The Noise Collector for Sparse Recovery in High Dimensions

Chrysoula Tsogka

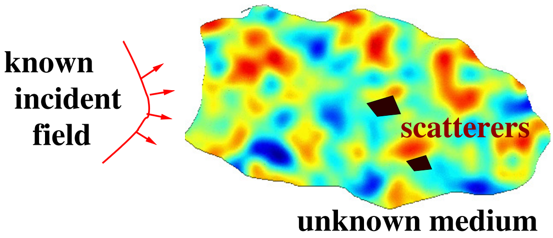
Collaborators: M. Moscoso (Univ. Carlos III de Madrid),
A. Novikov (Penn State),
G. Papanicolaou (Stanford University)

Department of Applied Mathematics, University of California, Merced

Support: AFOSR FA9550-17-1-0238,
AFOSR FA9550-18-1-0519.

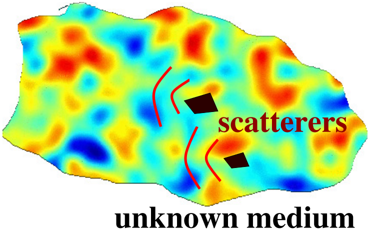
Table of contents

Inverse problems
Wave propagation



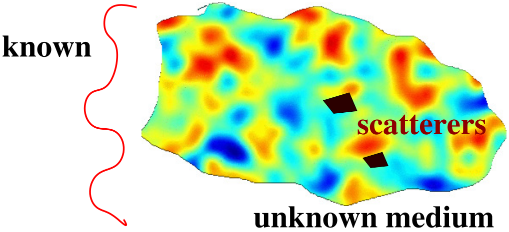
- Inverse problems aim to reconstruct a medium characteristics from knowledge of the response of the medium to a known incident field.
- In the context of the wave equation we seek to reconstruct the reflectivity by recording the medium's response to one or more known excitations.

Inverse problems
Wave propagation



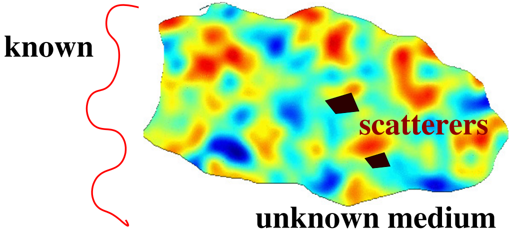
- Inverse problems aim to reconstruct a medium characteristics from knowledge of the response of the medium to a known incident field.
- In the context of the wave equation we seek to reconstruct the reflectivity by recording the medium's response to one or more known excitations.

Inverse problems
Wave propagation



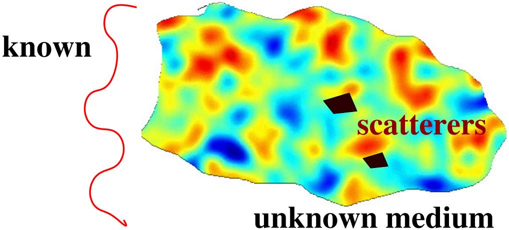
- Inverse problems aim to reconstruct a medium characteristics from knowledge of the response of the medium to a known incident field.
- In the context of the wave equation we seek to reconstruct the reflectivity by recording the medium's response to one or more known excitations.

Inverse problems
Wave propagation



- Inverse problems aim to reconstruct a medium characteristics from knowledge of the response of the medium to a known incident field.
- In the context of the wave equation we seek to reconstruct the reflectivity by recording the medium's response to one or more known excitations.
- We consider sparse reflectivities : often true in applications when the object to be imaged occupies a small part of the imaging scene

Inverse problems
Wave propagation

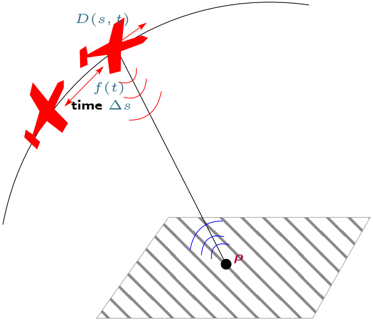


- Inverse problems aim to reconstruct a medium characteristics from knowledge of the response of the medium to a known incident field.
- In the context of the wave equation we seek to reconstruct the reflectivity by recording the medium's response to one or more known excitations.
- We consider sparse reflectivities : often true in applications when the object to be imaged occupies a small part of the imaging scene
- Study how uncertainty in the data affects the solution

Table of contents

Applications

Radar

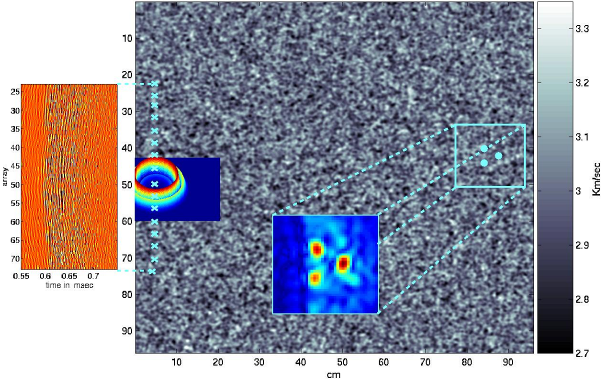


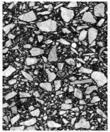
Here we want to retrieve the reflectivity on the ground. The measurements are obtained by sending pulses and collecting the echoes. The sensing matrix models how waves propagate from the airplane to the ground reflectors and back to the plane. The start-stop approximation is assumed.

Uncertainty in the data : measurement noise, measurement location

Applications

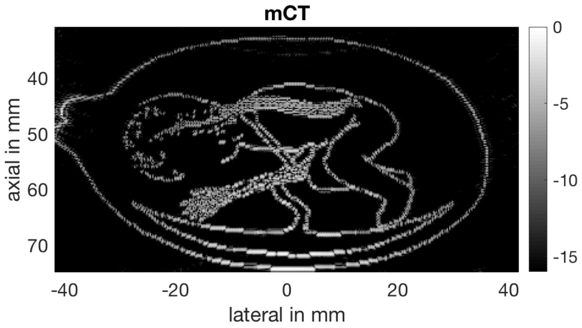
Non-destructive testing



Uncertainty : unknown micro-structure of material ; concrete sample 

Applications

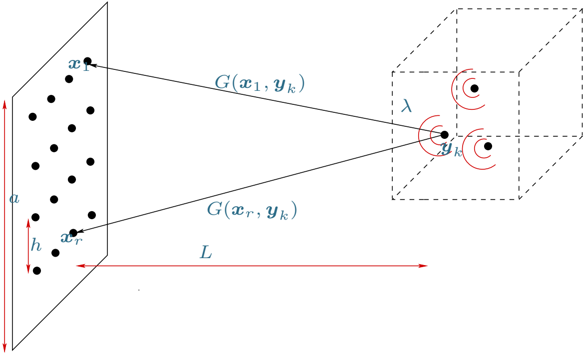
Medical Ultrasound



Uncertainty : measurement noise, low SNR

The imaging problem
Passive case

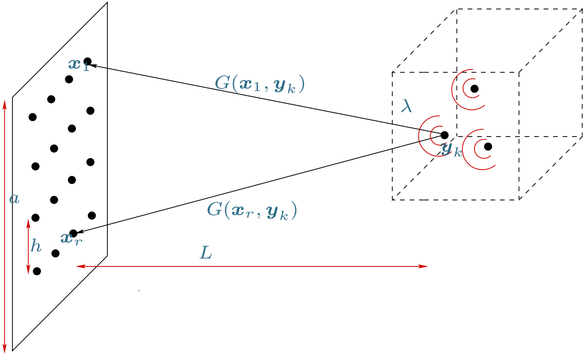
- We consider here the passive imaging problem, where we seek to reconstruct the positions $\boldsymbol{y}_j$ and complex-valued amplitudes ρ_j, $j = 1, \ldots, M$ of sources.



