Dictionary Learning Based on Sparse Distribution Tomography for 3D Deconvolution Microscopy

Emmanuel Soubies

Workshop Dictionary Learning on Manifolds - September 6th 2017

Biomedical Imaging Group, EPFL

Dictionary Learning Based on Sparse Distribution Tomography.
Pedram Pad, Farnood Salehi, Elisa Celis, Patrick Thiran and Michael Unser.
Proceedings of the 34th International Conference on Machine Learning 2017.

European Research Council

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline of the talk

1. Problem formulation and prior work

2. The sparse distribution tomography (SparsDT) algorithm [Pad et al 17]

3. Application to 3D Deconvolution Microscopy (Ongoing Work)

Problem formulation and prior work

Dictionary learning

Objective

Given a set of $K \in \mathbb{N}$ observations $\mathbf{y}_1, \ldots, \mathbf{y}_K$ in $\mathbb{R}^M$, we aim at finding a **mixture matrix (dictionary)** $\mathbf{A} \in \mathbb{R}^{M \times N}$ verifying the model

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i \quad \forall i \in [1 \ldots K],$$

where the **unknown** signals $\mathbf{x}_1, \ldots, \mathbf{x}_K$ in $\mathbb{R}^N$ fulfill some **hypotheses**.

Dictionary learning

Common hypotheses on the signals $\mathbf{x}_i$

- Statistical prior: the $\mathbf{x}_i$ are **realizations** of a random vector $\mathbf{X}$,

  $\rightarrow$ *e.g.* independent component analysis (ICA) [Hyvärinen et al 04],

  $\rightarrow$ generally $\mathbf{A}$ restricted to be square.

- Deterministic prior: the $\mathbf{x}_i$ are **sparse** ($\|\mathbf{x}_i\|_0 \ll N$)

  $\rightarrow$ *e.g.* sparse component analysis (SCA) [Zou et al 06, Gribonval et al 06].

SparsDT [Pad et al 17] $\longrightarrow$ best of both worlds.

The sparse distribution tomography (SparsDT) algorithm [Pad et al 17]

Fourier-domain techniques for classical probability theory

Probability density function (pdf) and characteristic function (cf)

A random vector $\mathbf{X}$ on $\mathbb{R}^N$ is fully characterized by

- its **probability density function** $p_{\mathbf{X}} : \mathbb{R}^N \mapsto \mathbb{R}_{\geqslant 0}$,
- or its **characteristic function** $\hat{p}_{\mathbf{X}} : \mathbb{R}^N \mapsto \mathbb{C}$.

both being linked through the **Fourier Transform**

$$\hat{p}_{\mathbf{X}}(\omega) = \mathbb{E}\{e^{j\langle \omega, \mathbf{X}\rangle}\} = \int_{\mathbb{R}^N} p_{\mathbf{X}}(\mathbf{x}) e^{j\langle \omega, \mathbf{x}\rangle} \mathrm{d}\mathbf{x} = \overline{\mathcal{F}\{p_{\mathbf{X}}\}(\omega)}$$

Useful properties

- **Joint cf of independant variables**: Let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ be a pair of two independant variables. Then, with $\omega = (\omega_1, \omega_2)$,

$$\hat{p}_{(\mathbf{X}_1,\mathbf{X}_2)}(\omega) = \hat{p}_{\mathbf{X}_1}(\omega_1)\hat{p}_{\mathbf{X}_2}(\omega_2).$$

- **Affine Transformation**: Let $\mathbf{H} \in \mathbb{R}^{M \times N}$ and $\mathbf{b} \in \mathbb{R}^M$. Then, the cf of $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{b}$ is

$$\hat{p}_{\mathbf{Y}}(\omega) = \hat{p}_{\mathbf{X}}(\mathbf{H}^T\omega)e^{j\mathbf{b}^T\omega}.$$

- **Sum of independant random vectors**: The cf of $\mathbf{Y} = a_1\mathbf{X}_1 + a_2\mathbf{X}_2$, for two reals a_1 and a_2, is

$$\hat{p}_{\mathbf{Y}}(\omega) = \hat{p}_{\mathbf{X}_1}(a_1\omega)\hat{p}_{\mathbf{X}_2}(a_2\omega)$$

Symmetric-α-Stable random variables

Definition ([Nikias and Shao 95])

A random variable X is **symmetric-α-stable** (SαS) if its **characteristic function** is of the form

$$\hat{p}_X(\omega) = \exp(-\gamma|\omega|^\alpha)$$

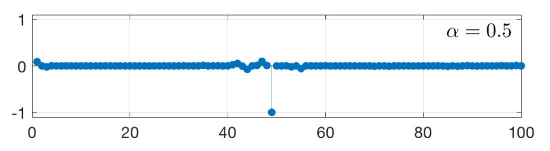
where

- $\gamma > 0$ is the **dispersion** parameter,
- $\alpha \in (0, 2]$ is the **stability** parameter.

