Computation of median orders: complexity results

Olivier Hudry
École nationale supérieure des télécommunications
46, rue Barrault, 75634 Paris cedex 13, France
(Olivier.Hudry@enst.fr)

Abstract. Given a set of individual preferences defined on a same finite set, we consider the problem of aggregating them into a collective preference minimizing the number of disagreements with respect to the given set and verifying some structural properties like transitivity. We study the complexity of this problem when the individual preferences as well as the collective one must verify different properties, and we show that the aggregation problem is NP-hard for different types of collective preferences, even when the individual preferences are linear orders.

Keywords. Complexity, partially ordered relations, median relations, aggregation of preferences.

1. Introduction

The problem that we deal with in this communication can be stated as follows: given a set (called a *profile*) Π of binary relations defined on the same finite set X, find a binary relation R^* defined on X verifying certain properties like transitivity and summarizing Π as accurately as possible. This problem occurs in different fields, for instance in the social sciences, in electrical engineering, in agronomy or in mathematics (see for example J.-P. Barthélemy *et alii* (1995), J.-P. Barthélemy and B. Monjardet (1981 and 1988), A. Guénoche *et alii* (1994), L. Hubert (1976) or M. Jünger (1985) for references). For example, in voting theory, X can be considered as a set of candidates, Π as a profile of individual preferences expressed by voters and R^* as the collective preference that we look for.

The aim of this communication is to study the complexity of finding R^*. We consider different types of ordered relations for the individual preferences of Π as well as for R^* and we show that for most cases, the decision problem associated with the determination of R^* is NP-complete. This problem has been already studied in some special cases, namely for the aggregation of a profile of linear orders into a linear order by J.B. Orlin (1988) and by J.J. Bartholdi III, C.A. Tovey and M.A. Trick (1989), and for the aggregation of a profile of binary relations into a linear order, a partial order, a complete preorder or a preorder (see below for the definitions of these structures) by Y. Wakabashi (1986 and 1998). We generalize these results by extending them to other cases.

This extended abstract is organized as follows. Section 2 recalls the definitions of the ordered relations that we take into account. In section 3, we show how the aggregation problem can be formulated in graph theoretical terms. Then we state our complexity results upon this aggregation problem without proofs in the last section (the proofs of these results can be found in O. Hudry (1989)).

2. The ordered relations

Given a finite set X, a binary relation R defined on X is a subset of $X \times X = \{(x, y): x \in X \text{ and } y \in X\}$. We note n the number of elements of X and we suppose that n is at least equal to 4. We note xRy instead of $(x, y) \in R$ and $x\overline{R}y$ instead of $(x, y) \notin R$. The following properties that a binary relation R can satisfy are basic:

- *reflexive*: $\forall x \in X$, xRx;
- *irreflexive*: $\forall x \in X$, $x\overline{R}x$;
- *antisymmetric*: $\forall (x,y) \in X^2$, $(xRy \text{ and } x \neq y) \Rightarrow y\overline{R}x$;
- *asymmetric*: $\forall (x,y) \in X^2$, $xRy \Rightarrow y\overline{R}x$;
- *transitive*: $\forall (x,y,z) \in X^3$, $(xRy \text{ and } yRz) \Rightarrow xRz$;
- *complete*: $\forall (x,y) \in X^2$ with $x \neq y$, xRy or (inclusive) yRx.

From a binary relation R, we may define an asymmetric relation A_R (called the *asymmetric part* of R) by: $xA_Ry \Leftrightarrow (xRy \text{ and } y\overline{R}x)$.

By combining the above properties, we may define different types of binary relations (see for instance J.-P. Barthélemy and B. Monjardet (1981) or P.C. Fishburn (1985)):

