Sparse representations in continuous dictionaries

Cédric Herzet

inventeurs du monde numérique

Main goal: fit a parametric model to data

$$\boldsymbol{y} = \sum_{i=1}^{k} x_i \, \boldsymbol{a}(\theta_i) + \text{noise} \qquad\qquad x_i \in \mathbb{R},\ \theta_i \in \Theta \subseteq \mathbb{R}^d$$

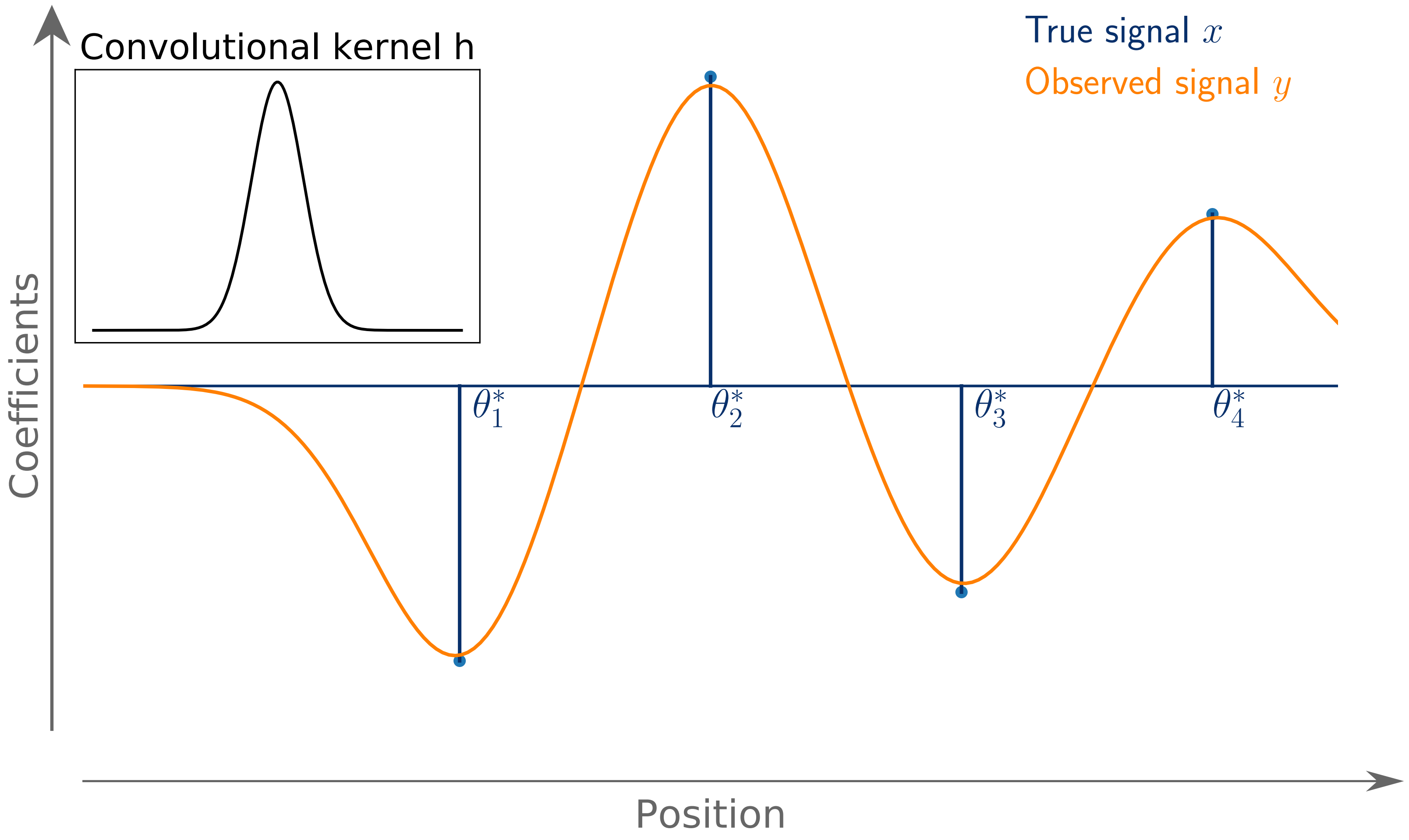
I know: $\boldsymbol{y}$ *(observations)*

