Comparing Aggregated Generalized Transportation Models Using Error Bounds

Igor S. Litvinchev
Computing Center Russian Academy of Sciences
Vavilov 40, Moscow 119991, Russia

Socorro Rangel∗
Department of Computer Science and Statistics, UNESP
Rua Cristóvão Colombo, 2265
S.J. Rio Preto, SP, 15054-000, Brazil

September 2002

Abstract

This paper presents a comparative study of aggregated generalized transportation models using aggregation error bounds. A priori and a posteriori bounds were derived and a computational experiment was designed to a) study the correlation between the a priori bounds, the a posteriori bounds and the actual error, and b) quantify the difference of the error bounds from the actual error. The experiment shows significant correlation between some of a priori bounds, the a posteriori bounds and the actual error. These preliminary results indicate that calculating the

∗Corresponding author. E-mail address: socorro@dcce.ibilce.unesp.br
a priori error bound is a useful strategy to select the appropriate aggregation level since the a priori bound varies in the same way the actual error does.

Keywords: Large Scale Optimization, Generalized Transportation Model; Aggregation; Aggregation Error Bounds;

1 Introduction

One of fundamental issues in complex systems modeling is the trade-off between the level of detail to employ and the ease of using and solving the model. Aggregation/disaggregation techniques have been developed to facilitate analyses of large problems and provide the decision-maker a set of tools to compare large models that have a high degree of detail with smaller aggregated models that have less detail. Aggregated models represent approximately the original complex model and are used in various stages of decision-making process.

In optimization-based approaches the difference between the actual optimal objective function value and the aggregated optimal objective function value can be used as a measure of the error introduced by aggregation. Many upper bounds on this error have been developed for the aggregated optimization models and are referred to as error bounds [5,6,8,11-14]. Two types of error bounds are considered, a priori error bounds and a posteriori error bounds. A priori error bounds may be calculated after the model has been aggregated but before the smaller aggregated model has been optimized. A posteriori error bounds are calculated after the aggregated model has been optimized. Note that in both cases the actual error is estimated without optimization of the original detailed model. Error bounds allow the modeler to compare various types of aggregation and, if the error bounds are not adequate, modify the aggregation strategy in an attempt to tighten the error bounds [11].

There are two main questions associated with the error bounds. The first one is addressed to the quantitative quality of the bound, i.e. the difference of an error bound
from the actual error. From this point of view the closer the error bound to the actual error, the better. The second question is how to compare different aggregation strategies based on the error bounds. It is not obvious that the aggregation yielding the tightest error bound has the smallest actual error. Moreover different bound calculation techniques may result in different correspondence between the error bound and the actual error.

The first question above is relatively well explored, at least for the case of a posteriori bounds for the linear programing (see [5,8,11] and the references therein). The second area is much less investigated. To our knowledge the first comprehensive study in this direction has been made recently in [10] for the classical transportation problem (TP) where they explored, among others, the following questions:

- If a model is aggregated using different methods to create several different smaller models and an error bound (a priori and/or a posteriori) is calculated for each aggregated model, does the model with the tightest error bound(s) give the closest approximation of the objective function value?

- Does the aggregated model with the tightest a priori error bounds have the tightest a posteriori error bounds (and vice versa)?

Many applications of aggregation/disaggregation theory in optimization involve linear network flow problems and, in particular, the TP. For this problem both a priori and a posteriori error bounds were developed and investigated [11,13]. In the network flow problems the nodes traditionally correspond to constraints, and arcs to variables. For most aggregation schemes, aggregation of nodes (constraints) implies or necessitates a corresponding aggregation of arcs (variables). Usually, only variables and constraints of the same type (e.g., sources or destinations) are aggregated.

The TP minimizes the cost of transporting supplies from the set of sources to the set of destinations, and it is assumed that the transportation flow is conserved on every
arc between the source and the destination. In the generalized transportation problem (GTP) the amount of flow that leaves the source can differ from the amount of flow that arrives to the destination. A certain multiplier is associated with each arc to represent a "gain" ("loss") of a unit flow on the way between the source and the destination. There are two common interpretations of the arc multipliers [1]. In the first one the multipliers are treated as modifying the amount of flow of some particular item. This way it is possible to model situations involving physical or administrative transformations. For example, evaporation, seepage, or monetary growth due to the interest rates. In the second interpretation, the multipliers are used to transform one type of item into another (manufacturing, machine loading and scheduling, currency exchanges, etc.). For more detailed discussion on the GTP and its applications see [1,3] In contrast to the TP, much less work has been done in aggregation for the GTP. Only a posteriori error bounds for aggregation in the GTP have been developed in [2].

