Improving the convergence of Schedule-Based Dynamic Transit Assignment Models with capacity constraints

Lory Michelle Bresciani Miristice*, Guido Gentile

Department of Civil, Building and Environmental Engineering,
University of Rome La Sapienza, Rome, Italy

lorymichelle.brescianimiristice@uniroma1.it, guido.gentile@uniroma1.it
* Corresponding author

Extended abstract submitted for presentation at the 12th Triennial Symposium on Transportation Analysis conference (TRISTAN XII)
June 22-27, 2025, Okinawa, Japan

February 28, 2025

Keywords: dynamic user equilibrium, congested public transport, passenger overcrowding and queuing, hyper-paths, fixed-point algorithms.

1 Introduction

Public transportation is essential for urban mobility, offering a sustainable alternative to private vehicles. However, transit networks often suffer from overcrowding and queuing, leading to discomfort and delays. Accurate simulation of passenger flows under dynamic congestion is crucial for optimizing transit performance and improving passengers' experience.

Dynamic Transit Assignment (DTA) models simulate passenger route choices and their interactions with the public transport network, taking into account the within-day evolution of travel demand and the effects of congestion. Macroscopic models, which aggregate travellers into flows, allow capturing broader system interactions, enhancing computational efficiency for large-scale networks (Bellei *et al.*, 2005). *Schedule-based* DTA models translate the temporal dimension into the structure of diachronic graphs (Nuzzolo & Russo, 1998), which represent the network over time. This approach assumes perfect knowledge of run timetables by passengers and thus differs from hyperpath-based *optimal strategies* (Nguyen & Pallottino, 1988, Spiess & Florian, 1989) where only the distribution of line frequency is considered for route choice. To capture dynamic passenger decision-making and their expectations under uncertainty, schedule-based models need to integrate fail-to-board probabilities (Hamdouch & Lawphongpanich, 2008), which account for passengers who cannot board arriving vehicles due to capacity constraints, by introducing a different kind of hyperpaths (Gentile & Noekel, 2016).

A meaningful comparison of design scenarios requires high algorithm precision that can be achieved by improving the speed of convergence to an equilibrium between demand and supply. Computational challenges arise due to the large size of diachronic graphs and the slow convergence of traditional fixed-point algorithms like the Method of Successive Averages (MSA), limiting real-time applicability for providing short-term forecasts of passenger volumes on vehicles and at stops. In this paper, we first formulate deterministic equilibrium on schedule-based transit networks with fail-to-board probabilities as a fixed-point problem by considering a one-to-one map through Gradient Projection (GP), instead of the classical one-to-many map of network loading. Then, we apply to this fixed-point problem a new solution method called Adaptive Trust Contraction (TC-A) algorithm (Gentile *et al.*, 2024). The result is linear convergence for the analysed DTA models with capacity constraints and strategic route choice.

2 Methodology

To model passenger flows under congestion, we formulate DTA as a fixed-point problem with iterate $\mathbf{p}$ (i.e., the vector of arc conditional probabilities or splitting rates by destination), where the mapping function models demand loading, network congestion and route choice. Arc conditional probabilities $\mathbf{p} = [\mathbf{r}, \mathbf{x}]$ reflect passenger route choices $\mathbf{r}$ at standard nodes and the outcome $\mathbf{x}$ of boarding at stops (fail-to-board probabilities), which may fail due to insufficient capacity.

Following the implicit path enumeration approach proposed in Gentile (2016), we define arc flows $\mathbf{q}$ as the result of loading the origin-destination demand $\mathbf{d}$ on the network based on given arc conditional probabilities $\mathbf{p}$:

$$\mathbf{q} = q(\mathbf{p}; \mathbf{d}). \tag{1}$$

Arc flows $\mathbf{q}$ allow computing arc costs $\mathbf{c}$ and fail-to-board probabilities $\mathbf{x}$ that take into account passenger congestion (respectively, overcrowding and queuing) based on the supply characteristics $\mathbf{s}$ (essentially, the capacity constraints, since line speeds are embedded into the structure of the diachronic graph):

