



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Transportation Research Part B 40 (2006) 351–367

TRANSPORTATION  
RESEARCH  
PART B

[www.elsevier.com/locate/trb](http://www.elsevier.com/locate/trb)

# Solving a non-convex combined travel forecasting model by the method of successive averages with constant step sizes

Hillel Bar-Gera<sup>a,\*</sup>, David Boyce<sup>b</sup>

<sup>a</sup> *Department of Industrial Engineering and Management, Ben-Gurion University, Be'er Sheva 84105, Israel*

<sup>b</sup> *Department of Civil and Materials Engineering, University of Illinois, Chicago, IL 60607, USA*

Received 20 March 2003; received in revised form 15 September 2003; accepted 17 May 2005

---

## Abstract

The method of successive averages is often used in iterative algorithms for solving various mathematical problems, and travel forecasting models in particular. In each iteration of these algorithms the current solution is averaged with an alternative solution generated by the algorithm. If the problem has an equivalent convex optimization formulation with a computable objective function, a line search can be used to determine the weight applied to the alternative solution, referred to as the step size. In the absence of a convex optimization formulation, the typical approach is to set the step size according to a predetermined sequence that is decreasing towards zero, such as  $1/k$  where  $k$  is the iteration index.

In this paper, we examine the use of a constant step size in the method of successive averages. A theoretical derivation shows that if the alternative solution generated by the algorithm is a linear function of the current solution, then using a constant step size is advantageous and substantially superior to using a sequence of decreasing step sizes. We conjecture that similar results may be expected in nonlinear differentiable problems as well.

Numerical results are presented for a travel forecasting model that combines user-equilibrium route choice with an origin-destination-mode (ODM) choice model. The proposed algorithm is based on origin-based assignment, in conjunction with successive averaging of constant step sizes for the differentiable ODM model. The results show that a properly chosen constant step size leads to excellent convergence for

---

\* Corresponding author. Tel.: +972 8 6461398; fax: +972 8 6472958.  
E-mail address: [bargera@bgu.ac.il](mailto:bargera@bgu.ac.il) (H. Bar-Gera).

both convex and non-convex models. A general strategy for choosing step sizes without a-priori knowledge is presented as well.

© 2005 Elsevier Ltd. All rights reserved.

*Keywords:* Successive averages; Travel forecasting; Combined models; Fixed point problem

---

## 1. Introduction

This paper examines a travel forecasting model that combines a user-equilibrium route choice model with a doubly constrained origin-destination-mode choice model with a compound deterrence function consisting of the product of a negative exponential term and a power term. Bar-Gera and Boyce (2003) showed that this model (without the power term) can be formulated as a convex optimization problem, and solved either by a link-based algorithm or more precisely and efficiently by an origin-based algorithm. Tanner (1961) suggested that a compound deterrence function can improve the model's ability to replicate observed travel patterns; see also the discussion in Erlander and Stewart (1990). While the change to a compound deterrence function may appear minor, the convex optimization formulation considered previously cannot accommodate it, nor can it accommodate many other variations that are often proposed to improve model realism. For brevity, we refer to these as non-convex models, even though it is theoretically possible that they are "convex" in some sense. In order to evaluate the implications of a compound deterrence function, or any other non-convex model, the first step is to develop precise and efficient algorithms for solving them, which is the main purpose of this paper.

The remainder of the paper is organized as follows. Section 2 describes the travel forecasting model, and demonstrates how it fits into the general fixed-point formulation presented in Bar-Gera and Boyce (2003). Section 3 contains a brief description of the origin-based algorithm for combined models presented in Bar-Gera and Boyce (2003). A discussion of the algorithm's ability to handle non-convex combined models follows, showing that the only necessary change is a different way to determine the step size when updating the ODM model. Choice of step size in the method of successive averages for fixed-point problems is discussed in Section 4. Section 5 presents computational results for one medium-size model and one large-scale model, under various scenarios. Finally, conclusions and suggestions for future research are presented in Section 6.

## 2. Combined travel forecasting models

The travel forecasting model examined in this paper is similar to the one considered in Bar-Gera and Boyce (2003), except for the introduction of a power term in Eqs. (9) and (10) below. The model seeks to represent typical traffic patterns during a certain time period of weekdays of a given year, in a study area which is divided into a set of *zones* denoted by  $Z$ . The expected number of travelers from origin  $p \in Z$  to destination  $q \in Z$  by mode  $m \in M = \{a, t\}$  (auto or transit), referred to as the ODM flow (persons/hour), is denoted by the variable  $d_{mpq}$ .  $\mathbf{d}$  denotes the array of ODM flows. Vehicle OD flows are person flows by auto divided by a constant auto occupancy factor,  $aof$ , with additional fixed OD truck flows (given as model input),

$$v_{pq} = d_{apq}/aof + d_{pq}^{\text{truck}}; \text{ or } \mathbf{v}(\mathbf{d}) = \mathbf{d}_a/aof + \mathbf{d}^{\text{truck}}.$$

The roadway system is described by a network consisting of a set of nodes  $N$  and a set of links  $A$ . Link  $a \in A$  connects its tail node  $a_t$  to its head node  $a_h$ . The set of available routes from origin  $p$  to destination  $q$  in the auto network is denoted by  $R_{pq}$ , and the set of these sets is  $\mathbf{R} = \{R_{pq}\}_{p,q \in Z}$ . The distribution of travelers from  $p$  to  $q$  among the routes in  $R_{pq}$  is described by a vector of non-negative route proportions (conditional probabilities)  $\gamma_{pq} = \{\gamma_{pqr}\}_{r \in R_{pq}}$ . The array of route proportion vectors is  $\gamma$ . Since route proportions must sum to one for each O–D pair, the set of all feasible route proportion arrays is

$$\Gamma(\mathbf{R}) = \left\{ \gamma \in [0, 1]^{|\mathbf{R}|} : \sum_{r \in R_{pq}} \gamma_{pqr} = 1 \quad \forall p, q \in Z \right\} \quad (1)$$

Given  $\gamma_{pq}$ , the implied vector of route flows is  $\mathbf{h}_{pq} = \{h_{pqr}\}_{r \in R_{pq}} = v_{pq} \cdot \gamma_{pq}$ ; the array of route flow vectors is denoted by  $\mathbf{h}$ . Defining the cross product to imply  $\mathbf{z} = \mathbf{x} \times \mathbf{y} \Rightarrow z_{pq} = x_{pq} \cdot y_{pq}$  provides a convenient way to define route flows by the following array notation  $\mathbf{h}(\mathbf{v}, \gamma) = \mathbf{v} \times \gamma$ . Flows on routes that pass through link  $a \in A$  are aggregated over all destinations to yield origin-based link flows  $f_{ap} = \sum_{q \in Z} \sum_{r \in R_{pq}: a \subseteq r} h_{pqr}$ ; these are further aggregated over all origins to yield total link flows  $f_a = \sum_{p \in Z} f_{ap}$ . Link travel time functions are of the BPR form

$$tt_a(f_a) = tt_a^0 \cdot (1 + 0.15 \cdot (f_a/k_a)^4) \quad (2)$$

where  $tt_a^0$  and  $k_a$  are the given link free flow travel time and the link capacity, respectively. Auto operating costs, including gasoline consumption, are a linear function of link length and travel time,

$$oc_a = \eta_1 \cdot tt_a(f_a) + \eta_2 \cdot l_a \quad (3)$$

where  $l_a$  is the given link length. Link generalized costs are

$$t_a(f_a) = \beta_a^t \cdot tt_a(f_a) + \beta_a^c \cdot oc_a(f_a) \quad (4)$$

where the  $\beta$ 's are calibration parameters. Parking costs by zone  $pc_z$  and walking times to or from the parking place  $wt_z$  in each zone are components of the additional auto costs, defined as

$$ac_{apq} = \beta_a^p \cdot \frac{pc_p + pc_q}{2} + \beta_a^w \cdot (wt_p + wt_q) \quad (5)$$

Route generalized costs by auto are  $c_{apqr} = ac_{apq} + \sum_{a \in r} t_a$ . This entire cost structure is written in short form as  $\mathbf{c} = \mathbf{C}(\mathbf{h})$ .

