EPFL

Dictionary Learning with Statistical Sparsity in the Presence of Noise

Shayan Aziznejad[1], Emmanuel Soubies[2], Michael Unser[1]

1) Biomedical Imaging Group, EPFL, Lausanne, Switzerland
2) IRIT, CNRS, Université de Toulouse, France

Swiss National Science Foundation

European Research Council
Established by the European Commission

Sparsity and Dictionary Learning

- Sparsity: A fundamental paradigm in modern signal processing.

- Natural signals and images have sparse representations on "certain basis".

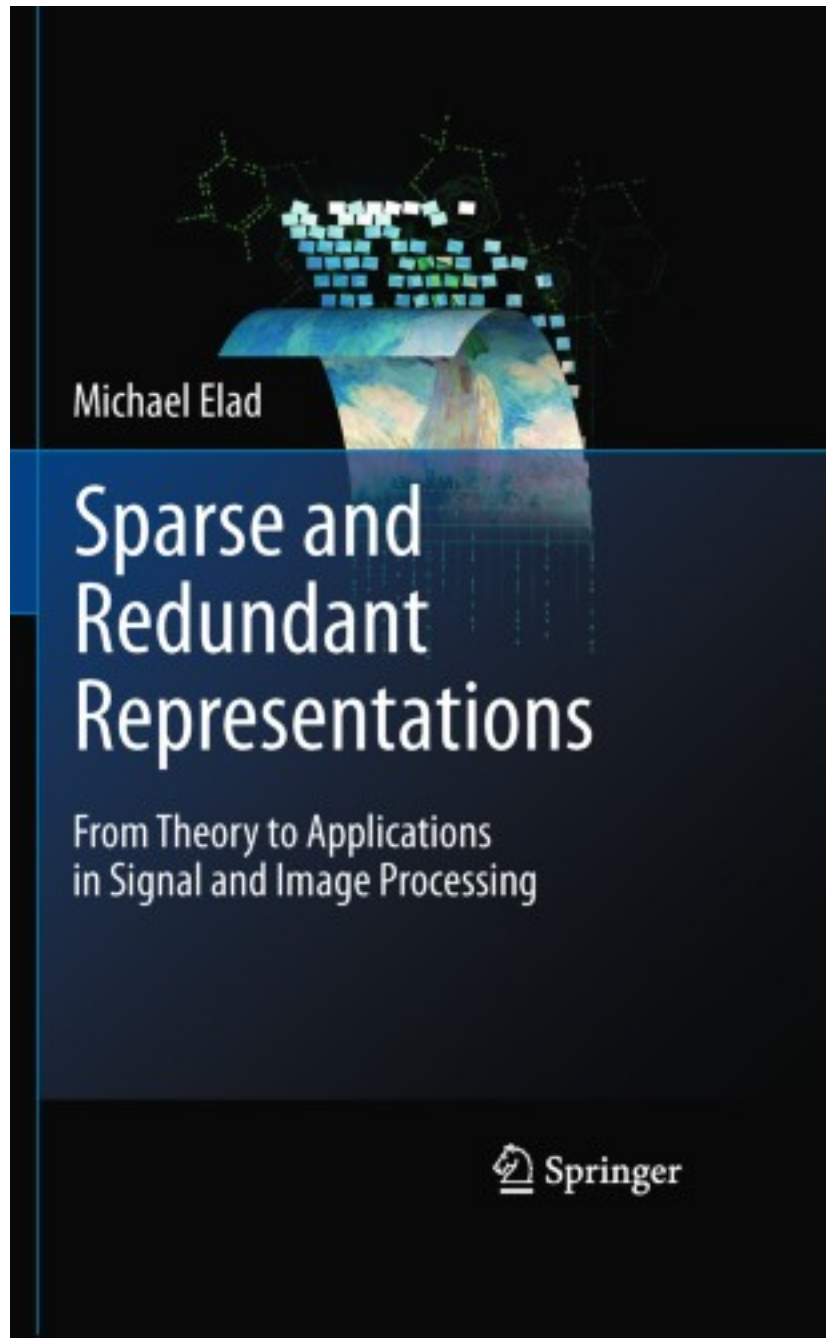
  $\mathbf{y} \in \mathbb{R}^M$: Data vector

  $\mathbf{A} \in \mathbb{R}^{M \times P}$: Dictionary

  $\Rightarrow \mathbf{y} \approx \mathbf{A}\mathbf{x},$ where $\mathbf{x} \in \mathbb{R}^P$ has K nonzero entries with $K << M$

- Applications in image processing: Reconstruction, Compression, Denoising, ...