$$\mathbf{c} = c(\mathbf{q}; \mathbf{s}), \tag{2}$$

$$\mathbf{x} = x(\mathbf{q}; \mathbf{s}). \tag{3}$$

Local (deterministic) route choices $\mathbf{r}$ are the result of shortest hyper-tree computations by destination, which yield the arc weights $\mathbf{w}$ (i.e., the cost to reach a destination using a given arc, conditional on being at its tail):

$$\mathbf{w} = w(\mathbf{c}, \mathbf{x}), \tag{4}$$

$$\mathbf{r} \in r(\mathbf{w}). \tag{5}$$

When two or more local alternatives have the same minimum cost there are infinite distribution of users that satisfy Wardrop principles (Wardrop, 1952), therefore the deterministic map $r(\mathbf{w})$ yields a set of points. Finally, arc conditional probabilities are updated using a convergence method. This process continues with the next iteration until convergence criteria are met.

Traditionally, the problem is formulated considering as fixed-point mapping the *one-to-many* operator provided by the deterministic network loading map just described, which is given by the composition of equations 1, 2, 3, and 4:

$$\mathbf{p} \in [r(\hat{w}(\mathbf{p})), x(q(\mathbf{p}))], \tag{6}$$

where

$$\hat{w}(\mathbf{p}) = w(c(q(\mathbf{p})), x(q(\mathbf{p}))). \tag{7}$$

The issue is that the typical all-or-nothing assignment to shortest hyper-paths $r^*(\mathbf{w})$ that is adopted to implement the deterministic network loading map leads to solutions away from equilibrium. MSA solves this kind of fixed-point problems by averaging all solutions obtained by applying the map, but it suffers from slow convergence due to decreasing step sizes:

$$\mathbf{p}^{(i+1)} = \mathbf{p}^{(i)} + \frac{1}{1 + \gamma \cdot i} \cdot \left(f^*(\mathbf{p}^{(i)}) - \mathbf{p}^{(i)} \right), \tag{8}$$

$$f^*(\mathbf{p}) = [r^*(\hat{w}(\mathbf{p})), x(q(\mathbf{p}))] \tag{9}$$

where $\mathbf{p}^{(i)}$ is the solution at iteration i. To mitigate this, a reduction factor $\gamma < 1$ can be adopted to speed up convergence, lowering the weight of older iterates. Anyhow, the fail-to-board mechanism introduces additional non-linearities, increasing the model complexity and causing oscillations in cost updates, which further slow convergence.

To overcome these limitations, we adopt the GP as a fixed-point mapping. We have equilibrium when the route choices $\mathbf{r}$ subtracted by the arc weights $\mathbf{w}$ (scaled by a factor σ) and projected onto the feasible region defined by flow conservation yields again the same point:

$$\mathbf{p} = f(\mathbf{p}) \equiv [\mathrm{Proj}_R \left(r(\hat{w}(\mathbf{p})) - \sigma \cdot \hat{w}(\mathbf{p}) \right), x(q(\mathbf{p}))], \tag{10}$$

Arc weights are the gradient of the Beckmann integral corresponding to our local deterministic model, i.e. the objective function of an optimization program whose solution is an equilibrium. Due to non-separable congestion effects in transit networks, the Beckmann integral is not directly applicable, and we can formulate the equilibrium conditions as a Variational Inequality problem. In this paper, we prefer to introduce directly the above equivalent fixed point problem. Pure GP algorithms may still face challenges with convergence, which is why the step size is further scaled by an MSA-like factor, proportional to the number of iterations (Gentile, 2016).

