SUR LA LIAISON ET LA DIVISION DES POINTS D'UN ENSEMBLE FINI

Travail collectif, rédigé par J. Łukaszewicz et contenant les résultats obtenus par le Groupe Général des Applications de l'Institut Mathématique de l'Etat: K. Florek, J. Łukaszewicz, J. Perkal, H. Steinhaus et S. Zubrzycki (Wrocław).

Ce travail est un sommaire géométrique de certaines applications de mathématiques aux sciences naturelles. Nous y exposons une méthode de déterminer la plus courte ligne polygonale dont les sommets se trouvent aux points d'un ensemble fini, donné d'avance; nous déterminons aussi la meilleure division d'un ensemble de n points en m parties ($m \leqslant n$) et indiquons une méthode effective de telle division [1]).

1. Soit Z un ensemble composé des points $A_1, A_2, \ldots, A_n$. Pour chaque couple A_i, A_j de points de Z soit définie la *distance* $\varrho(A_i, A_j) \geqslant 0$ telle que

(a) $\varrho(A_i, A_j) = 0$ lorsque $i = j$, et seulement dans ce cas,

(b) $\varrho(A_i, A_j) = \varrho(A_j, A_i)$,

(c) les distances non nulles ne se répètent pas dans l'ensemble Z, c'est-à-dire $\varrho(A_i, A_j) \neq \varrho(A_k, A_l)$ si, parmi les indices i, j, k, l, trois sont distincts.

On appellera tout couple A_i, A_j des points *segment* A_iA_j, eux-mêmes *extrémités* du segment. La *longueur du segment* A_iA_j sera la distance entre ses extrémités; on ne fera pas de distinction entre A_iA_j et A_jA_i.

Un ensemble quelconque de segments s'appellera *ligne polygonale*, et toutes les extrémités des segments — *sommets* de la ligne. La *longueur d'une ligne polygonale* sera la somme des longueurs de ses segments.

[1]) Ces méthodes de liaison et de division ont été proposées aux anthropologues pour ranger les crânes des fouilles. Elles s'appliquent aussi, avec effet, à des problèmes de biologie, d'agriculture, de technologie, même de linguistique. Voir *Taksonomia Wrocławska*, travail du Groupe Général des Applications de l'Institut Mathématique de l'Etat (Przegląd Antropologiczny, à paraître). Cf. aussi la communication collective *Une méthode taxonomique et ses applications aux sciences naturelles*, ce fascicule, p. 319.

Une ligne polygonale sera dite *connexe* (ou *liaison des sommets*) s'il existe, pour chaque couple A_i, A_j de ses sommets, une suite des segments lui appartenant

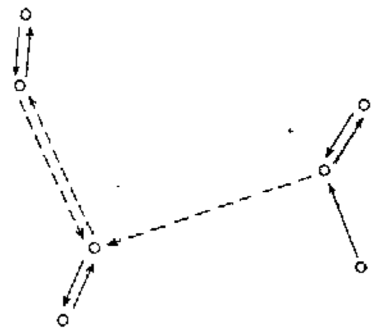
(1) $$A_{i_1}A_{i_2},\ A_{i_2}A_{i_3},\ \ldots,\ A_{i_{s-1}}A_{i_s},$$

où $A_{i_1} = A_i$, $A_{i_s} = A_j$.

Une ligne polygonale sera appelée *dendrite* s'il existe, pour chaque couple A_i, A_j de ses sommets, exactement une suite (1) réalisant la connexité et composée de segments différents deux-à-deux.

Soit $D(Z)$ la plus courte liaison des points de l'ensemble Z. La liaison $D(Z)$ est une dendrite, car s'il y avait deux suites des segments liant une paire des points, l'on pourrait obtenir une liaison plus courte en supprimant un segment.

Voici notre méthode pour former une dendrite avec les points de l'ensemble Z. Unissons, par un segment, chacun d'eux au point le plus proche; les segments ainsi obtenus seront appelés *liens du premier ordre*. Ils forment une ou plusieures lignes polygonales connexes qui sont des liaisons de points de certains sous-ensembles disjoints de Z. Nous appellerons ces sous-ensembles *groupements du premier ordre*. Lions chacun d'eux avec le groupement le plus proche (par *distance entre deux groupements* on comprend évidemment la plus petite distance entre leurs points deux-à-deux) à l'aide d'un segment qui réalisera la distance entre eux et que nous appellerons maintenant *lien du second ordre*. On procède ainsi, en employant des liens d'ordre de plus en plus élevé, jusqu'à ce qu'on obtienne une ligne polygonale connexe liant tous les points de l'ensemble Z. Désignons par $F(Z)$ cette liaison des points de Z. La liaison $F(Z)$ contient exactement $n-1$ liens, car chacun d'eux, dans la construction $F(Z)$, réduit le nombre de groupements d'une unité (chaque point de Z séparément étant considéré comme groupement d'ordre nul).



