

MULTI-FIDELITY DISCRETE OPTIMIZATION VIA SIMULATION

Dongyang Li
Haitao Liu
Xiao Jin
Haobin Li
Ek Peng Chew

Kok Choon Tan

Department of Industrial Systems
Engineering and Management
National University of Singapore
1 Engineering Drive 2
Singapore, 117576, SINGAPORE

Department of Analytics
and Operations
National University of Singapore
15 Kent Ridge Drive
Singapore, 119245, SINGAPORE

Yun Hui Lin

School of Business
University College Dublin
Belfield, Dublin 4, IRELAND

ABSTRACT

Digital simulation is a prevalent tool to evaluate the performance of complex industrial systems. In this paper, we consider the discrete optimization via simulation (DOvS) problems, where the design space is an integer lattice. In addition, we are interested in leveraging cheap low-fidelity models to enhance optimization efficiency. An innovative Gaussian Markov random field (GMRF), which adaptively evolves along with the optimization process, is introduced to exploit the spatial and inter-model relationships among the objective function values of designs. We then propose the Multi-Fidelity Gaussian Markov Improvement Algorithm (MFGMIA) under the Bayesian global optimization framework. The numerical experiments show that it can achieve significant performance improvement by properly using low-fidelity information.

1 INTRODUCTION

In recent decades, modern industrial systems have been increasing complexity. As a result, simple analytical models are usually no longer competent for analysis. Meanwhile, the advent and fast development of the computer make digital simulation a reality and a popular tool for modeling complex systems. Optimization via simulation (OvS) aims to find the best design (or decision variable) for a system. Here, the simulation usually refers to stochastic simulation with randomness, and the best design is defined to have the smallest (or largest) performance mean. Generally, the OvS problem can be formulated as

$$\min_{x \in \Theta} f(x) = \mathbb{E}[F(x)],$$

where the d -dimensional vector x denotes the feasible design and Θ is the design space. The $f(x)$ does not have a closed form and has to be estimated by stochastic sampling from a high-fidelity simulation model.

This paper focuses on the so-called DOvS, where the x is integer-ordered, and Θ is a finite subset of the integer lattice. There is a wide range of OvS problems falling into this category, e.g., machine allocation problems and inventory management problems. In most DOvS problems, the design space is so large that the computing budget can only exhaust a small fraction of feasible designs, let alone mitigating the intrinsic randomness of simulation. Therefore, finding the optimal design for DOvS problems is often challenging, and smart algorithms that can efficiently allocate computing budgets are highly desirable.

Apart from the high-fidelity simulation model, sometimes cheap low-fidelity models are available. The computation cost of the latter is often negligible, and thus, proper use of low-fidelity information is promising to enhance the efficiency of OvS in the high-fidelity models. The main drawback of low-fidelity models is the simulation biases (or errors) against the high-fidelity models. The biases are unknown and heterogeneous among different designs. The information that we can leverage is only the “belief” that designs with good performance in low-fidelity models are expected to be promising in high-fidelity models. In some cases, the low-fidelity models are not reliable and can mislead the optimization. Therefore, how to properly quantify the “belief” in a rigorous way is the key point for multi-fidelity OvS.

We consider DOvS using multi-fidelity models. Because the design space is integer-ordered and the number of low-fidelity models is also countable, a graph can naturally be constructed. Then, a discrete version of GMRF is proposed to quantify the prior belief of the unknown objective function. Both the intrinsic noise rooted in the stochastic simulation and extrinsic uncertainty from the random field can be captured, and an expected improvement (EI) like acquisition function is used to guide the optimization. The greatest novelty is that we devise a mechanism to update the prior belief so that the influence of low-fidelity information will gradually vanish as high-fidelity information becomes rich. As illustrated later, this feature is important for DOvS algorithms to get rid of possible misleading caused by low-fidelity models.

The remainder of this paper is structured as follows. In section 2, we review related literature. Section 3 introduces the optimization problem and the algorithm. In section 4, we compare the efficiency of MFGMIA with benchmark algorithms in several synthetic examples. Section 5 concludes this paper.

2 LITERATURE REVIEW

The broad topic of this paper is OvS. We refer interested readers to Fu (2015) for a thorough discussion on problems and algorithms. Specifically, this project is related to Bayesian optimization (BO), DOvS, and multi-fidelity optimization.

The BO dates back to the EGO algorithm (Jones et al. 1998), which aims to optimize expensive black-box functions without noise. The unknown objective function is regarded to be a Gaussian Random Field (prior distribution). As more objective function values are observed, the conditional (or posterior) distribution can be updated according to the Bayesian rule. Then, a famous criterion called EI, which can balance exploration and exploitation well, is used to guide the search. Later on, Huang et al. (2006) and Quan et al. (2013) extend to optimize objective functions with homoscedastic and heteroscedastic Gaussian noise, respectively. For those mentioned above, the objective function is usually assumed to be continuous among the design space. More details about BO can be found in Frazier (2018).

In DOvS problems, the discrete design space often contains a large number of feasible designs, and the use of adaptive random search (ARS) algorithms is prevailing in the literature. A survey of ARS algorithms can be found in Nelson (2010). There are few DOvS algorithms under the BO framework. Recently, Salemi et al. (2019) devise an innovative discrete version of GMRF as the prior belief and propose GMIA for DOvS problems. Semelhago et al. (2021) speed up the GMIA by employing a divide-and-conquer strategy to switch the search between a small promising subset and the global design space.