- There are N receivers at the array and K points in the image window IW.
  We assume that $N \ll K$ and that the unknown ρ has sparse support $M = \mathrm{supp}(\rho)$, $M < N$ and typically $M \ll K$.

The imaging problem
Passive case

- We consider here the passive imaging problem, where we seek to reconstruct the positions $\boldsymbol{y}_j$ and complex-valued amplitudes ρ_j, $j = 1, \ldots, M$ of sources.



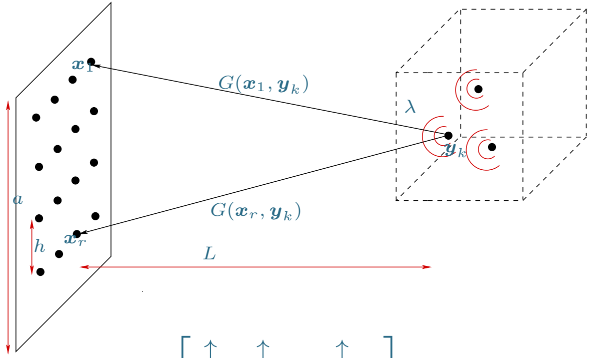
- The imaging problem consists in solving a linear system of the form

$$\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$$

with $\boldsymbol{b}$ the data, $\boldsymbol{\rho}$ the unknown, and $\mathcal{A}$ the sensing matrix

The imaging problem
The sensing matrix

- We consider here the passive imaging problem, where we seek to reconstruct the positions $\boldsymbol{y}_j$ and complex-valued amplitudes ρ_j, $j = 1, \ldots, M$ of sources.

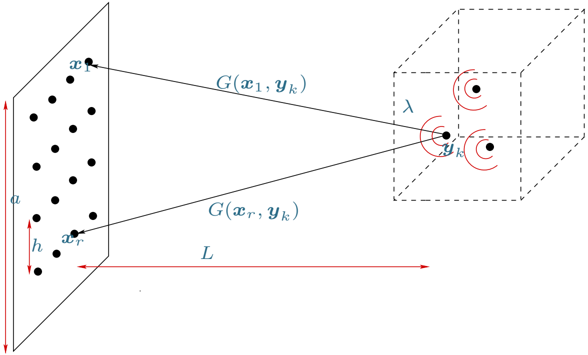


- In this case the sensing matrix is $\mathcal{A} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ a_1 & a_2 & & a_K \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$ with $a_k = g(\boldsymbol{y}_k)/\|g(\boldsymbol{y}_k)\|_2$ and

$$g(\boldsymbol{y}_k) = (G(\boldsymbol{x}_1, \boldsymbol{y}_k), G(\boldsymbol{x}_2, \boldsymbol{y}_k), \ldots, G(\boldsymbol{x}_N, \boldsymbol{y}_k))^T$$

The imaging problem
Active case

- We consider here the passive imaging problem, where we seek to reconstruct the positions $\boldsymbol{y}_j$ and complex-valued amplitudes ρ_j, $j = 1, \ldots, M$ of sources.



- The source imaging problem is considered here for simplicity. The active array imaging problem can be cast under the same linear algebra framework (i.e., solving $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$) even when multiple scattering is not negligible.

Table of contents

Compressive sensing

Our problem consist in finding the solution $\boldsymbol{\rho} \in \mathbb{C}^K$ of

$$\mathcal{A}\,\boldsymbol{\rho} = \boldsymbol{b}_\delta,$$

from highly incomplete ($1 \ll N < K$) measurement data $\boldsymbol{b}_\delta \in \mathbb{C}^N$

$$\boldsymbol{b}_\delta = \boldsymbol{b} + \boldsymbol{\delta b},$$

corrupted by the noise vector $\boldsymbol{\delta b}$.

There exist infinitely many solutions to our problem, and thus, it is a priori not possible to find the correct one without some additional information.

The sparsity of $\boldsymbol{\rho}$ changes the imaging problem substantially because we can formulate it as an optimization problem which seeks the sparsest vector in $\mathbb{C}^K$ that equates model and data.

Compressive sensing

$$\rho^{\ell_1}: \quad \min ||\rho||_1, \quad \text{subject to} \quad \mathcal{A}\rho = \boldsymbol{b}$$

Theorem[1] : In the noiseless case the ℓ_1 approach gives the exact solution ! under the assumption (mutual coherence) :

$$\max_{k \neq k'} |\langle \boldsymbol{a}_k, \boldsymbol{a}_{k'} \rangle| \leq 1/(2M), \quad \forall k, k' = 1, \dots, K.$$

Here a_k : normalized columns of $\mathcal{A}$, i.e., $\|a_k\|_2 = 1, \ \forall k$

Theorem[2] : The same result can be obtained assuming the sensing matrix $\mathcal{A}$ obeys the M-restricted isometry property which basically states that all sets of M-columns of $\mathcal{A}$ behave approximately as an orthonormal system.

In both cases the assumptions translate to conditions on the discretization of IW since in imaging $a_k = g(\boldsymbol{y}_k)/\|g(\boldsymbol{y}_k)\|_2$ with $g(\boldsymbol{y}_k, \omega_l) = (G(\boldsymbol{x}_1, \boldsymbol{y}_k), G(\boldsymbol{x}_2, \boldsymbol{y}_k), \dots, G(\boldsymbol{x}_N, \boldsymbol{y}_k))^T$

1. Donoho & Elad ‘03
2. Candès & Tao ‘05

Remarks

- A pair of columns $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ are coherent, if the corresponding grid-points are close to each other. Therefore the incoherence conditions can be satisfied only for coarse image discretizations that imply poor resolution.

Remarks

- A pair of columns $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ are coherent, if the corresponding grid-points are close to each other. Therefore the incoherence conditions can be satisfied only for coarse image discretizations that imply poor resolution.
- To achieve high resolution we propose to modify the classical theory so as to allow for some coherence in $\mathcal{A}$.

Remarks

- A pair of columns $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ are coherent, if the corresponding grid-points are close to each other. Therefore the incoherence conditions can be satisfied only for coarse image discretizations that imply poor resolution.
- To achieve high resolution we propose to modify the classical theory so as to allow for some coherence in $\mathcal{A}$.
- Let $\boldsymbol{\rho} \in \mathbb{C}^K$ be an M-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$. Define the index of its support T. For any $j \in T$ define the corresponding vicinity of $\boldsymbol{a}_j$ as

$$S_j = \left\{ k \ \text{ s.t. } \ |\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle| \geqslant \frac{1}{3M} \right\}.$$

Remarks

- A pair of columns $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ are coherent, if the corresponding grid-points are close to each other. Therefore the incoherence conditions can be satisfied only for coarse image discretizations that imply poor resolution.
- To achieve high resolution we propose to modify the classical theory so as to allow for some coherence in $\mathcal{A}$.
- Let $\boldsymbol{\rho} \in \mathbb{C}^K$ be an M-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$. Define the index of its support T. For any $j \in T$ define the corresponding vicinity of $\boldsymbol{a}_j$ as

$$S_j = \left\{ k \ \text{ s.t. } \ |\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle| \geqslant \frac{1}{3M} \right\}.$$

- For any vector $\boldsymbol{\xi} \in \mathbb{C}^K$ define its coherent misfit to $\boldsymbol{\rho}$ as

$$\mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\xi}) = \sum_{j \in T} \left| \rho_j - \sum_{k \in S_j} \langle \boldsymbol{a}_j, \boldsymbol{a}_k \rangle \xi_k \right|,$$

  and its incoherent remainder

$$\mathbf{In}(\boldsymbol{\rho}, \boldsymbol{\xi}) = \sum_{k \notin \Upsilon} |\xi_k|, \Upsilon = \cup_{j \in T} S_j.$$

