Journal of Convex Analysis 27 (2020) 053–078

J.-P. Aubin
VIMADES, 14 rue Domat, 75005 Paris, France
aubin.jp@gmail.com

Echelons of Sets on Dispersion Gaps: Tools for Cluster Analysis

Instead of analyzing time series of vectors or the problem of an allocation of vectors to clusters in vectors spaces, we shall investigate the same issues for time series and clusters of subsets K ranging over the hyperset $\mathcal{P}(X)$ of subsets of a set X of a plain set X deprived of any mathematical structure, let it be vectorial or topological. The arithmetic operations on vector spaces will be replaced by Boolean operations on hyperspaces. For that purpose, we rely on the ideas going back to Abraham de Moivre based on

1. *dispersion gaps* $[[A_1, A_2]](K)$ between two disjoint subsets A_1 and A_2 (called *gists* of the dispersion gap) of subsets K such that $A_1 \subset K \subset \complement A_2$ (instead of dispersion intervals $[v_1, v_2] \subset \mathbb{R}$); 2. *magnitudes* which are increasing hyperfunctions $\mu\colon K \in \mathcal{P}(X) \mapsto \mu(K) \in \mathbb{R}_+$ vanishing at the empty set (encompassing measure, capacities, etc.) of all denominations. The main instrument of measure of a set $K \in [[A_1, A_2]]$ between its two *gists* is its *echelon*

$$\mathbb{A}_\mu[[A_1, A_2]](K) \ := \ \frac{\mu(K \cap \complement A_1) - \mu(\complement(K) \cap \complement A_2)}{\mu(\complement A_1 \cap \complement A_2)} \ \in \ [-1, +1]$$

Its inverse associating with any echelon

$$e \in [-1, +1] \rightsquigarrow \mathbb{A}_\mu[[A_1, A_2]]^{-1}(e) \ \in \ [[A_1, A_2]]$$

the subsets sharing the same echelon. It plays the same role as the quantiles in statistics: it assigns the minimum value -1 to A_1 (instead of quantile 0) and $+1$ at $\complement A_2$, as the quantile 1. The subset $\mathbb{A}_\mu[[A_1, A_2]]^{-1}(0)$ plays the role of the median (instead of quantile 1/2). Actually, since we shall use the lattice operations sup and inf instead of the usual Kolmogorov measures $\int$, we were lead to use this new renormalization rule to compare all kind of magnitudes (Maslov measures, for instance).

We shall use magnitudes and echelons of sets to study time series of sets and clustering issues.

Keywords: Boolean structure, Choquet capacity, cluster, dispersion gap, echelon, extremal box, extremal envelope, extremal quantile, Galois filtration, hyperset, Kolmogorov measure, magnitude, Maslov measure, set filtration.

MSC: 34G25, 34A60, 49J27, 49J53, 93B03, 37B55, 28B20, 45N05.