A simulation optimization framework for the management of congested urban road networks

Carolina Osorio

Michel Bierlaire

STRC 2009 **September 2009**

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

STRC 2009

A simulation optimization framework for the management of congested urban road networks

Carolina Osorio
Transport and Mobility Laboratory
EPFL
1005 Lausanne
phone: +41 21 693 9327
fax: +41 21 693 8060
carolina.osoriopizano@epfl.ch

Michel Bierlaire
Transport and Mobility Laboratory
EPFL
1005 Lausanne
phone: +41 21 693 2537
fax: +41 21 693 8060
michel.bierlaire@epfl.ch

September 2009

Abstract

Deriving optimal traffic management schemes for urban road networks typically relies on the use of complex simulation tools, that capture in detail the behavior of drivers as well as their interaction with the network infrastructure. The integration of these traffic simulators within an optimization framework is an intricate task. Indeed, these simulators can be seen as stochastic nonlinear functions that are expensive to evaluate.

Simulation-based network optimization should therefore start with an important modeling effort, in order to exploit the structure of the problem at hand. In particular, we believe that in order to perform both fast and reliable simulation-based optimization for congested networks, information from the simulation tool should be combined with information from a metamodel (surrogate) that captures at a lower degree of detail the structure of the underlying problem.

In this paper, we propose a surrogate that combines information from a calibrated microscopic traffic simulation model with an analytical queueing network model. We integrate this surrogate within a derivative-free trust region optimization framework. We apply the framework to solve a fixed-time traffic signal control problem for a subnetwork of the Lausanne city center. We compare the performance of the derived signal plans with that of an existing signal plan for the city of Lausanne.

Keywords

Simulation-Optimization, Metamodel, Traffic control, Queueing

1 Introduction

Deriving optimal traffic management schemes for urban road networks typically relies on the use of microscopic simulation tools that capture in detail the behavior of drivers as well as their interaction with the network infrastructure. These simulation tools can provide accurate network performance estimates in the context of scenario-based analysis or sensitivity analysis. Nevertheless, their integration within an optimization framework remains an intricate process. A given traffic management scheme can be formulated as:

$$\min_{x,z \in \Omega} E[f(x, z, p, \epsilon)],$$

where the objective is to minimize the expected value of a suitable network performance measure f. This performance measure is a function of a decision or control vector x, endogenous variables z, exogenous parameters p and a random component ϵ. The feasible space Ω consists of a set of constraints that link x to z, p and f. For instance, a traffic signal control problem can take f as the travel time and x as the green splits for the signalized lanes. Elements such as the total demand or the network topology will be captured by p, while the distribution of the demand (route choice decisions) and the capacities of the signalized lanes will be captured by z. The random component ϵ describes the noise associated with a given realization of f.

The various traffic models embedded within the simulator make it a detailed and realistic model, but lead to nonlinear objective functions with no available closed form, and containing potentially several local minima. Since these are stochastic models, we can only derive estimates of $E[f]$. Additionally, computing these estimates is computationally expensive, since they involve running numerous replications. As a nonlinear stochastic and evaluation-expensive problem, it is complex to address. In practice, the aim of simulation optimization (SO) problems is to identify improved settings, rather than seek or proove optimality.

We believe that in order to perform both fast and reliable simulation optimization for congested networks, information from the simulation tool should be combined with information from a surrogate network model that analytically captures the structure of the underlying problem. In this paper, we propose such a surrogate. First, we present a literature review of surrogate-based SO methods (Section 2). In Section 3 we present the optimization framework and the surrogate model. We then show how this method applies to a fixed-time traffic signal optimization problem (Section 4). We comment on implementation issues (Section 5) and present empirical results in Section 6.

2 Literature review

Barton and Meckesheimer (2006) provide a classification and a review of simulation-optimization methods. Continuous SO problems fall into two categories: direct gradient and metamodel methods. Direct gradient methods estimate the gradient of the simulation response, and then resort to stochastic gradient-based techniques such as stochastic approximation (Spall, 2003). These methods do not attempt to fit a global approximation to the objective function. The simulation function's gradient can be estimated with direct methods (e.g. perturbation analysis), which require knowledge of the underlying probabilistic process (e.g. input probability distributions). In particular, automatic differentiation methods allow the exact evaluation of gradients but require the source code of the simulation model to be available (see Conn *et al.* (2000) and references herein). The gradient can also be estimated with indirect methods, which use only function evaluations (e.g. finite difference, simultaneous perturbation (Spall, 2003). Although there have been significant advances and novel approaches for gradient estimation (Fu *et al.*, 2005; Fu, 2006), methods that rely on direct derivative information often require more function evaluations, and their convergence is sensitive to the accuracy of the gradient estimation.

