Compressive sensing in impulsive environments

John Nolan George Tzagkarakis Panagiotis Tsakalides

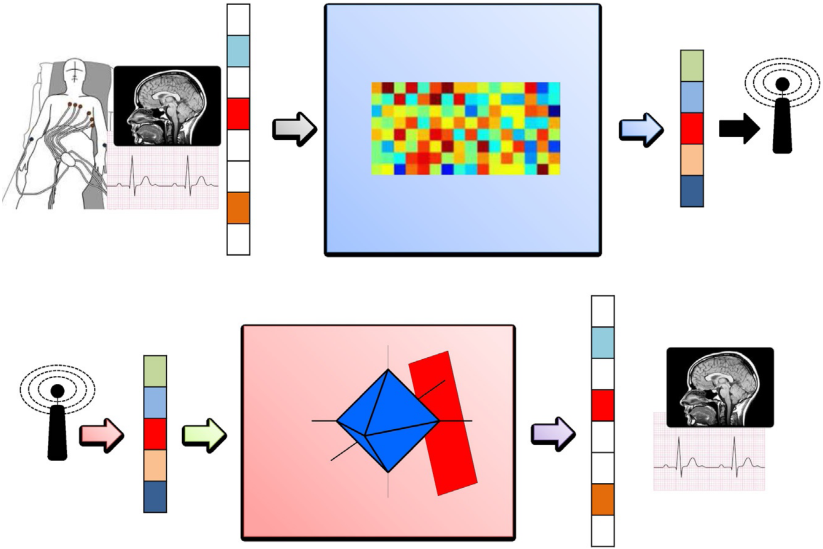
American University
Washington, DC USA

University of Crete
Heraklion, Greece

23rd Workshop on Applied Stochastic Processes
Shiraz, Iran
15 February 2022

Outline

Compressive Sensing Idea



From Sensors for Health Monitoring, Chapter 4, Cabral et al. (2019)

Here we examine this problem when the noise is heavy tailed. Engineers call this impulsive noise. Specifically, we will use symmetric α-stable noise, abbreviated $S\alpha S$

Split the problem into two parts:

- Sampling in the presence of impulsive noise (top half of previous page)
- Signal reconstruction in the presence of impulsive noise (bottom half of previous page)

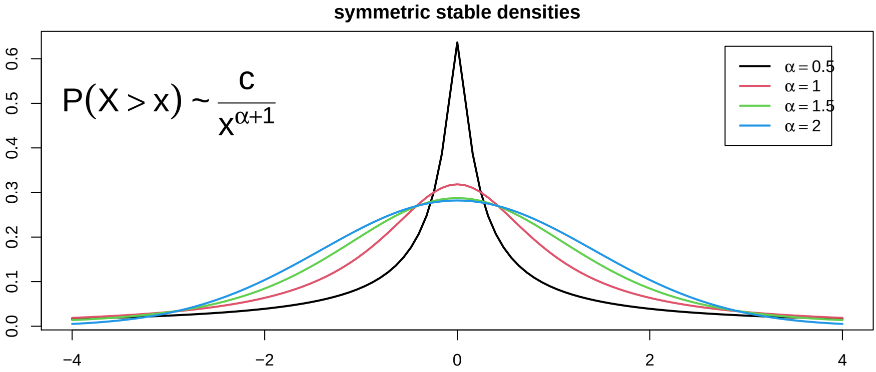
Stable distributions

A family of probability distributions, with four parameters:

- $\alpha \in (0, 2]$ is the index of stability
- $\beta \in [-1, 1]$ is the skewness parameter
- $\gamma > 0$ is the scale/dispersion parameter
- $\delta \in \mathbb{R}$ is a location parameter.

Stability property under addition/convolution, domain of attraction for sums, heavy tails, possible skewness.

No explicit formula for densities, now fast and reliable software to compute densities, cdf, simulate and estimate parameters. Here we will work with the symmetric case ($\beta = 0$), where the normalized (scale $\gamma = 1$, location $\delta = 0$) densities look like:



The notation $S\alpha S(\gamma)$ will be used for a symmetric $\alpha-$stable law with scale γ.

Compressive sampling

A vector $\mathbf{v}$ is *s-sparse* if

$\|\mathbf{v}\|_0 = s$ is the number of non-zero elements of $\mathbf{v}$.

