Hierarchical knowledge-gradient for sequential sampling

Citation for published version (APA):
Mes, M. R. K., Powell, W. B., & Frazier, P. I. (2009). *Hierarchical knowledge-gradient for sequential sampling*. (BETA publicatie : working papers; Vol. 290). Technische Universiteit Eindhoven.

Document status and date:
Published: 01/01/2009

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Hierarchical Knowledge-Gradient for Sequential Sampling

Martijn R.K. Mes[1], Warren B. Powell[2], Peter I. Frazier[1]

[1]Department of Operational Methods for Production and Logistics
University of Twente, Enschede, The Netherlands
[2]Department of Operations Research and Financial Engineering
Princeton University, Princeton, USA
[3]Department of Operations Research and Information Engineering
Cornell University, Ithaca, USA

October 7, 2009

Abstract

We consider the problem of selecting the best of a finite but very large set of alternatives. Each alternative may be characterized by a multi-dimensional vector and has independent normal rewards. This problem arises in various settings such as (i) ranking and selection, (ii) simulation optimization where the unknown mean of each alternative is estimated with stochastic simulation output, and (iii) approximate dynamic programming where we need to estimate values based on Monte-Carlo simulation.

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. Because the number of alternatives is large, we propose a hierarchical aggregation technique that uses the common features shared by alternatives to learn about many alternatives from even a single measurement, thus greatly reducing the measurement effort required. We demonstrate how this hierarchical knowledge-gradient policy can be applied to efficiently maximize a continuous function and prove that this policy finds a globally optimal alternative in the limit.

Keywords: sequential decision analysis, 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 that produces the fastest rate of learning, so that we can find the best value within a finite budget. 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 a family of aggregation functions $G^g : \mathcal{X} \rightarrow \mathcal{X}^g$, $g \in \mathcal{G}$, each of which maps $\mathcal{X}$ to a region $\mathcal{X}^g$, which is successively smaller than the original set of alternatives. We assume $g = 0$ corresponds to no aggregation, i.e., $\mathcal{X}^0 = \mathcal{X}$ and $G^0(x) = x$, which might be the finest discretization of a continuous set of alternatives. We do not require the aggregations to be hierarchical, although this will be common in practice. After each observation $\hat{y}^n_x = \theta_x + \epsilon^n$, we update a family of statistical estimates of θ at each level of aggregation. After n observations, we obtain a family of estimates $\mu^{g,n}_x$ of the function at different levels of aggregation, and we form an estimate μ^n_x of θ_x using

$$\mu^n_x = \sum_{g \in \mathcal{G}} w^{g,n}_x \mu^{g,n}_x, \tag{1}$$

where the weights $w^{g,n}_x$ sum to one over all the levels of aggregation for each point x. The estimates $\mu^{g,n}_x$ at more aggregate levels have lower statistical variance since they are based upon more observations, but will exhibit aggregation bias. The estimates $\mu^{g,n}_x$ 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 generalizes the ranking and selection problem to a very large set of potential alternatives. Rather than listing the alternatives $(1, 2, \ldots, M)$, we retain the multi-dimensional structure of an alternative x as a vector $(x_1, \ldots, x_D)$. This version of the 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 to maximize profits. We can also consider problems where x is a multi-dimensional set of continuous parameters. We will assume the parameters can be discretized to an arbitrarily fine level.

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 that exploits correlations in the belief structure, but where these correlations are assumed known, and where the number of alternatives is not too large. Instead of using a multivariate normal belief, 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 in (George et al., 2008).

This paper makes the following contributions. First, we extend the knowledge-gradient policy to problems where an alternative is described by a multi-dimensional vector in a computationally feasible way. We estimate a function using an appropriately weighted sum of estimates at different levels of aggregation. Second, we propose a version of the knowledge gradient that exploits aggregation structure and similarity between alternatives, without requiring that we specify an explicit covariance matrix for our belief. Instead, relationships between alternatives' values are derived from the structure of the weighted estimates. In addition to eliminating the difficulty of specifying an a priori covariance matrix, this avoids the computational challenge of working with large covariance matrices. Third, we show that a learning policy based on this method is optimal in the limit, i.e., 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 the numerical experiments in Section 6. We close with conclusions, remarks on generalizations, and directions for further research in Section 7.

2 Literature

There is by now a substantial literature on the general problem of finding the best 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 does not address the problem of rate of convergence, which is critical when measurements are expensive.

Some early works that deal with the challenge of determining how to optimally select measurements are (Raiffa and Schlaifer, 1968) and (De Groot, 1970), where it is formulated as a dynamic programming problem for which a practical solution has yet to be designed. An online version of the so-called multiarmed bandit problem was first solved in (Gittins and Jones, 1974) but an optimal policy for the offline ranking and selection problem remains out of reach, resulting in the development of a range of heuristics such as Boltzmann exploration, interval estimation, and hybrid exploration-exploitation policies such as epsilon-greedy, see (Frazier and Powell, 2008) for a review of these.

More formal search methods have 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 (Chen et al., 1996; He et al., 2007), and batch methods for linear loss (Chick and Inoue, 2001).

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 (Gupta and Miescke, 1996) under the name of the $(R_1, \dots, R_1)$ policy. In (Frazier et al., 2008) this policy was introduced as the 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., 2009) 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.

A significant development was the introduction in (Frazier et al., 2009) of a version 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.

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 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 distribution with unknown mean θ_x and known variance λ_x. 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 $\hat{y}_x^{n+1} = \theta_x + \varepsilon_x^{n+1}$. 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 conditional on a filtration $\mathcal{F}^n$ which is the sigma-algebra generated by $x^0, \hat{y}_{x^0}^1, x^1, \ldots, x^{n-1}, \hat{y}_{x^{n-1}}^n$. Further, we write $\mathbb{E}^n$ to indicate $\mathbb{E}[.|\mathcal{F}^n]$, the conditional expectation taken with respect to $\mathcal{F}^n$.

In this paper we follow a Bayesian approach, which offers a method of formalizing a priori beliefs and of combining them with the available observations to perform statistical inference. We assume that the different alternatives share common features, such that we learn about many alternatives from even a single measurement. A natural choice for a distribution of our belief about θ, that takes into account theses correlations in belief about the values $\theta_x, x \in \mathcal{X}$, is the multivariate normal with mean vector μ^0 and covariance matrix Σ^0,

$$\theta \sim \mathcal{N}\left(\mu^0, \Sigma^0\right). \tag{2}$$

As we show later on, our approach based on hierarchical aggregation requires a different belief distribution. However, to illustrate the Bayesion approach, let us temporarily consider (2).

Let μ^n be our estimate of θ after n measurements. This estimate will either be the Bayes estimate, which is the posterior mean $\mathbb{E}^n[\theta]$, or an approximation to this posterior mean as we will use later on. Similarly, let $\Sigma^n = Cov[\theta|\mathcal{F}^n]$ be the covariance matrix after n measurements. In the Bayesian approach we use Bayes' theorem to derive a posterior distribution using the prior distribution $p\left(\theta\right)$ as a function of μ^n and Σ^n, together with the likelihood function $p(\hat{y}_x^{n+1}|\theta)$, i.e., the likelihood of observing the data $\hat{y}_x^{n+1}$ due to the sampling decision $x^n = x$ given μ^n and Σ^n. The posterior distribution $p(\theta|\hat{y}_x^{n+1})$ is a function of μ^n, Σ^n, and conditional on the observed data $\hat{y}_x^{n+1}$ due to the sampling decision $x^n = x$. Since the prior on θ is multivariate normal and all samples are normally distributed, each of the posterior distributions on θ will be multivariate as well. Intuitively, we may view the learning that occurs from sampling as a narrowing of the conditional predictive distribution $\mathcal{N}\left(\mu^n, \Sigma^n\right)$ for θ, and as the tendency of μ^n, the center of the predictive distribution θ, to move toward θ as n increases.

After taking the N measurements, we make an implementation decision, which we assume is given by the alternative x^N that has the highest expected reward, i.e., $x^N = \arg\max_{x\in\mathcal{X}} \mu_x^N$. 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

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

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

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

3.1 Aggregation

Aggregation is performed using a set of aggregation functions $G^g : \mathcal{X} \rightarrow \mathcal{X}^g$, 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, $g = G$ 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 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.

$g=2$	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

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 referring to 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 $\mathcal{X}^1(4) = \{4, 5, 6\}$.

M^g $= |\mathcal{X}^g|$, with $M^0 = M$. In the example we have $M^1 = 3$.

Next, we introduce $\mu_x^{g,n}$ as being the estimate of the aggregated alternative $G^g(x)$ on the g^{th} aggregation level after n measurements. Using aggregation, we express μ_x^n (our estimate of θ_x) as a weighted combination of values $\mu_x^{g,n}$ for all aggregation levels $g \in \mathcal{G}$, i.e.,

$$\mu_x^n = \sum_{g \in \mathcal{G}} w_x^{g,n} \mu_x^{g,n}, \tag{3}$$

where $w_x^{g,n}$ are weights that govern the contribution of the aggregate estimate $\mu_x^{g,n}$ to the overall estimate μ_x^n of θ_x. Note that although all alternatives $x' \in \mathcal{X}^g(x)$ use the same aggregate estimate $\mu_x^{g,n}$ to estimate the value of the aggregated alternative $G^g(x) = G^g(x')$, they use separate weights $w_{x'}^{g,n}$ to determine how much the estimated value of this aggregated alternative contributes to the overall estimate μ_x^n.

We now describe the original frequentist interpretation of (3) found in (George et al., 2008). This interpretation provides specific values for the weights $w_x^{g,n}$. In the next section we also provide a Bayesian interpretation of (3) that results in the same expression for the weights.

