Using probabilistic information from macroscopic traffic models to derive computationally efficient simulation-based optimization algorithms

Carolina Osorio **

Department of Civil and Environmental Engineering,

Massachusetts Institute of Technology, Cambridge MA, USA

Hoda Bidkhori

Department of Civil and Environmental Engineering,

Massachusetts Institute of Technology, Cambridge MA, USA

*Email: osorioc@mit.edu

1 Introduction

This paper considers simulation-based optimization (SO) problems that rely on noisy, stochastic and computationally expensive evaluations of the underlying objective function. The focus of this paper is on the use of stochastic microscopic traffic simulators to solve nonlinear continuous constrained transportation problems. These are challenging SO problems. Additionally, we focus on developping methods with good short-term performance, that is we evaluate our algorithms under tight computational budgets (e.g. allowing only for a small and limited number of simulation runs or run-time).

This paper considers metamodel techniques for SO. Metamodels are simplified models of the underlying simulation model. The most common metamodels are analytical and deterministic functions, their functional form is typically chosen based on asymptotic properties or their computational efficiecy. Metamodels are typically used to approximate the stochastic objective function or constraints of the problem.

Metamodels have been classified as either functional (also called general-purpose) models and physical models (Søndergaard, 2003; Serafini, 1998). The latter are problemspecific approximations of the objective function or constraints. They have parameters that have a physical or structural interpretation. Reviews of metamodels include Barton and Meckesheimer (2006) and Kleijnen (2008).

In past work, we have proposed a metamodel that combines a physical metamodel with a functional metamodel (Osorio and Bierlaire, 2010). The functional component is a quadratic polynomial, which ensures asymptotic metamodel properties (needed to analyse asymptotic convergence properties), whereas the physical metamodel provides stuctural information about the underlying problem, and more importantly its functional form depends on the actual problem and objective function considered. The metamodel combines information from a low-resolution but computationally efficient analytical queueing model with high-resolution simulated data. The use of the combined metamodel has allowed us to achieve excellent short-term algorithmic performance.

The focus of this paper is to further enhance the short-term performance of SO algorithms. In this paper, we propose to use metamodels to go beyond the approximation of the problem formulation. We propose to use them to improve the point selection step (also known as selection procedure). At a given iteration, the algorithm must determine whether the newly identified point (called trial point) has improved performance compared to the point currently considered the best (called current iterate). This decision is called the point selection step.

2 Methodology

In this paper, we use the SO algorithm proposed in Osorio and Bierlaire (2007). We first describe the main steps of this algorithm, and detail its point selection step. We then present the novel point selection step.

The algorithm is a derivative-free trust region (TR) algorithm, which is based on the method proposed in Conn et al. (2009). For an introduction to TR methods, we refer the reader to Conn et al. (2000). They summarize the main steps of a TR method in the *Basic trust region algorithm*. The main idea of TR methods is to build, at each iteration, a model of the objective function which one "trusts" in a neighborhood of the current iterate (which is the point currently considered as the best), called the *trust region*.

The method proposed by Conn et al. (2009) builds upon the *Basic TR algorithm* by adding two additional steps: a model improvement step and a criticality step. For a detailed description, see Conn et al. (2009).

A given iteration k of the algorithm considers a metamodel m_k, an iterate x_k and a TR radius Δ_k. Hereafter, the subscript k refers to the iteration. Each iteration consists of 5 steps:

- **Criticality step.** This step may modify m_k and Δ_k if the measure of stationarity is close to zero.
- **Step calculation.** Approximately solve the TR subproblem to yield a trial point, which is a point that the metamodel predicts has improved performance compared to the current iterate.
- **Point selection step: acceptance or rejection of the trial point.** The actual (i.e. simulated) reduction of the objective function is compared to the reduction predicted by the model, this determines whether the trial point is accepted or rejected.
- **Model improvement.** Either certify that m_k is *fully linear* (i.e. satisfies Taylor-type bounds) in the TR or attempt to improve the accuracy of the metamodel.
- **TR radius update.**

Let x_k denote the current iterate, $x_k + s_k$ the current trial point, $\hat{f}(x_k)$ (respectively, $\hat{f}(x_k + s_k)$) the performance of the current iterate (resp. trial point) estimated by the simulator, $m_k(x_k)$ (resp. $m_k(x_k + s_k)$) the performance of the current iterate (resp. trial point) approximated by the metamodel. The simulated estimates are typically sample averages of the replications run at that point. The existing point selection step computes

$$\hat{\rho}_k = \frac{\hat{f}(x_k) - \hat{f}(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}.$$

If $\hat{\rho}_k \geq \eta_1$, then the trial point is accepted (i.e. it becomes the current iterate: $x_{k+1} = x_k + s_k$); otherwise it is rejected. The parameter η_1 is the trial point acceptance threshold.

