Signals, images and task-driven representation and dictionary learning

November 24, 2016

Keywords: dictionary learning, supervised dictionary learning, transform learning, sparse recovery, classification, feature engineering

1 Context

We live in an era of the data abundance: tremendous volume of data are available in regards to virtually everything. It includes biomedical signals, genomic data, industrial systems signals, weather and environment related data, astronomical surveys data, satellite earth images, and a plethora of human activities generated data. Also, spectacular breakthroughs in data analysis techniques and methodology have occured in signal processing and machine learning communities over the two past decades. One can get interesting insights on the convergence between these two areas from [Singh *et al.* 2015]. At the intersection of these two fields, *Dictionary Learning* (DL) has come to be a major topic of research on itself. It encompasses methods and algorithms which, for a given type of signal, aim at deriving a set of cardinal features which enables one to concisely describe signals of this type.

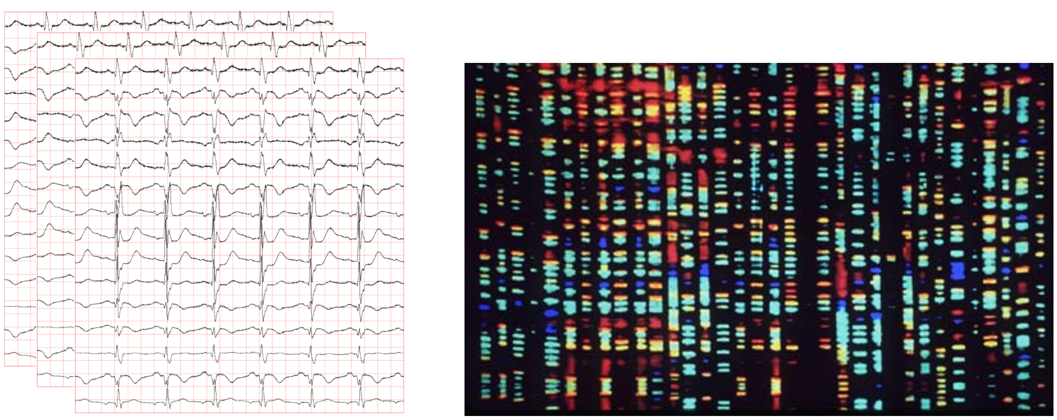

Figure 1: ECG signals on the left; genes expression levels sequences on the right.

Precisely, given a set of vectors $(\mathbf{x}_j)_{1\leq j\leq n}$ in $\mathbf{R}^p$ that might represent ECG signals or genes sequences as illustrated in Fig.1, DL techniques compute a set vectors $(\mathbf{d}_i)_{1\leq i\leq N}$ and a set of coefficients $(\alpha_{ij})_{\substack{1\leq i\leq N\\ 1\leq j\leq n}}$ so that

$$\mathbf{x}_j \approx \sum_{i=1}^{N} \alpha_{ij}\mathbf{d}_i,\ j = 1\cdots N \tag{1}$$

with a constraint of sparsity on the vectors $\boldsymbol{\alpha}_j = [\alpha_{1j}, \cdots, \alpha_{Nj}]^T$. As for the terminology,

- the matrix $\mathbf{D} = [\mathbf{d}_1, \cdots, \mathbf{d}_N]$ is the dictionary,

- its columns vectors ($\mathbf{d}_i$) are the atoms,
- and the vectors ($\boldsymbol{\alpha}_j$) are termed the codes as they represent each of the training set samples.

Hence, the learnt dictionary is specifically adapted for describing signals that are structurally similar to the training samples. The benefit of such dictionaries in sparsity-driven signal recovery has been shown in several applications (see for example [Elad & Aharon 2006, Ravishankar & Bresler 2011, Beckouche *et al.* 2013]).

10 years after the seminal contribution in [Aharon *et al.* 2006], major advances have been realized in DL regarding different important aspects including:

- discriminatory power: in numerous applications in machine learning, data are labeled and/or sampled from some regular manifold; thus it is suitable, for classification or interpolation tasks for instance, that the learned codes allow for a better discrimination of the data samples with respect to labels information or manifold's structure; the growing field of *supervised dictionary learning* precisely consists of DL methods that account for these additional information (a recent review can be found in [Gangeh *et al.* 2015]);
- geometric invariances: a considerable effort has been dedicated into to building DL methods that use the fact that similar structures can appear in the training set with different offsets, orientations (for bidimensional data and above) and scales; such methods rely either on training data registration [Beckouche *et al.* 2013, Yüzügüler *et al.* 2014] or on learning structured dictionaries [Barthélemy *et al.* 2012, Ophir *et al.* 2011];
- scalability: generally speaking, the main disavantage of learned dictionaries with respect to analytical dictionaries such as wavelets, DCT or curvelets related dictionaries is that the latter can be used through fast implicit transforms while the former are unstructured and potentially large matrices that one has to manipulate explicitly, which limits the practioner to learning reasonably small dictionaries on low dimensional data; this issue has been addressed in several works either from the data dimensionality angle[Xiang *et al.* 2011, Sulam *et al.* 2016] or from the computational efficiency of the dictionary perspective [Magoarou & Gribonval 2014, Chabiron *et al.* 2014].

