Received: 11 November 2020 | Accepted: 5 September 2021 | Published on: 7 October 2021

DOI: 10.1002/nav.22028

RESEARCH ARTICLE

WILEY

Technical note—Knowledge gradient for selection with covariates: Consistency and computation

Liang Ding[1] | **L. Jeff Hong**[2] | **Haihui Shen**[3] | **Xiaowei Zhang**[4]

[1]Department of Industrial Systems Engineering, Texas A&M University, College Station, Texas, USA
[2]School of Management and School of Data Science, Fudan University, Shanghai, China
[3]Sino-US Global Logistics Institute, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China
[4]Faculty of Business and Economics, University of Hong Kong, Pok Fu Lam, Hong Kong SAR

Correspondence
Haihui Shen, Sino-US Global Logistics Institute, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China.
Email: shenhaihui@sjtu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 72001140, 72091211, 71991473. "Chenguang Program" supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission, Grant/Award Number: 19CG14. Hong Kong Research Grants Council, Grant/Award Numbers: GRF 17201520, 16211417.

Abstract

Knowledge gradient is a design principle for developing Bayesian sequential sampling policies to solve optimization problems. In this paper, we consider the ranking and selection problem in the presence of covariates, where the best alternative is not universal but depends on the covariates. In this context, we prove that under minimal assumptions, the sampling policy based on knowledge gradient is consistent, in the sense that following the policy the best alternative as a function of the covariates will be identified almost surely as the number of samples grows. We also propose a stochastic gradient ascent algorithm for computing the sampling policy and demonstrate its performance via numerical experiments.

KEYWORDS

consistency, covariates, knowledge gradient, selection of the best

1 | INTRODUCTION

We consider the ranking and selection (R&S) problem in the presence of covariates. A decision maker is presented with a finite collection of alternatives. The performance of each alternative is unknown and depends on the covariates. Suppose that the decision maker has access to noisy samples of each alternative for any chosen value of the covariates, but the samples are expensive to acquire. Given a finite sampling budget, the goal is to develop an efficient sampling policy indicating locations as to which alternative and what value of the covariates to sample from, so that upon termination of the sampling, the decision maker can identify a decision rule that accurately specifies the best alternative as a function of the covariates.

The problem of R&S with covariates emerges naturally as the popularization of data and decision analytics in recent years. In clinical and medical research, for many diseases the effect of a treatment may be substantially different across patients, depending on their biometric characteristics (i.e., the covariates), including age, weight, lifestyle habits such as smoking and alcohol use, and so forth (Kim et al., 2011). A treatment regime that works for a majority of patients might not work for the others. Samples needed for estimating treatment effects may be collected from clinical trials or computer simulation. For example, in Hur et al. (2004) and Choi et al. (2014), a simulation model is developed to simulate the effect of several treatment regimens for Barrett's esophagus, a precursor to esophageal cancer, for patients with different biometric characteristics. Personalized medicine can then be developed to determine the best treatment regime that is customized to the particular characteristics of each individual patient. Similar customized decision making can be found in online advertising (Arora et al., 2008), where advertisements are displayed depending on consumers' web browsing history or buying behavior to increase the revenue of the advertising platform as well as to improve consumers' shopping experience.

Being a classic problem in the area of stochastic simulation, R&S has a vast literature. We refer to Kim and Nelson (2006) and Chen et al. (2015) for reviews on the subject with emphasis on frequentist and Bayesian

approaches, respectively. Most of the prior work, however, does not consider the presence of the covariates, and thus the best alternative to select is universal rather than varies as a function of the covariates. There are several exceptions, including Hu and Ludkovski (2017), Pearce and Branke (2017), and Shen et al. (2021). Among them Shen et al. (2021) take a frequentist approach to solve R&S with covariates, whereas the other two a Bayesian approach. The present paper adopts a Bayesian perspective as well.

This paper considers a sampling policy based on knowledge gradient (KG) for R&S with covariates. KG, introduced in Frazier et al. (2008), is a design principle that has been widely used for developing Bayesian sequential sampling policies to solve a variety of optimization problems, including R&S, in which evaluation of the objective function is noisy and expensive. In its basic form, KG begins with assigning a multivariate normal prior on the unknown constant performance of all alternatives. In each iteration, it chooses the sampling location by maximizing the increment in the expected value of the information that would be gained by taking a sample from the location. Then, the posterior is updated upon observing the noisy sample from the chosen location. The sampling efficiency of KG-type policies is often competitive with or outperforms other sampling policies; see Frazier et al. (2009), Scott et al. (2011), Ryzhov (2016), and Pearce and Branke (2018) among others.

A KG-based sampling policy for R&S with covariates is also proposed in Pearce and Branke (2017). The main difference here is that our treatment is more general. First, we allow the sampling noise to be heteroscedastic, whereas it is assumed to be constant for different locations of the same alternative in their work. Heteroscedasticity is of particular significance for simulation applications such as queueing systems. Second, we take into account possible variations in sampling cost at different locations, whereas the sampling cost is simply treated as constant everywhere in Pearce and Branke (2017). Hence, our policy, which we refer to as integrated knowledge gradient (IKG), attempts in each iteration to maximize a "cost-adjusted" increment in the expected value of information. These generalizations are straightforward when the variance of the sampling noise and the sampling cost are assumed to be known. We also briefly discuss and show how to deal with the case where they are unknown.

The first main contribution of this paper is to provide a theoretical analysis of the asymptotic behavior of the IKG policy, whereas Pearce and Branke (2017) conducted only numerical investigation. In particular, we prove that IKG is consistent in the sense that for any value of the covariates, the selected alternative upon termination of the sampling will converge to the true best almost surely as the sampling budget grows to infinity. Moreover, we consider a practical variant—termed quasi-IKG—which does not require the intermediate optimization problem in each iteration of IKG to be solved exactly, and prove its consistency under mild conditions.

Consistency of KG-type policies has been established in various settings, mostly for problems where the number of feasible solutions is finite, including R&S (Frazier et al., 2008, 2009; Frazier & Powell, 2011; Mes et al., 2011), and discrete optimization via simulation (Xie et al., 2016). KG is also used for Bayesian optimization of continuous functions in Wu and Frazier (2016), Poloczek et al. (2017), and Wu et al. (2017). However, in these papers the continuous domain is discretized first, which effectively reduces the problem to one with finite feasible solutions, in order to facilitate their asymptotic analysis. The finiteness of the domain is critical in the aforementioned papers, because the asymptotic analysis there boils down to proving that each feasible solution can be sampled infinitely often. This, by the law of large numbers, implies that the variance of the objective value estimate of each solution will converge to zero. Thus, the optimal solution will be identified ultimately since the uncertainty about the performances of the solutions will be removed completely in the end.

By contrast, proving consistency of KG-type policies for continuous solution domains demands a fundamentally different approach, since most solutions in a continuous domain would hardly be sampled even once after all. Among the several related papers, Scott et al. (2011) studies a KG-type policy for Bayesian optimization of continuous functions. Assigning a Gaussian process prior on the objective function, they established the consistency of the KG-type policy basically by leveraging the continuity of the covariance function of the Gaussian process, which intuitively suggests that if the variance at one location is small, then the variance in its neighborhood ought to be small too. Toscano-Palmerin and Frazier (2018) prove the consistency of a KG-type policy on a more general problem that can reduce to the problem in Scott et al. (2011), for both discrete and continuous domains.

We cast R&S with covariates to a problem of ranking a finite number of Gaussian processes, thereby having both discrete and continuous elements structurally. As a result, we establish the consistency of the proposed IKG policy by proving the following two facts—(i) each Gaussian process is sampled infinitely often, and (ii) the infinitely many samples assigned to a given Gaussian process drives its posterior variance at any location to zero, thanks to the assumed continuity of its covariance function. The theoretical analysis in this paper is partly built on the ideas developed for discrete and continuous problems, respectively, in Frazier et al. (2008) and Scott et al. (2011) in a federated manner.

Although our proofs share similar structures to those in Scott et al. (2011), our assumptions are substantially simpler and minimal. By contrast, for the proof in Scott et al. (2011) to be valid, technical conditions are imposed to regulate the asymptotic behavior of the posterior mean function and the posterior covariance function of the underlying Gaussian process. Nevertheless, the two conditions are difficult to verify. We do not impose such conditions. We achieve

the substantial simplification of the assumptions by leveraging the reproducing kernel Hilbert space (RKHS) theory. The theory has been used widely in machine learning (Steinwart & Christmann, 2008). But its use in the analysis of KG-type policies is less common. We develop several technical results based on RKHS theory to facilitate analysis of the asymptotic behavior of the posterior covariance function.[1]

The second main contribution of this paper is that we develop an algorithm to solve a stochastic optimization problem that determines the sampling decision of the IKG policy in its each iteration. In Pearce and Branke (2017), this optimization problem is addressed by the sample average approximation method with a derivative-free optimization solver. Instead, we propose a stochastic gradient ascent (SGA) algorithm, taking advantage of the fact that a gradient estimator can be derived analytically for many popular covariance functions. Numerical experiments demonstrate the finite-sample performance of the IKG policy in conjunction with the SGA algorithm.

