OPTIMAL POLITICAL DISTRICTING

Mehran Hojati†
Department of Finance and Management Science, College of Commerce, University of Saskatchewan, Saskatoon, Sask., Canada S7N 5A7

(Received July 1995; in revised form April 1996)

Scope and Purpose—Given an area and its population units, we wish to divide the area into m districts such that each district has almost the same population of eligible voters (within a given tolerance), is contiguous, compact, and has a minimum number of split population units. This fair representation problem has been a great concern of the public for decades. The districting problem is also used in the design of sales territories. Redistricting occurs often because of population shifts or for political reasons. The purpose of this paper was to find a practical and automated operations research computer solution method for this problem.

Abstract—For the political districting problem, I propose the following solution methodology: (a) use Lagrangian relaxation to determine the centres of the districts, then, (b) use the transportation technique to assign population units to centres, and finally, (c) resolve the splitting problem by solving a sequence of capacitated transportation problems. This method is applied to the problem of determining the provincial districts for the City of Saskatoon, Canada, and the results are compared with the actual districting done in 1993 by the Electoral Boundaries Commission. Copyright © 1996 Elsevier Science Ltd

1. INTRODUCTION AND LITERATURE REVIEW

Given an area, e.g. a province or a city, and its population units, e.g. subdivisions or census tracts, we wish to divide the area into m districts such that each district has almost the same population of eligible voters (within a given tolerance, e.g. 5%). In addition, it is desirable that a district consists of contiguous population units and to be compact (close to square or circle). Frequently, to obtain population parity within the tolerance, population units have to be split between two or more districts. However, to reduce the time-consuming task of actually splitting population units and the ensuing confusion of voters, this splitting is to be minimized.

Political re-districting occurs frequently because of population shifts and because of political desires to change the total number of representatives. This problem is related to a problem in marketing called sales territory alignment. A computer method to solve political districting is desirable because (a) constructing districts with population parity within the tolerance and (b) keeping the number of split population units to a minimum are conflicting objectives which cannot be easily resolved manually.

The main literature on this problem started with Hess et al. [1] who modelled it as a location-allocation problem, but due to the computational difficulty solved it by the following heuristic: (a) start with an arbitrary set of centres of districts (a set of population units), (b) use the transportation technique to allocate population units to the centres, where the demand for each population unit is its population of eligible voters, the supply of each centre is the population quota for each district, and the distance between a centre and a population unit is the square of the Euclidean distance between their centres of gravity, (c) assign any split population units to the centre supplying the largest fraction of its demand, and (d) calculate the centre of gravity of each derived district and resolve the transportation problem with these centres until convergence of centres occurs.

Fleischmann and Paraschis [2] used a similar method, except for the following heuristic used to resolve the split population units in the solution of the transportation problem: (a) exclude the full and zero assignments, and identify the subtree F consisting of the split population units and their suppliers, (b) for each arc (i,j) in F, determine if supplier i, after adjustment to any other demands of k with (i,k) in F, needs to supply to j to remain within the tolerance, or if demand point j, after adjustment to any other suppliers k with (k,j) in F, needs to obtain supplies from i. If the answer is yes to either of these two questions, then a (partial) assignment on (i,j) is performed, and (c) if the answer to both questions in (b) for every arc (i,j) in F is negative, then a full assignment is made on (i,j) if the tolerance is not exceeded for supplier i.

† Mehran Hojati is associate professor of Management Science in the Department of Finance and Management Science of the College of Commerce at the University of Saskatchewan. He received a Ph.D. in Management Science from the University of British Columbia. His research interests and publications are in the area of applications of Operations Research in networks and finance.
Thoreson and Liittschwager [3] used the following heuristic: (a) start with a reference population unit A, (b) find another population unit B furthest away from A, (c) build a district around B, including the closest population units to B until the district quota is just to be exceeded in which case include any contiguous population unit which will result in a population of eligible voters as close to the district quota as possible, (d) repeat with the furthest unassigned population unit B2 from A and build a district around B2, and (e) continue until all population units are assigned.

Garfinkel and Nemhauser [4] used the following exact method: (a) construct all sets of potential districts which are contiguous, compact, and have a total electorate within the tolerance, and (b) use a set partitioning integer linear programming model to minimize the maximum deviation of any chosen district from the district quota.