Using multi-fidelity models to improve optimization efficiency is first considered in the global optimization and computer science community (see, e.g., Forrester et al. (2007) and Huang et al. (2006)). Kandasamy et al. (2019) consider the multi-armed bandit problem using multi-fidelity models. Under a more general setting which does not assume a rank for low-fidelity models, here comes the so-called

multi-information source optimization, e.g., the misoKG (Poloczek et al. 2017), Ghoreishi and Allaire (2019), and Lam et al. (2015). Those algorithms usually use independent Gaussian random fields, e.g., the autoregressive co-kriging model (Kennedy and O’Hagan 2000), to fit the difference between high-fidelity and low-fidelity objective functions. In the OvS realm, the cost of low-fidelity samplings is assumed negligible, and the focus is thus on how to sample designs in the high-fidelity model. MO²TOS (Xu et al. 2016) uses the low-fidelity information to rank all designs, and then applies the OCBA rule (Chen et al. 2000) to guide the search in the ordinal space. Under a very different problem setting, Peng et al. (2019) considered multi-fidelity optimization for ranking and selection (R&S) problems.

3 PROBLEM FORMULATION

We consider a DOvS problem as

$$\min_{x \in \Theta} f(x) = \mathbb{E}[F(x)],$$

where $f(x)$, the objective function, is unknown but can be estimated by sampling from a high-fidelity simulation model. The simulation output $F(x)$ is random. Formally, we have

$$F_i(x) = f(x) + \varepsilon_i(x)$$

for the i -th replication, where $\{\varepsilon_i(x), i = 1, 2, \dots\}$ are independent and identically distributed (i.i.d.) Gaussian noise with mean zero and finite variance $\sigma_\varepsilon^2(x)$ that depends on x . The design space $\Theta = \Phi \cap \mathbb{Z}^d$ is finite and $|\Theta| = n$, where $\Phi \subset \mathbb{R}^d$ is a convex compact set (e.g., hyperrectangle), and \mathbb{Z}^d is the set of d -dimensional integer lattice. We assume there is a single optimal design $x^* = \arg \min_{x \in \Theta} f(x)$. In addition to the high-fidelity simulation model, we also have m low-fidelity models, whose output $g_l(x)$, $l = 1, 2, \dots, m$, can provide biased approximation for $f(x)$. Thus we have the following equation:

$$f(x) = g_l(x) + \delta_l(x), \quad l = 1, 2, \dots, m,$$

where the bias $\delta_l(x)$ is unknown and heterogeneous. Here we assume that those low-fidelity models are constructed by different methods; therefore, there is no hierarchical structure for their fidelity level. Furthermore, the sampling cost of low-fidelity models (e.g., calculations of analytical formulas) is negligible, and we already knew $g_l(x), l = 1, 2, \dots, m, x \in \Theta$ free of noise before the DOvS in the high-fidelity model.

In our problem setting, the design space Θ is relatively large, while the sampling budget is limited. Consequently, only a small fraction of the designs can be sampled. The goal is to smartly utilize low-fidelity information and efficiently allocate the sampling budgets to find the optimal design x^* . We aim to solve the problem under the BO framework, where the unknown objective function is modeled as a random field. As high-fidelity observations are collected, the conditional distribution is updated and used to guide the search. A skeleton of the framework for our multi-fidelity DOvS is given in Figure 1. In the following, we always call step 1 for initial sampling as iteration 0 and further cycles that go from step 2 to step 5 as iteration 1, iteration 2, and so on.

3.1 The GMRF for Multi-Fidelity DOvS Problems

We choose the GMRF to model the prior belief. All vectors in this paper are column vectors. A GMRF is a multivariate Gaussian random vector $Y = (Y_1, Y_2, \dots, Y_n)^T \in \mathbb{R}^n$ with respect to an undirected and labeled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each Y_i corresponds to a unique node i in \mathcal{V} . Two nodes are called the neighbor of each other if an edge connects them. The GMRF is “Markov” in the sense that, after knowing all the neighbors of a component Y_i , it is conditional independent from all rest components, and see Figure 2 for an example. Given the mean vector μ and the precision matrix (the inverse of the covariance matrix) Q , the probability density function of Y is

$$\phi(y|\mu, Q) = (2\pi)^{-n/2} |Q|^{1/2} \exp\left(-\frac{1}{2}(y - \mu)^T Q (y - \mu)\right).$$

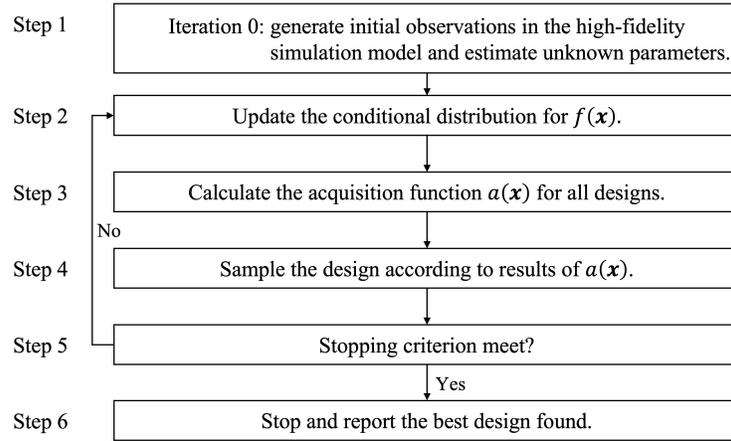


Figure 1: Bayesian optimization framework for DOvS using multi-fidelity models.

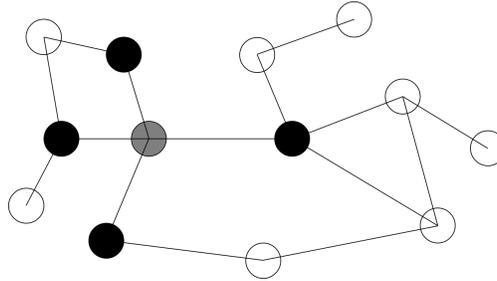


Figure 2: An example of the graph associated with a GMRF. After observing all the neighbors of the gray node, which are the four black nodes, the gray node is conditionally independent of other white nodes.

