Noname manuscript No.
(will be inserted by the editor)

Optimal Learning with a Local Parametric Belief Model

Bolong Cheng · Arta Jamshidi · Warren B. Powell

Received: date / Accepted: date

Abstract We are interested in maximizing smooth functions where observations are noisy and expensive to compute, as might arise in computer simulations or laboratory experimentations. We derive a knowledge gradient policy, which chooses measurements which maximize the expected value of information, while using a locally parametric belief model that uses linear approximations with radial basis functions. The method uses a compact representation of the function which avoids storing the entire history, as is typically required by nonparametric methods. Our technique uses the expected value of a measurement in terms of its ability to improve our estimate of the optimum, capturing correlations in our beliefs about neighboring regions of the function, without posing any assumptions on the global shape of the underlying function a priori. Experimental work suggests that the method adapts to a range of arbitrary, continuous functions, and appears to reliably find the optimal solution. Moreover, the policy is shown to be asymptotically optimal.

Keywords Ranking and selection · Optimal learning · Local parametric model · Stochastic search

1 Introduction

We consider the problem of maximizing an unknown function over a finite set of possible alternatives, where observations of the function are noisy and may be expensive. This problem arises under settings such as simulation-optimization,

Bolong Cheng
Department of Electrical Engineering, Princeton University, Princeton, NJ 08544
E-mail: bcheng@princeton.edu

Arta Jamshidi
Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544 E-mail: arta@princeton.edu

Warren B. Powell
Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544 E-mail: powell@princeton.edu

stochastic search, and ranking and selection. A popular strategy involves response surface methods which fit polynomial approximations to guide the search of the next observation (see [2], [8]). Our work was motivated by difficulties we encountered fitting parametric surfaces, even to relatively simple functions. Low order models can produce poor approximations, while higher order models quickly suffer from over fitting. This experience led us to consider a variety of statistical strategies, but ultimately produced a new local parametric procedure called Dirichlet Clouds with Radial Basis Functions (DC-RBF) [24]. This paper addresses the problem of doing stochastic search using the knowledge gradient ([12] and [34]), where the underlying belief model is represented using DC-RBF.

The optimization of noisy functions, or stochastic search, has been studied extensively since the seminal paper by [35] which introduces the idea of stochastic gradient algorithms for differentiable functions. [38] and [14] provide a thorough coverage of the literature for stochastic search methods. A separate line of research has evolved under the umbrella of active (or optimal) learning where observations are made specifically based on some measure of value of information (see [9] and [34] for reviews). For this problem class, we assume we do not have access to derivatives, and we assume that function evaluations are both noisy and expensive, as might arise when doing laboratory experiments. [19] was the first to introduce the idea of choosing measurements based on the marginal value of information to solve the ranking and selection problem. This policy is later extended under the name knowledge gradient using a Bayesian approach, where the value of information from an experiment is computed using the predictive distribution of the performance for a given set of parameters [12]. The knowledge gradient policy is shown to be the only stationary policy that is both myopically and asymptotically optimal. Most of the previous work in ranking and selection problems assumes the alternatives to be independent (alternatives close to each other do not exhibit correlation), see, e.g., [33].

There is a small literature that deals with correlated alternatives. [13] extends the knowledge gradient policy to take advantage of the covariance structure of alternatives using a lookup table belief model. [32] proposes an extension of the knowledge gradient where the beliefs can be represented by a linear parametric model. An adaptation of the knowledge gradient with correlated beliefs using kernel regression and aggregation of kernels for estimating the belief function is presented in [3]. The kernel version of knowledge gradient is built on the work presented in [29], where the estimates are the hierarchical aggregates of the values. A separate branch of statistical model based optimization research stems from the work of [28], where the true function is modeled as a realization of a Gaussian process. Earlier work on the optimization of one-dimensional functions use the Wiener process as prior, such as the algorithm proposed in [40]. [5]) due to its computational efficiency. A more general class of the Gaussian process is considered in the modern variations of this method. The sequential kriging optimization method from [22] fits a stationary Gaussian process onto the observed variables where the distance between alternatives dictates the correlation structure. Measurement points are selected by maximizing an expected improvement (EI) function. This is an extension of the well-known efficient global optimization (EGO) algorithm introduced in [25], which only considers noise-free observations. [39] introduces entropy minimization-based methods for determining the sampling decisions.

The online learning setting with discrete alternatives is studied in [18], using a method known as Gittins indices. However, Gittins indices are difficult to compute exactly, and only apply to a very specific problem (infinite horizon, independent beliefs). [6] proposes an approximation of the index for the infinite horizon discounted "Bayesian bandit" problem. An improved approximation is proposed in [7]. The knowledge gradient is extended to online (bandit) settings in [36]. The case of continuous decisions has been studied in [1], [17] and [37].

There are three major classes of function approximation methods: look-up tables, parametric models (linear or non-linear), and nonparametric models. Parametric regression techniques (such as linear regression, see [30]) assume that the underlying structure of the data is known a priori and is in the span of the regressor function. Due to the simplicity of this approach it is commonly used for regression. Neural networks have attracted considerable attention [21], but they are heavily dependent on the structure of the network and as a result require considerable tuning, just as parametric models are dependent on their assumed functional structure.

Nonparametric models ([10], [31]) offer the attraction of considerable generality by using the raw data to build local approximations of the function, producing a flexible but data-intensive representation. Nonparametric models are less sensitive to structural errors arising from a parametric model. Most nonparametric models require keeping track of all observed data points, which make function evaluations increasingly expensive as the algorithm progresses, a serious problem in stochastic search. Bayesian techniques for function approximation or regression are computationally intense and require storage of all the data points [15]. Local polynomial models [11] build linear models around each observation and keep track of all the data points.

Another class of approximation algorithms use local approximations around regions of the function, rather than each prior observation. [20] uses the notion of Dirichlet processes to create clusters using a Markov chain Monte Carlo method, and then fits linear models around each cluster. Radial basis functions have attracted considerable attention due to their simplicity and generality. One of the main attractions of the radial basis functions (RBFs) is that the resulting optimization problem can be broken efficiently into linear and nonlinear subproblems. Normalized RBFs are presented in [26] which perform well with limited training data. For a comprehensive treatment on various growing RBF techniques and automatic function approximation technique using RBF, see [23] and the references therein. See [24] for a more comprehensive review of various approximation techniques.

The Dirichlet Cloud Radial Basis Function approximation strategy (DC-RBF) is motivated by the need to approximate functions within stochastic search algorithms where new observations arrive iteratively. As we obtain new information from each iteration, DC-RBF provides a fast and flexible method for updating the approximation. DC-RBF is more flexible than classical parametric models, and provides a compact representation to minimize computational overhead. Unlike similar algorithms in the literature, our method has only one tunable parameter, assuming that the input data has been properly scaled.

This paper makes the following contributions: 1) We derive the knowledge gradient while using a DC-RBF belief model; 2) we develop a hierarchical version of DC-RBF, and derive the knowledge gradient for this belief model; 3) experimen-

tal testing shows that knowledge gradient with DC-RBF model, or its hierarchical version, generally outperforms the knowledge gradient using a nonparametric, kernel belief model (previously shown to be highly competitive against a wide range of algorithms), while scaling to higher dimensional problems.

The rest of the paper is organized as follows. Section 2 formulates the ranking and selection model and establishes the notation used in this paper. Section 3 reviews the knowledge gradient for correlated alternatives using both lookup table and a linear, parametric belief model. Section 4 reviews the DC-RBF approximation technique that is used for constructing the belief model in this paper and derives the knowledge gradient using the DC-RBF belief model. Section 5 proposes a hierarchical approach to the DC-RBF method to overcome shortcomings. In Section 6, we present an asymptotic convergence proof. Section 7 showcases the performance of the new KG-RBF using examples drawn from different problem classes.

2 Model

We consider a finite set of alternatives $\mathcal{X} = \{1, 2, ..., M\}$. Each alternative $x \in \mathcal{X}$ is associated with a normal distribution with an unknown mean μ_x and a known variance λ_x. We denote μ as the vector $[\mu_1, \ldots, \mu_M]'$. Now suppose we have a sequence of N measurement decisions, $x^0, x^1, \ldots, x^{N-1}$ to learn about these alternatives. At time n, if we measure alternative x, we observe

$$\hat{y}_x^{n+1} = \mu_x + \epsilon_x^{n+1},$$

with the sampling error ϵ_x^{n+1} being independent conditioned on $x^n = x$, and normally distributed with mean 0 and known variance λ_x. We also use the notation $\beta_x^\epsilon = (\lambda_x)^{-1}$ to denote the precision.

In this sequential sampling framework, it is natural to define the filtration $\mathcal{F}^n$ as the σ-algebra generated by $\{(x^0, \hat{y}_{x^0}^1), (x^1, \hat{y}_{x^1}^2), \ldots, (x^{n-1}, \hat{y}_{x^{n-1}}^n)\}$. It is apparent that any random variable with the superscript n is $\mathcal{F}^n$-measurable. Under the Bayesian setting, we assume a multivariate normal prior distribution on the value of μ as our initial estimate

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

Following the definition of $\mathcal{F}^n$, we denote $\theta^n := \mathbb{E}[\mu|\mathcal{F}^n]$, and $\Sigma^n := \mathrm{Cov}[\mu|\mathcal{F}^n]$ as the conditional mean and covariance of the posterior distribution of μ at time n.

For an offline ranking and selection problem, the objective is to find the optimal alternative after N measurements, where the final sampling decision is

$$x^N = \arg\max_{x \in \mathcal{X}} \theta_x^N.$$

Let Π be the set of all possible measurement policies that satisfies $x^n \in \mathcal{F}^n$, and $\pi = (x^0, \ldots, x^{N-1})$ be a generic policy in Π. We write $\mathbb{E}^\pi$ as the expectation taken when the policy π is used, we can express the problem of finding a sampling policy maximizing the expected reward as

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

Table 1 gives a summary of the notations used in this paper.