In the frequentist interpretation, the estimator $\mu_x^{g,n}$ of θ_x is given by the average of all observations of alternatives in $\mathcal{X}^g(x)$. We make two assumptions. First, we assume that the estimators $\{\mu_x^{g,n}, \ \forall g \in \mathcal{G}\}$ are independent and unbiased. Second, we assume that we know the variance of these estimators, $(\sigma_x^{g,n})^2 = Var[\mu_x^{g,n}]$. This variance is taken with respect to the "true" probability distribution, which has a fixed value of θ. Under these assumptions, the best (minimum variance) linear unbiased estimate (BLUE) of θ_x is given by $\mu_x^n = \sum_{g \in \mathcal{G}} w_x^{g,n} \mu_x^{g,n}$, where the weights $w_x^{g,n}$ are given by

$$w_x^{g,n} = \frac{\left(\sigma_x^{g,n}\right)^{-2}}{\sum_{g' \in \mathcal{G}} \left(\sigma_x^{g',n}\right)^{-2}}. \tag{4}$$

The variance $(\sigma_x^{g,n})^2$ of estimates $\mu_x^{g,n}$ at more aggregate levels is lower because these estimates are based on more measurements. However, this is done at the cost of introducing structural aggregation errors. The aggregation error in the estimate $\mu_x^{g,n}$ is given by the expected bias $\mathbb{E}\left[\mu_x^{g,n} - \theta_x\right]$. To cope with bias, George et al. (2008) proposes the following weights

$$w_x^{g,n} = \frac{\left((\sigma_x^{g,n})^2 + (\delta_x^{g,n})^2\right)^{-1}}{\sum_{g' \in \mathcal{G}} \left((\sigma_x^{g,n})^2 + \left(\delta_x^{g',n}\right)^2\right)^{-1}}, \tag{5}$$

where $\delta_x^{g,n}$ is an estimate of the bias given by

$$\delta_x^{g,n} = \mu_x^{g,n} - \mu_x^{0,n}. \tag{6}$$

These weights are still an approximation of the optimal weights since it effectively assumes the estimates at each level of aggregation are independent. However, George et al. (2008) have shown empirically that these weights produce near-optimal results.

Having derived the expressions (3) for the estimator μ_x^n and (5) for the weights, both using a frequentist interpretation, we now derive (in the next section) the same expressions using a Bayesian interpretation.

3.2 Bayesian updating equations

A very natural approach for integrating Bayesian updating within our aggregation structure is to use Bayesian regression (see Gelman et al., 2004). The aggregation function (3) can be seen as a form of linear regression where the dependent variables are the estimates of all alternatives given by the vector μ^n, the independent (or explanatory) variables are the weights $w_x^{g,n}$, and the regression parameters are the estimates $\mu_x^{g,n}$. In this case, our belief about θ is multivariate normal, $\theta \sim \mathcal{N}\left(W^0 v^0, \Sigma^0\right)$, where W^0 is a matrix with all the weights $w_x^{g,n}$ and v^0 a vector with the estimates $\mu_x^{g,n}$ for all $g \in \mathcal{G}$. A further derivation of this approach can be found in Appendix A. However, as mentioned in (Gelman et al., 2004, chap. 14), and also showed in Appendix A, this method would not be appropriate without prior information. To cope with this, we present an alternative approach where we have a belief on each aggregation level.

The idea of using separate beliefs on the values at each aggregation level is that the correlated multivariate normal (2) is replaced by a series of independent normal distributions for all $g \in \mathcal{G}$, and that these beliefs are combined using (3) to get μ_x^n. Just as with the multivariate normal, the normal prior with normally distributed observations will result in a normal posterior. Below we derive a Bayesian interpretation of (3). For this we assume independence among the aggregation levels.

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 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 probability 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, i.e., that have different values of $\mathcal{G}^g(x)$. The value ν_x^g is currently a fixed parameter of the model, and later we use the empirical Bayes approach, first estimating it from data and then using the estimated value as if it were given a priori.

When we measure alternative 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

$$\hat{y}_x^{g,n+1} \sim \mathcal{N}(\theta_x^g, 1/\beta_x^{g,n,\varepsilon}), \tag{7}$$

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 probability model for $\hat{y}_x^{g,n+1}$ in terms of θ_x^g induces a posterior on θ_x^g.

Momentarily fix g. We define $\mu_x^{g,n}$ and $\beta_x^{g,n}$ recursively 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

$$\mu_x^{g,n+1} = \left[\beta_x^{g,n}\mu_x^{g,n} + \beta_x^{g,n,\varepsilon}\hat{y}_x^{n+1}\right]/\beta_x^{g,n+1}, \tag{8}$$

$$\beta_x^{g,n+1} = \beta_x^{g,n} + \beta_x^{g,n,\varepsilon}, \tag{9}$$

where $\beta_x^{g,0} = 0$ and $\mu_x^{g,0} = 0$. We also define $(\sigma_x^{g,n})^2 = 1/\beta_x^{g,n}$, with $\sigma^{g,0} = \infty$. Note that $\mu_x^{g,n}$, $(\sigma_x^{g,n})^2$ and $\beta_x^{g,n}$ are 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.

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 3 (Appendix B), the posterior mean and precision are given by

$$\mu_x^n = \frac{1}{\beta_x^n}\left[\beta_x^0\mu_x^0 + \sum_{g\in\mathcal{G}}\left((\sigma_x^{g,n})^2 + \nu_x^g\right)^{-1}\mu_x^{g,n}\right], \tag{10}$$

$$\beta_x^n = \beta_x^0 + \sum_{g\in\mathcal{G}}\left((\sigma_x^{g,n})^2 + \nu_x^g\right)^{-1}. \tag{11}$$

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

$$\sigma_x^n = \left(\sum_{g \in \mathcal{G}} \left((\sigma_x^{g,n})^2 + \nu_x^g \right)^{-1} \right)^{-1}, \tag{12}$$

and the posterior mean μ_x^n is given by the weighted linear combination $\mu_x^n = \sum_{g \in \mathcal{G}} w_x^{g,n} \mu_x^{g,n}$, where the weights $w_x^{g,n}$ are given by

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

Thus, we see that the modeling assumption we made (that actually we had multiple independent observations instead of just one) has an identical effect to the one assumed in the frequentist derivation (in Section 3.1) because it causes the posterior means of θ_x^g across different values of g to be independent under the model when conditioned on θ_x. It also causes the posterior means to have the same form as the BLUE estimator (4) from the frequentist derivation.

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 prior.

First, we estimate ν_x^g by $(\delta_x^g)^2$. Obviously, this is an approximation, but it does result in the same weights as proposed by George et al. (2008). However, the aggregation error $\delta_x^{g,n}$ is undefined when $m_x^{0,n} = 0$. To cope with this, we introduce a base level g_x^* for each alternative x, being the lowest level g for which $m_x^{g,n} > 0$. Further, we set $w_x^{g,n} = 0$ and $\delta_x^{g,n} = 0$ for all levels $g < g_x^*$ and define the aggregation error in terms of the base levels, i.e., $\delta_x^{g,n} = \mu_x^{g_x^*,n} - \mu_x^{g,n}$ for $g \geq g_x^*$.

Next, we estimate $\beta_x^{g,n,\varepsilon}$ using $\beta_x^{g,n,\varepsilon} = (\sigma_x^{g,n,\varepsilon})^{-2}$ where $(\sigma_x^{g,n,\varepsilon})^2$ is the group variance (also called the population variance). The group variance $\left(\sigma_x^{0,n,\varepsilon}\right)^2$ 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

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

where $m_x^{g,n}$ is the number of measurements from the aggregated alternative $G^g(x)$ at the g^{th} aggregation level, i.e., 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 $(\sigma_x^{g,n,\varepsilon})^2$. In a way this makes sense 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

$$(\sigma_x^{g,n,\varepsilon})^2 = \frac{1}{|\mathcal{X}^g(x)|} \left(\sum_{\forall x' \in \mathcal{X}^g(x)} \lambda_{x'} + \left(\mu_{x'}^{0,n} - \mu_x^{g,n}\right)^2 \right). \tag{13}$$

We end this section by summarizing the Bayesian updating procedure. After sampling $x^n = x$, we use the resulting observation $\hat{y}_x^{n+1}$ in (8) and (9) to compute $\mu_x^{g,n+1}$ and $\beta_x^{g,n+1}$ for all $g \in \mathcal{G}$. Next, we use $\mu_{x'}^{g,n+1}$ in (6) to compute the biases $\delta_{x'}^{g,n+1}$ for all $x' \in \mathcal{X}$ and $g \in \mathcal{G}$. Finally, we use (5) and (13) to compute the weights $w_{x'}^{g,n+1}$ and group variances $(\sigma_x^{g,n+1,\varepsilon})^2$ for all $x' \in \mathcal{X}$ and $g \in \mathcal{G}$. The one task remaining is to derive a formal procedure for the measurement decisions x^n which we present in the next section.

As an aid to the reader, we briefly summarize the notation defined throughout this paper.

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 all aggregation levels that the alternatives x and x' have in common

$\mathcal{X}$ set of all alternatives

$\mathcal{X}^g$ set of all aggregated alternatives $G^g(x)$ at the g^{th} aggregation level

$\mathcal{X}^g(x)$ Set of all alternatives that share the same aggregated alternative $G^g(x)$ at the g^{th} aggregation level

N maximum number of measurements

M $= |\mathcal{X}|$

M^g $= |\mathcal{X}^g|$

θ_x unknown mean of the true value of alternative x

θ_x^g latent variable desribing the unknown mean of the "true" value of 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 the aggregated alternative $G^g(x)$ on the g^{th}aggregation level 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 $= 1/(\sigma_x^n)^2$, the precision of μ_x^n

$\beta_x^{g,n}$ $= 1/(\sigma_x^{g,n})^2$, the precision of $\mu_x^{g,n}$

$\beta_x^{g,n,\varepsilon}$ $= 1/(\sigma_x^{g,n,\varepsilon})^2$, the measurement precision of observations from alternatives $x' \in \mathcal{X}^g(x)$

$\delta_x^{g,n}$ estimate of the aggregation bias

$\nu_x^{g,n}$ variance of $\theta_x^g - \theta_x$

4 Measurement decision

Our goal is to maximize the expected reward $\mu_{x^N}^N$ of the implementation decision $x^N = \arg\max_{x\in\mathcal{X}} \mu_x^N$. During the sequence of N sampling decisions, $x^0, x^1, \ldots, x^{N-1}$ we gain information that increases our expected final reward $\mu_{x^N}^N$. 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 beliefs as introduced in (Frazier et al., 2008, 2009). Then, in Section 4.3, we adapt this policy to our hierarchical setting.