Symmetric-α-Stable random variables

α-stable distributions and sparsity

- When $\alpha < 2 \longrightarrow$ **heavy-tailed** distribution with **unbounded variance**,
- An **i.i.d. sequence** of α-stable random variables generate a **sparse** signal,
- The **smaller** the α the **sparser** the signal,



- For $\alpha = 2 \longrightarrow$ Gaussian distribution (non-sparse)

Symmetric-α-Stable random variables

Some fundamental properties

- **Stability under linear combination** [Nikias and Shao 95] Let
  - X_k, $k \in [1 \dots K]$, be i.i.d. SαS random variables with dispersion $\gamma > 0$ and stability $\alpha \in (0, 2]$.
  - $\mathbf{a} = [a_1 \dots a_K]$, be a vector of real numbers.

  Then

  $$\bar{X} = \sum_{k=1}^{K} a_k X_k$$

  SαS with dispersion $\gamma \|\mathbf{a}\|_\alpha^\alpha$ and stability α.
- **Generalized central limit theorem** [Meerschaert and Scheffler 01] Let
  - X_k, $k \in [1 \dots K]$, be i.i.d. copies of an heavy-tailed (decreasing as $|x|^{-\alpha-1}$) random variables X,
  - $\mathbf{a} = [a_1 \dots a_K]$, be a vector of real numbers.

  Then, when K become large,

  $$\bar{X} = \sum_{k=1}^{K} a_k X_k$$

  is well approximated by an α-stable random variable.

Main idea of the method

The model

We consider

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

where $\mathbf{x}$ is a vector with **SαS i.i.d. entries** (statistical prior) which make it **sparse** (deterministic prior).

Objective

Given $K \in \mathbb{N}$ realizations of $\mathbf{y}$ (K large), namely $\mathbf{y}_1 \ldots \mathbf{y}_K$, identify $\mathbf{A} \in \mathbb{R}^{M \times N}$.

Distribution of y

Using the cf of a random vector under a linear transformation, we get

$$\hat{p}_{\mathbf{y}}(\omega) = \exp(-\gamma \|\mathbf{A}^T \omega\|_\alpha^\alpha),$$

Main idea of the method

Let $\mathbf{u} \in \mathbb{R}^M$. Then from the stability under linear combination of SαS random variables,

$$\mathbf{u}^T\mathbf{y} = \mathbf{u}^T\mathbf{A}\mathbf{x} = (\mathbf{u}^T\mathbf{A}_{.1})\mathbf{x}_1 + \cdots + (\mathbf{u}^T\mathbf{A}_{.N})\mathbf{x}_N$$

is an **SαS random variable** with characteristic function

$$\hat{p}_{\mathbf{u}^T\mathbf{y}}(\omega) = \exp(-\underbrace{\gamma\|\mathbf{A}^T\mathbf{u}\|_\alpha^\alpha}_{\gamma(\mathbf{u})}|\omega|^\alpha),$$

Knowing the **dispersion** $\gamma(\mathbf{u})$ of $\mathbf{u}^T\mathbf{y}$ **for all** $\mathbf{u} \in \mathbb{R}^M$,
is $\mathbf{A}$ **identifiable** ?

Main idea of the method

Theorem ([Pad et al 17])

Let $\mathbf{A}$ *be an* $M \times N$ *matrix where columns are* **pairwise linearly independent**. *If* $\alpha \in (0, 2)$ *and* $\mathbf{B}$ *is an* $M \times N$ *matrix such that*

$$\|\mathbf{A}^T \mathbf{u}\|_\alpha^\alpha = \|\mathbf{B}^T \mathbf{u}\|_\alpha^\alpha$$

for all $\mathbf{u} \in \mathbb{R}^M$, *then* $\mathbf{B}$ *is* **equal** *to* $\mathbf{A}$ *up to negation and permutation of its columns.*

Given $\gamma(\mathbf{u}_l)$ for an appropriate finite set of vectors $[\mathbf{u}_1 \dots \mathbf{u}_L]$ (with $L \geq MN$), one can **identify** $\mathbf{A}$ by solving the set of **non-linear equations**

$$\gamma \|\mathbf{A}^T \mathbf{u}_1\|_\alpha^\alpha = \gamma(\mathbf{u}_1)$$

$$\vdots$$

$$\gamma \|\mathbf{A}^T \mathbf{u}_L\|_\alpha^\alpha = \gamma(\mathbf{u}_L)$$

Main idea of the method

- **Non-exact** estimation of the $\gamma(\mathbf{u}_l)$,
- Can lead to the **non-existence** of a solution to the system of equations,
- **Minimization** of a loss function,

$$\mathcal{F}(\mathbf{A}) = \frac{1}{\alpha L} \sum_{l=1}^{L} |\log(\|\mathbf{A}^T \mathbf{u}_l\|_\alpha^\alpha) - \log(\gamma(\mathbf{u}_l))|$$