The set of minimum cost routes for a given O–D pair  $p, q$  is denoted by  $R_{pq}^*(\mathbf{c}) = \text{argmin}\{c_{apqr} : r \in R_{pq}\}$ . The array of such sets is denoted by  $\mathbf{R}^*(\mathbf{c})$ . For any non-empty subset of routes  $\mathbf{R}'$ ,  $\emptyset \subsetneq \mathbf{R}'_{pq} \subseteq R_{pq}$ , define the set of feasible route proportion arrays that are limited to  $\mathbf{R}'$  as

$$\Gamma(\mathbf{R}') = \{\gamma \in \Gamma(\mathbf{R}) : \gamma_{pqr} = 0 \quad \forall r \notin \mathbf{R}'_{pq} \quad \forall p, q \in Z\} \quad (6)$$

We assume that route choice follows the user-equilibrium principle of Wardrop (1952), that as each traveler seeks to minimize the cost associated with their chosen route, at equilibrium the cost

of every used route cannot be greater than the cost of any alternative route. Hence we are looking for solutions where

$$\gamma \in \Gamma(\mathbf{R}^*(\mathbf{C}(\mathbf{h}(\mathbf{v}(\mathbf{d})), \gamma))) \quad (7)$$

The total flow of person-trips per hour from each origin  $\bar{d}_{p\bullet}$ ; and the total flow of person-trips per hour to each destination  $\bar{d}_{\bullet q}$  are given as part of the model input. These totals together with the model endogenous values of ODM generalized costs are used to determine ODM flows. Auto OD generalized costs are defined as the weighted average over all used routes,  $u_{apq} = \bar{U}_{apq}(\mathbf{c}, \gamma) = \gamma_{apq} \cdot c_{apq}$ ; or in array notation  $\mathbf{u}_a = \bar{\mathbf{U}}_a(\mathbf{c}, \gamma) = \mathbf{c}_a \times \gamma$ . Note that under the user-equilibrium assumption (7) average OD costs are in fact equal to the minimum OD costs. Transit OD generalized costs are a linear function of in-vehicle travel times  $c_{tpq}^i$ , fares  $c_{tpq}^f$ , and out-of-vehicle times  $c_{tpq}^o$

$$u_{tpq} = \beta_t^1 + \beta_t^i \cdot c_{tpq}^i + \beta_t^f \cdot c_{tpq}^f + \beta_t^o \cdot c_{tpq}^o \quad (8)$$

Transit cost components are all constants, independent of flow, as are transit OD generalized costs. ODM flows are given by

$$d_{apq} = A_p \cdot B_q \cdot \exp(-\mu \cdot u_{apq}) \cdot u_{apq}^{-\rho} \quad (9)$$

$$d_{tpq} = A_p \cdot B_q \cdot \exp(-\mu \cdot u_{tpq}) \cdot u_{tpq}^{-\rho} \quad (10)$$

where  $A_p$  and  $B_q$  are balancing factors determined by the method of Fratar (1954), with a correction as described in Appendix A, to guarantee that

$$\sum_{m \in M} \sum_{q \in Z} d_{mpq} = \bar{d}_{p\bullet} \quad (11)$$

$$\sum_{m \in M} \sum_{p \in Z} d_{mpq} = \bar{d}_{\bullet q} \quad (12)$$

When  $\rho = 0$  this function is the same logit-type ODM choice model discussed in Bar-Gera and Boyce (2003), which has a convex optimization formulation; however, that convex optimization formulation does not apply for  $\rho > 0$ . This ODM choice model is represented compactly by the continuous upper-bounded function  $\mathbf{d} = \Phi(\mathbf{u})$ . The entire combined ODM choice and route choice model is formulated as a fixed point problem as follows:

$$\{\mathbf{d}, \gamma\} \in F(\mathbf{d}, \gamma) = \{\Phi(\bar{\mathbf{U}}(\mathbf{C}(\mathbf{h}(\mathbf{v}(\mathbf{d})), \gamma)), \gamma)\} \times \Gamma(\mathbf{R}^*(\mathbf{C}(\mathbf{h}(\mathbf{v}(\mathbf{d})), \gamma))) \quad (13)$$

or equivalently

$$\mathbf{d} = \Phi(\bar{\mathbf{U}}(\mathbf{C}(\mathbf{h}(\mathbf{v}(\mathbf{d})), \gamma)), \gamma) \quad (14)$$

$$\gamma \in \Gamma(\mathbf{R}^*(\mathbf{C}(\mathbf{h}(\mathbf{v}(\mathbf{d})), \gamma))) \quad (15)$$

The precision of any given approximate solution  $\{\mathbf{d}, \gamma\}$  for (13) can be measured by two gap values. To measure deviations from (14), examine the so-called “subproblem” ODM flows as determined by the ODM model from current ODM costs,  $\mathbf{d}' = \Phi(\bar{\mathbf{U}}(\mathbf{C}(\mathbf{h}(\mathbf{v}(\mathbf{d})), \gamma)), \gamma)$ , and define the *total misplaced ODM flow*, in units of person-trips per hour, as the sum of absolute differences between these subproblem ODM flows and the ODM flows in the current solution; that is

$$\text{TMF} = \sum_{m \in \mathcal{M}} \sum_{p, q \in \mathcal{Z}} |d_{mpq} - d'_{mpq}| \quad (16)$$

The second measure is the *average auto excess cost*, in the same units used for generalized travel costs,

$$\text{AEC} = \frac{1}{v_{\bullet\bullet}} \cdot \sum_{r \in \mathcal{R}} h_r \cdot ec_r \quad (17)$$

$$ec_r = c_{apqr} - \min_{r' \in \mathcal{R}_{pq}} c_{apqr'} \quad (18)$$

$$v_{\bullet\bullet} = \sum_{p, q \in \mathcal{Z}} v_{pq} \quad (19)$$

Note that AEC measures assignment precision, while TMF measures ODM choice precision. Therefore, a solution is considered to be precise only if both measures are small enough, that is if average excess cost is less than, say, 0.001 vehicle-minutes, and total misplaced ODM flow is less than, say, 1000 person-trips/hour.