4.1 The knowledge-gradient policy

The knowledge-gradient policy was first introduced in (Gupta and Miescke, 1996) under the name $(R_1, \ldots, R_1)$ policy, 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 $S^n = (\mu^n, \Sigma^n)$. If we were to stop measuring now, our final expected reward would be $\max_{x\in\mathcal{X}} \mu_x^n$. Now, suppose we were allowed to make one more measurement x^n. Then, the observation $\hat{y}_{x^n}^{n+1}$ would result in an updated knowledge state S^{n+1} which might result in a higher expected reward $\max_{x\in\mathcal{X}} \mu_x^{n+1}$ at the next time unit. The expected incremental value due to measurement x is given by

$$\upsilon_x^{KG}(S^n) = \mathbb{E}\left[\max_{x'\in\mathcal{X}} \mu_{x'}^{n+1} | S^n, \hat{y}_x^{n+1}\right] - \max_{x'\in\mathcal{X}} \mu_{x'}^n. \tag{14}$$

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

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

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 (14) can be written as

$$\upsilon_x^{KG}\left(S^n\right) = \tilde{\sigma}_x\left(\Sigma^n, x\right) f\left(\frac{-|\mu_x^n - \max_{x' \neq x} \mu_{x'}^n|}{\tilde{\sigma}_x\left(\Sigma^n, x\right)}\right),$$

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

In the case of correlated beliefs, an observation $\hat{y}_x^{n+1}$ of alternative x may change our estimate $\mu_{x'}^n$ of alternatives $x' \neq x$. The knowledge gradient (14) can be written as

$$\upsilon_x^{KG,n}\left(S^n\right) = \mathbb{E}\left[\max_{x' \in \mathcal{X}} \mu_{x'}^n + \tilde{\sigma}_{x'}\left(\Sigma^n, x\right) Z | S^n, \hat{y}_x^{n+1}\right] - \max_{x' \in \mathcal{X}} \mu_{x'}^n, \tag{15}$$

where Z is a standard normal random variable and $\tilde{\sigma}_{x'}\left(\Sigma^n, x\right) = \Sigma_{x'x}^n / \sqrt{\lambda_x + \Sigma_{xx}^n}$ with $\Sigma_{x'x}^n$ the covariance between $\mu_{x'}^n$ and μ_x^n.

Solving (15) involves the computation of the expectation over the maximum of M linear functions. To do this, Frazier et al. (2009) provides an algorithm (Algorithm 2) which solves $h\left(a, b\right) = \mathbb{E}\left[\max_i a_i + b_i Z\right] - \max_i a_i$ as a generic function of any vectors a and b (where the elements of a and b are given by $\mu_{x'}^n$ and $\tilde{\sigma}_{x'}\left(\Sigma^n, x\right)$ respectively). The algorithm works as follows. First it sorts the sequence of pairs (a_i, b_i) such that the b_i are in non-decreasing 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 another pair (a_j, b_j), i.e., $a_i + b_i Z \leq a_j + b_j Z$ 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 $\tilde{M}$. Further, we use $\tilde{a}_i$ and $\tilde{b}_i$ to denote the i^{th} element of a and b respectively. The function $h\left(a, b\right)$ is computed using

$$h\left(a, b\right) = \sum_{i=1,\ldots,\tilde{M}} \left(\tilde{b}_{i+1} - \tilde{b}_i\right) f\left(-\left|\frac{\tilde{a}_i - \tilde{a}_{i+1}}{\tilde{b}_{i+1} - \tilde{b}_i}\right|\right). \tag{16}$$

Note that some variations of (16) are considered in (Frazier et al., 2009) to avoid rounding errors in the implementation. Further note that the knowledge gradient algorithm for correlated beliefs requires that the covariance matrix Σ^n be provided as an input. These correlations are typically attributed to physical relationships among the alternatives.

4.3 Hierarchical knowledge gradient

We derive the hierarchical knowledge-gradient (HKG) policy based on our choice of using separate beliefs on each aggregation level (see Section 3.1). For completeness, we added the knowledge-gradient sampling decision that exploits the Bayesian regression approach in Appendix D.

Our knowledge state is now given by $S^n = \{\mu_x^{g,n}, \beta_x^{g,n} : x \in \mathcal{X}, g \in \mathcal{G}\}$. From these parameters we are able to compute the knowledge gradient values. Before working out the knowledge gradient (14), we first focus on the aggregate estimate $\mu_x^{g,n+1}$. We rewrite the updating equation (8) as

$$\mu_x^{g,n+1} = \mu_x^{g,n} + \frac{\beta_x^{g,n,\varepsilon}}{\beta_x^{g,n} + \beta_x^{g,n,\varepsilon}} \left(\hat{y}_x^{n+1} - \mu_x^{g,n} \right).$$

For reasons that become clear later on, we further rewrite this equation by splitting the second term using

$$\mu_x^{g,n+1} = \mu_x^{g,n} + \frac{\beta_x^{g,n,\varepsilon}}{\beta_x^{g,n} + \beta_x^{g,n,\varepsilon}} \left(\hat{y}_x^{n+1} - \mu_x^{n} \right) + \frac{\beta_x^{g,n,\varepsilon}}{\beta_x^{g,n} + \beta_x^{g,n,\varepsilon}} \left(\mu_x^{n} - \mu_x^{g,n} \right).$$

Now, the new estimate is given by the sum of (i) the old estimate, (ii) the deviation of $\hat{y}_x^{n+1}$ from the current expectation μ_x^n times the relative increase in precision, and (iii) the deviation of the expectation μ_x^n from the aggregate estimate $\mu_x^{g,n}$ 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 lower this update will be because the precision of our belief on $\mu_x^{g,n}$ becomes 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 μ_x^n given by (12). So, $\left(\hat{y}_x^{n+1} - \mu_x^n\right) / \sqrt{(\sigma_x^n)^2 + \lambda_x}$ is a standard normal. Now we can write

$$\mu_x^{g,n+1} = \mu_x^{g,n} + \frac{\beta_x^{g,n,\varepsilon}}{\beta_x^{g,n} + \beta_x^{g,n,\varepsilon}} \left(\mu_x^{n} - \mu_x^{g,n} \right) + \tilde{\sigma}(g,x) Z, \tag{17}$$

where

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

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 (3) for the weighted estimate $\left\{\mu_{x'}^{n+1}\right\}$ 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 (17) and some rearrangement of terms we get

$$\begin{aligned} \mu_{x'}^{n+1} &= \sum_{g \in \mathcal{G}} w_{x'}^{g,n+1} \mu_{x'}^{g,n} + \sum_{g \in \mathcal{G}(x',x)} w_{x'}^{g,n+1} \frac{\beta_x^{g,n,\varepsilon}}{\beta_x^{g,n} + \beta_x^{g,n,\varepsilon}} \left(\mu_x^{n} - \mu_x^{g,n} \right) \\ &\quad + Z \sum_{g \in \mathcal{G}(x',x)} w_{x'}^{g,n+1} \tilde{\sigma}(g,x). \end{aligned} \tag{18}$$

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 $\bar{w}^{g,n}_{x'}(x)$, we use the updated precision $\beta^{g,n+1}_x$ due to sampling x in the weights (5). However, we use the current biases $\delta^{g,n}_x$ because the updated bias $\delta^{g,n+1}_x$ depends on the $\mu^{g,n+1}_x$ which we aim to estimate. The predictive weights $\bar{w}^{g,n}_{x'}(x)$ are given by

$$\bar{w}^{g,n}_{x'}(x) = \frac{\left(\left(\beta^{g,n}_{x'} + I^{g}_{x',x}\beta^{g,n,\varepsilon}_{x'}\right)^{-1} + \left(\delta^{g,n}_{x'}\right)^2\right)^{-1}}{\sum_{g'\in\mathcal{G}}\left(\left(\beta^{g',n}_{x'} + I^{g'}_{x',x}\beta^{g',n,\varepsilon}_{x'}\right)^{-1} + \left(\delta^{g',n}_{x'}\right)^2\right)^{-1}}, \tag{19}$$

where

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

After combining (14) with (18) and (19), we get the following knowledge gradient

$$\upsilon^{KG}_x\left(S^n\right) = \mathbb{E}\left[\max_{x'\in\mathcal{X}} a^n_{x'}(x) + b^n_{x'}(x)Z|S^n, \hat{y}^{n+1}_x\right] - \max_{x'\in\mathcal{X}}\mu^n_{x'}, \tag{20}$$

where

$$a^n_{x'}(x) = \sum_{g\in\mathcal{G}}\bar{w}^{g,n}_{x'}(x)\mu^{g,n}_{x'} + \sum_{g\in\mathcal{G}(x',x)}\bar{w}^{g,n}_{x'}(x)\frac{\beta^{g,n,\varepsilon}_x}{\beta^{g,n}_x + \beta^{g,n,\varepsilon}_x}\left(\mu^n_x - \mu^{g,n}_x\right), \tag{21}$$

$$b^n_{x'}(x) = \sum_{g\in\mathcal{G}(x',x)}\bar{w}^{g,n}_{x'}(x)\tilde{\sigma}\left(g,x\right). \tag{22}$$

Note that if we would have used the current weights $w^{g,n}_{x'}$ instead of the predictive weights $\bar{w}^{g,n}_{x'}(x)$, the convergence proofs of Section 5 would no longer hold.