This paper is focused on aggregation bounds for the GTP and is organized as follows. In Section 2, we state the aggregated GTP and study its properties. In the next section, a family of a priori bounds for a class of the GTP is derived, and new a posteriori bounds are proposed. It is shown that the a priori bounds can be tightened by searching the duals. Section 4 presents a numerical study carried out to compare the quality of the error bounds and to determine if there are relationships between the error bounds and the actual error. The results of this preliminary study show that there is a significant difference among the various methods of calculating the error bounds. Moreover, it is possible to point out the improved a priori and a posteriori bounds which have a statistically significant correlation with the actual error.

2 The original and the aggregated models

The original GTP before aggregation is considered in the following form [2]:

\begin{align*}
\text{Minimize} \quad & \sum_{(i,j) \in A} c_{ij} x_{ij} \\
\text{Subject to} \quad & \sum_{j \in J} x_{ij} - \sum_{i \in I} x_{ij} = f_i, \quad i \in I \setminus \{s, t\} \\
& \sum_{i \in I} x_{ij} = f_j, \quad j \in J \\
& x_{ij} \geq 0, \quad (i,j) \in A
\end{align*}
\[
    z^* = \min \sum_{i \in S} \sum_{j \in T} c_{ij} x_{ij} \quad \text{(OP)}
\]

\[
    \sum_{j \in T} d_{ij} x_{ij} \leq a_i, \ i \in S,
\]

\[
    \sum_{i \in S} x_{ij} = b_j, \ j \in T,
\]

\[
    x_{ij} \geq 0 \quad \forall i, j.
\]

It is assumed that all the coefficients in (OP) are positive. We will refer to \( S \) and \( T \) as the set of all sources and all destinations (customers), respectively.

The role of sources and destinations in (OP) can be reversed [2] by a simple transformation of variables: \( y_{ij} = d_{ij} x_{ij} \). This transformation removes the arc multipliers from the first group of constraints in (OP) and places them to the second group. Bearing in mind this correspondence between the sources and the destinations we restrict ourselves to the aggregation of the destinations (customers) in (OP). The aggregation of the sources can be treated similarly.

The aggregation is defined by a partition of the set \( T \) of the destinations into a set \( T^A \) of clusters \( T_k, k = 1, ..., K \) such that \( T_k \cap T_p = \emptyset, k \neq p \) and \( \cup_{k=1}^K T_k = T \). Note that the columns in the (OP) are also partitioned/clustered respectively. The fixed-weights variable aggregated linear programming problem [14] is derived by replacing the columns in each cluster by their weighted average. The objective coefficients are combined similarly.

For each \( k \) consider the non-negative weights \( g_{ij} \) fulfilling the following normalizing conditions

\[
    \sum_{j \in T_k} g_{ij} = 1, \ g_{ij} \geq 0 \quad \forall i, j \quad (1)
\]

The aggregated problem can be obtained by fixing the normalized weights \( g_{ij} \) and substituting the linear transformation (disaggregation)

\[
    x_{ij} = g_{ij} X_{ik}, \ j \in T_k \quad (2)
\]
into (OP) [6]. Thus the aggregated problem is as follows:

$$z = \min \sum_{i \in S} \sum_{k \in T^A} X_{ik} \sum_{j \in T_k} c_{ij} g_{ij}$$  \hspace{1cm} (3)$$

$$\sum_{k \in T^A} X_{ik} \sum_{j \in T_k} d_{ij} g_{ij} \leq a_i, \ i \in S,$$  \hspace{1cm} (4)$$

$$\sum_{i \in S} g_{ij} X_{ik} = b_j, \ j \in T_k, \ k \in T^A,$$  \hspace{1cm} (5)$$

$$X_{ik} \geq 0 \ \ \ \forall i, k.$$  \hspace{1cm} (6)$$

To write down the problem (3)-(6) we have used the observation that clustering is simply a rearrangement of the indexes and hence: $$\sum_{j \in T_k} = \sum_{k \in T^A} \sum_{j \in T_k}.$$ 

From (1) and (2) it follows that for any normalized weights we have $$X_{ik} = \sum_{j \in T_k} x_{ij}$$ and hence $$X_{ik}$$ can be treated as the supply from the source $$i$$ to the aggregated customer $$T_k.$$ However, the problem (3)-(6) cannot be considered as some GTP associated with $$S$$ sources and $$T^A$$ destinations because the number of the restrictions (5) is $$|T|,$$ the same as in (OP). The number of constraints (5) can be reduced as follows.

