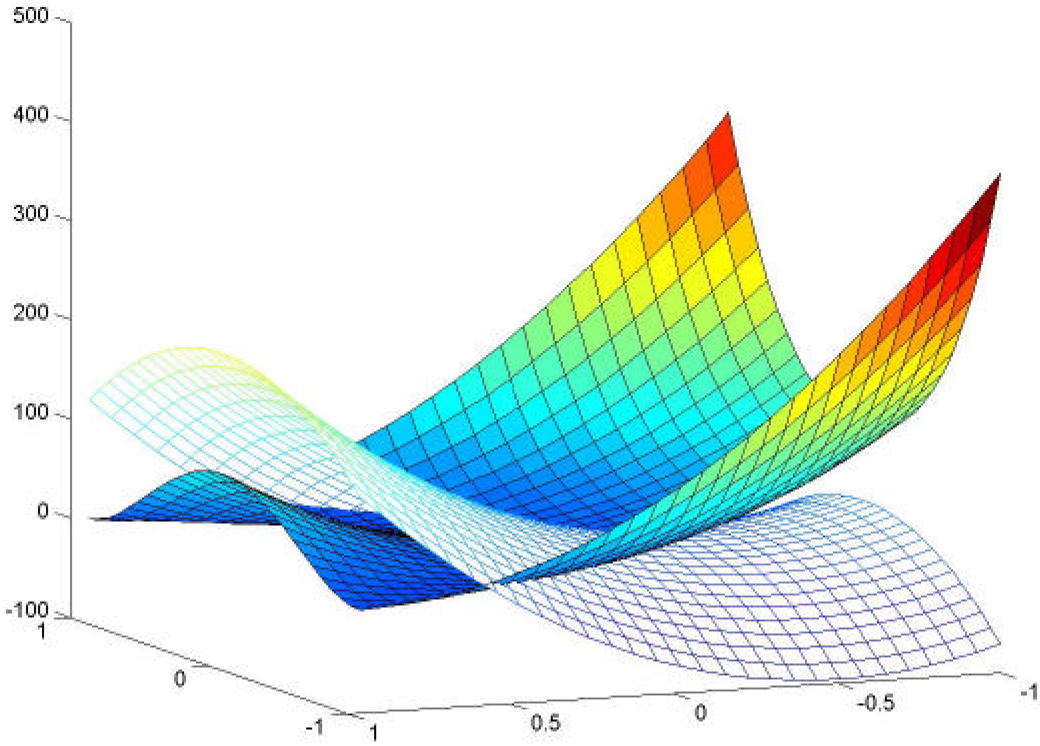


New Algorithmic Approaches for the Anticipatory Route Guidance Generation Problem

Frank Crittin, roso-dma-epfl
Michel Bierlaire, roso-dma-epfl

Conference paper STRC 2001
Session ITS

STRC | 1st Swiss Transport Research Conference
Monte Verità / Ascona, March 1.-3. 2001

New Algorithmic Approaches for the Anticipatory Route Guidance Generation Problem

Michel Bierlaire
Department of Mathematics
Ecole Polythechnique Federale
Lausanne

Phone: 021- 693 25 38
Fax: 021- 693 55 70
eMail: michel.bierlaire@epfl.ch

Frank Crittin
Department of Mathematics
Ecole Polythechnique Federale

Lausanne

Phone: 021- 693 81 00
Fax: 021- 693 55 70
eMail: frank.crittin@epfl.ch

Abstract

The emergence of Intelligent Transportation Systems (ITS) has considerably changed the field of transportation modeling during the past ten years. By Intelligent Transportation Systems (ITS) we refer to transportation systems which apply emerging information technologies to alleviate congestion problems (see, for example, Transportation Research Board 1999). They combine advanced surveillance systems collecting real-time traffic data, Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS) in order to achieve that objective. The complexity of these systems requires the development of sophisticated tools for their optimal exploitation.

Among the many ITS applications, providing real-time travel information is a particularly important one. It consists in using available data from the surveillance system in order to generate information that will be disseminated by the ATIS (variable message signs, highway advisory radio on standard broadcast bands, on-board GPS-based computers, etc.) The messages may contain complete or partial route recommendations, qualitative or quantitative information about traffic conditions over the network.

We are interested here in generating consistent anticipatory route guidance (CARG). Anticipatory guidance accounts for the probable evolution of traffic conditions over time and throughout the network. By consistent, we mean that the anticipated traffic conditions used to generate the guidance must be similar to the traffic conditions that the travelers are going to experience on the network. The problem is non trivial because, contrarily to weather forecast where the real system under consideration is not affected by information provision, the very fact of providing travel information may modify the future traffic conditions and, therefore, invalidate the prediction that has been used to generate it. Consequently, travelers response to information must be explicitly captured in the CARG generation process. If not, overreaction phenomena may be observed (Ben-Akiva, de Palma & Kaysi 1996). and the reliability of the information system will be affected.

Based on a detailed analysis framework Bottom, Ben-Akiva, Bierlaire, Chabini, Koutsopoulos and Yang (1999) and Bottom (2000) describe the CARG problem as a fixed point problem. In practice the computation of this fixed point is computationally expensive, because it involves complex simulation tools. Consequently classical algorithms solving fixed point problems, for example based on the Banach contraction principle, are inadequate, particularly for their use for real-time traffic management systems. Bottom (2000) has tested the most efficient fixed point algorithms on the CARG problem illustrating their slow behavior.

Our idea is to consider this problem as a large scale optimization problem without derivative, exploiting the specific characteristics of the CARG problem. Due to the absence of derivative information, we have designed a population-based procedure to gather variational information and define a local model of the objective function. Contrary to interpolation methods, for example DFO (Conn, Scheinberg & Toint1, 997), we do not force our model to interpolate the function. Indeed the large number of variables does not allow in practice to verify the geometric conditions required by these methods. Moreover, the stochastic character of the function makes the requirement of a perfect interpolation irrelevant.

Instead, at each iteration we incorporate available information about the function to update the model.

The proposed heuristic aims at decreasing the objective function as much as possible in an acceptable computational time. We present here the preliminary numerical results of this method for the consistent anticipatory route guidance problem and discuss their impact on operational dynamic Traffic Management System.

Keywords

Intelligence Transportation System (ITS) - Route Guidance Generation – Derivative Free Optimization – Fixed Point.

Swiss Transport Research Confe rence – STRC 2001 – Monte Verità

1. Introduction

Intelligent Transportation Systems have received a tremendous amount of attention in the last decade, both from practitioners and researchers. Interestingly, research in this field has yielded to new difficult mathematical problems. Among them, the consistent anticipatory route guidance generation (CARG) problem is very challenging.