Table 1: Table of Notation

Variable	Description
$\mathcal{X}$	Set of alternatives
N	Maximum number of measurements
M	Number of alternatives, $M = \|\mathcal{X}\|$
μ_x	Unknown true sampling mean of alternative x
λ_x	Known sampling variance of alternative x
β_x^ϵ	Precision of a measurement, where $\beta_x^\epsilon = (\lambda_x)^{-1}$
μ	Column vector $(\mu_1, \ldots, \mu_M)'$
θ^n	Mean of the posterior distribution on θ after n measurements
Σ^n	Covariance of the posterior distribution on the alternatives after n measurements
β_x^n	Precision of the posterior distribution of alternative x at time n, $\beta_x^n = (\sigma_x^{2,n})^{-1}$
$\hat{y}_x^{n+1}$	Sampling observation from measuring alternative x^n
ϵ^{n+1}	Measurement error of alternative x^n, $\epsilon^{n+1} \sim \mathcal{N}(0, \lambda_{x^n})$
x^n	Measurement decision at time n
α	Can be viewed as a vector of coefficients, typically of lower dimension than θ
X	Matrix in which each row is an alternative, where $\mu = X\alpha$
$\Sigma^{\theta,n}$	Covariance of posterior distribution on the parameters after n measurements
N_c	Number of clouds
D_T	Prespecified threshold distance separating clouds
c_i	Center of cloud U_i
W_i	Width of cloud U_i

2.1 Bayesian Inference

We work in the Bayesian setting to sequentially update our estimate of the alternatives. At time n, suppose we select $x^n = x$ and observe $\hat{y}_x^{n+1}$, we can compute the $n+1$ time posterior distribution with the following updating equations. ([15])

$$\theta^{n+1} = \theta^n + \frac{\hat{y}^{n+1} - \theta_x^n}{\lambda_x + \Sigma_{xx}^n} \Sigma^n e_x, \tag{1}$$

$$\Sigma^{n+1} = \Sigma^n - \frac{\Sigma^n e_x e_x' \Sigma^n}{\lambda_x + \Sigma_{xx}^n}, \tag{2}$$

where e_x is the standard basis vector. We can further rearrange equation (1) as the time n conditional distribution of θ^{n+1}, namely

$$\theta^{n+1} = \theta^n + \tilde{\sigma}(\Sigma^n, x^n) Z, \tag{3}$$

where

$$\begin{aligned} \tilde{\sigma}(\Sigma^n, x) &= \sqrt{\mathrm{Var}\left[\theta^{n+1} - \theta^n | \Sigma^n\right]}, \\ &= \frac{\Sigma^n e_x}{\sqrt{\lambda_x + \Sigma_{xx}^n}}, \end{aligned} \tag{4}$$

and Z is a standard normal random variable.

3 Knowledge Gradient with Correlated Beliefs

The knowledge gradient with correlated beliefs (KGCB) is introduced in [13] as a sequential decision policy for learning alternatives with correlation. We represent

our state of knowledge at time n as $S^n := (\theta^n, \Sigma^n)$. At time n, if we stop measuring now, the corresponding value of being in state S^n is represented by the estimate of the best alternative:

$$V^n(S^n) = \max_{x' \in \mathcal{X}} \theta^n_{x'}.$$

Suppose we are allowed one more measurement $x^n = x$, from which we observe $\hat{y}^{n+1}_x$, allowing us to use the Bayesian updating equations to reach the new knowledge state $S^{n+1}(x)$. The value of this new state is

$$V^{n+1}(S^{n+1}(x)) = \max_{x' \in \mathcal{X}} \theta^{n+1}_{x'}.$$

We would like to maximize the expected value of $V^{n+1}(S^{n+1}(x))$ at time n. The incremental value of the knowledge state due to measurement x is defined as:

$$\begin{aligned}
\nu^{KG,n}_x &= \mathbb{E}\left[V^{n+1}(S^{n+1}(x)) - V^n(S^n) | S^n, x^n = x\right], \\
\nu^{KG,n}_x &= \mathbb{E}\left[\max_{x' \in \mathcal{X}} \theta^{n+1}_{x'} | S^n, x^n = x\right] - \max_{x' \in \mathcal{X}} \theta^n_{x'}.
\end{aligned} \tag{5}$$

The knowledge gradient policy is a sampling decision that maximizes this expected incremental value, namely:

$$x^{KG,n} = \arg\max_{x \in \mathcal{X}} \nu^{KG,n}_x. \tag{6}$$

[13] provides an algorithm to compute the KG values for alternatives with correlated beliefs. First, we substitute equation (3) into the KG formula

$$\begin{aligned}
\nu^{KG,n}_x &= \mathbb{E}\left[\max_{x' \in \mathcal{X}} \theta^n_{x'} + \tilde{\sigma}_{x'}(\Sigma^n, x^n) Z | S^n, x^n = x\right] - \max_{x' \in \mathcal{X}} \theta^n_{x'} \\
&= h\left(\theta^n, \tilde{\sigma}(\Sigma^n, x)\right),
\end{aligned}$$

where $h(a, b) = \mathbb{E}\left[\max_i a_i + b_i Z\right] - \max_i a_i$ is a generic function of any vectors a and b of the same dimension. The expectation can be computed as the point-wise maximum of affine functions $a_i + b_i Z$ with an algorithm of complexity $O(M^2 \log(M))$.

The algorithm first sorts the alternatives with b_i in increasing order, then removes terms a_i, b_i if there is some i' such that $b_i = b_{i'}$ and $a_i > a_{i'}$ (i.e. removing parallel slopes with lower constant intercepts). Then it removes redundant $a_{i'}, b_{i'}$ if for all $Z \in \mathbb{R}$ there exists some i such that $i \neq i'$ and $a_{i'} + b_{i'} Z \leq a_i + b_i Z$. After all of the redundant components are dropped, we will have new vectors $\tilde{a}$ and $\tilde{b}$ of dimension $\tilde{M}$. Essentially, we are left with a concave set of affine functions. [13] then shows that the function $h(a, b)$ can be computed via

$$h(a, b) = \sum_{i=1,\ldots,\tilde{M}} (\tilde{b}_{i+1} - \tilde{b}_i) f\left(-\left|\frac{\tilde{a}_i - \tilde{a}_{i+1}}{\tilde{b}_{i+1} - \tilde{b}_i}\right|\right), \tag{7}$$

where $f(z) = \phi(z) + z\Phi(z)$. Here, $\phi(z)$ and $\Phi(z)$ are the normal density and cumulative distribution functions respectively.

3.1 Knowledge Gradient with Linear Belief Model

[32] further extends the KGCB to parametric beliefs using a linear model, namely representing the truth μ as a linear combination of a set of parameters i.e. $\mu = X\alpha$, where elements of α are the true coefficients of the parameters. Instead of maintaining a belief on the alternatives, we can maintain a belief on the coefficients. If we characterize α by a normal distribution, i.e. $\alpha \sim \mathcal{N}(\theta, \Sigma^\theta)$, we can generate a normal distribution on μ via linear transformation,

$$\mu \sim \mathcal{N}(X\theta, X\Sigma^\theta X^T).$$

Note that θ now represents the estimate of coefficients instead of the alternatives. Furthermore, we use Σ^θ to denote the covariance of the parameters. This linear transformation applies for prior and posterior distributions. At time n, if we measure alternative $x^n = x$, we can update θ^{n+1} and $\Sigma^{\theta,n+1}$ recursively via

$$\theta^{n+1} = \theta^n + \frac{\hat{\epsilon}^{n+1}}{\gamma^n}\Sigma^{\theta,n}x^n, \tag{8}$$

$$\Sigma^{\theta,n+1} = \Sigma^{\theta,n} - \frac{1}{\gamma^n}\left(\Sigma^{\theta,n}x^n(x^n)^T\Sigma^{\theta,n}\right), \tag{9}$$

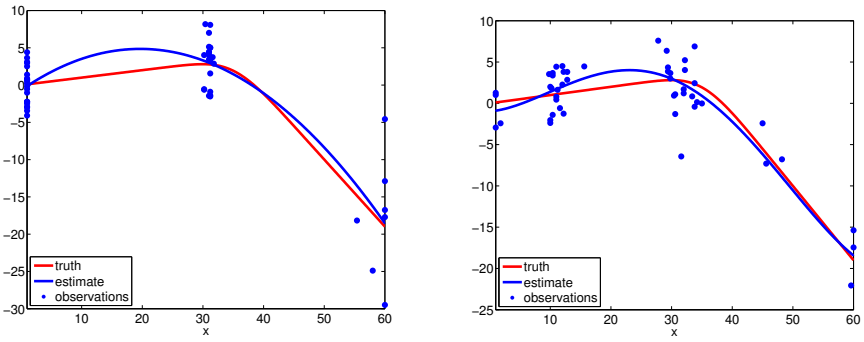
where we define $\hat{\epsilon}^{n+1} = \hat{y}^{n+1} - (\theta^n)^T x^n$ and $\gamma^n = \lambda_{x^n} + (x^n)^T\Sigma^{\theta,n}x^n$.

The parametric model allows us to represent the alternatives in a compact format since the dimensionality of the parameters is usually much smaller than that of the alternatives. Suppose we have tens of thousands of alternatives; using a lookup table belief model (as was used in [13]) would need to create and update the covariance matrix Σ^n with tens of thousands of rows and columns. With a parametric belief model, we only need to maintain the parameter covariance matrix $\Sigma^{\theta,n}$, which is determined by the dimensionality of the parameter vector θ. In addition, we never need to compute the full matrix $X\Sigma^{\theta,n}X^T$, although we will have to compute a row of this matrix.

KG with a linear, parametric model requires assuming that the parametric model is correct. However, knowing the right parametric model or parameter space is not always achievable in reality. For example, if we use a quadratic model to estimate an asymmetric unimodal function (details explained in Section 4.3), the knowledge gradient may not obtain the optimal alternative, as shown in Figure 1a. Performance can only be improved if we increase the number of parameters, as shown in in Figure 1b.

4 Dirichlet Cloud Radial Basis Function Model

The DC-RBF is a locally parametric model which introduces a cover over the input space to define local regions [24]. This scheme only stores a statistical representation of data in local regions and approximates the function locally with a low order polynomial such as a linear model, thereby avoiding the need to store the entire history of observations as required by classical nonparametric methods [11]. A nonlinear weighting system is associated with each local model which determines the contribution of this local model to the overall model output. The combined effect of the local approximations and the weights associated with them



(a) Quadratic parametric model (b) 4th-order parametric model

Fig. 1: Comparison of KG with different order polynomial models on an asymmetric unimodal function.

produces a nonlinear function approximation method. This technique is efficient, both in terms of computational time and memory requirements compared to classical nonparametric regression methods, which requires retaining the entire history.

Suppose now instead of the linear relationship $\mu = X\alpha$, we have the nonlinear relationship

$$\mu_x = \frac{\sum_{i=1}^{N_c} \varphi_i(x|c_i, W_i) x^T \alpha_i}{\sum_{i=1}^{N_c} \varphi_i(x|c_i, W_i)}, \tag{10}$$

where φ_i is a radial basis function with center c_i and width W_i. In our model, we employ the Gaussian kernel function

$$\varphi(x|c, W) = \exp(-\|x - c\|_W^2),$$

where $\|z\|_W = \sqrt{z^T W^{-1} z}$. N_c is the total number of clouds, which may grow with an increasing number of observations. Figure 2 shows the approximation of a quadratic function with locally linear models constructed by the DC-RBF method. The light blue lines represent the local linear parametric models. Vertical lines are the center of the clouds and the shaded area shows the standard deviation of each cloud.