Let the weights be chosen as $$g_{ij} = b_j / \sum_{j \in T_k} b_j, \ j \in T_k.$$ Then for each $$j \in T_k$$ the condition (5) turns out to $$\sum_{i \in S} X_{ik} = \sum_{j \in T_k} b_j$$ and the problem (3)-(6) becomes:

$$z = \min \sum_{i \in S} \sum_{k \in T^A} \bar{c}_{ik} X_{ik}$$  \hspace{1cm} (\widetilde{AP})$$

$$\sum_{k \in T^A} \bar{d}_{ik} X_{ik} \leq a_i, \ i \in S,$$  \hspace{1cm} $$

$$\sum_{i \in S} X_{ik} = b_k \text{ (repeted } |T_k| \text{ times), } k \in T^A,$$  \hspace{1cm} $$

$$X_{ik} \geq 0 \ \ \ \forall i, k,$$

where $$\bar{c}_{ik} = \sum_{j \in T_k} c_{ij} \bar{g}_{ij}, \ \bar{d}_{ik} = \sum_{j \in T_k} d_{ij} \bar{g}_{ij}, \ b_k = \sum_{j \in T_k} b_j.$$

Obviously, the redundant constraints in (\widetilde{AP}) can be relaxed without changing its primal optimal solution to form the problem which is usually referred to as the aggregated GTP [2]:

$$z = \min \sum_{i \in S} \sum_{k \in T^A} \bar{c}_{ik} X_{ik}$$  \hspace{1cm} (AP)
The primal solutions $\overline{X}_{ik}$ of $(\widetilde{AP})$ and (AP) are the same and the fixed-weight disaggregated solution $\overline{x}_{ij} = \overline{g}_{ij} \overline{X}_{ik}$, $j \in T_k$ is feasible to the original problem (OP). The following proposition shows that the problem dual to $(\widetilde{AP})$ always has multiple optimal solutions, even if the optimal dual solution to (AP) is unique.

**Proposition 1.** Let $\{\overline{u}_i, \overline{v}_k\}$ be an optimal dual solution to (AP). Denote

$$D(\overline{v}) = \{v_j, j \in T \mid \sum_{j \in T_k} v_j b_j = \overline{v}_k b_k, k = 1, \ldots, K\}.$$ Then any pair $\{\overline{u}, \hat{\overline{v}}\}$ with $\hat{\overline{v}} \in D(\overline{v})$ is an optimal dual solution to $(\widetilde{AP})$ and

$$z = -\sum_i a_i u_i + \sum_k \sum_j \hat{\overline{v}}_j b_j$$

for any $\hat{\overline{v}} \in D(\overline{v})$.(retirar no?) ja esta presente na linha anterior

**Proof.** Consider the dual to $(\widetilde{AP})$ or which is the same, the dual to (3)-(6) for $g_{ij} = \overline{g}_{ij}$:

$$-\sum_i a_i u_i + \sum_k \sum_{j \in T_k} v_j b_j \rightarrow \max$$

$$(\widetilde{DAP})$$

$$\overline{c}_{ik} + \overline{d}_{ik} u_i - \sum_{j \in T_k} v_j \overline{g}_{ij} \geq 0, \forall i, k,$$

$$u_i \geq 0, i \in S.$$ The dual to (AP) is as follows:

$$-\sum_i a_i u_i + \sum_k v_k \sum_{j \in T_k} b_j \rightarrow \max$$

$$(DAP)$$

$$\overline{c}_{ik} + \overline{d}_{ik} u_i - v_k \geq 0, \forall i, k,$$

$$\overline{u}_i \geq 0, i \in S.$$
These two dual problems should have equal optimal value since the primal problems (AP) and (\(\tilde{\text{AP}}\)) have the same optimal solutions. Let \(\{\tilde{u}_i, \tilde{v}_k\}\) be an optimal solution to (DAP). Construct \(\{\hat{u}_i, \hat{v}_j\}\) as follows: \(\hat{u}_i = u_i\) and \(\hat{v}_j\) is such that \(\sum_{j \in T_k} \hat{v}_j g_{ij} = \tilde{v}_k\). Comparing the restrictions of (DAP) and (D\(\tilde{\text{AP}}\)) we see that this \(\{\hat{u}_i, \hat{v}_j\}\) is feasible to (D\(\tilde{\text{AP}}\)). Moreover, since \(g_{ij} = b_j / \sum_{j \in T_k} b_j\) it is not hard to verify that the objective value of (D\(\tilde{\text{AP}}\)) for \(\{\hat{u}_i, \hat{v}_j\}\) is equal to the optimal objective of (DAP). Hence this \(\{\hat{u}_i, \hat{v}_j\}\) is optimal to (D\(\tilde{\text{AP}}\)).

Note, that the nonuniqueness of the optimal duals in (D\(\tilde{\text{AP}}\)) follows from the way the aggregation was performed, no matter whether the problem (AP) has a unique optimal dual solution or not. The nonuniqueness of the duals will be used further to derive a priori bounds for the GTP.

3 The error bounds

If the fixed weights \(g_{ij}\) are used to form the aggregated problem (AP), the optimal value \(\overline{z}\) of the objective function for (AP) is also an upper bound for the optimal objective \(z^*\) to the original problem, such that \(\overline{z} - z^* \geq 0\). This follows from the fact that the disaggregated solution \(\overline{x}_{ik} = \overline{g}_{ij} \overline{X}_{ik}, j \in T_k\) is feasible to the original problem.

