Offer #2024-08488

Master Internship - Revisiting PCA with norm-ratio sparsity penalties

Contract type : Internship

Level of qualifications required : Master's or equivalent

Fonction : Internship Research

Level of experience : Recently graduated

About the research centre or Inria department

The Inria Saclay-Île-de-France Research Centre was established in 2008. It has developed as part of the Saclay site in partnership with **Paris-Saclay University** and with the **Institut Polytechnique de Paris** .

The centre has **40 project teams**, 32 of which operate jointly with Paris-Saclay University and the Institut Polytechnique de Paris; Its activities occupy over 600 people, scientists and research and innovation support staff, including 44 different nationalities.

Context

In the context of the ERC MAJORIS, and in collaboration with IFPEN company, the aim of this internship is to investigate the problem of sparse principal component analysis (PCA), with norm-ratio sparsifying penalties.

Subject:

Principal component analysis (PCA) is a workhorse in linear dimensionality reduction [Jol02]. It is widely applied in exploratory data analysis, visualization, data preprocessing).
Principal components are usually linear combinations of all input variables. For high-dimension data, this may involve input variables that contribute very little to the understanding. Finding the few directions in space that explain best observations is desirable. Sparse PCA overcomes this disadvantage by finding linear combinations that contain just a few input variables, by adding sparsity constraints [CR24,ZX18]. One of such is formulated (cf. lasso) with the help of an absolute norm penalty/regularization. In [MBPS10], one designs this matrix factorization problem as:

$$\text{minimize}_{\alpha} || X - D \alpha ||^2_F + \lambda || \alpha ||_{1,1}$$

where: $X = [x_1,...,x_n]$ is the matrix of data vectors; D is a square matrix from a suitable basis set, $||.||_F$ denotes the Frobenius norm; $||.||_{1,1}$ denotes the sum of the magnitude of matrix coefficients, λ is a positive penalty weight.

A penalty such as $||.||_{1,1}$ is 1-homogeneous. This may only weakly emulate the sheer count of non-zero entries of a matrix, that would be scale-invariant or 0-homogeneous.

Recently, the SOOT/SPOQ family of penalties has been developed in our research group, as smooth emulations to the scale-invariant lp/lq norm ratios. The latter had been used for a while, as stopping-criteria, penalties or ``continuous'' sparsity count estimators [HR09]. They have been used successfully for the restoration/deconvolution/source separation of sparse signals [CCDP20,RPD+15].

The goal of the internship is to investigate the resolution of sparse PCA models, by replacing the standard l1 norm by such norm ratios. Convergence analysis of the proposed optimization algorithm, imlementation and validation over public benchmarks will be conducted.

[CCDP20] Afef Cherni, Emilie Chouzenoux, Laurent Duval, and Jean-Christophe Pesquet. SPOQ ℓp-over-ℓq regularization for sparse signal
recovery applied to mass spectrometry. IEEE Trans. Signal Process., 68:6070–6084, 2020.
[CR24] Fan Chen and Karl Rohe. A new basis for sparse principal component analysis. J. Comp. Graph. Stat.), 33(2):421–434, 2024.
[HR09] N. Hurley and S. Rickard. Comparing measures of sparsity. IEEE Trans. Inform. Theory,

55(10):4723–4741, Oct. 2009.

[Jol02] I. T. Jolliffe. Principal component analysis. Springer Series in Statistics, 2nd edition, 2002.

[MBPS10] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online learning for matrix factorization and sparse coding. J. Mach.
Learn. Res., 11:19–60, 2010.

[RPD+15] A. Repetti, M. Q. Pham, L. Duval, E. Chouzenoux, and J.-C. Pesquet. Euclid in a taxicab: Sparse blind deconvolution with smoothed
ℓ_1/ℓ_2 regularization. IEEE Signal Process. Lett., 22(5):539–543, May 2015.

[ZCD23] Paul Zheng, Emilie Chouzenoux, and Laurent Duval. PENDANTSS: PEnalized Norm-ratios Disentangling Additive Noise, Trend
and Sparse Spikes. IEEE Signal Process. Lett., 30:215–219, 2023.

[ZX18] Hui Zou and Lingzhou Xue. A selective overview of sparse principal component analysis. Proc. IEEE, 106(8):1311–1320, August
2018.

Assignment

Missions: The goal of this subject is to:
· investigate potential derivations using SOOT/SPOT penalties,
· implement the algorithmic work-flow in a scientific toolkit (eg scikit-learn),
· benchmark it against competing methods.

Environment: The intern will be supervised by Emilie Chouzenoux (Head of OPIS team, Inria Saclay) and Laurent Duval (Research Engineer, IFPEN, Rueil Malmaison). The intern student will join the Inria Saclay team OPIS (https://opis-inria.eu/). He/she will be located in the Centre de la Vision Numérique, in CentraleSupélec campus, Saclay, France. He/she will enjoy an international and creative environment where research seminars and reading groups take place very often. Informatic material expenses will be covered within the limits of the scale in force.

Organization: The proposed offer is dedicated to internship of Master 1 / Master 2 / Engineering students. The starting/end dates are flexible, with a minimum duration of 4 months.

Main activities

Main activities :

Bibliographical study

Programming in Python environment

Benchmark on public datasets

Scientific meetings

Writing of scientific reports

Skills

Languages : The candidate must be fluent in english and/or french languages.

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Remuneration

Gratification

General Information

- **Theme/Domain :** Optimization, machine learning and statistical methods
  Statistics (Big data) (BAP E)
- **Town/city :** Gif sur Yvette
- **Inria Center :** Centre Inria de Saclay
- **Starting date :** 2025-04-01
- **Duration of contract :** 5 months
- **Deadline to apply :** 2025-03-31

Contacts

- **Inria Team** : OPIS
- **Recruiter** :
  Chouzenoux Emilie / emilie.chouzenoux@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

We seek for a talented candidate in Master 1, Master 2, or Engineering studies, with a solid background in optimization, and signal processing, and a strong motivation for research and innovation. Experience in Python is necessary.

The candidates are requested to send a CV and a motivation letter to apply for this position.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.