Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

RATE ANALYSIS FOR OFFLINE SIMULATION ONLINE APPLICATION

Siyang Gao

Systems Engineering and Engineering Management
and School of Data Science
City University of Hong Kong
Kowloon, Hong Kong

Cheng Li

Statistics and Applied Probability
National University of Singapore
Singapore

Jianzhong Du

Department of Systems Engineering and Engineering Management
City University of Hong Kong
Kowloon, Hong Kong

ABSTRACT

We consider a recently proposed simulation-based decision-making framework, called offline-simulation-online-application (OSOA). In this framework, simulation experiments are not performed after the target problem is set up with all the input parameters; instead, they are performed before that with all the possible parameters that might come up in the target problem. Then, these computational results are directly used in real time for system evaluation or system optimization when the input parameters of the target problem are known. In this paper, we follow this framework and use stochastic kriging (SK) to model the system performance from the covariate space. Two measures, namely IMSE and IPFS, are proposed to evaluate the prediction errors for system evaluation and system optimization respectively. We establish the convergence rates of these two measures. They quantify the magnitude of prediction errors for the online application after a certain period of time is spent for the offline simulation.

1 INTRODUCTION

Large-scale complex systems broadly arise in engineering applications, e.g., in queueing systems, electric power grids, air and land traffic control systems, communication networks, manufacturing plants, supply chains, etc. These systems and problems usually do not satisfy the assumptions of analytical models, and the practitioners have to rely on the tool of stochastic simulation to evaluate the performance of them. In the meantime, since the evaluation of the system performance is accomplished by running simulation models rather than experimenting with the real systems, it provides the opportunity for practitioners to try different parameter settings (designs) for the systems and choose the best one in order to optimize their performance. In this paper, we will call these two purposes of simulation as system evaluation (SE) and system optimization (SO).

A major issue when using simulation is the computational efficiency. It is well known that simulation experiments can take a long time to produce accurate estimates and decisions for the problems under consideration (Law 2015; Ho et al. 2007). In contrast, a lot of practical problems are online, which require the decisions to be made in real time after the input parameters are revealed. For example, when there is a change in the arrival rate of patients to the hospital, the hospital manager wants to know the corresponding

978-1-7281-3283-9/19/$31.00 ©2019 IEEE

optimal medical staff configuration immediately without spending much time running simulation models and comparing different configuration designs. The doctor wants to decide the best medical treatment for a new patient without simulating different treatment methods.

To extend the applicability of simulation to online problems, Shen, Hong, and Zhang (2017) recently proposed a new framework for conducting simulation experiments. This framework is based on the observation that the possible values of the input parameters to the simulation model are typically known in advance through historical data or experience. Then, instead of starting the simulation after the input parameters are known, we can run simulation experiments in advance on the possible values of the input parameters and build predictive models that relate the system performance and the input parameters. Then, when the values of the parameters are revealed for online problems, the predictive models can be utilized to make real time decisions, eliminating the need to run simulation at this stage.

This framework is called offline-simulation-online-application (OSOA). The input parameters in OSOA have been known as the covariates, side information or auxiliary information in the literature (Perchet and Rigollet 2013; Bastani and Bayati 2016; Qiang and Bayati 2017; Farias and Li 2019). Note that the general idea of OSOA is, to some degree, similar to the concept of green simulation recently proposed, which also tries to improve the efficiency of evaluating the system performance at a new input by using historical simulation samples (Feng and Staum 2017). The main tool for the green simulation to achieve this goal is by change of probability measures.

Although OSOA is highly effective in reducing the decision time for online problems through simulation, the literature along this direction is sparse, rendering a lot of key questions under this framework largely unexplored. Shen et al. (2017) considered ranking and selection (R&S) in the presence of covariates. They used linear regression to characterize the relationship between the system performance of a certain design and the covariate values, and applied the indifference-zone method (Kim and Nelson 2001) to determine a budget allocation strategy on a set of covariate values and all the designs in order to provide a guarantee on the expected probability of correct selection (PCS) or the minimal PCS on the covariate space. Li et al. (2018) extended the result in Shen et al. (2017) to high-dimensional covariates and general dependence between the mean performance of a design and the covariates. Pearce and Branke (2017) and Pearce and Branke (2018) looked at estimating a function of covariates and decision variables, in order to provide an optimal decision in response to a set of covariate values that might come in the future, under the settings of R&S and continuous Bayesian optimization, respectively.

In this paper, we summarize recent results by Li et al. (2019) and aim to quantify the relationship between the offline simulation efforts and online prediction accuracy for both SE and SO. This is a fundamental problem in OSOA. Particularly, we consider continuous covariate space and a finite number of designs. We assume a fixed sampling distribution over the covariate space and build a stochastic kriging (SK) predictive model for the system performance of each design. The prediction error of SE is evaluated by the maximal integrated mean squared error (IMSE), where the maximum is taken on the set of designs and integration is over the covariate space. The maximal IMSE corresponds to the largest IMSE among the designs and is therefore an appropriate measure for assessing the worst-case SE error. The prediction error of SO is evaluated by the integrated probability of false selection (IPFS), where the integration is, similarly, over the covariate space and the selection for the best design is based on the predictions of the SK models. Note that PFS is equivalent to the probability of correct selection (PCS) commonly used in simulation optimization. We will investigate the convergence rates of the maximal IMSE and IPFS with the number of covariate points m ever collected in offline simulation, and our theoretical development indicates that both the maximal IMSE and IPFS converge at least at the rate of $\log^{\frac{1}{\kappa}} m/m$, where κ is a parameter associated with the covariance kernels. With additional assumptions on the MSE, IPFS enjoys an exponential convergence rate. It shows for practitioners the magnitude of prediction errors for online application after a certain period of time is spent for offline simulation.

This paper is related to two streams of literature. The first is kriging for building metamodels (Stein 2012; Kleijnen 2009), which has proven to be an effective tool for global metamodeling. Later (see, for

example, Opsomer et al. (1999), Kersting et al. (2007), Ankenman et al. (2010)), kriging was extended to allow heteroscedastic stochastic observation noise, known as the stochastic kriging. Chen et al. (2013) and Qu and Fu (2014) improved the SK model by incorporating the gradient information. The second stream of literature is ranking and selection. It seeks to determine an allocation of the simulation budget either to maximize the probability of correctly selecting the best design or guarantee this probability at a high-enough level. Representative methods for R&S include the optimal computing budget allocation (OCBA) (Chen et al. 2000; Fu et al. 2007; Gao and Chen 2016; Gao et al. 2017; Gao and Chen 2017), indifference-zone (IZ) method (Dudewicz and Dalal 1975; Kim and Nelson 2001; Nelson et al. 2001) and value of information procedure (VIP) (Frazier et al. 2008; Chick et al. 2010; Ryzhov 2016).