To get an upper bound on the value \(\overline{z} - z^*\) we will use the following result established in [6] for the fixed-weight aggregation in convex programming: if \(W\) is a closed convex and bounded set containing an optimal solution to the original problem (such a \(W\) is called a localization), then \(\overline{z} - z^* \leq \min_{x \in W} L(x, \tau)\). Here \(L(\cdot, \cdot)\) is the standard Lagrange function associated with the original problem and \(\tau\) is a vector of Lagrange multipliers. Below, for simplicity, we assume that the condition \(x \geq 0\) is always included in the definition of \(W\).

For the GTP the Lagrangian has the form

\[
L(x, u, v) = -\sum_{i \in S} u_i a_i + \sum_{j \in T} v_j b_j + \sum_{i \in S} \sum_{j \in T} (c_{ij} + u_i d_{ij} - v_j) x_{ij}, \quad u \geq 0
\]
and hence for any fixed dual multipliers \( u \geq 0 \) and \( v \) we have

\[
\bar{z} - z^* \leq (\bar{z} + \sum_{i \in S} u_i a_i - \sum_{j \in T} v_j b_j) + \max_{x \in W} \sum_{i \in S} \sum_{j \in T} (v_j - c_{ij} - u_i \delta_{ij}) x_{ij} \equiv \epsilon(u, v, W). \tag{7}
\]

In particular, if \( \hat{u}_i, \hat{v}_j \) is an optimal solution of the dual to \((\tilde{A}P)\), then

\[
- \sum_{i \in S} \hat{u}_i a_i + \sum_{j \in T} \hat{v}_j b_j = \bar{z} \quad \text{and} \quad \epsilon(\hat{u}, \hat{v}, W) = \max_{x \in W} \sum_{i \in S} \sum_{j \in T} (\hat{v}_j - c_{ij} - \hat{u}_i \delta_{ij}) x_{ij}. \tag{8}
\]

### 3.1 A priori bounds

To get an a priori bound for \( \bar{z} - z^* \) by (8) we need to estimate the objective coefficients of the maximization problem in (8) without solving the aggregated problem. For the TP this can be done by choosing \( \hat{u}_i = \bar{u}_i, \hat{v}_j = \bar{v}_k, j \in T_k \), such that \( \hat{v} \in D(\bar{v}) \). To see this, note that for \( \delta_{ij} = 1 \), the term in the brackets in (8) is \( \bar{v}_k - c_{ij} - \bar{u}_i \) and by the restrictions of \((\text{DAP})\) this is less than or equal to \( \bar{c}_{ik} - c_{ij} \), \( j \in T_k \), which immediately gives

\[
\bar{z} - z^* \leq \max_{x \in W} \sum_{i \in S} \sum_{j \in T} (\bar{c}_{ik(j)} - c_{ij}),
\]

where \( k(j) \) is the cluster to which \( j \) belongs, \( k(j) \in T^A \). If the localization \( W \) is defined before an aggregated problem has been solved (we will refer to such \( W \) as a priori localization), then inequality above provides the a priori error bound for the TP. However, if \( \delta_{ij} \neq 1 \), we can’t estimate the coefficients in (8) under the same choice of the duals \( \hat{u}, \hat{v} \). By this reason it was conjectured in [2,11] that ”...a priori bounds cannot be obtained for the GTP as is possible for simple transportation problem”. To cope with this problem, we will use the nonuniqueness of the optimal dual multipliers for the \((\hat{A}P)\), as stated in Proposition 1.
Proposition 2. Let $d_{ij} = p_i t_j$, where $p_i > 0$, $t_j > 0$. Then there exists $\hat{v} \in D(\bar{u})$ such that for the pair $\bar{u}, \hat{v}$ we have

$$\hat{v}_j - c_{ij} - \bar{u}_i d_{ij} \leq (\bar{c}_{ik} d_{ij} - c_{ij} \bar{d}_{ik})/\bar{d}_{ik}, \quad j \in T_k.$$ 

Proof. From restrictions of (DAP) we have $u_i \geq (v_k - c_{ik})/d_{ik}$ and hence

$$\hat{v}_j - c_{ij} - \bar{u}_i d_{ij} \leq [(\bar{c}_{ik} d_{ij} - c_{ij} \bar{d}_{ik}) - (\bar{v}_k d_{ij} - \hat{v}_j \bar{d}_{ik})]/\bar{d}_{ik}.$$ 

It is not hard to verify that by the definition of $\bar{d}_{ik}$ and the correspondence between $v_k$ and $\hat{v}_j, j \in T_k$ we have

$$\sum_{j \in T_k} b_j (v_k d_{ij} - \hat{v}_j \bar{d}_{ik}) = 0 \quad \forall i.$$ 

Since $b_j > 0$, then the case $(\bar{v}_k d_{ij} - \hat{v}_j \bar{d}_{ik}) > 0$ ($< 0$) for all $j \in T_k$ is impossible.

