ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE

Moments and Positive Polynomials for Machine Learning

Jean B. Lasserre

September 9-10 2019

Université Fédérale Toulouse Midi-Pyrénées

Agenda

Chair members

Context

Objective

Outline

Teaching

Chair members

- **Jean B. Lasserre** (PI, DR émérite au LAAS-CNRS)
- **Victor Magron** (CR au LAAS-CNRS)
- **Milan Korda** (CR au LAAS-CNRS): tentative
  - **Tong Chen**: PhD (starting October 2019)

Context

I. Many *ML* applications such as unsupervised clustering or deep learning, are formulated as large scale **nonconvex** problems. Still most algorithms are **first-order methods** and yet we lack theoretical tools to explain their success. In addition, we often face **average case problems** where data are drawn from distributions, and a better understanding of such situations is required. In particular **Robustness** of such algorithms has become a critical issue to address (e.g. in deep-learning).

Context (continued)

II. The **Christoffel function** is a powerful and simple tool from *Theory of Approximation* and *Orthogonal Polynomials*). In particular it permits to identify the **support of a measure** from the sole knowledge of finitely many **moments**.

However it is largely ignored in the Analytics, Statistics and Optimization research communities whereas we have already demonstrated that it can help solve important problems in data analysis (encode clouds of points, detect outliers, estimate a density from a sample, manifold learning)

Objective

ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE

- New algorithmic insights for **non-convex optimization in AI** via global optimization based on techniques from real algebraic geometry.
  - Certifiability of learning procedures
  - Efficiency and robustness of learning algorithms (à posteriori analysis). For instance:
    - Robustness of neural nets w.r.t. input perturbation
    - Stability of neural-based (data driven) control (e.g., in aeronautics applications)
- The Christoffel function (CF) for data analysis: outlier detection, support inference, data on manifold, manifold learning, etc.
  - Scalability and robustness
  - Asymptotics when the degree increases
  - Sampling strategies (data = clouds of points)

Outline

- General presentation

I. The **Moment-SOS hierarchy** based on *Positivity Certificates* from real algebraic geometry may be a promising tool to handle some ML problems, e.g. robustness certification of some (RELU) deep learning algorithms. However its scalability is still a scientific challenge. Specific properties like **sparsity** and/or **symmetries** must be exploited for practical implementation.

• This is our strategy for **certifying robustness of RELU-based neural nets**.

Outline (continued)

- General presentation (continued)

  II. The **Christoffel function:** Let μ be a measure on Ω (e.g., think of a cloud of 2D-points in $\mathbb{R}^2$). The degree-d Christoffel polynomial is a sum-of-squares polynomial defined by:

$$x \mapsto c_d(x) := \sum_{\alpha \in \mathbb{N}_d^n} P_\alpha(x)^2,$$

  where the P_α's are orthonormal polynomials (w.r.t. μ). It can be obtained from the moment matrix $M_d(\mu)$.

  Key property: The sublevel sets $S_d(\gamma) := \{x : c_d(x) \leq \gamma\}$ capture the geometric shape of the support of μ (again think of the cloud of points), even with relatively small $d = 2, 3, 4$.

$$n^d / c_d(x) \to 0 \text{ exponentially fast when } x \notin \text{support}(\mu)$$

- Some results
  - Scientific results: I. The Moment-SOS Hierarchy has been used in many contexts ranging from optimization (continuous & discrete), statistics (optimal design), signal processing (super-resolution), computational algebra (solving polynomial equations), computational geometry, quantum information.
  - Scientific results: II. We have already proved that the Christoffel function can be useful in data analysis:
    Lasserre J.B., Pauwels E. (2016) Sorting out typicality via the inverse moment matrix SOS polynomial, NIPS 2016.
    Lasserre J.B., Pauwels E. (2019) The empirical Christoffel function with applications in data analysis, Advances Comp. Math. 45, pp. 1439–1468

- Some results (Continued)
  - Related works:

    (i) The **Optimal Power Flow** problem (OPF) in the Management of Energy Networks: a large scale QCQP for which the (adapted) Moment-SOS Hierarchy has provided good results.

    (ii) **Deepcert** is an algorithm that implements the standard SOS-relaxation (first level of the moment-SOS Hierarchy) for certifying robustness of RELU-neural nets.

    (iii) **M. Korda & C. N. Jones.** Stability and performance verification of optimization-based controllers. Automatica, 78 (2017)

- Interaction with other chairs / industrial:

  • With **J. Bolte** (PI), **M. Teboulle** (PI), **S. Gratton** (PI), and **E. Pauwels** on Optimization

  • With **J. M Loubes** (PI), **F. Gamboa** (PI), and **E. Pauwels** on data analysis

Teaching

We propose a high-level course on OPTIMIZATION & APPROXIMATION (in a broad sense) (in view of ML applications)

- Targeted audience: Master 2 and Doctorate students, researchers, and engineers of ML companies
- Format:
  - Courses in standard format
  - High level seminars on specific topics by chair members, researchers with international reputation, engineers of ML oriented companies.
  - Seminars by course participants on a research article to analyze, or topic to discuss.