- How to choose $\mathbf{A}$?
  - Fixed dictionaries: Fourier-based (DCT, DFT), Wavelet-based (DWT), ...
  - Dictionary learning: Learning the transformation from the data



Symmetric-Alpha-Stable Distributions

- A parametric family of probability distributions

- Defined via their characteristic function

$$\Phi_X(\omega) = \mathbb{E}\left[\mathrm{e}^{\mathrm{j}\omega X}\right] = \exp\left(-\gamma|\omega|^{\alpha}\right)$$

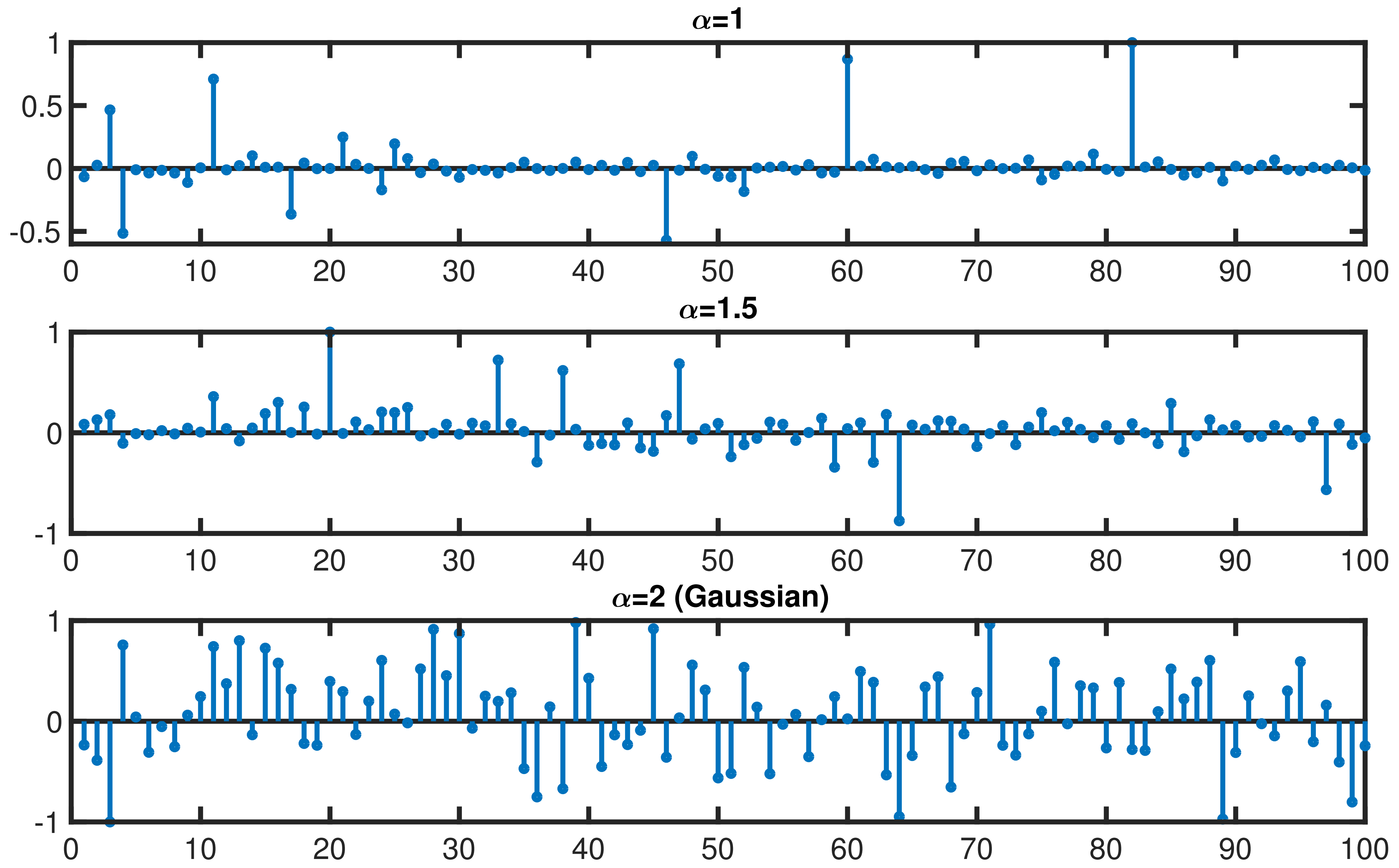
- Parameters: Stability $\alpha \in (0, 2]$ and dispersion $\gamma > 0$

- Tunable sparsity using the stability parameter α

- $\alpha = 2$: Zero mean Gaussian distributions with variance $\sigma^2 = 2\gamma$

- Closed under addition

$$X_n \overset{i.i.d.}{\sim} S\alpha S(\gamma) \text{ and } w_n \in \mathbb{R} \text{ for } n = 1, \ldots, N \quad \Rightarrow \quad X = \sum_{n=1}^{N} w_n X_n \sim S\alpha S\left(\gamma\|\mathbf{w}\|_{\alpha}^{\alpha}\right)$$



[1] G. Samoradnitsky and M.S. Taqqu, Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance, Chapman and Hall, New York, 2017.