4.1 Recursive Update of the Model

First, we define a cloud as a ball formed with the center c_i and with a (user specified) threshold radius of D_T. At time n, we measure alternative $x^n = x$, and observe $\hat{y}_x^{n+1}$. We then need to assign the data point x^n to a cloud; the resulting cloud I is determined by the closest L^2 distance

$$I = \arg\min_i D_i = \arg\min_i \|x^n - c_i\|. \tag{11}$$

We compare D_I to the threshold distance D_T. If $D_I > D_T$, we update the corresponding cloud I with the new observation; otherwise we spawn a new cloud and

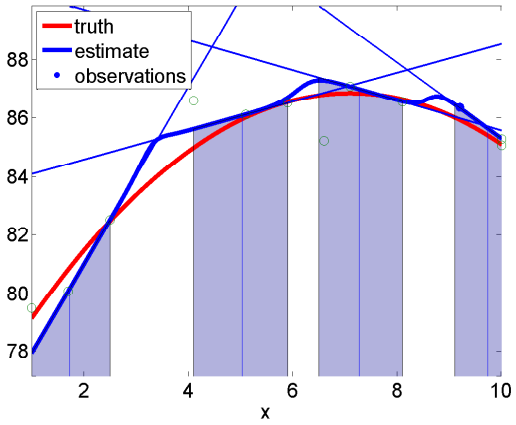


Fig. 2: Estimates of a quadratic function using DC-RBF with $D_T = 2$.

update the total number of clouds $N_c = N_c + 1$. To update a particular cloud I, we need to update the local regression model and the associated radial basis function. Let k_I be the number of data points for cloud I (note that $\sum_i k_i = n$). If the n-th data point is associated with cloud I, we can update the local parametric model with recursive least squares using:

$$\theta_I^{k_I+1} = \theta_I^{k_I} + \frac{\hat{\epsilon}^{k_I+1}}{\gamma_I^{k_I}} \Sigma_I^{\theta,k_I} x^n, \tag{12}$$

$$\Sigma_I^{\theta,k_I+1} = \Sigma_I^{\theta,k_I} - \frac{1}{\gamma_I^{k_I}} \left(\Sigma_I^{\theta,k_I} x^n (x^n)^T \Sigma_I^{\theta,k_I} \right), \tag{13}$$

where we define $\hat{\epsilon}^{k_I+1} = \hat{y}^{n+1} - (\theta_I^{k_I})^T x^n$ and $\gamma_I^{k_I} = \lambda_{x^n} + (x^n)^T \Sigma_I^{\theta,k_I} x^n$ as usual.

To update the basis function associated with cloud I, we have to update the center and the width that defines it. We first update the center of the basis function,

$$c_I^{k_I+1} = c_I^{k_I} + \frac{x^n - c_I^{k_I}}{k_I}. \tag{14}$$

Then we update the width of the kernel via the Welford formula for computing empirical variance [27]. Note that W_I is computed for each dimension separately,

$$S^{k_I} = S^{k_I-1} + (x^n - c^{k_I-1})(x^n - c^{k_I}). \tag{15}$$

The k_I-th estimate of the variance matrix is $W_I = \frac{S^{k_I}}{k_I - 1}$.

4.2 Knowledge Gradient with DC-RBF Belief Model

It is apparent that equation (10) is a weighted sum of multiple linear models. From the updating equations of DC-RBF, each data point is used to update the local estimate of one cloud, implying the coefficient vectors $\alpha_i, i = 1, \cdots, N_c$ are

independent. We impose a normal belief on each α_i such that $\alpha_i \sim \mathcal{N}(\theta_i, \Sigma_i^\theta)$. Let μ be the weighted estimate of the function at each point, which is a linear combination of independent normal random variables, implying that μ is also a normal random variable, where the distribution is expressed as

$$\mu \sim \mathcal{N}\left(\sum_{i=1}^{N_c}\left(\frac{\varphi_i}{\sum_{j=1}^{N_c}\varphi_j}\right)X\theta_i, \sum_{i=1}^{N_c}\left(\frac{\varphi_i}{\sum_{j=1}^{N_c}\varphi_j}\right)X\Sigma_i^\theta X^T\left(\frac{\varphi_i}{\sum_{j=1}^{N_c}\varphi_j}\right)^T\right).$$

To understand this expression, we first note that the mean of the distribution is equation (10) in vector form. Each row of X is a data point and $f(X|\theta_i) = X\theta_i$ maps the local parametric model from the parameter space (in $\mathbb{R}^d$) to the space of the alternatives (in $\mathbb{R}^M$, where $M \gg d$). Each φ_i is a $M \times M$ diagonal matrix capturing the weight on each local approximation relative to a particular cloud i (i.e. the basis function), given by

$$\varphi_i = \begin{bmatrix} \exp(-||x_1 - c_i||^2_{W_i}) & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \exp(-||x_M - c_i||^2_{W_i}) \end{bmatrix}.$$

The normalized RBF dictates where the local model is weighted in the data space, with the weight being close to 1 near the center of the cloud and diminishing quickly when moving away from the cloud. Since φ_i is diagonal, the division is performed element-wise. In the second part, the covariance of the normal distribution follows similarly. Each $\Sigma_i = X\Sigma_i^\theta X^T$ transforms the covariance of the parameters to the covariance of the alternatives. We need to multiply the covariance matrices by the respective radial basis functions to define the local weighting scheme and finally sum up all the clouds.

Let $\Theta^n = \{\theta_1^n, \ldots, \theta_{N_c}^n\}$ be the set of local parametric coefficients at time n. Working from our expression for the estimate of the function, we can derive the knowledge gradient using the RBF belief structure using

$$f(x|\Theta^{n+1}) = \frac{\sum_{j=1}^{N_c}\varphi_j^{n+1}x^T\theta_j^{n+1}}{\sum_{j=1}^{N_c}\varphi_j^{n+1}}. \tag{16}$$

To simplify our notation, we write $\varphi_j^{n+1}(x|c_i, W_i)$ as φ_i^{n+1}. Given x^n is selected at time n, and assuming x^n belongs to cloud I, we can rewrite θ_x^{n+1} as,

$$f(x|\Theta^{n+1}) = \frac{\sum_{j\neq I}\varphi_j^n x^T\theta_j^n}{\sum_{j=1}^{N_c}\varphi_j^{n+1}} + \frac{\varphi_I^{n+1}x^T\theta_I^{n+1}}{\sum_{j=1}^{N_c}\varphi_j^{n+1}}.$$

To simplify the expression, we let $K^{n+1} = \sum_{j=1}^{N_c}\varphi_j^{n+1}$. We also substitute θ_I^{n+1} with the recursive least square update from equation (12), where $\gamma_I^n = \lambda_{x^n} + (x^n)^T\Sigma_I^{\theta,n}x^n$. This allows us to rewrite the above expression as,

$$\begin{aligned} f(x|\Theta^{n+1}) &= \frac{\sum_{j\neq I}\varphi_j^n x^T\theta_j^n}{K^{n+1}} + \frac{\varphi_I^{n+1}x^T(\theta_I^n + \frac{\hat{y}^{n+1} - x^T\theta_I^n}{\gamma_I^n}\Sigma^{\theta,n}x^n)}{K^{n+1}} \\ &= \frac{\sum_{j=1}^{N_c}\varphi_j^{n+1}x^T\theta_j^n}{K^{n+1}} + \frac{\varphi_I^{n+1}(\hat{y}_{x^n}^{n+1} - x^T\theta_I^n)x^T\Sigma^{\theta,n}x^n}{K^{n+1}\gamma_I^n} \end{aligned}$$

$$= \frac{\sum_{j=1}^{N_c} \varphi_j^{n+1} x^T \theta_j^n}{K^{n+1}} + \frac{\varphi_I^{n+1}(f(x^n|\Theta^n) - x^T\theta_I^n)x^T\Sigma^{\theta,n}x^n}{K^{n+1}\gamma_I^n}$$
$$+ \frac{\varphi_I^{n+1}(\hat{y}_{x^n}^{n+1} - f(x^n|\Theta^n))x^T\Sigma^{\theta,n}x^n}{K^{n+1}\gamma_I^n}$$
$$= \frac{\sum_{j=1}^{N_c} \varphi_j^{n+1} x^T \theta_j^n}{K^{n+1}} + \frac{\varphi_I^{n+1}(f(x^n|\Theta^n) - x^T\theta_I^n)x^T\Sigma^{\theta,n}x^n}{K^{n+1}\gamma_I^n} + \tilde{\sigma}(x, x^n, I)Z,$$

where $Z = (\hat{y}_{x^n}^{n+1} - f(x^n|\Theta^n))/\sqrt{\lambda + \Sigma^n_{x^n x^n}}$ is a standard normal random variable and

$$\tilde{\sigma}(x, x^n, I) = \frac{\varphi_I^{n+1}(\sqrt{\lambda + \Sigma^n_{x^n x^n}})x^T\Sigma^{\theta,n}x^n}{K^{n+1}\gamma_I^n}.$$

We now have obtained the knowledge gradient,

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

where

$$a_x^n(x^n) = \frac{\sum_{j=1}^{N_c} \varphi_j^{n+1} x^T \theta_j^n}{K^{n+1}} + \frac{\varphi_I^{n+1}(f(x^n|\Theta^n) - x^T\theta_I^n)x^T\Sigma^{\theta,n}x^n}{K^{n+1}\gamma_I^n}, \quad (17)$$

$$b_x^n(x^n) = \tilde{\sigma}(x, x^n, I). \quad (18)$$

This is in the form of equation (7), therefore we can apply the computation method described for KG with correlated belief.