Now choose $\hat{v}_j$ to fulfill conditions $v_k d_{ij} - \hat{v}_j \bar{d}_{ik} = 0$ for all $j \in T_k$. Resolving the latter system of equations with respect to $\hat{v}_j$ for $d_{ij} = p_i t_j$ and $\bar{d}_{ik} = (\sum_{j \in T_k} b_j d_{ij})/(\sum_{j \in T_k} b_j)$ we obtain

$$\hat{v}_j = \bar{v}_k t_j \frac{\sum_{j \in T_k} b_j}{\sum_{j \in T_k} b_j t_j}, \quad j \in T_k.$$ 

It is not hard to verify that for this $\hat{v}_j$ we have $\sum_{j \in T_k} \hat{v}_j b_j = v_k \sum_{j \in T_k} b_j$ Hence $\hat{v} \in D(\bar{u})$ and the claim follows.

Proposition 2 together with (8) yields immediately the following expression for the a priori bound:

**Corollary.** For $d_{ij} = p_i t_j$ we have

$$\overline{z} - z^* \leq \max_{x \in W} \sum_{i \in S} \sum_{j \in T} \delta_{ij}^a x_{ij} = \varepsilon^a(W),$$

where

$$\delta_{ij}^a = \left( \frac{\bar{c}_{ik(j)} - c_{ij}}{\bar{d}_{ik(j)} \bar{d}_{ij}} \right) d_{ij}$$

and $k(j)$ is the cluster to which $j$ belongs, $k(j) \in T^A$.

**Remark 1.** Under the assumption $d_{ij} = p_i t_j$ we have $\sum_{j \in T} t_j x_{ij} \leq a_i/p_i, i \in S$ in the "source" restrictions of the original problem. This means that the "gain" ("loss") along
the arc \((i, j)\) is the same for all sources \(i\) connected with a destination \(j\). We can treat this case as an intermediate one between the TP, where the gains are the same for all arcs, and a general form of the GTP, where the gains vary from arc to arc. Interpreting the GTP as the machine loading problem (see, e.g., [1]), producing 1 unit of product \(j\) on machine \(i\) consumes \(d_{ij}\) hours of the machine’s time. If \(p_i\) is interpreted as a production ”velocity” of the machine \(i\), and \(t_j\) is a ”length” of the product \(j\) measured in suitable units, then the assumption \(d_{ij} = p_i t_j\) holds. In financial networks, sources of investments may be concentrated in one country or region, while destinations may be located in various countries/regions. Then \(t_j\) can be treated as a ”cross-the-boarder” fee, which is the same for all sources.

A priori localizations can be defined, for example, by manipulating the constraints of the original problem. Let

\[
W_b = \{x_{ij} \mid 0 \leq x_{ij} \leq r_{ij} \forall i, j\}, \quad r_{ij} = \min\{b_j, a_i/d_{ij}\},
\]

\[
W_s = \{x_{ij} \mid \sum_{j \in T} d_{ij} x_{ij} \leq a_i, \quad i \in S, \quad x_{ij} \geq 0 \forall i, j\},
\]

\[
W_d = \{x_{ij} \mid \sum_{i \in S} x_{ij} = b_j, \quad j \in T, \quad x_{ij} \geq 0 \forall i, j\},
\]

where the lower indices \(b, s,\) and \(d\) show that the restrictions defining the respective localization are associated with bounds, sources and destinations. Obviously \(W_s, W_d\) are localizations as well as \(W_{bs} = W_b \cap W_s\) and \(W_{bd} = W_b \cap W_d\). For these four localizations the optimization problem to calculate \(\varepsilon^a(W)\) can be solved analytically. Due to the decomposable structure of the localizations, the computation of \(\varepsilon^a(W_s), \varepsilon^a(W_d)\) results in a number of independent single-constrained continuous knapsack problems, which yields

\[
\varepsilon^a(W_s) = \sum_i a_i \max_j [\delta_{ij}^a] / d_{ij},
\]

\[
\varepsilon^a(W_d) = \sum_j b_j \max_i \delta_{ij}^a,
\]
where $[\alpha]^- = \max\{0, \alpha\}$. Similarly, the calculation of $\varepsilon^a(W_{bs}), \varepsilon^a(W_{bd})$ is reduced to the solution of independent single-constrained knapsack problems with upper bounds for the variables. Respective solutions can also be obtained analytically (see, e.g., [6]).