Our abstract result

Theorem

Let $\boldsymbol{\rho}$ be an M-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$, and suppose vicinities S_j do not overlap. Let ρ_δ be the minimal ℓ_1-norm solution of the noisy problem

$$\min \|\boldsymbol{\rho}_\delta\|_{\ell_1}, \ \textit{subject to } \mathcal{A}\boldsymbol{\rho}_\delta = \boldsymbol{b}_\delta,$$

with $\|\boldsymbol{b} - \boldsymbol{b}_\delta\|_{\ell_2} \leqslant \delta$.
Then

$$\mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\rho}_\delta) \leqslant 3\gamma\delta,$$

and

$$\mathbf{In}(\boldsymbol{\rho}, \boldsymbol{\rho}_\delta) \leqslant 5\gamma\delta.$$

If the noise level $\delta = 0$, and the number of vectors in $\Upsilon = \cup_{j \in T} S_j$ is smaller than the rank of $\mathcal{A}$, we have exact recovery : $\boldsymbol{\rho}_\delta = \boldsymbol{\rho}$.

Comments/Remarks

- This theorem says that when the vicinities associated to different sources do not overlap, ℓ_1 minimization determines their location with high precision.

Comments/Remarks

- This theorem says that when the vicinities associated to different sources do not overlap, ℓ_1 minimization determines their location with high precision.
- Allowing for the columns of $\mathcal{A}$ to be coherent inside the vicinities is crucial as it permits for a fine discretization to be used which in turn implies high resolution imaging.

Comments/Remarks

- This theorem says that when the vicinities associated to different sources do not overlap, ℓ_1 minimization determines their location with high precision.
- Allowing for the columns of $\mathcal{A}$ to be coherent inside the vicinities is crucial as it permits for a fine discretization to be used which in turn implies high resolution imaging.
- When there is noise in the data the location of the sources is not retrieved exactly but both the coherent and the incoherent misfit are small.

Comments/Remarks

- This theorem says that when the vicinities associated to different sources do not overlap, ℓ_1 minimization determines their location with high precision.
- Allowing for the columns of $\mathcal{A}$ to be coherent inside the vicinities is crucial as it permits for a fine discretization to be used which in turn implies high resolution imaging.
- When there is noise in the data the location of the sources is not retrieved exactly but both the coherent and the incoherent misfit are small.
- The incoherent misfit concerns the support of $\boldsymbol{\rho}_\delta$ far from the true sources location, *i.e.*, outside the vicinities. This type of noise in the reconstruction image is usually referred to as grass.

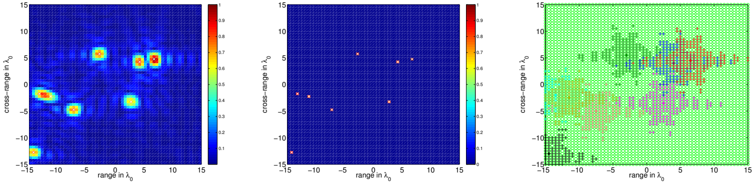
Comments/Remarks

- This theorem says that when the vicinities associated to different sources do not overlap, ℓ_1 minimization determines their location with high precision.
- Allowing for the columns of $\mathcal{A}$ to be coherent inside the vicinities is crucial as it permits for a fine discretization to be used which in turn implies high resolution imaging.
- When there is noise in the data the location of the sources is not retrieved exactly but both the coherent and the incoherent misfit are small.
- The incoherent misfit concerns the support of $\boldsymbol{\rho}_\delta$ far from the true sources location, *i.e.*, outside the vicinities. This type of noise in the reconstruction image is usually referred to as grass.
- The coherent misfit concerns the approximation of $\boldsymbol{\rho}$ by degrees of freedom inside the vicinities. The coherent misfit being small means that $\boldsymbol{\rho}$ can be well approximated by a linear combination of the values of $\boldsymbol{\rho}_\delta$ inside the vicinities of the true support.

ℓ_1 and super-resolution

- Exact recovery in the noiseless case

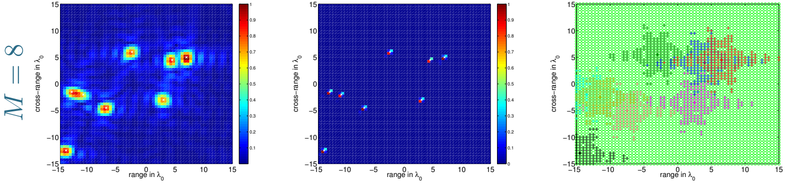
$M = \infty$



Vicinities overlap

ℓ_1 and super-resolution

- Exact recovery in the noiseless case



Vicinities overlap

- Robustness to noise (under smallness assumption) : the solution is decomposed in two parts, the coherent part, $\boldsymbol{\rho}_c$, which is supported in the vicinities S_j and is close to the true $\boldsymbol{\rho}$ and the incoherent part, $\boldsymbol{\rho}_i$, usually referred to as grass. The grass is supported away from the vicinities S_j and is shown to be small.

ℓ_1 and super-resolution

- Other works where super-resolution has been studied for highly coherent model matrices $\mathcal{A}$ that arise in imaging

A. Fannjiang and W. Liao, *Coherence pattern-guided compressive sensing with unresolved grids*, SIAM J. Imaging Sci. 5 (2012), pp. 179–202.

L. Borcea and I. Kocyigit, *Resolution analysis of imaging with ℓ_1 optimization*, SIAM J. Imaging Sci. 8 (2015), pp. 3015–3050.

L. Borcea and I. Kocyigit, *A multiple measurement vector approach to synthetic aperture radar imaging*, SIAM J. Imaging Sci. 11 (2018), pp. 770–801.

These works include results regarding the robustness of super-resolution in the presence of noise.

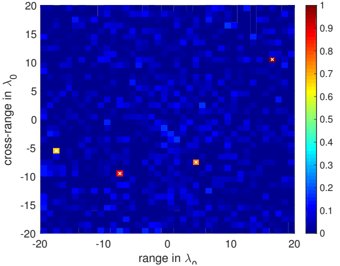
Table of contents

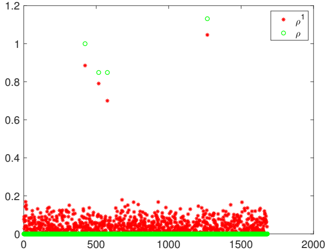
What about uncertainty in the data (noise)

Our theory implies that a key to control the noise is the constant γ,

$$\gamma = \sup_{\boldsymbol{b}} \frac{\|\boldsymbol{\rho}\|_{\ell_1}}{\|\boldsymbol{b}\|_{\ell_2}} \text{ where } \boldsymbol{\rho} \text{ is the minimal } \ell_1 - \text{solution of } \mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}.$$

For the imaging problem we have $\gamma = O(\sqrt{N})$ which is not satisfactory.





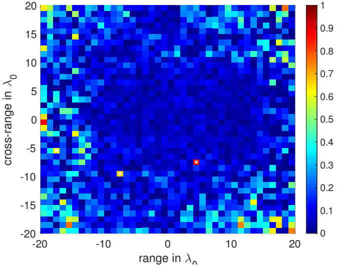
$N = 625$ measurements. $K = 1681$ pixels in the images. 100% noise.

