MATHÉMATIQUES ET SCIENCES HUMAINES

M. BARBUT

Échelles à distance minimum d'une partie donnée d'un treillis distributif fini

Mathématiques et sciences humaines, tome 18 (1967), p. 41-46

<http://www.numdam.org/item?id=MSH_1967__18__41_0>

M. BARBUT

ECHELLES A DISTANCE MINIMUM D'UNE PARTIE DONNEE D'UN TREILLIS DISTRIBUTIF FINI

En analyse hiérarchique, en théorie des choix à critères multiples, etc ..., on se pose souvent la question: étant donné une partie A du treillis distributif T produit direct de p ordres totaux finis, trouver l'échelle (la chaîne maximale) à distance minimum de A.

Ce problème est étudié ici pour un treillis distributif fini quelconque, et pour les distances "naturelles" sur les treillis distributifs, celles qui sont définies à partir d'une mesure (valuation au sens de G. BIRKHOFF) sur T. Cette note complète ainsi celle de C. LERMAN parue dans M.S.H. n° 17, p. 37; comme on l'avait alors annoncé. Les quatre premiers paragraphes de la présente note montrent que toute échelle répondant à la question contient nécessairement tous les éléments d'un certain ensemble construit à partir de A: la chaîne de A.

Le paragraphe 5 donne ensuite un algorithme de construction des échelles cherchées.

Enfin, le paragraphe 6 généralise au cas où les éléments de T sont pondérés.

Les résultats concernant la métrique des treillis distributifs sur lesquels on s'appuie ici sont dans une note ronéotypée (épuisée): "Médiane, distributivité, éloignements" de février 1961 (C.M.S.S. - E.P.H.E.).

T étant un treillis distributif fini, une <u>échelle</u> de T est ici une chaîne (partie totalement ordonnée) de T allant du minimum de T à son maximum. E est donc une échelle si:

$$E = \{m = x_0 < x_1 < x_2 < \dots < x_n = M\}$$

où n désigne la hauteur de T, m son minimum, M son maximum.

Une chaîne <u>partielle</u> est une partie totalement ordonnée de cardinal inférieur à n.

1) MESURES ET DISTANCES SUR T

a) - On appelle mesure sur T toute application $f : T \to R^*$ ($R^* = \{\text{réels positifs}\}$) satisfaisant à:

$$\forall x, \forall y : f(x \vee y) + f(x \wedge y) = f(x) + f(y)$$

- f est une mesure positive si f est strictement monotone, et $f(m) = 0$.
- Si f est une mesure positive, l'application $d : T^2 \longrightarrow R^*$ définie par:

$$d(x,y) = f(x \vee y) - f(x \wedge y)$$

est une distance; on sait que cette distance satisfait à:

$$d(x,y) + d(y,z) = d(x,z) \Longleftrightarrow x \wedge z \leqslant y \leqslant x \vee z$$

et que les seuls treillis où puisse être définie une telle distance sont les treillis distributifs (cette distance généralise pour les treillis distributifs la distance de la mesure de la différence symétrique pour le treillis booléen des parties d'un ensemble).

b) - Si $a \in T$ est un élément fixé, la fonction:

$$d(a,x) = g(x) \qquad (d(a,x) = f(a \vee x) - f(a \wedge x) \text{ f mesure positive})$$

est une mesure. En effet, on vérifie immédiatement que:

$$g(x \vee y) + g(x \wedge y) = g(x) + g(y).$$

c) - Si φ et ψ sont des mesures, $\varphi + \psi$ est une mesure (évident); $\lambda\varphi$ est une mesure. Les mesures sur **T** forment un vectoriel.

2) MESURE D'UNE PARTIE DE T ET CHAINE D'UNE PARTIE

$A = \{a_1, a_2, \ldots, a_k\}$ étant une partie de T, et f étant une mesure sur T, on appelle mesure * de A : $F(A) = \sum_{i=1}^{k} f(a_i)$.

Désignons par U_h l'infimum des supremums des parties de cardinal h de A (c'est aussi le supremum des infimums des parties de cardinal $n + 1 - h$)

$$U_1 = \wedge a_i = a_1 \wedge a_2 \wedge \ldots \wedge a_k$$

$$U_2 = \bigwedge_{i \neq j} (a_i \vee a_j) = \vee (a_{i_1} \wedge a_{i_2} \wedge \ldots \vee a_{i_{k-1}})$$

$$\vdots$$

$$U_k = \wedge (a_1 \vee a_2 \vee \ldots \vee a_k) = a_1 \vee a_2 \vee \ldots \vee a_k$$

On sait que : $U_1 \leqslant U_2 \leqslant U_3 \leqslant \ldots \leqslant U_k$; les U_j sont sur une <u>chaine</u> de T; ils déterminent la <u>chaine</u>(partielle) de A. f satisfait à:

$$\sum_{i=1}^{k} f(a_i) = \sum_{j=1}^{k} f(U_j)$$

(Cette identité peut se montrer par récurrence sur k , en se servant de l'égalité fonctionnelle des mesures).