### 3. Algorithms

Algorithms based on the method of successive averages consist of a procedure to generate subproblem solutions and a strategy for choosing step sizes. We consider two options for generating subproblem solutions; one is a link-based algorithm (Evans, 1976), and the other is an origin-based algorithm. The performances of these algorithms with step sizes that are determined according to a convex objective function, in a problem where such objective function exists, were examined in Bar-Gera and Boyce (2003). In this paper, we focus on other strategies for choosing step sizes, which are discussed in Section 4. For the sake of completeness, this section presents a brief review of the way subproblem solutions are generated in the two algorithms.

For the Evans algorithm, given the  $k$ th iteration solution  $(\mathbf{d}^k, \gamma^k)$ ;  $\mathbf{h}^k = \mathbf{v}(\mathbf{d}^k) \times \gamma^k$ , for each OD pair the subproblem route proportion of the minimum cost route equals 1, and equals 0 for all other routes, according to current iteration flows:  $\hat{\gamma}^k \in \Gamma(\mathbf{R}^*(\mathbf{C}(\mathbf{h}^k)))$ ; subproblem ODM flows are determined according to the costs of these routes,  $\hat{\mathbf{d}}^k = \Phi(\bar{\mathbf{U}}(\mathbf{C}(\mathbf{h}^k), \hat{\gamma}^k))$ , and then assigned according to subproblem route proportions to yield subproblem route flows  $\hat{\mathbf{h}}^k = \hat{\mathbf{d}}^k \times \hat{\gamma}^k$ . These are averaged with the current solution according to a step size  $0 \leq \lambda \leq 1$  to generate the next iteration solution  $\mathbf{h}^{k+1} = (1 - \lambda) \cdot \mathbf{h}^k + \lambda \cdot \hat{\mathbf{h}}^k$ .

The general scheme of the origin-based algorithm is presented in Fig. 1. It relies on the origin-based assignment algorithm presented in Bar-Gera (2002), which is briefly described next. The key addition to the algorithm for handling combined models is a procedure for updating ODM flows, while retaining the route proportions of the current solution. Given a current solution,  $\{\mathbf{d}^k, \gamma^k\}$ , subproblem ODM flows are determined according to average auto OD costs  $\hat{\mathbf{d}}^k = \Phi(\bar{\mathbf{U}}(\mathbf{C}(\mathbf{h}^k), \gamma^k))$ ; new ODM flows are obtained by a weighted average  $\mathbf{d}^{k+1} = (1 - \lambda) \cdot \mathbf{d}^k + \lambda \cdot \hat{\mathbf{d}}^k$ , where  $0 \leq \lambda \leq 1$  is a chosen step size. For convex models, the objective function can be used to determine the step size. A proof of convergence for the resulting algorithm is given in Bar-Gera

**Initialization:**

```

Let  $\mathbf{u} = \mathbf{U}^*(C(0))$ 
Let  $\mathbf{d}^0 = \Phi(\mathbf{u})$ 
for  $p$  in  $Z$  do
   $A_p =$  tree of minimum cost routes from  $p$ 
   $\mathbf{f}_p =$  all or nothing assignment using  $A_p$ 
end for

```

**Main loop:**

```

for  $n=1$  to number of main iterations
  Update O-D flows, retain route proportions
  for  $p$  in  $Z$  do
    update restricting subnetwork  $A_p$ 
    update origin-based approach proportions  $\alpha_p$ 
  end for
  for  $m=1$  to number of inner iterations
    for  $p$  in  $Z$  do
      update origin-based approach proportions  $\alpha_p$ 
    end for
  end for
end for

```

Fig. 1. An origin-based algorithm for combined models.

and Boyce (2003). As shown there, the use of average auto OD costs (rather than minimum auto OD costs as in the Evans algorithm) to determine subproblem ODM flows is critical for convergence.

Once the ODM flows are updated, route proportions are revised in an origin-based assignment iteration, while keeping O–D flows temporarily fixed. The main solution variables in the origin-based assignment algorithm are *origin-based approach proportions*:  $\alpha = \{\alpha_{ap}\}_{a \in A; p \in Z}$ ;  $0 \leq \alpha_{ap} \leq 1$ ;  $\sum_{a: a_h=j} \alpha_{ap} = 1 \forall j \in N \forall p \in Z$ . For every origin an a-cyclic restricting subnetwork is chosen,  $A_p \subseteq A$ ;  $a \notin A_p \Rightarrow \alpha_{ap} = 0$ . Initial restricting subnetworks are trees of minimum cost routes. To update the restricting subnetwork, unused links are removed,  $v_i$ —the maximum cost to node  $i$  within the restricting subnetwork is computed, and all links  $[i, j]$  such that  $v_i < v_j$  are added to the restricting subnetwork. Approach proportions for origin  $p$  are updated by shifting flows within the restricting subnetwork  $A_p$  according to a boundary (piece-wise linear) search in a direction determined by an approximate second order method.

Route proportions are determined by  $\gamma_{pqr} = \prod_{a \in r} \alpha_{ap}$ . It can be shown that  $f_{ap} = \alpha_{ap} \cdot g_{jp}$  where  $g_{jp} = \sum_{a: a_h=j} f_{ap}$  is the origin-based node flow, demonstrating that  $\alpha_{ap}$  is indeed the proportion of flow on approach  $a$  to node  $a_h$  for origin  $p$ . The availability of route proportions allows one to compute average O–D costs, and to assign new O–D flows by current route proportions. Due to the restriction to a-cyclic subnetworks, these computations can be done efficiently without route enumeration, in a time that is a linear function of the number of links times the number of origins. These properties are essential for the demand update procedure described above.

In both the Evans algorithm and the origin-based algorithm, the main obstacle to a general implementation for non-convex models is the determination of the step size. However, there is

a critical difference between the two algorithms. In the Evans algorithm, subproblem solutions are not continuous as a function of the current solution, as they are based on assignment to minimum cost routes. In contrast, when updating ODM flows with the origin-based algorithm the subproblem solution is a continuous smooth function of the current solution. This difference and its implications on step size choices are discussed in the next section.

#### 4. Choice of step size in the method of successive averages

To simplify the discussion and the notation, let us first examine a general fixed point problem  $\mathbf{x} = \mathbf{F}(\mathbf{x})$ , and then relate the conclusions of the general discussion to the travel forecasting model and to the algorithms described in the previous sections. Each iteration in the general method of successive averages (MSA) uses the current solution  $\mathbf{x}^k$  to find a subproblem solution,  $\mathbf{y}^k = \mathbf{F}(\mathbf{x}^k)$ . The next current solution is an average of these two solutions,  $\mathbf{x}^{k+1} = (\mathbf{1} - \lambda^k) \cdot \mathbf{x}^k + \lambda^k \cdot \mathbf{y}^k$ .

MSA was introduced in the seminal paper of Robbins and Monro (1951), who suggested in particular using step sizes that are predetermined as  $\lambda^k = 1/k$ , where  $k$  is the iteration index. Many researchers adopted this choice of step sizes (e.g. Powell and Sheffi, 1982). Polyak (1990) argues in the context of stochastic approximation techniques that asymptotic convergence is optimal in the sense that it is dominated by the stochastic noise if either a properly chosen constant step size is used, or step sizes of  $\lambda^k = k^{-\beta}$  where  $0 < \beta < 1$ . Averaging with a constant step size can also be viewed as a variant of the Landweber–Fridman method for inverse problems (Landweber, 1951), as described by Groetsch (Groetsch, 1993, pp. 96–97).