Such rich perspectives make DL an exciting and promising field of research, as far as data analysis is concerned; this Ph.D. thesis is precisely positioned at the confluent of these three sub-themes.

2 Research project

The goal of this thesis is to propose a dictionary and representation learning methodology that yields

- a good tradeoff between data description and discrimination
- and easily interpretable dictionary and codes.

As for the first point, the aim will be to take a step further in supervised dictionary learning by introducing geometric invariances. The possibility of learning a "fast transform like" dictionary in this context will be examined and ideas from the transform learning literature will be considered (see for instance [Pfister & Bresler 2015] and the references therein).

Supervised DL methods in general have important hyperparameters which choice is not straightforward and which are costly to tune. Proposing a rigorous framework for setting learning hyperparameters constitutes an important aspect of this thesis and will rely on existing solutions proposed in the machine learning and signal processing communities (see for example [Pedregosa 2016, Chaux & Blanc-Féraud 2012, Dang & Chainais 2015, Shervashidze & Bach 2015]).

An interesting line of investigation would be to evaluate the feasbility and benefit of adopting a nonlinear descriptive model in place of the model in Eq.1, in the same spirit as kernel-based DL method for instance (see [Golts & Elad 2016] and the references therein).

The second point is application oriented. Indeed, it is suitable to produce physically meaningful or interpretable atoms and codes in DL, especially when analyzing physiological signals. This notion of interpretability is both ubiquituous and ill-defined in machine learning as pointed out in [Lipton *et al.* 2016]. The reflection here will be on whether it is possible to define a quantitative interpretability criterion and use it in the learning procedure. Such criterion might be generic as the sharpness index introduced in [Leclaire & Moisan 2013] or the typical criteria used for points of interest detection in images (see for instance [Lowe 2004]). Alternately, it might rely on application specific quantities such as the Mel Frequency Cepstral Coefficients[Muda *et al.* 2010] in voice recognition.

A broad spectrum of applications can be considered including common classification tasks (handwritten digits, natural images...), biophysical signals analysis, hyperspectral earth images analysis, galaxies images or spectra classification, astronomical radio sources identification etc.

3 Organisation of the thesis work

The thesis will be organised in three parts:

1. **Supervised DL**
   - State-of-the-art: Supervised DL, DL with geometric invariance, fast transform like DL
   - Methodology: supervised DL+invariances
   - Methodology: supervised DL+invariances+computational efficiency
2. **Hyperparameters**
   - State-of-the-art: hyperparameters selection and optimization in machine learning and signal processing
   - Methodology: supervised DL+invariances+computational efficiency+automated parameters selection
3. **Interpretability**

The programming will be done in **Python**.

4 The candidate

- Education: Master 2 or equivalent in Signal Processing or Applied Mathematics or Data Science
- Skills: knowledge and practice of Optimization numerical schemes is required; an experience in Python programming is a plus
- Personal: autonomous, quick learner

5 Why apply?

5.1 The environment

This Ph.D. thesis will proceed in co-tutoring between the LADIS and the CosmoStat laboratories at CEA Saclay's Technological and Fundamental Research Directions respectively. These young and dynamic laboratories have a recognized expertise

- in data analysis methodology and industrial applications for the LADIS
- and in signal processing methodology and applications to astronomy for the CosmoStat.

This double culture will provide the Ph.D. candidate a wealthy and structuring learning ground and a high visibilty as for professional openings.

5.2 Learnings

The Ph.D. candidate will master modern and advanced data analysis tools, at the intersection of signal processing and machine learning. This thesis will provide the future Doctor a solid base for an orientation toward both academical research and research engineering in data science.

- *contacts*: Jean-Luc Starck (Ph.D. supervisor), jean-luc.starck@cea.fr; Ngolè Fred (tutor), fred-maurice.ngole-mboula@cea.fr;
- *Lab*: CEA/LADIS-CosmoStat in Saclay.

References

[Aharon *et al.* 2006] Michal Aharon, Michael Elad et Alfred Bruckstein. *K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation.* IEEE TRANSACTIONS ON SIGNAL PROCESSING, vol. 54, no. 11, page 4311, 2006.

[Barthélemy *et al.* 2012] Quentin Barthélemy, Anthony Larue, Aurélien Mayoue, David Mercier et Jérôme I Mars. *Shift & 2D rotation invariant sparse coding for multivariate signals.* IEEE Transactions on Signal Processing, vol. 60, no. 4, pages 1597–1611, 2012.

