How to Avoid Gerrymandering: A New Algorithmic Solution

Gregory B. Lush
Department of Electrical and Computer Engineering
University of Texas at El Paso
El Paso, TX 79968, USA
Email: lush@utep.edu

Esteban Gamez and Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
Email: vladik@utep.edu

Abstract—**Subdividing an area into voting districts is often a very controversial issue. If we divide purely geographically, then minority groups may not be properly represented. If we start changing the borders of the districts to accommodate different population groups, we may end up with very artificial borders – borders which are often to set up in such a way as to give an unfair advantage to incumbents. In this paper, we describe redistricting as a precise optimization problem, and we propose a new algorithm for solving this problem.**

I. Formulation of the Practical Problem

The notion of electoral districts. In the USA and in many other countries, voting is done by electoral districts:

- in elections to the US House of Representative, every federal voting district elects one representative;
- in elections to the state legislature, each state district elects one representative;
- in elections to the city council, every district elects one representative, etc.

To ensure equal representation, districts are drawn in such a way that all voting districts of the same level (federal, state, city, etc.) contain the exact same number of voters.

Need for redistricting. In time, demography changes. Some districts lose voters, some gain them. It is therefore necessary to change the districts in such a way that the new district contain equal number of people. In many places in the US, redistricting is performed every ten years, after a census provides an accurate count of the population in different areas.

How should redistricting be done? how is it done?

First consideration: the need to represent different geographic areas leads to compact districts. The whole point of electoral districts – as opposed to, e.g., nation-wide vote for political parties as in some countries – is that for each geographic region, there is a representative who is elected by people from this region and who, therefore, protects the interest of people from this region.

The closer people live to each other, the closer their regions. From this viewpoint, it makes sense to divide the country (state, city, etc.) into "compact" districts, so that voters living within the same district live as close to each other as possible.

Additional considerations affect the districts' shape. In addition to representing different geographic regions, we also need to represent different type of the population.

For example, rural areas have different problems than city areas, and it is desirable that their voice be heard. Farmers living near a city may have much more in common with farmers living somewhat far way than with the nearby city dwellers. So, instead of placing the nearby farmers into the same electoral district as the nearby towns, it may be desirable to combine them into one or more separate rural districts.

Similarly, ethnic and racial minorities have their own concerns, and it is desirable that these concerns be heard as well. If we simply divide people into equal-size geographic zones, then the minority population does not have much of an influence of elections in each of the resulting districts and may, thus, be under-represented in the legislature. The situation is somewhat similar to the case of rural population: minorities living somewhat far away may have more concerns in common that with the majority population living nearby. In view of this, it is desirable to form special minority districts which may not be perfectly geometrically compact, but which will enable the voice of the minority to be heard in the legislature.

There are many other reasons why we may want to deviate from the simple geometric shapes (corresponding to pure geographic consideration). For example, people in the border areas have specific concerns, with a larger emphasis on border security and interstate commerce and relations. People living in the regularly flooded areas are interested in flood protection, while for people who live in nearby higher-elevation area, flood protection is less of a concern.

All these are legitimate factors that need to be taken into account during redistricting.

Gerrymandering. In addition to legitimate concern, redistricting is also influences by politicking. In most states, redistricting is voted upon by the legislature. Whatever party is in power at the moment of redistricting tries to change the districts. As a result, in the next elections, the representation may be unfairly biased towards a larger representation for this party.

The possibility of this bias comes from the fact that if we have two parties A and B, then A votes in B-majority district are "lost" – since this district votes for B anyway. Similarly, B votes in A-majority districts are lost – since these districts vote for A anyway.

So, if party A is in power, it tries to divide all the A population into A-majority districts, and attach, to each district, as many B-voters as possible without violating these districts' A-majority character. The remaining B-voters are crammed into districts in which no one votes for A. In this manner, quite a lot of B votes are lost, while no A votes are lost. Once this plan is implemented, party A is guaranteed a much larger majority in the legislature than in the case when this representation was proportional to the number of voters voting for different parties.

For example, if we have 10 equal-size towns $A_1, \ldots, A_{10}$ voting for A and 8 same-size towns $B_1, \ldots, B_8$ voting for B, we can divide these 18 towns into the following 6 districts:

- the first district consists of A_1, A_2, and B_1;
- the second district consists of A_3, A_4, and B_2;
- the third district consists of A_5, A_6, and B_3;
- the fourth district consists of A_7, A_8, and B_4;
- the fifth district consists of A_9, A_{10}, and B_5;
- finally, the sixth district consists of the towns B_6, B_7, and B_8.