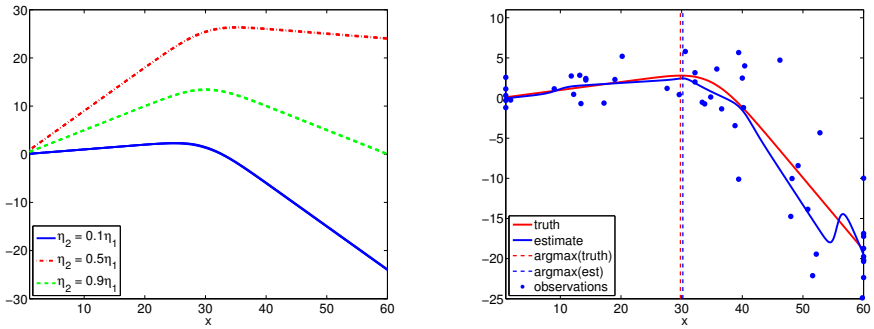
4.3 Demonstration of KG-RBF

In this section, we revisit the asymmetric unimodal function first illustrated in Figure 1, and demonstrate the advantage of local parametric models over the parametric model when optimizing this function with knowledge gradient. Our function is defined as

$$f(x) = \mathbb{E}\left[\eta_1 \min(x, D)\right] - \eta_2 x,$$

where D is a random variable and $\eta = (\eta_1, \eta_2)$ are fixed parameters with $\eta_2 < \eta_1$. Note that when $\eta_2 = \eta_1/2$, the function is symmetric. The function becomes asymmetric when η_2 is closed to η_1 or 0. Figure 3a shows three functions with different η_2 settings. Furthermore, we only get to observe a noisy realization of $f(x)$ at each iteration $\hat{f}(x) = \eta_1 \min(x, D) - \eta_2 x + \epsilon$. This makes the problem much harder, since we have heteroscedastic noise (the noise of a measurement is much higher when $x > D$). In Figure 3b, we pick $\eta_2 = 0.9$, and D is sampled from $\mathcal{N}(35, 4)$. A small variance of the distribution causes the function to have a "sharper" drop, whereas a larger variance yields a smoother decline. In this particular example, we again set the measurement noise variance $\lambda = 10$, along with prior beliefs $\theta^0 = \mathbf{0}$ and $\Sigma^0 = 10^4 I$.

For this particular problem, we set the distance threshold $D_T = 10$. There are five clouds generated in the DC-RBF model. We observe that KG-RBF quickly converges to the optimal alternatives after 50 iterations, evidenced by the sampling



(a) Asymmetric unimodal functions with $\eta_1 = 1$ and $\eta_2 = \{0.1, 0.5, 0.9\}$.

(b) Estimates of an asymmetric unimodal function using KG-RBF with $D_T = 10$.

Fig. 3: Examples of the asymmetric unimodal functions and an approximation of the function using KG-RBF.

decisions concentrating around its neighborhood. However, the estimation of the right hand side (the decreasing slope) is poor due to under-sampling and high measurement noise. Although this does not affect the outcome of the algorithm, we still want the local parametric model to be robust. In the subsequent section, we propose an hierarchical approach of estimation to avoid this behavior.

5 A Hierarchical Approach for Estimation

To address the issue caused by undersampling, we propose a hierarchical approach to estimate the posterior mean and variance using a weighted sum of multiple estimators. We further restrict the weights to be non-negative and sum up to one (convex combination).

We want to combine multiple levels of estimation with different D_T values. The motivation is to rely on clouds with large D_T in the early iterations to estimate a rough global model (i.e. exploration), then gradually use basis functions with smaller D_T values to construct finer local models and reduce bias of individual estimates (i.e. exploitation). This can be applied to any number of levels theoretically; however, we only use three levels to limit computational overhead.

We denote $(f(x|\Theta^0), \beta_x^0)$ as the prior mean and precision of the alternative. In practice we use an uninformative prior, where $f(x|\Theta^0) = 0, \beta_x^0 = 0$, for all $x \in \mathcal{X}$. We assume the measurements for the base level (independent belief) are unbiased. For all other levels, we need to account for bias.

Proposition 1 *Let $\mathcal{G}$ be the set of all aggregate levels, Θ^g be the set of parameters of level g, and $f(x|\Theta^g)$ be the estimates of x at level g, the final aggregated posterior mean and variance can be expressed as*

$$f(x|\Theta^{AG,n}) = \frac{1}{\beta_x^{AG,n}} \left(\beta_x^0 f(x|\Theta^0) + \sum_{g \in \mathcal{G}} ((\sigma_x^{g,n})^2 + (v_x^{g,n})^2)^{-1} f(x|\Theta^g) \right), \quad (19)$$

$$\beta_x^{AG,n} = \beta_x^0 + \sum_{g\in\mathcal{G}}((\sigma_x^{g,n})^2 + (v_x^{g,n})^2)^{-1}, \tag{20}$$

where $(\sigma_x^{g,n})^2 := Var(\mu_x^g|\mathcal{F}^n)$ *and* $(v_x^{g,n})^2 := (Bias(\mu_x^{g,n}|\mathcal{F}^n))^2 = (\mathbb{E}^n[\mu_x^{g,n} - \mu_x])^2$.

In practice, we approximate the bias $v_x^g = f(x|\Theta^g) - f(x|\Theta^0)$. We follow [29] by setting the weight on each level of aggregation as inversely proportional to the sum of bias squared and variance, namely:

$$w_x^{g,n} \propto \left((\sigma_x^{g,n})^2 + (v_x^{g,n})^2\right)^{-1}.$$

Proof We use induction to derive the weighting system. Let $\mathcal{C}$ be a generic subset of $\mathcal{G}$. We want to show that for $\mathcal{C}$ of any arbitrary size, equations (19) and (20) hold, then we can take $\mathcal{C} = \mathcal{G}$ to complete the proof.

First, consider $\mathcal{C} = \emptyset$, then the equations clearly hold since the posterior will be the same as the prior $(f(x|\Theta^0), \sigma_x^0)$ if there is no level. Now, suppose the equations hold for all $\mathcal{C}$ of size m, and consider $\mathcal{C}'$ with $m+1$ levels where $\mathcal{C} = \mathcal{C}' \setminus \{k\}$. By Bayes' rule, we obtain

$$P_{\mathcal{C}'}(\mu_x \in du) = P_{\mathcal{C}}(\mu_x \in du|Y_x^k = y) \propto P_{\mathcal{C}}(Y_x^k \in dy|\mu_x = u)P_{\mathcal{C}}(\mu_x \in du).$$

Y_x^k denotes the set of observations for level k. In practice there is only one set of observations for every level, in here it is for mathematical formality that we assume observations from each level are from a different distribution. Using our induction hypothesis, the second part of the product can be expressed as

$$P_{\mathcal{C}}(\mu_x \in du) = \phi\left((u - \mu_x^{\mathcal{C},n})/\sigma_x^{\mathcal{C},n}\right).$$

For the first part of the product, we assume independence between different levels, therefore

$$\begin{aligned}
P_{\mathcal{C}}(Y_x^k \in dh|\mu_x = u) &= P(Y_x^k \in dh|\mu_x = u) \\
&= \int_{\mathbb{R}} P(Y_x^k \in dh|\mu_x^g = w)P(\mu_x^g = w|\mu_x = u)dw \\
&\propto \int_{\mathbb{R}} \phi\left(\frac{f(x|\Theta^{k,n}) - w}{\sigma_x^{k,n}}\right)\phi\left(\frac{w-u}{v^k}\right)dw \\
&\propto \phi\left(\frac{f(x|\Theta^{k,n}) - u}{\sqrt{(\sigma_x^{k,n})^2 + (v_x^{k,n})^2}}\right).
\end{aligned}$$

Combining the two parts of the product we have

$$P_{\mathcal{C}'}(\mu_x \in du) \propto \phi\left(\frac{f(x|\Theta^{k,n}) - u}{\sqrt{(\sigma_x^{k,n})^2 + (v_x^{k,n})^2}}\right)\phi\left(\frac{u - \mu_x^{\mathcal{C},n}}{\sigma_x^{\mathcal{C},n}}\right) \propto \phi\left(\frac{u - f(x|\Theta^{C',n})}{\sigma_x^{\mathcal{C}',n}}\right),$$

which can be derived by completing the square.

Consequently, we need to update the KG derivation to reflect the aggregation in estimation. Equation (16) now only applies to an individual level; therefore we index the parameters of each level with a superscript g, e.g. $\Theta^{g,n}$. Furthermore, we let $\Theta^{\mathcal{G},n}$ be the union of $\Theta^{g,n}$ for all $g \in \mathcal{G}$. Following this convention, we see the $n+1$ time estimate for alternative x becomes

$$\begin{aligned} f(x|\Theta^{\mathcal{G},n+1}) &= \sum_{g\in\mathcal{G}} w_x^{g,n+1} f(x|\Theta^{g,n+1}) \\ &= \sum_{g\in G} w_x^{g,n+1} \frac{\sum_{j=1}^{N_c(g)} \varphi_j^{g,n+1} x^T \theta_j^{g,n+1}}{\sum_{j=1}^{N_c(g)} \varphi_j^{g,n+1}}. \end{aligned}$$

We see that the posterior $f(x|\Theta^{\mathcal{G},n+1})$ is a weighted sum of the estimates of the individual levels; therefore, the corresponding vectors $(a_x^n(x^n), b_x^n(x^n))$ are also weighted sums of,

$$a_x^n(x^n) = \sum_{g\in\mathcal{G}} w_x^{g,n} a_x^{g,n}(x^n), \tag{21}$$

$$b_x^n(x^n) = \sum_{g\in\mathcal{G}} w_x^{g,n} b_x^{g,n}(x^n), \tag{22}$$

where $a_x^{g,n}$ and $b_x^{g,n}(x^n)$ follow the same expression from equations (17) and (18) computed for each individual level $g \in \mathcal{G}$.

6 Convergence Results

In this section, we show that the HKG-RBF policy is asymptotically optimal almost surely. In other words, HKG-RBF learns the true value of each alternative as measurement budget $N \to \infty$, meaning that we will find the global optimum with probability one. The assumptions and lemmas (with proofs) used in the body of the paper are provided in Appendix A.

Theorem 1 *The HKG-RBF policy measures every alternative $x \in \mathcal{X}$ infinitely often, almost surely.*

Proof We start by defining Ω_0 as the almost sure event for which our assumptions and lemmas hold. For any $\omega \in \Omega_0$, let $\mathcal{X}'(\omega)$ be the random set of alternatives measured infinitely often by the KG-RBF policy. Assume that there is a set $G \subset \Omega_0$, with strictly positive probability such that for all $\omega \in G$, $\mathcal{X}'(\omega) \subset \mathcal{X}$ (strictly subset). In other words, we measure at least one alternative for a finite number of times with positive probability. Fix an $\omega \in G$, and let L be the last time we measure an alternative outside $\mathcal{X}'(\omega)$ for this ω.

