Symposium MATHEMATICS AND APPLICATION, Faculty of Mathematics, University of Belgrade, 2026

Complexity of lasso with normalized data: geometry and extremal combinatorics

Vladimir Dragović

The University of Texas at Dallas/MISANU
e-mail: Vladimir.Dragovic@utdallas.edu
ORCID: xxxx-xxxx-xxxx-xxxx

Abstract. We bridge statistics, convex and discrete geometry, extremal combinatorics, spectral graph theory, and coding theory in our study of complexity of the lasso regularization path under predictor normalization. We show that normalization imposes a new global restriction on the sign patterns encountered by the path: any two full-support patterns must agree in at least two coordinates. This reduces the highest-dimensional complexity of such paths in $\mathbb{R}^p$ to Kleitman's binary diameter problem of calculating $h_{p,r}$, the number of binary p-words that match at least at r places, for $r = 2$, and leads us to explicit finite-dimensional upper bounds. We provide an alternative way to calculate $h_{p,2}$ based on spectral graph theory and the Cvetković upper inertia bound for the independence number for a class of Cayley graphs. We obtain upper and lower bounds for the total complexity of normalized lasso paths. We show that the normalized lasso complexity remains exponential, while exhibiting combinatorial restrictions absent in the unrestricted problem. Furthermore, by using the concept of mutual coherence from statistics, we identify the data for which the lasso paths have $h_{p,r}$ as upper bounds for their highest dimension complexity for any given $r > 2$, while $r = 1$ case is identified with the case of unrestricted data studied by Mairal and Yu. The talk is based on a joint work with Borislav Gajić.

Keywords: lasso regularization paths; Hamming bounds; lasso type code; mutual coherence; realizability