Under this division, in the first five districts the majority votes for A, so A gets 5 votes and B gets only 1. In the resulting legislature, A gets 5/6 $\approx$ 83% of the votes – while among the population as a whole, A is supported by a much smaller majority of 10/18 $\approx 55\%$.

In principle, by placing every B-voter into an A-majority district, part A can achieve a redistricting in which B will get no representation at all.

How to avoid gerrymandering: known algorithms. Researchers have proposed many algorithms to avoid gerrymandering. Typically, these algorithms limit the "weirdness" of the district shapes: e.g., make them as round as possible, with the smallest possible length of the separation lines.

Limitations of the existing algorithms. The problem with this approach is that it only takes into account geographical closeness, and it does not allow to adjust the district in such a way as to give minorities or rural population an adequate representation.

Remaining problem. In this paper, we will show how we can avoid gerrymandering and still take into account not only geographical closeness and differences, but also other types of closeness and difference as well.

II. Toward Formulation of the Problem in Precise Mathematical Terms

How to represent voters and their representatives. Each person in the area can be described by the values of several relevant parameters $x_1, \ldots, x_n$ – parameters which are important for this person's political representation. As we have mentioned, these parameters may include geographic location (so that the corresponding x_i are geographic coordinates), income, rural vs. urban status, number of children, and many many other important parameters.

Utility approach: general reminder. To every voter, an ideal representative is a person from exactly the same geographic, ethic, cultural, etc, background, a person who can perfectly represent this voter's preferences. In reality, we have to compromise, to find a solution which is best for everyone.

Such compromises are studied by game theory, a theory specifically designed to combine preferences of different people in decision making.

In decision theory, each person's preference of different alternatives is described by a number called its *utility*; see, e.g., [5]. One of the possible objectives is then to maximize the overall utility, i.e., the sum of utilities of different participants.

Utility approach to our problem. In the partitioning problem, we must select districts, and we must select a representative from each district. The resulting utility of a voter depends on how close this voter is to the corresponding representative. If we denote the parameters characterizing the voter by $v = (v_1, \ldots, v_n)$, and the parameters characterizing the representative by $r = (r_1, \ldots, r_n)$, then the utility of the voter depends on how big the difference $d \stackrel{\text{def}}{=} r - v$ between these sets of parameters is, i.e., how big the values $d_1 = r_1 - v_1, \ldots, d_n = r_n - v_n$ are. So, the utility function must depend on the values of these n differences $u = u(d_1, \ldots, d_n)$.

Different characteristics are independent. Different characteristics represent independent quantities. This case has been actively analyzed in decision theory. In particular, it has been proven that the corresponding utility function can be represented as the sum of "marginal" utility functions representing different quantities, i.e., $u(d_1, \ldots, d_n) = u_1(d_1) + \ldots + u_n(d_n)$; see, e.g., [2], [3].

Difference are small. In a good representation, representatives are close to the voters, so the values of the differences $d_i = r_i - v_i$ are small. Since the values d_i are small, we can expand each function $u_i(d_i)$ in Taylor series

$$u_i(d_i) = u_i(0) + u_i'(0) \cdot d_i + \frac{u''(0)}{2} \cdot d_i^2 + \ldots$$

and keep only the main terms in this expansion.

As we have mentioned, the largest possible utility is attained when a representative is a perfect match for the voter, i.e., when $d_i = 0$. Thus, the function $u_i(d_i)$ attains a maximum at $d_i = 0$. Since it attains the maximum, its derivative $u_i'(0)$ at $d_i = 0$ is equal to 0. So, the first non-trivial term in the Taylor expansion is a quadratic one:

$$u_i(d_i) \approx u_i(0) + \frac{u''(0)}{2} \cdot d_i^2 + \ldots$$

Since we have the maximum, the second derivative is non-positive, so we can describe the corresponding terms as $u_i(d_i) \approx u_i(0) - w_i \cdot d_i^2$ for some "weight" $w_i \geq 0$.