Assume M, N, N' are positive integers with $M \ll N \leq N'$. Start with a (real, discrete time) signal $\mathbf{x} = (x_1, \ldots, x_N)^\top \in \mathbb{R}^N$. Let $A : \mathbb{R}^N \mapsto \mathbb{R}^M$ be a function that maps into a lower dimensional space.

In the standard model, A is linear, typically $A = \Phi\Psi^\top$ where Φ is an $(M \times N)$ random measurement matrix and Ψ is an $(N' \times N)$ sparsifying matrix. A common example for Ψ is the discrete cosine transform (DCT). Let $\mathbf{e} \in \mathbb{R}^N$ be the observational noise, then

$$\mathbf{y} = \Phi(\Psi^\top \boldsymbol{\alpha}_0 + \mathbf{e}) = \Phi\Psi^\top \boldsymbol{\alpha}_0 + \mathbf{n} = A\boldsymbol{\alpha}_0 + \mathbf{n},$$

where $\boldsymbol{\alpha}_0 = \Psi \mathbf{x}_0 \in \mathbb{R}^{N'}$ and $\mathbf{n} \in \mathbb{R}^M$ is transformed noise.

The goal is to find $\boldsymbol{\alpha}$ (so really Ψ) that satisfies an ℓ^0 - ℓ^2 constrained optimization problem

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \quad \text{s.t.} \quad \|\mathbf{y} - A\boldsymbol{\alpha}\|_2 \leq \epsilon,$$

where ϵ is some tolerance. This is a computationally hard problem.

Candés, Romberg, Tao and Donoho (2006) showed that for many problems, the ℓ^1 norm is equivalent to the ℓ^0 norm, so one usually solves the ℓ^1 - ℓ^2 constrained optimization problem

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - A\boldsymbol{\alpha}\|_2 \leq \epsilon, \tag{1}$$

which is more tractable.

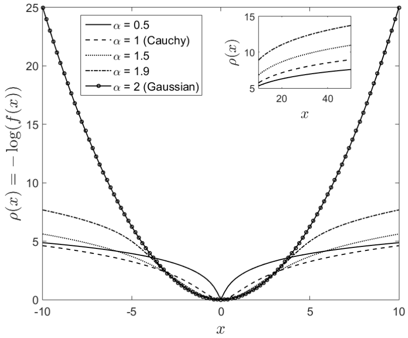
If $s = \|\boldsymbol{\alpha}\|_0 \approx \|\boldsymbol{\alpha}\|_1 \ll N$, then this greatly reduces the dimensionality. We can transmit an s vector instead of the original N vector without losing much of the signal.

However, if $\mathbf{e}$ has heavy tailed components, then $\mathbf{n}$ is heavy tailed and (1) is poorly behaved. Heavy tailed noise in $\mathbf{e}$ gets spread to many columns of $A\boldsymbol{\alpha}$.

So the goal here is to find a more robust choice of A.

Nonlinear stable filter for A

We propose that a nonlinear map $A : \mathbb{R}^N \mapsto \mathbb{R}^M$ be used. Specifically we use a stable filter that downweights extreme values. Let $\rho(\cdot) = -\log f_\alpha(\cdot; \gamma, 0)$, where $f_\alpha(\cdot; \gamma, 0)$ is the density of a symmetric stable $S\alpha S = S(\alpha, \beta = 0, \gamma, \delta = 0)$ r.v.



For $\mathbf{x} \in \mathbb{R}^N$ and weights $\mathbf{w} \in \mathbb{R}^N$, define the weighted stable matched filter (WSMF) is

$$WSMF(\alpha, \gamma; \mathbf{w}, \mathbf{x}) = \arg\min_{\theta} \sum_{j=1}^{N} \rho(w_j(x_j - \theta)).$$

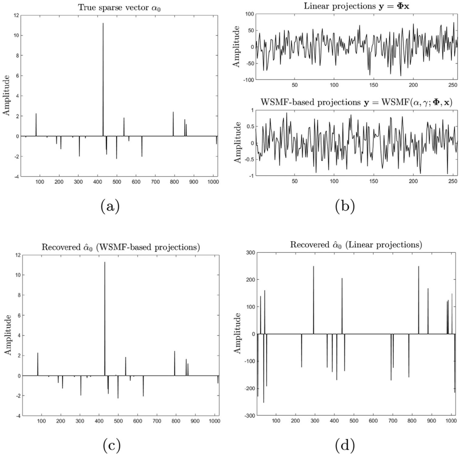
This was defined in Nolan (2008) using numerical calculation of $\rho(\cdot)$ and numerical minimization of (non-convex) objective function on the right.

