypercube design with $2p+2$ measurements, with p being the number of dimensions, as recommended by Huang et al. (2006). After this initial phase we continue to update the hyperparameters, using maximum likelihood estimation, during the first 50 measurements. The parameters are updated at each iteration.

KG for Correlated Beliefs (KGCB) This is the knowledge-gradient policy for correlated beliefs as presented in Section 4.1. We estimate the hyperparameters in the same way as done with SKO.

Hierarchical KG (HKG) This is the hierarchical knowledge-gradient policy as presented in this paper. This policy only requires an aggregation function as input. We present these functions in Section 6.3.

Hybrid HKG (HHKG) In this hybrid policy, we only exploit the similarity between alternatives in the updating equations and not in the measurement decision. As a result, this policy uses the measurement decision of IKG and the updating equations of HKG. The possible advantage of this hybrid policy is that it is able to cope with similarity between alternatives without the computational complexity of HKG.

Appendix E.

Here we show more detailed results for the experiments on one-dimensional problems. A complete overview of the results for the one-dimensional experiments with $N = 500$ can be found in Table 8 and with $N = 200$ in Table 9.

Besides the conclusions from the main text, we mention a few additional observations based on the more detailed results.

First, from Table 9 we see that the relative performance of KGCB and SKO depends on the value of ρ. On relatively smooth functions with $\rho \geq 2$, SKO outperforms KGCB, whereas the opposite holds for $\rho < 2$.

Second, it is remarkable to see that in the independent truth case (IT), the policies that exploit correlation (KGCB and HKG) are doing so well and outperform IKG. The explanation is the following. After M measurements, IKG has sampled each alternative once and the implementation decision is the one with the highest value observed so far. Obviously, this is not a reliable estimate, especially with $\lambda \geq 0.5$. The policies KGCB and HKG tend to resample promising alternatives. So, after M measurements, they have a more reliable estimate for their implementation decision.

Function	$\sqrt{\lambda}$	N	EXPL	IKG	KGCB	SKO	HKG	IE	UCB	BOLTZ
GP1R05	0.5	250	0.206	0.090	0.061	**0.029**	0.072	0.077	0.073	0.133
		500	0.169	0.044	0.037	**0.027**	0.053	0.038	0.040	0.075
	1	250	0.344	0.170	0.131	0.142	**0.111**	0.174	0.183	0.242
		500	0.332	0.108	**0.093**	0.111	**0.092**	0.106	0.113	0.155
GP1R02	0.5	250	0.152	0.041	**0.024**	**0.024**	0.032	0.046	0.043	0.069
		500	0.106	0.022	**0.014**	0.019	**0.017**	0.024	0.025	0.048
	1	250	0.308	0.103	**0.084**	0.129	**0.077**	0.112	0.111	0.151
		500	0.298	0.057	**0.050**	0.120	**0.044**	0.062	0.061	0.113
GP1R01	0.5	250	0.196	0.057	**0.019**	0.038	0.043	0.043	0.053	0.088
		500	0.158	0.033	**0.009**	0.024	0.027	0.022	0.024	0.058
	1	250	0.424	0.162	**0.107**	0.218	**0.114**	0.138	0.166	0.192
		500	0.348	0.084	**0.064**	0.165	**0.069**	0.069	0.088	0.143
GP1R005	0.5	250	0.253	0.065	**0.017**	0.047	0.049	0.053	0.058	0.100
		500	0.183	0.027	**0.008**	0.037	0.031	0.019	0.019	0.070
	1	250	0.483	0.162	**0.093**	0.189	**0.100**	0.145	0.178	0.210
		500	0.432	0.084	**0.046**	0.147	0.061	0.073	0.080	0.143
NSGP	0.5	250	0.249	0.052	0.070	0.146	**0.049**	**0.046**	**0.043**	0.122
		500	0.186	0.024	0.044	0.121	0.026	**0.019**	**0.019**	0.076
	1	250	0.539	0.193	0.184	0.240	**0.124**	0.150	0.175	0.220
		500	0.443	0.092	0.113	0.194	**0.067**	**0.068**	**0.073**	0.141
IT	0.5	250	0.182	0.075	0.066	0.107	**0.060**	0.075	0.074	0.113
		500	0.153	0.047	0.045	0.092	**0.040**	**0.042**	0.046	0.093
	1	250	0.306	0.155	0.144	0.207	**0.108**	0.151	0.162	0.188
		500	0.253	0.097	0.101	0.188	**0.087**	0.094	0.099	0.168
GP1	0.5	250	0.202	0.063	**0.030**	0.034	0.049	0.055	0.057	0.098
		500	0.154	0.032	**0.017**	0.027	0.032	0.026	0.027	0.063
	1	250	0.390	0.149	**0.104**	0.170	**0.101**	0.143	0.160	0.198
		500	0.352	0.083	**0.063**	0.136	**0.066**	0.078	0.086	0.138
NS0	0.5	250	0.215	0.064	0.068	0.126	**0.055**	0.060	**0.059**	0.118
		500	0.169	0.035	0.044	0.106	**0.033**	**0.031**	**0.032**	0.085
	1	250	0.423	0.174	0.164	0.224	**0.116**	0.150	0.168	0.204
		500	0.348	0.094	0.107	0.191	**0.077**	**0.081**	0.086	0.154

