Tropical linear regression and low-rank approximation — a first step in tropical data analysis

- Marianne Akian (INRIA and CMAP, École Polytechnique)
- Stéphane Gaubert (INRIA and CMAP, École Polytechnique)
- **Yang Qi** (INRIA and CMAP, École Polytechnique)
- Omar Saadi (INRIA and CMAP, École Polytechnique)

Keywords: Tropical geometry, mean payoff games, market equilibria, best approximation, auctions, disjunctive optimization.

Abstract: Tropical data arise naturally in many areas, such as control theory, phylogenetic analysis, machine learning, economics, and so on. However, many fundamental problems still deserve further investigations and more powerful tools need to be developed. In this talk, as a first step in tropical data analysis, we would like to introduce two useful models, namely tropical linear regression and tropical low-rank approximation, based on our work [1].

More precisely, given $a = (a_i) \in \mathbb{R}^n$, a *tropical hyperplane* is of the form

$$H_a := \{x \in \mathbb{R}^n \mid \max_{i\in[n]}(a_i + x_i) \text{ achieved at least twice}\}.$$

For a collection $\mathcal{V}$ of finitely many points in $\mathbb{R}^n$, the *tropical linear regression problem* is to find a best tropical hyperplane approximation H_a of $\mathcal{V}$, i.e., a vector $a \in \mathbb{R}^n$ minimizing the one-sided Hausdorff distance

$$\mathrm{dist}_H(\mathcal{V}, H_a) := \max_{v\in\mathcal{V}} \inf_{x\in H_a} \|v - x\| \ .$$

We will establish a strong duality theorem, showing the above distance coincides with the maximal radius of a Hilbert's ball contained in the tropical polyhedron spanned by $\mathcal{V}$. Algorithmically speaking, this regression problem is polynomial-time equivalent to mean payoff game. As an application, we illustrate our results by solving an inverse problem from auction theory.

Another important tool we will study is *tropical low-rank approximation*, i.e., given a subset $\mathcal{X} \subseteq \mathbb{R}^n$, for a fixed point $p \in \mathbb{R}^n$ and a fixed integer r, we would like to find r points $v_1, \dots, v_r \in \mathcal{X}$ and r real numbers $\alpha_1, \dots, \alpha_r$ minimizing

$$\inf_{\alpha_i\in\mathbb{R}, v_i\in\mathcal{X}} \|p - \max_{1\le i\le r}(\alpha_i + v_i)\| \ .$$

We will systematically discuss the relations among different notions of rank in the tropical setting. In particular, when $\mathcal{X}$ is the set of tropical rank-one matrices and $r = 2$, we will reveal a close relation between tropical linear regression and best rank-2 matrix approximation, which provides us an efficient algorithm for finding a best rank-2 matrix approximation for a given matrix.

References:

[1] M. Akian, S. Gaubert, Y. Qi, and O. Saadi. Tropical linear regression and mean payoff games: or, how to measure the distance to equilibria. *arxiv.org/abs/2106.01930*, 2021