2 PROBLEM STATEMENT

Suppose we have a finite number of system designs and a continuous covariate space. We introduce the following notation regarding the designs, covariates and simulation samples:

k: total number of designs;
d: dimension of the covariate space;
m: number of randomly-drawn covariate (design) points;
$\mathbf{X} = (X_1, \ldots, X_d)^\top$: the vector of random covariates;
$\mathscr{X} \subseteq \mathbb{R}^d$: support of covariates;
$y_i(\mathbf{x})$: mean of design i under covariate $\mathbf{x}$, $i \in \{1,2,...,k\}$ and $\mathbf{x} \in \mathscr{X}$;
$Y_{il}(\mathbf{x}) = y_i(\mathbf{x}) + \varepsilon_{il}(\mathbf{x})$: l-th simulation sample from design i under covariate $\mathbf{x}$, where $\varepsilon_{il}(\mathbf{x})$ is a mean-zero stochastic error;
n_j: number of simulation replications at covariate $\mathbf{x}_j$ for each of the k designs, $j \in \{1,2,...,m\}$;
$\overline{Y}_i(\mathbf{x}_j) = n_j^{-1} \sum_{l=1}^{n_j} Y_{il}(\mathbf{x}_j)$: sample mean of design i under covariate $\mathbf{x}_j$, $i \in \{1,2,...,k\}$ and $j \in \{1,2,...,m\}$;
$\mathbb{P}_{\mathbf{x}}$: sampling distribution for $\mathbf{X}$.

In this research, we use SK to model $y_i(\mathbf{x})$:

$$y_i(\mathbf{x}) = \mathbf{f}_i(\mathbf{x})^\top \beta_i + M_i(\mathbf{x}), \quad i = 1, \ldots, k,$$

where $\mathbf{f}_i(\mathbf{x}) = (\mathrm{f}_{i1}(\mathbf{x}), \ldots, \mathrm{f}_{iq}(\mathbf{x}))^\top$ and $\beta_i = (\beta_{i1}(\mathbf{x}), \ldots, \beta_{iq}(\mathbf{x}))^\top$ are a $q \times 1$ vector of known functions of $\mathbf{x}$ and a $q \times 1$ vector of unknown parameters. $M_i(\mathbf{x})$ is a realization of a mean zero stationary Gaussian process. We further introduce the following notation regarding the SK model:

$\Sigma_{M,i}(\mathbf{x}, \mathbf{x}') = \mathrm{Cov}[M_i(\mathbf{x}), M_i(\mathbf{x}')]$: covariance between $M_i(\mathbf{x})$ and $M_i(\mathbf{x}')$ for any $\mathbf{x}, \mathbf{x}' \in \mathscr{X}$;
$\mathbf{X}_m = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$: m randomly drawn covariate points;
$\overline{\varepsilon}_i(\mathbf{x}_j) = n_j^{-1} \sum_{l=1}^{n_j} \varepsilon_{il}(\mathbf{x}_j)$: averaged simulation error, $i \in \{1,2,...,k\}$ and $j \in \{1,2,...,m\}$;
$\mathbf{Y}_{ij} = (Y_{i1}(\mathbf{x}_j), \ldots, Y_{in_j}(\mathbf{x}_j))^\top$: an $n_j \times 1$ vector of simulation samples;
$\overline{\mathbf{Y}}_i = \left(\overline{Y}_i(\mathbf{x}_1), \ldots, \overline{Y}_i(\mathbf{x}_m)\right)^\top$: an $m \times 1$ vector of sample means;
$\mathbf{F}_i = (\mathbf{f}_i(\mathbf{x}_1)^\top, \ldots, \mathbf{f}_i(\mathbf{x}_m)^\top)^\top$: the $q \times m$ design matrix;
$\Sigma_{M,i}(\mathbf{X}_m, \mathbf{X}_m)$: the $m \times m$ covariance matrix across all covariate points $\mathbf{x}_1, \ldots, \mathbf{x}_m$, i.e. for $s = 1, \ldots, m$ and $t = 1, \ldots, m$, the (s,t) entry of $\Sigma_{M,i}(\mathbf{X}_m, \mathbf{X}_m)$ is $[\Sigma_{M,i}(\mathbf{X}_m, \mathbf{X}_m)]_{st} = \mathrm{Cov}[y_i(\mathbf{x}_s), y_i(\mathbf{x}_t)]$;
$\Sigma_{M,i}(\mathbf{X}_m, \mathbf{x}^*) = (\mathrm{Cov}[y_i(\mathbf{x}^*), y_i(\mathbf{x}_1)], \ldots, \mathrm{Cov}[y_i(\mathbf{x}^*), y_i(\mathbf{x}_m)])^\top$: an $m \times 1$ vector of covariances;
$\Sigma_{\varepsilon,i}$: the $m \times m$ covariance matrix of the averaged simulation errors, i.e. for $s = 1, \ldots, m$ and $t = 1, \ldots, m$, the (s,t) entry of $\Sigma_{\varepsilon,i}$ is $(\Sigma_{\varepsilon,i})_{st} = \mathrm{Cov}[\overline{\varepsilon}_i(\mathbf{x}_s), \overline{\varepsilon}_i(\mathbf{x}_t)]$;
$\Sigma_{y,i} = \Sigma_{M,i}(\mathbf{X}_m, \mathbf{X}_m) + \Sigma_{\varepsilon,i}$.

According to Stein (2012) and Ankenman et al. (2010), the MSE-optimal linear predictor of SK at a new point $\mathbf{x}^* \in \mathscr{X}$ is:

$$\widehat{y}_i(\mathbf{x}^*) = \mathbf{f}_i(\mathbf{x}^*)^\top \widehat{\beta}_i + \Sigma_{M,i}(\mathbf{X}_m, \mathbf{x}^*)^\top \Sigma_{y,i}^{-1} \left(\overline{\mathbf{Y}}_i - \mathbf{F}_i \widehat{\beta}_i \right), \tag{1}$$

where $\widehat{\beta}_i = \left(\mathbf{F}_i^\top \Sigma_{y,i}^{-1} \mathbf{F}_i \right)^{-1} \mathbf{F}_i^\top \Sigma_{y,i}^{-1} \overline{\mathbf{Y}}_i$. The optimal MSE from (1) at $\mathbf{x}^* \in \mathscr{X}$ is

$$\begin{aligned} \mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*) = {} & \Sigma_{M,i}(\mathbf{x}^*, \mathbf{x}^*) - \Sigma_{M,i}^\top(\mathbf{X}_m, \mathbf{x}^*) \left[\Sigma_{M,i}(\mathbf{X}_m, \mathbf{X}_m) + \Sigma_{\varepsilon,i} \right]^{-1} \Sigma_{M,i}(\mathbf{X}_m, \mathbf{x}^*) \\ & + \eta_i(\mathbf{x}^*)^\top \left[\mathbf{F}_i^\top \left(\Sigma_{M,i}(\mathbf{X}_m, \mathbf{X}_m) + \Sigma_{\varepsilon,i} \right)^{-1} \mathbf{F}_i \right]^{-1} \eta_i(\mathbf{x}^*), \end{aligned} \tag{2}$$

where $\eta_i(\mathbf{x}^*) = \mathbf{f}_i(\mathbf{x}^*) - \mathbf{F}_i^\top \left(\Sigma_{M,i}(\mathbf{X}_m, \mathbf{X}_m) + \Sigma_{\varepsilon,i} \right)^{-1} \Sigma_{M,i}(\mathbf{X}_m, \mathbf{x}^*)$.