$\boldsymbol{a} : \Theta \to \mathcal{H}$ *(functional form)*

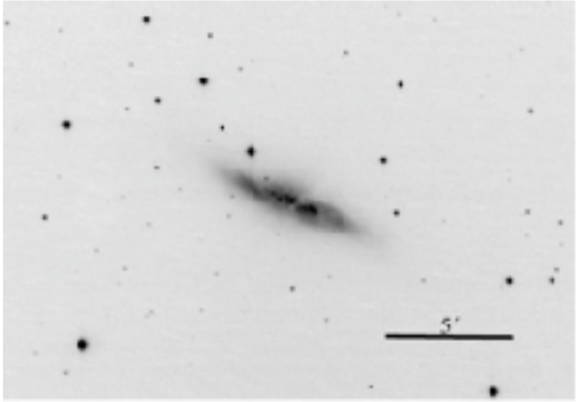
I want: $\{x_i\}_{i=1}^{k}, \{\theta_i\}_{i=1}^{k}$ *(component features)*

Example:
Gaussian deconvolution problem

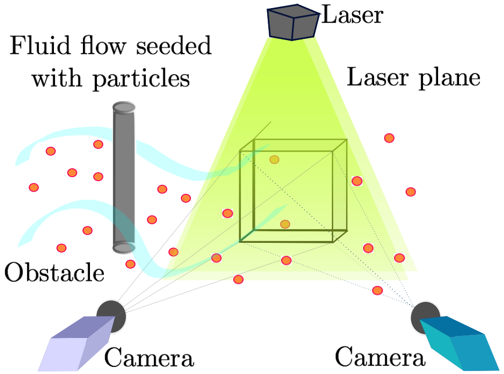
(Courtesy C. Elvira)



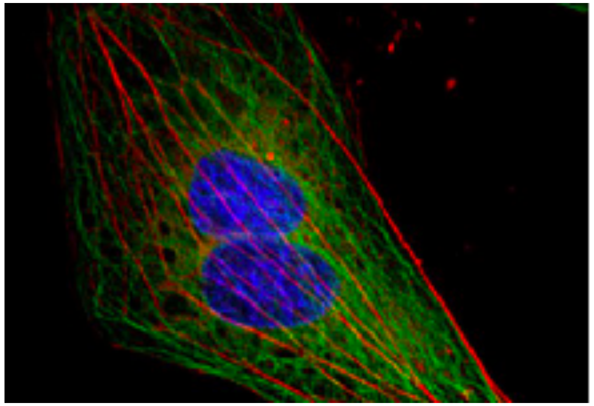
Other 2D/3D examples...



astrophysics



fluid mechanics



microscopy

The brute-force approach is likely to fail

$$\min_{\{x_i\}_{i=1}^{k},\{\theta_i\}_{i=1}^{k}} \underbrace{\left\| \boldsymbol{y} - \sum_{i=1}^{k} x_i \, \boldsymbol{a}(\theta_i) \right\|_{\mathcal{H}}^{2}}$$

this function is likely to have many local minima

Optimization problem over a *(d+1)k* dimensional space

Are the features identifiable from $\boldsymbol{y}$?

Can « sparse representations » provide an answer to these questions/problems ?

The sparse model

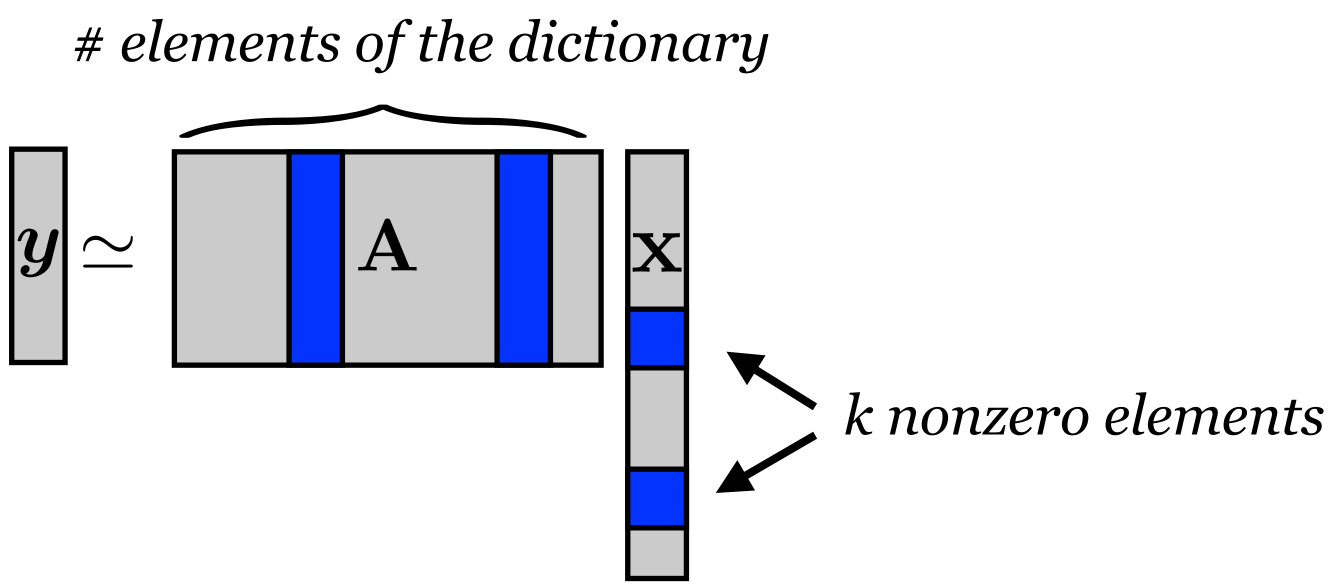
$$\boldsymbol{y} \simeq \sum_{i=1}^{k} x_i \, \boldsymbol{a}_i \qquad\qquad x_i \in \mathbb{R}, \; \boldsymbol{a}_i \in \mathcal{A}$$