[2] Nikias, C., and M. Shao, Signal processing with alpha stable distributions and applications, *John Wiley&Sons, New York,1995*.

Dictionary Learning with SaS Prior

- New Signal Model: $\mathbf{y} \approx \mathbf{A}\mathbf{x}$, where $\mathbf{x} = (x_1, \ldots, x_P)$ with $x_p \overset{i.i.d.}{\sim} S\alpha S(\gamma)$ for $p = 1, \ldots, P$

- Dictionary Learning Problem: Find $\mathbf{A}$ given the input data $\{\mathbf{y}_1, \ldots, \mathbf{y}_K\}$.

- SparsDT: Method for finding $\mathbf{A}$ in the noiseless scenario
  - $Z_{\mathbf{u}} = \mathbf{u}^T\mathbf{y} = \mathbf{u}^T\mathbf{A}\mathbf{x} \quad \Rightarrow \quad Z_{\mathbf{u}} \sim S\alpha S\left(\gamma\|\mathbf{A}^T\mathbf{u}\|_\alpha^\alpha\right)$
  - Estimating α and $\gamma(\mathbf{u}) = \gamma\|\mathbf{A}^T\mathbf{u}\|_\alpha^\alpha$ from i.i.d. realizations of $Z_{\mathbf{u}}$
  - Solving the nonlinear system of equations $\gamma(\mathbf{u}_\ell) = \gamma\|\mathbf{A}^T\mathbf{u}_\ell\|_\alpha^\alpha, \ell = 1, \ldots, L$
  - $L \geq M \times P \quad \Rightarrow \quad$ Solution is equal to $\mathbf{A}$ up to permutations and negation.

- Our contribution: Extending SparsDT to the case where we have additive Gaussian noise

[1] P. Pad, F. Salehi, E. Celis, P. Thiran, M. Unser, “Dictionary Learning Based on Sparse Distribution Tomography," Proceedings of the Thirty-Fourth International Conference on Machine Learning (ICML'17), Sydney, Australia, August 6-11, 2017, pp. 2731-2740.

Robust SparsDT

- Our proposed model: $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, where $x_p \overset{i.i.d.}{\sim} S\alpha S(\gamma)$ and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I})$
  - $Z_{\mathbf{u}} = \mathbf{u}^T\mathbf{y} = \mathbf{u}^T\mathbf{A}\mathbf{x} + \mathbf{u}^T\mathbf{n} \quad \Rightarrow \quad \Phi_{Z_{\mathbf{u}}}(\omega) = \exp\left(-\gamma\|\mathbf{A}^T\mathbf{u}\|_\alpha^\alpha|\omega|^\alpha - \frac{\sigma_n^2}{2}\|\mathbf{A}\mathbf{u}\|_2^2\omega^2\right)$
  - Dictionary learning problem is reduced to estimating parameters of $Z_{\mathbf{u}}$

- $X \sim S\alpha S(\gamma)$ and $Y \sim \mathcal{N}(0, \sigma^2)$ are independent and $Z = X + Y$.

- Estimation problem: Given i.i.d. realizations $\{Z_k\}_{k=1}^K$, find $\alpha \in (0, 2)$, $\gamma > 0$ and $\sigma > 0$.

