

# Biproportional Matrix Balancing with Upper Bounds

Heinz Spiess \*

November 1996

## Abstract

The purpose of the note is to look at the problem of biproportional matrix balancing when the upper bounds are imposed on the matrix elements. This problem can be formulated as a convex minimization problem. Using the Kuhn-Tucker optimality conditions the functional form of the resulting model is derived. The dual formulation of the problem is derived and it is shown how it can be solved by a cyclic coordinate descent method. This leads to the proposal of an efficient solution algorithm.

*→ First draft! — Subject to revision! ←*  
*Please let me know of any errors and typos you find - thanks!*

---

\*EMME/2 Support Center, Haldenstrasse 16, CH-2558 Aegerten, Switzerland

## 1 Introduction

The purpose of this note is to analyze an extension of the standard biproportional matrix balancing (Murchland 1977), also referred to as “Fratar” (Lamond and Steward 1981), “Furness method” (Furness 1970), “maximum entropy transportation model” or “maximum entropy spacial interaction model” (Florian ...?), and referred to as “two-dimensional matrix balancing” in the EMME/2 transportation planning package (Spiess 1984, INRO 1996).

As in the standard two-dimensional balancing, a given matrix (containing synthetic friction factors or “prior” matrix data) is “factored up” to a fit a given set of productions and attractions. But in addition to satisfying the given productions and attractions, upper bounds are now imposed on the elements of the resulting matrix. Such bounds can be thought to represent transport capacities, but they can also be simply used to limit the growth of individual matrix elements with respect to the prior matrix.

In the next section we briefly introduce at the standard biproportional matrix balancing model and its optimization formulation as a maximum entropy model. In section 3 the extended model is formulated and analyzed. In section 4, finally, an efficient solution algorithm is formulated, which is based on the coordinate descent method applied to the dual problem.

## 2 Standard Biproportional Matrix Balancing

The problem of the standard two-dimensional matrix balancing essentially consist in finding appropriate factors  $a_p$  and  $b_q$  for each origin  $p \in P$  and destination  $q \in Q$  so that when multiplied to the corresponding rows and columns of the prior matrix  $G_{pq}$  the marginal totals of resulting matrix  $g_{pq}$  correspond to the given productions  $O_p$  and attractions  $D_q$ . This can be expressed with the following set of equations:

$$g_{pq} = a_p b_q G_{pq}, \quad pq \in PQ, \quad (1)$$

$$\sum_{q \in Q} g_{pq} = O_p, \quad p \in P, \quad (2)$$

$$\sum_{p \in P} g_{pq} = D_q, \quad q \in Q. \quad (3)$$

It is easy to show that the above problem can also be expressed as the following convex optimization problem (also referred to as maximum entropy transportation problem):

$$\text{Min} \sum_{pq \in PQ} g_{pq} (\log g_{pq} - \log G_{pq} - 1) \quad (4)$$

subject to (2) and (3).

Defining  $\alpha_p$  and  $\beta_q$  as the dual variable associated with constraints (10) and (11), we obtain from the Kuhn-Tucker optimality conditions the following functional form

$$g_{pq} = e^{-\alpha_p - \beta_q} G_{pq}, \quad pq \in PQ, \quad (5)$$

which correspond exactly to (1) when using the following substitutions

$$a_p = e^{-\alpha_p}, \quad p \in P, \quad (6)$$

$$b_q = e^{-\beta_q}, \quad q \in Q. \quad (7)$$

The dual formulation of problem (4) can be written as the following unconstrained optimization problem:

$$\text{Min}_{\alpha, \beta} \sum_{pq \in PQ} G_{pq} e^{-\alpha_p - \beta_q} + \sum_{p \in P} \alpha_p O_p + \sum_{q \in Q} \beta_q D_q. \quad (8)$$

The standard biproportional matrix balancing model is known to be always feasible when  $\sum_{p \in P} O_p = \sum_{q \in Q} D_q$  and  $G_{pq} > 0$  for all  $pq \in PQ$ .

Problems which have  $G_{pq} = 0$  for some O-D pairs can also be handled in this framework. But in this case, the O-D pairs with  $G_{pq} = 0$  are not to be considered part of the set  $PQ$  and  $g_{pq} = 0$  is used for these O-D pairs where necessary, e.g. where it appears in a sum over origins or destinations.