The precision matrix Q is closely related to the graph and the conditional distribution of Y . To be specific, the diagonal entries of Q are the conditional precision of the corresponding Y_i , and the off-diagonal entries are proportional to the conditional correlation of the corresponding Y_i and Y_j , after observing all neighbors. Q is often very sparse since the graph of GMRFs usually does not have too many edges. Refer to Rue and Held (2005) for more details of GMRFs.

For DOvS problems, Salemi et al. (2019) introduce a graph structure that works well for the integer-ordered design space, and Figure 3 shows an example. Then, the mean vector of the GMRF contains a constant, i.e., $\mu = \mu_0 \mathbf{1}_{n \times 1}$, and the $n \times n$ precision matrix Q is given as

$$Q_{ij} = \begin{cases} \theta, & \text{if } x_i = x_j; \\ -\theta \lambda_k, & \text{if } |x_i - x_j| = e_k; \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where e_k is the d -dimensional basis vector with 1 at the k -th dimension and 0 otherwise, $k = 1, 2, \dots, d$. In simple words, a single parameter θ is introduced to be the conditional precision for all designs, and λ_k measures the conditional correlation between any two designs that are adjacent in the k -th dimension. The objective function of DOvS problems is usually non-smooth and fluctuates intensively, and thus, after knowing the neighbors of a design x_i , other designs can provide little information for predicting $f(x_i)$. GMRFs perfectly encapsulate this belief in the graph and precision matrix. By contrast, traditional Gaussian random fields with a continuous covariance kernel function are more suitable for fitting continuous and smooth objective functions. Besides, the precision matrix Q is sparse, where at most $(2d + 1)/n$ entries

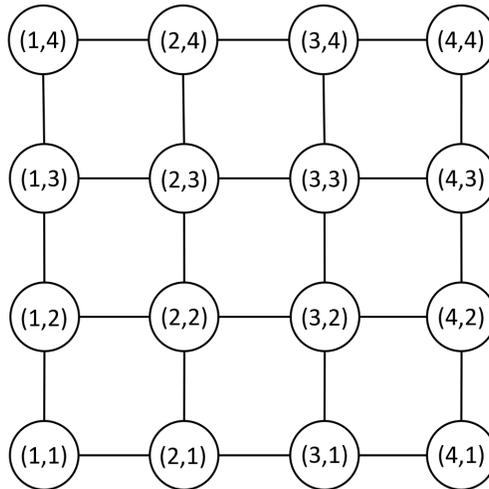


Figure 3: An example of the graph for a two-dimensional integer lattice, where the neighbors of a node just include those within a Euclidean distance of 1 in the space.

are nonzero, so sparse-matrix techniques can facilitate the calculation. If adopting continuous covariance kernels, a dense and ill-conditional matrix needs to be inverted in each iteration of the BO.

Motivated by the above, we propose a GMRF with a graph as in Figure 4 to integrate the low-fidelity information for DOvS, which is a straightforward extension of the graph in Figure 3. Note that there is no

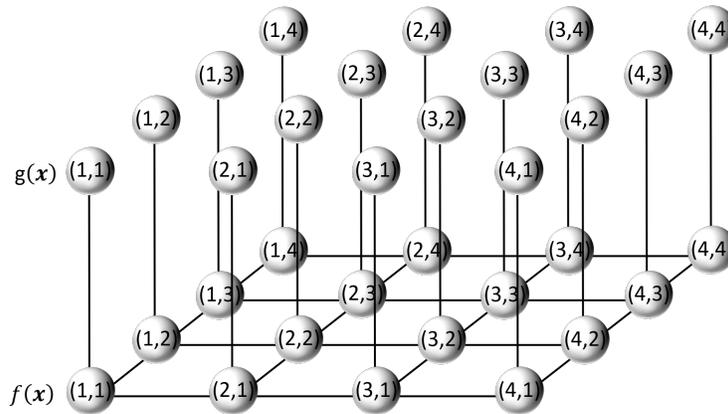


Figure 4: An example of the graph of the GMRF for a two-dimensional DOvS problem with one low-fidelity model. The nodes in the first layer denote the objective function values of the high-fidelity model, and those in the upper layer denote the objective function values of the low-fidelity model.

edge among the nodes of low-fidelity models themselves. In our problem setting, the objective function values of low-fidelity models are already known, and the focus is on the prediction of the high-fidelity objective function values. Deleting those redundant edges makes a negligible impact on the prediction and can introduce fewer parameters.

We pretend all objective function values, including those of low-fidelity models, are unknown, and consider the prior belief for

$$(f^T, g_1^T, \dots, g_m^T)^T = (f(x_1), \dots, f(x_n), g_1(x_1), \dots, g_1(x_n), \dots, g_m(x_1), \dots, g_m(x_n))^T.$$

Since the low-fidelity objective functions usually shift a bit away from the high-fidelity objective function, we introduce different constants $(\mu_0, \mu_1, \dots, \mu_m)$ for the prior means of each models, and it follows

$$\mu = (\mu_0 \mathbf{1}_{1 \times n}, \mu_1 \mathbf{1}_{1 \times n}, \dots, \mu_m \mathbf{1}_{1 \times n})^T. \quad (2)$$

The $nm \times nm$ precision matrix Q is given as

$$Q(g_l(x_i), g_{l'}(x_j)) = \begin{cases} \theta_l, & \text{if } x_i = x_j \text{ and } l = l'; \\ -\lambda_k \theta_0, & \text{if } |x_i - x_j| = e_k \text{ and } l = l' = 0; \\ -\rho_l \sqrt{\theta_0 \theta_l}, & \text{if } x_i = x_j \text{ and } l \neq 0, l' = 0; \\ -\rho_{l'} \sqrt{\theta_0 \theta_{l'}}, & \text{if } x_i = x_j \text{ and } l = 0, l' \neq 0; \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where $f(\cdot)$ is represented by $g_0(\cdot)$ here. $(\theta_0, \theta_1, \dots, \theta_m)$ are conditional precision for each model, $(\lambda_1, \dots, \lambda_d)$ play the same role as before, and (ρ_1, \dots, ρ_m) measure the conditional correlation between low-fidelity models and the high-fidelity model. All parameters can be estimate using the initial sampling observations. As a result, the prior distribution of $(f^T, g_1^T, \dots, g_m^T)^T$ is $N(\mu, Q^{-1})$ with μ and Q defined by (2) and (3).