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 (20) can now be computed using

$$\upsilon^{KG,n}_x = \sum_{i=1,\ldots,\tilde{M}-1}\left(\tilde{b}^n_{i+1}(x) - \tilde{b}^n_i(x)\right) f\left(-\left|\frac{\tilde{a}^n_i(x) - \tilde{a}^n_{i+1}(x)}{\tilde{b}^n_{i+1}(x) - \tilde{b}^n_i(x)}\right|\right), \tag{23}$$

where $\tilde{a}^n_i(x)$ and $\tilde{b}^n_i(x)$ follow from (21) and (22), after the sort and merge operation as described in Section 4.2.

The form of (23) 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 $O\left(M^2\log M\right)$.

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 (Frazier and Powell, 2008), the independent KG policy prefers to measure alternatives with a high mean and/or with a low precision. If the precision is the same, KG would select the alternative with the highest mean. If the means are the same, KG would select the alternative with the lowest precision. In the HKG policy, the means are given by weighted sums of estimates at all aggregation levels, and the precisions are given by weighted sums of precisions at all aggregation levels.

Now, consider a problem with eight alternatives and an aggregation structure given by a perfect binary tree, as illustrated in Figure 2. The first measurement will be random among the eight alternatives, in this case alternative 6. The next measurement will generally be chosen such that it shares the least number of aggregation levels with the first measurement, in this case random among alternatives 1 through 4, which results in alternative 1. The next measurement will generally be between alternatives 3, 4, 7, and 9 because they have the lowest precision. However, HKG prefers to measure alternatives 3 and 4 because they have a higher weighted mean due to observation 2.

In Figure 2, we have shown the weighted estimates after the first four measurements. In case of common measurement noise, the fifth measurement under the HKG policy will be either alternative 8 (highest mean of the four alternatives with lowest precision) or alternative 7 to gain more confidence (increase the precision) in the highest weighted estimate.

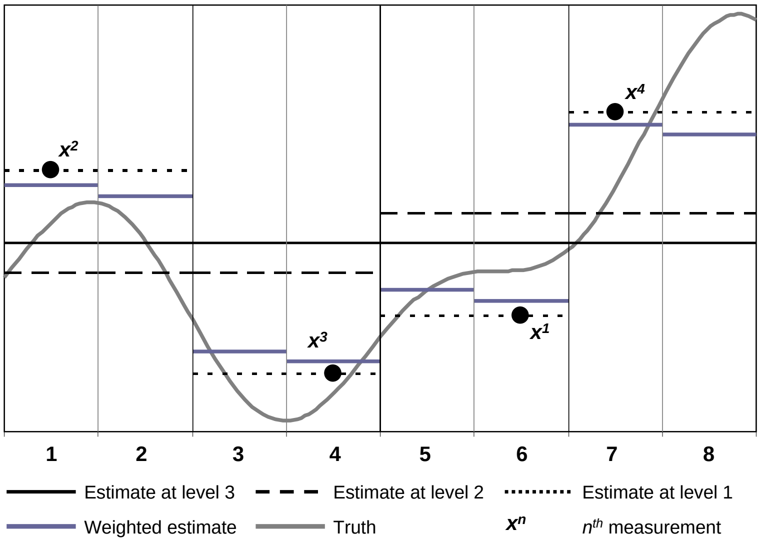

Figure 2: Illustration of the behavior of HKG

5 Convergence results

In this section we show that the value of the hierarchical knowledge-gradient policy converges to the value of the optimal policy in the limit as $N \to \infty$, and that the hierarchical knowledge-gradient policy eventually finds an optimal alternative almost surely. The theorem and corollary presented in this section depend on various lemmas that can be found in Appendix C.

Theorem 1 *In the limit, the HKG policy measures every alternative infinitely often, almost surely.*

Proof. Consider what happens as the number of measurements n we make under the HKG policy goes to infinity. Let $\mathcal{X}'$ 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 $\mathcal{X}' \neq \mathcal{X}$ with positive probability, i.e., 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 $\mathcal{X}'$. We compare the KG values $\upsilon_x^{KG,n}$ of those alternatives within $\mathcal{X}'$ to those outside $\mathcal{X}'$.

Let $x \in \mathcal{X}'$; we show that $\lim_n \upsilon_x^{KG,n} = 0$. Since f is an increasing 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 bound

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

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

Next, let $x \notin \mathcal{X}'$. We show that $\lim_{n\to\infty} \upsilon_x^{KG,n} > 0$. Define the set $\mathcal{I}$ to contain all indices i such that $\liminf_{n\to\infty} \tilde{b}_i^n(x) > 0$. From Lemma 8, we know there exists at least one x' for which $\liminf_{n\to\infty} b_{x'}^n(x) > 0$, namely $x' = x$ and there exists at least one x' for which $\lim_n b_{x'}^n(x) = 0$ since $\mathcal{X}'$ is nonempty. As a result, both $\mathcal{I}$ and its complement are nonempty. Thus, an $N_2 < \infty$ exists such that $\min_{i\in\mathcal{I}} \tilde{b}_i^n(x) > \max_{j\notin\mathcal{I}} \tilde{b}_j^n(x)$ for all $n > N_2$. By the ordering of $\tilde{b}_i^n(x)$ used to compute $\upsilon_x^{KG,n}$, and the monotonicity and nonnegativity of f, we have for all $n > N_2$,

$$\upsilon_x^{KG,n} \geq \min_{i\in\mathcal{I}, j\notin\mathcal{I}} \left(\tilde{b}_i^n(x) - \tilde{b}_j^n(x)\right) f\left(-\left|\frac{\tilde{a}_i^n(x) - \tilde{a}_j^n(x)}{\tilde{b}_i^n(x) - \tilde{b}_j^n(x)}\right|\right).$$

Now define $U = \sup_{n,i,x} |\tilde{a}_i^n(x)|$, which is almost surely finite since $\sup_n |a_{x'}^n(x)|$ is almost surely finite $\forall x, x' \in \mathcal{X}$ by Lemma 6. From this bound on $|a_i^n|$, we have the uniform bound $\sup_{n,i,x} |a_i^n(x) - a_{i+1}^n(x)| \leq 2U$. Then, for all $n > N_2$, the monotonicity of f implies

$$\upsilon_x^{KG,n} \geq \min_{i\in\mathcal{I}, j\notin\mathcal{I}} \left(\tilde{b}_i^n(x) - \tilde{b}_j^n(x)\right) f\left(\frac{-2U}{\tilde{b}_i^n(x) - \tilde{b}_j^n(x)}\right).$$

Taking limits, noting the continuity of f, and substituting $b^* = \min_{i\in\mathcal{I}} \tilde{b}_i^n(x) > 0$, we obtain

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

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

Corollary 2 *Under the HKG policy,* $\lim_n \mu_x^n = \theta_x$ *almost surely for each alternative* x.

Proof. Fix x and note that Theorem 1 implies that x is measured infinitely often. The estimate $\mu_x^{0,n}$ at the disaggregate level is a linear combination of the average of all observations of alternative x, and the prior value $\mu_x^{0,0}$. As $n \to \infty$, the weight placed on the prior value vanishes, and $\lim_n \mu_x^{0,n}$ is the same as the limit of the average of all observations of alternative x, which, by the law of large numbers, is almost surely equal to θ_x. This shows that $\lim_n \mu_x^{0,n} = \theta_x$ almost surely.

Turning our attention to the aggregate estimates $\mu_x^{g,n}$ and the overall estimate μ_x^n, define $\mathcal{G}' = \{g \in \mathcal{G} : \lim_n \delta_x^{g,n} = 0\}$ to be the levels of aggregation for which the bias is zero in the limit. Since the limiting bias on these levels is zero, we have $\lim_n \mu_x^{g,n} = \lim_n \mu_x^{0,n} = \theta_x$ almost surely for $g \in \mathcal{G}'$.

For each $g \notin \mathcal{G}'$, the statements $\delta_x^{0,n} = 0$ for all n and $\lim_n \beta_x^{g,n} = \infty$ (implied by Theorem 1) together imply that $\lim_{n\to\infty} w_x^{g,n} = 0$. Thus, in the limit, all weight is given to levels $g \in \mathcal{G}'$. This, together with the relation, $\mu_x^n = \sum_{g\in\mathcal{G}} w_x^{g,n} \mu_x^{g,n}$, implies that $\lim_n \mu_x^n = \lim_n \mu_x^{0,n} = \theta_x$ almost surely. ■

As an addendum to the proof, we note that usually the only level with zero limiting bias is the disaggregate level. In such cases, $\lim_{n\to\infty} w_x^{0,n} = 1$, i.e., we put full weight on the disaggregate level in the limit.

6 Numerical experiments

To evaluate the hierarchical knowledge-gradient policy, we perform a number of experiments using two different settings. First, we evaluate the performance of our approach in finding the maximum of a continuous one-dimensional function. Second, we consider an application in logistics where we aim to find the best multi-attribute vector out of a large set of possible attribute vectors, which can be seen as maximizing a multi-dimensional and possible non-continuous function. We present these experiments in Sections 6.1 and 6.2 respectively. We end, in Section 6.3 with some remarks on the choice of aggregation structure.

6.1 One-dimensional functions

First we test our approach on one-dimensional functions on a continuous domain. 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

$$Cov\,(i,j) = \sigma^2 \exp\left\{ -\left(\frac{|i-j|}{(M-1)\,\rho} \right)^{\eta} \right\}, \tag{24}$$

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

The idea is that higher values of ρ will have less peaks in the domain and higher values of η will result in smoother functions. Here we fix $\eta = 2$ and vary ρ. 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. The settings of ρ and λ are given in Table 1.