3) DISTANCE D'UNE ECHELLE A UNE PARTIE

a) - Si A est une partie quelconque de T, et d une distance sur T, la fonction d (a,x) étant une mesure, pour tout $a \in A$, posons:

$$d(A,x) = \sum_{a \in A} d(a,x) = g(x)$$

---

* $F : P(T) \longrightarrow R$ satisfait évidemment à la relation $\forall A, \forall B, F(A \cup B) + F(A \cap B) = F(A) + F(B)$

g est une mesure* (comme somme de mesures). Si $\mathcal{U} = \{U_1 < U_2 < \dots < U_k\}$ est la chaîne partielle déterminée par A, on a:

$$d(A,x) = d(\mathcal{U},x)$$

b) - Soit E une échelle de T, et A une partie quelconque de T. Posons:

$$d(E,A) = \sum_{\substack{x \in E \\ y \in A}} d(x,y)$$ (C'est une distance moyenne entre E et A).

Comme, pour chaque $x \in E$, $\sum_{y \in A} d(x,y) = d(x,A) = d(x,\mathcal{U})$

où $\mathcal{U}$ est la chaîne partielle de A, alors:

$$d(E,A) = d(E,\mathcal{U}).$$

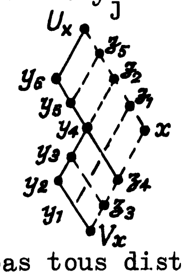
c) - Les échelles de T à distance minimum de A sont donc les échelles à distances minimum de la chaîne partielle $\mathcal{U}$ de A.

4) LES ECHELLES A DISTANCE MINIMUM D'UNE CHAINE DE T CONTIENNENT CETTE CHAINE

a) - Soit x un élément de T, E une échelle; si $x \notin E$, il existe une échelle E' passant par x et plus proche de x que E.

Soit $E = (m = Y_0 < y_1 < \dots < y_n = M)$; $U_x = \inf. y_i$, $y_i \geq x$
$V_x = \sup. y_i$, $y_i \leq x$.

Considérons les éléments $x \wedge y_j$ et $x \vee y_j$, $(v_x < y_j < U_x)$.



Ces éléments ne sont pas tous distincts; avec x, ils forment une chaine maximale de U_x à V_x ; comme toutes les chaines maximales de U_x à V_x ont même longueur, ils sont en nombre (h-1) (en nombre h avec x) si h est la différence des hauteurs entre U_x et V_x. Et l'on a:

$$\sum_{V_x < y_j < U_x} d(x,y_j) = \sum d(x,z_j) + \sum d(z_j,Y_j)$$

en notant $z_1, z_2, \dots, z_{h-1}$, ces éléments; en effet:

$$d(x,y_j) = d(x \wedge y_j, y_j) + d(x, x \wedge y_j)$$
$$= d(x \vee y_j, y_j) + d(x, x \vee y_j)$$

Donc l'échelle E' coïncidant avec E entre m et V_2, et U_x et M, et avec z_1 z_2 ... z_{h-1} entre U_x et V_x est telle que:

* A un facteur indépendant de x près, g(x) est la distance moyenne de x à la partie A.

$$d\ (x,E) = d\ (x,E') + \sum_{V_x<y_j<U_x}{}' d\ (z_j,y_j)$$

<u>Note</u>:

Si E est une échelle partielle, le résultat reste vrai; à condition toutefois de remplacer x par l'élément $x^* = (x \vee y_m) \wedge y_M$ ou y_m et y_M désignent le minimum le maximum de **E**; l'échelle E' passant par x^* et construite comme supra est plus proche de x que E.

b) - De 4°,a) et 3°c), il résulte immédiatement que les échelles à distance minimum d'une partie A donnée de T contiennent toutes la chaîne partielle U de A.

Dans le cas particulier où $f\ (x), = |x|$ où $|x|$ est la hauteur de x, (c'est à dire la cardinal commun des chaînes maximales de m à x) toutes les échelles passant par un point donné y de T sont à même distance de y; cette distance vaut:

$$\frac{1}{2}\left[|y|\ (\ |y| + 1) + (n - |y|\ )\ \ (\ n - |y| + 1\ )\right]$$

et l'ensemble des échelles à distance minimum de A est constitué par toutes les échelles contenant la chaîne partielle U de A.

