



# VI Congreso de Jóvenes Investigadores

Real Sociedad Matemática Española

León, Febrero de 2023

---

## Mathematical Optimization and Data Science: a back and forth relationship

Vanesa Guerrero \*

In an era in which the decision-making is often based on the analysis of complex and evolving data, it is crucial to have methodologies and tools which provide helpful support to the decider and that ease the interpretation of the models and results driving the process. Many of the problems arising in this context can be addressed combining the theoretical frameworks of mathematical optimization and data science, thus fostering knowledge exchange in both directions. On the one hand, addressing traditional data analysis modeling approaches from the perspective of mathematical optimization allows to incorporate desirable properties to the analysis. On the other hand, statistical modeling can help mathematical optimization to develop formulations which are more tractable in practice.

In this talk three bidirectional collaborations between mathematical optimization and data science will be discussed. First, the problem of enhancing interpretability in some dimensionality reduction techniques through sparsity [1] and using exogenous information [3] is addressed by means of Mixed Integer Nonlinear Programming (MINLP) approaches. Second, the usefulness of information visualization to improve interpretability of the unknown phenomena described by row data will be discussed. In particular, we focus on the use of global optimization to develop generalizations of the standard multidimensional scaling for which difference of convex functions and nonconvex quadratic binary optimization techniques are combined as solution approaches [2]. Lastly, we address the problem of reformulating general MINLPs using regression splines. The choice of these functional forms allows

us to approximate general non-convex MINLPs by a more tractable subclass of problems which can be efficiently solved by customized algorithms [4].

**Acknowledgments** The researcher is supported by projects PID2019-104901RB-I00, PID2019-110886RB-I00 and IJC2020-045220-I (MCIN/AEI/10.13039/501100011033 and the last also by European Union “NextGenerationEU/PRTR” funds), FQM-329, P18-FR-2369 and US-13811 (Junta de Andalucía) and PGMO program (Fondation Mathématique Jaques Hadamard). This support is gratefully acknowledged.

## References

- [1] E. Carrizosa, V. Guerrero, (2014). *rs-Sparse principal component analysis: A mixed integer nonlinear programming approach with VNS*. Computers & Operations Research **52** 349-354.
- [2] E. Carrizosa, V. Guerrero, D. Romero Morales (2018). *Visualizing data as objects by DC (difference of convex) optimization*. Mathematical Programming **169**(1) 119-140.
- [3] E. Carrizosa, V. Guerrero, D. Romero Morales, A. Satorra (2020). *Enhancing Interpretability in Factor Analysis by Means of Mathematical Optimization*. Multivariate Behavioral Research **55**(5) 748-762.
- [4] C. D’Ambrosio, J. Lee, A. Wächter. (2012) *An algorithmic framework for MINLP with separable non-convexity*. In J. Lee and S. Leyffer, editors, Mixed Integer Nonlinear Programming, volume 154 of The IMA Volumes in Mathematics and its Applications, pages 315-347. Springer, New York.

---

\*Departamento de Estadística, Universidad Carlos III de Madrid, C/ Madrid, 126, 28903 Getafe, Madrid. Email: vanesa.guerrero@uc3m.es