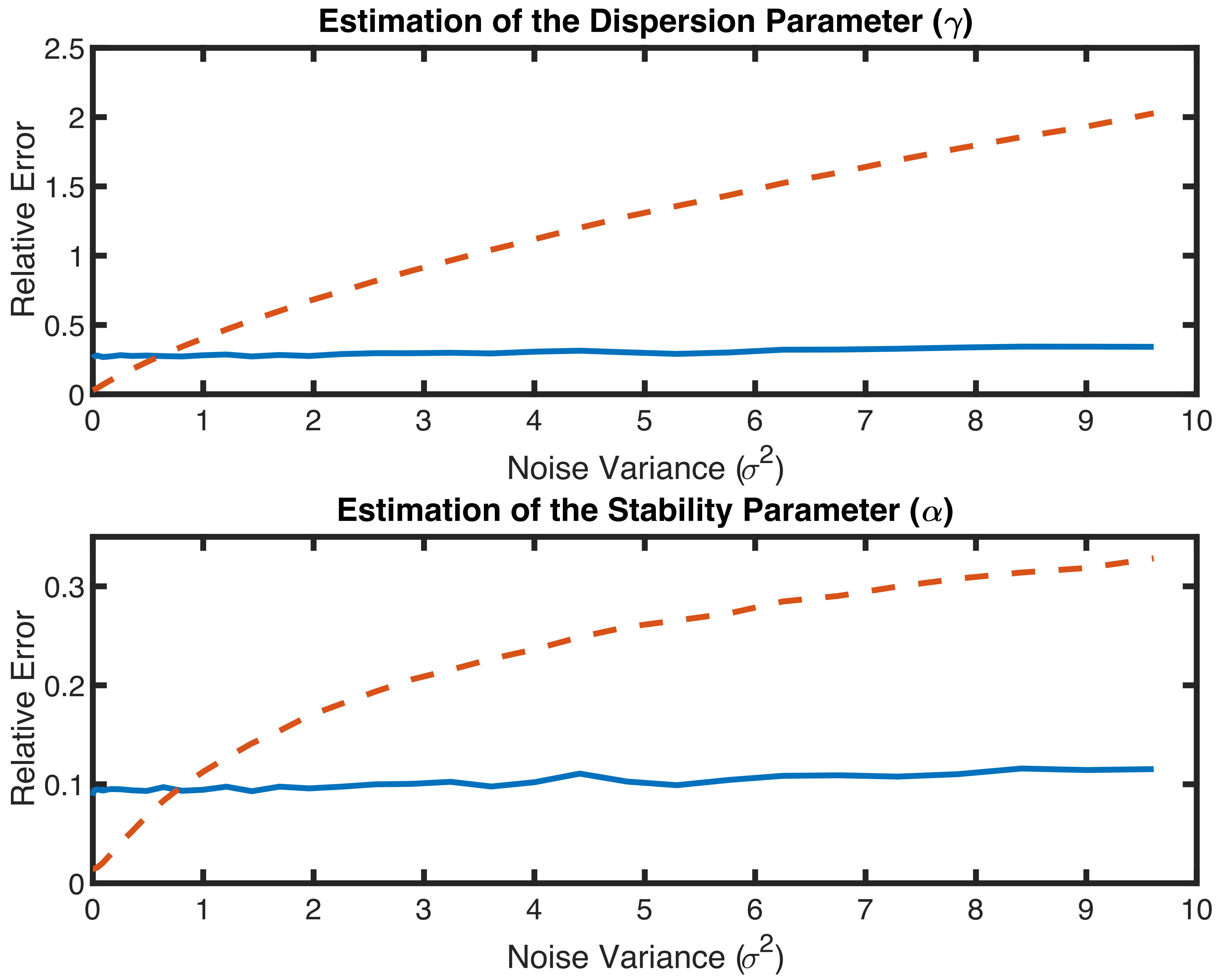
- The presented estimation module could be of independent interest.

Parameter Estimation Module

- Empirical characteristic function: $\tilde{\Phi}_Z(\omega) = \frac{1}{K}\sum_{k=1}^{K}\exp(\mathrm{j}\omega z_k)$
  - $\tilde{f}(\omega) = -\log_e\left(\tilde{\Phi}_Z(\omega)\right) \approx \gamma|\omega|^\alpha + \frac{\sigma^2}{2}\omega^2$
- Choose T frequencies $w_1, \ldots, w_T > 0$ and solve the minimization
  - $\min_{\substack{\alpha\in(0,2] \\ \gamma,\sigma\geq 0}} \sum_{t=1}^{T}\left(\tilde{f}(\omega_t) - \gamma|\omega_t|^\alpha - \frac{\sigma^2}{2}\omega_t^2\right)^2$
- For a fixed α, finding γ, σ can be done by solving a linear least squares problem.
  - Closed-form expression
- $\alpha \in (0,2] \quad \Rightarrow \quad$ We find α using grid search.

