A Gaussian Process Approach for High-Dimensional Simulation-based Transportation Optimization

Timothy Tay
Department of Civil and Environmental Engineering
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA, tayt@mit.edu

Carolina Osorio
Department of Civil and Environmental Engineering
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA, osorioc@mit.edu

Keywords: high-dimensional simulation-based optimization, Bayesian optimization, Gaussian processes

1. Introduction

Simulation-based urban traffic and mobility models are widely used by stakeholders to evaluate traffic management strategies or algorithms before implementation in the real-world. These simulators model the interactions between travelers by embedding detailed models of driver behavior. Concurrently, the simulators track numerous quantities of interest for every traveler in the network (Pell, Meingast, and Schauer 2017). To emulate this in terms of analytical expressions with the same accuracy is often very difficult. Hence, the level of detail provided by the simulators have made them popular tools for evaluating the performance of predetermined transportation strategies.

In recent times, there has been expanding interest in using simulation models with greater resolution and spatial coverage of urban networks to study the impacts of a transportation strategy across full city or metropolitan areas in detail. This translates into increasing computational demands of simulators. Furthermore, stakeholders are interested in jointly optimizing over entire urban networks, which leads to optimization problems of increasing dimensions. Despite significant increases in computational power over the years, this is outpaced by the demand needed to handle the simulators and higher dimensional optimization problems (Xu et al. 2016). The problem is exacerbated by the need for multiple simulation evaluations per iteration when considering stochastic models. Hence, this points to the need for more efficient optimization algorithms.

In this work, we consider high-dimensional simulation-based optimization (SO) problems that have continuous and general (i.e. non-convex) objective functions, with unknown analytical forms.

The constraints are assumed to be analytical and differentiable. Given the computation cost of evaluating the simulator, we also assume a limited computational budget (i.e.limited number of simulation runs). Since we are working with high-dimensional problems, the limited computational budget means that we are working in the regime with more decision variables than observations.

In this regime, it is crucial for the optimization algorithm to balance exploration and exploitation. In the context of optimization, exploration refers to searching in regions with few evaluated points, while exploitation refers to searching around the current best solution (Sun, Hong, and Hu 2014). Bayesian optimization with Gaussian process (GP) models provide a global optimization framework which systematically tries to balance the exploration-exploitation tradeoff (Jones, Schonlau, and Welch 1998), allowing it to be efficient in terms of the number of simulations needed. However, Bayesian optimization has been observed to be ineffective when tackling problems of high dimension (Kandasamy, Schneider, and Póczos 2015, Wang et al. 2016).

When designing efficient SO algorithms, analytical models can be used to provide prior information to quickly identify some good solutions. However, analytical models only approximates the objective function. As such, it is important that the optimization algorithm accounts for inaccuracies in the model. Our past work has focused on the formulation of metamodels which combine information from analytical models and the simulator, to approximate the objective function (Osorio and Bierlaire 2009, Osorio and Chong 2015). However, these past approaches do not explicitly balance exploration and exploitation.

To address these issues, we propose a method that combines ideas from Bayesian optimization, GPs and analytical modeling. Bayesian optimization provides a systematic way to balance exploitation and exploration, thus leading to more efficient optimization. At the same time, the use of GPs in combination with analytical models allows us make use of problem-specific information to guide the exploitation and exploration, potentially overcoming the challenges faced by Bayesian optimization and GPs in higher dimensions. Such a framework for optimization can be applied to a variety of transportation problems.

In the following section, we discuss the problem formulation and proposed method in more detail. In Section 3, we outline the empirical set up used to evaluate the proposed approach.

2. Proposed Method

Transportation SO problems can generally be formulated as:

$$\min_{\mathbf{x}\in\chi} f(\mathbf{x}, z; p) \equiv \mathbb{E}[F(\mathbf{x}, z; p)] \tag{1}$$

where f is the objective function, F represents the stochastic output of a simulation run, $\mathbf{x}$ is the high-dimensional vector of decision variables, χ is the feasible region, z denotes endogenous variables and p represents deterministic exogenous parameters.

In many transportation SO problems, we have access to some problem-specific information that can be used to formulate an analytical model. For instance, if we were working with an urban road network, we know the layout of the road network beforehand. Based on this knowledge, we can build an analytical model, such as a queueing network representation of the road network, which can then be used to approximate the objective function.

The analytical model can be used together with GPs, through the specification of the prior mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x'})$. Given a set of N previously evaluated points $X = [\mathbf{x_1}, \ldots, \mathbf{x_N}]^T$ and their function estimates $\mathbf{y} = (y_1, \ldots, y_N)$, the GP posterior estimate at a test point $\mathbf{x}_*$ is normally distributed with a predictive mean $\mu(\mathbf{x}_*)$ and variance $\sigma^2(\mathbf{x}_*)$ given by:

$$\mu(\mathbf{x}_*) = m(\mathbf{x}_*) + \mathbf{k}^T[\mathbf{K} + \tau^2 I]^{-1}(\mathbf{y} - m(\mathbf{x}_*)) \tag{2}$$

$$\sigma^2(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}^T[\mathbf{K} + \tau^2 I]^{-1}\mathbf{k} \tag{3}$$

$$\mathbf{k} = [k(\mathbf{x}_*, \mathbf{x_1}), \ldots, k(\mathbf{x}_*, \mathbf{x_N})] \tag{4}$$

and $\mathbf{K}$ is the covariance matrix for all pairs of points in the set X (see e.g. Williams and Rasmussen (2006)). We also assume that the function estimates $\mathbf{y}$ contain Gaussian noise with variance τ^2.