Learning algorithm (SparsDT)

Data: $N \in \mathbb{N}$, $L \in \mathbb{N}$, $\mathbf{y}_1 \dots \mathbf{y}_K$ in $\mathbb{R}^M$
Result: $\hat{\mathbf{A}} \in \mathbb{R}^{M \times N}$
Initialize $\mathbf{A}^0$ such that $\mathbf{A}_{.i} \sim \mathcal{N}(0, \mathbf{I}_{M \times M})$;
while *not converged* **do**

Generate $\mathbf{u}_1 \dots \mathbf{u}_L \sim \mathcal{N}(0, \mathbf{I}_{M \times M})$;
Estimate $\hat{\alpha}$, and $\hat{\gamma}(\mathbf{u}_l)$ for all $l \in [1 \dots L]$, from the data $\mathbf{y}_k$ and the $\mathbf{u}_l$;
Solve

$$\mathbf{A}^{k+1} = \arg\min_{\mathbf{A}} \mathcal{F}(\mathbf{A}) \ \text{ initialized with } \ \mathbf{A}^k$$

end
Algorithm 1: Sparse Distribution Tomography (SparsDT) algorithm [Pad et al 17].

Learning algorithm (SparsDT)

Estimation of $\hat{\gamma}(\mathbf{u}_l)$ and $\hat{\alpha}$

For $\mathbf{u} \in \mathbb{R}^M$, recall that $\mathbf{u}^T\mathbf{y}_1, \ldots, \mathbf{u}^T\mathbf{y}_K$ are **realizations of $\mathbf{u}^T\mathbf{y}$** which is SαS.
There exists several estimator from a set of realizations [Nolan 01] [Zolotarev 57][Nicolas 02] [Ma and Nikias 95] [Achim et al 08].

$$\hat{\alpha}(\mathbf{u}) = \left(\frac{6}{\pi^2 K} \sum_{k=1}^{K} (\log |\mathbf{u}^T\mathbf{y}_k| - \log \hat{\kappa}(\mathbf{u}))^2 - \frac{1}{2} \right)^{-\frac{1}{2}}$$

where $\log \hat{\kappa}(\mathbf{u}) = \frac{1}{K} \sum_{k=1}^{K} \log |\mathbf{u}^T\mathbf{y}_k|$. Then,

$$\hat{\alpha} = \frac{1}{L} \sum_{l=1}^{L} \hat{\alpha}(\mathbf{u}_l),$$

and

$$\log \hat{\gamma}(\mathbf{u}_l) = \hat{\alpha} \log \hat{\kappa}(\mathbf{u}_l) - (\hat{\alpha} - 1)\psi(1),$$

with $\phi(1)$ the negative of the Euler-Mascheroni constant.

Minimization of the cost $\mathcal{F}$

The inner minimization of $\mathcal{F}$ is performed using a (sub)gradient descent:

$$\mathbf{A}^{k+1} = \mathbf{A}^{k} - \eta \nabla \mathcal{F}(\mathbf{A}^{k}),$$

where $\eta > 0$ and

$$\nabla \mathcal{F}(\mathbf{A}) = \frac{1}{\hat{\alpha} L} \sum_{l=1}^{L} \mathrm{sgn}(\log(\gamma \|\mathbf{A}^T \mathbf{u}_l\|_{\hat{\alpha}}^{\hat{\alpha}}) - \log \hat{\gamma}(\mathbf{u}_l)) \cdot \frac{\nabla \|\mathbf{A}^T \mathbf{u}_l\|_{\hat{\alpha}}^{\hat{\alpha}}}{\|\mathbf{A}^T \mathbf{u}_l\|_{\hat{\alpha}}^{\hat{\alpha}}}$$

with,

$$\nabla \|\mathbf{A}^T \mathbf{u}_l\|_{\hat{\alpha}}^{\hat{\alpha}} = \alpha \begin{bmatrix} \mathrm{sgn}(\mathbf{A}_{\cdot 1}^T \mathbf{u}) |\mathbf{A}_{\cdot 1}^T \mathbf{u}|^{\alpha - 1} \mathbf{u}^T \\ \vdots \\ \mathrm{sgn}(\mathbf{A}_{\cdot N}^T \mathbf{u}) |\mathbf{A}_{\cdot N}^T \mathbf{u}|^{\alpha - 1} \mathbf{u}^T \end{bmatrix}^T$$

Learning algorithm (SparsDT)

Why Sparse Distribution Tomography (SparsDT) ?