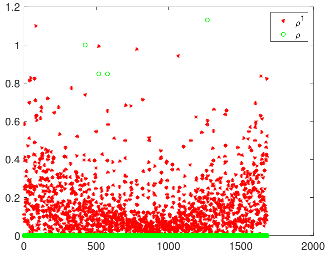
What about uncertainty in the data (noise)

Our theory implies that a key to control the noise is the constant γ,

$$\gamma = \sup_{\boldsymbol{b}} \frac{\|\boldsymbol{\rho}\|_{\ell_1}}{\|\boldsymbol{b}\|_{\ell_2}} \text{ where } \boldsymbol{\rho} \text{ is the minimal } \ell_1 - \text{solution of } \mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}.$$

For the imaging problem we have $\gamma = O(\sqrt{N})$ which is not satisfactory.





$N = 1369$ measurements. $K = 1681$ pixels in the images. 100% noise.

What about uncertainty in the data (noise)

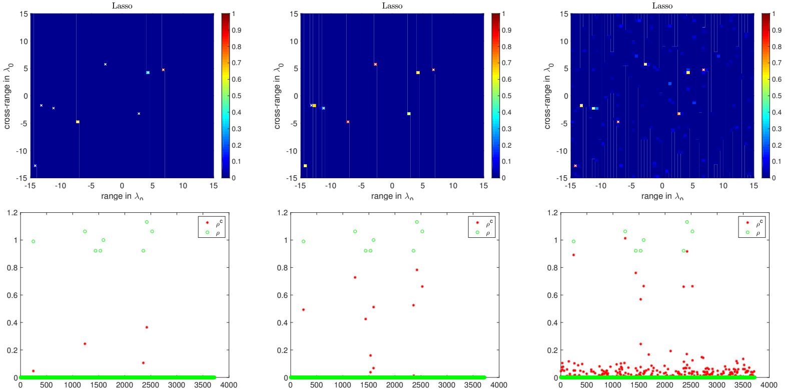
The most commonly used approach to deal with noise is to solve the ℓ_2-relaxed form of the optimization problem

$$\boldsymbol{\rho}_\lambda = \arg\min_{\boldsymbol{\rho}} \left(\lambda \|\boldsymbol{\rho}\|_{\ell_1} + \frac{1}{2} \|\boldsymbol{\mathcal{A}}\boldsymbol{\rho} - \boldsymbol{b}\|^2_{\ell_2} \right), \tag{1}$$

known as Lasso in the statistics literature [R. Tibshirani '96, Chen & D.Donoho '94, F.Santosa & W.Symes '86].

There are sufficient conditions, depending on the SNR and the value of λ (tuning parameter which should be adequately chosen), for the support of $\boldsymbol{\rho}_\lambda$ to be contained within the true support.

Tuning λ in Lasso



LASSO results with $\lambda = 1$, $\lambda = 0.5$ (optimal) and $\lambda = 0.1$.
100% noise. $N = 625$, $K = 3721$

▷ Develop a method for exact support recovery in the presence of noise. No tuning parameters. No a priory knowledge on the level of noise required.

The Noise Collector

We propose to solve instead

$$(\boldsymbol{\rho}_\tau, \boldsymbol{\eta}_\tau) = \arg\min_{\boldsymbol{\rho},\boldsymbol{\eta}} \left(\tau \|\boldsymbol{\rho}\|_{l_1} + \|\boldsymbol{\eta}\|_{l_1}\right), \\ \text{subject to } \mathcal{A}\boldsymbol{\rho} + \mathcal{C}\boldsymbol{\eta} = \boldsymbol{b} + \boldsymbol{\delta b},$$

where $\mathcal{C}$ is the *Noise Collector* matrix $\mathcal{C} \in \mathbb{C}^{N\times\Sigma}$, $\Sigma \gg K$ and τ is an $O(1)$ weight that is independent of the dimension of the problem and the level of noise.

This minimization problem can be understood as a relaxation of

$$\boldsymbol{\rho}_* = \arg\min_{\boldsymbol{\rho}} \|\boldsymbol{\rho}\|_{\ell_1}, \text{ subject to } \mathcal{A}\,\boldsymbol{\rho} = \boldsymbol{b} + \boldsymbol{\delta b},$$

It works by absorbing *all* the noise, and possibly some signal, in $\mathcal{C}\boldsymbol{\eta}_\tau$.

η does not correspond to a physical quantity. It is introduced to provide a fictitious source distribution (an appropriate linear combination of the columns of $\mathcal{C}$) that produces a good approximation to $\delta\boldsymbol{b}$.

Construction of the Noise Collector

(i) Columns of $\mathcal{C}$ should be sufficiently orthogonal to the columns of $\mathcal{A}$, so it does not absorb signals with meaningful information.

(ii) Columns of $\mathcal{C}$ should be uniformly distributed on the unit sphere $\mathbb{S}^{N-1}$ so that we could approximate well a typical noise vector.

(iii) The number of columns of $\mathcal{C}$ should grow slower than exponential with N, otherwise the method is impractical.

One way to guarantee all three properties is the *deterministic* approach that consists in filling up $\mathcal{C}$ imposing

$$|\langle \boldsymbol{a}_i, \boldsymbol{c}_j \rangle| < \frac{\alpha}{\sqrt{N}} \; \forall i, j \,, \text{ and } |\langle \boldsymbol{c}_i, \boldsymbol{c}_j \rangle| < \frac{\alpha}{\sqrt{N}} \; \forall i \neq j,$$

with $\alpha > 1$. Then it can be shown that the number Σ of columns in $\mathcal{C}$ grows at most polynomially : $N^{\alpha} \leqslant \Sigma \leqslant N^{\alpha^2}$.

We consider instead the *random* approach : the columns of $\mathcal{C}$ are drawn at random independently. Then the above inequalities hold but decoherence constraint is weakened by a logarithmic factor $\sqrt{\ln N}$.

Probabilistic design of NC

Choose $\beta > 1$, and pick $\Sigma = N^\beta$ vectors $\boldsymbol{c}_i$ at random and independently on $\mathbb{S}^{N-1}$.

Probabilistic design of NC

Choose $\beta > 1$, and pick $\Sigma = N^\beta$ vectors $\boldsymbol{c}_i$ at random and independently on $\mathbb{S}^{N-1}$.

Define the convex hulls

$$H_1 = \left\{ x \in \mathbb{R}^N \middle| x = \sum_{i=1}^{\Sigma} \xi_i \boldsymbol{c}_i, \ \sum_{i=1}^{\Sigma} |\xi_i| \leqslant 1 \right\}, \quad H_2 = \left\{ x \in \mathbb{R}^N \middle| x = \sum_{i=1}^{K} \xi_i \boldsymbol{a}_i, \ \sum_{i=1}^{\Sigma} |\xi_i| \leqslant 1 \right\},$$

Probabilistic design of NC

Choose $\beta > 1$, and pick $\Sigma = N^\beta$ vectors $\boldsymbol{c}_i$ at random and independently on $\mathbb{S}^{N-1}$.

Define the convex hulls

$$H_1 = \left\{ x \in \mathbb{R}^N \middle| x = \sum_{i=1}^{\Sigma} \xi_i \boldsymbol{c}_i, \ \sum_{i=1}^{\Sigma} |\xi_i| \leqslant 1 \right\}, \quad H_2 = \left\{ x \in \mathbb{R}^N \middle| x = \sum_{i=1}^{K} \xi_i \boldsymbol{a}_i, \ \sum_{i=1}^{\Sigma} |\xi_i| \leqslant 1 \right\},$$

Then, for any $\kappa > 0$ there exist constants $\alpha = \sqrt{(\beta-1)/2}$, $c_0(\kappa, \beta)$ and $N_0 = N_0(\kappa, \beta)$, such that $\forall N > N_0$

$$(i) \quad \max\left(\max_{i \leqslant K}(|\langle \boldsymbol{a}_i, \boldsymbol{\delta b}\rangle|), \max_{i \leqslant \Sigma}(|\langle \boldsymbol{c}_i, \boldsymbol{\delta b}\rangle|)\right) < c_0 \frac{\sqrt{\ln N}}{\sqrt{N}},$$

and

$$(ii) \quad \alpha \frac{\sqrt{\ln N}}{\sqrt{N}} \boldsymbol{\delta b} \in H_1,$$

with probability $1 - 1/N^\kappa$.