For the sampling distribution $\mathbb{P}_\mathbf{x}$ over $\mathscr{X}$ and design i, we let $\left\{ \phi_l^i(\mathbf{x}) : l = 1, 2, \ldots \right\}$ be an orthonormal basis with respect to $\mathbb{P}_\mathbf{x}$, with $\int_\mathscr{X} (\phi_l^i)^2(\mathbf{x}) d\mathbb{P}_\mathbf{x}(\mathbf{x}) = 1$ and $\int_\mathscr{X} \phi_l^i(\mathbf{x}) \phi_{l'}^i(\mathbf{x}) d\mathbb{P}_\mathbf{x}(\mathbf{x}) = 0$ for $l \neq l'$. According to Mercer's theorem, we can assume that the kernel $\Sigma_{M,i}$ has the series expansion $\Sigma_{M,i}(\mathbf{x}, \mathbf{x}') = \sum_{l=1}^\infty \mu_l^i \phi_l^i(\mathbf{x}) \phi_l^i(\mathbf{x}')$ with respect to $\mathbb{P}_\mathbf{x}$ for any $\mathbf{x}, \mathbf{x}' \in \mathscr{X}$, where $\mu_1^i \geq \mu_2^i \geq \ldots \geq 0$ are the eigenvalues of $\Sigma_{M,i}$. The trace of the kernel $\Sigma_{M,i}$ is defined as $\mathrm{tr}(\Sigma_{M,i}) = \sum_{l=1}^\infty \mu_l^i$. We let $L_2(\mathbb{P}_\mathbf{x})$ be the L_2 space under $\mathbb{P}_\mathbf{x}$. For any two generic functions $f_1, f_2 \in L_2(\mathbb{P}_\mathbf{x})$, their L_2-inner product is defined to be $\langle f_1, f_2 \rangle_{L_2(\mathbb{P}_\mathbf{x})} = \mathrm{E}_{\mathbb{P}_\mathbf{x}}[f_1(\mathbf{x}) f_2(\mathbf{x})]$. Any function $f^i \in L_2(\mathbb{P}_\mathbf{x})$ has the series expansion $f^i(\mathbf{x}) = \sum_{l=1}^\infty \theta_l^i \phi_l^i(\mathbf{x})$, where $\theta_l^i = \langle f^i, \phi_l^i \rangle_{L_2(\mathbb{P}_\mathbf{x})}$. The L_2 norm of f^i is given by $\|f^i\|_2^2 = \sum_{l=1}^\infty (\theta_l^i)^2$. The reproducing kernel Hilbert space (RKHS) $\mathbb{H}$ attached to $\Sigma_{M,i}$ is the space of all functions $f^i \in L_2(\mathbb{P}_\mathbf{x})$ such that its $\mathbb{H}$-norm $\|f^i\|_\mathbb{H}^2 = \sum_{l=1}^\infty (\theta_l^i)^2 / \mu_l^i < \infty$.

To estimate the SK model, we need the covariance kernels $\Sigma_{M,i}(\mathbf{x}, \mathbf{x}')$ to have more structures, and in this paper, we focus on the exponentially decaying kernels (Rasmussen and Williams 2006), a broadly-used class of kernels for kriging modeling. Exponentially decaying kernels satisfy $\mu_l^i \asymp \exp(-c_1^i l^{\kappa^i})$ for some constants $c_1^i > 0$ and $\kappa^i > 0$, where $a_l \asymp b_l$ means that for some $c_3 > 0$, $\lim_{l \to \infty} a_l / b_l = c_3$. A representative example of exponentially decaying kernels is the squared exponential kernel $\Sigma_{M,i}(x, x') = \exp\left\{ -\varphi^i (x - x')^2 \right\}$ for some $\varphi^i > 0$ and $\mathscr{X} = \mathbb{R}$. If the sampling distribution for $\mathbf{X}$ is $\mathbb{P}_\mathbf{x} = \mathscr{N}(0, \tau^2)$, then it has been shown in Rasmussen and Williams (2006) that for $l = 1, 2, \ldots$, the eigenvalues $\mu_l^i = \sqrt{2a^i / A^i} (B^i)^l = \sqrt{2a^i / A^i} \exp\{ -l \log(1/B^i) \}$, where $a^i = 1/(4(\tau^i)^2)$, $b^i = \varphi^i$, $c^i = \sqrt{(a^i)^2 + 2a^i b^i}$, $A^i = a^i + b^i + c^i$ and $B^i = b^i / A^i \in (0, 1)$. To make the theoretical analysis more tractable, in this paper, we fix $n_1 = \ldots = n_m = n$, i.e., each pair of design and covariate value will receive the same number of n replications.

For the purpose of SE, we will study the following measure

$$\mathscr{M} = \max_{i \in \{1, 2, \ldots, k\}} \mathrm{E}_{\mathbf{x}^*} \left[\mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*) \right],$$

which corresponds to the largest IMSE of all the designs. $\mathbf{x}^* \in \mathscr{X}$ is a random test point.

For the purpose of SO, we first define the best design and its mean performance, and the estimated best design below:

$$\begin{aligned} i^\circ(\mathbf{x}) &\in \arg \min_{i \in \{1, 2, \ldots, k\}} y_i(\mathbf{x}), \\ y^\circ(\mathbf{x}) &= \min_{i \in \{1, 2, \ldots, k\}} y_i(\mathbf{x}), \\ \widehat{i^\circ}(\mathbf{x}) &\in \arg \min_{i \in \{1, 2, \ldots, k\}} \widehat{y}_i(\mathbf{x}). \end{aligned}$$

Note that for a continuous covariate space, the optimal design might not be unique. To define a false selection, we adopt the formulation of IZ and introduce the IZ parameter $\Delta_0 > 0$. A false selection happens

when the mean of the estimated best design $y_{\widehat{i}^\circ(\mathbf{x}^*)}$ is no better than $y^\circ(\mathbf{x}^*)+\Delta_0$ for the test point $\mathbf{x}^* \in \mathscr{X}$. This definition allows some flexibility for determining the best design when the means of the top two designs are very close or exactly the same under some covariate value. Then the probability of false selection is given by

$$\mathrm{PFS}(\mathbf{x}^*) = \mathbb{P}_\varepsilon\left(y_{\widehat{i}^\circ(\mathbf{x}^*)}(\mathbf{x}^*) - y^\circ(\mathbf{x}^*) \geq \Delta_0\right).$$

This definition for the probability of false/correct selection has also been called probability of bad/good selection in some literature (Eckman and Henderson 2018). In this research, we will study the measure of IPFS

$$\mathscr{P} = \mathrm{E}_{\mathbf{x}^*}\mathrm{E}_M\left[\mathrm{PFS}(\mathbf{x}^*)\right],$$

where E_M arises from the randomness of the joint Gaussian measure on M_i, $i = 1,2,...,k$.