We conclude the introduction by reviewing briefly the most pertinent literature. A closely related problem is multi-armed bandit (MAB); see Bubeck and Cesa-Bianchi (2012) for a comprehensive review on the subject. The significance of covariates, thereby contextual MAB (or MAB with covariates), has also drawn substantial attention in recent years; see Rusmevichientong and Tsitsiklis (2010), Yang and Zhu (2002), Krause and Ong (2011), and Perchet and Rigollet (2013) among others. There are two critical differences between contextual MAB and R&S with covariates. First, the former generally assumes that the covariates arrive exogenously in a sequential manner, and the decision maker can choose at which arm (or alternative) to sample but not the value of covariates. By contrast, the latter assumes that the decision maker is capable of choosing both the alternative and the covariates when specifying sampling locations. A second difference is MAB focuses on minimizing the regret which is caused by choosing inferior alternatives and accumulated during the sampling process, whereas R&S focuses on identifying the best alternative eventually and the regret is not the primary concern.

The rest of the paper is organized as follows. In Section 2, we follow a nonparametric Bayesian approach to formulate the problem of R&S with covariates, introduce the IKG policy, and present the main result. In Section 3, we prove the consistency of our sampling policy in the sense that the estimated best alternative as a function of the covariates converges to the truth with probability one as the number of samples grows to infinity. We then propose to use SGA for computing our sampling policy in Section 4, and demonstrate its performance via numerical experiments in Section 5. We conclude in Section 6 and collect detailed proof and additional technical results and numerical experiments in the Online Appendix.

[1]Bect et al. (2019) adopt a supermartingale approach to study the asymptotic behavior of a general class of sequential sampling algorithms. Their analysis has a broader scope of applicability but it is technically more involved.

2 | PROBLEM FORMULATION

Suppose that a decision maker is presented with M competing alternatives. For each $i = 1, \dots, M$, the performance of alternative i depends on a vector of *covariates* $\boldsymbol{x} = (x_1, \dots, x_d)^\intercal$ and is denoted by $\theta_i = \theta_i(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$. The performances are unknown and can only be learned via sampling. In particular, for any i and $\boldsymbol{x}$, one can acquire possibly multiple noisy samples of $\theta_i(\boldsymbol{x})$. The decision maker aims to select the "best" alternative for a given value of $\boldsymbol{x}$, that is, identify $\operatorname{argmax}_i \theta_i(\boldsymbol{x})$. However, since the sampling is usually expensive in time and/or money, instead of estimating the performances $\{\theta_i(\boldsymbol{x}) : i = 1, \dots, M\}$ every time a new value of $\boldsymbol{x}$ is observed and then ranking them, it is preferable to learn *offline* the decision rule

$$i^*(\boldsymbol{x}) \in \operatorname*{argmax}_{1 \le i \le M} \theta_i(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{X}, \tag{1}$$

as a function of $\boldsymbol{x}$, through a carefully designed sampling process. Equipped with such a decision rule, the decision maker can select the best alternative upon observing the covariates in a timely fashion. In addition, the decision maker may have some knowledge with regard to the covariates. For example, certain values of the covariates may be more important or appear more frequently than others. Suppose that this kind of knowledge is expressed by a probability density function $\gamma(\boldsymbol{x})$ on $\mathcal{X}$.

During the offline learning period, we need to make a sequence of sampling decisions $\{(a^n, \boldsymbol{v}^n) : n = 0, 1, \dots\}$, where $(a^n, \boldsymbol{v}^n)$ means that the $(n+1)$-th sample, denoted by y^{n+1}, is taken from alternative a^n with covariates value $\boldsymbol{v}^n$ (refer it as location $\boldsymbol{v}^n$ for simplicity). We assume that given $\theta_{a^n}(\boldsymbol{v}^n)$, y^{n+1} is an unbiased sample having a normal distribution, that is,

$$y^{n+1} \,|\, \theta_{a^n}(\boldsymbol{v}^n) \sim \mathcal{N}\left(\theta_{a^n}(\boldsymbol{v}^n), \lambda_{a^n}(\boldsymbol{v}^n)\right),$$

where $y^n \mid \theta_i(\boldsymbol{x})$ is independent of $y^{n'} \mid \theta_{i'}(\boldsymbol{x}')$ for $(i, \boldsymbol{x}, n) \neq (i', \boldsymbol{x}', n')$. Here, $\lambda_i(\boldsymbol{x})$ is the variance of a sample of $\theta_i(\boldsymbol{x})$ given $\theta_i(\boldsymbol{x})$ and is assumed to be *known*. Moreover, suppose that the cost of taking a sample from alternative i at location $\boldsymbol{x}$ is $c_i(\boldsymbol{x}) > 0$, which is also assumed to be *known*. Suppose that the total sampling budget for offline learning is $B > 0$, and the sampling process is terminated when the budget is exhausted. Mathematically, we will stop with the $N(B)$-th sample, where

$$N(B) := \min\left\{ N : \sum_{n=0}^{N} c_{a^n}(\boldsymbol{v}^n) > B \right\}. \tag{2}$$

Consequently, the sampling decisions are $\{(a^n, \boldsymbol{v}^n) : n = 0, \dots, N(B)-1\}$ and the samples taken during the process are $\left\{y^{n+1} : n = 0, \dots, N(B)-1\right\}$. Notice that $N(B) = B$ if $c_i(\boldsymbol{x}) \equiv 1$ for $i = 1, \dots, M$, in which case the sampling budget is reduced to the number of samples.

Remark 1 The assumption of known $\lambda_i(\boldsymbol{x})$ is critical to the theoretical analysis in this paper. As we will see shortly, with known $\lambda_i(\boldsymbol{x})$, if we impose a Gaussian process as prior for θ_i, then its posterior will still be a Gaussian process, which makes the asymptotic analysis tractable. It would not be the case if $\lambda_i(\boldsymbol{x})$ also needs to be estimated. In practice, $\lambda_i(\boldsymbol{x})$ is usually unknown and it is a common issue in the experiment design. We suggest to follow the approach in Ankenman et al. (2010), which fits the surfaces of $\lambda_i(\boldsymbol{x})$ by running multiple simulations at certain design points and using the sample variances. See more details in the numerical experiments in the Online Appendix. The unknown sampling cost $c_i(\boldsymbol{x})$ in practice can be handled similarly.

We follow a nonparametric Bayesian approach to model the unknown functions $\{\theta_1, \dots, \theta_M\}$ as well as to design the sampling policy. We treat θ_i's as random functions and impose a prior on them under which they are mutually independent, although this assumption may be relaxed. Suppose that $\boldsymbol{x}$ takes continuous values and that under the prior, θ_i is a Gaussian process with mean function $\mu_i^0(\boldsymbol{x}) := \mathbb{E}[\theta_i(\boldsymbol{x})]$ and covariance function $k_i^0(\boldsymbol{x}, \boldsymbol{x}') := \mathrm{Cov}\left[\theta_i(\boldsymbol{x}), \theta_i(\boldsymbol{x}')\right]$ that satisfies the following assumption.

Assumption 1 For each $i = 1, \dots, M$, there exists a constant $\tau_i > 0$ and a positive continuous function $\rho_i : \mathbb{R}^d \mapsto \mathbb{R}_+$ such that $k_i^0(\boldsymbol{x}, \boldsymbol{x}') = \tau_i^2 \rho_i(\boldsymbol{x} - \boldsymbol{x}')$. Moreover,

(i) $\rho_i(|\boldsymbol{\delta}|) = \rho_i(\boldsymbol{\delta})$, where $|\cdot|$ means taking the absolute value component-wise;
(ii) $\rho_i(\boldsymbol{\delta})$ is decreasing in $\boldsymbol{\delta}$ component-wise for $\boldsymbol{\delta} \geq \mathbf{0}$;
(iii) $\rho_i(\mathbf{0}) = 1$, $\rho_i(\boldsymbol{\delta}) \to 0$ as $\|\boldsymbol{\delta}\| \to \infty$, where $\|\cdot\|$ denotes the Euclidean norm;
(iv) there exist some $0 < C_i < \infty$ and $\varepsilon_i, u_i > 0$ such that

$$1 - \rho_i(\boldsymbol{\delta}) \leq \frac{C_i}{|\log(\|\boldsymbol{\delta}\|)|^{1+\varepsilon_i}},$$

for all $\boldsymbol{\delta}$ such that $\|\boldsymbol{\delta}\| < u_i$.