Metamodel methods use an indirect-gradient approach by computing the gradient of a surrogate model (or metamodel), which is a deterministic function, instead of the gradient of the simulation response. The main advantage of a metamodeling approach is that the stochastic response of the simulation is replaced by a deterministic metamodel response function, then deterministic optimization techniques can be used. Metamodels are often a linear combination of basis functions from a parametric family. The most common approach is the use of low-order polynomials (e.g. linear or quadratic). Spline models have also been used, although their use within an SO framework has focused on univariate or bivariate functions, and as Barton and Meckesheimer (2006) mention: "unfortunately, the most popular and effective multivariate spline methods are based on interpolating splines, which have little applicability for SO". Radial basis functions have also been proposed (Oeuvray and Bierlaire, 2009). The existing metamodel methods fix apriori a functional form for the metamodel (e.g. quadratic). The functional forms considered are general-purpose forms, that are chosen based on their analytical tractability, but do not take into account any information with regards to the specific objective function, let alone the structure of the underlying problem.

In this paper, we use a metamodel method to perform SO. The metamodel of interest combines information from the simulator and from an analytical network model. For a given problem, the analytical model will yield a different functional form for the objective function. The metamodel proposed in this paper goes beyond existing metamodel approaches since the functional form is problem specific. This comes at the cost of deriving a framework that is particularly suited for network optimization but not intended for arbitrary optimization problems.

In order to integrate the proposed metamodel into an existing optimization method, we review the algorithms that allow for an arbitrary metamodel. These methods are called multi-model or hybrid methods. They share a common motivation, which is to combine the use of models with varying evaluation costs (low versus high-fidelity models, or coarse versus fine models).

A trust-region optimization framework for unconstrained problems allowing for multiple models was proposed by Carter (1986) (see references herein for previous multi-model frameworks). His work analyses the theoretical properties and derives a global convergence theory for several types of multi-model algorithms. It allows for nonquadratic models as long as at least one model is a standard quadratic with uniformly bounded curvature.

The Approximation and Model Management Optimization/Framework (AMMO or AMMF) is a trust-region framework for generating and managing a sequence of metamodels. There are several versions of the algorithm: for unconstrained problems (Alexandrov *et al.*, 1998), bound constrained (Alexandrov *et al.*, 2000), inequality constrained (Alexandrov *et al.*, 1999), generally constrained (Alexandrov *et al.*, 2001). Although no restrictions are imposed on the type of surrogates allowed, it is a first-order method that requires that the model and the objective function, as well as their first-order derivatives, coincide at each major (or accepted) iterate. Thus the metamodel must always behave as a first-order Taylor series approximation. This is a strong restriction if the function is noisy and expensive to evaluate.

The Surrogate-Management framework (SMF) proposed by Booker *et al.* (1999) is a derivative-free method for bound constrained problems. It is based on a direct search technique called pattern search. Since direct search techniques typically require many function evaluations, they use a surrogate model of the objective function to improve the performance of the algorithm. The surrogate model used is an interpolated kriging model. Nevertheless, interpolation techniques are inappropriate for noisy responses.

The Space Mapping (SM) technique and its many versions (Bandler *et al.*, 2006, 2004) is a simulation-based optimization technique that uses two metamodels: a fine and a coarse model. Both models are often simulation-based. The coarse model is constructed based on a transformation of the endogenous variables ("space mapping") that minimizes the error for a sampled set of high-fidelity response values. Nevertheless, SM relies on the assumption that via a transformation of the endogenous variables the coarse model will exhibit the physical/mathematical properties of the fine model (Alexandrov and Lewis, 2001) and as Bandler *et al.* (2004) mention "the required interaction between coarse model, fine model, and optimization tools makes SM difficult to automate within existing simulators". Alexandrov and Lewis (2001) give a comparison of the AMMO, the SMF and the SM methods.