We consider computational procedures that are not influenced by stochastic noise. We are particularly interested in cases where the subproblem solution is a continuous and smooth function of the current solution. In these cases, in the neighborhood of a fixed-point solution,  $\mathbf{x}^*$ , the relationship between the current solution and the subproblem solution is approximately linear. That is  $\mathbf{F}(\mathbf{x}) \approx \mathbf{x}^* + \mathbf{A} \cdot (\mathbf{x} - \mathbf{x}^*)$  for some matrix  $\mathbf{A}$ .

The true deviation from  $\mathbf{x}^*$  can be measured by  $\|\mathbf{x} - \mathbf{x}^*\|$ . During the iterative process, when  $\mathbf{x}^*$  is not known, the deviation can be estimated by  $\|\mathbf{x} - \mathbf{F}(\mathbf{x})\|$ . Under the linear approximation the two measures are closely related, since

$$\|\mathbf{x} - \mathbf{F}(\mathbf{x})\| \approx \|(\mathbf{I} - \mathbf{A}) \cdot (\mathbf{x} - \mathbf{x}^*)\| \leq \|\mathbf{I} - \mathbf{A}\| \cdot \|\mathbf{x} - \mathbf{x}^*\| \quad (20)$$

As is known from linear algebra, there exists a diagonalization matrix  $\mathbf{P}$  such that  $\mathbf{P}^{-1} \cdot \mathbf{A} \cdot \mathbf{P}$  is diagonal. We can assume w.l.o.g. that  $\mathbf{P}$  is orthonormal (norm preserving), so that  $\|\mathbf{x} - \mathbf{x}^*\| = \|\mathbf{P}^{-1} \cdot (\mathbf{x} - \mathbf{x}^*)\|$ . The eigenvalues of  $\mathbf{A}$  are the (possibly complex) diagonal values of  $\mathbf{P}^{-1} \cdot \mathbf{A} \cdot \mathbf{P}$ , denoted by the vector  $\omega$ . The deviation component along the  $i$ th eigenvector in the subproblem solution is the deviation in the current solution multiplied by  $\omega_i$ . In particular, if  $\omega_i$  is real, the correction proposed by the subproblem solution can be interpreted intuitively as follows: for  $\omega_i > 1$ , the correction is in the opposite direction, moving away from  $\mathbf{x}^*$ ; for  $0 < \omega_i < 1$ , the correction is in the right direction but too small; for  $\omega_i = 0$ , the correction is perfect, and one iteration eliminates any deviation; for  $-1 < \omega_i < 0$ , the correction is in the right direction but slightly too large, the deviation in the subproblem solution being smaller than the deviation in the current solution; for  $\omega_i < -1$ , the correction is in the right direction but too large, the deviation in subproblem solution being larger than the deviation in the current solution.

Denote  $\mathbf{A}(\lambda) = ((1 - \lambda) \cdot \mathbf{I} + \lambda \cdot \mathbf{A})$ . The iterative algorithm for step size  $\lambda$  is approximately  $\mathbf{x}^{k+1} \approx \mathbf{x}^* + \mathbf{A}(\lambda) \cdot (\mathbf{x}^k - \mathbf{x}^*)$ . Note that  $\mathbf{P}^{-1} \cdot \mathbf{A}(\lambda) \cdot \mathbf{P} = (\mathbf{I} - \lambda) \cdot \mathbf{I} + \lambda \cdot \mathbf{P}^{-1} \cdot \mathbf{A} \cdot \mathbf{P}$  is also diagonal. The magnitude of the  $i$ th eigenvalue of  $\mathbf{A}(\lambda)$  is  $r_i(\lambda) = \|(1 - \lambda + \lambda \cdot \omega_i)\|$ . Under the linear approximation, each eigenvector component of the deviation proceeds independently of all other components, according to

$$\|[\mathbf{P}^{-1} \cdot (\mathbf{x}^k - \mathbf{x}^*)]_i\| \approx \|[\mathbf{P}^{-1} \cdot (\mathbf{x}^1 - \mathbf{x}^*)]_i\| \cdot \prod_{j=1}^k r_i(\lambda^j) \quad (21)$$

Eq. (21) leads us to several important conclusions. If there exists an eigenvalue whose real component is greater or equal to 1,  $Re(\omega_i) \geq 1$ , then for any step size choice  $0 \leq \lambda \leq 1$  the magnitude of the corresponding eigenvalue of  $\mathbf{A}(\lambda)$  is greater or equal to 1 as well,  $r_i(\lambda) \geq 1$ . As a result, the magnitude of the deviation along that component cannot decrease,  $\|[\mathbf{P}^{-1} \cdot (\mathbf{x}^{k+1} - \mathbf{x}^*)]_i\| \geq \|[\mathbf{P}^{-1} \cdot (\mathbf{x}^k - \mathbf{x}^*)]_i\|$ , regardless of step size choices, and hence any form of MSA will fail.

If a constant step size  $\lambda$  is used, such that  $\max_i \{r_i(\lambda)\} = r_{i_{\max}}(\lambda) = r^*(\lambda) < 1$ , then all deviations converge to zero. Furthermore, asymptotically, deviations are dominated by the component along the  $i_{\max}$  eigenvector; therefore, the ratio of deviations between successive iterations approaches to  $r^*(\lambda)$ .

$$\frac{\|\mathbf{x}^{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}^k - \mathbf{x}^*\|} = \frac{\|\mathbf{P}^{-1} \cdot (\mathbf{x}^{k+1} - \mathbf{x}^*)\|}{\|\mathbf{P}^{-1} \cdot (\mathbf{x}^k - \mathbf{x}^*)\|} \approx \frac{\|[\mathbf{P}^{-1} \cdot (\mathbf{x}^{k+1} - \mathbf{x}^*)]_{i_{\max}}\|}{\|[\mathbf{P}^{-1} \cdot (\mathbf{x}^k - \mathbf{x}^*)]_{i_{\max}}\|} \rightarrow r^*(\lambda), \quad k \rightarrow \infty \quad (22)$$

The same result applies to estimated deviations as well, since for sufficiently large  $k$

$$\|\mathbf{x}^k - \mathbf{F}(\mathbf{x}^k)\| \approx \|(\mathbf{I} - \mathbf{A}) \cdot (\mathbf{x}^k - \mathbf{x}^*)\| \approx \|1 - \omega_{i_{\max}}\| \cdot \|[\mathbf{P}^{-1} \cdot (\mathbf{x}^k - \mathbf{x}^*)]_{i_{\max}}\| \quad (23)$$

and therefore

$$\frac{\|\mathbf{x}^{k+1} - \mathbf{F}(\mathbf{x}^{k+1})\|}{\|\mathbf{x}^k - \mathbf{F}(\mathbf{x}^k)\|} \rightarrow r^*(\lambda), \quad k \rightarrow \infty \quad (24)$$

The step size that minimizes  $r^*(\lambda)$  yields fastest convergence. Note that  $[r_i(\lambda)]^2$  is a quadratic convex function of  $\lambda$ . The maximum of convex functions is also convex, hence  $[r^*(\lambda)]^2$  is convex. Clearly  $r^*(0) = r_i(0) = 1$ . In addition,

$$\left. \frac{\partial ([r_i(\lambda)]^2)}{\partial \lambda} \right|_{\lambda=0} = 2 \cdot (Re(\omega_i) - 1) \quad (25)$$

so if MSA is applicable, then  $Re(\omega_i) < 1$  and  $r_i(\lambda)$  is decreasing at  $\lambda = 0$ . In conclusion,  $r^*(\lambda)$  is minimized at  $0 < \lambda^* \leq 1$  and  $r^*(\lambda^*) < 1$ . Furthermore,  $r^*(\lambda)$  decreases for  $0 \leq \lambda \leq \lambda^*$ , and increases for  $\lambda^* \leq \lambda \leq 1$ . Robust convergence can be obtained by any choice of a constant step size such that  $r^*(\lambda) < 1$ , and particularly by  $\lambda^*$ , which is the best choice possible for a constant step size.