Particularly, we will investigate the convergence rates of $\mathscr{M}$ and $\mathscr{P}$ with respect to the number of covariate points m ever sampled, which provides a general characterization about the level of accuracy for SE and SO-related online predictions after a certain period of offline simulation is conducted. To do so, we introduce the rate functions

$$R_i^{\mathrm{IMSE}}(m,n) = \max\left\{\frac{\log^{\frac{1}{\kappa^i}}(mn)}{mn}, \frac{\log^{\frac{r(\kappa^i+1)}{\kappa^i}}(mn)}{m^{\frac{r}{2}}}\right\};$$
$$\overline{R}^{\mathrm{IMSE}}(m,n) = \max_{i\in\{1,2,\ldots,k\}} R_i^{\mathrm{IMSE}}(m,n),$$

which play a key role in characterizing the convergence rates of $\mathscr{M}$ and $\mathscr{P}$. κ^i is a kernel parameter and r will be explained in Assumption 2 below.

To enable the theoretical analysis, we make the following technical assumptions in this paper:

Assumption 1 Stochastic noise $\varepsilon_{il}(\mathbf{x}_j)$'s are distributed with mean 0 and variance $\sigma_i^2(\mathbf{x}_j)$. They are independent across different i, j, and l, as well as independent of the Gaussian process $M_i(\mathbf{x})$ for all i, j, l and $\mathbf{x} \in \mathscr{X}$. In addition, there exist upper and lower bounds $\overline{\sigma}_0^2$ and $\underline{\sigma}_0^2$ of $\sigma_i^2(\mathbf{x}_j)$ for all i and $\mathbf{x} \in \mathscr{X}$.

Assumption 2 There exist positive constants ρ^i and $r > 2$, such that $\mathrm{E}_{\mathbb{P}_\mathbf{x}}\{(\phi_l^i)^{2r}(\mathbf{x})\} \leq (\rho^i)^{2r}$ for every $l = 1,2,\ldots,\infty$.

Assumption 3 $\mathrm{f}_{is} \in \mathbb{H}$ for each $i = 1,2,\ldots,k$ and $s = 1,\ldots,q$ and $\lambda_{\min}\left(\mathrm{E}_\mathbf{x}[\mathbf{f}_i(\mathbf{x})\mathbf{f}_i(\mathbf{x})^\top]\right) > 0$ if $\mathbf{x}$ follows the distribution $\mathbb{P}_\mathbf{x}$, where $\mathbb{H}$ is the reproducing kernel Hilbert space (RKHS) and $\lambda_{\min}(A)$ is the smallest eigenvalue of positive definite matrix A.

Assumption 4 The trace of $\Sigma_{M,i}$ is finite, i.e., $\mathrm{tr}\,(\Sigma_{M,i}) < \infty$.

Assumption 5 $\varepsilon_{il}(\mathbf{x})$ is normally distributed with $N(0,\sigma_i^2(\mathbf{x}))$ for all i, l and $\mathbf{x}$.

Assumption 6 For any given $\xi \in (0,0.5)$, there exist constants $d_1, d_2 > 0$ and $m_0 \geq 1$ dependent on ξ, such that when $m \geq m_0$, for any $t > 0$,

$$\mathbb{P}_{\mathbf{X}_m}\left\{\mathbb{P}_{\mathbf{x}^*}\left(\frac{\max_{i\in\{1,\ldots,k\}}\mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*)}{\overline{R}^{\mathrm{IMSE}}(m,n)} \geq t\right) \leq d_1\exp\left(-d_2 t\right)\right\} \geq 1-\xi.$$

Assumption 7 For any given $\xi \in (0,0.5)$, there exist constants $d_3 > 0$ and $m_0 \geq 1$ dependent on ξ, such that when $m \geq m_0$, for any $t > 0$,

$$\mathbb{P}_{\mathbf{X}_m}\left\{\frac{\max_{i\in\{1,\ldots,k\}}\sup_{\mathbf{x}^*\in\mathscr{X}}\mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*)}{\overline{R}^{\mathrm{IMSE}}(m,n)} \leq d_3\right\} \geq 1-\xi.$$

3 MAIN RESULTS

Our main results about the convergence rates of $\mathscr{M}$ and $\mathscr{P}$ can be established through a set of lemmas and theorems. We observe that $\mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*)$ in (2) can be decomposed into two parts

$$\mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*) = \mathrm{MSE}^{(M)}_{i,\mathrm{opt}}(\mathbf{x}^*) + \mathrm{MSE}^{(\beta)}_{i,\mathrm{opt}}(\mathbf{x}^*),$$

where

$$\begin{aligned}
\mathrm{MSE}^{(M)}_{i,\mathrm{opt}}(\mathbf{x}^*) &= \Sigma_{M,i}(\mathbf{x}^*,\mathbf{x}^*) - \Sigma^{\top}_{M,i}(\mathbf{X}_m,\mathbf{x}^*)\left[\Sigma_{M,i}(\mathbf{X}_m,\mathbf{X}_m)+\Sigma_{\varepsilon,i}\right]^{-1}\Sigma_{M,i}(\mathbf{X}_m,\mathbf{x}^*),\\
\mathrm{MSE}^{(\beta)}_{i,\mathrm{opt}}(\mathbf{x}^*) &= \eta_i(\mathbf{x}^*)^{\top}\left[\mathbf{F}_i^{\top}\left(\Sigma_{M,i}(\mathbf{X}_m,\mathbf{X}_m)+\Sigma_{\varepsilon,i}\right)^{-1}\mathbf{F}_i\right]^{-1}\eta_i(\mathbf{x}^*),
\end{aligned}$$

which arise from estimating $M_i(\mathbf{x})$ and β_i respectively.

The following lemma bounds $\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}^{(M)}_{i,\mathrm{opt}}(\mathbf{x}^*)\right]$ and $\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}^{(\beta)}_{i,\mathrm{opt}}(\mathbf{x}^*)\right]$ for the SK model associated with a fixed design i. It serves as the foundation for our further development on $\mathscr{M}$ and $\mathscr{P}$.