- Characteristic function of $\mathbf{y}$

$$\hat{p}_{\mathbf{y}}(\omega) = \exp(-\gamma \|\mathbf{A}^T\omega\|_\alpha^\alpha), \ \omega \in \mathbb{R}^M$$

- Restriction to one line $\omega = \mathbf{u}s$ for $s \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^M$,

$$\hat{p}_{\mathbf{y}}(\mathbf{u}s) = \exp(-\gamma \|\mathbf{A}^T\mathbf{u}s\|_\alpha^\alpha) = \underbrace{\exp(-\gamma \|\mathbf{A}^T\mathbf{u}\|_\alpha^\alpha |s|^\alpha)}_{\hat{p}_{\mathbf{u}^T\mathbf{y}}(s)}, \ s \in \mathbb{R}$$

- Analogy with the **Fourier slice theorem** used in **tomography**,

Experimental Results on synthetic SαS data

Data generation

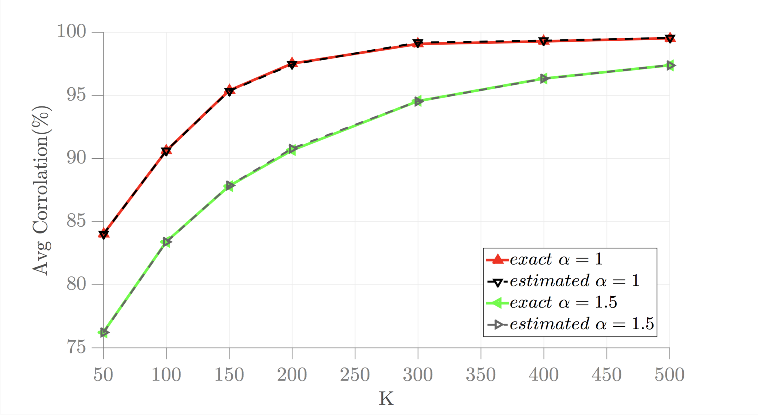
- Generate 50 random matrices $\mathbf{A}$,
- For each $\mathbf{A}$, generate $K \in \mathbb{N}$ SαS random vectors $\mathbf{x}_k$,
- Compute the corresponding data $\mathbf{y}_k = \mathbf{A}\mathbf{x}_k$,
- Given the $\mathbf{y}_k$, run the algorithm to estimate both α and $\mathbf{A}$.

Evaluation metrics

- **Average correlation**: Match the columns of $\mathbf{A}$ and $\hat{\mathbf{A}}$ such that the average correlation is maximized.
- **Exact recovery**: The dictionary $\mathbf{A}$ is considered to be exactly recovered if the average correlation is larger than 0.97.
- The metrics are averaged over the 50 generated dictionaries $\mathbf{A}$.

Experimental Results on synthetic SαS data

Influence of the number of samples K



Comparisons with SCA methods

- ℓ_2/ℓ_1:

$$\hat{\mathbf{A}}_{\ell_2/\ell_1} = \arg\min_{\mathbf{A}} \frac{1}{2K} \sum_{k=1}^{K} \left(\min_{\mathbf{x}_k} \|\mathbf{y}_k - \mathbf{A}\mathbf{x}_k\|_2^2 \ \text{ s.t. } \ \|\mathbf{x}_k\|_1 \leq \lambda_1 \right)$$

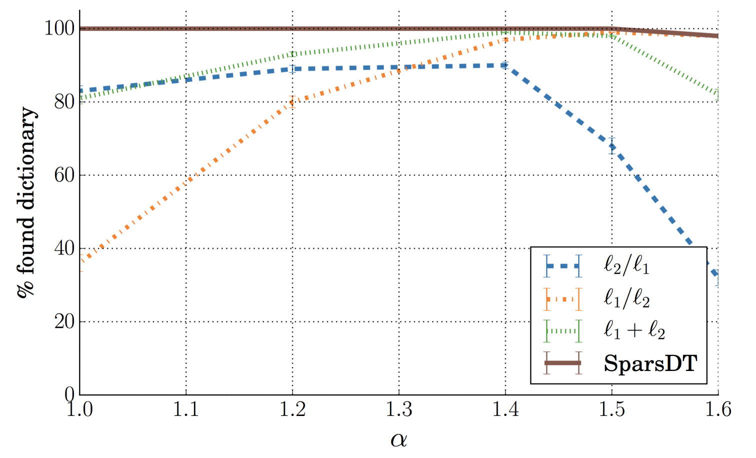
- ℓ_1/ℓ_2:

$$\hat{\mathbf{A}}_{\ell_1/\ell_2} = \arg\min_{\mathbf{A}} \frac{1}{2K} \sum_{k=1}^{K} \left(\min_{\mathbf{x}_k} \|\mathbf{x}_k\|_1 \ \text{ s.t. } \ \|\mathbf{y}_k - \mathbf{A}\mathbf{x}_k\|_2^2 \leq \lambda_2 \right)$$