Table 8: Results for the one-dimensional long experiments.

Function	$\sqrt{\lambda}$	N	EXPL	IKG	KGCB	SKO	HKG	HHKG
GP1R05	0.1	50	0.149	0.131	0.020	**0.001**	0.033	0.036
		200	0.102	0.008	0.006	**0.001**	0.008	0.008
	0.5	50	0.261	0.231	0.165	**0.078**	0.171	0.169
		200	0.216	0.097	0.075	**0.036**	0.085	0.080
	1	50	0.390	0.411	0.277	**0.210**	0.258	0.278
		200	0.359	0.222	0.150	0.148	**0.129**	0.162
GP1R02	0.1	50	0.039	0.038	0.010	**0.005**	0.026	0.050
		200	0.025	0.008	**0.003**	**0.002**	0.007	0.006
	0.5	50	0.203	0.187	0.079	**0.063**	0.092	0.126
		200	0.169	0.055	**0.029**	**0.029**	0.037	0.044
	1	50	0.396	0.389	**0.233**	**0.230**	**0.224**	0.257
		200	0.332	0.142	**0.096**	0.138	**0.097**	**0.087**
GP1R01	0.1	50	0.062	0.056	**0.007**	0.014	0.030	0.083
		200	0.036	0.006	**0.001**	0.008	0.008	0.005
	0.5	50	0.254	0.253	**0.121**	**0.117**	**0.132**	0.184
		200	0.218	0.065	**0.022**	0.043	0.055	0.054
	1	50	0.477	0.482	**0.303**	0.358	**0.294**	**0.283**
		200	0.441	0.182	**0.124**	0.235	**0.136**	**0.128**
GP1R005	0.1	50	0.111	0.099	**0.003**	0.011	0.047	0.144
		200	0.043	0.004	**0.000**	0.003	0.008	0.011
	0.5	50	0.342	0.336	**0.127**	0.157	0.170	0.222
		200	0.254	0.082	**0.021**	0.054	0.061	0.080
	1	50	0.577	0.482	0.329	0.411	**0.286**	0.401
		200	0.530	0.182	**0.118**	0.204	**0.123**	0.164
NSGP	0.1	50	0.168	0.143	0.087	0.135	**0.059**	0.184
		200	0.047	**0.003**	0.021	0.094	**0.005**	0.017
	0.5	50	0.391	0.373	0.235	0.265	**0.200**	0.294
		200	0.263	0.082	0.084	0.156	**0.066**	0.082
	1	50	0.692	0.627	0.428	0.451	**0.381**	0.440
		200	0.580	0.249	0.208	0.260	**0.153**	0.176
IT	0.1	50	0.053	0.050	**0.046**	0.052	**0.044**	**0.042**
		200	0.039	0.013	**0.012**	0.027	0.013	**0.011**
	0.5	50	0.212	0.203	0.144	0.178	0.141	**0.130**
		200	0.175	0.091	0.072	0.116	**0.065**	0.079
	1	50	0.305	0.310	**0.218**	0.298	**0.230**	**0.231**
		200	0.312	0.177	0.157	0.217	**0.128**	0.150

Table 9: Results for the one-dimensional normal experiments.

Obviously, there is a probability that KGCB and HKG do not measure the true optimal alternative after M measurements. However, given the way we generated this function, there are multiple alternatives close the the optimal one (we may expect 10% of the alternatives to be less then 0.1 from the optimum).