By construction we have $\min\{\varepsilon^a(W_b), \varepsilon^a(W_s)\} \geq \varepsilon^a(W_{bs})$ and similarly for $W_d$. Then $\varepsilon^a_{Best}$ is defined as

$$\varepsilon^a_{Best} = \min\{\varepsilon^a(W_{bs}), \varepsilon^a(W_{bd})\}.$$ 

It is interesting to compare the obtained a priori bounds with the results known for the TP [13]. Let $d_{ij} = 1 \ \forall i, j$ and all " $\leq $ " in the linear restrictions of (OP) be substituted for " $=$ ". Then $\delta^a_{ij} = \tau_{ik(j)} - c_{ij}$ and to get the estimation stated in proposition 2 we should put $\hat{v}_j = \tau_k, \ j \in T_k$. The expressions for our four a priori bounds will continue to be valid for the classical transportation problem if we change everywhere $[\delta^a_{ij}]^+ \ $ for $\delta^a_{ij}$. Then $\varepsilon^a(W_s)$ and $\varepsilon^a(W_d)$ coincide with the a priori bounds obtained in [13] for the customers aggregation in the TP. Our bound $\varepsilon^a_{Best}$ is at least as good as $\min\{\varepsilon^a(W_s), \varepsilon^a(W_d)\}$.

### 3.2 Tightening the a priori bounds

From the definition of $\varepsilon^a(W)$ it follows that the tighter the localization $W$, the better (smaller) the bound $\varepsilon^a(W)$. That is, for example, defining $W$ by the original constraints, we should retain as many constraints as possible. Alternatively, localization $W$ should be simple enough to guarantee that the computational burden associated with the calculation of the bound is not too expensive. A compromise decision is to utilize "easy" restrictions directly, while dualizing "complicated" restrictions by the Lagrangian.

Suppose that the localization $W_{all}$ is defined by all constraints of (OP) and the destination constraints are treated as "easy", while the source constraints are considered as "complicated". Then by the Lagrangian duality we have:

$$\varepsilon^a(W_{all}) \equiv \max_{x \in W_{all}} \sum_{i \in S} \sum_{j \in T} \delta^a_{ij} x_{ij} = \min_{u \geq 0} \left\{ \sum_{i} u_i a_i + \max_{x \in W_d} \sum_{i \in S} \sum_{j \in T} (\delta^a_{ij} - u_i d_{ij}) x_{ij} \right\} \equiv \min_{u \geq 0} \varepsilon^a(W_d, u) \leq \varepsilon^a(W_d, 0) \equiv \varepsilon^a(W_d),$$
where $\varepsilon^a(W_d, u)$ denotes the expression in the figure brackets.

After the bound $\varepsilon^a(W_d, 0) = \varepsilon^a(W_d)$ has been calculated, it is natural to look for another value of $u \geq 0$ to diminish the current bound. If $s$ is the search direction, we get $u = \rho s$, where the stepsize $\rho \geq 0$ and the components of the search direction should be nonnegative.

Denote by $\hat{x}(W)$ the optimal solution of the problem $\max_{x \in W} \sum_{i \in S} \sum_{j \in T} \delta_{ij} x_{ij} \equiv \varepsilon^a(W)$. Suppose $\varepsilon^a(W_d)$ has been calculated and let $\hat{x}(W_d)$ be unique. Then by the marginal value theorem, $\varepsilon^a(W_d, u)$ is differentiable at $u = 0$, such that $\nabla_u \varepsilon^a(W_d, 0) = a_i - \sum_{j \in T} d_{ij} \hat{x}_{ij}$. Now take one step of the projected gradient technique for the problem $\min_{u \geq 0} \varepsilon^a(W_d, u)$ starting from $u = 0$. This gives

$$\hat{u}_i(\rho) = \rho \max\{0, \sum_{j \in T} d_{ij} \hat{x}_{ij} - a_i\} \quad i \in S, \quad \rho \geq 0.$$  

The stepsize $\hat{\rho}$ associated with the steepest descent can be defined by a one-dimensional search: $\hat{\rho} = \arg \min \{\varepsilon^a(W_d, u(\rho)) | \rho \geq 0\}$. Under the uniqueness of $\hat{x}$ the choice of $\hat{u}_i(\hat{\rho})$ guarantees a strict decrease of the a priori bound, that is either $\varepsilon^a(W_d, \hat{u}) < \varepsilon^a(W_d, 0) = \varepsilon^a(W_d)$ or $\varepsilon^a(W_d) = \varepsilon^a(W_{all})$. The other a priori bounds can be strengthened similarly.

We will refer to the a priori bound $\varepsilon^a(W)$, tightened by searching the duals in the direction of the projected anti-gradient as $\varepsilon^a(W)$.

### 3.3 A posteriori bounds

Suppose now that (AP) has been solved and the values $\pi$ and $\tau$ are known. The expression for $\epsilon(\hat{u}, \hat{v}, W)$ in (8) still provides the valid bound and we only need to construct $\hat{u}, \hat{v}$ optimal to $(\widetilde{AP})$. By Proposition 1 we may put $\hat{u} = \pi$ and $\hat{v}_j = \tau_k$, $j \in T_k$.