3.2 Update of the GMRF

Traditional BO algorithms directly update the conditional distribution via the Bayesian rule. Many multi-fidelity algorithms under the BO framework, e.g., Huang et al. (2006) and Kandasamy et al. (2019), indeed follow this routine. However, they often need high-quality low-fidelity models, e.g., a correlation of at least 0.9, to perform well, because the low-fidelity information always influences the conditional distribution. In some cases, low-fidelity models can mislead the optimization. Since it is unlikely to ensure the low-fidelity models with high quality in practice, we are motivated to devise a mechanism that can properly deal with low-fidelity models. There are two features that we want our mechanism can realize:

- Allocating more influence to better low-fidelity models while less to worse low-fidelity models;
- Allocating more influence to low-fidelity models in the area with rare high-fidelity information, and decreasing the low-fidelity influence as the high-fidelity information becomes rich.

The first feature is already realized by the parameters (ρ_1, \dots, ρ_m) , which roughly measure the overall quality of each low-fidelity model. The motivation of the second feature is that high-fidelity information is always superior to low-fidelity information, so the latter is supposed to be gradually abandoned to avoid possible misleading. Fortunately, we are able to leverage the graph structure of GMRFs to realize it by modifying the precision matrix Q . Unlike the traditional BO, which only updates the conditional distribution, our framework also needs to update the prior belief.

To make it happen, we first introduce some new notation. Denote the information set collected before iteration t as \mathcal{F}_t , $t = 0, 1, \dots$, and we have $\mathcal{F}_0 = \{g_1, g_2, \dots, g_m\}$. Let $Q_\varepsilon^{(t)}$ be an $n \times n$ diagonal matrix, where the diagonal entries represent the intrinsic precision brought by the high-fidelity sampling, i.e.,

$$Q_\varepsilon^{(t)} = \text{diag}(r_t(x_1)/\sigma_\varepsilon^2(x_1), \dots, r_t(x_n)/\sigma_\varepsilon^2(x_n)), \quad (4)$$

where $r_t(x_i)$ is the number of high-fidelity sampling replications for design x_i before iteration t , $i = 1, 2, \dots, n$. The superscript (t) appearing here and in the following all indicates the term is updated before iteration t . The variance of sampling noise $\sigma_\varepsilon^2(x_i)$ is unknown but can be estimated by sample variance $s_\varepsilon^2(x_i)$. Then, another $n \times n$ auxiliary matrix A_t , which helps to update the precision matrix, is defined as

$$A_{ij}^{(t)} = \begin{cases} \frac{\lambda_k}{2 \sum_{r=1}^d \lambda_r} \max(S(Q_\varepsilon^{(t)}(i, i)), S(Q_\varepsilon^{(t)}(j, j))), & \text{if } |x_i - x_j| = e_k; \\ \sum_{k \neq i} A_{ik}^{(t)}, & \text{if } x_i = x_j; \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where $S(x) = \frac{2}{1+e^{-x}} - 1$ is a Sigmoid-like function mapping those non-negative intrinsic precision to $[0, 1)$. Then, denote the prior precision matrix before iteration t as $Q^{(t)}$, and it follows that

$$Q^{(t)}(g_l(x_i), g_{l'}(x_j)) = \begin{cases} \theta_l, & \text{if } x_i = x_j \text{ and } l = l'; \\ -(\lambda_k + A_{ij}^{(t)} \sum_{r=1}^m \rho_r) \theta_0, & \text{if } |x_i - x_j| = e_k \text{ and } l = l' = 0; \\ -\rho_l(1 - A_{ii}^{(t)})\sqrt{\theta_0\theta_l}, & \text{if } x_i = x_j \text{ and } l \neq 0, l' = 0; \\ -\rho_{l'}(1 - A_{ii}^{(t)})\sqrt{\theta_0\theta_{l'}}, & \text{if } x_i = x_j \text{ and } l = 0, l' \neq 0; \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Further, we can partition the precision matrix as

$$Q^{(t)} = \begin{bmatrix} Q_{ff}^{(t)} & Q_{fg}^{(t)} \\ Q_{gf}^{(t)} & Q_{gg}^{(t)} \end{bmatrix}.$$

As high-fidelity information grows, some off-diagonal entries corresponding to the conditional correlation between the low-fidelity and high-fidelity models decrease, while those corresponding to the conditional correlation inside the high-fidelity model increase accordingly. Furthermore, we keep the sum of conditional correlation from neighbors for any design in the high-fidelity model a constant, which makes the conditional mean of $f(x)$ a weighted average of its neighbors, although the weights will change adaptively. The following proposition ensures the positive definiteness of the precision matrix $Q^{(t)}$ under mild conditions.

Proposition 1 Suppose parameters $(\theta_0, \dots, \theta_m)$ are positive, $(\lambda_1, \dots, \lambda_d, \rho_1, \dots, \rho_m)$ are non-negative and satisfy $\sum_{l=1}^m \rho_l + 2\sum_{k=1}^d \lambda_k < 1$, then the precision matrix $Q^{(t)}$ defined by (6) is always positive definite.