Lemma 1 (Li et al. 2019) Suppose all the k designs have the sampling distribution $\mathbb{P}_{\mathbf{x}}$ for $\mathbf{X}_m$ and $\mathbf{x}^*$. Under Assumptions 1-4, for any $\delta_1,\delta_2,\delta_3 \in (0,0.5)$, the following inequalities hold

$$\begin{aligned}
\mathrm{E}_{\mathbf{X}_m}\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}^{(M)}_{i,\mathrm{opt}}(\mathbf{x}^*)\right] \le{}& \frac{1+\delta_2}{1-\delta_1}\frac{\overline{\sigma}_0^2}{mn}\gamma_i\left(\frac{\overline{\sigma}_0^2}{mn}\right)\\
&+\inf_{d\in\mathbb{N}}\left[\left\{\frac{mn}{\overline{\sigma}_0^2\delta_2}\mathrm{tr}(\Sigma_{M,i})+1\right\}\mathrm{tr}\left(\Sigma^{(d)}_{M,i}\right)+\mathrm{tr}(\Sigma_{M,i})\left\{88(\rho^i)^2\delta_1\frac{b(m,d,r)\gamma_i(\frac{\overline{\sigma}_0^2}{mn})}{\sqrt{m}}\right\}^r\right],\\
\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}^{(\beta)}_{i,\mathrm{opt}}(\mathbf{x}^*)\right] \lesssim_{\mathbb{P}_{\mathbf{X}_m}}{}& \frac{8q\,\mathrm{tr}(\Sigma_{M,i})}{\lambda_{\min}\left(\mathrm{E}_{\mathbf{x}}[\mathbf{f}_i(\mathbf{x})\mathbf{f}_i(\mathbf{x})^{\top}]\right)}\left\{\frac{2C^2_{\mathrm{f},i}}{(1-\delta_3)^2}\frac{\overline{\sigma}_0^2}{mn}\right.\\
&+\inf_{d\in\mathbb{N}}\left[\frac{2C^2_{\mathrm{f},i}}{(1-\delta_3)^2}\frac{mn\overline{\sigma}_0^2}{\underline{\sigma}_0^4}(\rho^i)^4\,\mathrm{tr}\left(\Sigma_{M,i}\right)\mathrm{tr}\left(\Sigma^{(d)}_{M,i}\right)+C^2_{\mathrm{f},i}\,\mathrm{tr}\left(\Sigma^{(d)}_{M,i}\right)\right.\\
&\left.\left.+C^2_{\mathrm{f},i}\,\mathrm{tr}\left(\Sigma_{M,i}\right)\left\{88(\rho^i)^2\delta_3\frac{b(m,d,r)\gamma_i(\frac{\overline{\sigma}_0^2}{mn})}{\sqrt{m}}\right\}^r\right]\right\},
\end{aligned}$$

where

$$\begin{aligned}
&b(m,d,r) = \max\left(\sqrt{\max(r,\log d)},\frac{\max(r,\log d)}{m^{1/2-1/r}}\right),\\
&\gamma_i(a) = \sum_{l=1}^{\infty}\frac{\mu_l^i}{\mu_l^i+a}\ \text{for any } a>0,\quad \mathrm{tr}\left(\Sigma^{(d)}_{M,i}\right) = \sum_{l=d+1}^{\infty}\mu_l^i,\\
&C_{\mathrm{f},i} = \max_{1\le s\le q}\|\mathrm{f}_{is}\|_{\mathbb{H}},
\end{aligned}$$

and for random variables U_m and V_m, $U_m \lesssim_{\mathbb{P}_{\mathbf{X}_m}} V_m$ means $|U_m/V_m|$ is bounded in $\mathbb{P}_{\mathbf{X}_m}-$ probability.

In Lemma 1, $\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}^{(M)}_{i,\mathrm{opt}}(\mathbf{x}^*)\right]$ and $\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}^{(\beta)}_{i,\mathrm{opt}}(\mathbf{x}^*)\right]$ are bounded in different ways. The upper bound of $\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}^{(\beta)}_{i,\mathrm{opt}}(\mathbf{x}^*)\right]$ is a bound in probability, which is weaker than that of $\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}^{(M)}_{i,\mathrm{opt}}(\mathbf{x}^*)\right]$, but it is enough for analyzing the rate of $\mathscr{M}$.

Proposition 1 (Li et al. 2019) Under Assumptions 1-4, we have

$$\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*)\right] \lesssim_{\mathbb{P}_{\mathbf{X}_m}} R_i^{\mathrm{IMSE}}(m,n) = \max\left\{\frac{\log^{\frac{1}{\kappa^i}}(mn)}{mn}, \frac{\log^{\frac{r(\kappa^i+1)}{\kappa^i}}(mn)}{m^{\frac{r}{2}}}\right\}.$$

Proposition 1 is built based on the bounds in Lemma 1 for a single SK model. It shows that the convergence rate of the IMSE of an SK model is governed by the rate function $R_i^{\mathrm{IMSE}}(m,n)$ defined in Section 2. Since in our setup of the problem, $r>2$ (according to Assumption 2) and n remains a constant in the offline sampling, Proposition 1 suggests that $\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*)\right]$ converges at the rate $\log^{\frac{1}{\kappa^i}} m/m$ in $\mathbb{P}_{\mathbf{X}_m}-$ probability.

Theorem 1 (Li et al. 2019) Suppose all the k designs have the sampling distribution $\mathbb{P}_{\mathbf{x}}$ for $\mathbf{X}_m$ and $\mathbf{x}^*$. Under Assumptions 1-4, the following result holds as $m\to\infty$,

$$\mathscr{M} \lesssim_{\mathbb{P}_{\mathbf{X}_m}} \overline{R}^{\mathrm{IMSE}}(m,n).$$

Theorem 1 identifies the convergence rate of $\mathscr{M}$, which is the same as the slowest rate for $\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*)\right]$'s, $i=1,2,...,k$. According to the definition of $\overline{R}^{\mathrm{IMSE}}(m,n)$ and the discussion about $R_i^{\mathrm{IMSE}}(m,n)$ above, $\overline{R}^{\mathrm{IMSE}}(m,n)$ takes the form $\log^{\frac{1}{\kappa}} m/m$, where κ is a parameter associated with the covariance kernels. It is not hard to see that Theorem 1 is a trivial extension of the result in Proposition 1.

Theorem 2 (Li et al. 2019) Suppose all the k designs have the sampling distribution $\mathbb{P}_{\mathbf{x}}$ for $\mathbf{X}_m$ and $\mathbf{x}^*$.
(i) Under Assumptions 1-4, $\mathscr{P} \lesssim_{\mathbb{P}_{\mathbf{X}_m}} \overline{R}^{\mathrm{IMSE}}(m,n)$ as $m\to\infty$;
(ii) Under Assumptions 1-6,

$$\mathscr{P} \lesssim_{\mathbb{P}_{\mathbf{X}_m}} \exp\left\{-\frac{1}{2} d_2^{1/2}\Delta_0\left[\overline{R}^{\mathrm{IMSE}}(m,n)\right]^{-1/2}\right\}$$

as $m\to\infty$;
(iii) Under Assumptions 1-5 and 7,

$$\mathscr{P} \lesssim_{\mathbb{P}_{\mathbf{X}_m}} \exp\left\{-\frac{1}{4} d_3^{-1}\Delta_0^2\left[\overline{R}^{\mathrm{IMSE}}(m,n)\right]^{-1}\right\}$$

as $m\to\infty$.