- A *partial order* is an asymmetric and transitive binary relation; $\mathcal{O}$ will denote the set of the partial orders defined on X;
- a *linear order* is a complete partial order; $\mathcal{L}$ will denote the set of the linear orders defined on X;
- a *tournament* is a complete and asymmetric binary relation; $\mathcal{T}$ will denote the set of the tournaments defined on X;
- a *preorder* is a reflexive and transitive binary relation; $\mathcal{P}$ will denote the set of the preorders defined on X;
- a *complete preorder* is a reflexive, transitive and complete binary relation; $\mathcal{C}$ will denote the set of the complete preorders defined on X;
- a *weak order* is the asymmetric part of a complete preorder; $\mathcal{W}$ will denote the set of the weak orders defined on X;
- an *interval order* is a partial order R satisfying: $\forall (x, y, z, t) \in X^4$, $(xRy \text{ and } zRt) \Rightarrow \{xRt$ or (inclusive) $zRy\}$; $\mathcal{I}$ will denote the set of the interval orders defined on X;
- a *semiorder* is an interval order R satisfying: $\forall (x, y, z, t) \in X^4$, $(xRy \text{ and } yRz) \Rightarrow \{xRt$ or (inclusive) $tRz\}$; $\mathcal{S}$ will denote the set of the semiorders defined on X;
- a *quasi-order* is a complete relation of which the asymmetric part is a semiorder; $\mathcal{Q}$ will denote the set of the quasi-orders defined on X;
- an *acyclic relation* is a relation R verifying: $\forall\, 1 \le k \le n$, $(x_iRx_{i+1}$ for $1 \le i \le k-1) \Rightarrow x_k\overline{R}x_1$; $\mathcal{A}$ will denote the set of acyclic relations defined on X.

It is possible to get other structures by adding or by removing reflexivity or irreflexivity from the above definition (and by changing asymmetry by antisymmetry). In fact, the distinction between reflexive and irreflexive relations is not relevant for our study (see O. Hudry (1989)):

the complexity results will remain the same. Thus, in the following, we do not take reflexivity or irreflexivity into account (for instance, we will consider that a linear order is also a preorder).

These types include the most studied and used partially ordered relations. We will also consider generic binary relations, without any particular property. The set of the binary relations will be noted $\mathcal{B}$. We may notice several inclusions between these sets, especially the following one: $\forall\ \mathcal{Z} \in \{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{S}, \mathcal{T}, \mathcal{W}\}$, $\mathcal{L} \subseteq \mathcal{Z}$; in other words, a linear order can be considered as a special case of any one of the other types.

3. Formulation of the aggregation problem

In order to get an optimization problem to deal with, it is necessary to explicit what we mean when we say that R^* must summarize Π "as accurately as possible". To do so, we consider the symmetric difference distance δ: given two binary relations R and S defined on the same set X, we have

$$\delta(R, S) = \left|\left\{(x, y) \in X^2 : \left[xRy \text{ and } x\bar{S}y\right] \text{ or } \left[x\bar{R}y \text{ and } xSy\right]\right\}\right|$$

This quantity $\delta(R, S)$ measures the number of disagreements between R and S. Though some authors consider sometimes another distance, δ is used widely and is appropriate for many applications. J.-P. Barthélemy (1979) shows that δ satisfies a number of naturally desirable properties and J.-P. Barthélemy and B. Monjardet (1981) recall that $\delta(R, S)$ is the Hamming distance between the characteristic vectors of R and S and point out the links between δ and the L_1 metric or the square of the Euclidean distance between these vectors (see also K.P. Bogart (1973 and 1975) and B. Monjardet (1979 and 1990)).

Then, for a profile $\Pi = (R_1, R_2, \ldots, R_m)$ of m relations, we can define the *remoteness* $\Delta(\Pi, R)$ (J.-P. Barthélemy and B. Monjardet (1981)) between a relation R and the profile Π by:

$$\Delta(\Pi, R) = \sum_{i=1}^{m} \delta(R, R_i)$$

The remoteness $\Delta(\Pi, R)$ measures the total number of disagreements between Π and R.

Our aggregation problem can be seen now as a combinatorial problem: given a profile Π, determine a binary relation R^* minimizing Δ over one of the sets $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{S}, \mathcal{T}, \mathcal{W}$. Such a relation R^* will be called a *median relation* of Π (J.-P. Barthélemy and B. Monjardet (1981)). According to the properties assumed for the relations belonging to Π or required from the median relation, we get many combinatorial problems. They are too numerous to state all of them explicitly; so we note them as follows:

Problems $P_m(\mathcal{Y}, \mathcal{Z})$. For a positive integer m, for $\mathcal{Y}$ belonging to $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{S}, \mathcal{T}, \mathcal{W}\}$ and $\mathcal{Z}$ belonging also to $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{S}, \mathcal{T}, \mathcal{W}\}$, $P_m(\mathcal{Y}, \mathcal{Z})$ denotes the following problem: given a finite set X, given a profile Π of m binary relations all belonging to $\mathcal{Y}$, find a relation R^* belonging to $\mathcal{Z}$ with : $\Delta(\Pi, R^*) = \underset{R \in \mathcal{Z}}{\text{Min}}\, \Delta(\Pi, R)$.