Of course, in practical applications the value of  $\lambda^*$  is not known a-priori. It may be estimated from past experience. In particular, the application of travel forecasting models often involves the analysis of various scenarios that differ only in the properties of a small number of roads; in that case it may be worthwhile to conduct several experiments and choose a step size accordingly. In general, choosing step sizes by trial and error may be possible, but not very desirable.

The simplest alternative is to choose a predetermined decreasing sequence of step sizes,  $\lambda^k \rightarrow 0$ . Clearly, from a certain iteration onwards  $\lambda^k < \lambda^*$  and therefore  $r^*(\lambda^k) < 1$ , thus leading to decreasing deviations. However, such strategy can be substantially inferior to using a problem-specific appropriate constant step size, because after some number of iterations the predetermined step sizes may be significantly smaller than  $\lambda^*$ . For example, if  $\lambda^k = 1/k$  and  $\omega_i = -a$  is a negative real value, then  $r_i(1/k) = \|\frac{k-1-a}{k}\|$ ; so if  $k_0 > a + 1$  then for  $k \gg k_0$

$$\frac{\|[\mathbf{P}^{-1} \cdot (\mathbf{x}^k - \mathbf{x}^*)]_i\|}{\|[\mathbf{P}^{-1} \cdot (\mathbf{x}^{k_0} - \mathbf{x}^*)]_i\|} \approx \prod_{j=k_0+1}^k \frac{j-1-a}{j} \approx \frac{\text{const}(k_0, a)}{k^{a+1}} \tag{26}$$

Hence, the residual deviation is a power function of the iteration index, rather than an exponential one in the case of a constant step size. Oddly enough, smaller values of  $a$ , which correspond to more precise subproblem solutions, yield inferior performance under this strategy for choosing step sizes.

We are looking for a strategy that will start from a given step size,  $\lambda^0$ , presumably an estimate of  $\lambda^*$  according to past experience, and decrease it as needed, that is if the reduction of deviations is not sufficient. Deviations are not necessarily reduced in every iteration, especially in earlier iterations when non-linear effects are significant; therefore, we propose to consider the *average deviation reduction ratio* over the last  $\delta k$  iterations prior to iteration  $k > \delta k$ , defined by

$$\hat{r}(k, \delta k) = \left[ \frac{\|\mathbf{x}^k - F(\mathbf{x}^k)\|}{\|\mathbf{x}^{k-\delta k} - F(\mathbf{x}^{k-\delta k})\|} \right]^{1/\delta k} \tag{27}$$

Naturally, averaging over more iterations increases the stability to fluctuations, but it also increases the time until the need for a smaller step size is detected.

As step sizes get smaller,  $\lambda \rightarrow 0$ , deviation reduction ratio in component  $i$  can be approximated by  $r_i(\lambda) \approx 1 - \lambda \cdot (1 - \text{Re}(\omega_i))$ ; and the overall deviation reduction ratio can be approximated by

$$r^*(\lambda) \approx 1 - \lambda \cdot \left( 1 - \max_i \{ \text{Re}(\omega_i) \} \right) \tag{28}$$

If  $\psi \geq 0$  is an estimated lower bound of  $1 - \max_i \{ \text{Re}(\omega_i) \} \geq 0$ , then it is reasonable to decrease the step size when  $\hat{r}(k, \delta k) > 1 - \psi \cdot \lambda$ . Choosing  $\psi = 0$  suggests to decrease the step size only if the algorithm did not make any progress at all during the last  $\delta k$  iterations. Larger values of  $\psi$  suggest decreasing the step size if the progress is not fast enough.

Once a decision to decrease the step size has been made, it only remains to determine the next step size. Applying approximation (28) to small step sizes suggests computation times that are inversely proportional to step size values; so using a step size that is 10% smaller than needed, increases computation time approximately by 10%. In view of this observation, it seems appropriate to set the next step size equal to the current step size multiplied by some decreasing factor,  $\xi$ , in the range of, say, 0.5–0.9.

There are of course numerous ways to choose  $\delta k$ ,  $\psi$ , and  $\xi$ , each leading to slightly different behavior. The main tradeoffs in these choices are quite clear: lower values of  $\delta k$  and  $\xi$  and higher values of  $\psi$  lead to faster decreasing step sizes, with the potential advantage of reaching  $r^*(\lambda) < 1$  in fewer iterations, and the possible disadvantage of using step sizes that are too small. In all cases, it is clear that the actual performance will be dominated by the gap between the initial step size  $\lambda^0$

and the optimal step size  $\lambda^*$ . It seems that most insights would be learned from testing various constant step sizes, which is the purpose of Section 5. Designing meaningful experiments to evaluate more sophisticated strategies like the one presented above is much more challenging, and remains a subject for future research.

To complete the discussion of this section, the connection to the travel forecasting model described in Section 2 and to the algorithms discussed in Section 3 should be clarified. If we assume that route proportions are predetermined by  $\gamma_0$ , then the combined model (13) reduces to

$$\mathbf{d} = \Phi(\bar{\mathbf{U}}(\mathbf{C}(\mathbf{h}(\mathbf{d}, \gamma_0)), \gamma_0)) = F(\mathbf{d}) \quad (29)$$

which is differentiable for most ODM choice models including the gravity model used here. The discussion about constant step sizes therefore applies to this problem directly. In the origin-based algorithm, during the update of ODM flows, route proportions are held fixed, as in (29). Indeed route proportions do change from one iteration to the next, thus changing the fixed-point function  $F$ . Therefore, while this case does not belong to the precise definition of a fixed-point problem analyzed in this section, it seems quite likely that the conclusions about the performance of different step size strategies will be applicable to the origin-based algorithm presented in Section 3.

In the Evans algorithm, subproblem solutions are based on all-or-nothing assignment to minimum cost routes. As such they are not continuous (and certainly not differentiable) as a function of the current solution. In this case using a constant step size is not expected to lead to convergence.

One more technical issue that deserves consideration is the measure of convergence. In this section we used  $\|\mathbf{x} - \mathbf{F}(\mathbf{x})\|_2 = (\sum_i (x_i - F_i(\mathbf{x}))^2)^{0.5}$  as it simplifies the analysis. In Section 2 we proposed the total misplaced ODM flow as a measure of convergence, which is equivalent to  $\|\mathbf{x} - \mathbf{F}(\mathbf{x})\|_1 = \sum_i |x_i - F_i(\mathbf{x})|$ . The latter is preferred as it is more intuitive and less sensitive to outliers, which is particularly important due to the high number of dimensions in our problem. Theoretically the difference between the two definitions could be significant, but practically it seems quite unlikely that this difference will make a substantial impact on the algorithm's convergence behavior.