By consistent, we mean that the anticipated traffic conditions used to generate the guidance must be similar to the traffic conditions that the travelers are going to experience on the network. The problem is non trivial because, contrarily to weather forecast where the real system under consideration is not affected by information provision, the very fact of providing travel information may modify the future traffic conditions and, therefore, invalidate the prediction that has been used to generate it. Consequently, travelers response to information must be explicitly captured in the generation process. If not, overreaction phenomena may be observed and the reliability of the information system will be affected.

In his recent thesis Bottom (2000) describe the CARG problem as a fixed point problem. In practice, the computation of this fixed point is computationally expensive, because it involves complex simulation tools. Consequently classical algorithms solving fixed point problems, for example based on the Banach contraction principle, seems inadequate, particularly in the context of real-time traffic management systems.

This paper describes a new algorithm considering the CARG problem as a large scale optimization problem. Due to the absence of derivative information, we design a population-based procedure to compute an approximation of the gradient and define a local model of the objective function. Contrary to interpolation methods, for example DFO (Conn, Scheinberg & Toint, 1997), we do not force our model to interpolate the function. Indeed the large number of variables does not allow in practice to verify the geometric conditions required by these methods. Moreover, the stochastic character of the function makes the requirement of a perfect interpolation irrelevant. Instead, at each iteration we incorporate available information about the function to update the model. The proposed heuristic aims at decreasing the objective function as much as possible in an acceptable computational time.

The purpose of this paper is to present the current version of the algorithm in what is likely to be a longer-term project. It is organized as follows: Section 3 introduces the problem and its fixed point interpretation. The description of existing methods and the heuristic are discussed in Section 4. Preliminary numerical results are presented in Section 5 and some conclusions and perspectives are outlined in Section 6.

2. Consistent Anticipatory Route Guidance

Route guidance refers to information provided to travelers in an attempt to facilitate their route choice decisions. Route guidance is said to be anticipatory or proactive if it is based on future traffic conditions, as opposed to reactive guidance, based on historical or on current traffic conditions.

Providing traffic information based on predicted traffic conditions may produce a phenomenon called overreaction. It happens when the number of drivers changing their behavior in response to the information is so high that the predicted traffic conditions used as a basis for the guidance generation becomes invalid. Such phenomenon is undesirable, as it jeopardizes the effectiveness and the reliability of the information system.

In order to avoid it, the guidance generation process must explicitly anticipate the drivers reaction to the information. As a consequence, we obtain a self-referenced problem, where the generated information depends on the predicted traffic conditions, that depend themselves on the generated information.

The consistent route guidance generation problem have been analyzed by several authors from a theoretical viewpoint. In his PhD dissertation, Kaysi (1992) has introduced for the first time the concept of consistency as a solution to the overreaction problem of anticipatory information systems. Other theoretical analysis has been proposed namely by Kaufman, Smith & Wunderlich (1998), Engelson (1997) and van Schuppen, (1997) The problem has also been considered in the context of the development of real-time DTA systems like DynaMIT (Ben-Akiva, Bierlaire, Bottom, Koutsopoulos & Mishalani, 1997), (Ben-Akiva, Bierlaire, Koutsopoulos & Mishalani, 2000), (Ben-Akiva, Bierlaire, Koutsopoulos & Mishalani, 1998), (Bottom, Ben-Akiva, Bierlaire, Chabini, Koutsopoulos & Yang, 1999), DYNASMART (Mahmassani, Hu, Peeta & Ziliaskopoulos, 1993) or TRANSIMS (Nagel 1997).

2.1 Fixed Point Formulation

The first formal formulation has been proposed by Bottom and al. (1999) and developed by Bottom (2000) in his recent PhD dissertation. This framework is based on exogenous parameters: a directed graph representing the traffic network, a discretized time horizon, a description of the transportation demand by origin, destination, departure time and access to information and behavioral class, an enumeration of feasible paths in the network, a list of decision points where en-route decisions can be implemented and finally on three sets of endogenous variables:

- **Path flows** *P*, representing the number of trips of a particular user class traveling from a given decision point to a given destination at a given time.
- **Network conditions** *C*, that are typically represented by time-dependent link impedances for each time interval in the given time horizon. In most cases, the impedance will be the link traversal time.
- **Guidance messages** *M* are quadruple involving message type, location, time and contents. The exact definition of these variables depends on the specific technology under consideration.

We note that both *P* and *C* are continuous, while *M* may be discrete or continuous, depending on the underlying information technology. Each pair of variable sets is related by causal relationship, expressed by the following map

- The **Network loading map**, denoted by *S: P ® C*, determines the traffic conditions that result from the assignment of a given set of time-dependent path flows P over the network.
- The **Guidance map**, denoted by *G: C® M*, represents the generation of actual messages *M* by the information system, based on predicted traffic conditions *C*. The map captures the technological characteristics of the information system.
- The **Routing map**, *D: M ® P* relates a given set of guidance messages to the resulting path flows. It captures the driver's response to the information, and is based on route choice and route switching model (see Ben-Akiva and Bierlaire, 1999).

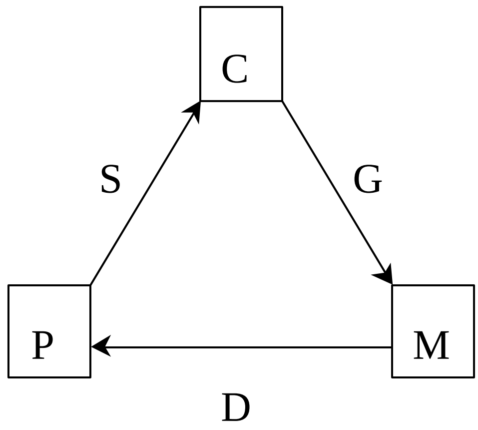
It is important to mention here that both the network loading map and the routing map are usually stochastic. Mainly, the routing map's stochasticity is related to the uncertainty associated to behavior modeling, and the network loading map's stochasticity to the discrete nature of traffic (see Bottom 2000 for a more detailed analysis of the sources of stochasticity).

The formulation of the consistent anticipatory guidance generation is based on the composition of these three maps. Actually, three such compositions can be defined:

$$D \circ G \circ S: \; P \rightarrow P$$

$$S \circ D \circ G: \; C \rightarrow C$$

$$G \circ S \circ D: \; M \rightarrow M$$



(2.1)

A consistent anticipatory guidance strategy is such that it corresponds to a fixed point of the $G \circ S \circ D: M \rightarrow M$ composite map. But the problem can be addressed with any of the three composite maps. If T is one of these maps and x^* the associated time-dependent set of variables, x^* is a fixed point of T if $x^* = T(x^*)$. These three composite maps are equivalent with respect to the existence of a fixed point: if one has a fixed point then they all do, and if one does not then none does.