Remark 2 Assumption 1 stipulates that k_i^0 is stationary, that is, it depends on $\boldsymbol{x}$ and $\boldsymbol{x}'$ only through the difference $\boldsymbol{x} - \boldsymbol{x}'$. In addition, τ_i^2 can be interpreted as the prior variance of $\theta_i(\boldsymbol{x})$ for all $\boldsymbol{x}$, and $\rho_i(\boldsymbol{x} - \boldsymbol{x}')$ as the prior correlation between $\theta_i(\boldsymbol{x})$ and $\theta_i(\boldsymbol{x}')$ which increases to 1 as $\|\boldsymbol{x} - \boldsymbol{x}'\|$ decreases to 0. The condition in part (iv) of Assumption 1 is weak. In conjunction with the continuity assumption of ρ_i, it implies that the sample paths of the Gaussian process θ_i are continuous almost surely if the mean function $\mu_i^0(\boldsymbol{x})$ is continuous; see, for example, Adler and Taylor (2007, Theorem 1.4.1). The sample path continuity will be used to establish the uniform convergence of the posterior mean functions.

A variety of covariance functions satisfy Assumption 1. Notable examples include the squared exponential (SE) covariance function

$$k_{\mathrm{SE}}(\boldsymbol{x}, \boldsymbol{x}') = \tau^2 \exp\left(-r^2(\boldsymbol{x} - \boldsymbol{x}')\right),$$

where $r(\boldsymbol{\delta}) = \sqrt{\sum_{j=1}^d \alpha_j \delta_j^2}$ and α_j's are positive parameters, and the Matérn covariance function

$$k_{\text{Matérn}}(\boldsymbol{x}, \boldsymbol{x}') = \tau^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} r(\boldsymbol{x} - \boldsymbol{x}')\right)^{\nu} \times K_\nu\left(\sqrt{2\nu} r(\boldsymbol{x} - \boldsymbol{x}')\right),$$

where ν is a positive parameter that is typically taken as half-integer (i.e., $\nu = p + 1/2$ for some nonnegative integer p), Γ is the gamma function, and K_ν is the modified Bessel function of the second kind. The covariance function reflects one's prior belief about the unknown functions. We refer to Rasmussen and Williams (2006, Chapter 4) for more types of covariance functions.

2.1 | Bayesian updating equations

For each $n = 1, 2, \dots$, let $\mathcal{F}^n$ denote the σ-algebra generated by $(a^0, \boldsymbol{v}^0), y^1, \dots, (a^{n-1}, \boldsymbol{v}^{n-1}), y^n$, the sampling decisions and the samples collected up to time n. Suppose that $(a^n, \boldsymbol{v}^n) \in \mathcal{F}^n$, that is, $(a^n, \boldsymbol{v}^n)$ depends only on the information available at time n. In addition, we use the notation $\mathbb{E}^n[\cdot] := \mathbb{E}[\cdot|\mathcal{F}^n]$, and define $\mathrm{Var}^n[\cdot]$ and $\mathrm{Cov}^n[\cdot]$ likewise.

Given the setup of our model, it is easy to derive that $\{\theta_1, \dots, \theta_M\}$ are independent Gaussian processes under the posterior distribution conditioned on $\mathcal{F}^n$, $n = 1, \dots, N(B)$. In particular, under the prior mutual independence, taking samples from one unknown function does not provide information on another. Let $\boldsymbol{V}_i^n := \{\boldsymbol{v}^\ell : a^\ell = i, \ell = 0, \dots, n-1\}$ denote the set of the locations of the samples taken from θ_i up to time n and define $\boldsymbol{y}_i^n := \{y^{\ell+1} : a^\ell = i, \ell = 0, \dots, n-1\}$ likewise. With slight abuse of notation, when necessary, we will also treat $\boldsymbol{V}_i^n$ as a matrix wherein the columns are corresponding to the points in the set and arranged in the order of appearance, and $\boldsymbol{y}_i^n$ as a column vector with elements also arranged in the order of appearance. Then, the posterior mean and covariance functions of θ_i are given by

$$\mu_i^n(\boldsymbol{x}) := \mathbb{E}^n[\theta_i(\boldsymbol{x})] = \mu_i^0(\boldsymbol{x}) + k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n) \times \left[k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i(\boldsymbol{V}_i^n)\right]^{-1} \left[\boldsymbol{y}_i^n - \mu_i^0(\boldsymbol{V}_i^n)\right], \quad (3)$$

$$k_i^n(\boldsymbol{x}, \boldsymbol{x}') := \mathrm{Cov}^n\left[\theta_i(\boldsymbol{x}), \theta_i(\boldsymbol{x}')\right] = k_i^0(\boldsymbol{x}, \boldsymbol{x}') - k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n) \times \left[k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i(\boldsymbol{V}_i^n)\right]^{-1} k_i^0(\boldsymbol{V}_i^n, \boldsymbol{x}'), \quad (4)$$

where for two sets $\boldsymbol{V}$ and $\boldsymbol{V}'$, $k_i^0(\boldsymbol{V}, \boldsymbol{V}') = \left[k_i^0(\boldsymbol{x}, \boldsymbol{x}')\right]_{\boldsymbol{x} \in \boldsymbol{V}, \boldsymbol{x}' \in \boldsymbol{V}'}$ is a matrix of size $|\boldsymbol{V}| \times |\boldsymbol{V}'|$, $\lambda_i(\boldsymbol{V}) = \mathrm{diag}\{\lambda_i(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{V}\}$ is a diagonal matrix of size $|\boldsymbol{V}| \times |\boldsymbol{V}|$, and $\mu_i^0(\boldsymbol{V}) =$

$\left(\mu_i^0(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{V}\right)$ is a column vector of size $|\boldsymbol{V}| \times 1$. Here $|\cdot|$ denotes the cardinality of a set. We refer to, for example, Scott et al. (2011, section 3.2) for details. Further, the following updating equation can be derived

$$\mu_i^{n+1}(\boldsymbol{x}) = \mu_i^n(\boldsymbol{x}) + \sigma_i^n(\boldsymbol{x}, \boldsymbol{v}^n) Z^{n+1}, \tag{5}$$

$$k_i^{n+1}(\boldsymbol{x}, \boldsymbol{x}') = k_i^n(\boldsymbol{x}, \boldsymbol{x}') - \sigma_i^n(\boldsymbol{x}, \boldsymbol{v}^n) \sigma_i^n\left(\boldsymbol{x}', \boldsymbol{v}^n\right), \tag{6}$$

where Z^{n+1} is a standard normal random variable independent to everything else, and

$$\sigma_i^n(\boldsymbol{x}, \boldsymbol{v}^n) \coloneqq \begin{cases} \widetilde{\sigma}_i^n(\boldsymbol{x}, \boldsymbol{v}^n), & \text{if } i = a^n, \\ 0, & \text{if } i \neq a^n, \end{cases} \quad \text{and}$$

$$\widetilde{\sigma}_i^n(\boldsymbol{x}, \boldsymbol{v}) \coloneqq \frac{k_i^n(\boldsymbol{x}, \boldsymbol{v})}{\sqrt{k_i^n(\boldsymbol{v}, \boldsymbol{v}) + \lambda_i(\boldsymbol{v})}}. \tag{7}$$

In particular, conditioned on $\mathcal{F}^n$ and prior to taking a sample at $(a^n, \boldsymbol{v}^n)$, the predictive distribution of $\mu_i^{n+1}(\boldsymbol{x})$ is normal with mean $\mu_i^n(\boldsymbol{x})$ and standard deviation $\sigma_i^n(\boldsymbol{x}, \boldsymbol{v}^n)$. Moreover, notice that

$$\begin{aligned} \mathrm{Var}^{n+1}\left[\theta_i(\boldsymbol{x})\right] &= k_i^{n+1}(\boldsymbol{x}, \boldsymbol{x}) = k_i^n(\boldsymbol{x}, \boldsymbol{x}) - \left[\sigma_i^n(\boldsymbol{x}, \boldsymbol{v}^n)\right]^2 \\ &\leq \mathrm{Var}^n\left[\theta_i(\boldsymbol{x})\right]. \end{aligned} \tag{8}$$

(Note that Equations (5)–(8) are still valid even if $k_i^n(\boldsymbol{v}^n, \boldsymbol{v}^n) = 0$, and/or $\lambda_i(\boldsymbol{v}^n) = 0$.) Hence, $\mathrm{Var}^n\left[\theta_i(\boldsymbol{x})\right]$ is nonincreasing in n. This basically suggests that the uncertainty about each unknown function under the posterior decreases as more samples from it are collected. It is thus both desirable and practically meaningful that such uncertainty would be completely eliminated if the sampling budget is unlimited, in which case one would be able to identify the decision rule Equation (1) perfectly. In particular, we define consistency of a sampling policy as follows.

Definition 1 A sampling policy is said to be consistent if it ensures that

$$\lim_{B \to \infty} \operatorname*{argmax}_{1 \leq i \leq M} \mu_i^{N(B)}(\boldsymbol{x}) = \operatorname*{argmax}_{1 \leq i \leq M} \theta_i(\boldsymbol{x}), \tag{9}$$

almost surely (a.s.) for all $\boldsymbol{x} \in \mathcal{X}$.