[Beckouche *et al.* 2013] Simon Beckouche, Jean-Luc Starck et Jalal Fadili. *Astronomical image denoising using dictionary learning.* Astronomy & Astrophysics, vol. 556, page A132, 2013.

[Chabiron *et al.* 2014] Olivier Chabiron, François Malgouyres, Jean-Yves Tourneret et Nicolas Dobigeon. *Learning a fast transform with a dictionary.* In international Traveling Workshop on Interactions between Sparse models and Technology-iTWIST'2014, pages pp–10, 2014.

[Chaux & Blanc-Féraud 2012] Caroline Chaux et Laure Blanc-Féraud. *Wavelet-based hyperparameter estimation for solving inverse problems.* 2012.

[Dang & Chainais 2015] Hong Phuong Dang et Pierre Chainais. *A Bayesian non parametric approach to learn dictionaries with adapted numbers of atoms.* In 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), pages 1–6. IEEE, 2015.

[Elad & Aharon 2006] Michael Elad et Michal Aharon. *Image denoising via learned dictionaries and sparse representation.* In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), volume 1, pages 895–900. IEEE, 2006.

[Gangeh *et al.* 2015] Mehrdad J Gangeh, Ahmed K Farahat, Ali Ghodsi et Mohamed S Kamel. *Supervised dictionary learning and sparse representation-a review.* arXiv preprint arXiv:1502.05928, 2015.

[Golts & Elad 2016] Alona Golts et Michael Elad. *Linearized kernel dictionary learning.* IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 4, pages 726–739, 2016.

[Leclaire & Moisan 2013] Arthur Leclaire et Lionel Moisan. *Blind deblurring using a simplified sharpness index.* In International Conference on Scale Space and Variational Methods in Computer Vision, pages 86–97. Springer, 2013.

[Lipton *et al.* 2016] Zachary C Lipton*et al.* *The Mythos of Model Interpretability.* IEEE Spectrum, 2016.

[Lowe 2004] David G Lowe. *Distinctive image features from scale-invariant keypoints.* International journal of computer vision, vol. 60, no. 2, pages 91–110, 2004.

[Magoarou & Gribonval 2014] Luc Le Magoarou et Rémi Gribonval. *Learning computationally efficient dictionaries and their implementation as fast transforms.* arXiv preprint arXiv:1406.5388, 2014.

[Muda *et al.* 2010] Lindasalwa Muda, Mumtaj Begam et I Elamvazuthi. *Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques.* arXiv preprint arXiv:1003.4083, 2010.

[Ophir *et al.* 2011] Boaz Ophir, Michael Lustig et Michael Elad. *Multi-scale dictionary learning using wavelets.* IEEE Journal of Selected Topics in Signal Processing, vol. 5, no. 5, pages 1014–1024, 2011.

[Pedregosa 2016] Fabian Pedregosa. *Hyperparameter optimization with approximate gradient.* arXiv preprint arXiv:1602.02355, 2016.

[Pfister & Bresler 2015] Luke Pfister et Yoram Bresler. *Learning sparsifying filter banks.* In SPIE Optical Engineering+ Applications, pages 959703–959703. International Society for Optics and Photonics, 2015.

[Ravishankar & Bresler 2011] Saiprasad Ravishankar et Yoram Bresler. *MR image reconstruction from highly undersampled k-space data by dictionary learning.* IEEE transactions on medical imaging, vol. 30, no. 5, pages 1028–1041, 2011.

[Shervashidze & Bach 2015] Nino Shervashidze et Francis Bach. *Learning the structure for structured sparsity.* IEEE Transactions on Signal Processing, vol. 63, no. 18, pages 4894–4902, 2015.

[Singh *et al.* 2015] Richa Singh, Mayank Vatsa, Angshul Majumdar et Ajay Kumar. Machine intelligence and signal processing. Springer Publishing Company, Incorporated, 1st édition, 2015.

[Sulam *et al.* 2016] Jeremias Sulam, Boaz Ophir, Michael Zibulevsky et Michael Elad. *Trainlets: Dictionary Learning in High Dimensions.* IEEE Transactions on Signal Processing, vol. 64, no. 12, pages 3180–3193, 2016.

[Xiang *et al.* 2011] Zhen J Xiang, Hao Xu et Peter J Ramadge. *Learning sparse representations of high dimensional data on large scale dictionaries.* In Advances in Neural Information Processing Systems, pages 900–908, 2011.

[Yüzügüler *et al.* 2014] A. C. Yüzügüler, E. Vural et P. Frossard. *Transformation-invariant dictionary learning for classification with 1-Sparse representations.* In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 3562–3566, May 2014.