Probabilistic design of NC

Choose $\beta > 1$, and pick $\Sigma = N^\beta$ vectors $\boldsymbol{c}_i$ at random and independently on $\mathbb{S}^{N-1}$.

Define the convex hulls

$$H_1 = \left\{ x \in \mathbb{R}^N \,\middle|\, x = \sum_{i=1}^{\Sigma} \xi_i \boldsymbol{c}_i,\ \sum_{i=1}^{\Sigma} |\xi_i| \leqslant 1 \right\}, \quad H_2 = \left\{ x \in \mathbb{R}^N \,\middle|\, x = \sum_{i=1}^{K} \xi_i \boldsymbol{a}_i,\ \sum_{i=1}^{\Sigma} |\xi_i| \leqslant 1 \right\},$$

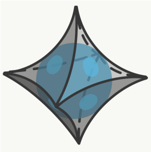
Then, for any $\kappa > 0$ there exist constants $\alpha = \sqrt{(\beta-1)/2}$, $c_0(\kappa,\beta)$ and $N_0 = N_0(\kappa,\beta)$, such that $\forall N > N_0$

$$(i) \quad \max\left(\max_{i\leqslant K}(|\langle \boldsymbol{a}_i, \boldsymbol{\delta b}\rangle|), \max_{i\leqslant \Sigma}(|\langle \boldsymbol{c}_i, \boldsymbol{\delta b}\rangle|)\right) < c_0 \frac{\sqrt{\ln N}}{\sqrt{N}},$$

and

$$(ii) \quad \alpha \frac{\sqrt{\ln N}}{\sqrt{N}} \boldsymbol{\delta b} \in H_1,$$

with probability $1 - 1/N^\kappa$.



(i) states that H_1 and H_2 are contained in the ℓ_2-ball of radius $c_0\sqrt{\ln N}/\sqrt{N}$ except for a few spikes in statistically insignificant directions.

Probabilistic design of NC

Choose $\beta > 1$, and pick $\Sigma = N^{\beta}$ vectors $\boldsymbol{c}_i$ at random and independently on $\mathbb{S}^{N-1}$.

Define the convex hulls

$$H_1 = \left\{ x \in \mathbb{R}^N \middle| x = \sum_{i=1}^{\Sigma} \xi_i \boldsymbol{c}_i, \ \sum_{i=1}^{\Sigma} |\xi_i| \leqslant 1 \right\}, \quad H_2 = \left\{ x \in \mathbb{R}^N \middle| x = \sum_{i=1}^{K} \xi_i \boldsymbol{a}_i, \ \sum_{i=1}^{\Sigma} |\xi_i| \leqslant 1 \right\},$$

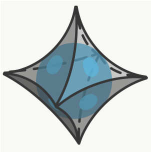
Then, for any $\kappa > 0$ there exist constants $\alpha = \sqrt{(\beta-1)/2}$, $c_0(\kappa, \beta)$ and $N_0 = N_0(\kappa, \beta)$, such that $\forall N > N_0$

$$(i) \quad \max(\max_{i \leqslant K}(|\langle \boldsymbol{a}_i, \boldsymbol{\delta b} \rangle|), \max_{i \leqslant \Sigma}(|\langle \boldsymbol{c}_i, \boldsymbol{\delta b} \rangle|)) < c_0 \frac{\sqrt{\ln N}}{\sqrt{N}},$$

and

$$(ii) \quad \alpha \frac{\sqrt{\ln N}}{\sqrt{N}} \boldsymbol{\delta b} \in H_1,$$

with probability $1 - 1/N^{\kappa}$.



(i) states that H_1 and H_2 are contained in the ℓ_2-ball of radius $c_0\sqrt{\ln N}/\sqrt{N}$ except for a few spikes in statistically insignificant directions.

(ii) states that H_1 contains an ℓ_2-ball of radius $\alpha\sqrt{\ln N}/\sqrt{N}$ except for a few statistically insignificant directions.

False Discovery Rate is zero

We have the following results :

Theorem (No phantom signal theorem)

Suppose there is no signal : $\boldsymbol{\rho} = 0$ *and* $\boldsymbol{\delta b}/\|\boldsymbol{\delta b}\|_{l_2}$ *is uniformly distributed on the unit sphere. Fix* $\beta > 1$*, and draw* $\Sigma = N^{\beta}$ *columns for* $\mathcal{C}$*, independently, from the uniform distribution on* $\mathbb{S}^{N-1}$*. For any* $\kappa > 0$ *there are constants* $\tau = \tau(\kappa, \beta)$ *and* $N_0 = N_0(\kappa, \beta)$ *such that,* $\forall N > N_0$*,* $\boldsymbol{\rho}_{\tau} = 0$ *with probability* $1 - 1/N^{\kappa}$*.*

We see no phantom signals when the algorithm is fed with pure noise. The no phantom weight τ may be chosen to be the smallest constant such that the no phantom signal theorem holds.

False Discovery Rate is zero

We have the following results :

Theorem (No phantom signal theorem)

Suppose there is no signal : $\boldsymbol{\rho} = 0$ and $\boldsymbol{\delta b}/\|\boldsymbol{\delta b}\|_{l_2}$ is uniformly distributed on the unit sphere. Fix $\beta > 1$, and draw $\Sigma = N^{\beta}$ columns for $\mathcal{C}$, independently, from the uniform distribution on $\mathbb{S}^{N-1}$. For any $\kappa > 0$ there are constants $\tau = \tau(\kappa, \beta)$ and $N_0 = N_0(\kappa, \beta)$ such that, $\forall N > N_0$, $\boldsymbol{\rho}_{\tau} = 0$ with probability $1 - 1/N^{\kappa}$.

We see no phantom signals when the algorithm is fed with pure noise. The no phantom weight τ may be chosen to be the smallest constant such that the no phantom signal theorem holds.

Theorem

Let $\boldsymbol{\rho}$ be an M-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$. Assume κ, β, the Noise Collector, and the noise are the same as in Theorem above. If the columns of $\mathcal{A}$ are decoherent : $|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| \leqslant \frac{1}{3M}$, then $supp(\boldsymbol{\rho}_{\tau}) \subseteq supp(\boldsymbol{\rho})$ for all $N > N_0$ with probability $1 - 1/N^{\kappa}$.

This means that we have zero false discovery rate for any level of noise and with probability that tends to one as the dimension of the data increases to infinity.
No false positives !

Supports of $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_\tau$ agree

Theorem

Suppose r *is the magnitude of smallest non-zero entry of* $\boldsymbol{\rho}$. *If* $\|\boldsymbol{\delta b}\|_{l_2} \leqslant c_2 \frac{\|\boldsymbol{b}\|_{l_2}^2}{\|\boldsymbol{\rho}\|_{l_1}} \frac{\sqrt{N}}{\sqrt{\ln N}}$, $c_2 = c_2(\kappa, \beta, r, M)$, *then for all* $N > N_0$, $supp(\boldsymbol{\rho}_\tau) = supp(\boldsymbol{\rho})$, *with probability* $1 - 1/N^\kappa$.

Exact support recovery when the noise is not too large.