Remark 3 Under the assumption that $\{\theta_1, \ldots, \theta_M\}$ are independent under the prior, collecting samples from θ_i does not provide information about θ_j if $i \neq j$. Therefore, a consistent policy under the independence assumption ought to ensure that the number of samples taken from each θ_i grows without bounds.

2.2 | Knowledge gradient policy

We first assume temporarily that $\boldsymbol{x}$ is given and fixed, and that $c_i(\boldsymbol{x}) = 1$ for $i = 1, \ldots, M$. Then, solving $\max_i \theta_i(\boldsymbol{x})$ is a selection of the best problem having finite alternatives, and each sampling decision is reduced to choosing an alternative i to take a sample of $\theta_i(\boldsymbol{x})$. The knowledge gradient (KG) policy introduced in Frazier et al. (2008) is designed exactly to solve such a problem assuming an independent normal prior. Specifically, the knowledge gradient at i is defined there as the increment in the expected value of the information about the maximum at $\boldsymbol{x}$ gained by taking a sample at i, that is,

$$\mathrm{KG}^n(i; \boldsymbol{x}) \coloneqq \mathbb{E}\left[\max_{1 \leq a \leq M} \mu_a^{n+1}(\boldsymbol{x}) \middle| \mathcal{F}^n, a^n = i\right] - \max_{1 \leq a \leq M} \mu_a^n(\boldsymbol{x}). \tag{10}$$

Then, each time the alternative i that has the largest value of $\mathrm{KG}(i; \boldsymbol{x})$ is selected to generate a sample of $\theta_i(\boldsymbol{x})$.

Let us now return to our context where (1) the covariates are present, (2) each sampling decision consists of both i and $\boldsymbol{x}$, and (3) each sampling decision may induce a different sampling cost. Since a sample of $\theta_i(\boldsymbol{x})$ would alter the posterior belief about $\theta_i(\boldsymbol{x}')$, we generalize Equation (10) and define

$$\begin{aligned} \mathrm{KG}^n(i, \boldsymbol{x}; \boldsymbol{v}) \coloneqq \frac{1}{c_i(\boldsymbol{x})} \Bigg\{ &\mathbb{E}\left[\max_{1 \leq a \leq M} \mu_a^{n+1}(\boldsymbol{v}) \middle| \mathcal{F}^n, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right] \\ &- \max_{1 \leq a \leq M} \mu_a^n(\boldsymbol{v}) \Bigg\}, \end{aligned} \tag{11}$$

which can be interpreted as the increment in the expected value of the information about the maximum at $\boldsymbol{v}$ gained per unit of sampling cost by taking a sample at $(i, \boldsymbol{x})$. Then, we consider the following *integrated KG* (IKG)

$$\begin{aligned} \mathrm{IKG}^n(i, \boldsymbol{x}) \coloneqq \frac{1}{c_i(\boldsymbol{x})} \int_{\mathcal{X}} \Bigg\{ &\mathbb{E}\left[\max_{1 \leq a \leq M} \mu_a^{n+1}(\boldsymbol{v}) \middle| \mathcal{F}^n, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right] \\ &- \max_{1 \leq a \leq M} \mu_a^n(\boldsymbol{v}) \Bigg\} \gamma(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}, \end{aligned} \tag{12}$$

and define the IKG sampling policy as

$$(a^n, \boldsymbol{v}^n) \in \operatorname*{argmax}_{1 \leq i \leq M, \boldsymbol{x} \in \mathcal{X}} \mathrm{IKG}^n(i, \boldsymbol{x}). \tag{13}$$

The integrand of Equation (12) can be calculated analytically, as shown in Lemma 1, whose proof is deferred to the Online Appendix.

Lemma 1 *For all $i = 1, \ldots, M$ and $\boldsymbol{x} \in \mathcal{X}$,*

$$\begin{aligned} \mathrm{IKG}^n(i, \boldsymbol{x}) = \frac{1}{c_i(\boldsymbol{x})} \int_{\mathcal{X}} \Bigg[&\left|\widetilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})\right| \phi\left(\left|\frac{\Delta_i^n(\boldsymbol{v})}{\widetilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})}\right|\right) \\ &- \left|\Delta_i^n(\boldsymbol{v})\right| \Phi\left(-\left|\frac{\Delta_i^n(\boldsymbol{v})}{\widetilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})}\right|\right) \Bigg] \gamma(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}, \end{aligned} \tag{14}$$

where $\Delta_i^n(\boldsymbol{v}) \coloneqq \mu_i^n(\boldsymbol{v}) - \max_{a \neq i} \mu_a^n(\boldsymbol{v})$, Φ is the standard normal distribution function, and ϕ is its density function.

We solve Equation (13) by first solving $\max_{\boldsymbol{x}} \mathrm{IKG}^n(i, \boldsymbol{x})$ for all i and then enumerating the results. The computational challenge in the former lies in the numerical integration in Equation (14). Notice that $\max_{\boldsymbol{x}} \mathrm{IKG}^n(i, \boldsymbol{x})$ is in fact a stochastic optimization problem if we view the integration in Equation (14) as an expectation with respect to the probability density $\gamma(\boldsymbol{x})$ on $\mathcal{X}$. One might apply the sample average approximation method to solve $\max_{\boldsymbol{x}} \mathrm{IKG}^n(i, \boldsymbol{x})$, but it would

be computationally prohibitive if $\mathcal{X}$ is high-dimensional. Instead, we show in Section 4 that the gradient of the integrand in Equation (14) with respect to $\boldsymbol{x}$ can be calculated explicitly, which is an unbiased estimator of $\nabla_{\boldsymbol{x}} \mathrm{IKG}^n(i,\boldsymbol{x})$ under regularity conditions, thereby leading to a stochastic gradient ascent method (Kushner & Yin, 2003).

We now present our main theoretical result—the IKG policy is consistent under simple assumptions. The proof will be sketched in Section 3 and all details are collected in the Online Appendix.

Assumption 2 The design space $\mathcal{X}$ is a compact set in $\mathbb{R}^d$ with nonempty interior.

Assumption 3 For each $i = 1, \ldots, M$, $\mu_i^0(\cdot)$, $\lambda_i(\cdot) > 0$ and $c_i(\cdot) > 0$ are all continuous on $\mathcal{X}$, and $\gamma(\cdot) > 0$ on $\mathcal{X}$.

Under Assumptions 1–3, the IKG policy (13) is well defined. This can be seen by noting that the maximum of $\mathrm{IKG}^n(i,\boldsymbol{x})$ over $\boldsymbol{x} \in \mathcal{X}$ is attainable since $\mathrm{IKG}^n(i,\boldsymbol{x})$ is continuous in $\boldsymbol{x}$ by Assumptions 1 and 3 together with Lemma 1, and $\mathcal{X}$ is compact by Assumption 2. Moreover, the IKG policy (13) is consistent as formally stated in the following Theorem 1.

Theorem 1 *If Assumptions* 1–3 *hold, then the IKG policy* (13) *is consistent, that is, under the IKG policy,*

(i) $k_i^{N(B)}(\boldsymbol{x},\boldsymbol{x}) \to 0$ *a.s. as* $B \to \infty$ *for all* $\boldsymbol{x} \in \mathcal{X}$ *and* $i = 1, \ldots, M$;
(ii) $\mu_i^{N(B)}(\boldsymbol{x}) \to \theta_i(\boldsymbol{x})$ *a.s. as* $B \to \infty$ *for all* $\boldsymbol{x} \in \mathcal{X}$ *and* $i = 1, \ldots, M$;
(iii) $\mathrm{argmax}_{1\le i\le M}\mu_i^{N(B)}(\boldsymbol{x}) \to \mathrm{argmax}_{1\le i\le M}\theta_i(\boldsymbol{x})$ *a.s. as* $B \to \infty$ *for all* $\boldsymbol{x} \in \mathcal{X}$.

We conclude this section by highlighting the differences between our assumptions and those in Scott et al. (2011), in which the consistency of a KG-type policy driven by a Gaussian process is proved. First and foremost, *they impose conditions on both the posterior mean function and the posterior covariance function to regulate their large-sample asymptotic behavior*. Specifically, they assume that uniformly for all n and $\boldsymbol{x},\boldsymbol{v} \in \mathcal{X}$ with $\boldsymbol{x} \neq \boldsymbol{v}$, (1) $|\mu^n(\boldsymbol{x}) - \mu^n(\boldsymbol{v})|$ is bounded a.s., and (2) $|\mathrm{Corr}^n[\theta(\boldsymbol{x}),\theta(\boldsymbol{v})]|$ is bounded above away from one, where Corr^n means the posterior correlation.[2] *The two assumptions are critical for their analysis but nontrivial to verify in practice*.

By contrast, we do not make such assumptions. Condition (1) is not necessary in our analysis because the "increment in the expected value of the information" is defined as Equation (12) in this paper, whereas in a different form without integration in Scott et al. (2011). There is no need for us to impose Condition (2) in order to regulate the asymptotic behavior of the posterior covariance function, because instead we achieve the same goal by utilizing reproducing kernel Hilbert space (RKHS) theory.