The most direct way to integrate problem-specific information into a GP model is to use an analytical model, $f^A(\mathbf{x})$, in the prior mean function of the GP: $m(\mathbf{x}) = \alpha f^A(\mathbf{x})$, where α is a scaling constant fitted based on the data. However, we also note that the posterior variance $\sigma^2(\mathbf{x})$ (Eq. (3)) represents uncertainty of the GP model, and plays an important role in driving the exploration of the feasible region. Based on Eq. (3), the uncertainty is only reduced in regions wtih nearby evaluated points. However, if a given region is known to have poor performance (e.g. as predicted by $f^A(\mathbf{x})$), it would also be desirable to have less uncertainty in the region, so that simulation budget is not wasted on evaluating points there. Instead, it would be beneficial to have greater uncertainty in unexplored regions with good predicted performance, so that the algorithm is encouraged to explore in these promising regions. This behavior can be achieved by incorporating the analytical model in the covariance function. For example, in this work, we propose a modified squared exponential covariance function with lengthscale ℓ and magnitude σ_0^2:

$$k(\mathbf{x}, \mathbf{x'}) = \frac{\sigma_0^2}{f^A(\mathbf{x}) + f^A(\mathbf{x'})} \exp\left(-\frac{1}{2\ell^2}||\mathbf{x} - \mathbf{x'}||_2^2\right) \tag{5}$$

Given the posterior GP model, the Bayesian optimization framework identifies the subsequent point to evaluate by maximizing an acquisition function. The acquisition function is defined in a way that tries to balance exploration and exploitation. It has a high value in regions where the posterior GP mean $\mu(\mathbf{x})$ is small (in a minimization problem) and in regions where the posterior variance $\sigma^2(\mathbf{x})$ is large. The expected improvement function (Jones, Schonlau, and Welch 1998) is a popular acquisition function. It has been shown to be effective in many studies, and is used in this work as well.

3. Empirical Analysis

The proposed method in Section 2 can be applied to a variety of continuous SO problems. In our work, we illustrate its effectiveness with a large-scale network-wide fixed cycle time traffic signal control problem, with the objective to minimize the expected trip travel time. The decision variables in this problem are the green splits (i.e. normalized green times) for each signal phase.

We consider a case-study using a simulation model of Midtown Manhattan. The Midtown Manhattan model is a large-scale network, and constitutes a high-dimensional optimization problem in the area of simulation optimization. It consists of 97 controlled intersections with a total of 259 signal phases. Using this simulation model, we compare the performance of our proposed method with other SO methods, in terms of the quality of proposed solutions, as well as the number of simulation evaluations required to obtain those solutions.

4. Conclusion

Optimization problems with simulation-based objective functions can be challenging to solve. The proposed method, combining Bayesian optimization, GPs and analytical modeling, shows promising preliminary results in finding good solutions quickly within the limited computational budget. There is also potential for it to be applied to a large variety of transportation problems.

Acknowledgments

The authors would like to thank the New York City Department of Transportation (NYCDOT), in particular Jingqin Gao, Mohamad Talas and Michael Marsico, for providing us with the simulation model for Midtown Manhattan. Timothy Tay would also like to thank the Agency for Science, Technology and Research (A*STAR) Singapore for funding his work.

References

Jones DR, Schonlau M, Welch WJ, 1998 *Efficient global optimization of expensive black-box functions. Journal of Global optimization* 13(4):455–492.

Kandasamy K, Schneider J, Póczos B, 2015 *High dimensional bayesian optimisation and bandits via additive models. International Conference on Machine Learning*, 295–304.

Osorio C, Bierlaire M, 2009 *An analytic finite capacity queueing network model capturing the propagation of congestion and blocking. European Journal of Operational Research* 196(3):996–1007.

Osorio C, Chong L, 2015 *A computationally efficient simulation-based optimization algorithm for large-scale urban transportation problems. Transportation Science* 49(3):623–636.

Pell A, Meingast A, Schauer O, 2017 *Trends in real-time traffic simulation. Transportation research procedia* 25:1477–1484.

Sun L, Hong LJ, Hu Z, 2014 *Balancing exploitation and exploration in discrete optimization via simulation through a gaussian process-based search. Operations Research* 62(6):1416–1438.

Wang Z, Hutter F, Zoghi M, Matheson D, de Feitas N, 2016 *Bayesian optimization in a billion dimensions via random embeddings. Journal of Artificial Intelligence Research* 55:361–387.

Williams CK, Rasmussen CE, 2006 *Gaussian processes for machine learning*, volume 2, chapter 2 (MIT press Cambridge, MA).

Xu J, Huang E, Hsieh L, Lee LH, Jia QS, Chen CH, 2016 *Simulation optimization in the era of industrial 4.0 and the industrial internet. Journal of Simulation* 10(4):310–320.