Among other heuristic methods which iteratively interchange population units between two districts are Bourjolly et al. [5] and Robertson [6]. Finally, Zoltners and Sinha [7] proposed a precedence network for population units to prevent discontiguity in districts. However, they assume that the centres of the districts are already fixed.

My approach is similar to [1] and [2], with the following differences: (a) I propose the use of Lagrangian relaxation to determine the centres of the districts, and (b) to resolve the splitting problem I propose the use of a sequence of capacitated transportation problems which at each stage tries to force one more shipment in F to zero. These methods are tested by using data for the City of Saskatoon, and the results are compared with the actual districting done in 1993 for the provincial legislature.

2. THE WAREHOUSE LOCATION MODEL

A warehouse location model for districting can be formulated as follows:

Let

\[ N = \text{number of population units} \]
\[ M = \text{number of districts required} \]
\[ Q_j = \text{number of eligible voters in population unit } j \] (demand of \( j \))
\[ Q = \text{district quotient} \] (= sum of \( Q_j M \))
\[ X_{ij} = \text{proportion of demand of population unit } j \text{ supplied by centre } i \] (in population unit \( i \)). \( (N^2 \text{ variables}) \)
\[ c_{ij}' = \text{square Euclidean distance from centre of population unit } i \text{ to the centre of population unit } j. \] This is similar to [1]. Other distance measures were also examined
\[ c_i = c_{ij}' \times Q_j \]
\[ Y_i = \text{indicator variable for population unit } i: \]
\[ = 1 \text{ if population unit } i \text{ is to be chosen as the centre of a district} \]
\[ = 0 \text{ otherwise.} \]

Minimize

\[ \sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij} X_{ij} \] (1)

Subject to:
Demand:

\[ \sum_{j=1}^{N} X_{ij} = 1 \text{ For each } j \] (2)

Supply:

\[ \sum_{j=1}^{N} Q_j X_{ij} = QY_i \text{ For each } i \] (3)

\[ \sum_{i=1}^{N} Y_i = M \] (4)

\[ X_{ij} \leq Y_i \text{ For each } j, \text{ for each } i \] (5)

\[ 0 \leq X_{ij} \leq 1 \text{ For each pair } i, j \] (6)

\[ Y_i = 0 \text{ or } 1 \text{ For each } i. \] (7)

Note that the population tolerance for each district is not used at this stage.
3. THE LAGRANGIAN RELAXATION METHOD

The procedure used here is a simplified version of that given in Beasley [8] who relaxes both the supply and the demand constraints. First, normalize (3) by dividing it by $Q$. Let $s_j$ and $t_i$ be the multipliers associated with (2) and the normalized (3). The Lagrangian lower bound programme (LLBP) is:

Minimize

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij} X_{ij} + \sum_{j=1}^{N} s_j \left( 1 - \sum_{i=1}^{N} X_{ij} \right) + \sum_{i=1}^{N} t_i \left[ - Y_i + \sum_{j=1}^{N} \left( \frac{Q}{Q} \right) X_{ij} \right]$$

Subject to: (4)–(7).

Defining

$$C_{ij}=c_{ij} - s_j + t_i \left( \frac{Q}{Q} \right)$$

LLBP becomes

Minimize

$$\sum_{i=1}^{N} \sum_{j=1}^{N} C_{ij} X_{ij} - \sum_{i=1}^{N} t_i Y_i + \sum_{j=1}^{N} s_j$$

Subject to: (4)–(7).

If $Y_i=0$, then the contribution of $X_{ij}$ and $Y_i$ to (8) is 0 (because $X_{ij}$ will be set to 0 too), whereas if $Y_i=1$, this contribution is given by

$$a_i=-t_i + \sum_{j=1}^{N} \min(0,C_{ij})$$

because it is optimal to set $X_{ij}=1$ if $C_{ij}\leq0$, and $X_{ij}=0$ if $C_{ij}>0$. Therefore, LLBP reduces to

Minimize

$$\sum_{i=1}^{N} a_i Y_i + \sum_{j=1}^{N} s_j$$

Subject to:

$$\sum_{i=1}^{N} Y_i = M$$

$$Y_i=0 \text{ or } 1 \quad \text{for each } i.$$ 

This zero/one programme can be solved by inspection as follows: Sort the $a_i$ in ascending order and set $Y_i=1$ for the first $M$ population units. Set $X_{ij}=1$ if $Y_i=1$ and $C_{ij}\leq0$, else set $X_{ij}=0$. Then, a valid lower bound on the optimal solution is given by

$$Z_{LB}=\sum_{i=1}^{N} a_i Y_i + \sum_{j=1}^{N} s_j$$

The near-optimal values for $t_i$, $i=1, \ldots, N$, and $s_j$, $j=1, \ldots, N$, are determined iteratively by the following subgradient optimization method:

Let

- $Z_{\text{max}}$ = maximum lower bound found
- $Z_{\text{UB}}$ = an upper bound for (1) (e.g. choose an arbitrary set of $M$ centres, solve by the transportation technique, then set $Z_{\text{UB}}$ = objective value)
- $n$ = number of subgradient iterations since $Z_{\text{max}}$ last increased
- $f$ = step length parameter.

Step 1. Initialize $Z_{\text{max}}=-\infty$; $n=0$; $f=2$; $t_i=0$, $i=1, \ldots, N$; $s_j=c_{ij}$, $j=1, \ldots, N$.

Step 2. Solve LLBP with the current set of multipliers’ and let the solution be $Z_{\text{LB}}$, $(Y_i)$, and $(X_{ij})$. 

$Z_{LB} > Z_{max}$, set $n=0$, and $Z_{max} = Z_{LB}$. Else, set $n=n+1$.

**Step 3.** If $n=30$, then 30 iterations of the subgradient procedure have been performed without an increase in $Z_{max}$. Hence, halve the step length parameter $f$ by setting $f=f/2$ and set $n=0$. Use the solution $Z_{LB}$, $(Y_i)$ and $(X_0)$ corresponding to the best lower bound $Z_{max}$ obtained so far.

**Step 4.** Calculate the subgradients $G_j$ and $H_i$ using

$$G_j = 1 - \sum_{j=1}^{N} X_{ij}$$  
$$j = 1, \ldots, N$$

and

$$H_i = -Y_i + \sum_{j=1}^{N} \left(\frac{Q_j}{Q} \right) X_{ij}$$  
$$i = 1, \ldots, N$$

**Step 5.** If $\sum_{j=1}^{N} G_j + \sum_{i=1}^{N} H_i = 0$ or $f < 0.005$, go to Step 7.

**Step 6.** Define the step size $T$ by

$$T = \frac{(f)(Z_{LB} - Z_{LB})}{\sum_{j=1}^{N} G_j + \sum_{i=1}^{N} H_i}$$

Update the Lagrange multipliers using

$$s_j = s_j + TG_j$$  
$$j = 1, \ldots, N$$

$$t_i = t_i + TH_i$$  
$$i = 1, \ldots, N$$

and go to Step 2.

**Step 7.** Solve the solution $(Y_i)$ associated with the current maximum lower bound $Z_{max}$ using a transportation technique. Terminate.

4. **RESOLVING THE SPLITS**

After determining the centres of districts by Lagrangian relaxation and solving the district composition problem by the transportation technique, the solution obtained will have at most $m-1$ population units which are split between two or more districts.

**Lemma 1.** The solution of the transportation technique will have at most $m-1$ split population units.

**Proof.** Because there are at most $m+N-1$ basic cells in this solution, at most $m-1$ demand points (population units) can have two or more cells in their associated columns, leaving at least $N-m+1$.

![Fig. 1. The two-partition reduction to (SRP) used in the proof of Lemma 2.](image-url)
Table 1. Population units (PU), number of eligible voters (1993), and the coordinates of centres of population units