We enhance the GP approach with the Adaptive Trust Contraction (TC-A) algorithm. The TC-A is an application of the feasible direction method (iterating a step along a descent direction) for nonlinear optimization on convex sets, with the objective function being the sum of squared residuals. It operates under the assumption that the residuals $f(\mathbf{p}) - \mathbf{p}$ indicate a descent direction relative to the sum of squared residuals. TC-A identifies the step α from $\mathbf{p}$ to $f(\mathbf{p})$ by escalating it by a factor $1 + \gamma_1 > 1$ if the last iterate enhanced the sum of squared residuals compared to the prior iteration; conversely, if no improvement is observed, the step is adjusted downwards by a factor $1 + \gamma_2 > 1$. In this scenario, both the current iterate and the search direction are consistently updated based on the assumption that the method will converge for a fixed, yet unknown, step size. The adjustment formula for $\alpha^{(i)}$ is articulated as following:

$$\alpha^{(i)} = \begin{cases} \min\left(1, \alpha^{(i-1)} \cdot (1 + \gamma_1)\right), & \text{if } y^{(i)} < y^{(i-1)} \\ \dfrac{\alpha^{(i-1)}}{1 + \gamma_2}, & \text{if } y^{(i)} \geq y^{(i-1)} \end{cases} \tag{11}$$

where $y^{(i)}$ is the sum of squared residuals at iteration i. Suggested values for the adjustment factors are $\gamma_1 = 0.1$ and $\gamma_2 = 0.5$.

3 Results and Discussion

We evaluated the performance of GP and TC-A versus pure MSA using a relative gap criterion. Testing was first conducted on a simple network, with 60 passengers arriving at a rate of 1 per minute over one hour. This network consisted of a single transit line connecting two stops, with four transit runs departing every 20 minutes, starting at minute 11. In an unconstrained scenario with a 20-passenger capacity per run, passengers distributed as expected across the four runs: 11 passengers boarded the first run, 20 the second and third runs, and 9 the last. With no boarding delays, both MSA and TC-A reached convergence in one iteration. To simulate congestion, we introduced capacity constraints of 18, 15, and 10 passengers per run, representing slight, moderate, and severe congestion. Figure 1 shows the convergence trends across these scenarios, underscoring TC-A's faster adaptability in real-time scenario optimization.

At an 18-passenger capacity, some fail-to-board events occurred, with passengers from the second and third runs needing to queue for later departures. TC-A handled this level efficiently, reaching convergence in 3 iterations, while MSA required 15 iterations to achieve a relative gap of 10^{-3}. With a reduced capacity of 15 passengers, queuing effects compounded, leading to delays across the second, third, and fourth runs. Here, TC-A reached convergence in 9 iterations, maintaining efficiency even as cumulative delays prevented 4 passengers from completing their trips; MSA required 80 iterations to reach a gap of 10^{-3}. Under a severe 10-passenger capacity, fail-to-board events intensified across all runs, leaving 20 passengers unserved due to demand exceeding capacity. TC-A reached convergence within 10 iterations under these conditions, while MSA took over 100 iterations to reach a relative gap of 10^{-3}.

We further tested TC-A on a medium-sized network simulating peak conditions in Singapore, with 38,000 passengers, 105 zones, and 72 transit lines. The assignment graph contained 688,362 arcs, representing 800 vehicle runs with 100-passenger capacity. On an Intel Core i7-7700 processor, TC-A reduced computation time by 14% compared to MSA, reaching a relative gap of