For the same localizations used in Section 3.1 we can calculate the a posteriori bounds $\varepsilon^p(\hat{u}, \hat{v}, W)$ analytically, since

$$\bar{z} - z^* \leq \max_{x \in W} \sum_{i \in S} \sum_{j \in T} \delta_{ij} x_{ij} \equiv \varepsilon^p(\hat{u}, \hat{v}, W),$$

13
where $\delta_{ij}^p \equiv \hat{v}_j - c_{ij} - \hat{u}_i d_{ij}$. Hence we only need to substitute $\delta_{ij}^p$ for $\delta_{ij}^a$ in the expressions for the a priori bounds obtained earlier. Note that the a posteriori bound is valid for the GTP in its general form, without any assumptions about the specifics of the arc multipliers.

For the choice of $\hat{u}, \hat{v}$ described above, the bounds $\varepsilon^p(\hat{u}, \hat{v}, W_s)$ and $\varepsilon^p(\hat{u}, \hat{v}, W_d)$ coincide with the a posteriori bounds obtained in [2]. The new bound $\varepsilon^p_{Best}$ obtained for the same choice of $\hat{u}, \hat{v}$ is at least as good as $\varepsilon^p(\hat{u}, \hat{v}, W_s)$, $\varepsilon^p(\hat{u}, \hat{v}, W_d)$, and a numerical study below demonstrates that in many cases it is significantly better.

After the a posteriori bound has been calculated for some fixed $\bar{u} \geq 0, \bar{v}$, it is natural to look for another duals to decrease the bound. Let the duals $\{\bar{u}, \bar{v}\}$ be changed in the direction $\{s, q\}$ with the stepsize $\rho$, such that $\bar{u} + \rho s \geq 0$. Then the best (i.e. bound minimizing) duals in this direction are defined from the problem:

$$\min_{\rho u + \rho s \geq 0} \varepsilon^p(\bar{u} + \rho s, \bar{v} + \rho q, W).$$

The dual-searching approaches to tightening the a posteriori bound are well known for the linear programming aggregation (see [5] and the references therein). Mendelssohn [9] proposed to look for the new duals in the direction of the dual aggregated solution, i.e. for our case $s = \bar{u}$, $q = \bar{v}$. Shetty and Taylor [12] used the right-hand side vector of the original problem as the search direction. In [6] one step of the projected gradient technique was used, similar to that was done in Section 3.2 for the a priori bounds. In all cases the stepsize $\rho$ can be obtained either by a one-dimensional search or by analyzing the non-differentiability points of the expression for $\varepsilon^p(\bar{u} + \rho s, \bar{v} + \rho q, W)$ as the function of the stepsize. We will denote these bounds by $\varepsilon^p_M(W)$, $\varepsilon^p_{ST}(W)$, and $\varepsilon^p_{\nabla}(W)$, respectively.

Thus for each of the four localizations $W_d$, $W_s$, $W_{bs}$, and $W_{bd}$ we can calculate two a priori bounds $\varepsilon^a(W)$, $\varepsilon^a_{\nabla}(W)$ and four a posteriori bounds $\varepsilon^p(W)$, $\varepsilon^p_M(W)$, $\varepsilon^p_{ST}(W)$, and $\varepsilon^p_{\nabla}(W)$. 

14
4 Numerical study

Within this section the results of a numerical study are described. To test the bounds derived in this paper, two classes of problems were considered: the general GTP (GTP) and the specific GTP with \(d_{ij} = p_i t_j\) (SGTP). Test problems were generated using the open source code GNETGEN (ftp://dimacs.rutgers.edu/pub/netflow/generators/network/netgen), which is a modification by F.Glover of the NETGEN generator [4]. For each class five problem sizes were considered, grouped in two sets: SET1 and SET2. The problems have the following dimensions: SET1 (30×50, 50×100, 100×150) and SET2 (50×1000, 50×2000), such that the number of destinations is greater (or significantly greater) than the number of sources. The destinations are clustered in 5 levels: 80, 60, 40, 20 and 10% of its total number. This clustering is refereed to as level 1 - level 5, respectively. Each cluster has \([|T|/|T^4|]\) destinations sorted by their index, except the last cluster containing all the rest destinations. Here \([\alpha]\) denotes the integer part of \(\alpha\). For each problem size ten problems were randomly generated, such that there are a total of \((5\text{ problem sizes})\times(5\text{ cluster levels})\times(10\text{ problems})=250\) observations for every class of problems.