<table>
<thead>
<tr>
<th>PU</th>
<th>Voters</th>
<th>X coordinate</th>
<th>Y coordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4880</td>
<td>11.5</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>2350</td>
<td>10.5</td>
<td>2.5</td>
</tr>
<tr>
<td>3</td>
<td>3455</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4020</td>
<td>8</td>
<td>1.5</td>
</tr>
<tr>
<td>5</td>
<td>4550</td>
<td>7.5</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>3200</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>2900</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>3315</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>2665</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>1205</td>
<td>7</td>
<td>5.5</td>
</tr>
<tr>
<td>11</td>
<td>2330</td>
<td>8</td>
<td>5.5</td>
</tr>
<tr>
<td>12</td>
<td>3440</td>
<td>7.5</td>
<td>4.5</td>
</tr>
<tr>
<td>13</td>
<td>4585</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>2635</td>
<td>11.5</td>
<td>3.5</td>
</tr>
<tr>
<td>15</td>
<td>2365</td>
<td>10.5</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>2235</td>
<td>11.5</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>4050</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>18</td>
<td>3205</td>
<td>14.5</td>
<td>5</td>
</tr>
<tr>
<td>19</td>
<td>2605</td>
<td>13</td>
<td>3.5</td>
</tr>
<tr>
<td>20</td>
<td>5155</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>3670</td>
<td>13</td>
<td>2.5</td>
</tr>
<tr>
<td>22</td>
<td>3070</td>
<td>12</td>
<td>7.5</td>
</tr>
<tr>
<td>23</td>
<td>3835</td>
<td>13.5</td>
<td>7</td>
</tr>
<tr>
<td>24</td>
<td>1725</td>
<td>15</td>
<td>7.5</td>
</tr>
<tr>
<td>25</td>
<td>505</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>26</td>
<td>3475</td>
<td>9.5</td>
<td>5.5</td>
</tr>
<tr>
<td>27</td>
<td>3950</td>
<td>8.5</td>
<td>7</td>
</tr>
<tr>
<td>28</td>
<td>1255</td>
<td>7</td>
<td>6.5</td>
</tr>
<tr>
<td>29</td>
<td>2665</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>30</td>
<td>2615</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>31</td>
<td>2630</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>32</td>
<td>3825</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>33</td>
<td>2510</td>
<td>3</td>
<td>8.5</td>
</tr>
<tr>
<td>34</td>
<td>2150</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>35</td>
<td>3110</td>
<td>6</td>
<td>7.5</td>
</tr>
<tr>
<td>36</td>
<td>3480</td>
<td>7</td>
<td>8.5</td>
</tr>
<tr>
<td>37</td>
<td>3120</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>38</td>
<td>3920</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>39</td>
<td>3540</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>40</td>
<td>3310</td>
<td>10.5</td>
<td>11.5</td>
</tr>
<tr>
<td>41</td>
<td>3935</td>
<td>10.5</td>
<td>12.5</td>
</tr>
</tbody>
</table>

Fig. 3. The optimal solution to the transportation problem; suppliers are bold-faced; each supplier supplies its own population unit (not shown).
columns having exactly one basic cell. This is because \(2(m - 1) + (N - m + 1) = N + m - 1\). If a population unit is split among more than two districts, then the number of split population units will be less than \(m - 1\).

The split-resolution problem (SRP) can be formulated as follows:

Let

\[
(x_j) = \text{solution of the transportation technique (in number of eligible voters not proportion)}
\]

\(J' = \text{set of split population units}\)

\(I' = \text{set of centres adjacent to any split area}\)

\(F = \text{the graph with vertex set } I' \cup J' \text{ and the above adjacency [a forest since } (x_j) \text{ is a basic solution]}\)

\(U_v = \text{set of vertices adjacent to vertex } v \text{ of } F\)

\(A_{\text{max}} = Q(1 + \text{tolerance}), \text{ tolerance, e.g., } =0.05\)

\(A_{\text{min}} = Q(1 - \text{tolerance})\)

\(a_i = \text{size of district } i \text{ without split population units.}\)

Min. number of arcs \((i,j)\) in \(F\) with \(x_{ij} > 0\)

Subject to:

\[
\sum_{i \in I'} x_{ij} = Q_j \quad j \text{ in } J'
\]

\[
A_{\text{min}} \leq a_i + \sum_{j \in U_v} x_{ij} \leq A_{\text{max}} \quad i \text{ in } I'
\]

\(X_{ij} \geq 0 \quad i \text{ in } I', j \text{ in } J'.\)

**Lemma 2.** (SRP) is NP-hard.

**Proof.** We will reduce the NP-complete two-partition problem to (SRP). Given \(n\) real numbers \(B_1, B_2, \ldots, B_n\), the two-partition problem asks if there is a set \(S \subset \{1, \ldots, n\}\) such that

\[
\sum_{j \in S} B_j = \sum_{j \notin S} B_j.
\]

This problem is equivalent to a (SRP) with total number of suppliers in \(F = 2n\), supplier \(s\) having \(A_{\text{max}} - a_i = B_i\) and \(A_{\text{min}} - a_i = B_i\), \(i = 1, \ldots, n\); supplier \(s'\) having \(A_{\text{max}} - a_i = B_i\) and \(A_{\text{min}} - a_i = 0\), \(i = 1, \ldots, n\); demand point \(d_j\) having \(Q_j = B_j\), \(j = 1, \ldots, n\), and demand point \(d\) having demand

\[
Q_d = \frac{\sum_{i=1}^{n} B_i}{2}
\]

and the relationship as depicted in Fig. 1.
(SRP) can be formulated as a zero/one integer linear programme. Fleischmann and Paraschis [2] used the following heuristic to solve a problem which is equivalent to (SRP):