## 5. Experimental results

This section presents computation results comparing the convergence of the proposed origin-based algorithm and the Evans algorithm for different step size strategies. The algorithms were applied to two test problems. One is based on the detailed Chicago Regional Network with 1790 zones, 12,982 nodes, and 39,018 road links. The model is for one hour during the morning peak with total ODM flow of about 1.4 million person-trips per hour. The second is based on an aggregated version of the previous network, with a slightly smaller area, and referred to as the Chicago Sketch Network. In this case there are 387 zones, 933 nodes, 2950 road links, and total ODM flow of about 1.4 million person-trips per hour. Both models were calibrated using household travel survey data, and validated using Census data; see [Boyce and Bar-Gera \(2003\)](#). Variations of the basic models included: multiplying the total ODM flow by a total flow factor (TFF) of 0.5, 1, or 2; changing the cost sensitivity  $\mu$  to 0.05, 0.1, or 0.2; and changing the power parameter  $\rho$  in the deterrence function of the gravity model (9, 10) to 0, 1, or 2; of course,  $\rho = 0$  corresponds to the omission of the power term.

Experiments with the Chicago Regional Network were conducted on a PC with a 2 GHz Xeon processor and 1 GB RAM, running under Windows XP<sup>TM</sup>. Experiments with the Chicago Sketch Network were conducted on a Compaq Alpha Unix Server model DS20E, with CPU speed of 666 MHz, and 256 MB RAM. All codes are written in C.

In all origin-based solutions AEC is practically negligible. Evans solutions yield substantially higher AEC values, but the convergence of ODM flows, as measured by TMF, is the main concern in these solutions as well. Therefore, the discussion in this section focuses on TMF.

Fig. 2 compares the convergence of the origin-based algorithm with the Evans algorithm for a problem that has a convex optimization formulation ( $\rho = 0$ ). When subproblem solutions are generated by the Evans algorithm, the use of convex line search yields solutions that are about twice as precise as those yielded by the predetermined sequence of step sizes  $\lambda^k = 1/k$ . After 10 min of CPU time these methods achieve a total misplaced ODM flow of 300 and 600 vph, respectively.

The origin-based algorithm with convex line search is very effective, achieving a total misplaced ODM flow of 10 vph in only 2 min of CPU time. When step sizes are predetermined by  $\lambda^k = 1/k$ , however, it performs much worse than both Evans-type algorithms, and after 10 min of CPU time the total misplaced ODM flow remains greater than 10,000 vph. Although surprising, this result is in agreement with the analysis in Section 4, showing that with  $\lambda^k = 1/k$  more precise subproblem solutions may lead to poorer convergence.

Figs. 3–5 show the performance of the origin-based solution with constant step sizes. Figs. 3 and 4 show for both networks that when a properly chosen step size is used ( $\lambda = 0.5$  in these examples), the performance is comparable to the performance obtained by using a convex line search. Fig. 5 demonstrates that properly chosen constant step sizes provide excellent convergence for non-convex models ( $\rho = 1$ ) as well. Similar trends were observed in all other scenarios examined. Results for the Chicago Regional Network in Fig. 3 show in some cases that the convex line search may cause fluctuations, whereas constant step sizes provide consistent improvement

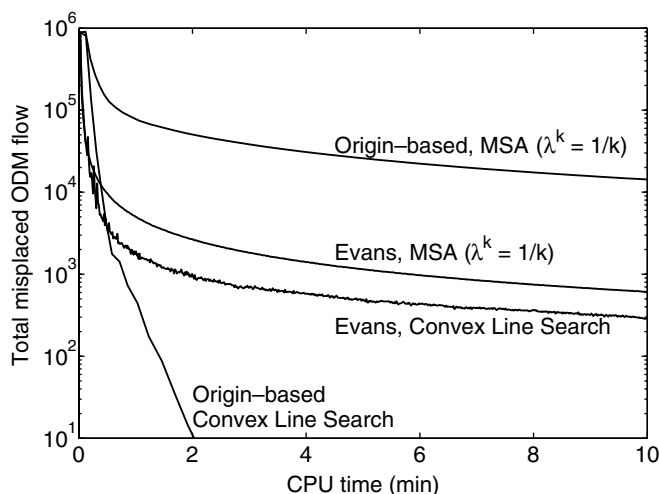


Fig. 2. Convergence comparison—Chicago Sketch Network.

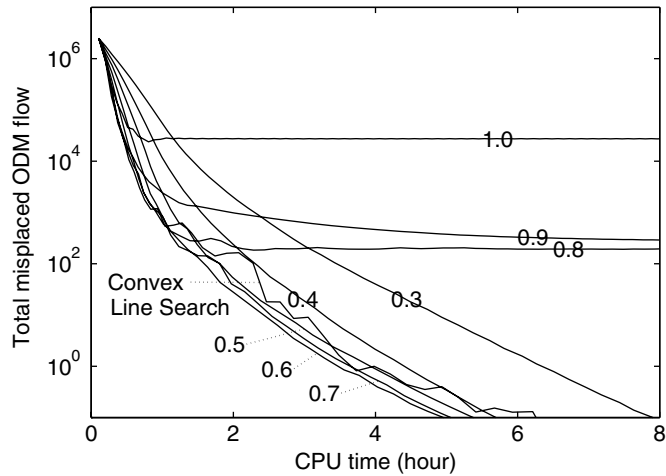


Fig. 3. OBA convergence for various constant step sizes, Chicago Regional Network,  $\rho = 0$ .

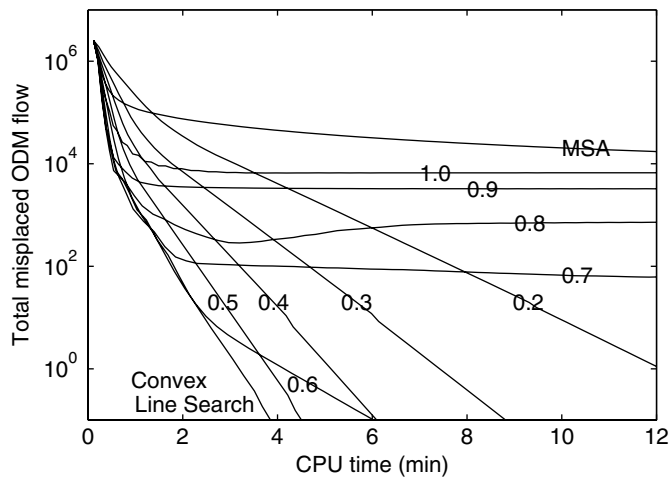


Fig. 4. OBA convergence for various constant step sizes, Chicago Sketch Network,  $\rho = 0$ .

without fluctuations. Fluctuations were rarely observed in more than 30 solutions for the Chicago Regional Model and 200 solutions for the Chicago Sketch Model, suggesting that decisions to decrease step sizes should be based on averaging of a relatively small number of iterations, such as  $3 \leq \delta k \leq 5$ .

Fig. 6 shows an attempt to use constant step sizes of various magnitudes with the Evans algorithm. As expected, regardless of how small the step size is, solutions do not improve beyond a certain point, which depends of course on the step size chosen.