Note that there is no guarantee that such fixed point exists, so in practice an approximate solution x_k is sought such that $|| x_k - T(x_k) || \leq \varepsilon$ for some $\varepsilon > 0$ and some norm $||.||$. The non-existence of a fixed point has a concrete implication on the guidance generation problem. Indeed, it means that whatever information is sent to travelers, it will never be consistent with the traffic conditions they will experience.

2.2 Characteristics of this problem

Using this analysis framework, the consistent anticipatory route guidance problem can be described as a fixed point problem with the following characteristics:

Expensive Computation The computation of the maps usually involves complex numerical simulations. Therefore the solution algorithm must be able to compute the fixed point using a minimum number of functions evaluations. Moreover, the complexity of ITS and the necessity of capturing traveler response to information impose the use of disaggregated models and simulation-based tools. Therefore, the formulation of the fixed-point problem cannot be stated as an analytical function. Instead, it can be described only by actually running a set of simulation tools with specific parameters values. As a consequence, optimization techniques that will be invoked may not rely on the availability of derivatives.

Large-scale problem The analysis framework proposes three different types of problem variables: path flows, link impedance and information messages. For non-trivial applications, each of these formulations involves a high number of variables. Indeed, the number of *C* variables is equal to the number of links in the network times the number of intervals in the time horizon. The number of *P* variables is equal to the number of paths in the network times the number of intervals in the time horizon, which for most practical applications is significantly larger than the number of *C* variables. The number of *M* variables again depends on the underlying technology.

Stochasticity The simulation-based approach motivated above also implies some sort of stochasticity in the fixed point problem computation. The level of stochasticity depends on the simulation tools used to compute the composite map. Bottom (2000) has shown that the level of stochasticity is strongly related to the level of granularity of the traffic's representation. The finer the granularity, the lower the stochasticity. When individual vehicles are explicitly represented, the level of stochasticity can be significant. We note that real-time systems (like DynaMIT, Ben Akiva and al. 2000) tend to aggregate vehicles into packets in order to increase the computation speed. This may consequently increase the level of stochasticity of the composite map. See (Bottom, 2000) for a detailed analysis of this phenomenon.

3. Existing methods

We briefly present two types of methods. The averaging methods are designed to solve fixed point problems. The derivative free algorithms are designed to solve small-scale, deterministic optimization problems when derivatives are not available. The proposed heuristics attempts to combine the two approaches in order to exploit their respective advantages.

3.1 Averaging methods

The formalization of the framework as a fixed point problem enables Bottom (2000) to identify some solution algorithms.

The functional iteration method, also known as the method of successive approximation, is based on the Banach contraction principle. It is a well known method for computing a fixed point of a contractive mapping. It starts at an arbitrary point x_0 and applies the iterative step $x_{k+1}=T(x_k)$ until a fixed point has been approximated to a sufficient degree of accuracy. It converges geometrically to a fixed point if T is contractive. Ortega and Rheinboldt (1970) have established other sufficient conditions for convergence of this method.

The simple averaging method involves the recursion

$$x_{k+1} = x_k + \alpha_k \left(T(x_k) - x_k\right)$$

where α_k is a sequence such that $\sum_{k=0}^{\infty} \alpha_k = \infty$ and $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$.

This class of algorithms was proved to converge to a root of $T(x)-x$ by Robbins and Monro (1951) and by Blum (1954) despite noisy function evaluations. Chung (1954) and Fabian (1973) also showed that Robbins and Monro recursive averaging procedure is an asymptotically optimal procedure for determining equation roots in the presence of noise. A typical example of a simple averaging is the method of successive averages (MSA) proposed by Sheffy and Powell (1982) and analyzed by Powell and Sheffi (1982), where $\alpha_k = 1/k$. For example, DynaMIT (Ben-Akiva, Bierlaire, Bottom, Koutsopoulos & Mishalani, 1997) uses solutions techniques similar to the MSA.

Magnanti and Perakis (1997) propose a scheme to determine a_k by minimizing a potential function $P(a_k)$ using either exact or inexact line search methods. They identify and investigate a variety of potential functions for the fixed point problem in the deterministic case and show that the theoretical rate of convergence can in some cases be faster than averaging with pre-determined a_k.

Averaging of iterates proposed by Polyak and Juditsky (1992) refers to computing a running average

$$x_k = \sum_{i=1}^{k} \frac{x_i}{k}$$

where x_i are calculated with the simple averaging process. They proved that the sequence x_k converges at an optimum rate if the simple averaging process involves a noisy function and converges see also (Kushner & Yang, 1993). Note that the calculations of x_k are done off-line and it begins after performing n steps of the simple averaging process.

In his conclusions, Bottom (2000) selects MSA and Polyak's averaging methods as best candidates to solve the consistent route guidance generation problem. When the stochasticity of the composite mapping is low, Polyak's method outperforms MSA by a factor of 2 to 4 in terms of run time efficiency. In the presence of high stochasticity, the performance of both approaches is comparable.

3.2 Derivative free algorithms

The fixed point problem based on the composite mapping T may be considered as a nonlinear minimization problem

$$\min_x d(x, T(x))$$

where d defines a suitable distance. As mentioned before, this is a large-scale problem for which no derivatives are available, and the function evaluation is particularly expensive. We briefly present here a review of derivative free algorithms that may be considered.

Evaluating derivatives Even if the derivatives are not analytically available, they could be numerically evaluated using finite difference techniques or algorithmic differentiation (Griewank, 2000). However, such techniques are not appropriate here, as they are expensive, and will differentiate the noise of the function instead of the function itself.

Direct Search methods Direct search methods try to find a descent direction at each iteration, namely a direction along which the objective function decreases with a suitable rate.

The first direct search methods were proposed during the 50s and the 60s (see Fletcher, 1965, for a review). The optimization community due to the lack of convergence results has scorned direct search methods. Such results were obtained in China by Yu (1979) and in Russia by Rykov (1980) and Rykov (1983). A regained interest in direct search methods appeared after the thesis by Torczon (1989), where she proposes a multi-directional search method and the associated convergence proof.

Torczon (1995) and Torczon (1997) formalized the concept of pattern search methods, and proposed a convergence analysis. Methods proposed previously by Box (1957), Hooke and Jeeves (1961), Powell (1964) and Powell (1965) can be categorized as pattern search methods. All of them are based on a smart exploration of a grid based on a predefined geometrical pattern. Wright (1996) presents an historical perspective.