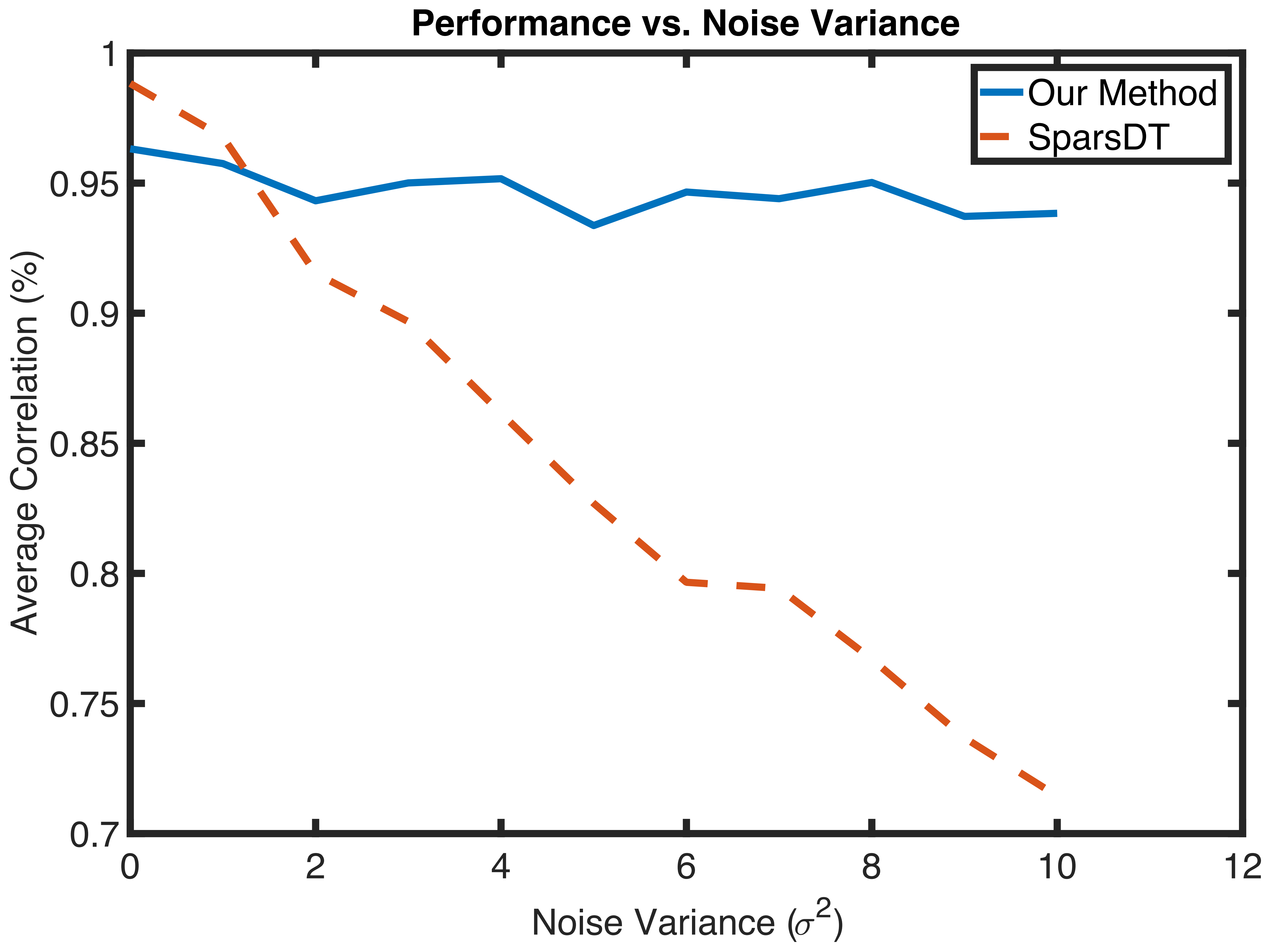
Numerical Example: Parameter Estimation

- True parameters: $\alpha = 1.255$ and $\gamma = 1$
- Sample size: $K = 10^4$
- Frequencey samples: $\boldsymbol{\omega} = (0.1, 0.11, 0.12, \ldots, 0.3)$
- Grid size for finding α: $h = 0.01$
- Results are averaged over 1000 runs.
  - Blue line: Our method.
  - Orange dash-dotted line: The estimation method used in SparsDT.



Numerical Example: Dictionary Learning

- Problem dimension: $M = 20$ and $P = 30$
- Rnadom Gaussian Dictionary
- Training size: $K = 1000$
- Signal model: $\alpha = 1.2$ and $\gamma = 1$



Conclusion

- A novel stochastic model for sparse signals with additive Gaussian noise.
- A new parameter estimation technique for the sum of SαS and Gaussian random variables.
- Making SparsDT robust to additive Gaussian noise
- Future directions:
  - Improving the parameter estimation technique
  - Applying on real problems: e.g. in Optical Coherence Tomography (OCT)

Scrivanti, G., Calatroni, L., Morigi, S., Nicholson, L., & Achim, A. (2020). Non-convex Super-resolution of OCT images via sparse representation. *arXiv preprint arXiv:2010.12576*.