Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

AN ADDITIVE DECOMPOSITION FOR DISCRETE SIMULATION OPTIMIZATION USING GAUSSIAN MARKOV RANDOM FIELDS

Harun Avci

Department of Industrial Engineering & Management Sciences
Northwestern University
Evanston, IL 60208, USA

ABSTRACT

We consider a discrete optimization via simulation problem with high-dimensional, integer-ordered decision variables. One of the methods to solve such a problem is Bayesian optimization (BO). Although BO can provide rapid solution improvement within a tight computational budget, the posterior update creates a significant computational overhead for large-scale problems. To overcome this challenge, we propose an algorithm that decomposes the prior distribution into an additive form as an approximation. Despite this approximation, our numerical analysis reveals that the algorithm can obtain rapid improvement.

1 INTRODUCTION

Optimization via simulation describes the methodologies for solving optimization problems for which the objective function can only be estimated by using stochastic simulation experiments. Such problems with discrete feasible solutions are called discrete optimization via simulation (DOvS) problems. For large-scale DOvS problems, where only a small fraction of feasible solutions can be simulated, one needs to exploit spatial structure among feasible solutions (e.g., neighboring solutions have similar values).

Salemi et al. (2019) model the objective function values as a realization of a Gaussian Markov random field (GMRF) and propose the Gaussian Markov improvement algorithm (GMIA). Using the posterior distribution of the GMRF, GMIA evaluates the complete expected improvement (CEI) at each solution to guide the search. Since the computational cost of posterior update increases at least quadratically in the number of feasible solutions, it limits the problem size. To reach beyond that limit, we propose additive GMIA (aGMIA) that decomposes the prior distribution into an additive form as an approximation.

2 METHODOLOGY

Consider a DOvS problem to minimize $y(\boldsymbol{x})$ subject to $\boldsymbol{x} \in \mathscr{X}$, where $\mathscr{X}$ is a finite subset of the D-dimensional integer lattice with $N = |\mathscr{X}|$ feasible solutions. The objective, $y(\boldsymbol{x})$, can only be estimated via stochastic simulation. Let $\boldsymbol{y}$ denote the vector of (unknown) objective function values. As motivation for our prior, assume that $y(\boldsymbol{x})$ decomposes into an additive model $y(\boldsymbol{x}) = \beta_0 + y^{(1)}(\boldsymbol{x}^{(1)}) + y^{(2)}(\boldsymbol{x}^{(2)}) + \cdots + y^{(G)}(\boldsymbol{x}^{(G)})$, where $\boldsymbol{x}^{(\rho)} \in \mathscr{X}^{(\rho)}$ represents a lower dimensional component of $\boldsymbol{x}$ for $\rho \in \mathscr{G} = \{1, 2, \ldots, G\}$, and β_0 is a location parameter. The superscripts in parentheses are used to index the "groups," which are disjoint and together form the decomposition. The decomposition of groups can be based on domain knowledge of the problem or changed from time to time during the search. Let $N^{(\rho)} = |\mathscr{X}^{(\rho)}|$; notice that $N^{(\rho)} \leq N$.

For each $\rho \in \mathscr{G}$, let $\boldsymbol{y}^{(\rho)}$ be the vector of ρ-components of the objective function. Using $\boldsymbol{y}^{(\rho)}$'s, the vector $\boldsymbol{y}$ of objective function values can be expressed as $\boldsymbol{y} = \beta_0 \vec{\mathbf{1}}_N + \sum_{\rho \in \mathscr{G}} \boldsymbol{T}^{(\rho)} \boldsymbol{y}^{(\rho)}$, where $\vec{\mathbf{1}}_N$ is an N-dimensional vector of ones, and $\boldsymbol{T}^{(\rho)}$ is the transformation matrix associated with group ρ. Since $\boldsymbol{y}^{(\rho)}$ is unknown, we model it as a realization of the GMRF $\mathbb{Y}^{(\rho)}$, a random vector of size $N^{(\rho)}$. Due to the repetitive rows in

