

4.0 Alternative Formulations of the Static O-D Problem

4.1 Introduction

Entropy maximization and information minimization techniques have been used to solve a number of transportation problems (Wilson 1970). The application of the entropy maximization principles to the static O-D estimation problem was initially proposed by Willumsen (1978). Willumsen demonstrated that by maximizing the entropy, the most likely trip matrix could be estimated subject to a set of constraints.

This section describes two formulations that were proposed by Willumsen and Van Zuylen (1980) to solve the static O-D problem. Furthermore, this section describes the assumptions that were made in order to solve the problem analytically.

4.2 Trip Formulation and Assumptions

Let the total number of O-D trips equal T and the number of trips traveling between origin i and destination j equal T_{ij} . Then the number of ways (defined as entropy) in which T trips can be divided into groups of T_{ij} trips (without replication) can be computed using Equation 4.1, subject to Equations 4.2a and 4.2b. All variables are described in the list of symbols, placed after the list of tables.

$$\text{Maximize: } Z(T_{ij}) = \prod_{ij} \binom{T_n}{T_{ij}} = \frac{T!}{\prod_{ij} T_{ij}!} \quad (4.1)$$

Subject To:

$$V_a = \sum_{ij} T_{ij} p_{ij}^a \quad \forall a \quad (4.2a)$$

$$T_{ij} \geq 0 \quad \forall i, j \quad (4.2b)$$

Where:

T_k Number of trips available in the k^{th} cell of a $p \times p$ O-D matrix.

$T_n = T - \sum_{k=1}^{k=n-1} T_k$ Total no. of trips available to be considered in selecting T_{ij} trips.

$$T = \sum_{ij} T_{ij}$$

$$k = i + (j - 1) \times p$$

The T_{ij} values that maximize the entropy function represent the most likely solution (highest probability of occurring). By taking the natural logarithm of the entropy function, the product terms are substituted for addition terms which simplifies the computation, as demonstrated in Equation 4.3.

$$\text{Maximize: } Z(T_{ij}) = \text{Ln} \left(\frac{T!}{\prod_{ij} T_{ij}!} \right) = \text{Ln}(T!) - \sum_{ij} \text{Ln}(T_{ij}!) \quad (4.3)$$

Stirling (Donald, 1976) proposed an approximation for $\text{Ln}(x!)$ when x is large, as demonstrated in Equation 4.4. The details of the approximation are provided in Appendix A.

$$\text{Ln}(x!) = x \text{Ln} x - x \quad (4.4)$$

Applying Stirling's approximation to Equation 4.3, the objective function is simplified to Equation 4.5.

$$\text{Maximize: } Z(T_{ij}) = T \text{Ln} T - T - \sum_{ij} (T_{ij} \text{Ln} T_{ij} - T_{ij}) \quad (4.5)$$

Willumsen (1984) makes the further assumption that T is constant. It must be noted that this assumption is not valid for all network and traffic characteristics. Based on this assumption, the first two terms can be dropped and the formulation becomes the objective function as defined in Equation 4.6, subject to the constraints of Equations 4.7a and 4.7b.

$$\text{Maximize: } Z(T_{ij}) = -\sum_{ij} (T_{ij} \ln T_{ij} - T_{ij}) \quad (4.6)$$

Subject To:

$$V_a = \sum_{ij} T_{ij} p_{ij}^a \quad \forall a \quad (4.7a)$$

$$T_{ij} \geq 0 \quad \forall i, j \quad (4.7b)$$

In the case where a prior matrix (seed matrix) t_{ij} is available, the entropy function is defined using Equation 4.8. Equation 4.8 is very similar to Equation 4.1 except that an additional term is added which includes the probability of a trip being between an origin/destination pair for each of the trips T_{ij} . The constraints defined in Equations 4.9a and 4.9b are identical to those defined in Equations 4.1a and 4.1b.

$$\text{Maximize: } Z(T_{ij}, t_{ij}) = \frac{T!}{\prod_{ij} (T_{ij}!)} \prod_{ij} \left(\frac{t_{ij}}{\sum_{ij} t_{ij}} \right)^{T_{ij}} \quad (4.8)$$

Subject To:

$$V_a = \sum_{ij} T_{ij} p_{ij}^a \quad \forall a \quad (4.9a)$$

$$T_{ij} \geq 0 \quad \forall i, j \quad (4.9b)$$

Where:

$$T = \sum_{ij} T_{ij}$$

The above function can be simplified to Equation 4.10.