Because our approach requires discretization, we use integer values for i and j in (24) with values between 1 and 128. We use a fixed aggregation structure for all test functions. This structure is given by a binary tree, i.e., $|\mathcal{X}^g(x)| = 2^g$ for all $x \in \mathcal{X}^g$ and $g \in \mathcal{G}$. Given $M = 128$ we have eight aggregation levels (including the disaggregate level).

Factor	Value
ρ	0.05, 0.1, 0.2, 0.5
λ	0.01, 0.25

Table 1: Values for ρ and λ

To generate test functions, we first generate a column vector $Z = (z_1, z_2, \ldots, z_{128})$, where the elements z_i are independent draws from a standard normal distribution. Then we compute a covariance matrix of size 128×128 with elements given by (24) and compute the Cholesky decomposition of the covariance matrix resulting in a lower-triangular matrix C. The test functions follow from $\theta = CZ$. Measuring from the test functions was done with normally distributed noise λ. To provide an illustration of the test functions, we show in Figure 3 one test function for each value of ρ.

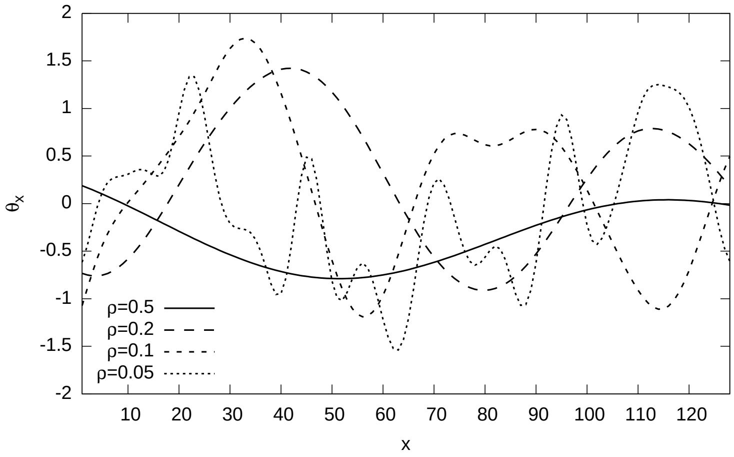

Figure 3: Illustration of one-dimensional test functions

We compare the hierarchical KG policy (HKG) against (i) a pure exploration policy (EXPL), i.e., we measure each alternative with the same probability, (ii) the independent KG policy (IKG) of (Frazier et al., 2008), and (iii) the correlated KG policy (CKG) of (Frazier et al., 2009) where the covariance function is assumed known. Because the CKG policy requires prior knowledge of the covariance function, we use the covariance function (24) with given parameters σ^2, η and ρ. Hence, CKG has perfect knowledge of the covariance matrix and, as a result, the performance of this policy can be regarded as a bound for HKG. We did not consider classical strategies such as interval estimation, Boltzmann exploration, and epsilon-greedy exploration since these require either an informed prior, or at least one measurement of each of the M alternatives. Our interest is primarily in problems where M may be much larger than the measurement budget. For further comparisons of IKG and CKG with several well known ranking and selection policies and Bayesian global optimization methods, we refer to (Frazier et al., 2008, 2009).

In our experiments we randomly generate 10 functions for all combinations of ρ and λ (resulting in $10 \times 4 \times 2 = 80$ test functions). Next, we test each policy on each function using 25 replications with different random number streams. We compare the policies for given values of ρ and λ, based on their average performance on the 25 replications and 10 test functions. As a primary performance indicator we use the opportunity cost which is defined as $(\max_i \theta_i) - \theta_{i^*}$, with $i^* \in \arg\max_x \mu_x^n$, i.e., the difference between the true maximum and the value of the best alternative found by the algorithm.

The results for various length scale parameters ρ can be found in Figures 4 and 5 for $\lambda = 0.01$ and $\lambda = 0.25$ respectively. From these figures we draw the following conclusions.

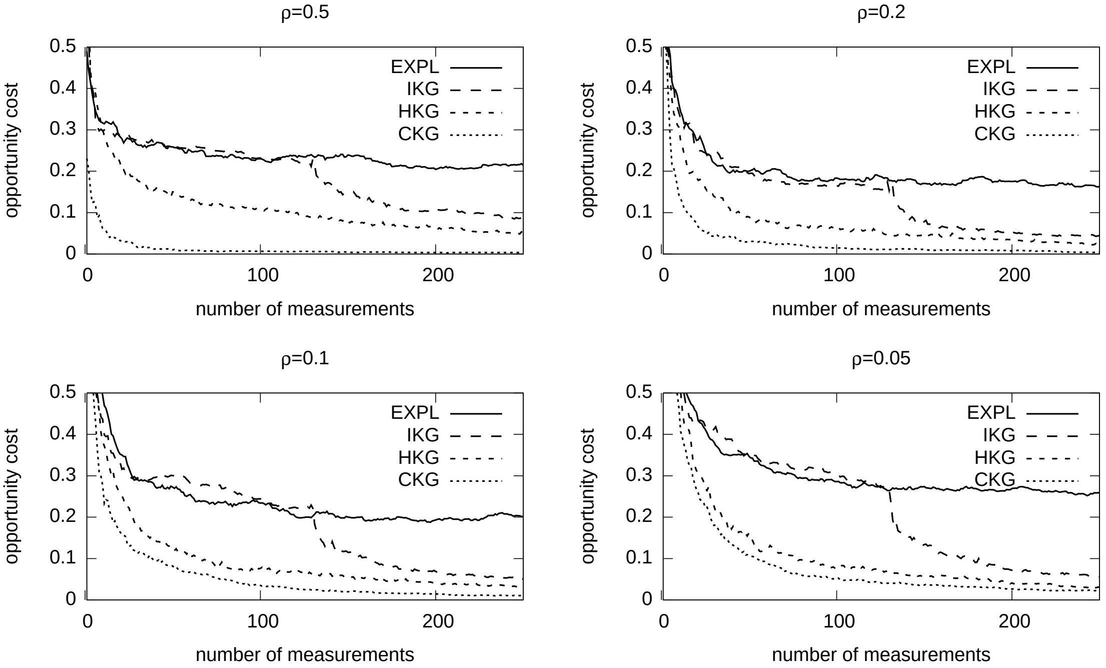

Figure 4: Comparison of EXPL, IKG, HKG, and CKG using $\lambda = 0.25$ and various settings for ρ.

First, we see that HKG consistently outperforms exploration and independent knowledge gradient, especially in the critical early iterations (since we are interested in doing as much as possible with very few iterations). Not surprisingly, CKG works best because it is given the true covariance function, something that HKG is not given.

Second, we see that HKG starts to approach the performance of CKG as the scale parameter ρ is decreased and/or the measurement noise λ is increased. For $\rho = 0.05$, the function fluctuates fairly rapidly, producing multiple local maxima even within relatively small areas (see Figure 3). For CKG this means that the prior covariance matrix contains relatively low covariances $Cov(i, j)$, especially when the difference between i and j is relatively big. As a result, a measurement from a single alternative provides relatively little information about the other alternatives. For HKG this means that within one aggregated set, we might have a local maximum and local minimum. HKG is able to deal with this by placing a relatively low weight on such sets. As a result, HKG is fairly robust against settings for ρ.

Third, we see that IKG and HKG seem to converge to each other. The IKG policy requires that we first measure all the 128 alternatives once. After this, the opportunity cost drops quickly with IKG. In some cases with low noise ($\lambda = 0.01$, shown in Figure 5) and a high number of measurements, we see that IKG result in slightly lower opportunity costs than HKG. The reason for this is that HKG still tends to put some weight on the aggregate levels. However, with increasing number of measurements, HKG tend to put all weight on the disaggregate level

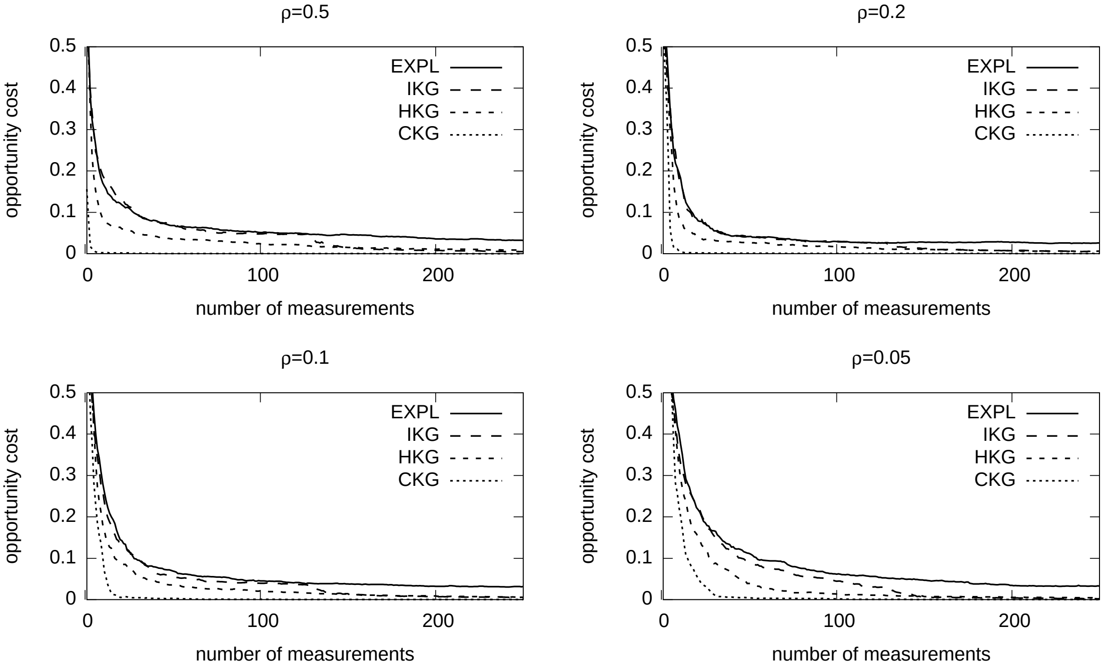

Figure 5: Comparison of EXPL, IKG, HKG, and CKG using $\lambda = 0.01$ and various settings for ρ.