A summary of the origin-based algorithm convergence performance near optimality for various scenarios is presented in Table 1. This table presents the deviation reduction per iteration, aver-

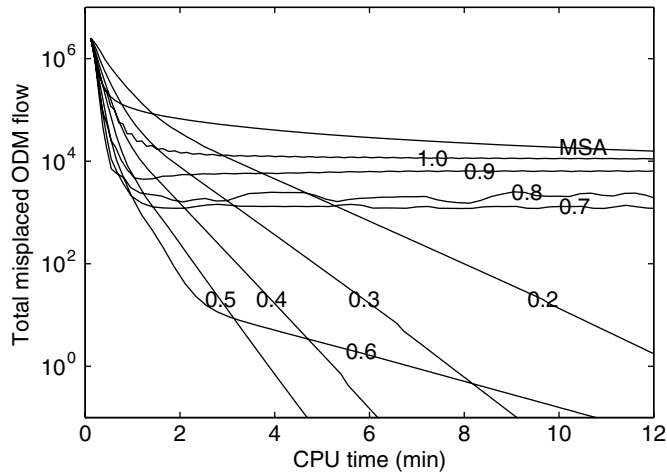


Fig. 5. OBA convergence for various constant step sizes, Chicago Sketch Network,  $\rho = 1$ .

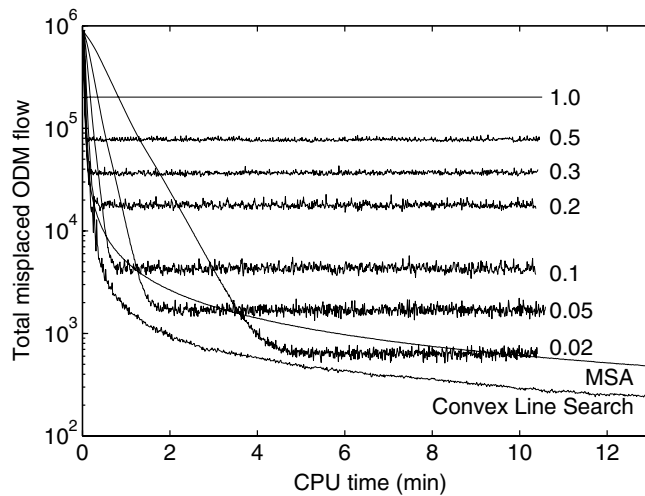


Fig. 6. Evans algorithm convergence for various constant step sizes, Chicago Sketch Network,  $\rho = 0$ .

aged over the last 10 iterations prior to the iteration that achieved a total misplaced ODM flow of 1 vph, that is  $\hat{r}(k_1, 10)$  where  $TMF(k_1) = 1$ . Interestingly, in all scenarios, when the step size is sufficiently small, the deviation reduction ratio is almost precisely  $1 - \lambda$  (between  $1 - 1.01\lambda$  and  $1 - 0.99\lambda$ ). According to Eq. (28), this suggests that  $\max_i \{Re(\omega_i)\} \approx 0$ . According to the intuitive characterization of eigenvalues in Section 4, if  $Re(\omega) \leq 0$  for all eigenvalues then the deviation in subproblem solutions is opposite in direction to the deviation of the current solution, meaning that the correction implied by the subproblem solution is larger than necessary. For example, if the equilibrium flow for a specific ODM combination is 100 vph while the current flow for that

Table 1  
Average deviation reduction ratio near optimality by step size and scenario for the Chicago Sketch Network

Scenario			$\lambda$											
	$\mu$	$\rho$	1	.9	.8	.7	.6	.5	.4	.3	.2	.1	.05	.02
1	.1	0	X	X	X	X	.76	.52	.61	.70	.801	.900	.9501	.9800
1	.1	1	X	X	X	X	.87	.64	.62	.70	.800	.900	.9501	.9800
1	.1	2	X	X	X	X	X	.73	.61	.74	.800	.900	.9500	.9800
.5	.1	0	.66	.50	.28	.30	.42	.52	.61	.70	.801	.900	.9502	.9801
.5	.1	1	.45	.37	.53	.69	.45	.50	.60	.70	.800	.900	.9501	.9800
.5	.1	2	.50	.25	.64	.38	.74	.50	.60	.70	.800	.900	.9500	.9800
2	.1	0	X	X	X	X	X	X	X	.71	.801	.900	.9501	.9800
2	.1	1	X	X	X	X	X	X	X	.80	.800	.900	.9501	.9800
2	.1	2	X	X	X	X	X	X	X	.93	.800	.900	.9500	.9800
1	.05	0	X	X	X	X	X	X	X	.93	.803	.902	.9507	.9802
1	.05	1	X	X	X	X	.60	.88	.89	.92	.802	.902	.9506	.9802
1	.05	2	X	X	X	X	.90	.53	.80	.77	.802	.902	.9506	.9802
1	.2	0	X	X	.85	.76	.75	.76	.77	.70	.800	.899	.9497	.9799
1	.2	1	X	X	X	X	X	.86	.68	.70	.800	.899	.9495	.9798
1	.2	2	X	X	X	X	X	X	.79	.87	.827	.899	.9497	.9799

X—solution did not converge; TFF—total flow factor;  $\mu$ —cost sensitivity;  $\rho$ —deterrence power parameter.

ODM combination is 110 vph then the flow for that ODM combination in the subproblem solution will be lower than 100 vph. This is quite reasonable considering the fact that additional flow for a specific ODM combination typically increases the cost of that ODM combination while reducing the cost of alternative ODM-s, thus making the shift more attractive than it really should be. The results of Table 1 also help in choosing  $\psi$  for the step size adjustment strategy presented in Section 4, which should be an estimate of  $1 - \max_i \{Re(\omega_i)\}$ . In view of the above observation  $\psi$  should be relatively close to 1, say in the range of 0.5 to 0.9.

Assuming that the linear approximation is valid for the average deviation reduction ratios presented in Table 1, we may expect that  $\hat{r}(k_1, 10) \approx r^*(\lambda)$ , and therefore  $\|1 - \lambda + \lambda \cdot \omega_i\| \leq \hat{r}(k_1, 10)$ . Graphically, the last inequality implies that in the complex plane all eigenvalues should lie inside the circle with a radius of  $\hat{r}(k_1, 10)/\lambda$ , centered at  $1 - 1/\lambda$ . Fig. 7 shows the resulting circles from several solutions with various step sizes for one scenario (Chicago Sketch Network, TFF = 0.5,  $\mu = 0.1$ ,  $\rho = 1$ ). Eigenvalues depend on the scenario, but not on the step size; therefore, they should lie at the intersection of all the circles. It seems that in this case convergence of the solutions with  $\lambda \leq 0.7$  is dominated by an eigenvalue of approximately 0, and the convergence of solutions with  $\lambda \geq 0.8$  is dominated by an eigenvalue of approximately  $-0.65$ .

The range of optimal step sizes in our results is quite moderate, between 0.1 and 0.5. These results cover six scenario variations of the Chicago Regional Model, and 15 scenario variations of the Chicago Sketch Model, with fairly radical changes such as Total Flow Factor between 0.5 and 2. We can expect, therefore, that the strategy presented in Section 4 would provide reasonable performance, and would not be too sensitive to the choices of  $\lambda_0$ ,  $\delta k$ ,  $\psi$ , and  $\xi$ . Thorough evaluation of this strategy remains a subject for future research.