To study the complexity of $P_m(\mathcal{Y}, \mathcal{Z})$, we develop the expression of Δ. For this, consider the characteristic vectors $r^i = \left(r^i_{xy}\right)_{(x,y) \in X^2}$ of the relations R_i ($1 \le i \le m$) defined by $r^i_{xy} = 1$ if xR_iy and $r^i_{xy} = 0$ otherwise, and similarly the characteristic vector $r = \left(r_{xy}\right)_{(x,y) \in X^2}$ of any binary relation R. Then, it is easy to get a linear expression of $\Delta(\Pi, R)$:

$$\delta(R, R_i) = \sum_{(x,y)\in X^2} \left| r_{xy} - r_{xy}^i \right| = \sum_{(x,y)\in X^2} \left| r_{xy} - r_{xy}^i \right|^2 = \sum_{(x,y)\in X^2} \left[r_{xy}\left(1 - 2r_{xy}^i\right) + r_{xy}^i \right]$$

hence $$\Delta(\Pi, R) = \sum_{i=1}^{m} \sum_{(x,y)\in X^2} \left| r_{xy} - r_{xy}^i \right|$$

and, after simplifications: $$\Delta(\Pi, R) = C - \sum_{(x,y)\in X^2} m_{xy} . r_{xy}$$

with $C = \sum_{i=1}^{m} \sum_{(x,y)\in X^2} r_{xy}^i$ and $m_{xy} = \sum_{i=1}^{m} \left(2r_{xy}^i - 1\right) = 2\sum_{i=1}^{m} r_{xy}^i - m$.

Notice that, with this expression of $\Delta(\Pi, R)$, it is easy to get a 0-1 linear programming formulation of the problems $P_m(\mathcal{Y}, \mathcal{Z})$ by adding the 0-1 linear constraints associated with each type of median relation; for example, the transitivity of R can be written: $\forall (x, y, z) \in X^3$, $r_{xy} + r_{yz} - r_{xz} \leq 1$ (see for instance Y. Wakabayashi (1986) or O. Hudry (1989) for details).

4. The complexity results

In this section, we pay attention to the complexity of $P_m(\mathcal{Y}, \mathcal{Z})$ for different types of profiles and different types of median relations.

For the profile Π, three cases will be distinguished below:

- in the first one, Π will be a profile of m binary relations, for any positive value of m;
- in the second one, Π will be a profile of m tournaments, for any even positive value of m;
- in the third one, Π will be a profile of a great enough number of relations belonging to any set $\mathcal{Y}$ including $\mathcal{L}$: $\mathcal{L} \subseteq \mathcal{Y}$; as noticed above, this allows all the previous sets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{I}$, $\mathcal{L}$, $\mathcal{O}$, $\mathcal{P}$, $\mathcal{Q}$, $\mathcal{S}$, $\mathcal{T}$, $\mathcal{W}$, as well as "mixed" sets such as $\mathcal{O} \cup \mathcal{P}$ or $\mathcal{A} \cup \mathcal{C} \cup \mathcal{T}$…

For the median relations, they belong to one of the sets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{I}$, $\mathcal{L}$, $\mathcal{O}$, $\mathcal{P}$, $\mathcal{Q}$, $\mathcal{S}$, $\mathcal{T}$, or $\mathcal{W}$.

We may summarize the complexity results by the following table. In this table, "NP-C" means that the decision problem associated with the considered problem $P_m(\mathcal{Y}, \mathcal{Z})$ is NP-complete, while "?" means that the complexity status of $P_m(\mathcal{Y}, \mathcal{Z})$ is unknown. The symbol ♥ (respectively ♣) shows the results got by Y. Wakabayashi (1986 and 1998) (respectively by J.B. Orlin (1988) and by J.J. Bartholdi III, C.A. Tovey and M.A. Trick (1989)). The NP-completeness of the case for which Π is reduced to only one binary relation while the median relation must be acyclic is directly given by the complexity of the well-known Feedback Arcset Problem (see M.R. Garey and D.S. Johnson (1979)); it is represented by the symbol ♠ in the table.