Theorem 2 identifies the convergence rate of $\mathscr{P}$. Part (i) states that the rate of $\mathscr{P}$ is also governed by the function $\overline{R}^{\mathrm{IMSE}}(m,n)$, same as $\mathscr{M}$. It can be obtained from Theorem 1 using the Markov's inequality. Lemma 1, Proposition 1, Theorem 1 and part (i) of Theorem 2 are all presented in a relatively general condition, which does not require the normal distribution assumption for the stochastic noise $\varepsilon_{il}(\mathbf{x})$ in Assumption 5 and the nice property on the tails of $\max_{i\in\{1,\ldots,k\}}\mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*)$ and $\max_{i\in\{1,\ldots,k\}}\sup_{\mathbf{x}^*\in\mathscr{X}}\mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*)$ in Assumptions 6 and 7.

Part (ii) states that with additional Assumptions 5 and 6, the convergence rate of $\mathscr{P}$ can be exponential, in the form of $\exp(-c\sqrt{m}\log^{-\frac{1}{2\kappa}} m)$, where $c=\frac{1}{2}d_2^{1/2}\Delta_0$, and part (iii) states that with additional Assumptions 5 and 7, the convergence rate of $\mathscr{P}$ can be exponential in the form of $\exp(-cm\log^{-\frac{1}{\kappa}} m)$, where $c=d_3^{-1}\Delta_0^2/4$. However, they are distinguished from the well-established exponential convergence rate of the PFS in R&S by comparing sample means of different designs (Dai 1996; Glynn and Juneja 2004), which takes the form of $\exp(-\nu n_{total})$, where n_{total} is the total number of simulation samples and ν is related to some large-deviations rate function. In practice, Assumptions 6 and 7 depend on the structure of $y_i(\mathbf{x})$ and the distribution of the stochastic noise, and are more difficult to verify.

4 NUMERICAL EXPERIMENTS

In this section, we numerically study the relationships between the maximal IMSE and IPFS with number of covariate points m. It provides a concrete demonstration of the theoretical results obtained in Section 3, and shows the rates that can be achieved on practical problems.

We use the following benchmark functions for testing. They have been extensively studied in the literature, e.g., in Duan et al. (1993), Potter and Jong (1994).

1. Ackley's function $(d = 1)$:

$$y(z) = f(z) + \varepsilon(z) = -a \cdot \exp(-b|z-x|) - \exp(\cos(c(z-x))) + a + \exp(1) + \varepsilon(z).$$

Parameters a, b and c are set as $a = 20$, $b = 0.2$ and $c = 2\pi$. For function $f(z)$, the global optimum z^* is obtained at $z^* = x$, with $f(z^*) = 0$. We consider $k = 9$ discrete designs $-4, -3, ..., 3, 4$ and covariate x is uniformly sampled in $[0,4]$. The variance of $\varepsilon(z)$ is 1 for all z and x.

2. Rosenbrock's valley $(d = 2)$:

$$y(\mathbf{z}) = f(\mathbf{z}) + \varepsilon(\mathbf{z}) = 100[(z_2 - x_2) - (z_1 - x_1)^2]^2 + [1 - (z_1 - x_1)^2]^2 + \varepsilon(\mathbf{z}).$$

For function $f(\mathbf{z})$, the global optimum $\mathbf{z}^*$ is obtained at $z_1^* = x_1 + 1$ and $z_2^* = x_2 + 1$, with $f(z^*) = 0$. We consider $k = 9$ discrete designs $(-4,-4), (-3,-3), ..., (3,3), (4,4)$ and covariate $\mathbf{x}$ is uniformly sampled in $[0,4] \times [0,4]$. The variance of $\varepsilon(z)$ is 1 for all $\mathbf{z}$ and $\mathbf{x}$.

3. Easom's function $(d = 2)$:

$$y(\mathbf{z}) = f(\mathbf{z}) + \varepsilon(\mathbf{z}) = -\cos(z_1 - x_1)\cos(z_2 - x_2)\exp(-(z_1 - x_1 - \pi)^2 - (z_2 - x_2 - \pi)^2) + \varepsilon(\mathbf{z}).$$

For function $f(\mathbf{z})$, the global optimum $\mathbf{z}^*$ is obtained at $z_1^* = x_1 + \pi$ and $z_2^* = x_2 + \pi$, with $f(z^*) = -1$. We consider $k = 9$ discrete designs $(-4,-4), (-3,-3), ..., (3,3), (4,4)$ and covariate $\mathbf{x}$ is uniformly sampled in $[0,4] \times [0,4]$. The variance of $\varepsilon(z)$ is 1 for all $\mathbf{z}$ and $\mathbf{x}$.

In the test examples introduced above, we have modified the original benchmark functions in the following two aspects to fit into the framework of OSOA: 1) added the observation noise $\varepsilon(\mathbf{z})$ turn the function output from deterministic to stochastic; and 2) introduced covariate $\mathbf{x}$ as parameters in test example. We have used notation $y(\mathbf{z})$ and $f(\mathbf{z})$ to emphasize that the decision variable is $\mathbf{z}$; their performance, of course, is also affected by the covariate $\mathbf{x}$. In the special case of $\mathbf{x} = 0$, $f(\mathbf{z})$ degenerates to the original benchmark function.

To enable graphical presentation, we will test the convergence rates of $\tilde{\mathscr{M}}$ and $\tilde{\mathscr{P}}$ with

$$\tilde{\mathscr{M}} = \mathrm{E}_{\mathbf{X}_m}\left[\max_{i\in\{1,2,...,k\}} \mathrm{E}_{\mathbf{x}^*}\left[\mathrm{MSE}_{i,\mathrm{opt}}(\mathbf{x}^*)\right]\right],$$
$$\tilde{\mathscr{P}} = \mathrm{E}_{\mathbf{X}_m}\left[\mathrm{E}_{\mathbf{x}^*}\left[\mathrm{PFS}(\mathbf{x}^*)\right]\right].$$

In the test, we fix the number of m, sample m points in the covariate space, generate n simulation samples for each pair of design and covariate point, build the SK models and use the SK outputs to find MSE and PFS. This is repeated for multiple runs for estimating $\tilde{\mathscr{M}}$ and $\tilde{\mathscr{P}}$, and is then repeated for different values of m for showing their convergence rates. For each test example, we will study three distributions for $\varepsilon(\mathbf{z})$: normal distribution with $N(0,1)$, uniform distribution with $U[-\sqrt{3},\sqrt{3}]$, and Bernoulli distribution with $P(\varepsilon(\mathbf{z}) = 1) = 0.5$ and $P(\varepsilon(\mathbf{z}) = -1) = 0.5$. We let the number of replications at each pair of design and covariate value be 5, the number of test points $\mathbf{x}^*$ for each SK model be 1,600 and the number of macro-replications be 500. The IZ parameter for a correct selection is set as $\Delta_0 = 0.01$. Figure 1 shows the convergence rates of $\tilde{\mathscr{M}}$ and $\tilde{\mathscr{P}}$.

