Internship in data science (Feb.-Jul. 2020)
Model learning with rational optimization and constraints

2019/12/22

Reference: http://laurent-duval.eu/job-2020-internship-data-science-rational-models.html

Summary

Title: Model learning with rational optimization and constraints
Tentative start/end: February 1st — July 31st 2020 (**6 months**)
Funding: DataIA (https://dataia.eu/appel-projets/appel-annuel-stagiaires-master-2-ou-equivalent)
Team:

- **Marc Castella** (PI): marc.castella@telecom-sudparis.eu
  Laboratoire SAMOVAR, Telecom SudParis
  9, rue Charles Fourier, 91011 Evry Cedex
- Jean-Christophe Pesquet: jean-christophe.pesquet@centralesupelec.fr
  OPIS, INRIA Paris Saclay, CentraleSupélec (internship location)
  9, rue Joliot Curie, 91190, Gif-sur-Yvette
- Laurent Duval: laurent.duval@ifpen.fr
  ESIEE Paris, IFP Energies nouvelles
  1 avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex

Subject

This subject is in close connection with Arthur Marmin's PhD thesis (Rational models optimized exactly for chemical processes improvement), co-supervised by Marc Castella, Jean-Christophe Pesquet and Laurent Duval [CPP05, CBCP06, CCP10].

The context of the subject is the modeling of experimental data — $\mathbf{y} = A_\theta(\mathbf{x})$ — where A_θ denotes a parametric family of functions. For a long time, polynomial and rational functions have been playing a key role in that respect (calibration, saturation, system modeling, interpolation, signal reconstruction. . .) Their use is classically confined to least-square or least-absolute value regression. However, it is highly desirable in real-world applications to incorporate:

- data-based penalties (positivity, interval bounds),
- model and noise priors, such as power laws or variance stabilizing transforms [Ans48],
- statistical priors, such as robustness or sparsity [Sob81, CCDP20].

Basically, from the modeling perspective, any cost function of practical interest can be approximated as accurately as desired by a polynomial. Nonetheless, optimizing multivariate rational function under polynomial constraints is a difficult problem when no convexity property holds. Recent mathematical breakthroughs have made it possible to solve problems of this kind in an exact manner by building a hierarchy of convex relaxations [Las01, CPM19].

The objective of this internship is to pursue the developments in [MCPD18, MCP18, MCPD20, MCD+20] applied so far to signal restoration. We address here the quest for best model selection in experimental data fitting. We minimize a criterion $\mathcal{J}$ composed of two terms. The first one is a fit measure between the model and recorded measurements $\mathbf{x}$ and $\mathbf{y}$. The second terms are sparsity-promoting (approximating the ℓ_0 pseudo-norm) and interval bound penalizations, weighted by positive parameters λ and μ.

$$\mathcal{J}(\mathbf{x}) = \underbrace{\|\mathbf{y} - A_\theta(\mathbf{x})\|_2^p}_{\text{data fidelity}} + \underbrace{\lambda\Psi(\theta) + \mu\Phi(A_\theta(\mathbf{x}))}_{\text{penalization}} .$$

We choose Φ, Ψ and A_θ as rational functions. The latter will be dealt with by subclasses (e.g. polynomials, homographic functions). The proposed methodology will be evaluated on the many models and experimental data available at IFP Energies nouvelles. A particular attention will be paid to the usability to data practitioners, by embedding the algorithms into a user-friendly interface, with help in choosing parameters p, λ and μ.

Internship information

- Candidate profile: second/third year engineering school and/or master of science with strong skills in statistics, machine learning, optimization. Languages: Python/Matlab; web services is a plus
- Duration: 4 to 6 months
- Location: Center for Visual Computing, CentraleSupélec, Université Paris-Saclay.
- Supervision: Marc Castella (marc.castella@telecom-sudparis.eu) and Jean-Christophe Pesquet.
- Co-supervision: Laurent Duval (laurent.duval@ifpen.fr), IFP Energies nouvelles
- Application: send resume and motivation letter to supervisors; please name documents as **LASTNAME_Firstname-Resume.*** or **LASTNAME_Firstname-Letter.***

References

[Ans48] F. J. Anscombe. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4):246–254, Dec. 1948.

[CBCP06] M. Castella, P. Bianchi, A. Chevreuil, and J.-C. Pesquet. A blind source separation framework for detecting CPM sources mixed by a convolutive MIMO filter. *Signal Process.*, 86(8):1950–1967, 2006.

[CCDP20] Afef Cherni, Emilie Chouzenoux, Laurent Duval, and Jean-Christophe Pesquet. SPOQ ℓ_p-over-ℓ_q regularization for sparse signal recovery applied to mass spectrometry. *PREPRINT*, 2020.

[CCP10] M. Castella, A. Chevreuil, and J.-C. Pesquet. Convolutive mixtures. In P. Comon and C. Jutten, editors, *Handbook of Blind Source Separation. Independent Component Analysis and Applications.* Academic Press, Oxford UK, Burlington USA, 2010.

[CPM19] Marc Castella, Jean-Christophe Pesquet, and Arthur Marmin. Rational optimization for nonlinear reconstruction with approximate ℓ_0 penalization. *IEEE Trans. Signal Process.*, 67(6):1407–1417, 2019.

[CPP05] M. Castella, J.-C. Pesquet, and A.P. Petropulu. A family of frequency- and time-domain contrasts for blind separation of convolutive mixtures of temporally dependent signals. *IEEE Trans. Signal Process.*, 53(1):107–120, Jan. 2005.

[Las01] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, Jan. 2001.

[MCD+20] Arthur Marmin, Marc Castella, Laurent Duval, Bodgan Dumitrescu, and Jean-Christophe Pesquet. Rational optimization: Old problems and new solutions. *PREPRINT*, 2020.

[MCP18] Arthur Marmin, Marc Castella, and Jean-Christophe Pesquet. Sparse signal reconstruction with a sign oracle. In *Proc. Signal Processing with Adaptive Sparse Structured Representations (SPARS) workshop*, 2018.

[MCPD18] Arthur Marmin, Marc Castella, Jean-Christophe Pesquet, and Laurent Duval. Signal reconstruction from sub-sampled and nonlinearly distorted observations. In *Proc. Eur. Sig. Image Proc. Conf.*, 2018.

[MCPD20] Arthur Marmin, Marc Castella, Jean-Christophe Pesquet, and Laurent Duval. Nonlinear sparse signal reconstruction via piecewise rational optimization: application to chromatography. *PREPRINT*, 2020.

[Sob81] E. Sober. The principle of parsimony. *Brit. J. Phil. Sci.*, 32:145–156, 1981. Isn't it ironic, don't you think? (Sober as an author made for parsimony).