We use this filter with the measurement matrix Φ having rows $\phi_1, \ldots, \phi_M$ and define the nonlinear transformation

$$A(\mathbf{x}) = (c_1 WMSF(\alpha, \gamma; \phi_1, \mathbf{x}), \ldots, c_M WMSF(\alpha, \gamma; \phi_M, \mathbf{x}))^\top,$$

where $c_i = \sum_{j=1}^{N} |\phi_{i,j}|$. In words: filter out extremes in each column of 'wide matrix' on page 4.

Simulated run with $N = 1024$, $s = \lceil 0.05N \rceil = 51$, $\Psi =$DCT, $\Phi =$ i.i.d. $\{-1, 1\}$ with $p = 1/2$, OMP reconstruction.

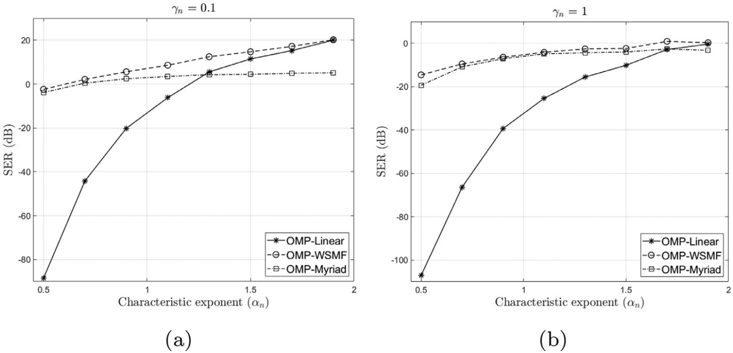


To assess performance across multiple runs, we use the signal-to-error (SER). For exact $\mathbf{x}$ and OMP reconstructed $\widehat{\mathbf{x}}$, the SER is

$$\mathrm{SER}(\mathbf{x}, \widehat{\mathbf{x}}) = 10 \log_{10} \left(\frac{\sum_{j=1}^{N} x_j^2}{\sum_{j=1}^{N} (x_j - \widehat{x}_j)^2} \right).$$

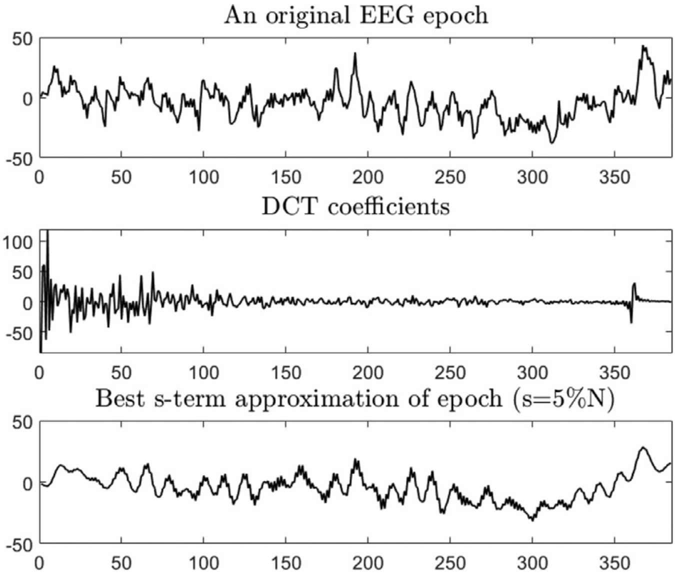
Better reconstruction gives a larger SER.

Simulation with 500 Monte Carlo runs

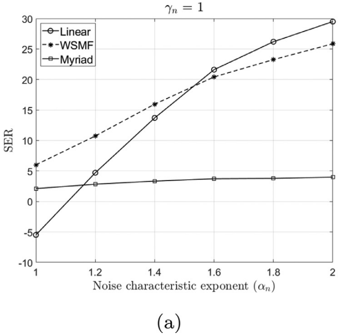


EEG data

$N = 384$, $s = \lceil 0.05N \rceil = 20$.



SER for EEG data with varying $\alpha \in \{1, 1.2, 1.4, 1.6, 1.8, 2\}$ and OMP over 500 simulations.



The paper gives an iterative algorithm for estimating α and γ.