(see the proof of Corollary 2) such that HKG coincides with IKG.

To provide an indication of the significance of the results, we display the standard deviation of the performance of the policies after the 128^{th} measurement. The standard deviation is computed using the average performance for each of the 10 individual functions for each value of ρ. The results can be found in Figure 6. We see that for low values of ρ, CKG performs significantly better than HKG (but of course requires prior knowledge on the covariance matrix). However, with increasing ρ, the difference between HKG and CKG clearly declines.

6.2 Multi-dimensional functions

Next, we consider an application that arose in a transportation application (see Simao et al., 2009) where we had to 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) which of six fleets to which 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) were divided into 100 regions (by the company) which is further discretized into 10 areas. At the most disaggregate level, there are $100 \times 100 \times 6 = 60,000$ attributes. Our problem is to find which of these 60,000 attributes is best.

To reduce computation time, we divide the spatial attributes into 25 regions and 5 areas. Further, 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

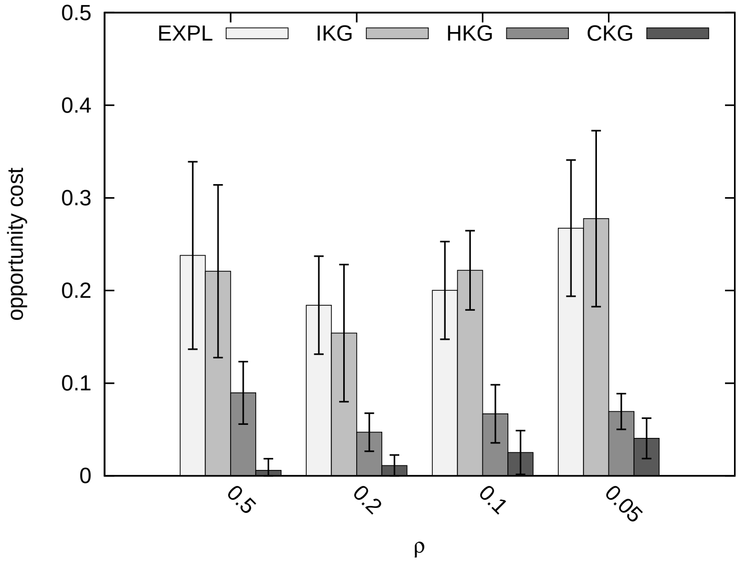

Figure 6: Comparison of EXPL, IKG, HKG, and CKG at the 128^{th} measurement using $\lambda = 0.25$.

represent location as one of 5 areas.

To evaluate the quality of our search, we have to use a known function that describes the underlying truth. We describe the expected single period reward using a standard test function called the six-hump camel back from (Branin, 1972) which is given by

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

with $x_1 \in [-1.6, 2.4]$ and $x_2 \in [-0.8, 1.2]$.

We let x_1 be the location and x_2 be the driver domicile, which are both discretized into 25 pieces to represent the regions and into 5 pieces to represent the areas. 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.

x_3	CAN	WR	US_S	US_T	US_IS	US_IT
$p_1(x_3)$	4800	4800	4700	4500	4200	4000
$p_2(x_3)$	100	100	200	0	200	0

Table 2: Parameter settings

To cope with the fact that some drivers (CAN and WR) are bounded by certain locations, we exclude combinations $\{x_3 = \text{CAN} \wedge x_1 < 1.8\}$ and $\{x_3 = \text{WR} \wedge x_1 > -0.8\}$. As a result, the number of different attributes is 2725. The maximum of $g(x_1, x_2, x_3)$ is attained at $g(0, 0, \text{US\_S})$ with value 4700.

To provide an indication of the resulting function, we show $\max_{x_3} g(x_1, x_2, x_3)$ in Figure 7. This function has similar properties as the six-hump, except for the presence of discontinuities due to the capacity types CAN and WR, and a twist at $x_1 = x_2$. We compare EXPL, IKG, and HKG using the opportunity cost. To get reliable results, we perform 10 replications with IKG and 50 replications with EXPL and IKG.

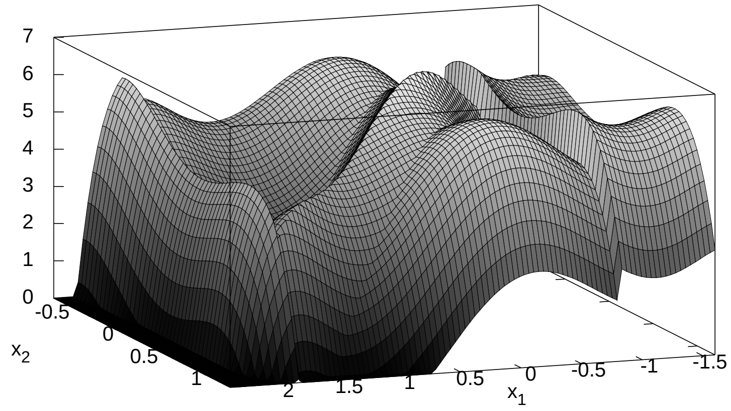

Figure 7: $max_{x_3} g(x_1, x_2, x_3)$

The results can be found in Figure 8. Again we see, that HKG outperforms EXPL and IKG. In fact, in all 10 replications, HKG converged to the optimum solution, i.e., it finds the best out of 2725 alternatives in less than 1200 measurements.

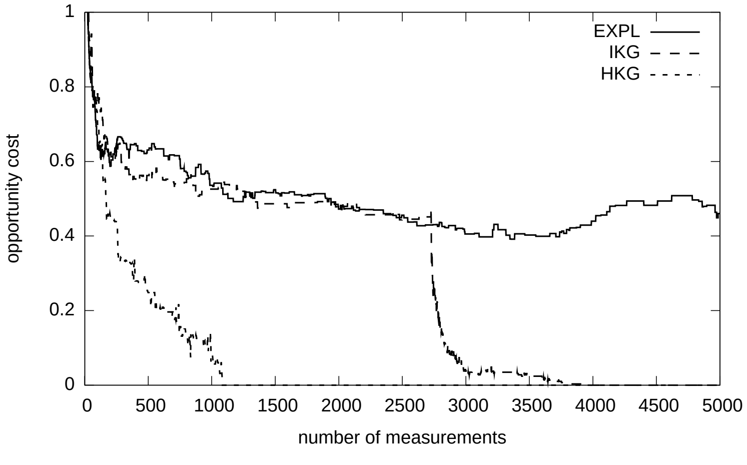

Figure 8: Comparison of IKG with HKG using various aggregation structures and various settings for ρ.

6.3 Remarks on the aggregation structure

We end this section with a short note on the choice of aggregation structure since it is the only "tunable parameter" in HKG. Without showing the results, we experienced that HKG is relatively robust to the choice of aggregation structure. For example, we tested HKG on the one-dimensional test functions using a minimal aggregation structure where we use only one aggregated estimate $\mu_x^{g,n}, g \geq 1$. We found that HKG still significantly outperforms EXPL and IKG in the case where $N \leq M$. The reason for this is as follows. The mean and precision of the unmeasured alternatives equal the grand mean and the precision of the grand mean. The precision of the alternatives we measured is most of the time bigger than those we did not measure, and the mean of some of these are above the grand mean and some of them are below the grand mean. As long as there are unmeasured alternatives, HKG tends not to measure alternatives with mean below the grand mean. So, the grand mean forms a kind of threshold in the sampling decision. In general, a finer aggregation structure with more levels is always better. If it appears that higher aggregation levels are not required, or even not appropriate, HKG automatically puts higher weights to the lower levels.

7 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 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. This method requires the use of a known covariance matrix. In our strategy, we compute covariances from our statistical model which estimates the value of the function using weighted estimates.

The hierarchical knowledge-gradient (HKG) algorithm shares the inherent steepest ascent property of the knowledge gradient algorithm which makes observations that produce the greatest improvement in our estimate of the maximum of the function. We also prove that the algorithm is guaranteed to produce the optimal solution in the limit, since the HKG algorithm shares the inherent characteristic of knowledge-gradient policies of measuring every alternative infinitely often, in the limit. This feature, however, was not automatic and required the careful design of the updating strategy to handle the fact that we are approximating the covariance structure from data rather than assuming it as input.

We close with experimental results on a class of scalar functions and a multi-attribute problem drawn from a transportation application. The scalar functions were randomly generated using a specified covariance structure, allowing us to compare the performance of HKG against the knowledge-gradient algorithm which takes the covariance structure as input. HKG was shown to produce fast convergence in the early iterations, a feature that is critical in many applications. For the transportation application, we showed that, HKG finds the best of 2725 alternatives in all replications in less than 1200 measurements, despite the presence of noisy observations.

Our HKG policy has several limitations. First, 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.

Second, the HKG policy requires enumerating all possible choices before determining the next measurement. The logic in this paper can handle perhaps thousands of choices, but not millions. Our own work is motivated by applications where we need to make good choices with a small number of measurements, typically far smaller than the set of potential measurements. The HKG policy can work quite well even when we sample only a portion of all potential measurements (of course this performance depends on the structure of the problem), but specialized algorithms would need to be designed if $|\mathcal{X}|$ is extremely large.

Third, we assumed the measurement noise to be known. We might overcome this by placing a normal-gamma prior on the unknown means and variances at each aggregation level (see Chick et al., 2009). In this case we basically rely on the sample variances. As a result, it would take some measurements before we get reliable estimates of the variances.