Conn *et al.* (2009a) recently proposed a trust-region derivative-free framework for unconstrained problems. This framework allows for arbitrary metamodels and makes no assumption on how these metamodels are fitted (interpolation or regression). To ensure global convergence a model improvement algorithm guarantees that the models have a uniform local behavior (i.e. satisfy Taylor-type bounds) in a finite number of steps.

Derivative-free (DF) methods do not require nor do they explicitly approximate derivatives. Resorting to a DF algorithm, rather than to first or second order algorithms, is therefore appropriate for noisy problems where the derivatives are difficult to obtain and often unreliable. This is also the case when the evaluation of the objective function is computationally expensive, or when the simulation source code is unavailable, the simulator must then be treated as a black box (Moré and Wild, 2009). In the field of transportation, the simulators fall into all three of these categories. Thus we will opt for a DF approach.

Among the two main strategies used to ensure global convergence, line search and trust region methods, the latter are more appropriate for our context since they "extend more naturally than line search methods to models that are not quadratics with positive Hessians" (Carter, 1986). Trust-region (TR) methods when both first and second-order derivatives are unavailable is a relatively recent topic (see Conn *et al.* (2009b) for references). Additionally, the most common approach for fitting metamodels within a TR framework is interpolation. Nevertheless, for noisy functions we believe that regression is more appropriate since it is less sensitive to the inaccuracy of the observations.

The framework proposed by Conn *et al.* (2009a), as a derivative-free TR method that allows for arbitrary models and does not impose interpolation, is therefore particularly appealing. We will therefore integrate our metamodel within this framework.

3 Method

In this section, we first describe the main ideas of the optimization algorithm that will be used. We then present the metamodel.

3.1 Algorithmic franework

For an introduction to trust region (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 method proposed by Conn *et al.* (2009a) builds upon the *Basic TR algorithm* by adding two additional steps: a model improvement step and a criticality step. We present the main steps of the algorithm. For a detailed description see Conn *et al.* (2009a). A given iteration k of the algorithm considers a metamodel m_k, an iterate x_k and a TR radius Δ_k. 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.
- **Acceptance of the trial point.** The actual 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* in the TR or carry out improvement steps.
- **TR radius update.**

3.2 Metamodel

The metamodel combines information from two models: a simulation model and an analytical queueing model. We first present these two models, we then describe how they are combined.

Simulation model. We use a calibrated microscopic traffic simulation model of the Lausanne city center. A detailed description of this model is given in Dumont and Bert (2006). It is implemented with the AIMSUN simulator (TSS, 2008). It contains a total of 652 roads and 231 intersections, 49 of which are signalized. For a given decision vector x the simulator provides a realization of the random variable $f(x, z, p, \epsilon)$.

Analytical queueing model. This model resorts to *finite capacity queueing theory* to capture the key traffic dynamics and the underlying network structure, e.g. how upstream and downstream queues interact, how this interaction is linked to network congestion. The model consists of a system of nonlinear equations. It is formulated based on a set of exogenous parameters θ that capture the network topology, the total demand, as well as the turning probabilities. A set of endogenous variables y describe the traffic dynamics, e.g. spillback probabilities, the average rates at which a spillback diffuses, queue length stationary distributions. For a given decision vector x the network model yields the objective function $T(x, y, \theta)$.

A detailed description of the queueing model and a case study illustrating how the endogenous variables describe the formation and diffusion of congestion is given in Osorio and Bierlaire (2009a). Its formulation for an urban road network appears in Osorio and Bierlaire (2009b). It has been successfully used to solve a fixed-time traffic signal control problem in Osorio and Bierlaire (2009b).

We recall here the notation that we have introduced so far:

x	decision vector;
T	estimate of the objective function derived by the queueing model;
f	simulation response;
y	endogenous queueing model variables;
θ	exogenous queueing model parameters;
z	endogenous simulation variables;
p	exogenous simulation parameters;
ϵ	random component of the simulation response.