Let $x \in \mathcal{X}'(\omega)$. We need to show that $\lim_n \nu_x^{KG,n} = 0$. From [13], we know that $f(z) = \phi(z) + z\Phi(z)$ is a nondecreasing function. Moreover, by the KGCB procedure, $b_j^n(x)$ is sorted in increasing order. Then,

$$\nu_x^{KG,n} \leq \sum_{j=1,\ldots,|\mathcal{X}|-1} (b_{j+1}^n(x) - b_j^n(x)) f(0). \tag{23}$$

Then by lemma 3, we know that $\lim_n b^n_{x'}(x) = 0$, $\forall x' \in \mathcal{X}$, and for $j = 1, \ldots, |\mathcal{X}|$, we have $\lim_n b^n_j(x) = 0$. By letting $n \to \infty$ in equation (23), we obtain $\lim_n \nu^{KG,n}_x = 0$. This is equivalent to saying that the knowledge gradient value for infinitely sampled alternatives converges to zero.

We now consider some $x \notin \mathcal{X}'(\omega)$, which is an alternative that is not measured infinitely often. We want to show that $\lim_n \nu^{KG,n}_x > 0$. We define the set $\mathcal{J} := \{j | \liminf_n b^n_j(x) > 0\}$. By lemma 3, it follows that $\liminf_n b^n_x(x) > 0$. Now we know that $\mathcal{J}$ is nonempty. Since at least one alternative is measured infinitely often, $\mathcal{X}'(\omega)$ is nonempty. By lemma 3, this means that there exists at least one alternative x'' such that $\liminf_n b^n_{x''}(x) = 0$. In other words, we also have $\mathcal{J}^C$ being nonempty. Then, there exists an $L' < \infty$ such that $\min_{j\in\mathcal{J}} b^n_j(x) > \max_{j\notin\mathcal{J}} b^n_j(x)$ for all $n > L'$. Recalling that $f(z)$ is non-decreasing and positive, we have

$$\nu^{KG,n}_x \geq \min_{j\in\mathcal{J}, j'\in\mathcal{J}^C} (b_j(x) - b_{j'}(x)) f\left(-\left|\frac{a^n_j - a^n_{j+1}}{b^n_{j+1}(x) - b^n_j(x)}\right|\right). \tag{24}$$

Now using lemma 2, we define the upper bound $U := \sup_{n,j,x} |a^n_j(x)|$, with $U < \infty$. It follows that $\sup_{n,j,x} |a^n_j(x) - a^n_{j+1}(x)| \leq 2U$. So for all $n \geq L'$, combining with the monotonicity of $f(z)$, we can derive

$$\nu^{KG,n}_x \geq \min_{j\in\mathcal{J}, j'\in\mathcal{J}^C} (b_j(x) - b_{j'}(x)) f\left(-\frac{2U}{b^n_{j+1}(x) - b^n_j(x)}\right). \tag{25}$$

We define $\bar{b} := \min_{j\in\mathcal{J}} b_j(x) > 0$. By the continuity of $f(z)$, we have

$$\lim_n \nu^{KG,n}_x \geq \bar{b} f\left(-\frac{2U}{\bar{b}}\right) > 0. \tag{26}$$

Now, for $x' \notin \mathcal{X}'$, we have $\lim_n \nu^{KG,n}_{x'} > 0$. On the other hand, for $x \in \mathcal{X}'$, we have $\lim_n \nu^{KG,n}_x = 0$. For $x' \in \mathcal{X}'$, there exists some $n > L$ such that $\lim_n \nu^{KG,n}_{x'} > \lim_n \nu^{KG,n}_x$. In other words, we will start measuring an alternative outside of $\mathcal{X}'$ some time after L. However, this contradicts with our assumption that $\mathcal{X}'(\omega) \subset \mathcal{X}$ and that L is the last time we measured an alternative outside of $\mathcal{X}'(\omega)$. Therefore $\mathcal{X}'(\omega) = \mathcal{X}$ for all $\omega \in \Omega_0$.

Corollary 1 *Under the HKG-RBF policy,* $\lim_{n\to\infty} \theta^n_x = \mu_x$ *almost surely for every alternative* $x \in \mathcal{X}$*. In other words, HKG-RBF will eventually find the true optimal alternative.*

Proof By Theorem 1, every alternative x is measured infinitely often. By the strong law of large numbers, we have

$$\lim_{n\to\infty} f(x|\Theta^{0,n}) = \mu_x \quad \text{almost surely.}$$

Since we measure all of the alternatives infinitely often, we also have

$$\lim_{n\to\infty} (\sigma^{g,n}_x)^2 \to 0,$$

for all $g \in \mathcal{G}$ and for all $x \in \mathcal{X}$. Now we consider the weights $w^{g,n}_x$. Fixing $x \in \mathcal{X}$ and $\omega \in \Omega$, and by assumption 2, all hierarchical levels $g \neq 0$ have bias,

i.e. $\lim_n v_x^{g,n}(\omega) \neq 0$. This means that $\lim_n w_x^{g,n} \to 0$ for all $g \neq 0$, implying $\lim_n w_x^{0,n} = 1$. In other words, we have

$$\lim_n \sum_{g \in \mathcal{G}} w_x^{g,n} f(x|\Theta^{g,n}) = \lim_n w_x^{0,n} f(x|\Theta^{0,n}) = \mu_x.$$

This result shows that HKG-RBF converges asymptotically to the true optimal alternative. However, it is infeasible to measure the alternatives infinitely often in practice. In the following section, we will demonstrate the convergence of our method from empirical results.

7 Numerical Experiments

We ran our algorithm on continuous functions on $\mathbb{R}^d$ to find the global maximum of the function. We chose from test functions for similar procedures as well as common functions from applications. Since our algorithm is developed for problems with finite alternatives, we need to discretize the set of alternatives. Moreover, we assume a non-informative Bayesian prior for all the experiments. We set the prior mean to zero and prior variance to some arbitrarily large number. At each time step, we observe a noisy realization of the true function. In section 7.1, we compare our algorithm on one-dimensional Gaussian processes to two other offline learning methods. In section 7.2, we experiment on multi-dimensional test functions. In section 7.3, we present a generalized version of the asymmetric unimodal function in $\mathbb{R}^d$.

7.1 Performance on One-Dimensional Test Functions

In this section, we compare KG-RBF to two other methods. The first is pure exploration (Expl), where each alternative is selected with probability $1/M$ at every time step. The second is the knowledge gradient with non-parametric beliefs (KGNP) presented in [3], which uses aggregates over a set of kernel functions to estimate the truth. We test these policies on two types of Gaussian processes. All GP functions are defined on the discretized points $x = 1, 2, \ldots, 100$. For all the policies, we select the prior mean to be $\theta^0 = \mathbf{0}$, and prior covariance to be $\Sigma^0 = 10^4 I$. For KGNP, we use a dictionary of Epanechnikov kernels with width $h = \{5, 10, 20, 40, 80\}$. For KG-RBF, we select the threshold distance $D_T = 5$ and $D_T = 2$. We also test the hierarchical version, HKG-RBF, with three levels $D_T = \{100, 5, 1\}$. This setting corresponds to the combination of a global parametric model, a local parametric model, and a equivalent look-up table model.

7.1.1 Gaussian Process with Homoscedastic Covariance Functions

First we test our policies on Gaussian processes with the covariance function

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

which produces a stationary process with variance σ^2 and length scale ρ. Higher values of ρ produce fewer numbers of peaks in the domain, resulting in a smoother

function. The variance σ^2 scales the function vertically. Figure 4 illustrates randomly generated Gaussian processes with different values of ρ.

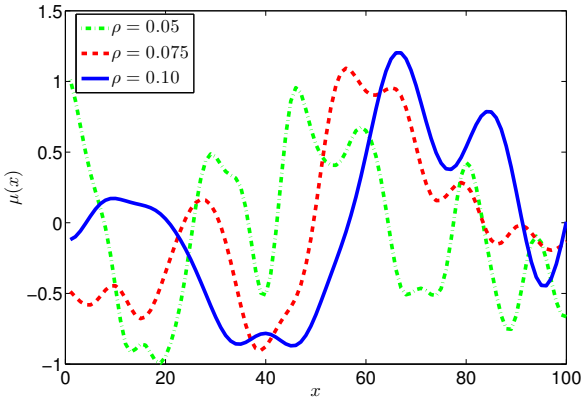


Fig. 4: Stationary Gaussian processes with different ρ

For our test functions, we fix $\sigma^2 = 0.5$ and vary $\rho \in \{0.05, 0.1\}$. For both of these values, we generate 500 test functions to test our policy. We use the average opportunity cost as the performance benchmark, where

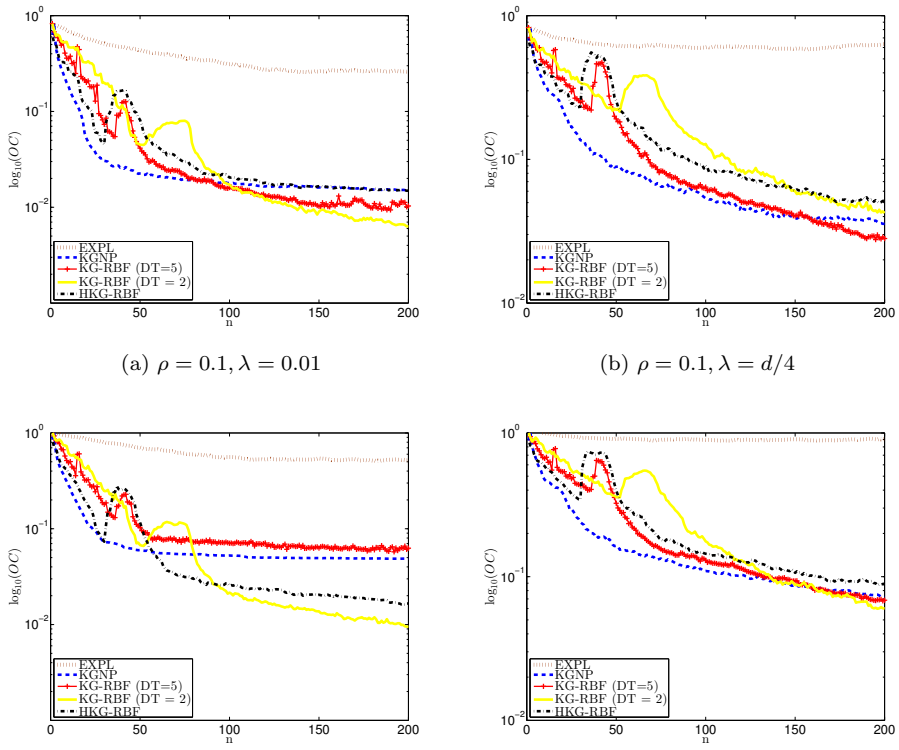
$$OC(n) = \max_i(\mu_i) - \mu_{i^\star},$$

with $i^\star = \arg\max_x \theta_x^n$. In other words, it is the difference between the maximum of the truth and the value of the best alternative found by the algorithm. We also test the policies on two different sampling noise levels: the lower one with $\lambda = 0.01$ and the high noise with $\lambda = \frac{1}{4}(\max(\mu) - \min(\mu))$. The opportunity costs on a log scale for different policies are given in Figure 5.