As a result, our prior belief for $(f^T, g_1^T, \dots, g_m^T)^T$ before iteration t is $N(\mu, (Q^{(t)})^{-1})$ with μ and $Q^{(t)}$ defined in (2) and (6) respectively. Let $\bar{F}^{(t)}$ be an $n \times 1$ vector such that each element is either the sample mean of the high-fidelity observations of the corresponding design, if it has been sampled, or μ_0 otherwise. Let $g = (g_1^T, \dots, g_m^T)^T$ and $\mu_g = (\mu_1 1_{1 \times n}, \dots, \mu_m 1_{1 \times n})^T$. Then we have the following theorem for the conditional distribution of f .

Theorem 1 The conditional distribution of $f = (f(x_1), \dots, f(x_n))^T$ given \mathcal{F}_t is

$$f | \mathcal{F}_t \sim N(\mu_f^{(t)}, \Sigma_f^{(t)}),$$

with

$$\mu_f^{(t)} = \left(\mu_0 1_{n \times 1} - (Q_{ff}^{(t)})^{-1} Q_{fg}^{(t)}(g - \mu_g) \right) + \Sigma_f^{(t)} Q_\varepsilon^{(t)} \left(\bar{F}^{(t)} - \mu_0 1_{n \times 1} + (Q_{ff}^{(t)})^{-1} Q_{fg}^{(t)}(g - \mu_g) \right) \quad (7)$$

and

$$\Sigma_f^{(t)} = (Q_{ff}^{(t)} + Q_\varepsilon^{(t)})^{-1}. \quad (8)$$

Note that $(Q_{ff}^{(t)})^{-1} Q_{fg}^{(t)}(g - \mu_g)$ is the contribution from the low-fidelity information. As more and more high-fidelity observations are collected, some nonzero entries in $Q_{fg}^{(t)}$ will go to zero, and $(Q_{ff}^{(t)})^{-1} Q_{fg}^{(t)}(g - \mu_g)$ gradually disappears as well. Moreover, both $Q_{ff}^{(t)}$ and $(Q_{ff}^{(t)} + Q_\varepsilon^{(t)})$ are $n \times n$ symmetric diagonal-dominant sparse matrix, and the fraction of nonzero entries is also bounded above by $(2d + 1)/n$.

3.3 Acquisition Function

The acquisition function plays an important role in BO. We adopt the complete expected improvement (CEI) as in Salemi et al. (2019) and Semelhago et al. (2021), which quantifies both the intrinsic noise and extrinsic uncertainty of the unknown objective function values.

Denote $\tilde{x}^{(t)}$ the best sampled design before iteration t , i.e., the one with the lowest sample mean. Then, the CEI is defined as

$$\text{CEI}^{(t)}(x) = \mathbb{E} \left[\max(f(\tilde{x}^{(t)}) - f(x), 0) \mid \mathcal{F}_t \right]. \quad (9)$$

Note that the conditional distribution of $(f(\tilde{x}^{(t)}), f(x))$ is bivariate normal. Denote the conditional mean and variance of their difference $(f(\tilde{x}^{(t)}) - f(x))$ by $M^{(t)}(x)$ and $V^{(t)}(x)$. Then, we have

$$M^{(t)}(x) = \mu_f^{(t)}(\tilde{x}^{(t)}) - \mu_f^{(t)}(x) \quad \text{and} \quad V^{(t)}(x) = \Sigma_f^{(t)}(\tilde{x}^{(t)}) + \Sigma_f^{(t)}(x) - 2\Sigma_f^{(t)}(\tilde{x}^{(t)}, x).$$

After some simple algebra, the CEI defined in (9) can be calculated as

$$\text{CEI}^{(t)}(x) = M^{(t)}(x) \Phi \left(\frac{M^{(t)}(x)}{\sqrt{V^{(t)}(x)}} \right) + \sqrt{V^{(t)}(x)} \phi \left(\frac{M^{(t)}(x)}{\sqrt{V^{(t)}(x)}} \right), \quad (10)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution function and probability density function of the standard normal distribution. Note that $\text{CEI}^{(t)}(\tilde{x}^{(t)})$ is always 0. Moreover, to compute CEI, we only need to know the diagonal entries and the column corresponding to $\tilde{x}^{(t)}$ of the conditional covariance matrix $\Sigma_f^{(t)}$.

3.4 MFGMIA

Now, we are ready to give the whole picture of our optimization algorithm, MFGMIA. There are several

Algorithm 1: MFGMIA

- 1 Choose n_0 designs and sample r replications for each of them in the high-fidelity model;
 - 2 Use the initial outputs to estimate $(\mu_0, \dots, \mu_m, \theta_0, \dots, \theta_m, \lambda_1, \dots, \lambda_d, \rho_1, \dots, \rho_m)$;
 - 3 **while** *Stopping criterion not met* **do**
 - 4 Find the current best sampled design $\tilde{x}^{(t)}$;
 - 5 Update $Q_\varepsilon^{(t)}$, $A^{(t)}$ and $Q^{(t)}$ according to (4), (5) and (6);
 - 6 Compute Cholesky factor of $Q_{ff}^{(t)}$ and $(Q_\varepsilon^{(t)} + Q_{ff}^{(t)})$: $L_{ff}^{(t)}$ and $L_{f\varepsilon}^{(t)}$;
 - 7 Compute $\mu_f^{(t)}$ according to (7), using $L_{ff}^{(t)}$ and $L_{f\varepsilon}^{(t)}$;
 - 8 Compute the diagonal entries of $\Sigma_f^{(t)}$ according to (8), using $L_{f\varepsilon}^{(t)}$;
 - 9 Compute the column of $\Sigma_f^{(t)}$ corresponding to $\tilde{x}^{(t)}$ according to (8), using $L_{f\varepsilon}^{(t)}$;
 - 10 Compute $\text{CEI}^{(t)}(x)$ by (10), $\forall x \in \Theta/\tilde{x}^{(t)}$; Find $x_{\text{CEI}}^{(t)} = \arg \max_{x \in \Theta} \text{CEI}^{(t)}(x)$;
 - 11 Sample both $\tilde{x}^{(t)}$ and $x_{\text{CEI}}^{(t)}$ for r replications respectively;
 - 12 **end**
 - 13 Return the best sampled design \tilde{x} as the estimated optimal solution.
-