We now describe how f and T are combined to derive the metamodel m. The functional form of m is:

$$m(x, y, \theta, \alpha, \beta) = \alpha T(x, y, \theta) + \phi(x, \beta),$$

where ϕ is a quadratic polynomial, α and β are parameters of the metamodel. The polynomial ϕ is quadratic with diagonal second derivative matrix. This choice is based on existing numerical experiments for derivative-free TR methods which show that they are often more efficient than full quadratic models (Powell, 2003).

$$\phi(x,\beta) = \beta_0 + \sum_{j=1}^{d} \beta_j x_j + \sum_{j=1}^{d} \beta_{p+j} x_j^2,$$

where d is the dimension of x, and x_j is the j^{th} component of x.

At a given iteration k of the algorithm (described in Section 3.1), the parameters β and α of the metamodel are fitted using the current sample by solving the least squares problem:

$$\min_{\alpha,\beta} \sum_{i=1}^{n_k} (w_{ki}(\hat{f}(x^i, z^i, p, \epsilon^i) - m(x^i, y^i, \theta, \alpha, \beta)))^2,$$

where x^i represents the i^{th} point in the sample, with the corresponding simulated observation $\hat{f}(x^i, z^i, p, \epsilon^i)$, n_k is the sample size and w_{ki} is the weight associated to the i^{th} observation at iteration k.

The weights capture the importance of each point with regards to the current iterate. The work of Atkeson *et al.* (1997) gives a survey of weight functions and analyzes their theoretical properties. We use what is known as the *inverse distance* weight function, along with the Euclidean distance, this leads to the following weight parameters:

$$w_{kj} = \frac{1}{1 + \|x_k - x_j\|_2^2}$$

The weight of a given point is therefore inversely proportional to its distance from the current iterate. This will allow us to approximately have a Taylor-type behavior, where local points have more weight.

The least squares problem is solved using the Matlab routine *lsqnonlin* (The Mathworks, 2008).

4 Optimization Problem

4.1 Traffic signal control

We illustrate the use of this framework with a signal control problem for a subnetwork of the city of Lausanne. A review of the different formulations, as well as the definitions of the traffic signal terms used hereafter, is given in Appendix A of Osorio and Bierlaire (2009b). We consider a fixed-time signal control problem where the offsets, the cycle times and the all-red durations are fixed. The stage structure is also given. In other words, the set of lanes associated with each stage as well as the sequence of stages are both known. To formulate this problem we use the following notation:

b_i	available cycle ratio of intersection i;
$x(p)$	green split of phase p;
x_L	vector of minimal green splits for each phase;
$\mathcal{I}$	set of intersection indices;
$\mathcal{P}_{\mathcal{I}}(i)$	set of phase indices of intersection i.

The problem is traditionally formulated as follows:

$$\min_{x,z} E[f(x, z, p, \epsilon)] \tag{1}$$

subject to

$$\sum_{p \in \mathcal{P}_{\mathcal{I}}(i)} x(p) = b_i, \ \forall i \in \mathcal{I} \tag{2}$$

$$x \geq x_L. \tag{3}$$

In this problem the decision vector x consists of the green splits for each phase. The objective is to minimize the expected travel time (Equation (1)). The linear constraints (2) link the green times of the phases with the available cycle time for each intersection. The bounds (3) correspond to minimal green time values for each phase. These have been set to 4 seconds according to the Swiss standard (VSS, 1992).

4.2 TR subproblem

At a given iteration k the TR subproblem includes three more constraints than the previous problem. It is formulated as follows:

$$\min_{x,y} \; m_k(x, y, \theta, \alpha_k, \beta_k) \tag{4}$$

subject to

$$\sum_{p \in \mathcal{P}_{\mathcal{I}}(i)} x(p) = b_i, \; \forall i \in \mathcal{I} \tag{5}$$

$$\ell(x, y, \theta) = 0 \tag{6}$$

$$\|x - x_k\|_2 \leq \Delta_k \tag{7}$$

$$y \geq 0 \tag{8}$$

$$x \geq x_L, \tag{9}$$

where x_k is the current iterate and ℓ denotes the queueing model. Equation (6) consists of the system of nonlinear equations that define the queueing model, the corresponding endogenous variables are subject to positivity constraints (Equation (8)). This system is given explicitly and detailed in Osorio and Bierlaire (2009b) (Equations (9), (10) and (12) of that paper). The analytical form of T is also detailed in Section 4 of that paper. Constraint (7) is the TR constraint. It uses the Euclidean norm (Conn *et al.*, 2009a). Thus the TR subproblem consists of a nonlinear objective function subject to nonlinear and linear equalities, a nonlinear inequality and bound constraints. This problem is solved with the Matlab routine for constrained nonlinear problems, *fmincon*, which resorts to a sequential quadratic programming method (Coleman and Li, 1996, 1994).