The single level KG-RBF with $D_T = 5$ performs comparably to the KGNP asymptotically in all of the examples; performance improves when we reduce $D_T = 2$. However, smaller D_T has a slower initial convergence rate. This is expected since smaller D_T requires more exploration in the beginning. HKG-RBF balances both exploitation and exploration; it has the same asymptotic performance as the smaller D_T scenario without compromising the initial convergence rate. We also want to point out that the opportunity cost for all KG-RBF methods does not decrease monotonically; there is a "bump" after approximately 50 iterations. This behavior is possibly due to poor estimates resulting from interpolation between different clouds.

7.1.2 Gaussian Process with Heteroscedastic Covariance Functions

We now consider non-stationary covariance functions, particularly the Gibbs covariance function [16]. It has a similar structure to the exponential covariance



(a) $\rho = 0.1, \lambda = 0.01$ (b) $\rho = 0.1, \lambda = d/4$

(c) $\rho = 0.05, \lambda = 0.01$ (d) $\rho = 0.05, \lambda = d/4$

Fig. 5: Comparison of policies on homoscedastic GP, where $d = \max(\mu) - \min(\mu)$.

function but is heteroscedastic. The Gibbs covariance function is given by

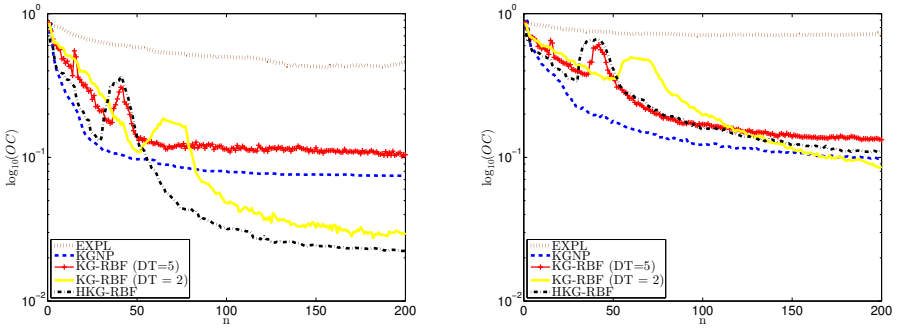
$$Cov(i,j) = \sigma^2 \sqrt{\frac{2l(i)l(j)}{l(i)^2 + l(j)^2}} \exp\left(-\frac{(i-j)^2}{l(i)^2 + l(j)^2}\right),$$

where $l(i)$ is an arbitrary positive function in i. Here we chose a horizontally shifted periodic sine function

$$l(i) = 10\left(1 + \sin\left(2\pi\left(\frac{i}{100} + u\right)\right)\right) + 1,$$

where u is a uniform random number from $[0, 1]$. We compute the opportunity cost as previously described. The opportunity costs on a log scale for different policies are given in Figure 6.

In both examples, KG-RBF with $D_T = 5$ underperforms when comparing to KGNP; it out performs when we decrease $D_T = 2$, especially in the low noise scenario. Again, we notice similar behaviors for HKG-RBF, where it balances the trade-off between exploration and exploitation.



(a) $\lambda = 0.01$ (b) $\lambda = d/4$

Fig. 6: Comparison of policies on heteroscedastic GP, where $d = \max(\mu) - \min(\mu)$.

7.2 Performance on Two Dimensional Test Functions

Next we test our methods with two test functions: the Six-hump camel back [4] and Tilted Branin [22].

The Six-hump camel back (SHCB) function is given by

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

over the domain $x \in [-1.6, 2.4] \times [-0.8, 1.2]$. We discretize the domain and use a 33×33 on $\mathbb{R}^2$, which gives a total of 1089 different alternatives. For the SHCB function, we run on two different noise variance $\lambda = 0.12^2, 2.5^2$.

The Tilted Branin (TB) function is given by

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

over the domain $x \in [-5, 10] \times [0, 15]$. We discretize the domain and use a 30×30 on $\mathbb{R}^2$, which gives a total of 961 different alternatives. For the SHCB function, we run on two different noise variances $\lambda = 2^2, 5^2$. For each measurement noise setting, we test both KG-RBF and KGNP policies for 50 iterations for a total of 100 times. We provide the opportunity cost and the standard error in Table 2.

Table 2: Expected opportunity cost after 50 iterations for two dimensional test functions

		HKG-RBF		KGNP	
Test functions	λ	$\mathbb{E}[OC(50)]$	SE	$\mathbb{E}[OC(50)]$	SE
Six-hump camel back	0.12^2	0.0411	0.0170	0.0310	0.0012
	2.5^2	1.6716	2.2166	1.1596	0.2284
Tilted Branin	2^2	0.0115	0.2308	0.8414	0.2661
	5^5	1.0094	0.6665	0.8469	0.1857

In the low noise settings, the HKG-RBF is comparable to KGNP; it even outperforms for the tilted branin problem. In the high noise setting, the HKG-RBF slightly underperforms when comparing to KGNP for both functions. This behavior is expected since these two functions are very "bumpy." Traditional non-parametric methods benefit from evaluating the weighted average estimate at every single point. It is difficult for the HKG-RBF to build accurate local models with so few data points, since each "bump" requires at least three RBF's (or hyperplanes) for estimation.

7.3 Application Example

We return to our asymmetric unimodal functions; this type of function appears in many applications such as tuning black-box simulators or optimizing physical experiments where a tunable parameter has an optimal setting interval, where results get worse as you move away from the right setting. It is common in such settings that the cost of being a little too high (or too low) is much higher than the cost of being a little too low (or too high).

In the one-dimensional experiments, we fix $\eta_1 = 1$ and vary $\eta_2 \in \{0.1, 0.3, 0.7, 0.9\}$, which determines the slope shapes. For each η combination, we generate D from a normal distribution with fixed mean 35 and vary the variance $\sigma^2 \in \{0.5, 1, 2, 4\}$. We also have additive noise $\epsilon \sim \mathcal{N}(0, 10)$. The domain of the function is $[1, 60]$, discretized by 0.2; in other words, there are 296 alternatives. We want to compare the hierarchical KG-RBF to KG with parametric models. For the HKG-RBF we set $D_T = 6$. We consider two parametric models: quadratic and 4-th order polynomial. We compare the opportunity cost after 50 iterations for all methods and compare them in Table 3.

Table 3: Expected opportunity cost after 50 iterations for asymmetric unimodal functions

.

		HKG-RBF		KGLinP (quadratic)		KGLinP (4th order)	
η_2	σ^2	$\mathbb{E}[OC(50)]$	SE	$\mathbb{E}[OC(50)]$	SE	$\mathbb{E}[OC(50)]$	SE
0.1	0.5	**0.6765**	0.5187	1.4449	0.1345	0.8325	0.2262
	1	**0.6984**	0.6546	1.3891	0.2181	0.8299	0.4833
	2	0.6882	0.6948	1.1812	0.1311	**0.5451**	0.1840
	4	0.6010	0.5596	0.8673	0.2361	**0.4460**	0.4892
0.3	0.5	**0.7721**	0.7561	2.0998	0.1860	1.0516	0.3406
	1	**0.6138**	0.7053	1.9214	0.4031	0.9615	0.6504
	2	**0.6621**	0.7866	1.5709	0.2030	0.5895	0.3293
	4	0.5974	0.8239	0.9544	0.3298	**0.3614**	0.4334
0.7	0.5	**0.9611**	0.7132	2.6741	0.1389	1.3643	0.2512
	1	**0.6780**	0.6843	2.4656	0.2155	1.2047	0.3898
	2	**0.6261**	0.6734	2.1213	0.1332	0.8321	0.2875
	4	**0.4050**	0.4258	1.4185	0.2494	0.4245	0.3542
0.9	0.5	**0.6216**	0.5629	1.4288	0.0533	0.8732	0.1065
	1	**0.6480**	0.5445	1.3523	0.0840	0.8565	0.2607
	2	**0.5965**	0.5494	1.1607	0.0578	0.6359	0.1222
	4	0.5893	0.6533	0.9091	0.0893	**0.4399**	0.3066

We see that HKG-RBF outperforms the KG with parametric models when σ is small, where the true function has a "sharp" drop at the optimal region. However, KG with parametric models begins to outperform HKG-RBF when $\sigma^2 = 4$. This behavior is expected since polynomial parametric models benefit when there is more curvature at the optimal region. However, we see that HKG-RBF is still comparable in performance in most cases except when $\sigma^2 = 4$. When σ^2 increases, the function will have a smoother curve but with higher noise variance. Consequently, estimates of the local linear models become unreliable, but global parametric model can better capture the optimal point.

Next, we generalize our asymmetric unimodal function to a multi-dimensional scenario, where the true function is

$$f(x_1, \ldots, x_k) = \sum_{i=1}^{k} \eta_{1,i} \mathbb{E}\left[\min\left(x_i, \left(D - \sum_{j=1}^{i-1} x_j\right)^+\right)\right] - \sum_{i=1}^{k} \eta_{2,i} x_i.$$

Similarly, we restrict that $\eta_{1,i} \leq \eta_{1,i+1}$ and $\eta_{1,i} > \eta_{2.i}$, for all i. This function has a simple interpretation: it is equivalent to a multi-item newsvendor problem, where $\eta_{1,i}$ is the price item i, and $\eta_{2,i}$ is its cost. We want to find the optimal combination of supplies $(x_1^*, \ldots, x_k^*)$ to satisfy the uncertain demand D. Once again, we test our method against KG with a quadratic model. Even in the simple quadratic case, the number of parameters increases proportionally to the dimension squared. In the two-dimensional case, we pick $\eta_{1,i} = \langle 2, 1.8\rangle$ and $\eta_{2,i} = \langle 1.8, 1.5\rangle$. We select our domain to be $[0, 15] \times [0, 15]$, discretized to 31×31 grid. In the three-dimensional example, we select $\eta_{1,i} = \langle 2, 1.8, 1.7\rangle$ and $\eta_{2,i} = \langle 1.9, 1.6, 1.6\rangle$. Similarly, the domain is $[0, 15] \times [0, 15] \times [0, 15]$, discretized to $31 \times 31 \times 31$ grid. For both experiments, we sample D from normal distributions with mean 20 and standard deviations of $\sigma = \{1, 4\}$, which represent different surface smoothness. We also test on two different measurement noise settings, $\lambda = \{1, 9\}$. Table 4 shows the opportunity costs of the 2D experiments after 50 iterations and Table 5 shows the opportunity costs of the 3D experiments after 100 iterations.