Note that according to Theorem 1 and part (i) of Theorem 2, we have

$$\log \mathscr{M} \le -\log m + c_M,$$
$$\log \mathscr{P} \le -\log m + c_P,$$

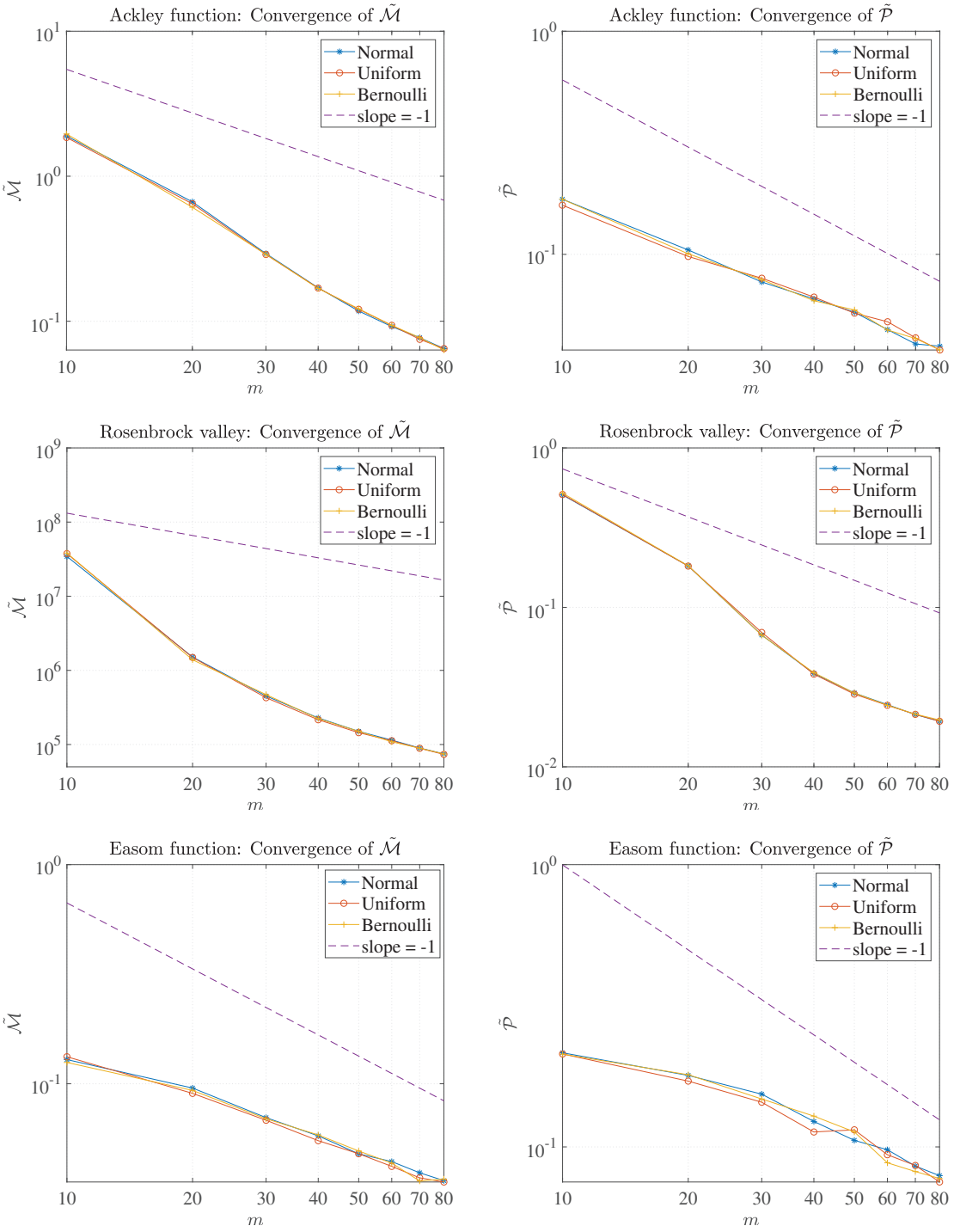

Figure 1: Convergence rates of $\tilde{\mathscr{M}}$ and $\tilde{\mathscr{P}}$.

where c_M and c_P are both $o(\log m)$. That is, when m is large enough, $\log \mathscr{M}$ and $\log \mathscr{P}$ are approximately negative linear with respect to $\log m$. Therefore, in each subfigure of Figure 1, we have added a reference line $-\log m + c_0$ for a clear presentation, where c_0 is a constant independent of m.

It can be observed that, for all the three functions tested, $\tilde{\mathscr{M}}$ and $\tilde{\mathscr{P}}$ converge. The convergence pattern of $\tilde{\mathscr{M}}$ for the Rosenbrock's valley is not so obvious compared to the rest because in this test function, the range of the function values is large, and $\tilde{\mathscr{M}}$ is still far from 0 when m is 80. Both $\log \tilde{\mathscr{M}}$ and $\log \tilde{\mathscr{P}}$ are at least negative linear with $\log m$, which is in line with our theoretical results. Different distributions of the observation noise do not cause a great influence on the convergence rates of the two terms. It is probably because the three distributions tested are all symmetric, and their variances are not very large. In these experiments, we have only tested cases with dimension $d = 1$ and $d = 2$. For problems with higher dimensions, the value of m for $\tilde{\mathscr{M}}$ and $\tilde{\mathscr{P}}$ to have stable performance is usually much larger.

5 CONCLUSIONS

Offline-simulation-online-application (OSOA) is a recently proposed framework for conducting simulation experiments. It advocates to start the simulation (long) before the parameters (covariate values) of the problems are revealed, and therefore has the potential to significantly reduce the time for simulation-based decisions, making simulation applicable for online problems. In this paper, we focus on an important question in OSOA, the relationship between the offline simulation efforts and the online application performance. We introduce two measures for evaluating the prediction errors in the online application, namely the maximal IMSE for system performance and IPFS for system optimization. By utilizing the stochastic kriging (SK) models, we show that the maximal IMSE and IPFS both converge at least at the rate of $\log^{\frac{1}{\kappa}} m/m$, where m is the number of covariate points ever sampled. The convergence rate of IPFS can be exponential as $\exp(-c\sqrt{m}\log^{-\frac{1}{2\kappa}} m)$ or $\exp(-cm\log^{-\frac{1}{\kappa}} m)$ in some more restrictive situations. These results for practitioners provide good insight into the online application performance that can be achieved from certain amount of offline simulation.

ACKNOWLEDGEMENTS

This research was supported in part by Hong Kong Research Grant Council under Grants 11214317 and 21204316, National Science Foundation of China under Grant 61603321 and National University of Singapore under Grant R155000172133.

REFERENCES

Ankenman, B. E., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58(2):371–382.

Bastani, H., and M. Bayati. 2016. "Online Decision-making with High-Dimensional Covariates". *Operations Research*. Forthcoming. https://ssrn.com/abstract=2661896.