Simplex search methods are based on the vertices of a simplex, which is updated to reflect the local geometry of the objective function. The first simplex method was proposed by Spendley, Hext and Himsworth (1962), but the most popular has probably been proposed by Nelder and Mead (1965), even if it has been shown not to converge in some circumstances (Mckinnon, 1998).

We note here that these methods are designed for small problems. We have made preliminary tests with these methods for solving the consistent route guidance generation problem. Their inefficiency, compared to fixed point methods as MSA or Polyak averaging, was clearly demonstrated.

Interpolation methods The first ideas for interpolation methods seem to have been proposed by Winfield (1969). The work by Powell (1970) is usually considered as seminal in this field. The idea is to replace the expensive objective function by a quadratic model that interpolates the function at specific points. The main difficulty is to maintain good geometrical properties of the interpolation points (De Boor and Ron ,1992; Powell, 1994). An efficient trust-region based algorithm (DFO) has been proposed by Conn and Toint (1995), and the convergence properties have been analyzed by Conn et al. (1997). This algorithm has been shown to perform well with practical applications (Booker, Dennis, Frank, Serafini, Torczon & Trosset, 1998). A variant of DFO based on Lanczos method has been proposed by Gould, Lucidi, Roma and Toint (1999). There are also hybrid approaches combining interpolation and direct search methods (see, for example, Lucidi and Sciandrone, 1997).

Unfortunately, these methods are also designed for small scale problems. Indeed, $(n+1)(n+2)/2$ interpolation points are necessary to compute a model of the objective function.

4. Framework for the new algorithm

The new algorithm is designed to account for the problem properties. Its structure is consistent with the classical framework of trust region methods (Conn, Gould and Toint, 2000.) This kind of algorithm uses a model of the true objective function around the current iterate which is easier to minimize than the objective. This model is assumed to sufficiently approximate the objective function in a given trust region, traditionally a ball centered at the current iterates. The radius of this ball is called the trust region radius, denoted by *D*. The algorithm minimizes the model in the trust region to find a new iterate. The objective function is then evaluated at this point. The trust region radius is updated to reflect the model accuracy at the new iterate. This process is repeated until convergence.

The first ingredient of a trust region algorithm is the choice of an adequate model. A common choice is to define a local quadratic model q of f in the neighborhood of a given point x_0 that is

$$q(x; x_0, g, H) = f(x_0) + (x - x_0)^T g + \frac{1}{2}(x - x_0)^T H(x - x_0)$$

We note that, in the classical trust region context, g represents the gradient of the objective function, and H the Hessian or a quasi-Newton approximation of it, both evaluated at x_0. Here, due to the absence of derivatives information, we need to design alternative procedures for defining g and H. This method will be presented bellow.

4.1 Gradient model

The method attempts to compute a direction g capturing available variationnal information about the objective function. It is designed to a approximate the gradient projection onto the manifold generated by previous iterates using. The direction g is derived in the following way.

At each iteration, a population $P^k=\{ x_0, x_1, ..., x_{k-1}\}$ is considered, such that $f(x_i)$ is available for every i. Moreover we assume that $x_0 = argmin_{x Î P}\ f(x)$.

Let D be the subspace of dimension s £ $k\text{-}1$ generated by the directions $d_i=x_i\text{-}x_0$, for $i=1,\ldots,k\text{-}1$, and let b_i, $i=1,\ldots,s$ be a basis of that subspace. We define

$$g(a)=\sum_{j=1}^{s} a_j\, b_j \qquad (4.1)$$

where the coefficients a_j are the solution of the following minimization problem:

$$\min_{a} \sum_{\hat{\imath}=1}^{k-1} \frac{\left(f(x_i)-f(x_0)-d_i^T g(a)\right)^2}{\|d_i\|^2} \qquad (4.2)$$

This method uses relatively large steps to approximate the directional derivatives, contrarily to finite differences techniques, and captures mostly variations in F. The robustness of this approach, and the trade-off with the quality of g for optimization purposes, has been analyzed.

4.2 Quadratic model

In order to obtain a quadratic model of the objective function, we use an extension of the technique suggested in the context of interpolation methods (Conn and Toint, 1995), when not enough points are available for a valid model interpolation. Among the infinite number of models that interpolate the objective function, they select the one such that

$$\|g\|^2 + \|H\|_F \qquad (4.3)$$

is minimal. Our extension does not impose on the model to interpolate the objective function. Instead, we consider a priori approximation of g and H, say g_0 and H_0, and select g^* and H^* such that

$$\begin{aligned}&\left(g^*, H^*\right)= \\ &\arg\min_{g,H} \sum_{i=1}^{N-1} v_i \left(f(x_i)-q(x_0;x_i-x_0,g,H)\right)+a\,\|\,g-g_0\,\|^2+b\,\|\,H-H_0\,\|_F\end{aligned} \qquad (4.4)$$

where x_i , $i = 1,...,N-1$ are points where the function has already been evaluated, w_i, a and b are weight parameters. The main motivation of not imposing perfect fit of the model at the known points is again related to the relevance of these points (points far from x_0 are less relevant for the model) and to the stochasticity of f. Note that this problem is a very large scale least square problem and can be written as: $Ax = b$ where x is a vector with $(n(n+1))/2 + n$ values.

The gradient model described above is used to obtain g_0 . The choice of H_0 is more difficult. Currently, we use $H_0 = I$.

4.3 Trust region update

The basic trust region radius update (Conn and al. 2000) is based on a comparison of the improvement predicted by the quadratic model, and the actual improvement in the objective function.

In our context, we would like to incorporate more information collected during the iterations to update the radius. Contrarily to methods where g and H are the exact gradient and Hessian, our model may be improved after an unsuccessful iteration, as a new point will be used for its estimation. Therefore, assuming that the new model (incorporating the new point) is better, there is no need to immediately decrease the trust region radius. So we define a measure of actual improvement of the model when a new point is incorporated, and base the update strategy on that measure. If the iteration is unsuccessful, but the new point x_{k+1} increases the precision of our model, the trust region radius remains the same.

4.4 The complete algorithm

The values for D_0, w, a, b, $h_1, g_i > 0$, for $i=1,2,3$ and the initial population P^0 are given.