$$\text{Maximize: } Z(T_{ij}, t_{ij}) = \frac{T! \left(\sum_{ij} t_{ij} \right)^{-\sum T_{ij}}}{\prod_{ij} (T_{ij}!)} \prod_{ij} (t_{ij})^{T_{ij}} \quad (4.10)$$

Given that T is the summation of T_{ij} and t is the summation of t_{ij} , the objective function is simplified to Equation 4.11.

$$\text{Maximize: } Z(T_{ij}, t_{ij}) = \frac{T!(t)^{-T}}{\prod_{ij} (T_{ij}!)} \prod_{ij} (t_{ij})^{T_{ij}} \quad (4.11)$$

Based on Willumsen's (1984) assumption of constant total number of trips (i.e. T and t are constants), Equation 4.11 can be simplified to Equation 4.12.

$$\text{Maximize: } Z(T_{ij}, t_{ij}) = \frac{\prod_{ij} (t_{ij})^{T_{ij}}}{\prod_{ij} (T_{ij}!)} \quad (4.12)$$

By taking the natural logarithm of Equation 4.12 and expanding Equation 4.12 using Stirling's approximation, the objective function is simplified to Equation 4.13.

$$\text{Maximize: } Z(T_{ij}, t_{ij}) = -\sum_{ij} \left(T_{ij} \text{Ln} \left(\frac{T_{ij}}{t_{ij}} \right) - T_{ij} \right) \quad (4.13)$$

Subject To:

$$V_a = \sum_{ij} T_{ij} p_{ij}^a \quad \forall a \quad (4.14a)$$

$$T_{ij} \geq 0 \quad \forall i, j \quad (4.14b)$$

Where:

$$T = \sum_{ij} T_{ij}$$

$$t = \sum_{ij} t_{ij}$$

Willumsen (1981, 1984) subtracted a constant term from the objective function (summation of t_{ij}), as demonstrated in Equation 4.15. This constant term does not alter the solution; however, it simplifies the objective function, as will be demonstrated later.

$$\text{Maximize: } Z(T_{ij}, t_{ij}) = - \sum_{ij} \left(T_{ij} \text{Ln} \left(\frac{T_{ij}}{t_{ij}} \right) - T_{ij} + t_{ij} \right) \quad (4.15)$$

Substituting the Taylor series for $\text{Ln}(x)$ when $x \geq 0.5$ (Equation 4.16) into Equation 4.15 and only including the first two terms assuming that T_{ij} is sufficiently close to t_{ij} ($T_{ij}/t_{ij} \approx 1$) results in Equation 4.17. After simplifying Equation 4.17, the problem becomes a minimization problem, as demonstrated in Equation 4.20 subject to a number of constraints (Equations 4.21a, 4.21b, and 4.21c; Willumsen and Van Zuylen, 1980).

$$\text{Ln}(x) = \left(\frac{x-1}{x} \right) + \frac{1}{2} \left(\frac{x-1}{x} \right)^2 + \frac{1}{3} \left(\frac{x-1}{x} \right)^3 + \dots \quad (4.16)$$

$$\text{Maximize: } Z(T_{ij}, t_{ij}) = - \sum_{ij} \left(T_{ij} \left(\frac{\frac{T_{ij}}{t_{ij}} - 1}{\frac{T_{ij}}{t_{ij}}} \right) + \frac{T_{ij}}{2} \left(\frac{\frac{T_{ij}}{t_{ij}} - 1}{\frac{T_{ij}}{t_{ij}}} \right)^2 - T_{ij} + t_{ij} \right) \quad (4.17)$$

$$\text{Maximize: } Z(T_{ij}, t_{ij}) = - \sum_{ij} \left(T_{ij} - t_{ij} + \frac{T_{ij}}{2} \left(\frac{\frac{T_{ij}}{t_{ij}} - 1}{\frac{T_{ij}}{t_{ij}}} \right)^2 - T_{ij} + t_{ij} \right) \quad (4.18)$$

$$\text{Maximize: } Z(T_{ij}, t_{ij}) = -\sum_{ij} \left(T_{ij} - t_{ij} + \frac{1}{2T_{ij}} (T_{ij} - t_{ij})^2 - T_{ij} + t_{ij} \right) \quad (4.19)$$

$$\text{Minimize: } Z(T_{ij}, t_{ij}) = \sum_{ij} \left(\frac{1}{2T_{ij}} (T_{ij} - t_{ij})^2 \right) \quad (4.20)$$