$\boldsymbol{T}^{(\rho)}$'s, the elements of $\boldsymbol{y}$ might be linearly dependent. Therefore, modeling $\boldsymbol{y}$ with $\boldsymbol{y}^{(\rho)}$'s replaced by $\mathbb{Y}^{(\rho)}$'s forms a random vector that has linearly dependent rows. To avoid such linear dependence, we use some group $g \in \mathscr{G}$ to remove the dependency of $y^{(g)}(\boldsymbol{x}^{(g)})$ on $\boldsymbol{x}^{(g)} \in \mathscr{X}^{(g)}$, and model it as a realization of a random variable. Therefore, the vector $\boldsymbol{y}$ of objective function values can be modeled as the realization of

$$\mathbb{Y} = \beta_0 \vec{\mathbf{1}}_N + \sum_{\rho \in \mathscr{G} \setminus \{g\}} \boldsymbol{T}^{(\rho)} \mathbb{Y}^{(\rho)} + \mathbb{W}^{(g)}, \tag{1}$$

where $\mathbb{W}^{(g)}$ is a zero-mean random vector with a diagonal covariance matrix. Group g is intended to have limited impact on the objective function for the model in (1) to be accurate. Here are the key contributions:

Hyperparameters: The hyperparameters of $\mathbb{Y}^{(\rho)}$'s and $\mathbb{W}^{(g)}$ need to be estimated. However, only the noisy simulation output $Y(\boldsymbol{x}) = y(\boldsymbol{x}) + \varepsilon(\boldsymbol{x})$ is observed. We develop a maximum likelihood estimator based on taking differences among outputs to isolate the effect of other groups.

Posterior update: Given the hyperparameters, we derive the posterior distributions of $\mathbb{Y}^{(\rho)}$'s and $\mathbb{W}^{(g)}$, and thus that of $\mathbb{Y}$. The posterior update requires inverting a matrix of size $N^{(\rho)} \times N^{(\rho)}$ for each $\rho \in \mathscr{G} \setminus \{g\}$ and G matrices of size $|\mathscr{D}| \times |\mathscr{D}|$, where $\mathscr{D}$ is the set of feasible solutions that have been simulated. Without employing the additive decomposition, the posterior update requires inverting a matrix of size $N \times N$. Since $N^{(\rho)}$'s and $|\mathscr{D}|$ are much smaller than N, it becomes possible to solve large-scale, high-dimensional DOvS problems.

Two-stage sampling criterion: We develop a two-stage sampling criterion. The first stage restricts the candidate solutions to a much smaller subset of all feasible solutions by fixing $\boldsymbol{x}^{(\rho)}$ for $\rho \in \mathscr{G} \setminus \{g\}$. The second stage identifies the most promising solution (with the largest CEI) to simulate from the restricted set of candidate solutions.

Finding the best CEI solution: To construct the restricted set of candidate solutions in the first stage of the sampling criterion, we need to identify a solution with the largest CEI, requiring us to evaluate the CEI for each solution. When the number of feasible solutions is very large this calculation alone can become a huge burden. Using relations between the CEIs of feasible solutions, we quickly identify a subset, denoted by $\mathscr{F}$, of feasible solutions for which the CEI should be computed.

3 ILLUSTRATION

We illustrate the performance of the aGMIA in Figure 1 by using a five-product inventory cost minimization problem with $N = 15^{10}$ feasible solutions. This problem is a variant of the well-known (s,S) inventory problem. We decompose the objective function into 5 groups, each representing a product.

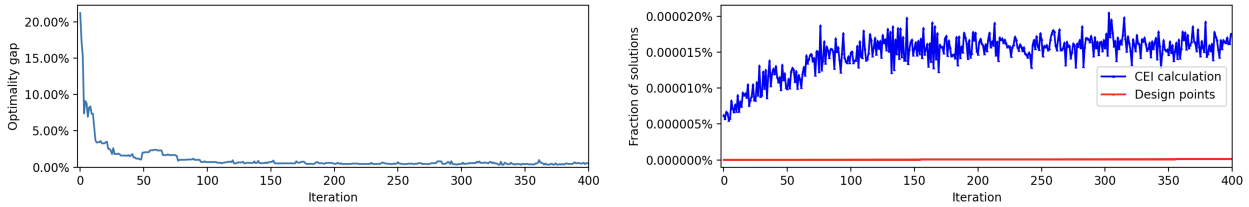

Figure 1: The left plot shows the optimality gap vs. the number of iterations. The right plot shows $|\mathscr{F}|/N$ (in blue) and $|\mathscr{D}|/N$ (in red) vs. the number of iterations. Although $|\mathscr{D}| < 300$ in the first 100 iterations, there is a rapid progress. At each iteration, the CEI is computed for no more than 120,000 feasible solutions.

REFERENCES

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.