Initialisation and Warm up

Define $D=D_0$ and $P^k=\{ x_0,x_1,...,x_{k-1}\}$
Compute $f(x_i)$

Start a new iteration

Check stopping criteria

Set $x_0 = argmin_{x\hat{I}\,P}\ f(x)$

Compute $d_i=(x_i-x_0),\ i=1,...,k-1$

Compute the a priori gradient

Compute a basis $G = span(d_1,...,d_{k-1})$

Compute the a priori gradient using (4.1)

Construct the quadratic model

Compute (g^*, H^*) using (4.4) with weightings w, a, b.

Construct $m(x_0+s) = f(x_0) + s^T g^* + \frac{1}{2} s^T H^* s$

Solve the trust region problem

Find $s_k = arg\ min_{s \le D}\left(m(x_0+s)\right)$

$x_{k+1} = x_k + s_k$

Evaluate the objective function

Compute $f(x_{k+1})$

Add x_{k+1} to P_k

Update the trust region radius

Compute $r_k = \dfrac{f(x_k) - f(x_{k+1})}{m(x_k) - m(x_{k+1})}$ (4.5)

If the iteration is successful

If $r_k > h_1$ then $D_k = ,g_1\ D_k$ (increase the radius)
else $D_k = D_k$.

If the two last iterations are unsuccessful

If $r_k > r_{k-1}$ then $D_k = g_2\ D_k$
(x_{k+1} increases precision of model, the radius is slightly reduced)
else $D_k = g_3\ D_k$
(deteriorate the model, the radius is reduced)

$k = k + 1$ and go to point 1.

4.4.1 Practical Issues

We now need to comment some features introduced in the algorithm, but not discussed above:

Starting Point and Initial Population The generation of route guidance is performed frequently, on the same network, but with traffic conditions that vary in time. In most cases, when the traffic conditions do not vary drastically, the solution of previously solved instances of the problem will be a good approximation of the solution of the next instance. At this moment we start with an admissible point and run a few iterations of MSA algorithm to create an initial population. It is called the warm-up phase.

Initial Trust-Region Radius We assume that the initial trust-region radius D_0 is given. In practice, it is quite difficult to arbitrary choose a value for this radius. We currently use a technique proposed by Sartenaer (1997) that take into account the size of H^* at x_0 computed with initial population. It is given by:

$$D_0 = \frac{\|g\|^2}{\left|g^T H^* g\right|}$$

Geometrics properties Of course, the quality of the model as an approximation of the objective function around x_0 depends on geometric properties of points in our population. So when an unsuccessful iteration occurs, to ensure progress of our algorithm we have two options to improve the model. First reduce the trust-region radius, so the next point will be nearer to x_0 and local information will be added to the model. Secondly keep the same radius improving the quality of our model introducing a better point in the population. This point must satisfy some geometric properties, similar to the poisedness mentioned by Conn and Toint (1995).

In the current version of the algorithm, the quality of the new point is tested as we construct the basis of the subspace generated by the directions d_j. If the point degenerates the basis, it is removed from the population and a new point is computed. Of course, better measures of the geometric quality of our population will improve the precision of our model, and hopefully increase the speed of the algorithm.

Intensification versus Diversification Intensification is the integration in the population of a point which is in the manifold spanned by directions d_j . Diversification integrates a point outside this manifold. It is important to combine both in order to improve the model accuracy (intensification) and explore many dimension of the problem (diversification).

Stopping criteria A well-known difficulty associated with derivative-free methods is optimality identification. Indeed, in the absence of derivatives, classical stopping criteria (see, for example, Dennis and Schnabel, 1983) cannot be used. In our context, however, the value of the objective function at the solution is 0 (if a fixed point exists), but due to the large-scale character of the problem, and the sophisticated simulation tools that may be involved in the composite map computation, we believe that the computation of an exact fixed point is hopeless for most practical applications. Moreover, the exact fixed point is not necessarily requested for real implementations. So we introduce a *computation budget*, which represents the number of evaluations of the composite map we can afford in the context of an application. The objective is to find the most consistent guidance within that budget.

5. Preliminary Numerical Results

We now report some preliminary numerical experience using the proposed algorithm in order to illustrate the ideas of this paper and to demonstrate their practical inputs. The algorithm has been implemanted in *C++* and compared with MSA and Polyak algorithms.

The computation of the route guidance is based on a simple simulation tool implemented by Bottom (2000). This software implements the framework developed in section 3. It provides simple versions of the supply, demand and guidance map. The network is composed of 14 links with a single origin-destination (OD) pair and 11 OD paths.

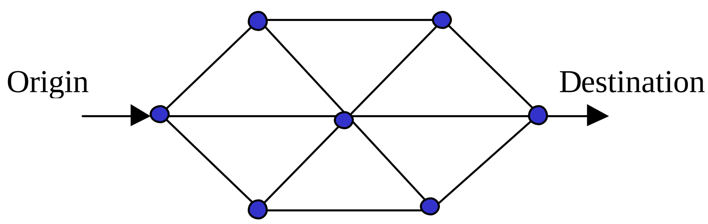


Figure1

The network is driven by a demand rate of 10800 trips/hours from Origin node to Destination node over a period of 20 minutes, after which there are no further network entries. The numbers of informed travelers is 50 %. Informed travelers receive the latest estimates of time-dependent link traversal times from the point algorithm. The simulation time step is 1 second. All simulation start with an empty network and continue until the last vehicle has left the network; this generally takes around 40 minutes of simulated time.

All guidance generation calculations work with the composite link condition map $S \circ D \circ G : C \rightarrow C$ as it involves the smallest number of variables of the tree fixed point formulation. In this case we can write the fixed point problem as finding $x \in C$ such that

$$x = S \circ D \circ G(x).$$

Variables of this problem are time-dependent link impedance for all time interval in the given horizon, it can be view as a table of size 33600. Initial point for all algorithms is a table of free-flow link traversal times.

MSA is implemented with the common value for a_k set to *1/k*. In practice Polyak iterate averaging method is only started after MSA stabilizes, this is called the *window of averaging*. With the simple simulator this window of averaging appears after 100 iterations. In Figure 5.1 Polyak's method is run after performing 100 steps of MSA, these results are qualitatively equivalent to Bottom (2000).