5 Implementation notes

Constraints As described in Section 2, DF TR methods are a relatively recent topic (Conn *et al.*, 2009b). The algorithms developed so far are derived based on sound theoretical properties that lead to a solid global convergence theory, but they are mostly formulated for unconstrained problems. Unfortunately, the optimization problems encountered in practice are rarely unconstrained. Conn *et al.* (2009b) reviews constrained DF algorithms, and confirms that for constrained problems "currently, there is no convergence theory developed for TR interpolation-based methods", not to mention TR methods that allow for regression models. Conn *et al.* (1998) extends the use of a TR method for unconstrained problems to problems with general constraints. The traffic management problems that we are interested in solving fall into the category of what they denote as *easy* constraints. These are general constraints that are continuously differentiable and whos first order partial derivatives can be computed relatively cheaply (with regards to the cost of evaluating the objective function). In their approach they include such constraints in the TR subproblem, which ensures that all trial points are feasible. Conn *et al.* (2009b) mention that such an approach is often sufficient in practice. Here we use the method proposed by Conn *et al.* (2009a) for unconstrained methods, and extend its use to constrained problems as Conn *et al.* (1998) propose.

Limited computational budget The main motivation to go beyond a pure quadratic surrogate is to improve the short term performance of a given DF algorithm, since near convergence a quadratic will asymptotically provide an adequate approximation for a second-order Taylor series model. Recently, the importance of evaluating the short-term behavior of DF algorithms has been emphasized by Moré and Wild (2009) and Zhang *et al.* (2009). Furthermore, DF applications often involve a limited computational budget. In many practical situations an improved solution rather than a local optimum may be all that is required or that can be computed for a given budget (Zhang *et al.*, 2009). We will therefore focus on the performance of this approach given a fixed and tight computational budget.

Criticality step Since we are interested in the short term behavior of this approach, the theoretical considerations needed to ensure global convergence are not our main focus. We assume that the limited resources are not sufficient to approach an optimal point, i.e. the measure of stationarity will not go under a given threshold. Thus we do not consider the criticality step of the original algorithm. We assume throughout that the model is not *certifiably fully linear* (which is required when approaching a stationary point so that the stationary measure of the model can be trusted). If at a given iteration, the measure of stationarity does go under this threshold then a purely quadratic metamodel can be used (so that within a finite number of steps we can ensure that it will satisfy Taylor-type bounds).

Model improvement step At each iteration we obtain one observation of the simulated objective function (associated to the trial point), no further improvement steps are carried out. In order to improve the performance of the algorithm, diversification sampling should be carried out. Determining when and how this diversification should take place is currently being studied.

TR radius update There are 2 cases where the TR radius is reduced in the algorithm: (1) if it is known that the model is *fully linear*, but it has over-predicted the reduction in the objective function; (2) when approaching a stationary point (so that the model becomes more accurate and the stationary measure can be trusted). Since we assume throughout that the model is not *certifiably fully linear* and we focus on the short-term performance of the algorithm, the TR radius is never reduced in this implementation.

Initial sample Since our focus is on problems with a limited and tight computational budget, we assume that there are no initial observations available. Although the least squares routine used allows for underdetermined systems, which in our case occur when the dimension of the sampled space is smaller than the number of parameters to estimate $2d+2$, we use augmented data to make the least-squares matrix of full rank. These artificial observations are chosen so that the parameters are near an initial value (chosen as zero) and are attributed a small weight (10^{-2}).

Algorithmic parameters The following values are used for the parameters of the TR algorithm: $\Delta_0 = 10^3, \Delta_{max} = 10^{10}, \eta_1 = 10^{-3}, \gamma_{inc} = 1.2$. Typical values for TR parameters are given in Carter (1986). For the algorithm used to solve the TR subproblem we set the tolerance for relative change in the objective function to 10^{-3} and the constraint tolerance to 10^{-2}. We limit the computational budget to 50 iterations, and use a random feasible point as the initial point.