In this approximation, the overall utility function takes the form

$$u(d_1, \ldots, d_n) = \sum_{i=1}^{n} u_i(d_i) = \sum_{i=1}^{n} u_i(0) - \sum_{i=1}^{n} w_i \cdot d_i^2.$$

Thus, maximizing utility is equivalent to minimizing the following "disutility" function

$$U(d) = \sum_{i=1}^{n} w_i \cdot (r_i - v_i)^2.$$

If for some characteristic i, we have $w_i = 0$, this means that this characteristic does not affect the preferences and can therefore be ignored. So, in the following text, we will assume that all the weights are non-zero, i.e., that $w_i > 0$.

Resulting formulation of the problem. Maximizing overall utility is equivalent to minimizing overall disutility. As a result, we arrive at the following formulation of the problem. We have:

- an integer n (number of characteristics), and n positive real numbers $w_1, \ldots, w_n$ (weights of these characteristics);
- a collection of finitely many (N) n-dimensional vectors $x^{(k)} = (x_1^{(k)}, \ldots, x_n^{(k)})$, $1 \le k \le N$; we will say that the vector $x^{(k)}$ describes the k-th voter; and
- an integer c; this integer will be called the number of voting districts.

Our objective is to subdivide N voters $\{1, 2, \ldots, N\}$ into c groups $D_1, \ldots, D_c$, and select a vector $v(1), \ldots, v(c)$ within each group in such a way that the overall disutility

$$\sum_{j=1}^{c} \sum_{k \in D_j} \rho(x^{(k)}, v(j))$$

attains the smallest possible value, where we denoted

$$\rho(x, v) = \sum_{i-1}^{n} w_i \cdot (x_i - v_i)^2.$$

III. Clustering Approach to Solving this Problem: Main Idea and Limitations

Relation to clustering. In qualitative terms, we want voters within each districts to be close to each other, and we want each voter to be closer to other voters from this district than to voters from other districts.

In general, this subdivision is called *clustering*; see, e.g., [7]. So, our problem is a particular case of a well-studied problem of clustering. It is therefore reasonable to use clustering techniques to solve the voter districting problems.

Iterative clustering. In this paper, we will use one of the most natural clustering algorithms – an iterative clustering algorithm.

Main idea behind iterative clustering. The main idea behind this algorithm is as follows. Once we know the clusters $D_1, \ldots, D_v$, we can easily find the optimal representations for each cluster. Indeed, the optimal representative v_j for a cluster D_j can be determined from the condition that the disutility

$$U_j \stackrel{\text{def}}{=} \sum_{k \in D_j} \rho(x^{(k)}, v(j))$$

of this cluster be as small as possible. Differentiating this disutility with respect to the i-th component $v(j)_c$ and equating the derivative to 0, we conclude that $v(j)_c$ is the arithmetic average of all the values $x_j^{(k)}$ for all the elements $k \in D_c$ of this structure. (For simplicity, we assume that there is a person with exactly these average characteristics.)

Similarly, if we know the representatives $v(1), \ldots, v(c)$, then to minimize the overall disutility, we assign, to each voter k, the district D_j whose representative $v(j)$ is the closest to this voter, i.e., for which the disutility $\rho(x^{(k)}, v(j))$ attains the smallest possible value.

Thus, we arrive at the following algorithm.

Iterative clustering algorithm. We start with some representations $v(1), \ldots, v(c)$ (in politics, we can start with representatives of the existing districts).

We subdivide the voters into c groups corresponding to the representative: each point $x^{(k)}$ is assigned to the group D_j if is the value $\rho(x^{(k)}, v(j))$ is the smallest among all c values $\rho(x^{(k)}, v(1)), \ldots, \rho(x^{(k)}, v(c))$.

After that, for each group D_j, we re-calculate the value $v(j)$: the new value is the arithmetic average of the vectors $x^{(k)}$ corresponding to all $k \in D_j$.

Then, with the new starting point $v(1), \ldots, v(c)$, we repeat these two steps. Iterations continue until the process converges.

Limitations. The main limitation of this process is that we can end up with districts of different population size. For example, if we have a large city and a small town nearby, and only geographic division is relevant, then a natural 2-cluster subdivision will result in two unequal clusters:

- the first (larger) cluster contains all the voters from the big city, and
- the second (smaller) cluster contains all the voters from the small town.

IV. Towards a More Adequate Solution

Main idea. In the previous section, we have mentioned that a cluster approach can lead to a disproportional representation. To avoid this problem, in the above optimization problem, we must limit ourselves only to groups $D_1, \ldots, D_c$ of equal size N/c.

How can we modify the clustering algorithm to take this restriction into account?