$\downarrow$

« dictionary » of elementary signals

The « sparse » hypothesis: $k \ll \dim(\boldsymbol{y})$

Finite dictionaries allow for a matrix formulation



Mathematical formulation of the sparse representation problem

$$(P_0): \qquad \hat{\mathbf{x}} = \arg\min \|\mathbf{z}\|_0 \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{z}$$

(P_o) is NP-hard

«Convex relaxation» enables a tractable implementation

$$(P_1): \qquad \hat{\mathbf{x}} = \arg\min \|\mathbf{z}\|_1 \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{z}$$

 *(P_1)* is a convex problem!

Noise-aware modification of *(P_1)*: the « LASSO » problem

$$\hat{\mathbf{x}} \in \arg\min_{\mathbf{z}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_{\mathcal{H}}^2 + \lambda \|\mathbf{z}\|_1 \right\}$$

Sparse representations
in continuous dictionaries

Continuous dictionaries contain an infinite number of elements

$$\boldsymbol{y} \simeq \sum_{i=1}^{k} x_i \, \boldsymbol{a}_i \qquad\qquad x_i \in \mathbb{R},\ \boldsymbol{a}_i \in \mathcal{A}$$

where $\mathcal{A} = \{\boldsymbol{a}(\theta) : \theta \in \Theta\}$

NB: $\mathrm{card}(\mathcal{A}) = \mathrm{card}(\Theta)$

$\mathcal{A}$ may thus contain an infinite number of elements !

The « BLASSO » problem: a continuous counterpart to LASSO

$$\hat{\mu} \in \arg\min_{\mu \in \mathcal{M}(\Theta)} \left\{ \frac{1}{2} \|\boldsymbol{y} - A\mu\|_{\mathcal{H}}^2 + \lambda \|\mu\|_{TV} \right\}$$

where $\mathcal{M}(\Theta)$: set of Radon measures on Θ

$A : \mathcal{M}(\Theta) \to \mathcal{H}$ linear operator

$\|\mu\|_{TV}$: « total variation » norm

Pro & con of the « BLASSO » problem

The BLASSO problem is convex

Convex optimality conditions can be used to derive recovery results

The optimisation variable is a measure

Optimization over an infinite dimensional space can be tedious

Some selected recovery results of the literature

Recovery in the noiseless setting

[de Castro *et al*, 2012]: sum of some positive functions (Markov system)

[Candes *et al*, 2015]: sum of cosines under separability condition

[Elvira *et al*, 2019]: signed sum of completely monotone functions

Stability in the noisy setting

[Denoyelle *et al*, 2018]: stability of the number of components in high SNR regime

Under some conditions, the solution is a discrete measure

$$\hat{\mu} = \sum_{i=1}^{k} \hat{x}_i \delta_{\hat{\theta}_i}$$

We can focus on discrete measures in the resolution of BLASSO

A practical implementation: the « sliding » Frank-Wolfe algorithm

Init: $\hat{\mu} = 0$

Repeat: 1) Set $\hat{\mu} = \hat{\mu} + x^\star \delta_{\theta^\star}$

with $$\theta^\star = \arg\max_{\theta \in \Theta} |\langle \boldsymbol{a}(\theta), \boldsymbol{y} - A\hat{\mu} \rangle_{\mathcal{H}}|$$
$$x^\star = \text{“smart choice”}$$

NB: at this point, $\hat{\mu}$ has the form $\hat{\mu} = \sum_{i=1}^{k} x_i \delta_{\theta_i}$

2) Change $\hat{\mu}$ as you want as long as the cost decreases

The search of the new spike to add may be computationally demanding

$$\theta^\star = \arg\max_{\theta \in \Theta} |\langle \boldsymbol{a}(\theta), \boldsymbol{y} - A\hat{\mu} \rangle_{\mathcal{H}}|$$

d-dimensional « linear » optimization problem

We propose efficient procedures to prune the optimization domain Θ *[Herzet et al, 18], [Dorffer et al, 18]*

Conclusions

- Sparse representation in continuous dictionaries can bring new answers to pretty old problems
- The battle for the design of sparse representation algorithms in continuous dictionaries is open
- We are interested in applications !