6 Empirical Analysis

We now evaluate the performance of the proposed method by considering a subnetwork of the Lausanne city center. The subnetwork (Figure 1) contains 48 roads and 15 intersections. Nine intersections are signalized and control the flow of 30 roads. There are a total of 51 phases that are considered variable. The intersections have a cycle time of either 90 or 100 seconds. The considered demand scenario consists of the evening peak period (17h-18h). Within this time period congestion gradually increases.

The queueing model of this subnetwork consists of 102 queues. The TR subproblem consists of 621 endogenous variables with their corresponding lower bound constraints, 408 nonlinear equality constraints, 171 linear equality constraints and 1 nonlinear inequality constraint.

For a given computational budget, our method yields an 'optimal' signal plan for the subnetwork. We then use the simulation model to evaluate the effect of this signal plan upon the entire Lausanne network. We run 100 replications to evaluate the performance of these 'optimal' plans. Each replication is preceded by a 15 minute warm-up period.

We compare the performance of the plans derived by this method with that of an existing

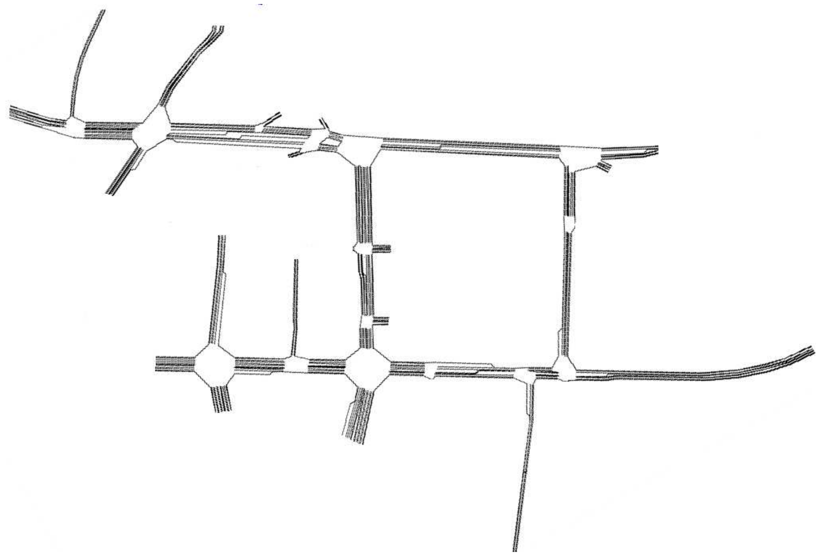

Figure 1: Subnetwork of the Lausanne city center

signal plan for the city of Lausanne. For more information concerning this existing control plan we refer the reader to Dumont and Bert (2006). It is quite a challenge to compare to this existing plan, since its a coordinated plan (i.e. green waves exist on the main arterials).

Figure 2 displays the empirical cumulative distribution function (cdf) of the average travel times across the 100 replications for four signal plans. The two thin solid lines correspond to the 'optimal' plans derived by the proposed method, the thick solid line corresponds to the existing plan, and the two dotted lines correspond to the random initial plans. The plans derived based on the first initial plan are labeled on the figure as x_1. The labeled cdf's show that starting off from a poorly performing initial point, our model leads to a plan with very good performance. The other initial point has a performance similar to that of the existing plan. Our method still yields a minor improvement. By comparing the performance of the plans derived by the proposed method to that of the existing plan, these preliminary results illustrate the added value of our approach. With no initial sample, and a tight computational budget, our method is able to identify signal plans that improve the distribution of the average travel time.

We have also run the algorithm using a purely quadratic metamodel. Nevertheless, as mentioned in Section 5 the algorithm is initialized with no initial sample and a diversification strategy has not yet been integrated. Thus the method based on a purely quadratic model does not search at all the feasible space, and yields as 'optimal' points the initial random points. Without a diversification strategy comparing these two metamodel methods directly is of little interest.