(a) For any arc \((i,j)\) in \(F\), an obligatory partial assignment \(x_{ij} = \delta_j\) takes place if

\[
\delta_j = \max(\delta_i, \delta_j) > 0
\]

where

\[
\delta_i = Q_j - \sum_{k \in (i,j)} (A_{mk} - a_k)
\]

\[
\delta_k = A_{mk} - a_i - \sum_{k \in (i,j)} Q_k.
\]

Then, \(a_i\) and \(Q_j\) are updated by \(a_i = a_i + x_{ij}\) and \(Q_j = Q_j - x_{ij}\) and another arc \((k,l)\) is examined.

(b) If for every arc \((i,j)\) in \(F\), \(\delta_j \leq 0\), then an arbitrary full assignment can take place for any arc \((i,j)\) if \(a_i + Q_j \leq A_{mk}\). After this, repeat steps (a) (some \(\delta_j\) may become positive) and (b) until all arcs are assigned.

The application of Fleischmann and Paraschis heuristic may not reduce the number of arcs in \(F\) with

<table>
<thead>
<tr>
<th>COST</th>
<th>d10</th>
<th>d7</th>
<th>d36</th>
<th>d39</th>
<th>d28</th>
<th>d13</th>
<th>d14</th>
<th>d15</th>
</tr>
</thead>
<tbody>
<tr>
<td>s33</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s33e</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s9</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s9e</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s8</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s8e</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s37</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s37e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s41</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s41e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s12</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s12e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s5</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s5e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s16</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
</tr>
<tr>
<td>s16e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
</tr>
<tr>
<td>s18</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
</tr>
<tr>
<td>s18e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
</tr>
<tr>
<td>s24</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s24e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s21</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
<tr>
<td>s21e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
</tr>
</tbody>
</table>

DEMAND

<table>
<thead>
<tr>
<th>COST</th>
<th>d10</th>
<th>d7</th>
<th>d36</th>
<th>d39</th>
<th>d28</th>
<th>d13</th>
<th>d14</th>
<th>d15</th>
</tr>
</thead>
</table>

DEMAND 2665.0 3090.0 3110.0 3920.0 3950.0 3440.0 4585.0 2635.0 DEMAND

Fig. 5. Continued overleaf.
### Optimal political districting

<table>
<thead>
<tr>
<th>COST</th>
<th>d19</th>
<th>d22</th>
<th>DD</th>
<th>SUPPLY</th>
</tr>
</thead>
<tbody>
<tr>
<td>s33</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
</tr>
<tr>
<td>s33e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>915.00</td>
</tr>
<tr>
<td>s9</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>3449.0</td>
</tr>
<tr>
<td>s9e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>1190.0</td>
</tr>
<tr>
<td>s8</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>2211.0</td>
</tr>
<tr>
<td>s8e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>1190.0</td>
</tr>
<tr>
<td>s37</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>4712.0</td>
</tr>
<tr>
<td>s37e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>1190.0</td>
</tr>
<tr>
<td>s41</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>503.00</td>
</tr>
<tr>
<td>s41e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>1190.0</td>
</tr>
<tr>
<td>s12</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>4319.0</td>
</tr>
<tr>
<td>s12e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>1190.0</td>
</tr>
<tr>
<td>s5</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>3796.0</td>
</tr>
<tr>
<td>s5e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>1190.0</td>
</tr>
<tr>
<td>s16</td>
<td>9.0000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>3156.0</td>
</tr>
<tr>
<td>s16e</td>
<td>9.0000</td>
<td>9.0000</td>
<td>.00000</td>
<td>1190.0</td>
</tr>
<tr>
<td>s18</td>
<td>.00000</td>
<td>.00000</td>
<td>9.0000</td>
<td>2416.0</td>
</tr>
<tr>
<td>s18e</td>
<td>.00000</td>
<td>.00000</td>
<td>.00000</td>
<td>1190.0</td>
</tr>
<tr>
<td>s24</td>
<td>.00000</td>
<td>9.0000</td>
<td>9.0000</td>
<td>2161.0</td>
</tr>
<tr>
<td>s24e</td>
<td>.00000</td>
<td>9.0000</td>
<td>.00000</td>
<td>1190.0</td>
</tr>
<tr>
<td>s21</td>
<td>9.0000</td>
<td>.00000</td>
<td>9.0000</td>
<td>1277.0</td>
</tr>
<tr>
<td>s21e</td>
<td>9.0000</td>
<td>.00000</td>
<td>.00000</td>
<td>1190.0</td>
</tr>
</tbody>
</table>