The basic model used in compressed sensing reconstruction is

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{n},$$

where

- $\mathbf{y} = (y_1, \ldots, y_M)^\top$ is the received signal
- Φ is a random $M \times N$ measurement matrix
- $\mathbf{x} = (x_1, \ldots, x_N)^\top$ is the original signal with $\|\mathbf{x}\|_0 \leq s =$ number of non-zero terms in x
- $\mathbf{n} = (n_1, \ldots, n_M)^\top$ is i.i.d. additive noise

Typically $M < N$ - so underdetermined linear system, $s \ll N$ - so sparse, and $\mathbf{n}$ has a light tailed distribution, say $\text{Var}(n_j) < \infty$.

Choice of Φ

Typical choice of Φ is random, say i.i.d. Gaussian; it is known that this choice satisfies the Restriction Isometry Property (RIP) with high probability.

Does it make sense to use i.i.d. stable terms in Φ? Numerical experiments argue against this. The presence of very large terms in Φ seems to make the RIP constant bigger.

Signal reconstruction

Here we focus on the recovery part of the problem. To recover $\mathbf{x}$ from received $\mathbf{y}$, the standard solution is to find $\widehat{\mathbf{x}}$ that minimizes

$$\|\mathbf{y} - \Phi\mathbf{x}\|_2 \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq s.$$

When the noise term $\mathbf{n}$ is heavy tailed, the first term above is poorly behaved, so we seek a different approach.

The rest of this talk will focus on two topics:

- Replacing the 2-norm above with a $p-$norm, where $0 < p < 1$. This has the sparsity property of the ℓ^1 norm and it also more robust to heavy tails.
- Developing a reliable way to numerically solve the resulting problem.

Performance will be demonstrated with simulated data and with EEG data.

Fractional moments and reconstruction

For X $SaS(\gamma)$, it is known that the fractional lower order moments (FLOM) are given by

$$E|X|^p = \begin{cases} c(\alpha, p)\gamma^p & 0 < p < \alpha \\ +\infty & p \geq \alpha \end{cases} \tag{2}$$

For $0 < p < \alpha$, a sample estimate of $E|X|^p$ is $\left(\sum_{i=1}^{N} |x_i|^p\right)/N$.

We will reconstruct a signal by solving the minimization problem for $\mathbf{x}$:

$$\min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_p^p \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq s. \tag{3}$$

Because of (2), the left hand side above is the (p-th power of the) dispersion of $\mathbf{y} - \Phi\mathbf{x}$, so (3) is a minimum dispersion criterion.

Solving the optimization problem

We approximately solve the minimization problem using a greedy algorithm that is gradient based and automatically satisfies the constraint.

Since $p < 1$, the objective/cost function $\|\mathbf{y} - \Phi\mathbf{x}\|_p^p$ is not differentiable, we replace that term with a smoothed version. For $\epsilon > 0$, define

$$\|\mathbf{x}\|_{p,\epsilon}^p = \sum_{j=1}^{N} \left(x_j^2 + \epsilon\right)^{p/2}$$

We replace the previous objective function (3) with the following:

$$\min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_{p,\epsilon}^p \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq s. \tag{4}$$

MD-IHT: Minimum dispersion - iterative hard thresholding

Hard thresholding operator: $\mathbf{z} = H_s(\mathbf{x})$ has components $z_i = x_i$ if x_i is one of the s largest elements of $\mathbf{x}$; otherwise $z_i = 0$.

Calculations shows the gradient of $\|\mathbf{y} - \Phi\mathbf{x}\|_{p,\epsilon}^p$ w.r.t. $\mathbf{x}$ is

$$\mathbf{g} = -p\Phi^\top W \left(\mathbf{y} - \Phi\mathbf{x}\right),$$

where W is a diagonal matrix with $W_{i,i} = \left((y_i - \Phi_{i\cdot}\mathbf{x})^2 + \epsilon\right)^{-(1-p/2)}$.

The iterative approach to minimizing the smoothed objective function is start with an valid (sparsity s) initial value $\mathbf{x}^{(0)}$ and set

$$\mathbf{x}^{(t+1)} = H_s\left(\mathbf{x}^{(t)} + \mu^{(t)}\mathbf{g}^{(t)}\right),$$

where $\mu^{(t)}$ is a step size determined below. Repeat until relative change is less than some tolerance, call the result $\mathbf{x}_0^*$.

Parameter settings and convergence analysis

Choice of step size $\mu^{(t)}$: Proposition in the paper gives a value that cannot increase the dispersion error.

Other parameters: estimate α and γ from $\mathbf{y}$ using FLOMs; $\mathbf{x}^{(0)}$ suggestions, etc.