Second, in Scott et al. (2011) the prior covariance function of the underlying Gaussian process is of SE type. We relax it to Assumption 1, which allows a great variety of covariance functions. We also take into account possibly varying sampling costs at different locations.

[2]The subscript i is ignored because there is only one Gaussian process involved in Scott et al. (2011).

3 | CONSISTENCY

It is straightforward to show that $N(B) \to \infty$ if and only if $B \to \infty$, since $c_i(\cdot)$ is bounded both above and below away from zero on $\mathcal{X}$ for each $i = 1, \ldots, M$ under Assumptions 2 and 3. Thus, Theorem 1 is equivalent to Theorem 2 as follows.

Theorem 2 *If Assumptions* 1–3 *hold, then under the IKG policy,*

(i) $k_i^n(\boldsymbol{x},\boldsymbol{x}) \to 0$ *a.s. as* $n \to \infty$ *for all* $\boldsymbol{x} \in \mathcal{X}$ *and* $i = 1, \ldots, M$;
(ii) $\mu_i^n(\boldsymbol{x}) \to \theta_i(\boldsymbol{x})$ *a.s. as* $n \to \infty$ *for all* $\boldsymbol{x} \in \mathcal{X}$ *and* $i = 1, \ldots, M$;
(iii) $\mathrm{argmax}_{1\le i\le M}\mu_i^n(\boldsymbol{x}) \to \mathrm{argmax}_{1\le i\le M}\theta_i(\boldsymbol{x})$ *a.s. as* $n \to \infty$ *for all* $\boldsymbol{x} \in \mathcal{X}$.

The bulk of the proof of consistency of the IKG policy lies in part (i) of Theorem 2, that is, to show that $\lim_{n\to\infty} \mathrm{Var}^n[\theta_i(\boldsymbol{x})] = 0$ a.s. for all $\boldsymbol{x} \in \mathcal{X}$ and $i = 1, \ldots, M$. It consists of two steps, which are summarized into the later Propositions 2 and 3. However, both Propositions 2 and 3 critically relies on the asymptotic behavior of the posterior covariance function, which is characterized in the following Proposition 1.

Proposition 1 *Fix* $i = 1, \ldots, M$. *If* k_i^0 *is stationary, then for any* $\boldsymbol{x} \in \mathcal{X}$, $k_i^n(\boldsymbol{x},\boldsymbol{x}')$ *converges to a limit, denoted by* $k_i^\infty(\boldsymbol{x},\boldsymbol{x}')$, *uniformly in* $\boldsymbol{x}' \in \mathcal{X}$ *as* $n \to \infty$.

Proposition 1 shows that *irrespective* of the allocation of the design points $\{\boldsymbol{v}^\ell : \ell = 0, \ldots, n-1\}$, $k_i^n(\boldsymbol{x},\cdot)$ converges *uniformly* as $n \to \infty$ for all $\boldsymbol{x} \in \mathcal{X}$. (Note that this does not mean the limit is necessarily zero.) Not only is this result of interest in its own right, but also is crucial for proving the consistency of IKG policy under assumptions weaker than those imposed for previous related problems (Scott et al., 2011). For example, the uniform convergence preserves the continuity of $k_i^n(\boldsymbol{x},\cdot)$ in the limit, a property that is crucial for the proof of Proposition 2. A more general version of Proposition 1 is given in Bect et al. (2019, Proposition 2.9), but we present a different proof built on RKHS theory in the Online Appendix. Proposition 1 sets a foundation for analyzing the asymptotic behavior of Bayesian sequential sampling policies based on Gaussian processes with minimal assumptions.

Before we formally state Propositions 2 and 3, the following definitions are required. For each i, let η_i^n denote the (random) number of times that a sample is taken from alternative i regardless of the value of $\boldsymbol{x}$ up to the n-th sample, that is,

$$\eta_i^n := \sum_{\ell=0}^{n-1} \mathbb{I}_{\{a^\ell = i\}}.$$

Further, let $\eta_i^\infty := \lim_{n\to\infty} \eta_i^n$, which is well defined since it is a limit of a nondecreasing sequence of random variables.

Proposition 2 *Fix $i = 1, \dots, M$. Assumptions 1–3 hold and $\eta_i^\infty = \infty$ a.s., then for any $\boldsymbol{x} \in \mathcal{X}$, $k_i^\infty(\boldsymbol{x}, \boldsymbol{x}) = 0$ a.s. under the IKG policy.*

Proposition 3 *If Assumptions 1–3 hold, then $\eta_i^\infty = \infty$ a.s. for each $i = 1, \dots, M$ under the IKG policy.*

Part (i) of Theorem 2 is an immediate consequence of Propositions 2 and 3. The proofs of parts (ii) and (iii) of Theorem 2 and Propositions 2 and 3 are all collected in the Online Appendix.

In practice, the IKG policy (13) can only be solved numerically, as discussed in the next section, in which case the obtained solution $\left(\widetilde{a}^n, \widetilde{\boldsymbol{v}}^n\right)$ is not exactly equal to the true solution $(a^n, \boldsymbol{v}^n)$. Inspired by Bect et al. (2019), we consider the *quasi*-IKG sampling policy, which chooses the sampling decision $\left(\widetilde{a}^n, \widetilde{\boldsymbol{v}}^n\right)$ such that

$$\mathrm{IKG}^n\left(\widetilde{a}^n, \widetilde{\boldsymbol{v}}^n\right) \geq \mathrm{IKG}^n\left(a^n, \boldsymbol{v}^n\right) - \varepsilon_n, \tag{15}$$

where $\{\varepsilon_n\}$ is a sequence of nonnegative real numbers such that $\varepsilon_n \to 0$ as $n \to \infty$. It is not difficult to see that such quasi-IKG policy is also consistent, as formally stated in the following Theorem 3, whose proof is collected in the Online Appendix.

Theorem 3 *If Assumptions 1–3 hold, then the quasi-IKG policy as defined in Equation (15) is consistent.*

4 | STOCHASTIC GRADIENT ASCENT

We now discuss computation of Equation (13) under Assumptions 1–3. It primarily consists of two steps.

(i) For each $i = 1, \dots, M$, solve $\max_{\boldsymbol{x}\in\mathcal{X}} \mathrm{IKG}^n(i, \boldsymbol{x})$ to find its maximizer, say $\boldsymbol{v}_i^n$.
(ii) Set $a^n = \mathrm{argmax}_{1\leq i\leq M} \mathrm{IKG}^n\left(i, \boldsymbol{v}_i^n\right)$ and set $\boldsymbol{v}^n = \boldsymbol{v}_{a^n}^n$.

Let $\boldsymbol{\xi}$ denote a $\mathcal{X}$-valued random variable with density $\gamma(\cdot)$, and

$$h_i^n(\boldsymbol{v}, \boldsymbol{x}) := \left|\widetilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})\right| \phi\left(\left|\frac{\Delta_i^n(\boldsymbol{v})}{\widetilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})}\right|\right) - \left|\Delta_i^n(\boldsymbol{v})\right| \Phi\left(-\left|\frac{\Delta_i^n(\boldsymbol{v})}{\widetilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})}\right|\right). \tag{16}$$

Then, we may rewrite Equation (14) as

$$\mathrm{IKG}^n(i, \boldsymbol{x}) = [c_i(\boldsymbol{x})]^{-1} \mathbb{E}\left[h_i^n(\boldsymbol{\xi}, \boldsymbol{x})\right], \tag{17}$$

which suggests the following sample average approximation,

$$\widehat{\mathrm{IKG}}^n(i, \boldsymbol{x}) = \frac{1}{c_i(\boldsymbol{x}) J} \sum_{j=1}^{J} h_i^n\left(\boldsymbol{\xi}_j, \boldsymbol{x}\right), \tag{18}$$

where $\boldsymbol{\xi}_j$'s are independent copies of $\boldsymbol{\xi}$ and J is the sample size. In particular, we will use Equation (18) in step (ii) above for computing a^n for given $\boldsymbol{v}_i^n$'s. However, the sample average approximation method can easily become computationally prohibitive when applied to solve $\max_{\boldsymbol{x}} \mathrm{IKG}^n(i, \boldsymbol{x})$ in step (i) if the domain $\mathcal{X}$ is high-dimensional. Hence, we consider instead the stochastic gradient ascent method to complete step (i).