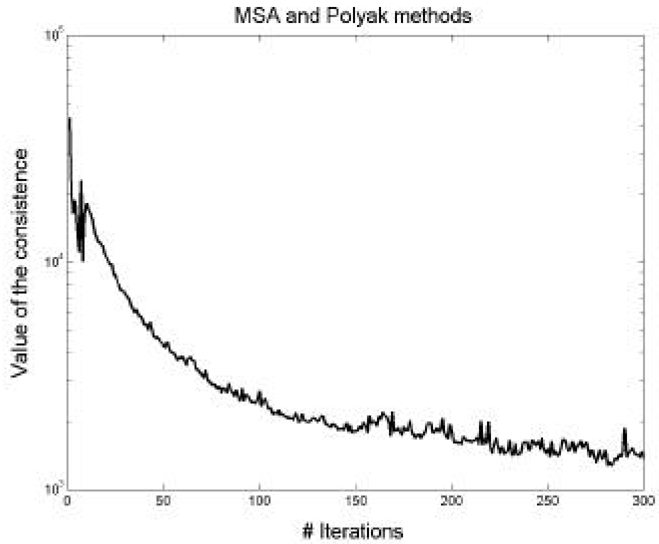


Figure 5.1

We will refer at Nitro (New Investigation in Traffic Recommendation & Optimization) as name of the algorithm described in this paper. We will consider an hybrid heuristic combining Nitro and MSA, referred to as "Hybrid". This method simply takes, at each iteration, the best one among those generated by Nitro and MSA.

In Figure 5.2 we compare the 50 first iterates of Nitro algorithm and MSA. The first iteration coincides with the 15th evaluation of the warm up period. We see that Nitro algorithm has a good model of the function and decreases the objective function faster than MSA. Figure 5.2 also shows that Nitro algorithm struggles after about 20 iterations, even though MSA finds better solutions. The interpretation is certainly that the algorithm is stuck in a subspace where no improvement can be achieved anymore.

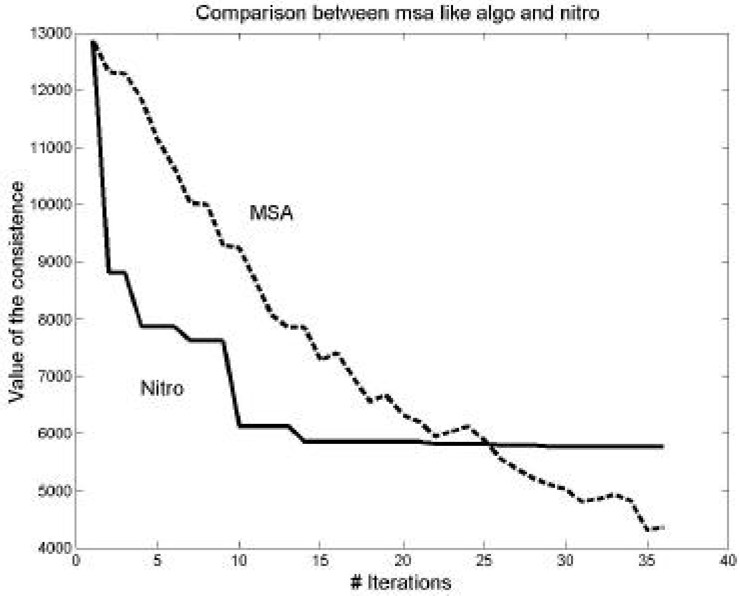


Figure 5.2

Real-time considerations

In Figure 5.2 after 10 evaluations of the function the difference of consistence between MSA and Nitro is about 3000. In terms of real-time applications, if the budgetary constraint is tight, *i.e.* maximum number of function evaluations is small, this algorithm seems to be a very good alternative to averaging methods. Moreover, in practice, the CARG problem will be performed frequently on the same network. The solution of a previous instance of the problem is a good approximation of the solution of the next instance. Also the quadratic model of the previous solution is a good approximation of the model of the objective function at the starting point of a new instance. In figure 5.3 we take a better initial point than the classical free flow link traversal and begin Nitro algorithm with its associated model. We observe that nitro improves the consistency pretty soon, while MSA needs 9 iterations to do as well as Nitro.

.

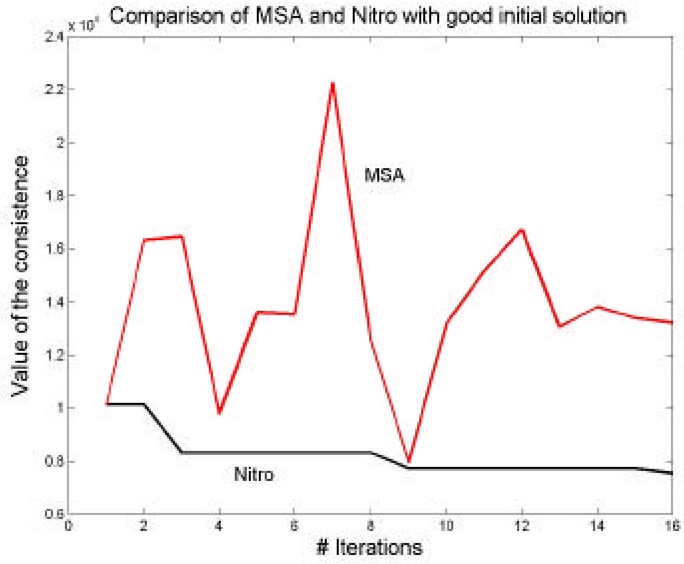


Figure 5.3

Sensitivity

The behavior of Nitro algorithm is quite sensitive to parameters. Particularly for the value assigned to the parameters of the least square problem (4.4) that computes the quadratic model. We vary the value of parameter a, the weight on the a priori gradient $g_{0.}$ Trade-off between the a priori gradient and interpolation points has to be well defined, otherwise the method will struggle faster through lack of information, as shown in Figure 5.4 for the value $\alpha = 2$.

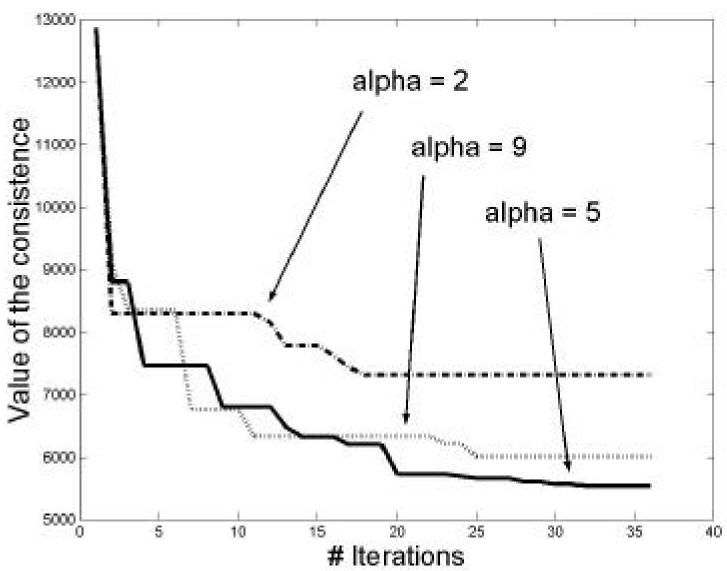


Figure 5.4

Performance

Figure 5.5 shows the performances using the relative improvement of Hybrid algorithm compared with MSA for 100 iterations.

$$\text{Relative improvement} = \frac{x_k^{MSA} - x_k^{Hybrid1}}{x_k^{MSA}}$$

The good performance of Nitro algorithm is here confirmed. We also see that Hybrid algorithm is similar in performance to MSA after 50 iterations. This behavior can be explained by the size of the trust region, which is probably too large. A shrewder trust region update policy as to be implemented to solve this problem.