Note, however, that problems with  $G_{pq} = 0$  for some O-D pairs are not always feasible. In this case, there is no easy method to determine the feasibility of a problem a priori, so that the primal feasibility of a solution found by a dual algorithm always has to be checked explicitly.

### 3 Biproportional Matrix Balancing with Upper Bounds

We now turn our attention to an extension of the model defined in the previous section in which explicit upper bounds  $U_{pq}$  are imposed on the resulting matrix elements  $g_{pq}$ . Since imposing upper bounds will have a direct impact on the functional form of the resulting model, it is not possible to formulate the extended model by simply adding a further constraint to problem (1)-(11).

However, the minimization problem (4) lends itself very well for an extension by adding upper bound constraints. This leads to the following extended problem formulation:

$$\text{Min} \sum_{pq \in PQ} g_{pq} (\log g_{pq} - \log G_{pq} - 1) \quad (9)$$

subject to

$$\sum_{q \in Q} g_{pq} = O_p, \quad p \in P, \quad (10)$$

$$\sum_{p \in P} g_{pq} = D_q, \quad q \in Q, \quad (11)$$

$$g_{pq} \leq U_{pq}, \quad pq \in PQ. \quad (12)$$

Using again  $\alpha_p$  and  $\beta_q$  as the dual variables associated with the production (=refprod) and attraction constraints (=refattr) and the new dual variables  $\mu_{pq} > 0$  for the upper bound constraints (12), we can state the Kuhn-Tucker optimality conditions for the above problem as follows:

$$\log g_{pq} - \log G_{pq} + \alpha_p + \beta_q + \mu_{pq} = 0, \quad pq \in PQ, \quad (13)$$

$$\mu_{pq}(U_{pq} - g_{pq}) = 0, \quad pq \in PQ. \quad (14)$$

The optimality condition (13) can be rewritten to yield the following functional form for the resulting model:

$$g_{pq} = G_{pq} e^{-\alpha_p - \beta_q - \mu_{pq}}, \quad pq \in PQ. \quad (15)$$

By applying the complementary slackness conditions (14) for the dual variable  $\mu_{pq}$  and distinguishing between the cases  $\mu_{pq} = 0$  and  $\mu_{pq} > 0$ , (15) becomes

$$g_{pq} = \left\{ \begin{array}{l} G_{pq} e^{-\alpha_p - \beta_q} = a_p b_q G_{pq} \quad \text{if } \mu_{pq} > 0 \\ U_{pq} \quad \text{otherwise} \end{array} \right\} \quad (16)$$

which can be written even more concisely as

$$g_{pq} = \text{Min}\{G_{pq} e^{-\alpha_p - \beta_q}, U_{pq}\} = \text{Min}\{a_p b_q G_{pq}, U_{pq}\}, \quad pq \in PQ. \quad (17)$$

Note that the above model formulation does no longer explicitly contain the dual variables  $\mu_{pq}$ .

Without loss of generality we can assume that  $U_{pq} > 0$  for all  $pq \in PQ$ , as matrix elements  $g_{pq}$  can always be forced to zero by setting  $G_{pq} = 0$  (and removing the corresponding O-D pairs from the set  $PQ$ , as explained in section 2).

The next step is to formulate the Lagrangian dual for problem (9) which becomes

$$\text{Min}_{\substack{\alpha, \beta \\ \mu \geq 0}} -L(\alpha, \beta, \mu) = \sum_{pq \in PQ} G_{pq} e^{-\alpha_p - \beta_q - \mu_{pq}} + \sum_{p \in P} \alpha_p O_p + \sum_{q \in Q} \beta_q D_q + \sum_{pq \in PQ} \mu_{pq} U_{pq} \quad (18)$$

Note that, except for the non-negativity of  $\mu_{pq}$ , the dual problem is essentially unconstrained. For this type of problem the a simple method of successive coordinate descent

is known to converge to the optimal solution (see e.g. Luenberger 1984). Thus, it can be used to solve the dual problem (18) and find the optimal values of the dual variables. If the primal problem (9) is feasible, the optimal values of the dual variables  $\alpha_p$  and  $\beta_q$  can be inserted into (17) to find the optimal solution of the primal problem.