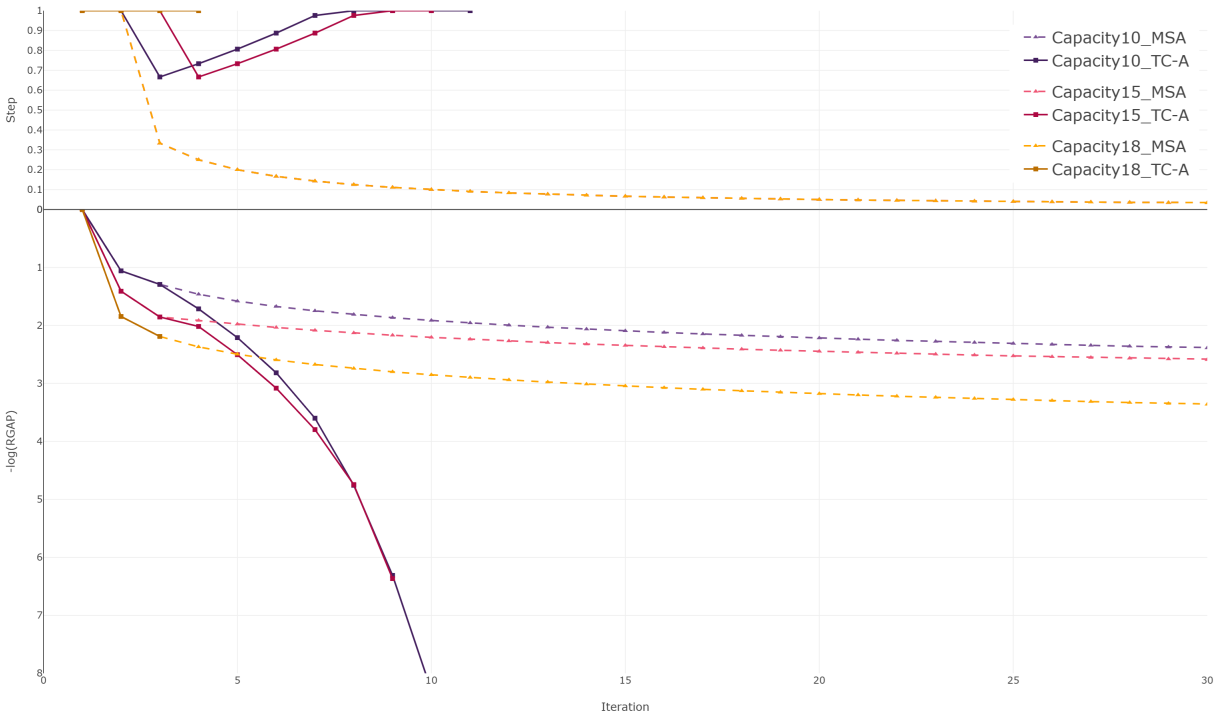

Figure 1 – *Convergence and step size trends of MSA and TC-A across different congestion levels. The trends of step size for congested MSA (capacity 18, 15, and 10) overlap.*

10^{-3} in 74 seconds (9 iterations), while MSA took 83 seconds (17 iterations). TC-A also achieved higher precision, reaching a relative gap of 10^{-4} in 522 seconds (47 iterations), while MSA did not meet this threshold within 100 iterations.

The results demonstrate that TC-A outperforms MSA in both convergence speed and precision, particularly under capacity constraints and in larger, congested networks. TC-A's adaptive step sizing effectively manages non-linearities from fail-to-board probabilities, reducing oscillations and computation time, making it ideal for real-time high-demand transit optimization.

References

Bellei, Giuseppe, Gentile, Guido, & Papola, Natale. 2005. A within-day dynamic traffic assignment model for urban road networks. *Transportation Research Part B: Methodological*, **39**(1), 1–29.

Gentile, Guido. 2016. Solving a Dynamic User Equilibrium model based on splitting rates with Gradient Projection algorithms. *Transportation Research Part B: Methodological*, **92**(10), 120–147.

Gentile, Guido, & Noekel, Klaus. 2016. Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems.

Gentile, Guido, Eldafrawi, Mohamed, & Bresciani Miristice, Lory Michelle. 2024. *Applying the Trust Contraction Algorithm for Solution of Stochastic User Equilibrium.* Presented at SIDT XXXVI Scientific Seminar, Cagliari, Italy, June 12–14, 2024.

Hamdouch, Younes, & Lawphongpanich, Siriphong. 2008. Schedule-based transit assignment model with travel strategies and capacity constraints. *Transportation Research Part B: Methodological*, **42**(7), 663–684.

Nguyen, S., & Pallottino, S. 1988. Equilibrium traffic assignment for large scale transit networks. *European Journal of Operational Research*, **37**, 176–186.

Nuzzolo, A., & Russo, F. 1998. A dynamic network loading model for transit services. *Proceedings of Tristan III.*

Spiess, Heinz, & Florian, Michael. 1989. Optimal strategies: A new assignment model for transit networks. *Transportation Research Part B: Methodological*, **23**(2), 83–102.

Wardrop, John Glen. 1952. Road Paper. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, **1**(3), 325–362.