parameters that we need to prespecify. For the initial experimental designs, a rule of thumb in the literature is $n_0 = 10d$, and the Latin hypercube sampling method is recommended. For the number of replications r , it cannot be too small because it may lead to poor estimation of $\sigma_\varepsilon^2(x)$. A suggestion is somewhere between 5 and 20 (Chen and Lee 2011), and we adopt $r = 10$ in the numerical experiments.

In step 5, only the entries related to the sampled design of the previous iteration need to be updated. Step 6 is the most expensive one, where factorization techniques for sparse diagonal-dominant matrix can

facilitate calculation. In step 11, the current found best design $\tilde{x}^{(t)}$ is sampled because it is promising, and a good estimate of the optimal objective function value is crucial for OvS algorithms to work well. For the stopping criterion, we suggest the fixed-budget.

4 NUMERICAL EXPERIMENTS

In this section, we compare the MFGMIA with several competitors. The main goal is to check whether the MFGMIA can efficiently and properly make use of the low-fidelity information. For comparison, we create two versions of GMIA: (i) using the GMRF to fit $f(x)$ only and applying CEI (GMIA1); (ii) using the GMRF to fit $f(x) - g_l(x)$ and applying CEI (GMIA2). The second one is to let the GMIA use low-fidelity information, but unlike MFGMIA which can adaptively abandon the low-fidelity information, the influence of $g_l(x)$ always exists in GMIA2. Hence, GMIA2 can be regarded to make use of the spirit of many multi-fidelity optimization algorithms in the literature, e.g., Huang et al. (2006). Besides, the random sampling algorithm is introduced as a naive benchmark. The following two experiments are both adapted from Xu et al. (2016).

4.1 Synthetic Function Examples

The high-fidelity function $f(x)$ and three low-fidelity functions $g_l(x), l = 1, 2, 3$, are plotted in Figure 5. The design space is $\Theta = \{1, 2, \dots, 1000\}$ and the optimal design is $x^* = 164$ with objective function value

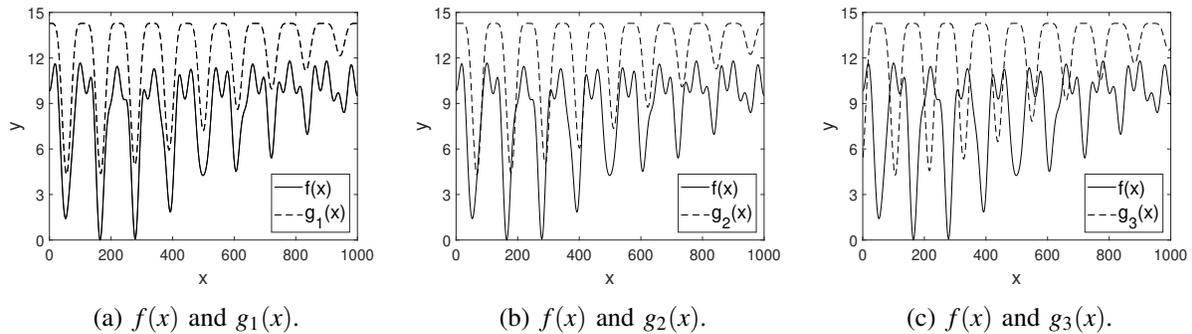


Figure 5: Multi-fidelity objective functions.

$f(x^*) = 0$. All three low-fidelity functions have the same shape and shift upwards from the high-fidelity function. Nevertheless, their horizontal shifts are different, which leads to different correlations with the high-fidelity function, i.e., 0.95 for $g_1(x)$, 0.71 for $g_2(x)$, and -0.50 for $g_3(x)$. We can regard $g_1(x)$ as a high-quality low-fidelity model, since it has almost the same local minimum region as $f(x)$. $g_2(x)$ is middle-quality since it regards some sub-optimal designs as top ones, and it can probably mislead the optimization. Besides, the information of $g_3(x)$ is low-quality, which identifies good designs as poor ones.

Consequently, we create three test cases using each low-fidelity function, respectively. The sampling noise applied to $f(x)$ is universally $\varepsilon(x) \sim N(0, 0.01)$. We increase the sampling budget T from 100 to 2,000 and compare the finite-budget performance of each algorithm. The comparison criterion is the optimality gap, i.e., the difference between the objective function values of the reported best design and the true best design. All results are reported based on 1,000 independent macro replications, where the four algorithms start from the same initial sampling in each replication.

The results are plotted in Figure 6. In Figure 6a, we can see both MFGMIA and GMIA2 converge to the top designs in the very beginning, which of course owes to the high quality of $g_1(x)$. The second test case depicted in Figure 6b shows the strength of MFGMIA. Here, the low-fidelity function $g_2(x)$ can provide some useful information, so MFGMIA and GMIA2 both achieve better performance than GMIA1. However, because the information of $g_2(x)$ is a bit misleading, GMIA2 tends to get stuck in some

sub-optimal designs and converge slower than MFGMIA. Figure 6c further illustrates how GMIA2, which does not abandon low-fidelity information, can get into serious trouble. The low-fidelity function $g_3(x)$ cannot provide any useful information, so the best practice is to ignore it. As a result, our MFGMIA can identify the poor quality of $g_3(x)$ and perform equivalently well as GMIA1, while GMIA2 is even worse than random sampling in this scenario.