Choice of p: The constant $c(\alpha, p) \to \infty$ as p approaches α, so we recommend picking $p = \alpha/2 - \delta < 1$.

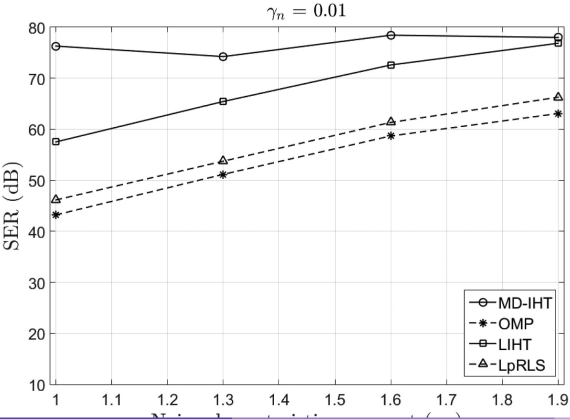
Bound on the reconstruction error: With assumptions on the RIP constant and upper bound on dispersion of the noise, say $\gamma \leq \eta$,

$$\|\mathbf{x}_0 - \mathbf{x}_0^*\|_2 \leq c(\text{RIP constant}, M, \alpha, p) \cdot \eta.$$

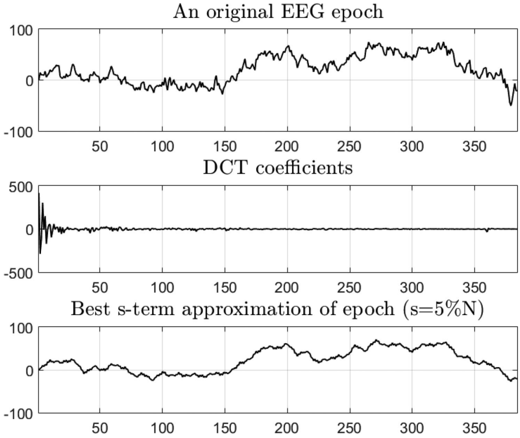
Performance with Synthetic Data

Compare MD-IHT to OMP = Orthogonal Matching Pursuit, LIHT = Lorentzian iterative hard thresholding, & LpRLS= ℓ^p-reweighted least squares

$N = 1024 =$ dimension of $\mathbf{x}$, $M = 0.25N = 256$, $s = 0.02\,N = 21$, with 500 simulations, plot the average SER=signal to error rate



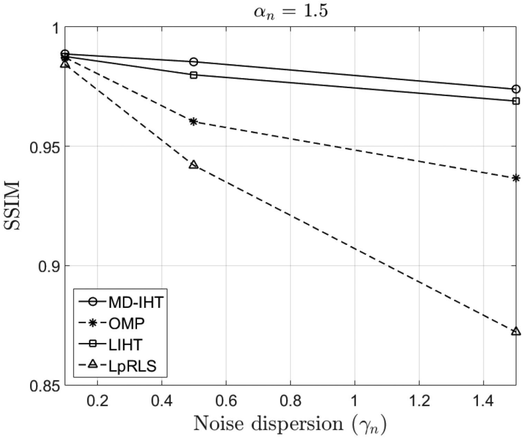
Example with EEG recordings



Here we use the Discrete Fourier Transform (DCT) in the middle row as a "sparsifying dictionary". Data from EEGLAB: $N = 384$, $M = \lceil 0.5N \rceil = 192$, sparsity $s = \lceil 0.05N \rceil = 20$.

EEG Data, cont'd

500 simulations contaminating the above data with $\alpha = 1.5$-stable noise and varying dispersion γ, plot the average SSIM=structural similarity index measure between the original data and the reconstructed signal



References

Nolan, J. P. Advances in nonlinear signal processing for heavy tailed noise, Proceeding of the International Workshop in Applied Probability (2008)

G. Tzagkarakis, J. P. Nolan, and P. Tsakalides. Compressive sampling using symmetric α-stable distributions for robust sparse signal reconstruction, *IEEE Transactions on Signal Processing*, **67**, 808-820 (2019), doi.org/10.1109/TSP.2018.2887400.

G. Tzagkarakis, J. P. Nolan, and P. Tsakalides. Robust nonlinear compressive sampling using symmetric α-stable distributions *Signal Processing, 182* (2021), /doi.org/10.1016/j.sigpro.2020.107944

Any Questions?