The current algorithm merely compares sample averages, and thus does not account for the stochasticity of $\hat{f}(x_k)$ and $\hat{f}(x_k + s_k)$ when testing for improvement. In this paper, we replace the existing point selection step with a probabilistic metric that accounts for this stochasticity.

Accounting for this stochasticity is particularly important when the sample averages are computed based on small samples. This is typically the case when SO algorithms are used under tight computational budgets, e.g. when the number of simulation runs is limited and small. This is the case of the urban transportation applications that motivate this work.

The new point selection step proceeds as follows. If

$$Pr(\hat{\rho}_k \geq \eta_1) \geq p_0, \tag{1}$$

then we accept the trial point; otherwise we reject the trial point. The parameter p_0 is an exogenous threshold probability.

The main challenge we face is that this probability is usually estimated by combining sample average and sample variance information. Nonetheless, when the SO algorithms are used under tight computational budgets the number of replications of each point is very small (e.g. only a couple of replications for each point). Thus, the sample variance may be large, and thus the method is not effective at identifying trial points with improved performance.

To overcome this, we propose to interpret the probability in Equation (1) as that arising from a posterior distribution of a Bayesian framework. We use higher-order information from a probabilistic analytical metamodel to estimate the parameters of the corresponding prior distributions.

This paper uses the Bayesian framework proposed in Inoue (2000), along with information provided by the queueing model (i.e. the physical component of the metamodel) to fit the parameters of the prior distributions. We expect this to improve the effectiveness of the point selection step, particularly for small sample sizes. Thus, we expect this approach to improve the short-term performance of the considered SO algorithm.

3 Empirical analysis

To evalaute the short-term performance of this approach, we consider a fixed-time traffic signal control problem as formulated in detail in Osorio and Bierlaire (2007) and in Osorio and Bierlaire (2009).

We evaluate and illustrate the use of this framework with a case study based on the road network of the Swiss city of Lausanne. We use a calibrated microscopic traffic simulation

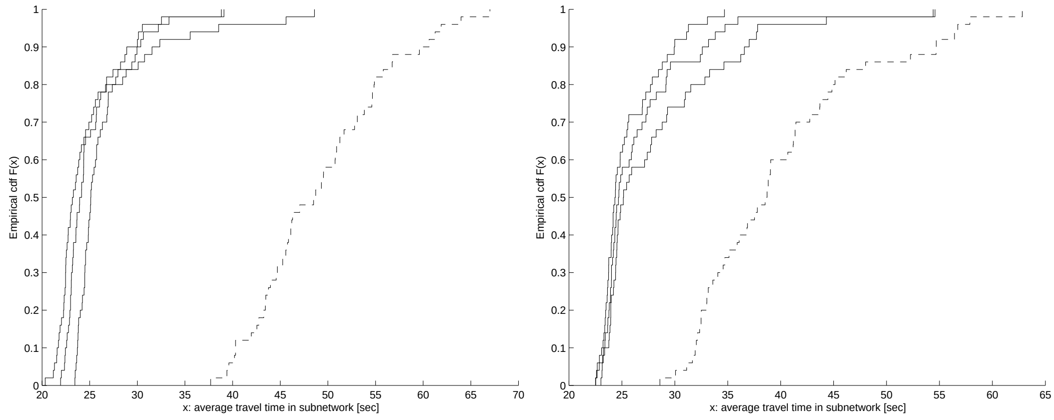

Figure 1: Empirical cdf's of the average travel times in the subnetwork.

model of the Lausanne city center. This model (Dumont and Bert, 2006) is implemented with the AIMSUN simulator (TSS, 2008). Details regarding the Lausanne network are given in Osorio and Bierlaire (2009). This network considers demand for the evening peak period.

We consider a set of 48 roads and 15 intersections. The signalized intersections have a cycle time of either 90 or 100 seconds. Nine intersections are signalized and control the flow of 30 roads.

We consider two different initial points, which are randomly drawn signal plans. For each initial point, we proceed as follows. We consider a tight computational budget, which is defined as a maximum number of simulation runs that can be carried out. The computational budget is set to 150 runs. We run each algorithm 3 times, and each time allow for these 150 simulation runs. This yields three different "optimal" solutions. We then use the simulator to evaluate in detail the performance of these solutions.

To evaluate the performance of a given signal plan, we run 50 replications of the simulation model, and plot the empirical cumulative distribution function (cdf) of the average travel times over these 50 runs. The empirical cdf's of the different signal plans are then compared.