Finally, even though HKG seems to be quite competitive, there are some results that suggest future extensions of HKG. Specifically, HKG seems to have convergence problems in the low noise case ($\lambda = 0.1$). We see this from (i) the settings with $\lambda = 0.1$ and $n = 200$ where HKG underperforms IKG on three cases (two of them with significant differences), (ii) the settings with the one-dimensional long experiments where HKG is outperformed by IKG in three cases, each of them having a low value for λ and a large number of measurements, and (iii) the hybrid policy HHKG is outperformed by IKG on most of the $\lambda = 0.1$ cases. We believe that the source of this problem lies in the use of the base level $\tilde{g}_x^n$, that is, the lowest level g for which we have at least one observation on an aggregate alternative that includes alternative x ($m_x^{g,n} > 0$). We introduced this base level because we need the posterior mean μ_x^n and the posterior variance $(\sigma_x^n)^2$ for all alternatives, including those we have not measured. When λ is relatively small, the posterior variance on the aggregate levels $(\sigma_x^{g,n})^2$ increases relatively quickly; especially because the squared bias $(\delta_x^{g,n})^2$, which we use as an estimate for ν_x^g, is small at the base level (equal to the lower bound $\underline{\delta}$). As a result, we may become too confident about the value of an alternative we never measured. We may be able to resolve this by adding a prior on these functions, which obviously requires prior knowledge about the truth or additional measurements, or by tuning $\underline{\delta}$.

References

Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, pages 263–274, 2008.

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.

Peter L. Bartlett, Varsha Dani, Thomas P. Hayes, Sham Kakade, Alexander Rakhlin, and Ambuj Tewari. High-probability regret bounds for bandit online linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, pages 335–342, 2008.

Russell R. Barton and Martin Meckesheimer. Metamodel-based simulation optimization. In Shane G. Henderson and Barry L. Nelson, editors, *Simulation*, volume 13 of *Handbooks in Operations Research and Management Science*, pages 535 – 574. Elsevier, 2006.

Robert E. Bechhofer. A single-sample multiple decision procedure for ranking means of normal populations with known variances. *The Annals of Mathematical Statistics*, 25(1):16–39, 1954.

Robert E. Bechhofer, Thomas J. Santner, and David M. Goldsman. *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons*. John Wiley & Sons, New York, NY, 1995.

Dimitri P. Bertsekas and David A. Castanon. Adaptive aggregation methods for infinite horizon dynamic programming. *IEEE Transactions on Automatic Control*, 34(6):589–598, 1989.

Dimitri P. Bertsekas and John N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA, 1996.

Ronen I. Brafman and Moshe Tennenholtz. R-MAX - a general polynomial time algorithm for near-optimal reinforcement learning. *Journal of Machine Learning Research*, 3:213–231, 2003.

Franklin H. Branin. Widely convergent method for finding multiple solutions of simultaneous nonlinear equations. *IBM Journal of Reseach and Development*, 16(5):504–522, 1972.

Erik Brochu, Mike Cora, and Nando de Freitas. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. Technical Report TR-2009-023, Department of Computer Science, University of British Columbia, 2009.

Sébastien Bubeck, Réemi Munos, Gilles Stoltz, and Csaba Szepesvári. Online optimization in X-armed bandits. In *Advances in Neural Information Processing Systems*, pages 201–208, 2009a.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *Proceedings of the 20th International Conference on Algorithmic Learning Theory*, pages 23–37, 2009b.

Chun-Hung Chen, Hsiao-Chang Chen, and Liyi Dai. A gradient approach for smartly allocating computing budget for discrete event simulation. In *Proceedings of the 28th Conference on Winter Simulation*, pages 398–405, 1996.

Stephen E. Chick and Koichiro Inoue. New two-stage and sequential procedures for selecting the best simulated system. *Operations Research*, 49(5):732–743, 2001.

Stephen E. Chick, Jurgen Branke, and Christian Schmidt. Sequential sampling to myopically maximize the expected value of information. *INFORMS Journal on Computing*, 22(1):71–80, 2010.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. PAC bounds for multi-armed bandit and Markov decision processes. In *Proceedings of the 15th Annual Conference on Computational Learning Theory (COLT)*, pages 193–209, 2002.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for reinforcement learning. In *Proceedings of the 20th International Conference on Machine Learning*, pages 162–169, 2003.

Abraham D. Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, pages 385–394, 2005.

Peter I. Frazier, Warren B. Powell, and Savas Dayanik. A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439, 2008.

Peter I. Frazier, Warren B. Powell, and Savas Dayanik. The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, 21(4):599–613, 2009.

Abraham George, Warren B. Powell, and Sanjeev R. Kulkarni. Value function approximation using multiple aggregation for multiattribute resource management. *Journal of Machine Learning Research*, 9:2079–2111, 2008.

Mark N. Gibbs. *Bayesian Gaussian Processes for Regression and Classification*. PhD thesis, University of Cambridge, 1997.

Steffen Grünewälder, Jean-Yves Audibert, Manfred Opper, and John Shawe-Taylor. Regret bounds for gaussian process bandit problems. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2010.