Supports of $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_\tau$ agree

Theorem

Suppose r *is the magnitude of smallest non-zero entry of* $\boldsymbol{\rho}$*. If* $\|\boldsymbol{\delta b}\|_{l_2} \leqslant c_2 \frac{\|\boldsymbol{b}\|_{l_2}^2}{\|\boldsymbol{\rho}\|_{l_1}} \frac{\sqrt{N}}{\sqrt{\ln N}}$*,* $c_2 = c_2(\kappa, \beta, r, M)$*, then for all* $N > N_0$*,* $supp(\boldsymbol{\rho}_\tau) = supp(\boldsymbol{\rho})$*, with probability* $1 - 1/N^\kappa$*.*

Exact support recovery when the noise is not too large.

Theorem (Exact Recovery)

If there is no noise $\boldsymbol{\delta b} = 0$*. Then,* $\boldsymbol{\rho}_\tau = \boldsymbol{\rho}$ *for all* $M < \frac{2\sqrt{N}}{3c_0\tau\sqrt{\ln N}}$ *with probability* $1 - 1/N^\kappa$*.*

Exact recovery in the noise free case.

GelMa

To find the minimizer $\boldsymbol{\rho}_\tau, \boldsymbol{\eta}_\tau$,

$$(\boldsymbol{\rho}_\tau, \boldsymbol{\eta}_\tau) = \arg\min_{\boldsymbol{\rho},\boldsymbol{\eta}} \left(\tau\|\boldsymbol{\rho}\|_{l_1} + \|\boldsymbol{\eta}\|_{l_1}\right),$$
$$\text{subject to } \mathcal{A}\boldsymbol{\rho} + \mathcal{C}\boldsymbol{\eta} = \boldsymbol{b} + \boldsymbol{\delta b},$$

we consider a variational approach. We define the function

$$F(x, \eta, z) = \lambda\left(\tau\|x\|_{\ell_1} + \|\eta\|_{\ell_1}\right) + \frac{1}{2}\|\mathcal{A}x + \mathcal{C}\eta - b\|^2_{\ell_2} + \langle z, b - \mathcal{A}x - \mathcal{C}\eta\rangle$$

for an $O(1)$ weight τ, and determine the solution as

$$\max_{z} \min_{x,\eta} F(x, \eta, z).$$

This variational principle finds the minimum exactly for all values of the regularization parameter λ.

GelMa

To determine the exact extremum, we use the iterative soft thresholding algorithm GeLMA which is a semi-implicit version of the primal-dual method of Chambolle & Pock.

M. Moscoso, A. Novikov, G. Papanicolaou and L. Ryzhik, A differential equations approach to l1-minimization with applications to array imaging, Inverse Problems 28 (2012).

A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, Journal of Mathematical Imaging and Vision 40 (2011), pp 120–145.

GelMa

Pick a value for the regularization parameter λ, e.g. $\lambda = 1$. Choose step sizes $\Delta t_1 < 2/\|[\mathcal{A}\,|\,\mathcal{C}]\|^2$ and $\Delta t_2 < \lambda/\|\mathcal{A}\|$. Set $\boldsymbol{\rho}_0 = \mathbf{0}$, $\boldsymbol{\eta}_0 = \mathbf{0}$, $\boldsymbol{z}_0 = \mathbf{0}$, and iterate for $k \geqslant 0$:

$$
\begin{aligned}
\boldsymbol{r} &= \boldsymbol{b} - \mathcal{A}\,\boldsymbol{\rho}_k - \mathcal{C}\,\boldsymbol{\eta}_k\,,\\
\boldsymbol{\rho}_{k+1} &= \mathcal{S}_{\tau\,\lambda\Delta t_1}\left(\boldsymbol{\rho}_k + \Delta t_1\,\mathcal{A}^*(\boldsymbol{z}_k + \boldsymbol{r})\right)\,,\\
\boldsymbol{\eta}_{k+1} &= \mathcal{S}_{\lambda\Delta t_1}\left(\boldsymbol{\eta}_k + \Delta t_1\,\mathcal{C}^*(\boldsymbol{z}_k + \boldsymbol{r})\right)\,,\\
\boldsymbol{z}_{k+1} &= \boldsymbol{z}_k + \Delta t_2\,\boldsymbol{r}\,,
\end{aligned}
$$

where $\mathcal{S}_\lambda(y_i) = \text{sign}(y_i)\max\{0, |y_i| - \lambda\}$.

Note : Choosing two step sizes instead of the smaller one Δt_1 improves the convergence speed.

Cost of the Noise Collector

- The *Noise Collector* matrix $\mathcal{C}$ is constructed theoretically by drawing $\Sigma = N^\beta$ normally distributed N-dimensional vectors, normalized to unit length.

Cost of the Noise Collector

- The *Noise Collector* matrix $\mathcal{C}$ is constructed theoretically by drawing $\Sigma = N^{\beta}$ normally distributed N-dimensional vectors, normalized to unit length.
- In practice we construct $\mathcal{C}$ by drawing $N^{\beta-1}$ normally distributed N-dimensional vectors. These are the generating vectors of the *Noise Collector*. From each one of them a circulant $N \times N$ matrix $\mathcal{C}_i$, $i = 1, \ldots, N^{\beta-1}$, is constructed. Typically $\beta \approx 1.5$.

Cost of the Noise Collector

- The *Noise Collector* matrix $\mathcal{C}$ is constructed theoretically by drawing $\Sigma = N^{\beta}$ normally distributed N-dimensional vectors, normalized to unit length.
- In practice we construct $\mathcal{C}$ by drawing $N^{\beta-1}$ normally distributed N-dimensional vectors. These are the generating vectors of the *Noise Collector*. From each one of them a circulant $N \times N$ matrix $\mathcal{C}_i$, $i = 1, \dots, N^{\beta-1}$, is constructed. Typically $\beta \approx 1.5$.
- The full *Noise Collector* matrix can be constructed by concatenation,

$$\mathcal{C} = [\mathcal{C}_1 \,|\mathcal{C}_2 \,|\dots|\mathcal{C}_{N^{\beta-1}} \,]$$

The full Noise Collector matrix is never formed, only the $N^{\beta-1}$ generating vectors are stored.

Cost of the Noise Collector

- The *Noise Collector* matrix $\mathcal{C}$ is constructed theoretically by drawing $\Sigma = N^\beta$ normally distributed N-dimensional vectors, normalized to unit length.
- In practice we construct $\mathcal{C}$ by drawing $N^{\beta-1}$ normally distributed N-dimensional vectors. These are the generating vectors of the *Noise Collector*. From each one of them a circulant $N \times N$ matrix $\mathcal{C}_i$, $i = 1, \ldots, N^{\beta-1}$, is constructed. Typically $\beta \approx 1.5$.
- The full *Noise Collector* matrix can be constructed by concatenation,

$$\mathcal{C} = [\mathcal{C}_1 \,|\mathcal{C}_2 \,| \ldots |\mathcal{C}_{N^{\beta-1}} \,]$$

  The full Noise Collector matrix is never formed, only the $N^{\beta-1}$ generating vectors are stored.
- Exploiting the circulant structure of the matrices $\mathcal{C}_i$, we perform the matrix vector multiplications $\mathcal{C}\boldsymbol{\eta}_k$ and $\mathcal{C}^*(\boldsymbol{z}_k + \boldsymbol{r})$ using the FFT.