Figures 1 and 2 display results for two different intial points. Figure 1 displays the cdf's of the average travel time in the controlled subnetwork. This figure contains two plots, one for each initial signal plan. Each plot displays 4 empirical cdf's. The solid curves are the cdf's of the signal plans proposed by our method. The dashed curve is the cdf of the initial signal plan. For both initial points, the proposed methodology systematically identifies

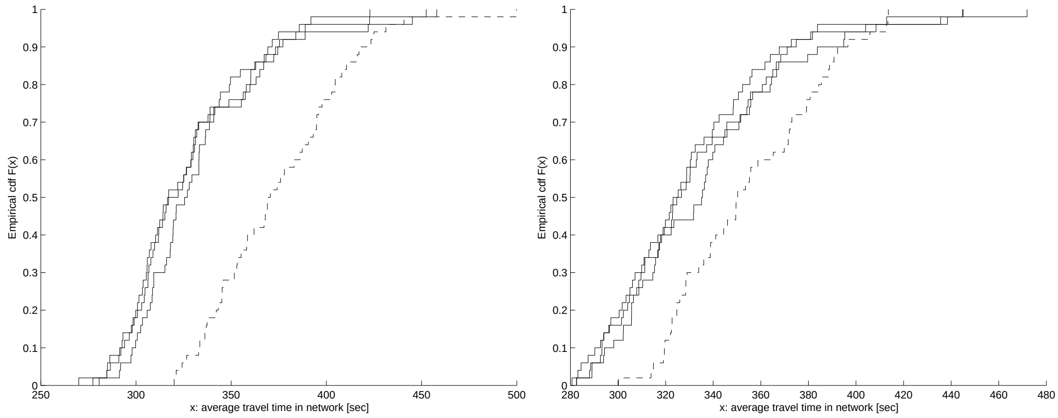

Figure 2: Empirical cdf's of the average travel times in the full network (entire city of Lausanne).

signal plans with improved performance even under tight computational budgets.

Figure 2 represents the results for the same 2 intial points, but displays the average travel time of the full network (i.e. accounting also for the roads otuside the subnetwork). This figure indicates that the proposed plans systematically provide improvement of travel times at the full city-scale.

References

Barton, R. R. and Meckesheimer, M. (2006). Metamodel-based simulation optimization, *in* S. G. Henderson and B. L. Nelson (eds), *Handbooks in operations research and management science: Simulation*, Vol. 13, Elsevier, Amsterdam, chapter 18, pp. 535–574.

Conn, A. R., Gould, N. I. M. and Toint, P. L. (2000). *Trust-region methods*, MPS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics and Mathematical Programming Society, Philadelphia, PA, USA.

Conn, A. R., Scheinberg, K. and Vicente, L. N. (2009). Global convergence of general derivative-free trust-region algorithms to first- and second-order critical points, *SIAM Journal on Optimization* **20**(1): 387–415.

Dumont, A. G. and Bert, E. (2006). Simulation de l'agglomération Lausannoise SIMLO, *Technical report*, Laboratoire des voies de circulation, ENAC, Ecole Polytechnique Fédérale de Lausanne.

Inoue, K. (2000). *Decision-theoretic comparison of alternate system configurations using stochastic simulation*, PhD thesis, University of Michigan.

Kleijnen, J. P. C. (2008). *Design and Analysis of Simulation Experiments*, Springer, New York, USA.

Osorio, C. and Bierlaire, M. (2007). An analytic finite capacity queueing network capturing congestion and spillbacks, *Proceedings of the Sixth Triennial Symposium on Transportation Analysis, TRISTAN VI*, Phuket, Thailand.

Osorio, C. and Bierlaire, M. (2009). A surrogate model for traffic optimization of congested networks: an analytic queueing network approach, *Technical Report 090825*, Transport and Mobility Laboratory, ENAC, Ecole Polytechnique Fédérale de Lausanne.

Osorio, C. and Bierlaire, M. (2010). A simulation-based optimization approach to perform urban traffic control, *Proceedings of the Triennial Symposium on Transportation Analysis (TRISTAN)*, Tromsø, Norway.

Serafini, D. B. (1998). *A framework for managing models in nonlinear optimization of computationally expensive functions*, PhD thesis, Rice University.

Søndergaard, J. (2003). *Optimization using surrogate models - by the Space Mapping technique*, PhD thesis, Technical University of Denmark.

TSS (2008). *AIMSUN NG and AIMSUN Micro Version 5.1*, Transport Simulation Systems.