Sparse unmixing for active molecular imaging

– Ph.D. proposal in statistical signal/image processing –

Henrique GOUART1, Nicolas DOBIGEON1 and Cyril RUCKEBUSCH2

1University of Toulouse, IRIT/INP-ENSEEIHT, 31000 Toulouse, France
2University of Lille, LASIRE, 59650 Villeneuve d’Ascq
henrique.goulart@irit.fr, nicolas.dobigeon@irit.fr, cyril.ruckebusch@univ-lille.fr

Context

This Ph.D. position focuses on a central problem in spectral imaging, known as spectral unmixing, which consists in jointly decomposing a set of $N$ acquired $F$-valued spectra into $K$ common spectral signatures associated with “pure sources” (e.g., materials or chemical components) weighted by pixel-specific coefficients describing the local mixture of these sources in the $N$ pixels. Mathematically, under a linear mixture assumption, this amounts to recovering two nonnegative matrix factors $W \in \mathbb{R}_{+}^{F \times K}$ and $H \in \mathbb{R}_{+}^{K \times N}$ such that the observation matrix $Y \in \mathbb{R}_{+}^{F \times N}$ decompose as

$$Y \approx WH,$$

for a suitable notion of distance depending on the noise model (e.g., $\ell_2$ for white Gaussian noise). In (1), the matrix $W$ contains the (discretized) spectra of the $K$ sources in its columns, whereas $H$ is composed of coefficients $h_{k,n} \in [0, 1]$ describing the relative proportion of the $k$th source in the $n$th pixel. In other words, each column of $H$ lies on the $(K - 1)$-dimensional simplex. Under these constraints this problem has a natural geometric interpretation: it amounts to finding the convex hull enclosing a set of $N$ points (the columns of $Y$) on the nonnegative orthant of $\mathbb{R}^F$. Often, the number $K$ of elementary components is much smaller than the number of pixels $N$ and the observation dimension $F$, meaning that there is a high level of redundancy in the information conveyed by the whole set of pixels. The nonnegative factorization (1) thus leads to a low-rank approximation of $Y$.

Objectives

The ability of acquiring an image with high spatial resolution is limited by many practical factors such as time and, in the case of biological imaging, the risk of damaging the target specimen by a long exposure to an incoming excitation laser. The IMAGIN project aims at devising a strategy for an efficient estimation of $W$ (that is, the vertices of the sought convex hull) without ever acquiring the “full matrix” $Y$, but performing instead a targeted sparse sampling. This amounts to selecting the most informative pixels (according to a precise criterion to be defined), and then solving the sparse unmixing problem on that reduced set of pixels, which should carry enough information on $W$. To this end, we may rely on the statistical analysis of auxiliary low-(spatial)-resolution images $\tilde{Y} \in \mathbb{R}_{+}^{F \times n}$, with $n \ll N$, to guide the selection process.

In this context, the Ph.D. student shall be mainly focused on:

1. The formulation of an appropriate optimization problem to address the unmixing problem (on a reduced number of pixels selected according to a given strategy) and devise corresponding (efficient) optimization algorithm(s). In particular, the algorithm(s) should be able to process the acquired observations in real time. Relevant regularizations may be included into the objective function of this optimization problem, such as a $\ell_{1,2}$-based group-sparsity or total variation applied to $H$ and a minimum volume constraint imposed on $W$. The resulting problem could be of the form

$$\min_{W \geq 0, H \geq 0} D(\tilde{Y} \mid WHM) + \lambda_w R_w(W) + \lambda_h R_h(H),$$

where $D(\cdot \mid \cdot)$ is a suitable distance or divergence (e.g., a squared $\ell_2$ loss), $M \in \{0, 1\}^{N \times n}$ is a binary mask selecting the acquired pixels such that $\tilde{Y} = YM$ and the functionals $R_w(\cdot)$ and $R_h(\cdot)$ act as regularizers. Its solution will call for the
use of possibly nonsmooth optimization methods based, e.g., on proximal algorithms. Then, the results produced by this scheme will be compared to those obtained by state-of-the-art methods from the literature. Finally, the theoretical properties of the proposed formulation and algorithm (notably convergence) will be studied.

2. A theoretical study which will seek precise mathematical results (possibly of a probabilistic nature) characterizing the unmixing performance that one can achieve by relying on the developed algorithms (for pixel selection and unmixing).

**Scientific environment**

The Ph.D. candidate will be a member of the IMAGIN project, funded by ANR. The two main teams involved in this thesis are the Signal & Communications (SC) group from IRIT (CNRS and Toulouse INP) and the Dynamics, Nanoscopy & Chemometrics (DyNaChem) group from LASIRE (CNRS and University of Lille). The SC group brings its expertise in the development of state-of-the-art statistical signal & image processing methods, in particular for multivalued images for various applications (medical imaging, remote sensing, microscopy). The DyNaChem group is interested in micro- and nano-imaging of photoactive bio-systems, with a particular focus on hyperspectral and super-resolved nanometer-scale imaging of ultrafast processes, and on the development of new instrumentation and methodologies for the analysis of these hyperspectral super-resolved data.

The Ph.D. student will therefore benefit from a scientifically rich environment and will be able to acquire a solid background on the most recent results and advances in statistical signal & image processing for molecular imaging. He/she will be mainly co-advised by

- **Henrique Goulart**, Assistant Professor within the SC group at IRIT laboratory (UMR CNRS 5505, Toulouse);
- **Nicolas Dobigeon**, Professor within the SC group at IRIT laboratory (UMR CNRS 5505, Toulouse) and AI Research Chair at the Artificial and Natural Intelligence Toulouse Institute (ANITI);

in collaboration with

- **Cyril Ruckebusch**, Professor within the DyNaChem group at LASIRE laboratory (UMR CNRS 8516, Villeneuve d’Ascq).

The student’s workplace will be the INP-ENSEEIHT campus of IRIT (where the SC group is located), in a lively neighbourhood of the Toulouse city center. Nevertheless, he/she may also have short-period visits in the other groups (notably the DyNaChem group in Lille) involved in the IMAGIN project.

**Funding**

This position will be fully funded by the ANR project IMAGIN.

**Period**

The Ph.D. shall start in 2022, with a duration of 3 years. The precise starting date can be adjusted according to the availability of the selected candidate.

**Keywords**

Nonnegative matrix factorization, statistical signal/image processing, multi-band imaging.

**Profile & requirements**

Graduate students with major in applied mathematics, statistics, computer science or electrical engineering. The knowledge needed for this work includes a strong background in **signal & image processing, statistics, linear algebra** and **optimization**. Knowledge of machine learning and/or random matrix theory and/or high-dimensional statistics is a plus. Experience and/or interests in microscopy will be appreciated.
**Contact & application procedure**

Applicants are also invited to send (as pdf files)

- a detailed curriculum,
- official transcripts from each institution you have attended (in French or English).

to the co-advisors

- Henrique Goulart, henrique.goulart@irit.fr
- Nicolas Dobigeon, nicolas.dobigeon@irit.fr
- Cyril Ruckebusch, cyril.ruckebusch@univ-lille.fr

You will be contacted if your profile meets the expectations. Review of applications will be closed when the position is filled.

**References**