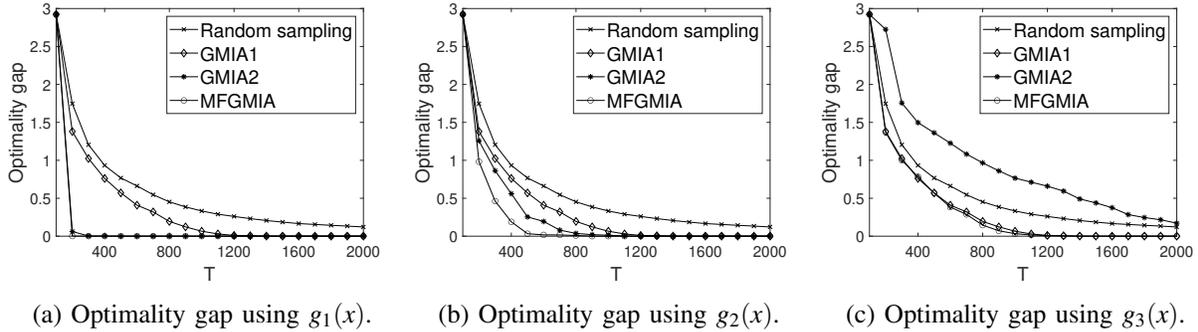


Figure 6: Optimality gap of each case.

4.2 A Machine Allocation Problem

This problem is about a manufacturing system, where 37 machines need to be allocated to 5 workstations so that the average processing time for each product is minimized. After considering all deterministic constraints, there are in total 780 feasible designs spreading over a 4-dimensional integer space. The high-fidelity model is a discrete-event simulation model constructed in computers, while the low-fidelity model is an analytical approximation by treating the system as a Jackson network. Figure 7 displays the sketch of the manufacturing system and the simulation output of multi-fidelity models, where the designs are listed in terms of their low-fidelity ranks.

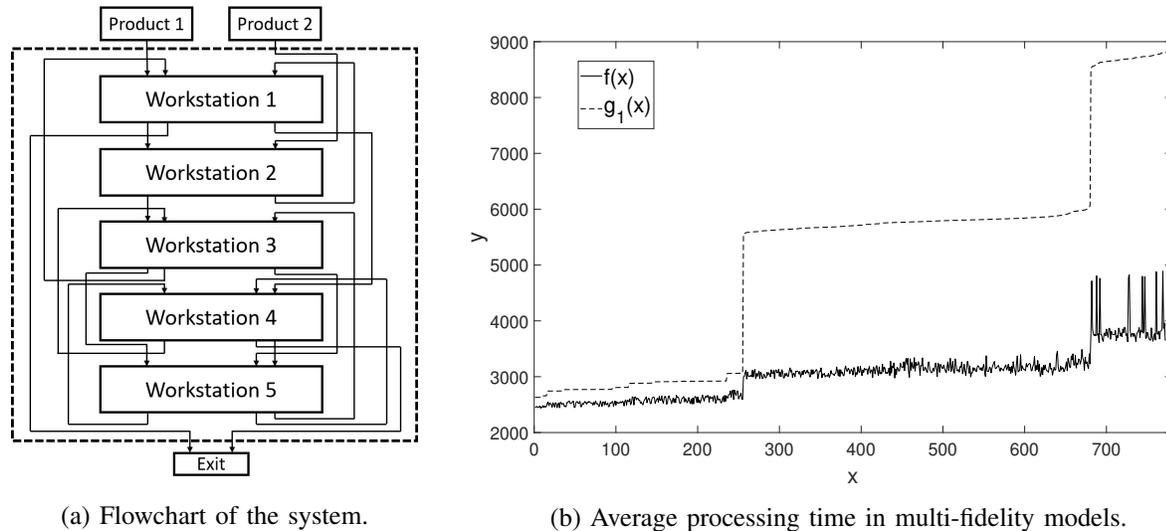


Figure 7: The manufacturing system.

Again, we add $\epsilon(x) \sim N(0, 0.01)$ to $f(x)$ to mimic the simulation noise. The initial sampling size n_0 is set to be 20, and we increase the sampling budget T from 200 to 2,000. The results are plotted in Figure

8. We can see both MFGMIA and GMIA2 quickly converge to the optimal design, which is essentially

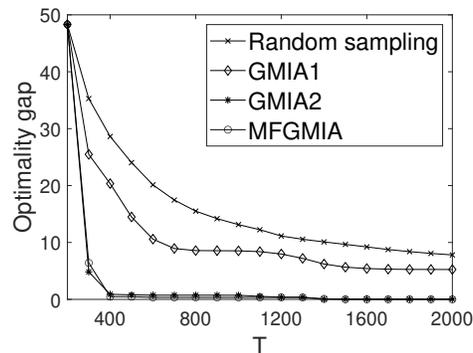


Figure 8: Optimality gap for machine allocation problem.

due to the high quality of the low-fidelity approximation.

In summary, our MFGMIA can efficiently utilize valuable low-fidelity information. It can also properly deal with the misleading information brought by the low-fidelity models. Furthermore, when the low-fidelity model has poor quality, the MFGMIA can perform at least equivalently well as the single-fidelity algorithm.

5 CONCLUSION

DOvS problems with an integer-ordered design space are prevalent in industries. In this paper, we devise an optimization algorithm for DOvS problems where cheap low-fidelity models are available. Specifically, our algorithm is developed under the BO framework, and an innovative GMRF is constructed to model the prior belief for the unknown objective function values. To better deal with the low-fidelity information, we adaptively update the prior distribution so that the high-fidelity information gradually dominates the prediction. In the numerical experiments, we show the strength of the MFGMIA in multi-fidelity DOvS problems.