Median relation ($\mathcal{Z}$)	$\Pi \in \mathcal{B}^m$ ($\mathcal{Y} = \mathcal{B}$)	$\Pi \in \mathcal{T}^m$ ($\mathcal{Y} = \mathcal{T}$)	$\Pi \in \mathcal{Y}^m$ with $\mathcal{L} \subseteq \mathcal{Y}$
binary relation ($\mathcal{B}$)	(trivially) polynomial	(trivially) polynomial	(trivially) polynomial
tournament ($\mathcal{T}$)	(trivially) polynomial	(trivially) polynomial	(trivially) polynomial
acyclic relation ($\mathcal{A}$)	NP-C for any $m \geq 1$♠	NP-C for any m even	NP-C for m great
complete preorder ($\mathcal{C}$)	NP-C for m great♥	NP-C for m great	NP-C for m great
interval order ($\mathcal{I}$)	NP-C for any $m \geq 2$	NP-C for m great	NP-C for m great
linear order ($\mathcal{L}$)	NP-C for any $m \geq 1$♥	NP-C for any m even	NP-C for m great♣
partiel order ($\mathcal{O}$)	NP-C for any $m \geq 2$♥	?	?
preorder ($\mathcal{P}$)	NP-C for m great♥	?	?
quasi-order ($\mathcal{Q}$)	NP-C for m great	NP-C for m great	NP-C for m great
semiorders ($\mathcal{S}$)	NP-C for any $m \geq 2$	NP-C for m great	NP-C for m great
weak order ($\mathcal{W}$)	NP-C for m great	NP-C for m great	NP-C for m great

From this table, it appears that some cases are still unsolved, for instance for the computation of a median preorder of a profile of linear orders. It is also the case for special values of m. One such interesting case is the one for which Π is reduced to one tournament ($m = 1$) while the median relation must be a linear order. This problem is known as the Slater problem (P. Slater (1961)).

5. References

J.-P. Barthélemy (1979), Caractérisations axiomatiques de la distance de la différence symétrique entre des relations binaires, *Mathématiques et Sciences Humaines* 67, 85-113.

J.-P. Barthélemy, O. Hudry, G. Isaak, F. S. Roberts, B. Tesman (1995), The reversing number of a digraph, *Discrete Applied Mathematics* 60, 39-76.

J.-P. Barthélemy, B. Monjardet (1981), The median procedure in cluster analysis and social choice theory, *Mathematical Social Sciences* 1, 235 - 267.

J.-P. Barthélemy, B. Monjardet (1988), The median procedure in data analysis: new results and open problems, in *Classification and related methods of data analysis*, H.H. Bock (ed.), North-Holland, Amsterdam.

J.J. Bartholdi III, C.A. Tovey, M.A. Trick (1989), Voting schemes for which it can be difficult to tell who won the election, *Social Choice and Welfare* 6, 157-165.

K.P. Bogart (1973), Preference structures I : distance between transitive preference relations, *Journal of Mathematical Sociology* 3, 49-67.

K.P. Bogart (1975), Preference structures II : distances between asymetric relations, *SIAM Journal of Applied Mathematics* 29 (2), 254-262.

P.C. Fishburn (1985), *Interval orders and interval graphs, a study of partially ordered sets*, John Wiley, New York.

M.R. Garey, D.S. Johnson (1979), *Computers and intractability, a guide to the theory of NP-completeness*, Freeman, New York.

A. Guénoche, B. Vandeputte-Riboud, J.-B. Denis (1994), Selecting varieties using a series of trials and a combinatorial ordering method, *Agronomie* 14, 363-375.

L. Hubert (1976), Seriation using asymmetric proximity measures, *Br. J. Math. Statist. Psychol.* 29, 32-52.

O. Hudry (1989), *Recherche d’ordres médians : complexité, algorithmique et problèmes combinatoires*, thèse de l’ENST, Paris.

M. Jünger (1985), *Polyhedral combinatorics and the acyclic subdigraph problem*, Heldermann Verlag, Berlin.

B. Monjardet (1979), Relations à éloignement minimum de relations binaires, note bibliographique, *Mathématiques et Sciences Humaines* 67, 115-122.

B. Monjardet (1990), Sur diverses formes de la « règle de Condorcet » d’agrégation des préférences, *Math. Inf. Sci. hum* 111, 61-71.

J.B. Orlin, 1988, unpublished manuscript.

P. Slater (1961), Inconsistencies in a schedule of paired comparisons, *Biometrika* 48, 303-312.

Y. Wakabayashi (1986), *Aggregation of binary relations: algorithmic and polyhedral investigations*, PhD thesis, Augsbourg.

Y. Wakabayashi (1998), The Complexity of Computing Medians of Relations, *Resenhas* 3 (3), 323-349.