- $\ell_2 + \ell_1$:

$$\hat{\mathbf{A}}_{\ell_2/\ell_1} = \arg\min_{\mathbf{A}} \frac{1}{2K} \sum_{k=1}^{K} \left(\min_{\mathbf{x}_k} \|\mathbf{y}_k - \mathbf{A}\mathbf{x}_k\|_2^2 + \lambda_3 \|\mathbf{x}_k\|_1 + \lambda_4 \|\mathbf{x}_k\|_2^2 \right)$$

Exact recovery rate in function of α



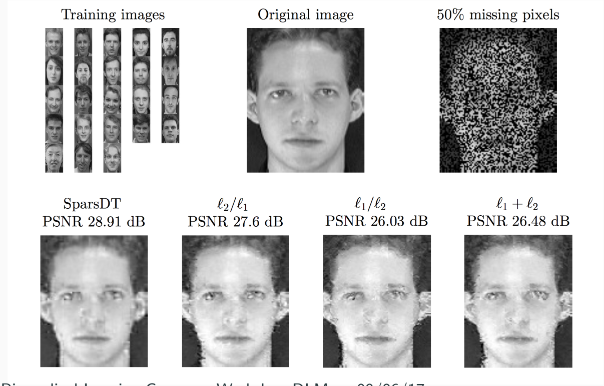
Experimental Results on synthetic SαS data

Toward real imaging problems

- The SαS model may be **too restrictive** to describe real data,
- This assumption can be relaxed by considering **heavy-tailed distribution** and the generalized central limit theorem,
- **Elimination** of the generated $\mathbf{u}$ leading to $\hat{\alpha}(\mathbf{u})$ greater than 2,
- **Good performaces** of the method have been shown in [Pad et al 17] for inpainting and denoising.

Inpainting problem

- Learning using 23 images,
- 230 atoms in the dictionary, 7×7 patches,
- Remove 50% of the pixels of a new image,
- Reconstruct the image using the learned dictionaries.



Application to 3D Deconvolution Microscopy (Ongoing Work)

3D Deconvolution Microscopy

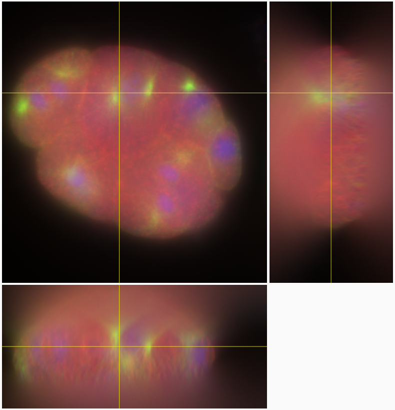
Model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

Main issues

- Large volumes,
- Non isotropic resolution,
- Non exact knowledge of the PSF,

C. Elegans embryo (Widefield)



Learning strategy

Motivations

- Benefit from the better lateral (x, y) resolution
- Reduce the computational complexity

$\Longrightarrow$ **Dictionary learned from 2D (x, y) patches and applied in every 2D sections**

Learning strategy

1. **Classical deconvolution**

$$\hat{\mathbf{x}}_{\mathrm{init}} \in \left\{ \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{H}\mathbf{x}-\mathbf{y}\|_2^2 + \lambda_{\mathrm{init}}\mathcal{R}(\mathbf{L}\mathbf{x}) + i_{\geqslant 0}(\mathbf{x}) \right\},$$

2. **Dictionary learning** using $M_p \in \mathbb{N}$ patches extracted from the lateral sections (x, y) of $\hat{\mathbf{x}}_{\mathrm{init}}$,
3. **Final deconvolution** in a plug-and-play [Chan et al 17, Sreehari et al 16] fashion
   3.1 Denoising, sparse coding
   3.2 Generalized Tikonov deconvolution

Implementation details

- The first and last deconvolution is implemented using the inverse problem library (*GlobalBioIm*) developped in our group [Unser et al 17],
- The sparse-coding step is performed using orthogonal matching pursuit (OMP) [Pati et al 93],
- The dictionary learning is performed either by SparsDT or $\ell_1 + \ell_2$ using the SPAMS library [Mairal et al 10].

The GlobalBioIm Matlab library

Motivations

- **Simplifying** the process of solving inverse problems,
- Capitalizing on the **strong commonalities** between various image formation **models**,
- **Flexibility** and **modularity** of the framework,
- **Reproductible** research.

Modular formulation of inverse problems: Three key ingredients

- **Linear Operators** (LinOp)
- **Cost Functions** (Cost)
- **Optimization algorithms** (Opti)

General Philosophy

- Build your forward model,
- Build your cost function (data term + regularization),
- Minimize the cost with an optimization routine.