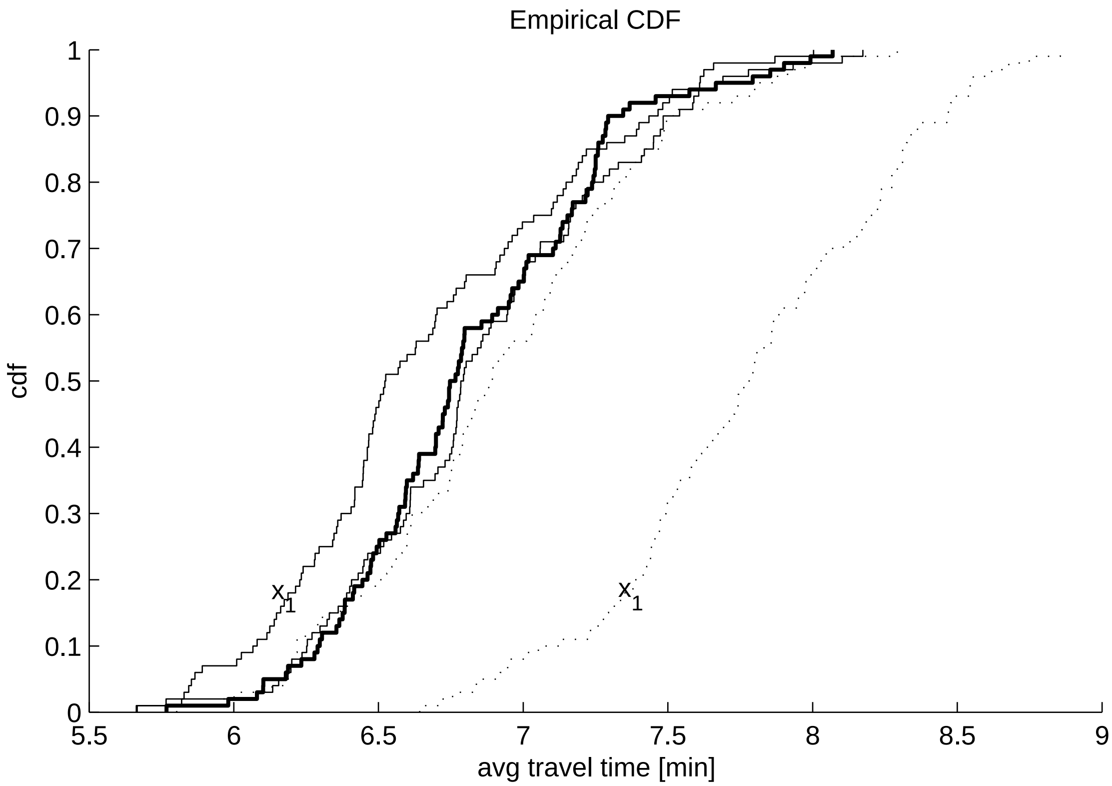

Figure 2: Empirical cumulative distribution function of the average travel time

7 Conclusion

This paper presents an simulation optimization framework for the management of congested networks. It proposes a metamodel that combines information from a traffic simulation tool and an analytical network model. The framework is illustrated by solving a fixed-time signal control problem for a subnetwork of the Lausanne city center. The performance of the derived plans is compared to that of an existing plan for the city of Lausanne. Although the method is run with no initial sample and a tight computational budget, it derives well performing signal plans.

These are preliminary results, but they indicate that this approach may be suitable for high dimensional problems (more than 100 variables) that would otherwise require a large sample size to initially fit the metamodel of interest. Efficiently tackling constrained high dimensional problems is one of the main limitations of existing DF methods. The main component of this methodology that we are currently working on, is the definition of a diversification sampling strategy, that would refine the model improvement step of the algorithm. Furthermore, the sensitivity of the method to the numerous algorithmic parameters needs to be evaluated.

Acknowledgments

This research was supported by the Swiss National Science Foundation grant 205320-117581.

References

Alexandrov, N. M., J. E. Dennis, R. M. Lewis and V. Torczon (1998) A trust region framework for managing the use of approximation models in optimization, *Structural Optimization*, **15**, 16–23.

Alexandrov, N. M. and R. M. Lewis (2001) An overview of first-order model management for engineering optimization, *Optimization and Engineering*, **2**, 413–430.

Alexandrov, N. M., R. M. Lewis, C. R. Gumbert, L. L. Green and P. A. Newman (1999) Optimization with variable-fidelity models applied to wing design, *Technical Report*, **CR-1999-209826**, NASA Langley Research Center, Hampoton, VA, USA, December 1999.