We mention two areas for further research. HKG is designed to work on functions which depend on a multiattribute vector, and as we have presented it, requires that we scan all possible measurements before making a decision. When the number of measurements is large (something we would expect with a multidiensional vector) this step becomes prohibitive. As an alternative, we can use HKG to choose regions to measure at successively finer levels of aggregation, corresponding to the family of aggregation functions. More specifically, we might first make an aggregated sampling decision $x^{g,n} = x$ with $x \in \mathcal{X}^g$. Because the aggregated sets $\mathcal{X}^g$ for $g > 0$ have fewer elements than the disaggregated set $\mathcal{X}$ we might gain some computational advantage. Preliminary experiments have shown that this method can drastically reduce computation time without harming the performance too much. In addition, this option scales much better in the number of alternatives. However, there is still more research required on this issue. Another area for further research is the applicability of HGK for approximate dynamic programming. The main challenge here is to find a way to cope with the bias in downstream values.

References

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

Bertsekas, D. P. and Tsitsiklis, J. N. (1996). *Neuro-Dynamic Programming.* Athena Scientific, Belmont, MA.

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

Chen, C.-H., Chen, H.-C., and Dai, L. (1996). A gradient approach for smartly allocating computing budget for discrete event simulation. *Simulation Conference Proceedings, 1996. Winter*, pages 398–405.

Chick, S. E., Branke, J., and Schmidt, C. (2009). Sequential sampling to myopically maximize the expected value of information. *INFORMS Journal on Computing*, To appear.

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

De Groot, M. H. (1970). *Optimal statistical decisions.* McGraw-Hill, New York.

Frazier, P. I. and Powell, W. B. (2008). Optimal Learning. In *TutORials in Operations Research*, pages 213–246, Hanover, Md. Informs.

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

Frazier, P. I., Powell, W. B., and Dayanik, S. (2009). The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, to appear.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis.* CRC Press, 2nd edition.

George, A., Powell, W. B., and Kulkarni, S. R. (2008). Value function approximation using multiple aggregation for multiattribute resource management. *Journal of Machine Learning Research*, 9:2079–2111.

Gittins, J. C. and Jones, D. M. (1974). A dynamic allocation index for the sequential design of experiments. In Gani, J., editor, *Progress in Statistics*, pages 241–266.

Gupta, S. S. and Miescke, K. J. (1996). Bayesian look ahead one-stage sampling allocations for selection of the best population. *Journal of statistical planning and inference*, 54(2):229–244.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). *The Elements of Statistical Learning.* Springer series in Statistics, New York, NY.

He, D., Chick, S. E., and Chen, C.-H. (2007). 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.

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

Raiffa, H. and Schlaifer, R. (1968). *Applied Statistical Decision Theory.* M.I.T. Press.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Annals of Math. Stat.*, 22:400–407.

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

Simao, H. P., Day, J., George, A. P., Gifford, T., Nienow, J., and Powell, W. B. (2009). An approximate dynamic programming algorithm for large-scale fleet management: A case application. *Transportation Science*, 43(2):178–197.

Snijders, T. A. and Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Sage Publications Ltd.

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

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

Appendix A

In the Bayesian regression approach we assume the truth can be expressed by a system of linear equations $\theta = Wv$. Conditioned on W, v, and $x^n = x$, the sample now has conditional distribution

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

Now assume W is a given matrix with weights where $W = \begin{bmatrix} W^0 & \cdots & W^G \end{bmatrix}$, where the elements W^g are matrices themselves consisting of all the weights at aggregation level g. Each row in W^g consists of at most one non-zero. So, the number of nonzero weights at a given row of W is at most $G+1$. The vector v contains the regression parameters which are unknown. Therefore we formulate a normal belief on v with mean vector v^n and covariance matrix $\Sigma^{v,n}$

$$v \sim N\left(v^n, \Sigma^{v,n}\right),$$

where v^n is the column vector $v^n = \begin{bmatrix} \mu^{0,n} & \cdots & \mu^{G,n} \end{bmatrix}^T$ where $\mu^{g,n} = \left(\mu_1^{g,n}, \ldots, \mu_{|\mathcal{X}^g|}^{g,n}\right)$, i.e., the vector of unique estimates at level g. Further, $\Sigma^{v,n}$ is the covariance matrix of v^n with size $\left(M^0 + \ldots + M^G\right) \times \left(M^0 + \ldots + M^G\right)$. We write $\Sigma^{v,n}\left(g, g'\right)$ to indicate a submatrix of the covariance matrix of size $M^g \times M^{g'}$ which provides the covariances between $\mu_x^{g,n}$ and $\mu_{x'}^{g',n}$.

Suppose we are at time n and we are going to measure one additional alternative x resulting in the observation $\hat{y}_x^{n+1}$. The posterior distribution of v can be computed by treating the prior as additional data points, and then weighing their contribution to the posterior (Gelman et al., 2004). Using the techniques described in (Gelman et al., 2004), it can be shown that the posterior mean and covariance are given by

$$\begin{aligned} v^{n+1} &= \left(\left(\Sigma^{v,n}\right)^{-1} + \left(W_x\right)^T W_x \frac{1}{\lambda_x}\right)^{-1} \left(\left(\Sigma^{v,n}\right)^{-1} v^n + \left(W_x\right)^T \frac{\hat{y}_x^{n+1}}{\lambda_x}\right), \\ \Sigma^{v,n+1} &= \left(\left(\Sigma^{v,n}\right)^{-1} + \left(W_x\right)^T W_x \frac{1}{\lambda_x}\right)^{-1}, \end{aligned}$$

where W_x denotes the x^{th} row of W.

The vector v^{n+1} of regression parameters contains predictions because $\hat{y}_x^{n+1}$ is still unknown. We are interested in the value $\mu^{n+1} = Wv^{n+1}$ where we, for illustrative purposes, still use the current weights W. Using the Sherman–Morrison formula we can rewrite the posterior mean of the vector v^{n+1} as follows

$$\begin{aligned}
v^{n+1} &= \left(\Sigma^{v,n} - \frac{\Sigma^{v,n}\left(W_x\right)^T W_x \frac{1}{\lambda_x}\Sigma^{v,n}}{1 + W_x\Sigma^{v,n}\left(W_x\right)^T \frac{1}{\lambda_x}}\right)\left(\left(\Sigma^{v,n}\right)^{-1} v^n + \frac{1}{\lambda_x}\left(W_x\right)^T \hat{y}_x^{n+1}\right) \\
&= v^n - \frac{\Sigma^{v,n}\left(W_x\right)^T W_x v^n}{W_x\Sigma^{v,n}\left(W_x\right)^T + \lambda_x} + \left(\Sigma^{v,n} - \frac{\Sigma^{v,n}\left(W_x\right)^T W_x\Sigma^{v,n}}{W_x\Sigma^{v,n}\left(W_x\right)^T + \lambda_x}\right)\frac{\hat{y}^x_{n+1}}{\lambda_x}\left(W_x\right)^T \\
&= v^n - \frac{\Sigma^{v,n}\left(W_x\right)^T W_x v^n}{W_x\Sigma^{v,n}\left(W_x\right)^T + \lambda_x} + \hat{y}^x_{n+1}\left(\frac{\Sigma^{v,n}\left(W_x\right)^T}{W_x\Sigma^{v,n}\left(W_x\right)^T + \lambda_x}\right) \\
&= v^n + \frac{\Sigma^{v,n}\left(W_x\right)^T}{W_x\Sigma^{v,n}\left(W_x\right)^T + \lambda_x}\left(\hat{y}_x^{n+1} - W_x v^n\right).
\end{aligned}$$

The predictive mean $\mu_{x'}^{n+1}$ of alternative x' after observing the value $\hat{y}_x^{n+1}$ of alternative x is given by

$$\begin{aligned}
\mu_{x'}^{n+1} &= W_{x'}^n v^{n+1} \\
&= W_{x'}^n v^n + \frac{W_{x'}^n\Sigma^{v,n}\left(W_x\right)^T}{W_x\Sigma^{v,n}\left(W_x\right)^T + \lambda_x}\left(\hat{y}_x^{n+1} - W_x v^n\right) \\
&= \mu_{x'}^n + \frac{W_{x'}^n\Sigma^{v,n}\left(W_x\right)^T}{W_x\Sigma^{v,n}\left(W_x\right)^T + \lambda_x}\left(\hat{y}_x^{n+1} - \mu_x^n\right).
\end{aligned}$$

Using the Sherman–Morrison formula, the posterior covariance matrix $\Sigma^{v,n+1}$ can be written as

$$\begin{aligned}
\Sigma^{v,n+1} &= \left(\left(\Sigma^{v,n}\right)^{-1} + \left(W_x\right)^T W_x \frac{1}{\lambda_x}\right)^{-1} \\
&= \Sigma^{v,n} - \frac{\Sigma^{v,n}\left(W_x\right)^T W_x\Sigma^{v,n}}{W_x\Sigma^{v,n}\left(W_x\right)^T + \lambda_x}.
\end{aligned}$$

Obviously, the treatment of weights as independent variables is not appropriate since they will be influenced by the measurement. Another disadvantage of this approach is that it requires a prior on v^n as noted in (Gelman et al., 2004, chap. 14). This can also be seen from the above updating equations. When no prior information is available, we have to rely on a non-informative prior where the diagonal elements of $\Sigma^{v,n}$ are infinite. Whenever we have an observation for some alternative x at some aggregation level g with weight less then one, the posterior variance remains infinite. Hence, this approach would not be appropriate without prior information.