Equation (17) means that in step (i) above, we solve the stochastic optimization problem

$$\boldsymbol{v}_i^n \in \operatorname*{argmax}_{\boldsymbol{x}\in\mathcal{X}} \ [c_i(\boldsymbol{x})]^{-1} \mathbb{E}\left[h_i^n(\boldsymbol{\xi}, \boldsymbol{x})\right],$$

for each $i = 1, \dots, M$. If $g_i^n(\boldsymbol{\xi}, \boldsymbol{x})$ is an unbiased estimator of $\frac{\partial}{\partial \boldsymbol{x}} \left\{[c_i(\boldsymbol{x})]^{-1} \mathbb{E}\left[h_i^n(\boldsymbol{\xi}, \boldsymbol{x})\right]\right\}$, then $\boldsymbol{v}_i^n$ can be computed approximately using the stochastic gradient ascent (SGA) method; see Kushner and Yin (2003) for a comprehensive treatment and Newton et al. (2018) for a recent survey on the subject. Given an initial solution $\boldsymbol{x}_1 \in \mathcal{X}$ and a maximum iteration limit K, SGA iteratively computes

$$\boldsymbol{x}_{k+1} = \Pi_{\mathcal{X}}\left[\boldsymbol{x}_k + b_k g_i^n\left(\boldsymbol{\xi}_k, \boldsymbol{x}_k\right)\right], \quad k = 1, \dots, K, \tag{19}$$

where $\Pi_{\mathcal{X}} : \mathbb{R}^d \mapsto \mathcal{X}$ denotes a projection mapping points outside $\mathcal{X}$ back to $\mathcal{X}$,[3] and b_k is referred as the step size that satisfies $\sum_{k=1}^\infty b_k = \infty$ and $\sum_{k=1}^\infty b_k^2 < \infty$. In general, the choice of b_k is crucial for the practical performance of SGA, and it is commonly set as $b_k = \alpha/k^\beta$ for some constants α and β.

Note that $g_i^n(\boldsymbol{\xi}, \boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}\left[h_i^n(\boldsymbol{\xi}, \boldsymbol{x})/c_i(\boldsymbol{x})\right]$ under mild regularity conditions (L'Ecuyer, 1995). The explicit forms of $g_i^n(\boldsymbol{\xi}, \boldsymbol{x})$ for several common covariance functions are collected in the Online Appendix. Besides, in the implementation of SGA algorithm, practical modifications such as mini batch and *Polyak-Ruppert averaging* (Polyak & Juditsky, 1992) can be adopted to achieve better performance. Detailed discussion is collected in the Online Appendix, together with other implementation issues of IKG policy.

5 | NUMERICAL EXPERIMENTS

In this section, we evaluate the performance of the IKG policy via numerical experiments due to two reasons. First, the theoretical analysis, albeit establishing the consistency of the IKG

[3]For example, one may set $\Pi_{\mathcal{X}}(\boldsymbol{x})$ to be the point in $\mathcal{X}$ closest to $\boldsymbol{x}$.

policy in a large-sample asymptotic regime, does not provide a guarantee on the finite-sample performance of the policy. Second, the analysis has implicitly assumed that the sampling decisions of the IKG policy in Equation (13) can be computed exactly, while in practice it needs to be solved numerically via methods such as SGA that we have proposed. Additional numerical experiments on other issues, including the computational cost comparison between SGA versus the sample average approximation and the effect of estimated $\lambda_i(\boldsymbol{x})$, are collected in the Online Appendix. All the numerical experiments are implemented in MATLAB and the source code is available at https://github.com/shenhaihui/ikg.

5.1 | Finite-sample performance

The numerical experiments are conducted on synthetic problems, with the number of alternatives $M = 5$ and the dimensionality $d = 1, 3, 5, 7$. For each $i = 1, \dots, M$, the true performance of alternative i is the revised Griewank function,

$$\theta_i(\boldsymbol{x}) = \sum_{j=1}^{d} \frac{x_j^2}{4000} - 1.5^{d-1} \prod_{j=1}^{d} \cos\left(\frac{x_j}{\sqrt{ij}}\right), \ \boldsymbol{x} \in \mathcal{X} = [0, 10]^d.$$

Further, we set sampling variance $\lambda_i(\boldsymbol{x}) \equiv 0.01$, and take prior $\mu_i^0(\boldsymbol{x}) = \mu^0(\boldsymbol{x}) \equiv 0$, and $k_i^0(\boldsymbol{x}, \boldsymbol{x}') = k^0(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{1}{d}\|\boldsymbol{x} - \boldsymbol{x}'\|^2\right)$. We set the cost function $c_i(\boldsymbol{x}) \equiv 1$ for each $i = 1, \dots, M$, but will investigate the impact of a different cost function later.

We consider two density functions for the covariates: (1) uniform distribution on $\mathcal{X}$: $\gamma(\boldsymbol{x}) = 1/|\mathcal{X}|$; (2) multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $4^2\boldsymbol{I}$ truncated on $\mathcal{X}$: $\gamma(\boldsymbol{x}) = \phi\left(\boldsymbol{x}; \mathbf{0}, 4^2\boldsymbol{I}\right) / \int_{\mathcal{X}} \phi\left(\boldsymbol{v}; \mathbf{0}, 4^2\boldsymbol{I}\right) \mathrm{d}\boldsymbol{v}$. For convenience, we call the above specifications Problem 1 (P1) and Problem 2 (P2), respectively, depending on the choice of $\gamma(\boldsymbol{x})$.

The parameters involved in the SGA algorithm (see details in the Online Appendix) are given as follows: $K = 100d$, $K_0 = K/4$, $b_k = 200d/k^{0.7}$, $m = 20d$, and $J = 500d^2$. Moreover, the algorithm is started with a random initial solution. The performance of the IKG policy with respect to the sampling budget B is evaluated via the *opportunity cost* (OC), that is, the integrated difference in performance between the best alternative and the alternative chosen by the IKG policy upon exhausting the sampling budget.

$$\mathrm{OC}(B) := \mathbb{E}\left[\int_{\mathcal{X}} \left(\theta_{i^*(\boldsymbol{x})}(\boldsymbol{x}) - \theta_{\widehat{i^*}(\boldsymbol{x};\omega)}(\boldsymbol{x})\right) \gamma(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}\right],$$

where $\widehat{i^*}(\boldsymbol{x}; \omega) \in \mathrm{argmax}_{1 \le i \le M} \mu_i^{N(B)}(\boldsymbol{x}; \omega)$ is the learned decision rule up to the budget B under the IKG policy, ω denotes the samples taken under the policy, and the expectation is with respect to ω. Clearly, $\mathrm{OC}(B) \to 0$ as $B \to \infty$, since the IKG policy is consistent. We estimate $\mathrm{OC}(B)$ via

$$\widehat{\mathrm{OC}}(B) = \frac{1}{L} \sum_{l=1}^{L} \left[\frac{1}{J'} \sum_{j=1}^{J'} \left(\theta_{i^*(\boldsymbol{x}_j)}\left(\boldsymbol{x}_j\right) - \theta_{\widehat{i^*}(\boldsymbol{x}_j;\omega_l)}\left(\boldsymbol{x}_j\right)\right)\right],$$

where $L = 30$ is the number of replications, ω_l denotes the samples for replication $l = 1, \dots, L$, and $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{J'}\}$ is a random sample of the covariates generated from a given density function $\gamma(\boldsymbol{x})$ with $J' = 1000d^2$ for the purpose of evaluation.

We compare the IKG policy against three other polices:

- *IKG with Random Covariates (IKGwRC).* Recall that in the computation of IKG policy, random solution is used to initiate the SGA algorithm. To check whether such random initialization is a main cause for the effectiveness of IKG, we consider the IKGwRC policy as follows. Let $\boldsymbol{x}_1^n, \dots, \boldsymbol{x}_M^n$ be the initial solutions for M alternatives used in the SGA algorithm when computing $\left(\widehat{a}^n, \widehat{\boldsymbol{v}}^n\right)$, $n = 0, 1, \dots$ Then the IKGwRC policy will sample at $(a^n, \boldsymbol{v}^n)$ given by

  $$a^n = \underset{1 \le i \le M}{\mathrm{argmax}} \log \widehat{\mathrm{IKG}}^n\left(i, \boldsymbol{x}_i^n\right) \ \text{ and } \ \boldsymbol{v}^n = \boldsymbol{x}_{a^n}^n,$$

  where the same samples are used to compute $\widehat{\mathrm{IKG}}^n$ as in the IKG policy.

- *Binned Successive Elimination (BSE).* The BSE policy is proposed by Perchet and Rigollet (2013) for solving nonparametric MAB problems with covariates. In their setting, values of the covariates arrive randomly, and the policy only determines which alternative to select. To implement BSE in our setting, we randomly generate $\boldsymbol{v}^n$ from uniform distribution on $\mathcal{X}$, and then apply the BSE policy to determine a^n. The BSE policy divides $\mathcal{X}$ into m^d parts, where m is the number of uniformly divided regions on each coordinate. For each problem, m is tuned within the set $\{1, \dots, 10\}$, while other parameters follow the suggestion in Perchet and Rigollet (2013).
- *Pure Random Search (PRS).* The PRS policy will sample at $(a^n, \boldsymbol{v}^n)$, where a^n is randomly generated from the uniform distribution on $\{1, \dots, M\}$ and $\boldsymbol{v}^n$ is generated from the uniform distribution on $\mathcal{X}$.