Table 4: Expected opportunity cost after 50 iterations for 2D asymmetric unimodal functions, where $\eta_{1,i} = \langle 2, 1.8\rangle$ and $\eta_{2,i} = \langle 1.8, 1.5\rangle$, and $D \sim \mathcal{N}(20, \lambda)$.

.

		HKG-RBF		KGLinP	
σ	λ	$\mathbb{E}[OC(50)]$	SE	$\mathbb{E}[OC(50)]$	SE
1	1	1.3814	0.7852	2.1461	0.1474
	9	1.5320	1.0476	2.3571	0.6400
4	1	0.7868	0.3752	1.4371	0.1860
	9	0.9267	0.6104	1.5583	0.6119

We see that HKG-RBF outperforms the parametric model in all settings, especially when σ is small. Small σ represents a "sharper" hump at the optimal region, for which KG with a parametric model cannot estimate correctly. As σ increases (i.e., smoother surface), we see that the performance of the parametric model is almost comparable to that of HKG-RBF, especially in the 3D examples. In the 2D experiments, increasing the measurement noise deteriorates the performance

Table 5: Expected opportunity cost after 100 iterations for 3D asymmetric unimodal functions, where $\eta_{1,i} = \langle 2, 1.8, 1.7\rangle$ and $\eta_{2,i} = \langle 1.9, 1.6, 1.6\rangle$, and $D \sim \mathcal{N}(20, \lambda)$.

.

		HKG-RBF		KGLinP	
σ	λ	$\mathbb{E}[OC(100)]$	SE	$\mathbb{E}[OC(100)]$	SE
1	1	1.6209	1.6840	2.6219	0.4232
	9	1.5931	1.1334	2.6359	0.4409
4	1	1.8853	1.8964	2.1362	0.3394
	9	1.5758	1.3175	2.1063	0.3662

of HKG-RBF. This behavior is expected since the local linear models are sensitive to noise. However, the high noise does not affect the performance of HKG-RBF in the 3D cases.

8 Conclusion

In this paper, we introduced an optimal learning policy to optimize an unknown function. Our strategy is an adaptation of the knowledge gradient algorithm [13] for correlated alternatives. Instead of a known covariance matrix, we use a statistical method called Dirichlet cloud radial basis function [24] to define local regions and approximate the local covariance structures. DC-RBF then uses a weighted sum of the local models to estimate the global function. In addition, we propose a hierarchical approach that combines multiple levels of DC-RBF with different threshold distances D_T's.

We proved the algorithm is asymptotically optimal, since it measures every alternative infinitely often. Finally, we showed the performance of the HKG-RBF on two types of one-dimensional Gaussian processes and two commonly used two dimensional test functions. HKG-RBF generally performs better than the competing methods in these experiments. We also demonstrated the advantage of HKG-RBF over KG with parametric models on a multi-dimensional application with asymmetric unimodal functions.

While the HKG-RBF policy performs well in simulations, there are several shortcomings. Although the hierarchical structure improves the estimation and removes the need to find the optimal D_T, it adds computational burden to the problem. In practice, we have to limit our models to a maximum of three levels. Second, the number of data points required to initialize a new cloud increases linearly with the number of dimension of the data. This becomes efficient when we are dealing with high-dimensional data and small number of measurements. We would like to develop a new scheme in initializing clouds that are independent of the number of dimensions.

References

1. Agrawal, R.: The continuum-armed bandit problem. SIAM Journal on Control and Optimization **33**(6), 1926–1951 (1995)

2. Barton, R.R., Meckesheimer, M.: Chapter 18 metamodel-based simulation optimization. In: S.G. Henderson, B.L. Nelson (eds.) Simulation, *Handbooks in Operations Research and Management Science*, vol. 13, pp. 535 – 574. Elsevier (2006)
3. Barut, E., Powell, W.B.: Optimal learning for sequential sampling with non-parametric beliefs. Journal of Global Optimization pp. 1–27 (2013)
4. Branin, F.H.: Widely convergent method for finding multiple solutions of simultaneous nonlinear equations. IBM J. Res. Dev. **16**(5), 504–522 (1972)
5. Calvin, J., Žilinskas, A.: One-dimensional global optimization for observations with noise. Computers and Mathematics with Applications **50**(1-2), 157–169 (2005). DOI 10.1016/j.camwa.2004.12.014. URL http://www.scopus.com/inward/record.url?eid=2-s2.0-27544449903partnerID=40md5=6f2a79009b74d74bef2523430b25c56e. Cited By 8
6. Chang, F., Lai, T.L.: Optimal stopping and dynamic allocation. Advances in Applied Probability **19**(4), pp. 829–853 (1987)
7. Chick, S.E., Gans, N.: Economic analysis of simulation selection problems. Management Science **55**(3), 421–437 (2009)
8. Cochran, W., Cox, G.: Experimental Deisgns. Wiley, New York (1957)
9. Cohn, D.A., Ghahramani, Z., Jordan, M.I.: Active learning with statistical models. Journal of Artificial Intelligence Research **4**, 129–145 (1996)
10. Eubank, R.L.: Spline Smoothing and Nonparametric Regression, 2 edn. Marcel Dekker (1988)
11. Fan, J., Gijbels, I.: Local Polynomial Modelling and Its Applications. Monographs on statistics and applied probability series. Chapman and Hall/CRC (1996)
12. Frazier, P.I., Powell, W.B., Dayanik, S.: A knowledge-gradient policy for sequential information collection. SIAM J. Control Optim. **47**(5), 2410–2439 (2008)
13. Frazier, P.I., Powell, W.B., Dayanik, S.: The Knowledge-Gradient Policy for Correlated Normal Beliefs. INFORMS Journal on Computing **21**(4), 599–613 (2009)
14. Fu, M.C.: Chapter 19 gradient estimation. In: S.G. Henderson, B.L. Nelson (eds.) Simulation, *Handbooks in Operations Research and Management Science*, vol. 13, pp. 575 – 616. Elsevier (2006)
15. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian Data Analysis, 2 edn. Chapman and Hall/CRC (2004)
16. Gibbs, M.N.: Bayesian gaussian processes for regression and classification. Ph.D. thesis, University of Cambridge (1997)
17. Ginebra, J., Clayton, M.K.: Response surface bandits. Journal of the Royal Statistical Society. Series B (Methodological) **57**(4), pp. 771–784 (1995)
18. Gittins, J.C.: Bandit processes and dynamic allocation indices. Journal of the Royal Statistical Society. Series B (Methodological) **41**(2), pp. 148–177 (1979)
19. Gupta, S.S., Miescke, K.J.: Bayesian look ahead one-stage sampling allocations for selection of the best population. Journal of Statistical Planning and Inference **54**(2), 229–244 (1996)
20. Hannah, L.A., Blei, D.M., Powell, W.B.: Dirichlet process mixtures of generalized linear models. J. Mach. Learn. Res. **12**, 1923–1953 (2011)
21. Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice Hall PTR, Upper Saddle River, NJ, USA (1998)
22. Huang, D., Allen, T., Notz, W., Zeng, N.: Global optimization of stochastic black-box systems via sequential kriging meta-models. Journal of Global Optimization **34**(3), 441–466 (2006)
23. Jamshidi, A.A., Kirby, M.J.: Towards a black box algorithm for nonlinear function approximation over high-dimensional domains. SIAM Journal on Scientific Computation **29**(3), 941–964 (2007)
24. Jamshidi, A.A., Powell, W.B.: A recursive local polynomial approximation method using dirichlet clouds and radial basis functions. SIAM Journal on Scientific Computation:submitted (2013)
25. Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. Journal of Global Optimization **13**(4), 455–492 (1998). DOI 10.1023/A:1008306431147
26. Jones, R., Lee, Y.C., Barnes, C.W., Flake, G., Lee, K., Lewis, P., Qian, S.: Function approximation and time series prediction with neural networks. In: Neural Networks, 1990., 1990 IJCNN International Joint Conference on, pp. 649–665 vol.1 (1990)
27. Knuth, D.E.: The Art of Computer Programming, vol. 2: Seminumerical Algorithms, 3 edn. Addison-Wesley, Boston (1998)

28. Kushner, H.J.: A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. Journal of Fluids Engineering **86**(1), 97–106 (1964). URL http://dx.doi.org/10.1115/1.3653121
29. Mes, M.R., Powell, W.B., Frazier, P.I.: Hierarchical knowledge gradient for sequential sampling. J. Mach. Learn. Res. **12**, 2931–2974 (2011)
30. Montgomery, D.C., Peck, E.A., Vining, G.G.: Introduction to Linear Regression Analysis (4th ed.). Wiley & Sons, Hoboken (2006)
31. Muller, H.G.: Weighted local regression and kernel methods for nonparametric curve fitting. Journal of the American Statistical Association **82**(397), pp. 231–238 (1987)
32. Negoescu, D.M., Frazier, P.I., Powell, W.B.: The Knowledge-Gradient Algorithm for Sequencing Experiments in Drug Discovery. INFORMS Journal on Computing **23**(3), 346–363 (2010)
33. Nelson, B.L., Swann, J., Goldsman, D., Song, W.: Simple procedures for selecting the best simulated system when the number of alternatives is large. Operations Research **49**, 950–963 (1999)
34. Powell, W.B., Ryzhov, I.O.: Optimal Learning. Wiley (2012)
35. Robbins, H., Monro, S.: A stochastic approximation method. The Annals of Mathematical Statistics **22**(3), pp. 400–407 (1951)
36. Ryzhov, I.O., Powell, W.B., Frazier, P.: The knowledge gradient algorithm for a general class of online learning problems. Operations Research **60**(1), 180–195 (2012)
37. Scott, W., Frazier, P., Powell, W.B.: The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression. SIAM Journal on Optimization **21**(3), 996–1026 (2011)
38. Spall, J.C.: Introduction to Stochastic Search and Optimization, 1 edn. John Wiley & Sons, Inc., New York, NY, USA (2003)
39. Villemonteix, J., Vazquez, E., Walter, E.: An informational approach to the global optimization of expensive-to-evaluate functions. Journal of Global Optimization **44**(4), 509–534 (2009)
40. Žilinskas, A.: Two algorithms for one-dimensional multimodai minimization. Mathematische Operationsforschung und Statistik. Series Optimization **12**(1), 53–63 (1981). DOI 10.1080/02331938108842707. URL http://dx.doi.org/10.1080/02331938108842707

A Proofs

Assumption 1 *For any local parametric model θ, let $\theta(i)$ denote the i-th element of the θ. For any $i \neq j$, we have $\limsup_{n\to\infty} |Corr^n[\theta(i), \theta(j)]| \leq 1$ almost surely. Essentially, we claim the correlation between parameters to be bounded by 1.*

Assumption 2 *For any level $g \in \mathcal{G} \setminus \{0\}$, we have $\liminf_n |v_x^{g,n}| > 0$. In other words, estimates for all the levels, except when $g = 0$, will have bias.*

Lemma 1 *If we have a prior on each parameter, then for any $x, x' \in \mathcal{X}$, we have $\sup_n |x^T \theta_j^{g,n}| < \infty$ and $\sup_n |x^T \Sigma_j^{\theta,g,n} x'| < \infty$ almost surely.*

Proof We can show that for any cloud j, $(\theta_j^n, \Sigma_j^{\theta,n})$ is a uniformly integrable martingale; therefore, it converges to some integrable random variable $(\theta_j^\infty, \Sigma_j^{\theta,\infty})$ almost surely.