Appendix B

Proposition 3 *The posterior belief on θ_x given observations up to time n for all aggregation levels is normally distributed with mean and precison*

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

Proof. Let $Y_x^{g,n} = \left\{\hat{y}_{x^{m-1}}^{g,m} : m \leq n, G^g(x) = G^g(x^{m-1})\right\}$. 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,

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

The second term may be rewritten using the induction statement as

$$\mathbb{P}_H\left\{\theta_x \in du\right\} = \varphi\left((u - \mu_x^{H,n})/\sigma_x^{H,n}\right).$$

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

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

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

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

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

The variance of the sample observation $\hat{y}_x^{n+1} = W_x^n v^{n+1} + \varepsilon$ is given by the measurement error plus a prediction error

$$\begin{aligned}
Var\left(\hat{y}_x^{n+1}\right) &= Var\left(W_x^n v^{n+1}\right) + \lambda_x \\
&= W_x^n \Sigma^{v,n+1} \left(W_x^n\right)^T + \lambda_x.
\end{aligned}$$

Therefore, the random variable $Z = \left(\hat{y}_x^{n+1} - W_x^n v^{n+1}\right) / \sqrt{W_x^n \Sigma^{v,n+1} \left(W_x^n\right)^T + \lambda_x}$ is a standard normal. So we write:

$$\mu^{n+1} = \mu^n + \tilde{\sigma}\left(\Sigma^{v,n+1}, x\right) Z,$$

where

$$\tilde{\sigma}\left(\Sigma^{v,n+1}, x\right) = \frac{W^n \Sigma^{v,n+1} \left(W_x^n\right)^T}{\sqrt{W_x^n \Sigma^{v,n+1} \left(W_x^n\right)^T + \lambda_x}}.$$

The knowledge-gradient policy can now be rewritten as

$$x^n = \arg\max_{x \in \mathcal{X}} \mathbb{E}\left[\max_{x' \in \mathcal{X}} \mu_{x'}^n + \left(e_{x'}\right)^T \tilde{\sigma}_{x'}\left(\Sigma^{v,n+1}, x\right) Z\right] - \max_{x' \in \mathcal{X}} \mu_{x'}^n,$$

where $\tilde{\sigma}_{x'}(\Sigma^{v,n+1}, x)$ denotes the row $\tilde{\sigma}(\Sigma^{v,n+1}, x)$ corresponding with alternative x'.

The sampling decision x^n is equal to the one proposed in (Frazier et al., 2009) with the exception of the column vector $\tilde{\sigma}(\Sigma^{v,n+1}, x)$. However, we can use the same approach to solve the sampling decision.

Note that without aggregation, which basically means we are only using the aggregation level $g = 0$ with weights $w^0_{x'} = 1$, $\tilde{\sigma}_{x'}(\Sigma^{v,n+1}, x)$ would simply be given by

$$\tilde{\sigma}_{x'}\left(\Sigma^{v,n+1}, x\right) = \Sigma^{v,n+1}_{x',x}\left(0,0\right) / \sqrt{\lambda_x + \Sigma^{v,n+1}_{xx}\left(0,0\right)}.$$

Given that $\Sigma^{v,n+1}_{x',x}\left(0,0\right) = \Sigma^{n+1}_{x',x}$ the resulting equation coincides with the results of Frazier et al. (2008).

Appendix D

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

Lemma 4 *If $z_1, z_2, \ldots$ is a sequence of non-negative real numbers bounded above by a constant $a < \infty$, and $s_n = \sum_{k \leq n} z_k$, then $\sum_n (z_n/s_n)^2 \mathbf{1}_{\{s_n > 0\}}$ is finite.*

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

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

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

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

The second term may be bounded by

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

■

Lemma 5 *Fix $g \in \mathcal{G}$ and $x \in \mathcal{X}$ and let*

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

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

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

Also let $M^n = (\bar{y}_x^n - \theta_x)^2 + \sum_{m=n}^{\infty}\mathbf{1}_{\{x^m=x\}}\lambda_x(\alpha^m)^2$, and note that Lemma 4 together with the upper bound $(\min_{x'}\lambda_{x'})^{-1}$ on $\beta_x^{g,m,\varepsilon}$ imply that M^0 is finite. We will show that M^n is a supermartingale with respect to the filtration generated by $(\hat{y}_x^n)_{n=1}^{\infty}$.

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

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

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 $\mathbb{E}^n\left[\left(\hat{y}_x^{n+1} - \theta_x\right)^2\right] = \lambda_x$, imply

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

Since $M^n \geq 0$ 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,

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

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 $\sup_n |\bar{y}_x^n| \leq \sup_n |\bar{y}_x^n - \theta_x| + |\theta_x| < \infty$ almost surely. ■

Lemma 6 *For each x, x', and g, the following quantities are almost surely finite:* $\sup_n |\mu_x^{g,n}|$, $\sup_n |a_{x'}^n(x)|$, *and* $\sup_n |b_{x'}^n(x)|$.

Proof. We begin by showing that $\sup_n |\mu_x^{g,n}|$ is almost surely finite. 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 5 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

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

By Lemma 5, $\sup_n |\bar{y}_{x'}^n|$ is almost surely finite, and hence so is $\sup_n |\mu_x^{g,n}|$.

We now turn our attention to $a_{x'}^n(x)$ and $b_{x'}^n(x)$. Both $a_{x'}^n(x)$ and $b_{x'}^n(x)$ are weighted linear combinations of the terms $\mu_{x'}^{g,n}$ (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 both $\sup_n |a_{x'}^n(x)|$ and $\sup_n |b_{x'}^n(x)|$ are almost surely finite. ■

Lemma 7 *Let $\mathcal{X}'$ be the (random) set of alternatives measured infinitely often by a policy. Then, for each $x \notin \mathcal{X}'$ and $x' \neq x$ that we measure at least once, and for each $g \in \mathcal{G}(x', x)$ for some $x' \neq x$, we have* $\liminf_n |\delta_x^{g,n}| \neq 0$.

Proof. The bias $\delta_x^{g,n}$ is given by $\mu_x^{0,n} - \mu_x^{g,n}$. Because $x \notin \mathcal{X}'$, there exists some $N < \infty$ such that $\mu_x^{0,n} = \mu_x^{0,N}$ for all $n \geq N$. Note that the estimate $\mu_x^{0,N}$ is a linear combination of finitely many normally distributed random variables, and is thus a continuous random variable.

Since $x \neq x'$ is measured at least once, no cluster point of the sequence $(\mu_x^{g,n})_{n\in\mathbb{N}}$ is perfectly correlated with $\mu_x^{0,N}$. Since the probability of equality between a continuous random variable and another random variable whose values are not perfectly correlated is 0, the probability that $\mu_x^{0,N}$ is equal to any cluster point of $(\mu_x^{g,n})_{n\in\mathbb{N}}$ is 0. This implies that $\liminf_n |\delta_x^{g,n}| \neq 0$ almost surely. ■

Lemma 8 *Let $\mathcal{X}'$ be the (random) set of alternatives measured infinitely often by a policy. Then, for each $x', x \in \mathcal{X}$, the following statements hold almost surely,*

- *if $x \in \mathcal{X}'$ then* $\lim_n b_{x'}^n(x) = 0$ *and* $\lim_n b_x^n(x') = 0$.
- *if $x \notin \mathcal{X}'$ then* $\liminf_n b_x^n(x) > 0$.

Proof. Let x' and x be any pair of alternatives. First consider the case $x \in \mathcal{X}'$.

When x' and x do not share any aggregation levels, the set $\mathcal{G}(x', x)$ is empty and hence $b_{x'}^{n+k}(x) = b_x^{n+k}(x') = 0$, from which $\lim_{n\to\infty} b_{x'}^n(x) = 0$ and $\lim_{n\to\infty} b_x^n(x') = 0$ follows trivially. When x' and x share one or more aggregation levels g, the precisions $\left\{\beta_x^{g,n} = \beta_{x'}^{g,n},\ \forall g \in G(x', x)\right\}$ are updated $m_{x'}^{0,n} + m_x^{0,n} = \infty$ times. As a consequence $\lim_{n\to\infty} \beta_x^{g,n} = \lim_{n\to\infty} \beta_{x'}^{g,n} = \infty$. Hence, in this case as well, we have $\lim_{n\to\infty} b_{x'}^n(x) = 0$ and $\lim_{n\to\infty} b_x^n(x') = 0$.

Now consider the case $x \notin \mathcal{X}'$. We show that $\liminf_n b_x^n(x) > 0$. Alternative x has at least one aggregation level, namely the disaggregate level $g = 0$, that is not shared with alternatives in $\mathcal{X}'$. As a consequence,

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

which is finite because $x \notin \mathcal{X}'$, i.e., $\beta_x^{0,n} \leq \beta_x^{N_1,0}$. In fact, if we started with a non-informative prior, $\beta_x^{0,n}$ would be given by $m_x^{0,n} (\lambda_x)^{-1}$. Therefore, we can write

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

where the weights are given by

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

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 that $\limsup_n \psi_x^{g,n} < \infty$ for all $g \in \mathcal{G} \setminus \{0\}$ by considering two cases. Define the set $\mathcal{G}'$ to contain all $g \in \mathcal{G}$ for which there exists an $x' \in \mathcal{X}'$ such that $\mathcal{G}(x', x)$ is not empty. This is the set of aggregation levels shared by x and an alternative measured infinitely often. For each $g \in \mathcal{G}'$ we have $\lim_n \beta_x^{g,n} = 0$, which implies $\limsup_n \psi_x^{g,n} = \limsup_n \left(\delta_x^{g',n} \right)^{-2}$, which is finite by Lemma 7. For each $g \notin \mathcal{G}'$ we know $\psi_x^{g,n}$ is constant for all n after the last measurement within $\mathcal{X}^g(x)$ and thus $\limsup_n \psi_x^{g,n}$ is finite in this case as well.

Finally, $\limsup_n \psi_x^{g,n} < \infty$ and $(\lambda_x)^{-1} > 0$ together imply that $\liminf_n \bar{w}_x^{0,n}(x) > 0$. This shows $\liminf_n b_x^n(x) > 0$. ■