Analysis of the new problem. If we fix the districts D_j, then the optimal location of each point $v(j)$ is still the arithmetic average of the corresponding vectors.

The problem is when we have the centers $v(j)$; what is then the optimal subdivision into the regions? The problem of finding an optimal set (or optimal sets) has been analyzed in [6]. By applying the approach from [6], we conclude that in the optimal subdivision, for every two neighboring districts D_i and D_j, the division corresponds to some fixed value of the ratio $\rho(x, v(i))/\rho(x, v(j))$:

- points x for which this ratio is smaller than a certain threshold are assigned to the class D_i, while
- points x for which this ratio is larger than the threshold are assigned to the class D_j.

In the absence of equal-size restriction, this threshold is equal to 1.

In general, we can therefore conclude that in the optimal solution, there exist some weights α_j such that each point x is assigned to the class D_j if the the value $\alpha_j \cdot \rho(x^{(k)}, v(j))$ is the largest among all c values $\alpha_1 \cdot \rho(x^{(k)}, v(1)), \ldots, \alpha_c \cdot \rho(x^{(k)}, v(c))$.

In these terms, the problem is to find the values α_j for which this subdivision leads to equal-size classes D_j.

Observation: multiplying all the coefficients α_j by the same constant. One can easily check that if we multiply all the values α_j by the same constant, we get exactly the same subdivision into classes. In the following analysis, we will use this property to simplify the formulas.

Towards an iterative process for computing the coefficients α_j. Let us start with some values $\alpha_1^{(p)}, \ldots, \alpha_c^{(p)}$. In the beginning, we can start with the values equal to 1, or with the values corresponding to the previous iteration of the whole algorithm.

For each j, we then find the auxiliary value β_j for which there are exactly N/c points x for which $\beta_j \cdot \rho(x, v(j))$ is smaller than all the values $\alpha_k^{(p)} \cdot \rho(x, v(k))$ for all $k \neq j$. This value β_j can be easily found by bisection (see, e.g., [1]). Indeed:

- if for some β_j, we get fewer than N/c points, this means that we have to decrease this value (and thus, increase the number of points that satisfy the desired property);
- on the other hand, if for some β_j, we get more than N/c points, this means that we have to increase this value (and thus, decrease the number of points that satisfy the desired property).

Once we find such values β_j, we must then find the new values α_j for which

$$\frac{\alpha_j}{\alpha_k} \approx \frac{\beta_j}{\alpha_k^{(p)}}$$

for all $j \neq k$.

Analysis of this auxiliary problem. This condition is non-linear in terms of the unknowns α_j. It can be simplified to a linear one if we consider logarithms instead of the original unknowns, i.e., if we consider $A_j \stackrel{\text{def}}{=} \ln(\alpha_j)$, $B_j \stackrel{\text{def}}{=} \ln(\beta_j)$, and $A_j^{(p)} \stackrel{\text{def}}{=} \ln(\alpha_j^{(p)})$, and take the logarithms of both sides of the desired condition. Since the logarithm of a ratio is equal to the difference between the logarithms, we conclude that

$$A_j - A_k \approx B_j - A_k^{(p)}.$$

This problem is linear in terms of the unknown A_j, so we can use the Least Squares Method to find these unknowns, i.e., find A_j for which the sum

$$\sum_j \sum_{k \neq j} (A_j - A_k - B_j + A_k^{(p)})^2$$

attains the smallest possible value.

Differentiating this expression with respect to A_j and equating the derivative to 0, we conclude that

$$\sum_{k \neq j} (A_j - A_k - B_j + A_k^{(p)}) + \sum_{k \neq j} (A_j - A_k - A_j^{(p)} + B_k) = 0.$$

The terms corresponding to $k = j$ add up to 0:

$$(A_j - A_j - B_j + A_j^{(p)}) + (A_j - A_j - A_j^{(p)} + B_j) = 0.$$

So, adding such terms to both sums, we get a simplified formula

$$\sum_k (A_j - A_k - B_j + A_k^{(p)}) + \sum_k (A_j - A_k - A_j^{(p)} + B_k) = 0,$$

or, equivalently,

$$2n \cdot A_j - 2 \sum_k A_k - n \cdot B_j + \sum_k B_k - n \cdot A_j^{(p)} + \sum_k A_k^{(p)} = 0.$$