Applying the coordinate descent method to the dual problem (18) means solving cyclically the first order optimality conditions for the corresponding dual variables. This implies finding the zeros for the partial derivatives of the dual objective function with respect to all  $\alpha_p$  and  $\beta_q$ , and all non-zero  $\mu_{pq}$ :

$$\partial L(\alpha, \beta, \mu) / \partial \alpha_p = -\sum_{q \in Q} G_{pq} e^{-\alpha_p - \beta_q - \mu_{pq}} + O_p = 0, \quad p \in P, \quad (19)$$

$$\partial L(\alpha, \beta, \mu) / \partial \beta_q = -\sum_{p \in P} G_{pq} e^{-\alpha_p - \beta_q - \mu_{pq}} + D_q = 0, \quad q \in Q, \quad (20)$$

$$\partial L(\alpha, \beta, \mu) / \partial \mu_{pq} = -G_{pq} e^{-\alpha_p - \beta_q - \mu_{pq}} + U_{pq} = 0, \quad \text{for all } \mu_{pq} > 0. \quad (21)$$

By observing that (21) is equivalent to (17), it is easy to see that conditions (19) and (21) can be combined into the simpler production conditions

$$\sum_{q \in Q} \text{Min}\{a_p b_q G_{pq}, U_{pq}\} = O_p, \quad p \in P, \quad (22)$$

and the conditions (20) and (21) combine into the simpler attraction conditions

$$\sum_{p \in P} \text{Min}\{a_p b_q G_{pq}, U_{pq}\} = D_q, \quad q \in Q. \quad (23)$$

Solving (22) for  $\alpha_p$  will simultaneously satisfy the first order conditions (19) for origin  $p$  **and** (21) for all  $\mu_{pq} > 0$  related to origin  $p$ . In the same manner solving (23) for  $\beta_q$  will simultaneously satisfy the first order conditions (20) for destination  $q$  **and** (21) for all  $\mu_{pq} > 0$  related to destination  $q$ .

Before turning our attention to the solution algorithm, let us briefly look at the question of feasibility. The introduction of upper bounds influences the feasibility in a similarly complex manner as do the zeroes in the matrix  $G_{pq}$ . While of course the conditions

$$\sum_{q \in Q} U_{pq} \geq O_p \quad p \in P, \quad (24)$$

$$\sum_{p \in P} U_{pq} \geq D_q \quad q \in Q \quad (25)$$

are necessary conditions for feasibility, they are by no means sufficient.

## 4 Solution Algorithm

The proposed solution algorithm can be classified as a coordinate descent method applied to the dual problem formulation. As it is based on iteratively solving equations (22) and

(23), it actually does not follow optimize along dual coordinates one by one but always satisfies the first order conditions simultaneously for all dual variables associated with either one origin or one destination, (which should result in an improved convergence compared to a one-by-one coordinate descent method).

Algorithm 1 describes the basic algorithm proposed to solve the biproportional matrix balancing problem with upper bounds (9). This algorithm assumes that equations of the type  $\sum_i \min\{x f_i, u_i\} = T$  can be solved for  $x$ . How this can be done efficiently is shown further on in Algorithm 2.

### Algorithm 1:

#### 0. Initialization

Set  $b_q^0 = 1$  for  $q \in Q$  and set  $k := 1$ .

#### 1. Balance origins

For each origin  $p \in P$  solve the equation

$$\sum_{q \in Q} \min\{a_p^k b_q^{k-1} G_{pq}, U_{pq}\} = O_p$$

for variable  $a_p^k$  by applying Algorithm 2.

#### 2. Balance destinations

For each destination  $q \in Q$  solve the equation

$$\sum_{p \in P} \min\{a_p^k b_q^k G_{pq}, U_{pq}\} = D_q$$

for variable  $b_q^k$  by applying Algorithm 2.

#### 3. Test stopping criteria

If convergence is reached for the the multipliers  $a_p$  and  $b_q$ , e.g.  $\|a^k - a^{k-1}\| + \|b^k - b^{k-1}\| < \epsilon_1$ , then go to step 4, otherwise set  $k := k + 1$  and return to step 1.