Shanti S. Gupta and Klaus J. Miescke. Bayesian look ahead one-stage sampling allocations for selection of the best population. *Journal of Statistical Planning and Inference*, 54(2):229–244, 1996.

Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman. *The Elements of Statistical Learning*. Springer series in Statistics, New York, NY, 2001.

Donghai He, Stephen E. Chick, and Chun-Hung Chen. Opportunity cost and OCBA selection procedures in ordinal optimization for a fixed number of alternative systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(5):951–961, 2007.

Deng Huang, Theodore T. Allen, William I. Notz, and Ning Zheng. Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of Global Optimization*, 34 (3):441–466, 2006.

Frank Hutter. *Automated Configuration of Algorithms for Solving Hard Computational Problems*. PhD thesis, University of British Columbia, 2009.

Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998.

Leslie P. Kaelbling. *Learning In Embedded Systems*. MIT Press, Cambridge, MA, 1993.

Michael Kearns and Satinder Singh. Near-optimal reinforcement learning in polynomial time. *Machine Learning*, 49(2-3):209–232, 2002.

Seong-Hee Kim and Barry L. Nelson. *Handbook in Operations Research and Management Science: Simulation*, chapter Selecting the best system. Elsevier, Amsterdam, 2006.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the 40th Annual ACM symposium on Theory of Computing*, pages 681–690, 2008.

Robert D. Kleinberg. *Online Decision Problems With Large Strategy Sets*. PhD thesis, MIT, 2005.

Harold J. Kushner. A new method of locating the maximum of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering*, 86:97–106, 1964.

Tze L. Lai. Adaptive treatment allocation and the multi-armed bandit problem. *The Annals of Statistics*, 15(3):1091–1114, 1987.

Michael LeBlanc and Robert Tibshirani. Combining estimates in regression and classification. *Journal of the American Statistical Association*, 91(436):1641–1650, 1996.

Nick Littlestone. From on-line to batch learning. In *Proceedings of the Second Annual Workshop on Computational learning theory*, pages 269–284, 1989.

Daniel J. Lizotte. *Practical Bayesian Optimization*. PhD thesis, University of Alberta, 2008.

Omid Madani, Daniel J. Lizotte, and Russell Greiner. The budgeted multi-armed bandit problem. In *Proceedings of the 17th Annual Conference on Computational Learning Theory (COLT)*, pages 643–645, 2004.

Volodymyr Mnih, Csaba Szepesvári, and Jean-Yves Audibert. Empirical Bernstein stopping. In *Proceedings of the 25th International Conference on Machine Learning*, pages 672–679, 2008.

Jonas Mockus. On Bayesian methods for seeking the extremum. In G. Marchuk, editor, *Optimization Techniques IFIP Technical Conference Novosibirsk, July 17, 1974*, volume 27 of *Lecture Notes in Computer Science*, pages 400–404. Springer Berlin / Heidelberg, 1975.

Warren B. Powell and Peter I. Frazier. Optimal learning. In *TutORials in Operations Research*, pages 213–246. INFORMS, 2008.

Carl E. Rasmussen and Christopher Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22(3):400–407, 1951.

David F. Rogers, Robert D. Plante, Richard T. Wong, and James R. Evans. Aggregation and disaggregation techniques and methodology in optimization. *Operations Research*, 39(4):553–582, 1991.

Michael J. Sasena. *Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations*. PhD thesis, University of Michigan, 2002.

Shai Shalev-Shwartz. *Online learning: Theory, algorithms, and applications*. PhD thesis, The Hebrew University of Jerusalem, 2007.

Hugo P. Simao, Jeff Day, Abraham P. George, Ted Gifford, John Nienow, and Warren B. Powell. An approximate dynamic programming algorithm for large-scale fleet management: A case application. *Transportation Science*, 43(2):178–197, 2009.

Tom A.B. Snijders and Roel J. Bosker. *Multilevel Analysis: An Introduction To Basic And Advanced Multilevel Modeling*. Sage Publications Ltd, 1999.

James C. Spall. *Introduction to Stochastic Search and Optimization*. Wiley-Interscience, Hoboken, NJ, 2003.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings International Conference on Machine Learning (ICML)*, 2010.

Emmanuel Vazquez and Julien Bect. Convergence properties of the expected improvement algorithm with fixed mean and covariance functions. *Journal of Statistical Planning and Inference*, 140(11):3088–3095, 2010.

Julien Villemonteix, Emmanuel Vazquez, and Eric Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509–534, 2009.

Yuhong Yang. Adaptive regression by mixing. *Journal of American Statistical Association*, 96 (454):574–588, 2001.