So, we conclude that

$$A_j = \frac{1}{2} \cdot (B_j + A_j^{(p)}) + \text{const},$$

where the constant is the same for all j. Adding a constant to all the values $A_j = \ln(\alpha_j)$ means multiplying all the values α_j by the same constant. We have already mentioned that this multiplication does not change our subdivision into regions. Thus, for simplicity, we can simply take

$$A_j = \frac{1}{2} \cdot (B_j + A_j^{(p)}),$$

or, equivalently,

$$\alpha_j = \sqrt{\beta_j \cdot \alpha_j^{(p)}}.$$

V. Example

To illustrate our idea, let us consider a simple example of unequal distribution. Let us assume that we have a uniform population distribution on the interval $[0, 1]$, and that we start with 3 centers $v(1) = 0$, $v(2) = 0.5$, and $v(3) = 1.0$.

If we simply assign each point to the cluster corresponding to the nearest center, then:

- the points from the interval $[0, 0.25]$ (closest to $v(1) = 0$) are assigned the cluster D_1,
- the points from the interval $[0.25, 0.75]$ (closest to $v(2) = 0.5$) are assigned the cluster D_2, and
- the points from the interval $[0.75, 1]$ (closest to $v(3) = 1.0$) are assigned the cluster D_3.

In this arrangement, the clusters D_1 and D_3 have a quarter of the population, while the cluster D_2 has a half and is thus twice as large.

In this example, the (squared) distance from the first center is x^2, the squared distance from the second center is $(0.5-x)^2$, and the squared distance from the third center is $(1-x)^2$.

We started with the weights $\alpha_1^{(1)} = \alpha_2^{(1)} = \alpha_3^{(1)} = 1$.

To avoid the difference in size, let us find the value β_1 for which there are exactly 1/3 of the points for which

$$\beta_1 \cdot x^2 \le (0.5 - x)^2.$$

(We should not worry about the third class since it does not have a common boundary with the first one.) For $\gamma_1 \stackrel{\text{def}}{=} \sqrt{\beta_1}$, this condition means $\gamma_1 \cdot x \le 0.5 - x$, i.e., equivalently,

$$(\gamma_1 + 1) \cdot x \le 0.5$$

and $x \le \dfrac{0.5}{1+\gamma_1}$. Thus, the portion of voters who satisfy this inequality is exactly $\dfrac{0.5}{1+\gamma_1}$. We want this portion to be 1/3, so we must have $\dfrac{0.5}{1+\gamma_1} = 1/3$ and thus, $\gamma_1 = 0.5$. So, $\beta_1 = (\gamma_1)^2 = 0.25$.

Similarly, for the second cluster, the condition

$$\beta_2 \cdot (0.5 - x)^2 \le x^2$$

is equivalent to $\gamma_2 \cdot (0.5 - x) \le x$ for $\gamma_2 \stackrel{\text{def}}{=} \sqrt{\beta_2}$ (and to a similar inequality on the boundary with the third cluster). This is equivalent to $\gamma_2 \cdot (0.5+x) \le x$, i.e., to $\gamma_2 \cdot 0.5 \le (1+\gamma_2) \cdot x$ and to $x \ge \dfrac{\gamma_2 \cdot 0.5}{1+\gamma_2}$. To make districts equal in size, we need to make the borderline 1/3, so we must have $\dfrac{\gamma_2 \cdot 0.5}{1+\gamma_2} = 1/3$ and thus, $\gamma_2 = 2$. Hence, $\beta_2 = (\gamma_2)^2 = 4$.

Similarly, we get $\beta_3 = 0.25$. Now, we can compute the next iteration to α_j:

- we have $\alpha_1^{(2)} = \sqrt{\beta_1 \cdot \alpha_1^{(1)}} = \sqrt{0.25 \cdot 1} = 0.5$;
- we have $\alpha_2^{(2)} = \sqrt{\beta_2 \cdot \alpha_2^{(1)}} = \sqrt{4 \cdot 1} = 2$;
- we have $\alpha_3^{(2)} = \sqrt{\beta_3 \cdot \alpha_3^{(1)}} = \sqrt{0.25 \cdot 1} = 0.5$.