Alexandrov, N. M., R. M. Lewis, C. R. Gumbert, L. L. Green and P. A. Newman (2001) Approximation and model management in aerodynamic optimization with variable-fidelity models, *Journal of Aircraft*, **38** (6) 1093–1101.

Alexandrov, N. M., E. J. Nielsen, R. M. Lewis and W. K. Anderson (2000) First-order model management with variable-fidelity physics applied to multi-element airfoil optimization, paper presented at *Proceedings of the 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA, USA, September 6-8 2000.

Atkeson, C. G., A. W. Moore and S. Schaal (1997) Locally weighted learning, *Artificial Intelligence Review*, **11**, 11–73.

Bandler, J. W., Q. Cheng, A. Dakroury, A. Mohamed, M. Bakr, K. Madsen and J. Søndergaard (2004) Space mapping: The state of the art, *IEEE Transactions on Microwave Theory and Techniques*, **52** (1) 337–360.

Bandler, J. W., S. Koziel and K. Madsen (2006) Space mapping for engineering optimization, *SIAG/Otimization Views-and-News*, **17** (1) 19–26.

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

Booker, A. J., J. E. Dennis, P. D. Frank, D. B. Serafini, V. Torczon and M. W. Trosset (1999) A rigorous framework for optimization of expensive functions by surrogates, *Structural Optimization*, **17**, 1–13.

Carter, R. G. (1986) Multi-model algorithms for optimization, Ph.D. Thesis, Rice University.

Coleman, T. F. and Y. Li (1994) On the convergence of reflective newton methods for large-scale nonlinear minimization subject to bounds, *Mathematical Programming*, **67** (2) 189–224.

Coleman, T. F. and Y. Li (1996) An interior, trust region approach for nonlinear minimization subject to bounds, *SIAM Journal on Optimization*, **6**, 418–445.

Conn, A. R., N. I. M. Gould and P. L. Toint (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., K. Scheinberg and P. L. Toint (1998) A derivative free optimization algorithm in practice, paper presented at *Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, St. Louis, MO, USA, September 2-4 1998.

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

Conn, A. R., K. Scheinberg and L. N. Vicente (2009b) *Introduction to derivative-free optimization*, MPS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics and Mathematical Programming Society, Philadelphia, PA, USA.

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

Fu, M. C. (2006) Gradient estimation, in S. G. Henderson and B. L. Nelson (eds.) *Handbooks in operations research and management science: Simulation*, vol. 13, chap. 19, 576–616, Elsevier, Amsterdam.

Fu, M. C., F. W. Glover and J. April (2005) Simulation optimization: a review, new developments, and applications, paper presented at *Proceedings of the 2005 Winter Simulation Conference*, 83–95, Piscataway, New Jersey, USA.

Moré, J. and S. Wild (2009) Benchmarking derivative-free optimization algorithms, *SIAM Journal on Optimization*, **20** (1) 172–191.

Oeuvray, R. and M. Bierlaire (2009) Boosters: a derivative-free algorithm based on radial basis functions, *International Journal of Modelling and Simulation*, **29** (1) 26–36.

Osorio, C. and M. Bierlaire (2009a) 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. and M. Bierlaire (2009b) 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, August 2009.

Powell, M. J. D. (2003) On trust region methods for unconstrained minimization without derivatives, *Mathematical Programming*, **97** (3) 605–623.

Spall, J. C. (2003) *Introduction to stochastic search and optimization: estimation, simulation, and control*, Wiley-Interscience series in discrete mathematics and optimization, John Wiley & Sons, New Jersey, USA.

The Mathworks, I. (2008) *Optimization Toolbox Version 4. User's Guide Matlab*, Natick, MA, USA.

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

VSS (1992) *Norme Suisse SN 640837 Installations de feux de circulation; temps transitoires et temps minimaux*, Union des professionnels suisses de la route, VSS, Zurich, Mai 1992.

Zhang, H., A. Conn and K. Scheinberg (2009) A derivative-free algorithm for the least-square minimization, *Technical Report*, **CCT-TR-2009-8**, Louisiana State University, Center for Computation and Technology, April 2009.