Cost of the Noise Collector

- The *Noise Collector* matrix $\mathcal{C}$ is constructed theoretically by drawing $\Sigma = N^\beta$ normally distributed N-dimensional vectors, normalized to unit length.
- In practice we construct $\mathcal{C}$ by drawing $N^{\beta-1}$ normally distributed N-dimensional vectors. These are the generating vectors of the *Noise Collector*. From each one of them a circulant $N \times N$ matrix $\mathcal{C}_i$, $i = 1, \ldots, N^{\beta-1}$, is constructed. Typically $\beta \approx 1.5$.
- The full *Noise Collector* matrix can be constructed by concatenation,

$$\mathcal{C} = [\mathcal{C}_1 \,|\mathcal{C}_2 \,|\ldots|\mathcal{C}_{N^{\beta-1}}\ ]$$

  The full Noise Collector matrix is never formed, only the $N^{\beta-1}$ generating vectors are stored.
- Exploiting the circulant structure of the matrices $\mathcal{C}_i$, we perform the matrix vector multiplications $\mathcal{C}\boldsymbol{\eta}_k$ and $\mathcal{C}^*(\boldsymbol{z}_k + \boldsymbol{r})$ using the FFT.
- This makes the complexity associated to the *Noise Collector* $O(N^\beta \log(N))$.

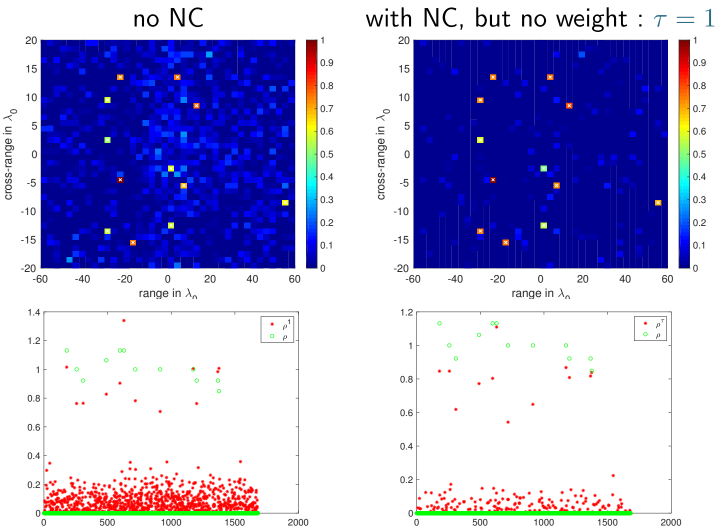
Cost of the Noise Collector

- The *Noise Collector* matrix $\mathcal{C}$ is constructed theoretically by drawing $\Sigma = N^{\beta}$ normally distributed N-dimensional vectors, normalized to unit length.
- In practice we construct $\mathcal{C}$ by drawing $N^{\beta-1}$ normally distributed N-dimensional vectors. These are the generating vectors of the *Noise Collector*. From each one of them a circulant $N \times N$ matrix $\mathcal{C}_i$, $i = 1, \ldots, N^{\beta-1}$, is constructed. Typically $\beta \approx 1.5$.
- The full *Noise Collector* matrix can be constructed by concatenation,

$$\mathcal{C} = [\mathcal{C}_1 \,|\mathcal{C}_2 \,|\ldots|\mathcal{C}_{N^{\beta-1}} \,]$$

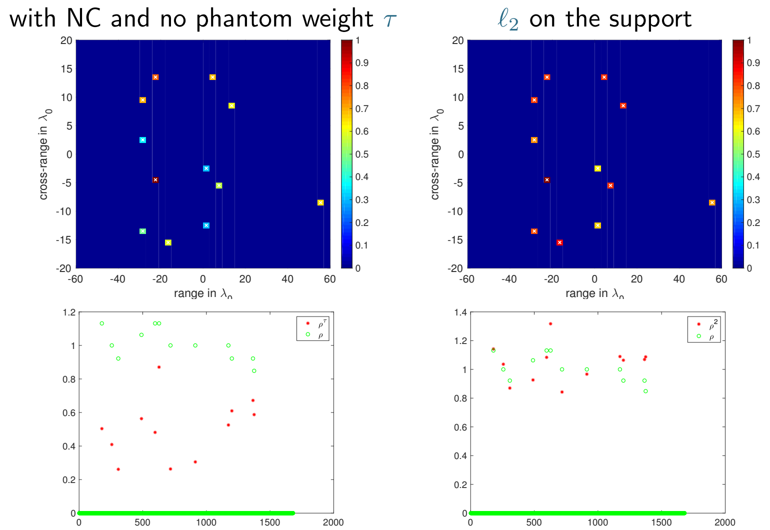
  The full Noise Collector matrix is never formed, only the $N^{\beta-1}$ generating vectors are stored.
- Exploiting the circulant structure of the matrices $\mathcal{C}_i$, we perform the matrix vector multiplications $\mathcal{C}\boldsymbol{\eta}_k$ and $\mathcal{C}^*(\boldsymbol{z}_k + \boldsymbol{r})$ using the FFT.
- This makes the complexity associated to the *Noise Collector* $O(N^{\beta} \log(N))$.
- The cost of performing the matrix vector multiplication $\mathcal{A}\,\boldsymbol{\rho}_k$ is NK and, thus, the additional cost due to the Noise Collector is negligible as, typically, $K \gg N^{\beta-1}$.

Results



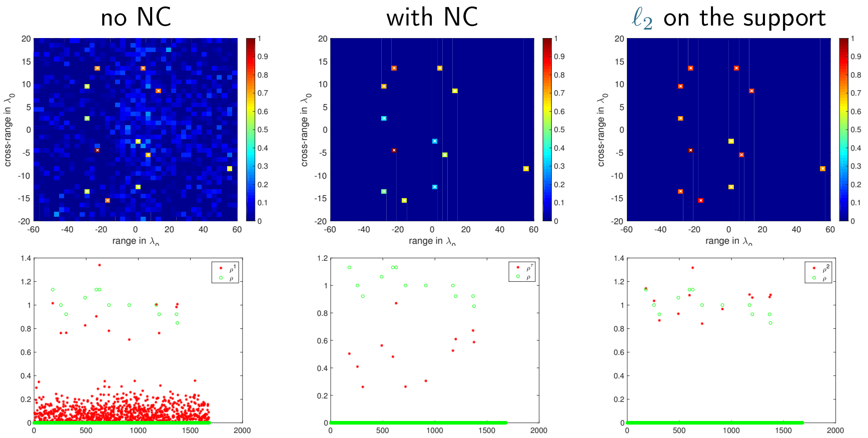
Top row : the images. Bottom row : solution vector with red stars and the true solution vector with green circles.

Results



Top row : the images. Bottom row : solution vector with red stars and the true solution vector with green circles.

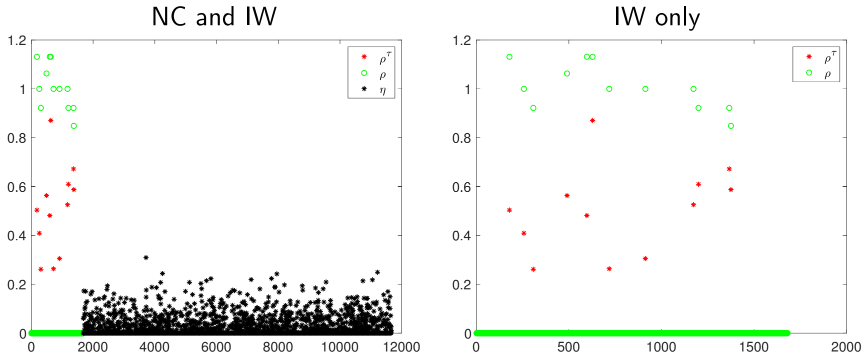
Results



Top row : the images. Bottom row : solution vector with red stars and the true solution vector with green circles.