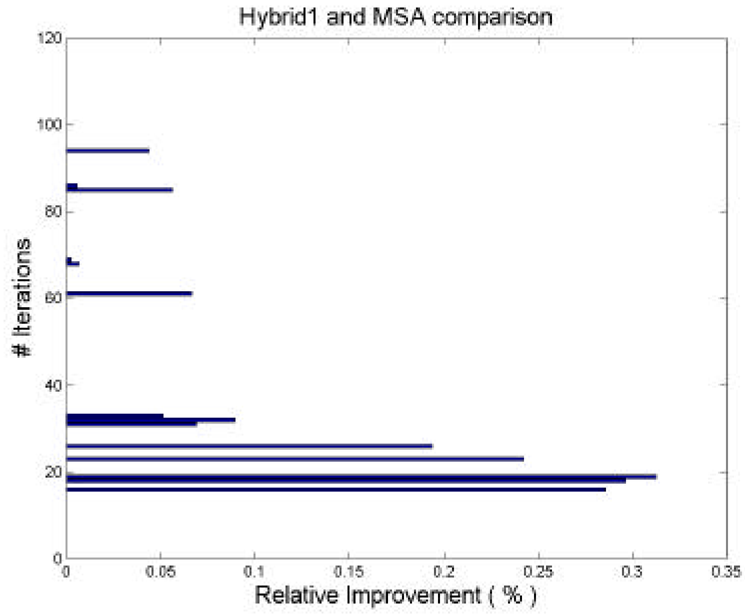


Figure 5.5

Those preliminary results are encouraging, even if a deeper analysis is required to better understand the algorithm behavior

6. Conclusions and perspectives

We have presented the current status of development of a new algorithm to solve the CARG problem. This algorithm is based on trust-region framework, computing at each iteration an approximation of the gradient in the manifold spanned by the previous iterates and a quadratic model of the objective function. The algorithm seems efficient compared to the classical averaging methods and its performance are promising, particularly in the context of real-time application. Some additional development is necessary before the ideas presented here can result in a final applicable algorithm. In particular, the following research directions seem to be of interest.

- In this paper we apply an optimization formulation of the fixed point problem, *i.e.* $\min \| T(x) - x \|$. We believe that a formulation in term of search of a root, *i.e.* $h(x) = 0$ with $h(x) = T(x) - x$, using a Newton-like method will allow us to get around some problem inherent to the optimization formulation. For example, appearance of local minimums, which are not fixed point of $T(x)$. Another possibility will be to use several optimization formulations.

- Other formulation of the CARG problem should be investigated, including the two other composite maps define in (2.1) that characterizes the fixed point formulation. As Bottom (2000) proposed, the path time variable appeared to involve the least amount of stochasticity and the good behavior of our algorithm will increase.

- As said before a better measure on the geometric properties of the points in our population will improve the quality of the quadratic model and so the performance of the algorithm. For example, building a sub-quadratic model, that is a model in which not not all degrees of freedom of a full quadratic model are exploited, such model can be obtained by multivariate interpolation algorithms proposed by Sauer and Xu (1995).

- The development of a "convergence" theory covering the final version of our algorithm is desirable in itself and also for a better understanding of its internal behavior.

The idea presented here are a first step in the development of a robust and efficient algorithm for solving CARG problem. It is clear that further numerical experiences are needed to analyze the performance of our algorithm in a more general context.

7. References

Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H. Mishalani, R. 2000. Real time simulation of traffic demand-supply interactions within DynaMIT, Technical Report RO-000301 , ROSO-DMA-EPFL Swiss Institute of Technology, CH-1015 Lausanne. http://rosowww.epfl.ch/mbi/dynamit-supply.pdf

Ben-Akiva, M. Bierlaire, M. 1999. Discrete choice methods and their applications to short-term travel decisions, in R.~Hall (ed.), Handbook of Transportation Science, Kluwer, pp. 5-34.

Ben-Akiva, M. E., Bierlaire, M., Koutsopoulos, H. N. Mishalani, R. G. 1998. DynaMIT : a simulation-based system for traffic prediction, Proceedings of the DACCORD Short-Term forecasting workshop.

Ben-Akiva, M. E., de Palma , A. Kaysi, I. 1996. The impact of predictive information on guidance efficiency: an analytical approach, in L. Bianco P. Toth (eds), Advanced Methods in Transportation Analysis , Springer-Verlag, pp. 413--432.

Bottom, J. 2000. Consistent Anticipatory Route Guidance, PhD thesis,Massachusetts Institute of Technology.

Booker, A. J., Dennis, J. E., Frank, P. D., Serafini, D. B., Torczon, V. Trosset, M. W. 1998. A rigourous framework for optimization of expensive functios by surrogates, Technical report, Mathematics \& Engineering Analysis, Boeing Shared Services Group. http://www.cs.wm.edu/\ va/research/rfroot.ps.gz

Bottom, J., Ben-Akiva, M., Bierlaire, M., Chabini, I., Koutsopoulos, H. Yang, Q. 1999, Investigation of route guidance generation issues by simulation with DynaMIT , in A. Ceder (ed.), Transportation and Traffic Theory. Proceedings of the 14th ISTTT, Pergamon, pp. 577--600.

Bottom, J., Ben-Akiva, M., Bierlaire, M., Chabini, I., Koutsopoulos, H. Yang, Q. 1999. Investigation of route guidance generation issues by simulation with DynaMIT , in A. Ceder (ed.), Transportation and Traffic Theory. Proceedings of the 14th ISTTT , Pergamon, pp. 577--600.

Box, G. E. P. 1957. Evolutionary operation: a method for increasing industrial productivity, Appl. Statist. 6: 81--101.

Chung, K. L. 1954. On a stochastic approximation method, Annals of mathematical statistics 25 (3): 463--483.

Conn, A., Gould, N. Toint, P. 2000. Trust region methods, MPS--SIAM Series on Optimization, SIAM.