Chen, C. H., J. Lin, E. Yücesan, and S. E. Chick. 2000. "Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization". *Discrete Event Dynamic Systems* 10(3):251–270.

Chen, X., B. E. Ankenman, and B. L. Nelson. 2013. "Enhancing Stochastic Kriging Metamodels with Gradient Estimators". *Operations Research* 61(2):512–528.

Chick, S. E., J. Branke, and C. Schmidt. 2010. "Sequential Sampling to Myopically Maximize the Expected Value of Information". *INFORMS Journal on Computing* 22(1):71–80.

Dai, L. 1996. "Convergence Properties of Ordinal Comparison in the Simulation of Discrete Event Dynamic Systems". *Journal of Optimization Theory and Applications* 91(2):363–388.

Duan, Q., V. K. Gupta, and S. Sorooshian. 1993. "Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization". *Journal of Optimization Theory and Applications* 76(3):501–521.

Dudewicz, E. J., and S. R. Dalal. 1975. "Allocation of Observations in Ranking and Selection with Unequal Variances". *Sankhya* 37B(1):28–78.

Eckman, D. J., and S. G. Henderson. 2018. "Guarantees on the Probability of Good Selection". In *Proeedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 351–365. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Farias, V., and A. Li. 2019. "Learning Preferences with Side Information". *Management Science* 65(7):3131–3149.

Feng, M., and J. Staum. 2017. "Green Simulation: Reusing the Output of Repeated Experiments". *ACM Transactions on Modeling and Computer Simulation* 27(4):1–28.

Frazier, P. I., W. B. Powell, and S. Dayanik. 2008. "A Knowledge-Gradient Policy for Sequential Information Collection". *SIAM Journal on Control Optimization* 47(5):2410–2439.

Fu, M., J. Hu, C.-H. Chen, and X. Xiong. 2007. "Simulation Allocation for Determining the Best Design in the Presence of Correlated Sampling". *INFORMS Journal on Computing* 19(1):101–111.

Gao, S., and W. Chen. 2016. "A New Budget Allocation Framework for Selecting Top Simulated Designs". *IIE Transactions* 48(9):855–863.

Gao, S., and W. Chen. 2017. "Efficient Feasibility Determination with Multiple Performance Measure Constraints". *IEEE Transactions Automatic Control* 62(1):113–122.

Gao, S., W. Chen, and L. Shi. 2017. "A New Budget Allocation Framework for the Expected Opportunity Cost". *Operations Research* 65(3):787–803.

Glynn, P., and S. Juneja. 2004. "A Large Deviations Perspective on Ordinal Optimization". In *Proceedings of the 2004 Winter Simulation Conference*, edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 577–585. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ho, Y., Q. Zhao, and Q. Jia. 2007. *Ordinal Optimization: Soft Optimization for Hard Problems*. New York: Springer.

Kersting, K., C. Plagemann, P. Pfaff, and W. Burgard. 2007. "Most Likely Heteroscedastic Haussian Process Regression". In *Proceedings of the 24th International Conference on Machine Learning*, edited by Q. Yang and M. Wooldridge, 393–400. Corvalis, Oregon: Association for Computing Machinery.

Kim, S. H., and B. L. Nelson. 2001. "A Fully Sequential Procedure for Indifference-Zone Selection in Simulation". *ACM Transactions on Modeling and Computer Simulation* 11(3):251–273.

Kleijnen, J. P. C. 2009. "Kriging Metamodeling in Simulation: A Review". *European Journal of Operational Research* 192(3):707–716.

Law, A. M. 2015. *Simulation Modeling and Analysis*. 5th ed. New York: McGraw-Hill.

Li, C., S. Gao, and J. Du. 2019. "Rate Analysis for Offline Simulation Online Application". Under review. http://blog.nus.edu.sg/stalic/rate-analysis-for-offline-simulation-online-application/.

Li, X., X. Zhang, and Z. Zheng. 2018. "Data-driven Ranking and Selection: High-Dimensional Covariates and General Dependence". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1933–1944. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Nelson, B. L., J. Swann, D. Goldsman, and W. Song. 2001. "Simple Procedures for Selecting the Best Simulated System When the Number of Alternatives is Large". *Operations Research* 49(6):950–963.

Opsomer, J. D., D. Ruppert, M. P. Wand, U. Holst, and O. Hössjer. 1999. "Kriging with Nonparametric Variance Function Estimation". *Biometrics* 55:704–710.

Pearce, M., and J. Branke. 2017. "Efficient Expected Improvement Estimation for Continuous Multiple Ranking and Selection". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 2161–2172. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Pearce, M., and J. Branke. 2018. "Continuous Multi-Task Bayesian Optimisation with Correlation". *European Journal of Operational Research* 270(3):1074–1085.

Perchet, V., and P. Rigollet. 2013. "The Multi-Armed Bandit Problem with Covariates". *The Annals of Statistics* 41(2):693–721.

Potter, M. A., and K. A. D. Jong. 1994. "A Cooperative Coevolutionary Approach to Function Optimization". In *Proceedings of the Third Conference on Parallel Problem Solving from Nature*, edited by Y. Davidor, H. Schwefel, and R. Männer, 249–257. Berlin, Heidelberg: Springer.

Qiang, S., and M. Bayati. 2017. "Dynamic Pricing with Demand Covariates". Under review. http://dx.doi.org/10.2139/ssrn.2765257.

Qu, H., and M. C. Fu. 2014. "Gradient Extrapolated Stochastic Kriging". *ACM Transactions on Modeling and Computer Simulation* 24(4):23:1–23:25.

Rasmussen, C. E., and C. K. Williams. 2006. *Gaussian Process for Machine Learning*. MIT press.

Ryzhov, I. O. 2016. "On the Convergence Rates of Expected Improvement Methods". *Operations Research* 64(6):1515–1528.

Shen, H., L. J. Hong, and X. Zhang. 2017. "Ranking and Selection with Covariates". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 2137–2148. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Stein, M. L. 2012. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media.

AUTHOR BIOGRAPHIES

SIYANG GAO is an Assistant Professor in the Department of Systems Engineering and Engineering Management at the City University of Hong Kong. He holds a Ph.D. in industrial engineering from University of Wisconsin-Madison. His research interests include simulation optimization, large-scale optimization and their applications to healthcare. He is a member of INFORMS and IEEE. His e-mail address is siyangao@cityu.edu.hk.

CHENG LI is an Assistant Professor in the Department of Statistics and Applied Probability at the National University of Singapore. He holds a Ph.D. in statistics from Northwestern University. His research interests include scalable Bayesian methods, Bayesian nonparametrics, high dimensional modeling, Bayesian analysis for heavy-tailed data, quasi-Bayesian inference and stochastic simulation. His e-mail address is stalic@nus.edu.sg.

JIANZHONG DU is a PhD student at the Department of Systems Engineering and Engineering Management, City University of Hong Kong. His research interests include simulation optimization and convergence analysis. His email address is jianzhodu2-c@my.cityu.edu.hk.