Subject To:

$$V_a = \sum_{ij} T_{ij} p_{ij}^a \quad \forall a \quad (4.21a)$$

$$T_{ij} \geq 0 \quad \forall i, j \quad (4.21b)$$

Where:

$$\frac{T_{ij}}{t_{ij}} \approx 1$$

Given that Equation 4.20 is based on the assumption the $T_{ij} \approx t_{ij}$, the T_{ij} term in the denominator can be replaced with t_{ij} to generate a simple second-degree minimization problem, as defined in Equation 4.22 subject to the constraints 4.23a and 4.23b. The final formulation presented in Equation 4.22 is not provided in the literature; instead it is proposed here. The advantage of this formulation is that the denominator term (t_{ij}) is known *a priori*, unlike the formulation of Equation 4.20 in which the denominator includes variables being solved for (T_{ij}). Consequently, unlike in the case of Equation 4.20, Equation 4.22 does not require an iterative type of algorithm in order to find the optimum solution.

The objective function presented in Equation 4.22 represents a weighted regression in which the squared error about the seed matrix is minimized. For a uniform seed, the terms t_{ij} all become equal, and the objective function becomes a standard regression problem.

$$\text{Minimize: } Z(T_{ij}, t_{ij}) = \sum_{ij} \left(\frac{1}{2t_{ij}} (T_{ij} - t_{ij})^2 \right) \quad (4.22)$$

Subject To:

$$V_a = \sum_{ij} T_{ij} p_{ij}^a \quad \forall a \quad (4.23a)$$

$$T_{ij} \geq 0 \quad \forall i, j \quad (4.23b)$$

Where:

$$\frac{T_{ij}}{t_{ij}} \approx 1$$

Using Lagrangian Multiplier, the problem can be transformed from a constrained optimization to an unconstrained optimization, as demonstrated in Equation 4.24. The objective function includes additional terms equivalent to the number of links for which observed flows are available. These terms are multiplied by variables (λ_a) that are solved for. However, due to the fact that flow continuity does not necessarily exist in most practical applications, Equation 4.24 is not necessarily solvable. Consequently, the formulation that is presented in Equation 4.25 is proposed. In this formulation, the objective function is to minimize a weighted combination of the link flow and trip squared error. A constant weighting factor is applied to the link flow error term. This formulation (Equation 4.25) is easily solved using standard regression tools (spreadsheets) and most importantly, provides a unique solution.

$$\text{Minimize: } Z'(T_{ij}, t_{ij}) = \sum_{ij} \left(\frac{1}{2t_{ij}} (T_{ij} - t_{ij})^2 \right) + \sum_a \lambda_a \left(V_a - \sum_{ij} T_{ij} p_{ij}^a \right) \quad (4.24)$$

$$\text{Minimize: } Z''(T_{ij}, t_{ij}) = \sum_{ij} \left(\frac{1}{2t_{ij}} (T_{ij} - t_{ij})^2 \right) + \lambda \sum_a \left(V_a - \sum_{ij} T_{ij} p_{ij}^a \right)^2 \quad (4.25)$$

In summary, the initial formulation (Equation 4.8 with constraints 4.9a and 4.9b) provides the most likely O-D matrix subject to constraints of flow continuity and non-negativity. Applying Stirling's approximation and taking the natural logarithm of the objective function, the objective function is simplified (Equation 4.15). By expanding the

terms of Equation 4.15 using a Taylor series and based on the assumption that the estimated O-D table is close to the seed O-D table ($T_{ij} \approx t_{ij}$), the formulation can be further simplified (Equation 4.20 with constraints 4.21a and 4.21b). Furthermore, it is proposed in this paper that the term T_{ij} in the denominator be replaced by the term t_{ij} , allowing the problem to be solved using regression (Equation 4.22 with constraints 4.23a and 4.23b).

4.3 Volume Formulation and Assumptions

An alternative approach to solving the O-D problem is to formulate the problem in terms of the link volumes instead of O-D trips. The volume formulation was proposed by Willumsen and Van Zuylen (1980), as demonstrated in Equations 4.30, 4.31a, 4.31b and 4.31c. However, a systematic comparison of the two approaches has not been undertaken to date. The approach followed by Van Zuylen (1978) is based on Brillouin's measure of information. This formulation is similar to the volume formulation (Equation 3.3) in Chapter 3. The difference being in the interpretation of the formulation.