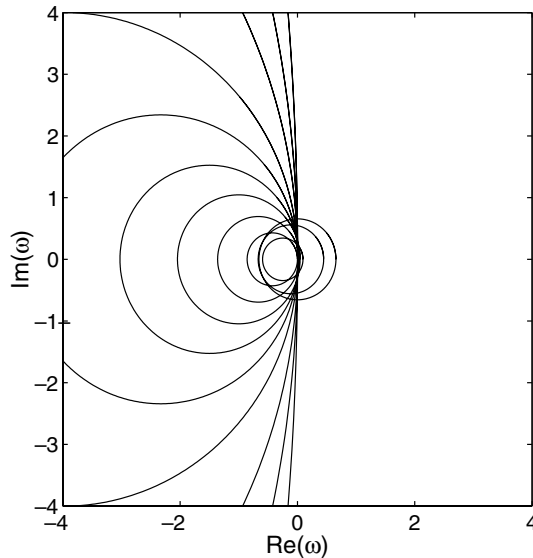


Fig. 7. Eigenvalues boundaries according to actual convergence rates.

## 6. Conclusions

We demonstrated that non-convex combined models can be solved efficiently and precisely with an origin-based algorithm, using a properly chosen constant step size for averaging demand between iterations. We also presented a general strategy to adjust the step size along the solution process, according to the actual performance of the algorithm.

The results presented here show clearly that if the algorithm used to generate subproblem solutions is continuous and precise, using step sizes from a predetermined sequence such as  $\lambda^k = 1/k$  may lead to extremely poor convergence, substantially inferior to using a properly chosen constant step size.

The method of successive averages with a properly chosen constant step size could be useful for solving other continuous fixed-point problems as well. This remains a subject for future research.

## Acknowledgement

The authors wish to thank the Chicago Area Transportation Study, Chicago, IL for providing the network data.

## Appendix A

Gravity ODM models, like the one presented in (9)–(12), are typically solved by iterative balancing (Fratat, 1954), that is by

$$A_p^1 = 1 \quad \forall p \in Z \tag{30}$$

$$B_q^1 = 1 \quad \forall q \in Z \tag{31}$$

$$A_p^j = \frac{\bar{d}_{p\bullet}}{\sum_{q \in Z} B_q^{j-1} \cdot \hat{d}_{pq}} \quad \forall j > 1; p \in Z \tag{32}$$

$$B_q^j = \frac{\bar{d}_{\bullet q}}{\sum_{p \in Z} A_p^j \cdot \hat{d}_{pq}} \quad \forall j > 1; p \in Z \tag{33}$$

$$eo_p^j = \bar{d}_{p\bullet} - \sum_{q \in Z} A_p^j \cdot B_q^j \cdot \hat{d}_{pq} \tag{34}$$

$$ed_q^j = \bar{d}_{\bullet q} - \sum_{p \in Z} A_p^j \cdot B_q^j \cdot \hat{d}_{pq} \tag{35}$$

where  $\hat{d}_{pq} = \exp(-\mu \cdot u_{apq}) \cdot u_{apq}^{-\rho} + \exp(-\mu \cdot u_{tpq}) \cdot u_{tpq}^{-\rho}$  is the composite deterrence;  $eo$ , and  $ed$  are the origin and destination residual errors, respectively. At the end of every iteration, destination constraints are satisfied precisely,  $ed_q^j = 0$ . Origin residual errors converge to 0,  $\max_p \{eo_p^j\} \rightarrow_{j \rightarrow \infty} 0$ , usually fairly quickly, especially if the matrix  $\mathbf{\hat{d}}$  is reasonable. In early iteration solutions of combined models  $\mathbf{\hat{d}}$  can be problematic, leading to fairly slow convergence of the balancing process, which may require hundreds of iterations to achieve even a moderate precision of  $\max_p \{e_p^j\} < 0.01$ . Waiting for the balancing process to achieve higher precision is clearly inefficient. Stopping the balancing process with imprecise solutions is also problematic, since these imprecisions propagate to solutions of subsequent iterations through the averaging of demand, thus making all subsequent solutions infeasible as well.

We devised a simple correction procedure that appears to be fairly robust. In the last balancing iteration,  $j = n$ , let

$$A_p^n = \begin{cases} \frac{\bar{d}_{p\bullet}}{\sum_{q \in Z} B_q^{n-1} \cdot \hat{d}_{pq}}, & \text{if } eo_p^{n-1} < 0 \\ A_p^{n-1}, & \text{if } eo_p^{n-1} \geq 0 \end{cases} \tag{36}$$

$$B_q^n = B_q^{n-1} \tag{37}$$

As a result, all residual errors are non-negative,  $eo_p^n \geq 0 \forall p \in Z; ed_q^n \geq 0 \forall q \in Z$ , and can be divided proportionally among all OD pairs. That is, setting OD flows to

$$\tilde{d}_{pq} = A_p^n \cdot B_q^n \cdot \hat{d}_{pq} + \frac{eo_p^n \cdot ed_q^n}{\bar{e}} \tag{38}$$

where  $\bar{e} = \sum_{p \in A} eo_p^n = \sum_{q \in Z} ed_q^n$ . Finally, ODM flows can be easily determined by  $d_{apq} = \tilde{d}_{pq} \cdot \exp(-\mu \cdot u_{apq}) \cdot u_{apq}^{-\rho}; d_{tpq} = \tilde{d}_{pq} \cdot \exp(-\mu \cdot u_{tpq}) \cdot u_{tpq}^{-\rho}$ .

**References**

Bar-Gera, H., 2002. Origin-based algorithm for the traffic assignment problem. *Transportation Science* 36 (4), 398–417.  
 Bar-Gera, H., Boyce, D., 2003. Origin-based algorithms for combined travel forecasting models. *Transportation Research Part B* 37 (5), 405–422.

- Boyce, D., Bar-Gera, H., 2003. Validation of multiclass urban travel forecasting models combining origin–destination, mode, and route choices. *Journal of Regional Science* 43 (3), 517–540.
- Evans, S.P., 1976. Derivation and analysis of some models for combining trip distribution and assignment. *Transportation Research* 10, 37–57.
- Erlander, S., Stewart, N.F., 1990. *The Gravity Model in Transportation Analysis—Theory and Extensions*. VSP, Utrecht.
- Fratrar, T.S., 1954. Vehicular trip distribution by successive approximation. *Traffic Quarterly* 8, 53–56.
- Groetsch, C.W., 1993. *Inverse Problems in the Mathematical Sciences*. Vieweg, Wiesbaden.
- Landweber, L., 1951. An iteration formula for Fredholm equations of the first kind. *American Journal of Mathematics* 73 (3), 615–624.
- Polyak, B.T., 1990. New method of stochastic approximation type. *Automation and Remote Control* 51 (7), 937–946.
- Powell, W.B., Sheffi, Y., 1982. The convergence of equilibrium algorithms with predetermined step sizes. *Transportation Science* 16, 45–55.
- Robbins, H., Monro, S., 1951. A stochastic approximation method. *The Annals of Mathematical Statistics* 22, 400–407.
- Tanner, J.C., 1961. Factors affecting the amount of travel. Road Research Laboratory, Technical Paper No. 51, Her Majesty's Stationery Office, London.
- Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, Part II 1, 325–378.