DEMAND 3205.0 3670.0 6545.0

DEMAND

**Fig. 5.** The capacitated transportation model for F (arc capacities not shown).

Let \( x_{ij} > 0 \), as will be seen in Section 6. Therefore, I propose the following heuristic for (SRP):

1. Identify all arcs \((i,j)\) in \( F \) with \( \delta_{ij} \leq 0 \), and call this set \( \mathcal{N} \). These arcs will always have a positive flow in them.
2. Give all arcs in \( F \) an infinite capacity.
3. Choose an arc \((i,j)\) in \( F \) not in \( \mathcal{N} \), set its capacity to zero, and solve a capacitated transportation problem as described below.
4. If the problem obtained is feasible, go to next step; else reset the capacity of \((i,j)\) back to infinity and continue.
5. Choose another arc \((k,l)\) in \( F \) not in \( \mathcal{N} \), set its capacity to zero, and solve the capacitated transportation problem. Repeat as in (d) until all arcs in \( F \) not in \( \mathcal{N} \) are examined.

### 5. The Capacitated Transportation Problem

The following formulation is similar to that given by Marlin [9], who used it to solve a (servicemen) districting problem with fixed district centers. Note that Marlin did not solve the split area problem. To represent the split resolution problem (which has a minimum and a maximum for each supplier) as a
transportation problem, we divide the shipments of a supplier \( i \) into two types: regular shipments \( x_{ij} \) which is part of the minimum supply \( A_{\text{min}} - a_i \), and the extra shipments \( x'_{ij} \) (if needed, up to a total of \( A_{\text{max}} - A_{\text{min}} \)). Therefore, for each supplier \( s_i \), we introduce a duplicate supplier \( s'_i \). Also, in addition to all demand points \( J' \), we introduce a dummy demand point \( DD \) to receive any extra shipments \( x'_{ij} \) which is not shipped to any demand point in \( J' \). Demand of \( DD \) is equal to (number of suppliers in \( F \))(\( A_{\text{max}} - A_{\text{min}} \))/2.

Let

\[
\begin{align*}
    x_{ij} &= \text{shipment from } s_i \text{ to } d_j \text{ to satisfy (part of) } A_{\text{min}} - a_i \\
    x'_{ij} &= \text{shipment from } s'_i \text{ to } d_j \\
    x'_{ij,DD} &= \text{shipment from } s'_i \text{ to } DD \\
    c_{ij} &= 9 \text{ if } (i,j) \text{ not in } F \text{ (any positive cost will work)} \\
    &= 0 \text{ if } (i,j) \text{ in } F \\
    c'_{ij} &= 9 \text{ if } (i,j) \text{ not in } F \\
    &= 0 \text{ if } (i,j) \text{ in } F \text{ or } j = DD
\end{align*}
\]

Minimize

\[
\sum_{i \in F, j \in F} c_{ij} x_{ij} + \sum_{i \in F, j \in F} c'_{ij} x'_{ij} + \sum_{i \in F} c'_{i,DD} x'_{i,DD}
\]
Subject to:

\[
\sum_{i \in I'} x_{ij} + \sum_{j \in J'} x_{ij} = Q_j \quad \text{For each } j \text{ in } J'
\]

\[
\sum_{i \in I'} x_{ij} = \frac{1}{2} |J'| |I'| (A_{\text{max}} - A_{\text{min}})
\]

\[
\sum_{j \in J'} x_{ij} = A_{\text{min}} - a_i \quad \text{For each } i \in I'
\]

\[
\sum_{j \in J'} x_{ij} + x_{ij} = A_{\text{max}} - A_{\text{min}} \quad \text{For each } i \in I'
\]

Each cell's shipment is capacitated.