The information contained in a set of N observations where the state k has been observed n_k times is defined by Brillouin (1956) as:

$$I = -Ln \left(N! \prod_k \left(\frac{q_k^{n_k}}{n_k!} \right) \right) \quad (4.26)$$

where q_k is the a priori probability of observing state k . If the observations are counts on a particular link then one can define state ij as the state in which the vehicle observed has been travelling between origin i to destination j . So,

$$n_{ij}^a = T_{ij} p_{ij}^a \quad (4.27)$$

We can also express the a priori probability of observing state ij on link a as a function of a priori information about the O-D matrix as:

$$q_{ij}^a = \frac{t_{ij} p_{ij}^a}{\sum_{ij} t_{ij} p_{ij}^a} \quad (4.28)$$

where t_{ij} is the a priori number of trips between i and j . The information contained in V_a counts on link is then:

$$I_a = -Ln \left(\frac{V_a! \prod_{ij} \left(\frac{t_{ij} p_{ij}^a}{v_a} \right)^{T_{ij} p_{ij}^a}}{\prod_{ij} (T_{ij} p_{ij}^a)!} \right) \quad (4.29)$$

Now the objective function becomes:

$$\text{Minimize: } Z(T_{ij}, t_{ij}) = -\sum_a Ln \left(\frac{V_a! \prod_{ij} \left(\frac{t_{ij} p_{ij}^a}{v_a} \right)^{T_{ij} p_{ij}^a}}{\prod_{ij} (T_{ij} p_{ij}^a)!} \right) \quad (4.30)$$

Subject To:

$$V_a = \sum_{ij} T_{ij} p_{ij}^a \quad \forall a \quad (4.31a)$$

$$v_a = \sum_{ij} t_{ij} p_{ij}^a \quad \forall a \quad (4.31b)$$

$$T_{ij} \geq 0 \quad \forall i, j \quad (4.31c)$$

As was the case with the trip formulation, using Stirling's approximation, the objective function is simplified to Equation 4.36.

$$\text{Minimize: } Z(T_{ij}, t_{ij}) = -\sum_a \text{Ln} \left(\frac{V_a! \prod_{ij} \left(\frac{t_{ij} p_{ij}^a}{v_a} \right)^{T_{ij} p_{ij}^a}}{\prod_{ij} (T_{ij} p_{ij}^a)!} \right) \quad (4.32)$$

$$\text{Minimize: } Z(T_{ij}, t_{ij}) = -\sum_a \left(V_a \text{Ln}(V_a) - V_a + \sum_{ij} T_{ij} p_{ij}^a \text{Ln} \left(\frac{t_{ij} p_{ij}^a}{v_a T_{ij} p_{ij}^a} \right) + \sum_{ij} T_{ij} p_{ij}^a \right) \quad (4.33)$$

$$\text{Minimize: } Z(T_{ij}, t_{ij}) = -\sum_a \left(\sum_{ij} T_{ij} p_{ij}^a \text{Ln}(V_a) - V_a + \sum_{ij} T_{ij} p_{ij}^a \text{Ln} \left(\frac{t_{ij} p_{ij}^a}{v_a T_{ij} p_{ij}^a} \right) + V_a \right) \quad (4.34)$$

$$\text{Minimize: } Z(T_{ij}, t_{ij}) = -\sum_a \sum_{ij} T_{ij} p_{ij}^a \text{Ln} \left(\frac{V_a t_{ij}}{T_{ij} v_a} \right) \quad (4.35)$$

$$\text{Minimize: } Z(T_{ij}, t_{ij}) = \sum_a \sum_{ij} T_{ij} p_{ij}^a \text{Ln} \left(\frac{T_{ij} v_a}{V_a t_{ij}} \right) \quad (4.36)$$

4.4 Summary

In this Chapter we have seen the derivations of the various trip and volume formulations as well as their assumptions. In the following Chapter, two different approaches to obtain the most likely O-D trip table, taking into account the lack of flow continuity at nodes due to inconsistencies in data is explored. The first approach tries to solve the problem by effectively dividing the problem into two sub problems, and then integrating the output of the first to be a part of the input in the second and solving the second part. The other approach is to effectively solve the problem in one pass.