Under the new values, the condition that $\alpha_1 \cdot \rho(x, v(1))$ is the smallest of the three weighted distances is now equivalent to $0.5 \cdot x^2 \le 2 \cdot (0.5 - x)^2$, i.e., to $x^2 \le 4 \cdot (0.5 - x)$ and to $x \le 2 \cdot (0.5 - x) = 1 - 2x$ and $3x \le 1$. So, we have exactly 1/3 of the voters in the first cluster. Similarly, we have exactly 1/3 of the voters in the second and in the third clusters. Thus, here, we will have $\beta_j = \alpha_j^{(2)}$ for all j and thus, $\alpha_j^{(3)} = \sqrt{\beta_j \cdot \alpha_j^{(2)}} = \alpha_j^{(2)}$ – the same value as on the previous iteration. The process has converged.

Comment. Our preliminary experiments show that the process does converge in more realistic cases as well. (Of course, the above fast and perfect convergence only happens for "toy"" (simplified) examples.)

VI. Resulting Iterative Algorithm

We start with some representations $v(1), \ldots, v(c)$ (in politics, we can start with representatives of the existing districts).

We subdivide the voters into c groups corresponding to the representative. This subdivision is based on the values α_j of the weights assigned to each group (to equalize the group sizes). On each iteration of the clustering algorithm, we use several iterations to find these values. To make a description clearer, we will call the iterations of the whole clustering process *big* iterations, and the iterations needed to find the values α_j *small* iterations.

We start small iterations with the values

$$\alpha_1^{(1)} = \ldots = \alpha_c^{(1)} = 1$$

(or with the values for the previous big iteration). Once we have the values $\alpha_j^{(p)}$ corresponding to the current small iteration, then, for each group j, we find the value β_j for which there are exactly N/c points $x^{(k)}$ for which the value $\beta_j \rho(x^{(k)}, v(j))$ is small than all the $c-1$ values $\alpha^{(p)} \cdot \rho(x^{(k)}, v(l))$ corresponding to other clusters $l \ne j$. This value β_j can be found by using bisection.

Then, we take $\alpha_j^{(p+1)} = \sqrt{\beta_j \cdot \alpha_j^{(p)}}$. Iterations continues until the process converges, i.e., until the differences

$$\alpha_j^{(p+1)} - \alpha_j^{(p)}$$

become smaller than a given small value ε.

Once we reach convergence, we subdivide the voters into c groups corresponding to the representative: each point $x^{(k)}$ is assigned to the group D_j if is the value $\alpha_j \cdot \rho(x^{(k)}, v(j))$ is the smallest among all c values

$$\alpha_1 \cdot \rho(x^{(k)}, v(1)), \ldots, \alpha_c \cdot \rho(x^{(k)}, v(c)).$$

After that, for each group D_j, we re-calculate the value $v(j)$: the new value is the arithmetic average of the vectors $x^{(k)}$ corresponding to all $k \in D_j$.

Then, with the new starting point $v(1), \ldots, v(c)$, we repeat the two steps of the big iteration process. Iterations continue until the process converges, i.e., until the differences between the values $v(j)$ at the two consequent iterations become smaller than a given small value ε.

VII. Conclusion

In this paper, we describe a new algorithm for dividing an area into voting districts, an algorithm that can take into account not only geographic closeness, but also common interests of voters. This algorithm does not provide a single division: it provides a division for every assignment of weights w_i to different factors. These weights must be determine empirically, to best reflect the population preferences.

Of course, since several different re-districtings are possible, it will still be possible to argue which re-districting is the best, but this time, the argument will be about exact and empirically checkable things: weights of different factors.

Acknowledgments

This work was supported in part by NSF grants HRD-0734825, EAR-0225670, and EIA-0080940, by Texas Department of Transportation grant No. 0-5453, by the Japan

Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08, and by the Max Planck Institut für Mathematik.

The authors are thankful to the anonymous referees for valuable suggestions.

References

[1] Th. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, MA, 2001.

[2] P.C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.

[3] P.C. Fishburn, *Nonlinear preference and utility theory*, The John Hopkins Press, Baltimore, MD, 1988.

[4] International Institute for Democracy and Electoral Assistance (IDEA), *Electoral Systems Design: A New IDEA Handbook*, International IDEA, Stockholm, Sweden, 2005.

[5] R. D. Luce and H. Raiffa, *Games and Decisions*, Dover, New York, 1989.

[6] H. T. Nguyen and V. Kreinovich, "How to Divide a Territory? A New Simple Differential Formalism for Optimization of Set Functions", *International Journal of Intelligent Systems*, 1999, Vol. 14, No. 3, pp. 223–251.

[7] H. C. Romesburg, *Cluster Analysis for Researchers*, 2004.