Conn, A. R. Toint, P. L. 1995. An algorithm using quadratic interpolation for unconstrained derivative free optimization, in G. D. Pillo F. Gianessi (eds), Nonlinear Optimization and Applications , Plenum Publishing, p. (to appear).Also available as Report 95/6, Dept of Mathematics, FUNDP, Namur, Belgium.

Conn, A., Scheinberg, K. Toint, P. L. 1997 On the convergence of derivative-free methods for unconstrained, in A. Iserles M. Buhmann (eds), Approximation Theory and Optimization: Tributes to M.J.D. Powell, Cambridge University Press, Cambridge, UK, pp. 83--108

De Boor , C. Ron, A. 1992. Computational aspects of polynomial interpolation in several variables, Mathematics of Computation 58 (198): 705--727.

Dennis, J. E. Schnabel, R. B. 1983. Numerical methods for unconstrained optimization and nonlinear equations, Prentice-Hall, Englewood Cliffs, USA.

Engelson, L. 1997. Self-fulfilling and recursive forecasts --- an analytical perspective for driver information systems, Proceedings of 8th IATBR meeting. Austin, Texas.

Fabian, V. 1973. Asymptotically efficient stochastic approximations: the RM case, Annals of statistics 1 (3): 486--495.

Fletcher, R. 1965. Function minimisation without evaluating derivatives --- a review, Computer Journal 8.

Gould, N., Lucidi, S., Roma, M. Toint, P. L. 1999. Solving the trust region subproblem with the L anczos method, Journal on Optimization .

Griewank, A. 2000. Evaluationg derivatives. Principles and Techniques of algorithmic differentiation , Frontiers in Applied Mathematics, Siam.

Hooke, R. Jeeves, T. A. 1961. Direct search solution of numerical and statistical problems, Journal of the ACM 8 : 212--229.

Kaufman, D. E., Smith, R. L. Wunderlich, K. E. 1998. User-equilibrium properties of fixed points in iterative dynamic routing/assignment methods, TRC 6 : 1--16.

Kaysi, I. 1992. Framework and Models for the Provision of Real-Time Driver Information, PhD thesis, Massachusetts Institute of Technology, Cambridge, Ma.

Kushner, H. J. Yang, J. 1993. Stochastic approximation with averaging of the iterates: optimal asymptotic rate of convergence for general process, SICON 31 (4): 1045--1062.

Lucidi, S. Sciandrone, M. 1997. On the global convergence of derivative free methods for unconstrained optimization, Technical report , Universita di Roma, La Sapienza.

Magnanti, T. Perakis, G. 1997. Averaging schemes for variational inequalities and systems of equations, Mathematics of operations research 22 (3): 568--587.

Mahmassani, H., Hu, T., Peeta, S. Ziliaskopoulos, A. 1993. Development and testing of dynamic traffic assignment and simulation procedures for ATIS/ATMS applications, Technical Report DTFH61-90-R-00074-FG, Center for Transportation Research, University of Texas at Austin.

Mckinnon, K. I. 1998. Convergence of the N elder- M ead simplex method to a nonstationary point, SIOPT 9 (1): 148--158.

Nagel, K. 1997. Experiences with iterated traffic microsimulations in Dallas, in M. Schrekenberg D. E. Wolf (eds), Proceedings of the workshop on traffic and granular flow , SV, Duisburg, Germany.

Nelder, J. A. Mead, R. 1965. A simplex method for function minimization Computer Journal 7: 308--313.

Polyak, B. T. Juditsky, A. B. 1992. Acceleration of stochastic approximation by averaging, SICON 30 (4): 838--855.

Powell, M. J. D. 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives, Computer Journal 17 : 155--162.

Powell, M. J. D. 1965. A method for minimzing a sum of squares of nonlinear functions without calculating derivatives, Computer Journal 7 : 303--307.

Powell, M. J. D. 1970. A new algorithm for unconstrained optimization, in J. B. Rosen, O. L. Mangasarian K. Ritter (eds), Nonlinear Programming , Academic Press, New York.

Powell, M. J. D. 1994. A direct search optimization method that models the objective by quadratic interpolation, Presentation at the 5th Stockholm Optimization Days.

Powell, W. Sheffi, Y. 1982. The convergence of equilibrium algorithms with predetermined step size, Transportation Science 16 : 45--55.

Robbins, H. Monro, S. 1951. A stochastic approximation method, Annals of mathematical statistics 22 (3): 400--407.

Rykov, A. 1980. Simplex direct search algorithms, Automation and Robot Control, 41: 784-793.

Rykov, A. 1983. Simplex algorithms for unconstrained optimization, Problems of Control and Information Theory 12: 195--208.

Sartenaer, A. 1997. Automatic determination of an initial trust region in nonlinear programming. SIAM Journal on scientific Computing, 18(6), 1788-183.

Sauer, Th. Xu, Y. 1995. On multivariate Lagrange interpolation. Mathematics of Computation, 64:1147-1170.

Sheffi, Y. Powell, W. 1982. An algorithm for the equilibrium assignment problem with random link times, Networks 12: 191--207.

Spendley, W., Hext, G. R. Himsworth, F. R. 1962. Sequential application of simplex designs in optimization and evolutionary operation, Technometrics 4 . Transportation Research Board 1999. Research on intelligent transportation systems, human factors, and advanced traveler information system design and effects, National Academy Press, National Research Council. Washington, D.C.

Torczon, V. 1995. Pattern search methods for nonlinear optimization, SIAG/OPT Views-and-News, A Forum for the SIAM Activity Group on Optimization, 6: 7--11.

Torczon, V. 1997. On the convergence of pattern search algorithms, SIAM Journal on Optimization 7 (1): 1--25.

Torczon, V. J. 1989. Multi-directional search: a direct search algorithm for parallel machine, PhD thesis, Rice University, Houston, Texas. http://www.cs.wm.edu/\ va/research/thesis.ps.gz

van Schuppen , J. H. 1997. Annex D : routing control of motorway networks. DACCORD deliverable D06.1 : co-ordinated control strategies, Hague Consulting Group, The Hague.

Winfield, D. 1969. Function and functional optimisation by interpolation in a data table, PhD thesis, Harvard University, Cambridge, MA.

Wright, M. H. 1996. Direct search methods: once scorned, now respectable, in D. F. Griffith G. A. Watson (eds), Numerical Analysis 1995(Proceedings of the 1995 Dundee Biennial Conference in Numerical Analysis), Addison Weslay, Lomgman, Harlow, UK, pp. 191--208.