The problems were obtained using random number seeds to generate the problem. The available supplies \(a_i\) were randomly generated with \(\sum a_i = 10^4\) for all instances in SET1 and \(\sum a_i = 1.8 \cdot 10^6\) in SET2. Customer demands \(b_j\) are also randomly generated. The cost coefficients \(c_{ij}\) and the arc multipliers \(d_{ij}\) were randomly generated from the intervals \([2.0, 20.0]\) and \([0.5, 3.0]\) respectively.

Four localizations were used to calculate the aggregation bounds: \(W_d, W_s, W_{bs},\) and \(W_{bd}\). For each localization we calculated two a priori bounds \(\varepsilon^a(W), \varepsilon^p_{\nabla}(W)\) for the SGTP and four a posteriori bounds \(\varepsilon^p(W), \varepsilon^p_{\nabla}(W), \varepsilon^p_{ST}(W)\) and \(\varepsilon^p_M(W)\) for both SGTP and GTP.

The dimension of the original problem allows the exact solution for all problems sizes. Correlations between the a priori (a posteriori) error bounds and the actual error \(\bar{z} - z^*\) are presented in Tables 1,4 for all localizations. Table 2 gives intercorrelations between the
best a priori bound $\varepsilon_{\text{best}}^a$ and the a posteriori bounds for the SGTP. Tables 3, 5 show the correlation by level of aggregation. Due to space limitations only averaged characteristics are provided. The $p$-values associated with respective correlations are also omitted, since for all localizations (except for $W_s$) they were less than 0.0001. For more detailed numerical results see [7].

Similar to [5,6], the numerical quality of the bounds was characterized by the upper and the lower bounds on $z^*$: $\bar{z} \geq z^* \geq LB_\varepsilon$, where $LB_\varepsilon = \bar{z} - \varepsilon(W)$. We used the indicators $(\bar{z} - z^*)/z^*$ and $(z^* - LB_\varepsilon)/z^*$. The average values of these indicators, together with standard deviations, are presented in Tables 6-9, sorted by localizations and levels of clustering.

The computational results show that the bounds calculated for the localization $W_{bd}$ have the strongest correlation with the actual error. This takes place for all classes and sizes of the problems, as well as for all types of the bounds and levels of clustering. We would like to stress that the a priori bound $\varepsilon_{\text{best}}^a(W_{bd}) = \varepsilon_{\nabla}^a(W_{bd})$ is highly and statistically significantly correlated with the actual error and with the a posteriori bounds (see Tables 2, 3).

The a posteriori error bounds are more correlated with the actual error. Mendelssohn’s bound, $\varepsilon_{M}^p(W_{bd})$, has the strongest correlation. It is also superior in correlation to all other bounds for any fixed localization.

(retirei esta parte do texto)

It is interesting to note that this approach is less sensitive to the type of localization used.

5 Conclusions
In this paper we studied some aspects of the aggregation applied to the GTP. The main focus was on the bounds for the aggregation error. Only customers (destinations) were aggregated and the clustering level was the only element of the aggregation strategy that was varied in the experimental study.

A priori error bounds were derived for the special case of the GTP and the new a posteriori were proposed for the GTP in its general form. The experimental study demonstrates that there exists a localization \((W_d \cap W_b)\) which gives a high and statistically significant correlation between any of the bounds and the actual error. It is important to note that the a priori bound for this localization is also highly correlated with the actual error. That is, if the customers in the GTP are aggregated using the strategy outlined in this study, and the prescribed localization is chosen, the a priori bound is an appropriate guide to use in selecting the clustering level.

The a posteriori error bounds are more correlated with the actual error than the a priori bounds. Moreover, the dual searching approaches result in a higher correlation with the actual error. In particular, Mendelssohn’s approach [9] demonstrates the highest correlation.

Concerning the quantitative characteristics, the a posteriori bounds provide much tighter approximation of the actual error than the a priori bounds. Again, the tightest approximation was achieved for the localization mentioned above. Comparing various techniques to calculate the a posteriori bounds, the tightest bounds were obtained by the Mendelssohn’s approach.

In summary, these results indicate that for our set of randomly generated problems, the aggregation with the tightest a priori bound provides the best approximation of the optimal objective value. Under our conditions the aggregated model with the tightest a priori error bound tends to have the tightest a posteriori bound. Moreover, the aggregated model with the tightest a posteriori error bounds provide the closest approximation of the objective function value.
Based on our experimental study we may conclude that the choice of the localization plays an important role in the aggregation bounds computations. The proper localization not only improve the value of the error bound, but also increase the correlation between the aggregation bound and the actual error. In this paper we have studied only the localizations obtained by simple constraint manipulations. An interesting direction for the future research is the use of more complex localizations including ones obtained after the aggregated problem has been solved.

6 Acknowledgements

The authors would like to thank Russian foundation RFBR (02-01-00638), Brazilian foundation FAPESP (2000/09971-9), and Mexican foundation CONACyT (38870-A) for partial financial support of this work.

References