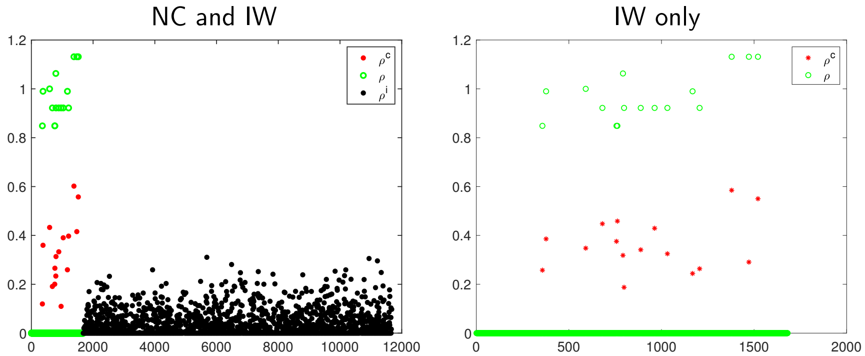
$$\beta = 1.5,\ SNR = \frac{\|\boldsymbol{b}\|_{l_2}}{\|\boldsymbol{\delta b}\|_{l_2}} = 1,\ N = 625,\ K = 1681,\ \tau = 2.$$

The Noise Collector at work



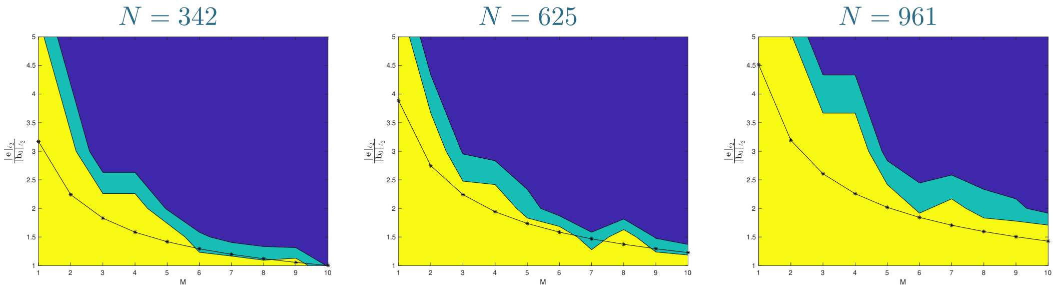
Coefficients of the solution. $M = 12$

The Noise Collector at work



Coefficients of the solution. $M = 16$

Failure to recover



Algorithm performance for exact support recovery. Success corresponds to the value 1 (yellow) and failure to 0 (blue). The small phase transition zone (green) contains intermediate values. Ordinate and abscissa are the sparsity M and $\frac{\|\boldsymbol{\delta b}\|_2}{\|\boldsymbol{b}\|_2}$. The black line is $\frac{1}{\sqrt{\ln N}} \frac{\sqrt{N}}{\sqrt{M}}$.

Table of contents

Concluding remarks

- Provided a theoretical framework to examine under what conditions the ℓ_1-minimization problem admits a solution that is close to the exact one when coherence in the columns of $\mathcal{A}$ is allowed inside the vicinities.

Concluding remarks

- Provided a theoretical framework to examine under what conditions the ℓ_1-minimization problem admits a solution that is close to the exact one when coherence in the columns of $\mathcal{A}$ is allowed inside the vicinities.

- To increase the robustness of ℓ_1-minimization when imaging with noisy data we propose to solve

$$(\boldsymbol{\rho}_\tau, \boldsymbol{\eta}_\tau) = \arg\min_{\boldsymbol{\rho},\boldsymbol{\eta}} \left(\tau\|\boldsymbol{\rho}\|_{l_1} + \|\boldsymbol{\eta}\|_{l_1}\right), \quad \text{subject to } \mathcal{A}\boldsymbol{\rho} + \mathcal{C}\boldsymbol{\eta} = \boldsymbol{b} + \boldsymbol{\delta b},$$

instead of

$$\boldsymbol{\rho}_* = \arg\min_{\boldsymbol{\rho}} \|\boldsymbol{\rho}\|_{\ell_1}, \text{ subject to } \mathcal{A}\,\boldsymbol{\rho} = \boldsymbol{b} + \boldsymbol{\delta b},$$

Here the $N \times N^\beta$ matrix $\mathcal{C}$ is the *Noise Collector* matrix and τ is an $O(1)$ weight for *any* level of noise.

Concluding remarks

- Provided a theoretical framework to examine under what conditions the ℓ_1-minimization problem admits a solution that is close to the exact one when coherence in the columns of $\mathcal{A}$ is allowed inside the vicinities.

- To increase the robustness of ℓ_1-minimization when imaging with noisy data we propose to solve

$$\begin{aligned}(\boldsymbol{\rho}_\tau, \boldsymbol{\eta}_\tau) = \arg\min_{\boldsymbol{\rho},\boldsymbol{\eta}} \left(\tau\|\boldsymbol{\rho}\|_{l_1} + \|\boldsymbol{\eta}\|_{l_1}\right), \\ \text{subject to } \mathcal{A}\boldsymbol{\rho} + \mathcal{C}\boldsymbol{\eta} = \boldsymbol{b} + \boldsymbol{\delta b},\end{aligned}$$

  instead of

$$\boldsymbol{\rho}_* = \arg\min_{\boldsymbol{\rho}} \|\boldsymbol{\rho}\|_{\ell_1}, \text{ subject to } \mathcal{A}\,\boldsymbol{\rho} = \boldsymbol{b} + \boldsymbol{\delta b},$$

  Here the $N \times N^\beta$ matrix $\mathcal{C}$ is the *Noise Collector* matrix and τ is an $O(1)$ weight for *any* level of noise.

- The additional cost of the Noise Collector is negligible compared to the cost of solving the original problem.

Concluding remarks

- The most important features of the Noise Collector are that :

Concluding remarks

- The most important features of the Noise Collector are that :
  1. no calibration is necessary with respect to the level of noise,

Concluding remarks

- The most important features of the Noise Collector are that :

  1. no calibration is necessary with respect to the level of noise,
  2. exact support recovery is obtained for relatively large levels of noise

$$\|\delta b\|_{\ell_2} \leqslant c_1 \frac{\|\boldsymbol{b}\|_{\ell_2}^2 \sqrt{N}}{\|\boldsymbol{\rho}\|_{\ell_1} \sqrt{\ln N}},$$

Concluding remarks

- The most important features of the Noise Collector are that :

  1. no calibration is necessary with respect to the level of noise,
  2. exact support recovery is obtained for relatively large levels of noise

     $$\|\delta b\|_{\ell_2} \leqslant c_1 \frac{\|\boldsymbol{b}\|_{\ell_2}^2 \sqrt{N}}{\|\boldsymbol{\rho}\|_{\ell_1} \sqrt{\ln N}},$$

  3. we have zero false discovery rates for all levels of noise, with high probability.

Concluding remarks

- The most important features of the Noise Collector are that :
  1. no calibration is necessary with respect to the level of noise,
  2. exact support recovery is obtained for relatively large levels of noise

$$\|\delta b\|_{\ell_2} \leqslant c_1 \frac{\|\boldsymbol{b}\|_{\ell_2}^2 \sqrt{N}}{\|\boldsymbol{\rho}\|_{\ell_1} \sqrt{\ln N}},$$

  3. we have zero false discovery rates for all levels of noise, with high probability.
- More on the Noise Collector and its theoretical analysis in

M. Moscoso, A. Novikov, G. Papanicolaou, CT, *Imaging with highly incomplete and corrupted data*, Inverse Problems, 36(3), p. 035010, 2020. https://doi.org/10.1088/1361-6420/ab5a21

M. Moscoso, A. Novikov, G. Papanicolaou, CT, *The Noise Collector for sparse recovery in high dimensions*, Proceedings of the National Academy of Sciences, 117 (21), p. 11226-11232, 2020. https://doi.org/10.1073/pnas.1913995117