Regularized least-squares deconvolution

$$\hat{\mathbf{x}} \in \left\{ \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\mathcal{R}(\mathbf{L}\mathbf{x}) + i_{\geqslant 0}(\mathbf{x}) \right\},$$

When $\mathcal{R}(\mathbf{L}\mathbf{x})$ is the **TV** regularizer:

```
%% Building the Cost
H=LinOpConv(psf);
Fdata=CostL2(H,y);
L=LinOpGrad(size(y));
R=CostMixNorm21([4]);
P=CostNonNeg(size(y));
```

```
%% Minimizing the Cost (ADMM)
Fn={lamb*R,P};
Hn={L,LinOpIdentity(size(y))};
rho_n=[1e-1,1e-1];
ADMM=OptiADMM(Fdata,Fn,Hn,rho_n);
ADMM.maxiter=50;
ADMM.run(y);
```

The GlobalBioIm Matlab library

General Philosophy

- Build your forward model,
- Build your cost function (data term + regularization),
- Minimize the cost with an optimization routine.

Regularized least-squares deconvolution

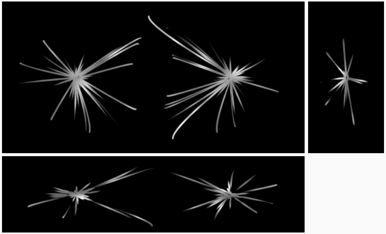
$$\hat{\mathbf{x}} \in \left\{ \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{Hx}-\mathbf{y}\|_2^2 + \lambda\mathcal{R}(\mathbf{Lx}) + i_{\geqslant 0}(\mathbf{x}) \right\},$$

When $\mathcal{R}(\mathbf{Lx})$ is the **Hessian Shatten-norm** regularizer:

```
%% Building the Cost
H=LinOpConv(psf);
Fdata=CostL2(H,y);
--> L=LinOpHess(size(y));
--> R=CostMixNormShatt1([],1);
P=CostNonNeg(size(y));
```

```
%% Minimizing the Cost (ADMM)
Fn={lamb*R,P};
Hn={L,LinOpIdentity(size(y))};
rho_n=[1e-1,1e-1];
ADMM=OptiADMM(Fdata,Fn,Hn,rho_n);
ADMM.maxiter=50;
ADMM.run(y);
```

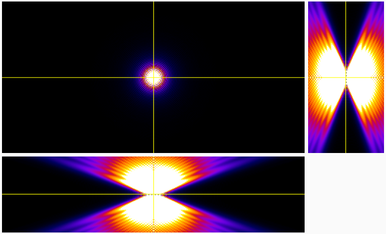
Ground Truth



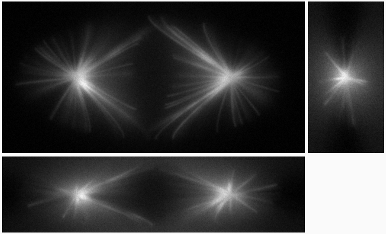
Data simulated using a realistic model [Sage et al 13, Vonesch and Lefkimmiatis 14]

- Born and Wolf PSF model,
- Shot noise and readout noise,
- Background fluorescence,
- Quantization.

PSF



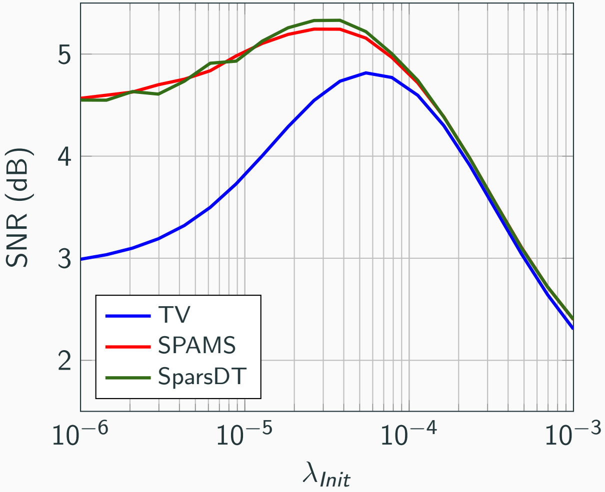
Corrupted data



Reconstruction Settings

- Initial TV deconvolution for different λ_{init},
- Dictionary learning with SPAMS ($\ell_2 + \ell_1$) and SparsDT (patch size 8×8, 10000 iterations, 150 atoms in the dictionary)
- Final deconvolution:
  - max atoms for OMP fixed to 10,
  - $\lambda_{\text{final}} = 5$.
  - 20 iterations.