6. AN APPLICATION

The City of Saskatoon is to be partitioned into 11 provincial constituencies (districts). The population units were chosen to be census tracts, see Statistics Canada [10]. A census tract is generally a local community of population between 1000 to 5000 residents. The 42 census tracts were numbered population unit 1 to 42 as shown in Fig. 2.

The number of eligible voters in each census tract was obtained by summing all age groups 20 years or older residing in the tract from [10], and the city was grided horizontally and vertically to determine the coordinates of the centres of each census tract, see Table 1.

Then, the square Euclidean distance between all centres were computed, and all data were input to a computer programme which performed the Lagrangian relaxation method. The solution i.e. centres of districts, after 35 min on a 486 personal computer, in terms of population unit numbers is: 5, 8, 9, 12, 16, 18, 21, 24, 33, 37, and 41.

Then, a transportation software was used with the above centres as suppliers each having a capacity of \( Q = 130 \ 865/11 = 11 \ 896.8 \), demands as given in Table 1, and square Euclidean distances. The optimal solution is displayed in Fig. 3.

Next, the split resolving method was used. The \( F \) subtree corresponding to split population units in the above solution and their split shipments are displayed in Fig. 4. The first capacitated transportation model to solve the split problem is displayed in Fig. 5. Note the tolerance was 5% or 595 voters.

Applying Fleischmann's \( \delta_q \) results in Step (a) identified the following arcs (the set NA):
Fig. 9. Districts derived for Saskatoon in this paper.
Finally, the capacitated transportation model was repeatedly solved as follows (feasible has the objective value=0, infeasible has objective value>0):

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Capacity</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(s33,d10)</td>
<td>feasible</td>
</tr>
<tr>
<td>2</td>
<td>(s8,d7)</td>
<td>infeasible (reset capacity=∞)</td>
</tr>
<tr>
<td>3</td>
<td>(s36,d8)</td>
<td>feasible</td>
</tr>
<tr>
<td>4</td>
<td>(s37,d36)</td>
<td>infeasible (reset capacity=∞)</td>
</tr>
<tr>
<td>5</td>
<td>(s37,d28)</td>
<td>feasible</td>
</tr>
<tr>
<td>6</td>
<td>(s5,d13)</td>
<td>infeasible (reset capacity=∞)</td>
</tr>
<tr>
<td>7</td>
<td>(s18,d15)</td>
<td>feasible</td>
</tr>
<tr>
<td>8</td>
<td>(s18,d19)</td>
<td>infeasible (reset capacity=∞)</td>
</tr>
</tbody>
</table>

The $F$ subtree has now changed to that displayed in Fig. 6. The number of split population units has been decreased by four, which happens to be the optimal solution to (SRP) in this case. In contrast, the Fleischmann and Paraschis heuristic results in a solution (see Fig. 7) with no elimination of arcs of $F$.

The final solution including the full assignments and those assignments given by my heuristic in Fig. 6 is displayed in Fig. 8. This solution is displayed in terms of city area in Fig. 9. The actual constituencies determined by the Electoral Boundaries Commission [11] are displayed in Fig. 10.

The following statistics compare the two solutions:

- **My Solution**
  - Maximum deviation from quotient: 5%
  - Compactness measure: 15
  - No. of discontiguous districts: 1
  - No. of split areas: 6

- **Actual Solution**
  - Maximum deviation from quotient: 1%
  - Compactness measure: 20
  - No. of discontiguous districts: 20
  - No. of split areas: 17

where compactness is measured as \[ \sum_{i=1}^{n} |L_i - W_i|, \] where $L_i$ is the maximum length of district $i$, and $W_i$ is the maximum width of district $i$.

7. CONCLUSION

For the political districting problem I have proposed the following three-step solution methodology: (a) use Lagrangian relaxation to determine the centres of the districts, then, (b) use the transportation technique to assign population units to centres, and finally, (c) resolve the splitting problem by solving a sequence of capacitated transportation problems. The contributions of this paper are, (i) the use of Lagrangian relaxation to determine the centres of districts in the districting problem is new, and (ii) the use of a sequence of capacitated transportation problems to resolve the split area problem is also new.

As can be seen in Section 6, this three-step method is efficient and practical. It resulted in more compact districts and considerably less split areas (6) than the actual districts implemented (17). Compact areas are more convenient for the voters and the elected representatives in terms of travelling distances. A smaller number of split areas would mean reduced administrative costs in terms of record keeping and organisation of the district. Furthermore, our method needs a minimum number of manual intervention and is least affected by local politics.

REFERENCES