#### 4. Compute primal solution

If the maximum production constraint violation for all  $p \in P$  is smaller than some predefined tolerance value, i.e.

$$\left| \sum_{q \in Q} \min\{a_p^k b_q^k G_{pq}, U_{pq}\} - O_p \right| < \epsilon_2,$$

then compute the optimal primal solution

$$g_{pq} = \min\{a_p^k b_q^k G_{pq}, U_{pq}\},$$

otherwise the primal problem is infeasible. STOP.

Now let us consider the subproblem how to find the factor  $x$  which solves an equation of the form

$$\sum_i \min\{xf_i, u_i\} = T. \quad (26)$$

This type of problem occurs in step 1 of Algorithm 1 with  $i = q$ ,  $f_i = b_q^{k-1}G_{pq}$ ,  $u_i = U_{pq}$ , and in step 3 with  $i = q$ ,  $f_i = a_p^k G_{pq}$ ,  $u_i = U_{pq}$ . Without loss of generality we can assume that  $f_i > 0$  and  $u_i > 0$  for all  $i \in I$ . If there were any elements  $i$  with  $f_i = 0$  and/or  $u_i = 0$ , these can simply be dropped, since they do not influence the solution of (26) at all.

This type of problem can be solved by first sorting the elements  $i$  according to their  $u_i/f_i$  ratios and then scanning the elements linearly in this order and checking if the optimal value  $x$  lies in the range between of two consecutive  $u_i/f_i$  ratios. This leads to the formulation of the following algorithm:

**Algorithm 2:** (Solve  $\sum_i \min\{xf_i, u_i\} = T$ )

**0. Initialization**

Set  $F = \sum_i f_i$  and  $U = T$

**1. Sort elements**

Sort the elements  $i$  in increasing order of  $u_i : f_i$  ratios.

**2. Scan elements**

Using the order established in step 1, do for each  $i$ :

If  $u_i F \leq f_i U$  then

set  $F := F - f_i$  and  $U := U - u_i$ ,

else

go to step 3.

If this point is reached after having scanned all element, then the problem is infeasible, STOP.

**3. Compute optimal solution**

Set the solution to  $x = U/F$  and STOP.

So far, the above algorithms have only been implemented in a “quick and dirty” manner purely for testing purposes. In the few examples tested so far (all based on data taken from the standard EMME/2 Winnipeg data bank using 154 traffic zones), the algorithm has been found to converge quite rapidly (4-7 iterations) to the optimal solution. But of course, further tests are needed.

## 5 Conclusions

The methodology of introducing upper bounds is presented in this note only for the biproportional matrix balancing. However, it can most likely be applied directly to other related problems, such as the multi-proportional matrix balancing or Evans and Kirby's three-dimensional matrix balancing.

## References

- [1] Bregman L. (1967) *The Relaxation Method of Finding the Common Point of Convex Sets and its Application to the Solution of Problems in Convex Programming*. **U.S.S.R. Computational Math. Mathematical Phys.** 7, pp 200-217.
- [2] Evens S.P. and Kirby H.R. (1974) *A Three-Dimensional Furness Procedure for Calibrating Gravity Models*. **Transportation Research**, Vol 8, pp 105-122.
- [3] Furness K.P. (1970) *Time Function Interaction*. **Traffic Engineering and Control** Vol 7, No 7, pp19-36.
- [4] INRO Consultants Inc. (1996), *EMME/2 User's Manual*.
- [5] Lamond B. and Stewart N.F. (1981) *Bregman's Balancing Method*. **Transportation Research**, Vol 15B, pp 239-248.
- [6] Luenberger D.G. (1984) *Linear and Nonlinear Programming*. Second Edition, Addison-Wesley.
- [7] Murchland J. (1977) *The Multi-proportional Problem*. University College London, research note JDM 263.
- [8] Spiess H. (1984). *Contributions à la théorie et aux outils de planification de réseaux de transport urbain*. Ph.D. thesis, Département d'informatique et de recherche opérationnelle, Centre de recherche sur les transports, Université de Montréal, Publication 382.