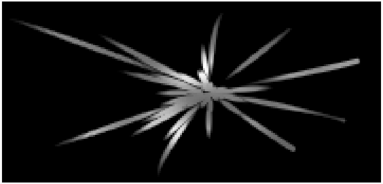
Numerical Experiments — Results in terms of SNR



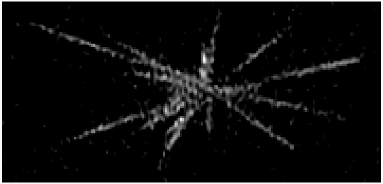
Numerical Experiments — Reconstructions for $\lambda_{Init} = 10^{-6}$

Zooms

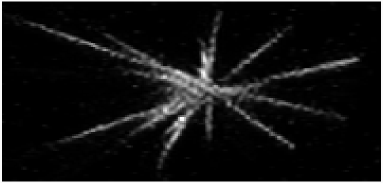
Ground Truth



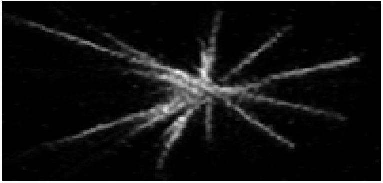
TV Regularization



SPAMS

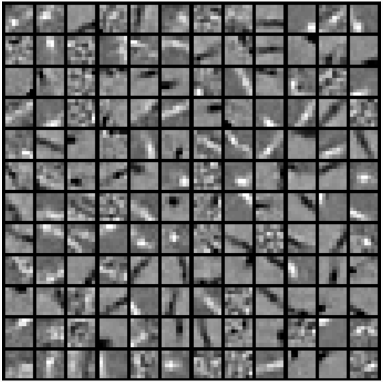


SparsDT

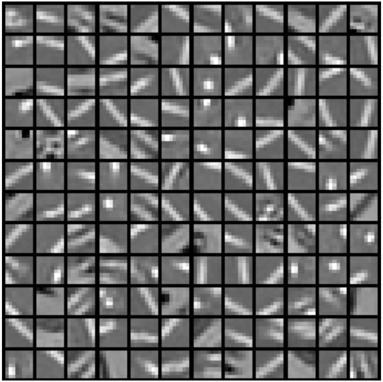


Example of learned dictionary ($\lambda_{Init} = 10^{-6}$)

SparsDT



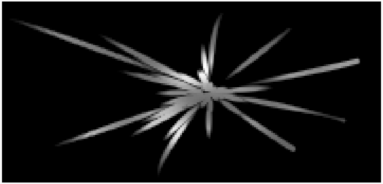
SPAMS



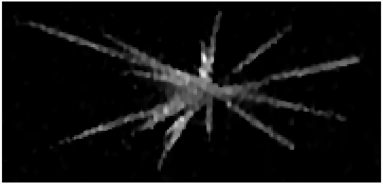
Numerical Experiments — Reconstructions for $\lambda_{Init} = 2.6 \cdot 10^{-5}$

Zooms

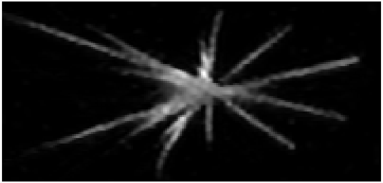
Ground Truth



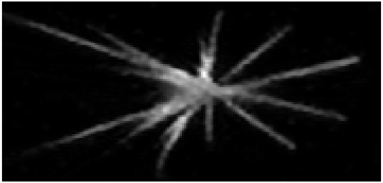
TV Regularization



SPAMS

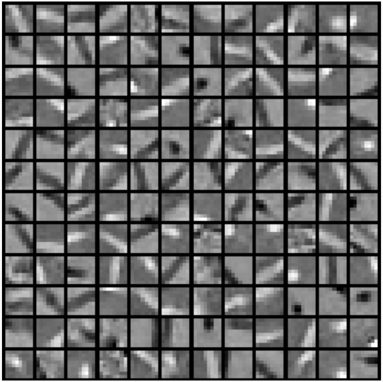


SparsDT

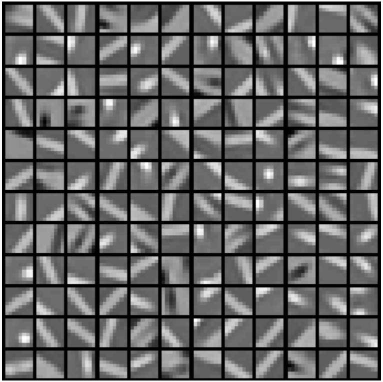


Example of learned dictionary ($\lambda_{Init} = 2.6 \cdot 10^{-5}$)

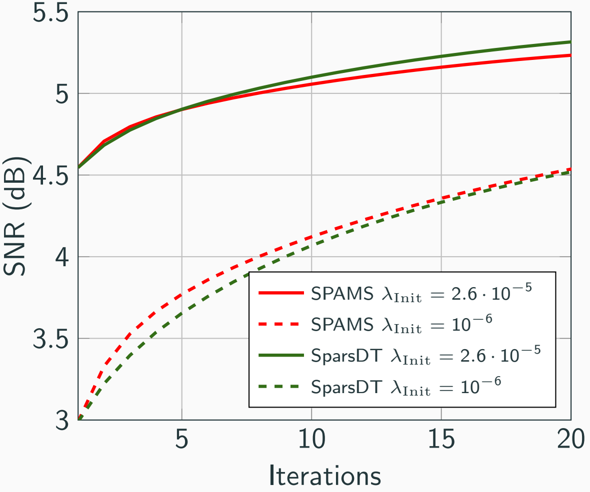
SparsDT



SPAMS



Numerical Experiments — SNR evolution final deconvolution loop



Take Home Message

SparsDT dictionary learning algorithm

- Completely **novel approach** with a new philosophy,
- Has a **rigourous** mathematical justification,
- Combine both **statistical** and **deterministic** priors,
- **No parameter** to tune,
- **Competitive performances** with state of the art methods that have been studied for the last ten years,

Dictionary Learning Based on Sparse Distribution Tomography.
Proceedings of the 34th International Conference on Machine Learning 2017.
Pedram Pad, Farnood Salehi, Elisa Celis, Patrick Thiran and Michael Unser.