Fix any g and j. we see that

$$\Sigma^{\theta,n+1} - \Sigma^{\theta,n} = -\frac{\Sigma^{\theta,n} x^n (x^n)^T \Sigma^{\theta,n}}{\lambda + \Sigma_{x^n,x^n}},$$

$$\mathrm{diag}(\Sigma^{\theta,n+1} - \Sigma^{\theta,n}) = -\frac{1}{\lambda + \Sigma_{x^n,x^n}} \begin{bmatrix} \zeta_{x^n}^2(1) & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \zeta_{x^n}^2(d) \end{bmatrix},$$

where $\zeta_x^n(i)$ is the i-th element of the product $\Sigma^{\theta,n} x^n$. By definition, we have $\lambda + \Sigma_{x^n,x^n} > 0$. We also note that the $\mathrm{diag}(\Sigma^{\theta,n})$ is in fact the variance of each parameter, namely $\mathrm{Var}^n[\theta]$.

It is apparent that $\mathrm{Var}^n[\theta(i)]$ is a non-increasing sequence and bounded by zero below. For non-diagonal elements of $\Sigma^{\theta,n}(i,j)$, where $i \neq j$, we define it as the covariance of element i and j. By the definition of correlation, we have

$$\begin{aligned}\Sigma^{\theta,n}(i,j) &= \mathrm{Corr}^n[\theta(i),\theta(j)]\sqrt{\mathrm{Var}^n[\theta(i)]\mathrm{Var}^n[\theta(j)]},\\ |\Sigma^{\theta,n}(i,j)| &\leq |\sqrt{\mathrm{Var}^n[\theta(i)]\mathrm{Var}^n[\theta(j)]}| < \infty.\end{aligned}$$

It follows that $\sup_n |x^T \Sigma_j^{\theta,g,n} x'| < \infty$, since all the elements of $\Sigma_j^{\theta,g,n}$ are finite, as well as all the alternatives $x, x' \in \mathcal{X}$.

Lemma 2 *If we have a prior on each alternative, then for any $x, x' \in \mathcal{X}$, the following are finite almost surely:* $\sup_n |\theta_x^{g,n}|$, $\sup_n |a_{x'}^{g,n}(x)|$, *and* $\sup_n |b_{x'}^{g,n}(x)|$.

Proof We see that $\sup_n |\theta_x^{g,n}|$ is finite follows from the fact that $\theta_x^{g,n}$ is a convex combination of $x^T\theta_j^{g,n}$. By 1, the convex combination of finite variables is also finite.

To show that $\sup_n |a_{x'}^{g,n}(x)|$ is finite, we need bound both parts of the sum as the following,

$$|a_{x'}^n(x)| \leq \left|\frac{\sum_{j=1}^{N_c}\varphi_j^{n+1}(x)x^T\theta_j^n}{K^{n+1}(x)}\right| + \left|\frac{\varphi_I^{n+1}(x)(f(x^n|\Theta^n) - x^T\theta_I^n)x^T\Sigma_I^{\theta,n}x'}{K^{n+1}(x)\gamma_I^n}\right|.$$

For the first part, we a have convex combination of all the local parametric model, therefore

$$\left|\frac{\sum_{j=1}^{N_c}\varphi_j^{n+1}(x)x^T\theta_j^n}{K^{n+1}(x)}\right| \leq \frac{\sum_{j=1}^{N_c}\varphi_j^{n+1}(x)\left|x^T\theta_j^n\right|}{K^{n+1}(x)} < \infty.$$

By Lemma 1, the convex combination of finite values is also finite. For the second part, we see that

$$\begin{aligned}\left|\frac{\varphi_I^{n+1}(x)(f(x^n|\Theta^n) - x^T\theta_I^n)x^T\Sigma_I^{\theta,n}x'}{K^{n+1}(x)\gamma_I^n}\right| &= \left|\frac{\varphi_I^{n+1}(x)}{K^{n+1}(x)}\right|\left|\frac{(f(x^n|\Theta^n) - x^T\theta_I^n)x^T\Sigma_I^{\theta,n}x'}{\gamma_I^n}\right|,\\ &\leq \left|\frac{(f(x^n|\Theta^n) - x^T\theta_I^n)x^T\Sigma_I^{\theta,n}x'}{\gamma_I^n}\right|.\end{aligned}$$

Using Lemma 1 again, this expression is also finite since γ_I^n is bounded by $\lambda_x > 0$ below. Now we will show that $\sup_n |b_{x'}^{g,n}(x)|$ is finite. We temporarily remove the index g for clarity, then for any cloud I,

$$\begin{aligned}|b_{x'}^n(x)| &= \left|\frac{\varphi_I^{n+1}(\sqrt{\lambda + \Sigma_{x'x'}^n})x^T\Sigma_I^{\theta,n}x'}{K^{n+1}\gamma_I^n}\right|\\ &\leq \left|\frac{\varphi_I^{n+1}}{K^{n+1}}\right|\left|\frac{(\sqrt{\lambda + \Sigma_{x'x'}^n})x^T\Sigma_I^{\theta,n}x'}{\gamma_I^n}\right|\\ &\leq \left|\frac{(\sqrt{\lambda + \Sigma_{x'x'}^n})x^T\Sigma_I^{\theta,n}x'}{\gamma_I^n}\right|.\end{aligned}$$

We know that $\sup_n |\Sigma_{x'x'}^n| < \infty$. It follows from Lemma 1 that $\sup_n |b_{x'}^{g,n}(x)|$ is finite almost surely.

Lemma 3 *For any $\omega \in \Omega$, let $\mathcal{X}'(\omega)$ be the random set of alternatives measured infinitely often by the KG-RBF policy. For all $x, x' \in \mathcal{X}$, 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 We first consider the case where $x \in \mathcal{X}'$. We know that $\lim_n (\sigma_x^{g,n})^2 \to 0$ for all $g \in \mathcal{G}$. We focus on the asymptotic behavior of weights $w_x^{g,n}$ first. By definition, the bias of the base level is $\lim_n v_x^{0,n} = 0$. For $g \neq 0$, we have

$$w_g^{g,n} \leq \frac{((\sigma_x^{g,n})^2 + (v_x^{g,n})^2)^{-1}}{(\sigma^{0,n})^{-2} + ((\sigma_x^{g,n})^2 + (v_x^{g,n})^2)^{-1}}.$$

By assumption 2, all hierarchical levels $g \neq 0$ have bias, i.e. $\lim_n v_x^{g,n} \neq 0$. This means that $\lim_n w_x^{g,n} \to 0$ for all $g \neq 0$, implying $\lim_n w_x^{0,n} = 1$. In other words, we have

$$\lim_n b_{x'}^n(x) = \lim_n \sum_{g \in \mathcal{G}} w_{x'}^{g,n} b_{x'}^{g,n}(x) = \lim_n b_{x'}^{0,n}(x).$$

The DC-RBF model for the base level is equivalent to a look up table model with independent alternatives, since we let D_T to be infinitesimally small (or equivalently we let $N_c = M$ for the base level). Note that independence of the alternatives on the base level comes from the fact that all of the RBF's are independent of each other by definition. Following equation (4), we have

$$b_{x'}^{0,n}(x) = \frac{e_x \Sigma^{0,n} e_{x'}}{\sqrt{\lambda_{x'} + \Sigma_{x'x'}^n}},$$

where $\Sigma^{0,n}$ is the covariance matrix of the alternatives and is diagonal. It is apparent that $b_{x'}^{0,n}(x) = 0$ when $x' \neq x$, since the alternatives are independent. For $x' = x$, we know that the numerator $e_x \Sigma^n e_x = (\sigma_x^n)^2$, which is the posterior variance of alternative x. Since $x \in \mathcal{X}'$ is measured infinitely many times, we have $(\sigma_x^n)^2 \to 0$. Therefore, $\lim_n b_{x'}^n(x) = 0$ under this case as well.

If $x \notin \mathcal{X}'$, it is equivalent to show that for any cloud I and level $g \in \mathcal{G}$, we have

$$\liminf_n \tilde{\sigma}(x, x, I) = \frac{\varphi_I^{n+1}(\sqrt{\lambda + \Sigma_{xx}^n}) x^T \Sigma_I^{\theta,n} x}{K^{n+1} \gamma_I^n} > 0.$$

This becomes evident when we rearrange the expression as

$$\begin{aligned}
\liminf_n \tilde{\sigma}(x, x, I) &= \left(\frac{\varphi_I^{n+1}}{K^{n+1}}\right) \left(\frac{\sqrt{\lambda + \Sigma_{xx}^n}}{\gamma_I^n}\right) \left(x^T \Sigma_I^{\theta,n} x\right) \\
&> \left(\frac{\varphi_I^{n+1}}{K^{n+1}}\right) \left(\frac{\sqrt{\lambda}}{\gamma_I^n}\right) \left(x^T \Sigma_I^{\theta,n} x\right) \\
&> 0.
\end{aligned}$$

The first part of the product is greater than zero by our use of the Gaussian kernel, where $\varphi(x) > 0$, for all $x \in \mathcal{X}$. By definition $\Sigma_{xx}^n = \sigma_x^2 > 0$ for a finitely sampled alternative; therefore, the second part of the product is also lower bounded by 0. The last part of the sum is also lower bounded by zero since $\Sigma_I^{\theta,n}$ is a semi-definite matrix by definition. We have $b_x^n(x)$ as a convex combination of the above expression in equation (22); hence $\liminf_n b_x^n(x) > 0$ for finitely sampled alternatives $x \notin \mathcal{X}'$.