5) ALGORITHME DE DETERMINATION DES ECHELLES A DISTANCE MINIMUM

a) - Dans le cas général où $f\ (y)$, la mesure positive qui sert à définir la distance, n'est pas la hauteur $|y|$, posons encore, pour chaque élément $x \in T$:

$g\ (x) = d\ (x,A) = d\ (x,\mathcal{U})$, A étant une partie donnée de T, et $\mathcal{U}$ sa chaîne.

Il faut trouver une échelle:

$$E = \left\{y_0 = m < y_1 < \ldots < y_n = M\right\}$$

telle que: $\sum_{i=1}^{n} g\ (y_i)$ soit minimum.

Pour chaque $x \in T$, soit:

$$G\ (x) = \min \sum_{y_i < x} g\ (y)$$

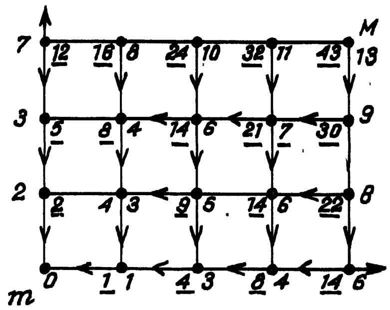
minimum pris sur l'ensemble des chaînes maximales de m à x.

Soit C_x l'ensemble des éléments <u>couverts</u> par x; alors

$$G\ (x) = g\ (x) + \min_{y \in C_x} G\ (y\ )$$

Par suite, on peut déterminer de façon récurrente G (x) en chaque $x \in T$, et ipso facto toutes les échelles à distance minimum de A.

Exemple:



T, produit direct de deux ordres totaux.

Nombre non souligne: valeur de g (x)
Nombre souligné : valeur de G (x)
Les flèches indiquent les cheminements optimaux.

b) - Comme on sait que les échelles optimales passent par les points de $\mathcal{U}$, l'algorithme peut être allégé (si $\mathcal{U}$, est de cardinal élevé) car:

1 - Il suffit de déterminer la fonction pour les seuls éléments de T situés dans les intervalles:

$$[m, U_1], \quad [U_1, U_2], \quad \dots, \quad [U_k, M]$$

2 - Il suffit ensuite de déterminer G pour ces seuls éléments.

3 - Mais la fonction g étant une mesure, il suffit de la calculer pour une partie génératrice de T (ou des intervalles de T nous intéressent) pour l'avoir partout au moyen d'additions rapides: dans l'exemple supra, la donnée de g pour les "marges" suffit. D'une façon générale, il suffit lorsque T est produit direct de p ordres totaux, de connaître g sur chacun de ces ordres totaux.

6) CAS OU LES ELEMENTS DE T SONT PONDÉRÉS

A chaque $x \in T$ on associe le nombre p_x; en pratique, on a le plus souvent $p_x \in \{0, 1, 2, \dots\}$

(p_x pourrait être, en analyse hiérarchique par exemple, le nombre de "sujets" fournissant un "patron" x); $p_x = 0$ si le patron x n'est pas représenté. On peut encore définir:

$$\forall y \in T, \quad g(y) = \sum_{x \in T} p_x \; d(x, y)$$

où d est une distance définie comme en 1°). D'après 1°, b et c), g est une mesure:

$$g(y) + g(z) = g(y \vee z) + g(y \wedge z)$$

Elle est donc calculable aisément pour tout $y \in T$ dès que l'on connait sa valeur pour chaque élément d'une partie génératrice.

La méthode de 5°) a) s'applique alors sans changement pour déterminer les échelles E

$$E = \left\{ m = y_0 < y_1 < y_2 < \ldots < y_n = M \right\}$$

telles que $\sum_{i=0}^{n} g(y_i)$ soit minimum.

Critique:

1) Ce qui rend le problème si simple, c'est le fait que les valeurs (les p_x, puis les $g(x)$) sont attachées aux sommets (éléments) du treillis, et non aux arcs. C'est aussi le fait que la distance d'un élément à une partie, et de deux parties entre elles étant calculées comme des moyennes (ou des sommes); on reste dans le cadre de l'Algèbre linéaire.

2) Le problème traité dans cette note est distinct de celui qui a fait l'objet de ma note "Sur les ordres totaux à distance minimum d'une relation binaire" donnée dans M.S.H. n° 17, p. 47. Le lecteur intéressé pourra comparer les deux problèmes pour le cas où ils sont comparables, c'est-à-dire celui où le treillis distributif T de la présente note est celui des parties d'un ensemble fini.