The performances of the four policies for problems P1 and P2 with $d = 1, 3, 5, 7$ are shown in Figures 1 and 2, respectively. Several findings are made as follows.

First, the estimated opportunity cost in all the test problems exhibits a clear trend of convergence to zero. This, from a practical point view, provides an assurance that the IKG policy in conjunction with the SGA algorithm indeed works as intended, that is, the uncertainty about the performances of the competing alternatives will vanish eventually as the sampling budget grows. Second, the IKG policy can quickly reduce the opportunity cost when the sampling budget is relatively small, but the reduction appears to slow down

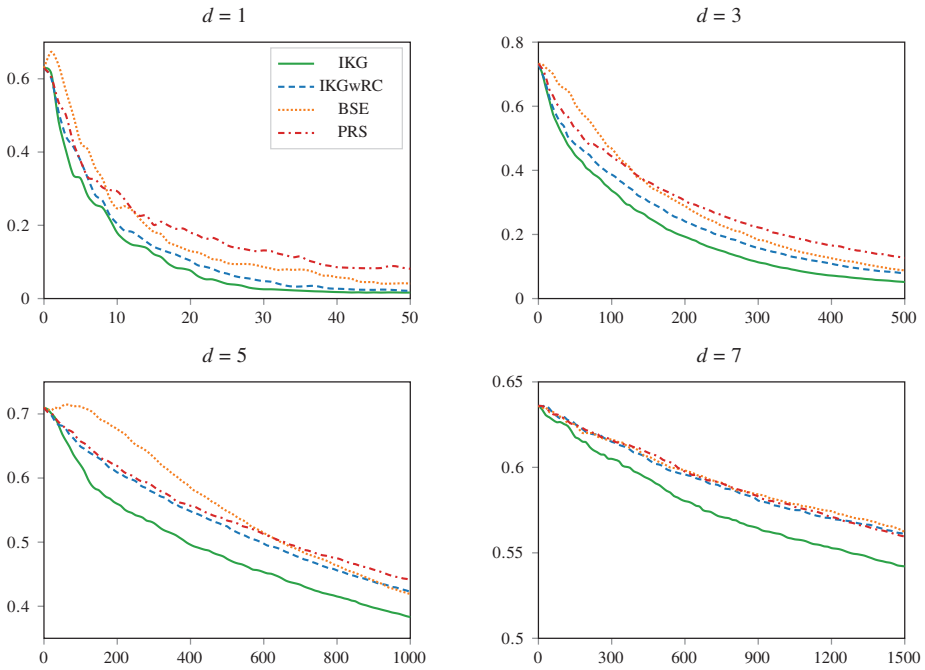

FIGURE 1 Estimated opportunity cost (vertical axis) as a function of the sampling budget (horizontal axis) for P1

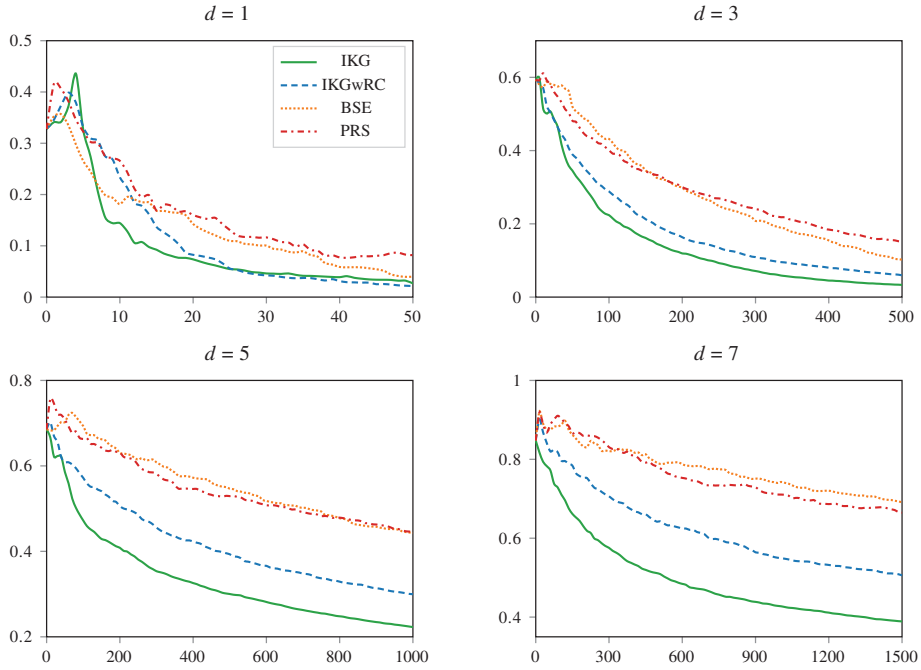

FIGURE 2 Estimated opportunity cost (vertical axis) as a function of the sampling budget (horizontal axis) for P2

as the sampling budget increases. This finding is consisted with prior research on other KG-type policies such as Frazier et al. (2009), Frazier and Powell (2011), and Xie et al. (2016). Third, the learning task of identifying the best alternative becomes substantially more difficult when the dimensionality of the covariates is large. This can be seen from the growing sampling budget and the slowing reduction in the opportunity cost as d increases.

Overall, IKG outperforms the other three policies. Specific comparisons are as follows. First, IKG has better performance than IKGwRC, especially when the dimensionality is high, which indicates that the SGA algorithm in IKG for solving $\boldsymbol{v}_i^n$ (see Section 4) indeed works well and has a significant effect in IKG. Second, BSE has inferior performance than IKG, which may be caused by the fact that BSE only optimizes a^n given randomly observed $\boldsymbol{v}^n$, while IKG optimizes both a^n and $\boldsymbol{v}^n$ at the same time. Third, PRS overall has the worst performance, which is not surprising since it does not utilize any information gained from previous sampling. Note that PRS is a consistent policy, but the consistency does not guarantee any finite-sample performance. This reflects the value of IKG—it is not only provably consistent, but also

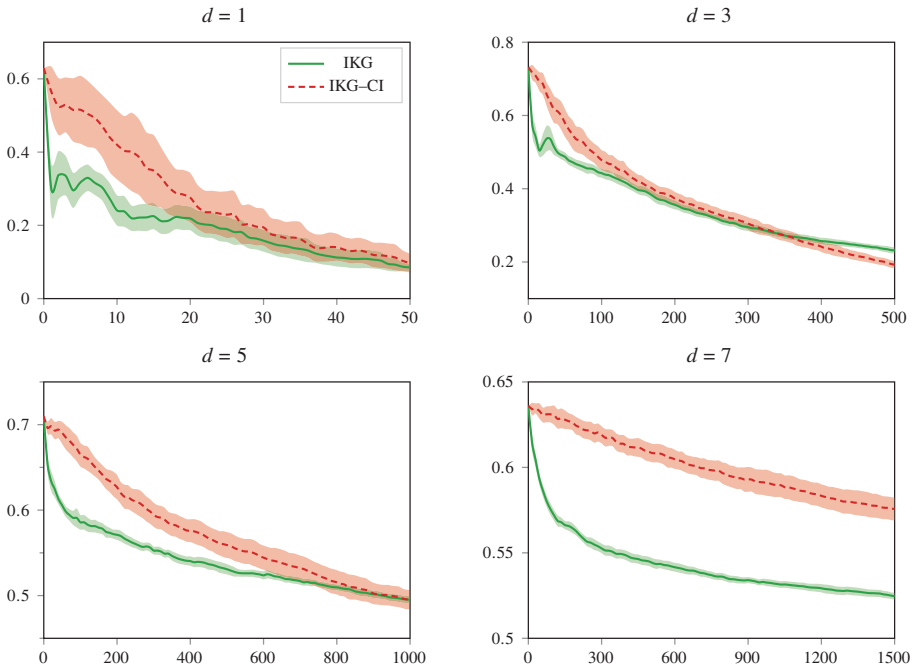

FIGURE 3 Estimated opportunity cost (vertical axis) as a function of the sampling budget (horizontal axis) for P3. IKG–CI means sampling costs are ignored when implementing the IKG policy. The shaded regions represent the 99% confidence intervals

takes advantage of information gained from previous samples to yield good finite-sample performance.

5.2 | Effect of sampling cost

We are also interested in the effect of sampling costs on the IKG policy. In particular, we consider a different cost function other than the unit cost function: $c_i(\boldsymbol{x}) = 2^{3-i}\left(1 + \|\boldsymbol{x} - \mathbf{5}\|^2/(10d)\right)$, where $\mathbf{5}$ is a $d \times 1$ vector of all fives. We set $\gamma(\boldsymbol{x})$ to be the uniform density[4] and call this specification Problem 3 (P3). We compare two scenarios: (i) the sampling cost is incorporated correctly; and (ii) one ignores variations in the sampling cost at different locations and mistakenly uses the unit sampling cost when implementing the IKG policy (but the actual sampling consumption follows $c_i(\boldsymbol{x})$). The comparison is illustrated in Figure 3.