Théorème I.

$$F(Z)=D(Z).$$

Démonstration. Comme $D(Z)$ est la plus courte liaison des points de Z, il suffit de démontrer que $F(Z)\subset D(Z)$. Considérons un segment arbitraire $a=A_iA_j$ appartenant à la liaison $F(Z)$. Supposons, que ce segment soit, dans la construction $F(Z)$, un lien d'ordre k attachant le groupement U d'ordre $k-1$ à V, le plus proche groupement de même ordre. Si a n'appartenait pas à $D(Z)$, celle-ci devrait contenir une suite de segments

(2) $$A_{i_1}A_{i_2},\ A_{i_2}A_{i_3},\ \ldots,\ A_{i_{s-1}}A_{i_s},$$

où $A_{i_1}=A_i$, $A_{i_s}=A_j$. Dans cette suite figurerait le segment $b=A_{i_r}A_{i_{r+1}}$ dont une extrémité seulement appartient au groupement U. Le segment b serait plus long que a puisque ce dernier, vu la construction de $F(Z)$, est le plus court qui lie U avec les autres groupements. En enlevant de $D(Z)$ le segment b et ajoutant à $D(Z)-b$ le segment a, on aurait, contrairement à l'hypothèse, une liaison des points de Z plus courte que $D(Z)$. Le segment a doit donc appartenir à $D(Z)$, ce qui termine la démonstration.

Le théorème ci-dessus implique immédiatement:

(a) il existe exactement une liaison $D(Z)$ qui est la plus courte et elle se compose de $n-1$ segments;

(b) la liaison $F(Z)$ est une dendrite.

2. La somme des longueurs des liaisons $D(Z_k)$ s'appelle *longueur de la division* de l'ensemble Z en m sous-ensembles disjoints

(3) $$Z=Z_1+Z_2+\ldots+Z_m.$$

La division de l'ensemble Z en m parties sera dite *la meilleure* lorsque sa longueur est minimum[2]).

Pour lier les points à l'intérieur de tous les sous-ensembles Z_k, $n-m$ segments suffisent. Réciproquement, $n-m$ segments arbitraires, qui ne forment pas des polygones fermés, déterminent une division de l'ensemble Z en m sous-ensembles joints à l'intérieur. En particulier, $n-m$ segments arbitraires de la plus courte liaison $D(Z)$ déterminent une telle division. Nous allons démontrer que l'on obtient la meilleure division en enlevant de $D(Z)$ les $m-1$ segments les plus longs. A cet effet nous sert le lemme suivant:

Lemme. Si la division (3) *est la meilleure division de l'ensemble Z en m parties, alors*

$$\sum_{k=1}^{m} D(Z_k)\subset D(Z).$$

Démonstration. Soit $a=A_iA_j\epsilon\sum_{k=1}^{m}D(Z_k)$. Nous allons montrer que $a\epsilon D(Z)$. Vu la disjonction des ensembles (3), il existe exactement un h tel que $a\epsilon D(Z_h)$. Si l'on élimine le segment a de $D(Z)$, l'ensemble Z_h se décomposera en deux ensembles U et V. Si l'on avait $a\bar{\epsilon} D(Z)$, $D(Z)$ contiendrait la suite de segments

(4) $$A_{i_1}A_{i_2},\ A_{i_2}A_{i_3},\ \ldots,\ A_{i_{s-1}}A_{i_s},$$

où $A_{i_1}=A_i$, $A_{i_s}=A_j$. Le segment $b=A_{i_r}A_{i_{r+1}}$, dont une extrémité seulement appartient à U, figurerait dans cette suite. Si b était plus court que a, on aurait, en enlevant a de $\sum_{k=1}^{m}D(Z_k)$ et en ajoutant b à $\sum_{k=1}^{m}D(Z_k)-a$, m liaisons disjointes dont la somme de longueurs serait plus petite que la longueur $\sum_{k=1}^{m}D(Z_k)$, contrairement à l'hypothèse. Si b était plus long que a, alors $D(Z)-b+a$ serait une liaison plus courte que $D(Z)$. Le segment a doit donc appartenir à $D(Z)$, ce qui termine la démonstration.

De ce lemme résulte immédiatement le théorème annoncé:

Théorème II. On obtient la meilleure division de l'ensemble Z en m parties en enlevant de la liaison $D(Z)$ les $m-1$ segments les plus longs, les liaisons qui restent étant les plus courtes liaisons $D(Z_k)$ des parties respectives.

Państwowy Instytut Matematyczny
Institut Mathématique de l'Etat

[2]) On peut aussi faire usage d'une division en groupements d'ordre k. Nous l'avons appliquée, par exemple, à l'étude des groupements de villes.