ACKNOWLEDGMENTS

This research paper has been made possible by the funding support from the Singapore Maritime Institute and the Centre of Excellence in Modelling and Simulation for Next Generation Ports (Singapore Maritime Institute grant: SMI-2017-SP-02).

REFERENCES

- Chen, C.-H., and L. H. Lee. 2011. *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*. Hackensack, New Jersey: World Scientific.
- 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.
- Forrester, A. I., A. Söbester, and A. J. Keane. 2007. "Multi-Fidelity Optimization via Surrogate Modelling". In *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, edited by M. Lockwood, 3251–3269. London: Royal Society.
- Frazier, P. I. 2018. "Bayesian Optimization". In *INFORMS Tutorials in Operations Research*, 255–278. INFORMS.
- Fu, M. C. 2015. *Handbook of Simulation Optimization*. New York: Springer.
- Ghoreishi, S. F., and D. Allaire. 2019. "Multi-Information Source Constrained Bayesian Optimization". *Structural and Multidisciplinary Optimization* 59(3):977–991.
- Huang, D., T. T. Allen, W. I. Notz, and R. A. Miller. 2006. "Sequential Kriging Optimization Using Multiple-Fidelity Evaluations". *Structural and Multidisciplinary Optimization* 32(5):369–382.
- Huang, D., T. T. Allen, W. I. Notz, and N. Zeng. 2006. "Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models". *Journal of Global Optimization* 34(3):441–466.

- Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization* 13(4):455–492.
- Kandasamy, K., G. Dasarathy, J. Oliva, J. Schneider, and B. Póczos. 2019. "Multi-Fidelity Gaussian Process Bandit Optimisation". *Journal of Artificial Intelligence Research* 66:151–196.
- Kennedy, M. C., and A. O'Hagan. 2000. "Predicting the Output from a Complex Computer Code When Fast Approximations are Available". *Biometrika* 87(1):1–13.
- Lam, R., D. L. Allaire, and K. E. Willcox. 2015. "Multifidelity Optimization Using Statistical Surrogate Modeling for Non-Hierarchical Information Sources". In *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. January 5th-9th, Kissimmee, Florida, 0143.
- Nelson, B. L. 2010. "Optimization via Simulation Over Discrete Decision Variables". In *INFORMS Tutorials in Operations Research*, 193–207. INFORMS.
- Peng, Y., J. Xu, L. H. Lee, J. Hu, and C.-H. Chen. 2019. "Efficient Simulation Sampling Allocation Using Multifidelity Models". *IEEE Transactions on Automatic Control* 64(8):3156–3169.
- Poloczek, M., J. Wang, and P. I. Frazier. 2017. "Multi-Information Source Optimization". In *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*. December 4th-9th, Long Beach, California, 4291–4301.
- Quan, N., J. Yin, S. H. Ng, and L. H. Lee. 2013. "Simulation Optimization via Kriging: A Sequential Search Using Expected Improvement with Computing Budget Constraints". *IIE Transactions* 45(7):763–780.
- Rue, H., and L. Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, Florida: Chapman and Hall.
- Salemi, P. L., E. Song, B. L. Nelson, and J. Staum. 2019. "Gaussian Markov Random Fields for Discrete Optimization via Simulation: Framework and Algorithms". *Operations Research* 67(1):250–266.
- Semelhago, M., B. L. Nelson, E. Song, and A. Wächter. 2021. "Rapid Discrete Optimization via Simulation with Gaussian Markov Random Fields". *INFORMS Journal on Computing* 33(3):915–930.
- Xu, J., S. Zhang, E. Huang, C.-H. Chen, L. H. Lee, and N. Celik. 2016. "MO²TOS: Multi-Fidelity Optimization with Ordinal Transformation and Optimal Sampling". *Asia-Pacific Journal of Operational Research* 33(03):1650017.

AUTHOR BIOGRAPHIES

DONGYANG LI is a Ph.D. candidate in the Department of Industrial Systems Engineering and Management at National University of Singapore. His research interests include simulation optimization, stochastic optimization, and reinforcement learning. His email address is dongyang.li@u.nus.edu.

HAITAO LIU is a Ph.D. candidate in the Department of Industrial Systems Engineering and Management at National University of Singapore. His research interests include simulation optimization and statistic learning. His email address is haitao.liu@u.nus.edu.

XIAO JIN is a Research Fellow in the Centre of Excellence in Modelling and Simulation for Next Generation Ports. He received his Ph.D. degree in the Department of Industrial Systems Engineering and Management from National University of Singapore. His research interest includes simulation optimization and artificial intelligence with applications on logistics and maritime studies. His email address is isejinx@nus.edu.sg.

HAOBIN LI is a Senior Lecture in the Department of Industrial Systems Engineering and Management at National University of Singapore, and he received his Ph.D. degree in 2014 from the same department. He has research interests in operations research, simulation optimization with application on logistics and maritime studies. His email address is li_haobin@nus.edu.sg.

EK PENG CHEW is a Professor and Deputy Head for the Department of Industrial Systems Engineering and Management, National University of Singapore. He received his Ph.D. degree from the Georgia Institute of Technology, USA. His research interests include logistics and inventory management, system modeling and simulation. His email address is isecep@nus.edu.sg.

KOK CHOON TAN is an Associate Professor and Deputy Head for the Department of Analytics and Operations, National University of Singapore. He received his Ph.D. degree from the Massachusetts Institute of Technology, USA. His research interests include logistics and supply chain optimization, and system modeling. His email address is biztkc@nus.edu.sg.

YUN HUI LIN is an Assistant Professor in Management Information Systems, University College Dublin. He obtained his Ph.D. degree from the Department of Industrial Systems Engineering and Management, National University of Singapore. His research focuses on operations research and management science. His email address is yunhui.lin@ucd.ie.