There are two observations. On one hand, despite the misspecification in the sampling cost function, the IKG policy is still consistent, with the associated opportunity cost converging to zero. This is not surprising, because using the unit sampling cost function, that is, $c_i(\boldsymbol{x}) \equiv 1$, is exactly the setup of Theorem 2. On the other hand, however, the finite-sample performance of the IKG policy indeed deteriorates as a result of the misspecification. Further, the deterioration appears to become more significant as the dimensionality of the covariates increases.

6 | CONCLUSIONS

In this paper, we study sequential sampling for the problem of selection with covariates which aims to identify the best alternative as a function of the covariates. Each sampling decision involves choosing an alternative and a value of the covariates, from the pair of which a sample will be taken. We design a sequential sampling policy via a nonparametric Bayesian approach. In particular, following the well-known KG design principle for simulation optimization, we develop the IKG policy that attempts to maximize the "one-step" integrated increment in the expected value of information per unit of sampling cost.

We prove the consistency of the IKG policy under minimal assumptions. Compared to prior work on asymptotic analysis of KG-type sampling policies, our assumptions are simpler and significantly more general, thanks to technical machinery that we develop based on RKHS theory. Nevertheless, to compute the sampling decisions of the IKG policy requires solving a multidimensional stochastic optimization problem. To that end, we develop a numerical algorithm based on the SGA method. Numerical experiments illustrate the finite-sample performance of the IKG policy and provide a practical assurance that the developed methodology works as intended.

[4]Setting $\gamma(\boldsymbol{x})$ to be the truncated normal density leads to similar findings.

ACKNOWLEDGMENTS

The authors would like to thank the editor-in-chief, associate editor and reviewers for their insightful and detailed comments that have significantly improved this paper. This work was sponsored by the National Natural Science Foundation of China (Grants 72001140, 72091211, and 71991473), the "Chenguang Program" supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission (Grant 19CG14), and the

Hong Kong Research Grants Council (GRF 17201520 and 16211417).

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

L. Jeff Hong https://orcid.org/0000-0001-7011-4001
Haihui Shen https://orcid.org/0000-0002-4157-1278
Xiaowei Zhang https://orcid.org/0000-0002-5798-646X

REFERENCES

Adler, R. J., & Taylor, J. E. (2007). Random fields and geometry. Springer.

Ankenman, B., Nelson, B. L., & Staum, J. (2010). Stochastic kriging for simulation metamodeling. *Operations Research*, *58*(2), 371–382.

Arora, N., Dreze, X., Ghose, A., Hess, J. D., Iyengar, R., Jing, B., Joshi, Y., Kumar, V., Lurie, N., Neslin, S., Sajeesh, S., Su, M., Syam, N., Thomas, J., & Zhang, Z. J. (2008). Putting one-to-one marketing to work: Personalization, customization, and choice. *Marketing Letters*, *19*(3–4), 305–321.

Bect, J., Bachoc, F., & Discourager, D. (2019). A supermartingale approach to Gaussian process based sequential design of experiments. *Bernoulli*, *25*(4A), 2883–2919.

Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, *5*(1), 1–122.

Chen, C.-H., Chick, S. E., Lee, L. H., & Pujowidianto, N. A. (2015). *Ranking and selection: Efficient simulation budget allocation*. In M. C. Fu (Ed.), Handbook of simulation optimization (pp. 45–80). Springer.

Choi, S. E., Perzan, K. E., Tramontano, A. C., Kong, C. Y., & Hur, C. (2014). Statins and aspirin for chemoprevention in Barrett's esophagus: Results of a cost-effectiveness analysis. *Cancer Prevention Research*, *7*(3), 341–350.

Frazier, P., Powell, W., & Dayanik, S. (2009). The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, *21*(4), 599–613.

Frazier, P. I., Powell, W., & 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. (2011). Consistency of sequential Bayesian sampling policies. *SIAM Journal on Control and Optimization*, *49*(2), 712–731.

Hu, R., & Ludkovski, M. (2017). Sequential design for ranking response surfaces. *SIAM/ASA Journal on Uncertainty Quantification*, *5*(1), 212–239.

Hur, C., Nishioka, N. S., & Gazelle, G. S. (2004). Cost-effectiveness of aspirin chemoprevention for Barrett's esophagus. *Journal of the National Cancer Institute*, *96*(4), 316–325.

Kim, E. S., Herbst, R. S., Wistuba, I. I., Lee, J. J., Blumenschein, G. R., Jr., Tsao, A., Stewart, D. J., Hicks, M. E., Erasmus, J., Jr., Gupta, S., Alden, C. M., Liu, S., Tang, X., Khuri, F. R., Tran, H. T., Johnson, B. E., Heymach, J. V., Mao, L., Fossella, F., … Hong, W. K. (2011). The BATTLE trial: Personalizing therapy for lung cancer. *Cancer Discovery*, *1*(1), 44–53.

Kim, S.-H., & Nelson, B. L. (2006). *Selecting the best system*. In S. G. Henderson & B. L. Nelson (Eds.), Handbooks in operations research and management science (Vol. *13*, pp. 501–534). Elsevier.

Krause, A., & Ong, C. S. (2011). *Contextual Gaussian process bandit optimization*. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), Advances in neural information processing systems. (Vol. *24*, pp. 2447–2455). Curran Associates, Inc.

Kushner, H. J., & Yin, G. G. (2003). Stochastic approximation and recursive algorithms and applications. Springer-Verlag.

L'Ecuyer, P. (1995). Note: On the interchange of derivative and expectation for likelihood ratio derivative estimators. *Management Science*, *41*(4), 738–747.

Mes, M. R., Powell, W. B., & Frazier, P. I. (2011). Hierarchical knowledge gradient for sequential sampling. *Journal of Machine Learning Research*, *12*, 2931–2974.

Newton, D., Yousefian, F., & Pasupathy, R. (2018). *Stochastic gradient descent: Recent trends*. In E. Gel & D. Lewis (Eds.), TutORials in operations research (Vol. *7*, pp. 193–220) INFORMS. Institute for Operations Research and the Management Sciences (INFORMS).

Pearce, M., & Branke, J. (2017). *Efficient expected improvement estimation for continuous multiple ranking and selection*. In Chan W. K. V., D'Ambrogio A., Zacharewicz G, Mustafee N., Wainer G., Page E., Proceedings of the 2017 winter simulation conference (pp. 2161–2172). IEEE Press.

Pearce, M., & Branke, J. (2018). Continuous multi-task Bayesian optimisation with correlation. *European Journal of Operational Research*, *270*(3), 1074–1085.

Perchet, V., & Rigollet, P. (2013). The multi-armed bandit problem with covariates. *The Annals of Statistics*, *41*(2), 693–721.

Poloczek, M., Wang, J., & Frazier, P. I. (2017). *Multi-information source optimization*. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), Advances in neural information processing systems (Vol. *30*, pp. 4288–4298). Curran Associates, Inc.

Polyak, B. T., & Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, *30*(4), 838–855.

Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. MIT Press.

Rusmevichientong, P., & Tsitsiklis, J. N. (2010). Linearly parameterized bandits. *Mathematics of Operations Research*, *35*(2), 395–411.

Ryzhov, I. O. (2016). On the convergence rates of expected improvement methods. *Operations Research*, *64*(6), 1515–1528.

Scott, W., Frazier, P., & Powell, W. (2011). The correlated knowledge gradient for simulation optimization of continuous parameters using Gaussian process regression. *SIAM Journal on Optimization*, *21*(3), 996–1026.

Shen, H., Hong, L. J., & Zhang, X. (2021). Ranking and selection with covariates for personalized decision making. *INFORMS Journal on Computing* (Forthcoming). https://doi.org/10.1287/ijoc.2020.1009.

Steinwart, I., & Christmann, A. (2008). Support vector machines. Springer.

Toscano-Palmerin, S., & Frazier, P. I. (2018). *Bayesian optimization with expensive integrands*. arXiv:1803.08661.

Wu, J., & Frazier, P. I. (2016). *The parallel knowledge gradient method for batch Bayesian optimization*. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), Advances in neural information processing systems 29 (pp. 3126–3134). Curran Associates, Inc.

Wu, J., Poloczek, M., Wilson, A. G., & Frazier, P. I. (2017). *Bayesian optimization with gradients*. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), Advances in neural information processing systems 30 (pp. 5267–5278). Curran Associates, Inc.

Xie, J., Frazier, P. I., & Chick, S. E. (2016). Bayesian optimization via simulation with pairwise sampling and correlated prior beliefs. *Operations Research*, *64*(2), 542–559.

Yang, Y., & Zhu, D. (2002). Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates. *The Annals of Statistics*, *30*(1), 100–121.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ding, L., Hong, L. J., Shen, H., & Zhang, X. (2022). Technical note—Knowledge gradient for selection with covariates: Consistency and computation. *Naval Research Logistics (NRL)*, *69*(3), 496–507. https://doi.org/10.1002/nav.22028