GlobalBioIm: A Unifying Computational Framework for Solving Inverse Problems.
Proceedings of the OSA Congress on Computational Optical Sensing and Imaging 2017.
Michael Unser, Emmanuel Soubies, Ferréol Soulez, Michael McCann, and Laurène Donati.
Website: http://bigwww.epfl.ch/algorithms/globalbioim/

Thank you for your attention!

[Hyvärinen et al 04] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, John Wiley & Sons, 2004.

[Zou et al 06] H. Zou, T. Hastie, and R. Tibshirani, *Sparse principal component analysis*, Journal of computational and graphical statistics, 15(2), 265-286, 2006.

[Gribonval et al 06] R. Gribonval and S. Lesage, *A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges.*, Proc ESANN, 323-330, 2006.

[Pad et al 17] P. Pad, F. Salehi, E. Celis, P. Thiran and M. Unser, *Dictionary Learning Based on Sparse Distribution Tomography.*, In International Conference on Machine Learning (pp. 2731-2740), 2017.

[Nikias and Shao 95] C. L. Nikias and M. Shao, *Signal processing with alpha-stable distributions and applications.*, Wiley-Interscience, 1995.

References II

[Meerschaert and Scheffler 01] M. M. Meerschaert and H. P. Scheffler, *Limit distributions for sums of independent random vectors: Heavy tails in theory and practice.*, John Wiley & Sons, 2001.

[Vonesch and Lefkimmiatis 14] C. Vonesch and S. Lefkimmiatis, *Summary of the ISBI 2013 Grand Challenge on 3D Deconvolution Microscopy.*, Proc ISBI 2014.

[Sage et al 13] D. Sage, H. Kirshner, C. Vonesch, S. Lefkimmiatis, and M. Unser , *Benchmarking Image-Processing Algorithms for Biomicroscopy: Reference Datasets and Perspectives.*, Proc EUSIPCO 2013.

[Achim et al 08] A. Achim, A. Loza, D. Bull, and N. Canagarajah , *Statistical modelling for wavelet-domain image fusion.*, Image Fusion: Algorithms and Applications, 119-138, 2008

References III

[Chan et al 17] S.H. Chan, X. Wang, and O.A. Elgendye , *Plug-and-play admm for image restoration: Fixed-point convergence and applications.*, IEEE Transactions on Computational Imaging, 3(1):84 98, 2017.

[Pati et al 93] Y.C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, *Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition.*, Twenty-Seventh Asilomar Conference on Systems and Computers, pages 4044, 1993.

[Sreehari et al 16] S. Sreehari, S.V. Venkatakrishnan, B. Wohlberg, G. T Buzzard, L.F. Drummy, J.P. Simmons, and C. A. Bouman., *Plug-and-play priors for bright field electron tomography and sparse interpolation.*, IEEE Transactions on Computational Imaging, 2(4):408423, 2016.

[Unser et al 17] M. Unser, E. Soubies, F. Soulez, M. McCann, and L. Donati. , *GlobalBioIm: A Unifying Computational Framework for Solving Inverse Problems.*, Proceedings of the OSA Congress on Computational Optical Sensing and Imaging 2017.

[Mairal et al 10] J. Mairal, F. Bach, J. Ponce and G. Sapiro. , *Online Learning for Matrix Factorization and Sparse Coding.* , Journal of Machine Learning Research, volume 11, pages 19-60. 2010.

[Nolan 01] J.P. Nolan, *Maximum likelihood estimation and diagnostics for stable distributions.* , Lvy processes: theory and applications, 379-400, 2001.

[Zolotarev 57] V. M. Zolotarev, *Mellin-Stieltjes transforms in probability theory* , Theory of Probability & Its Applications, 2(4), 433-460, 1957.

[Nicolas 02] J. M Nicolas, *Introduction aux statistiques de deuxime espce: Applications des logs-moments et des logs-cumulants l'analyse des lois d'images radar.*, Traitement du signal, 19(3), 139-167, 2002

[Ma and Nikias 95] X. Ma and C.L. Nikias, *Parameter estimation and blind channel identification in impulsive signal environments.* , IEEE transactions on signal processing, 43(12), 2884-2897, 1995