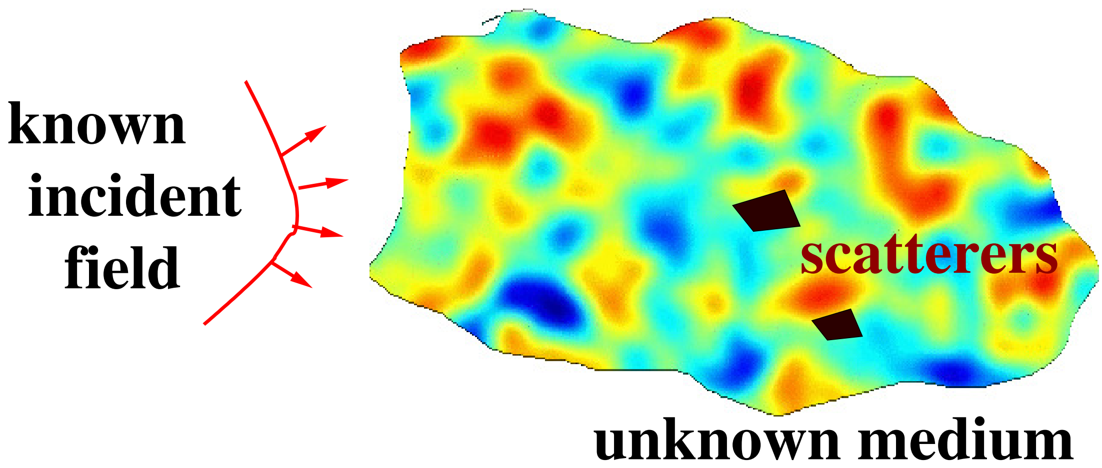
The Noise Collector for sparse recovery in high dimensions

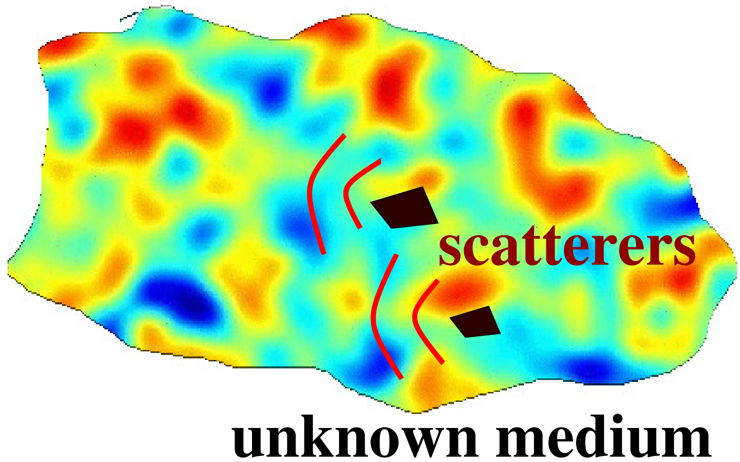
Alexei Novikov
Department of Mathematics
Penn State University, USA

with M.Moscoso (Madrid), G.Papanicolaou (Stanford), and C.Tsogka (UC Merced).
Supported by NSF and AFOSR.

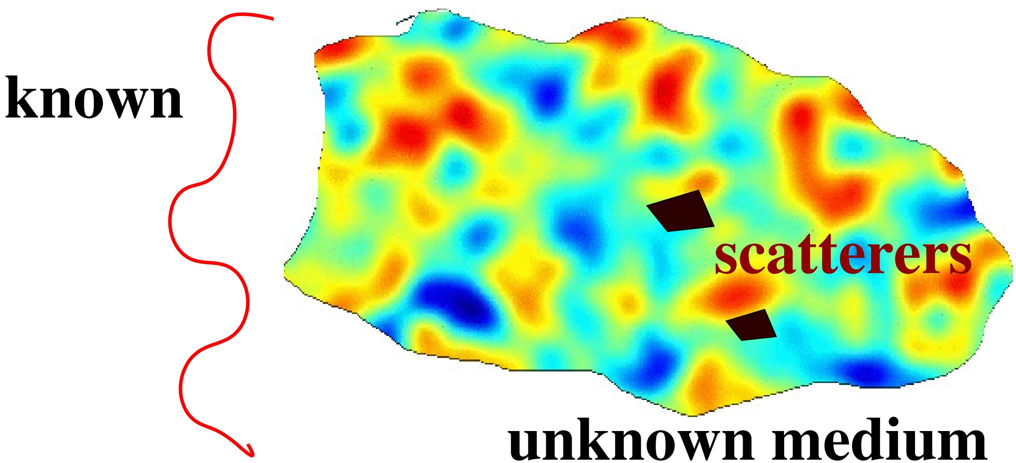
Inverse problems in wave propagation



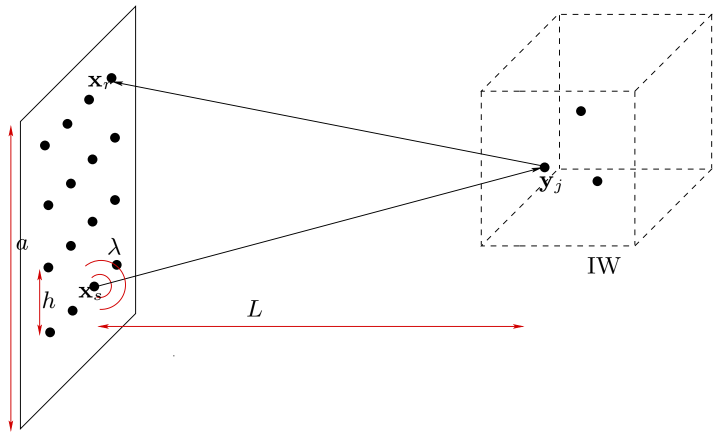
Inverse problems in wave propagation



Inverse problems in wave propagation

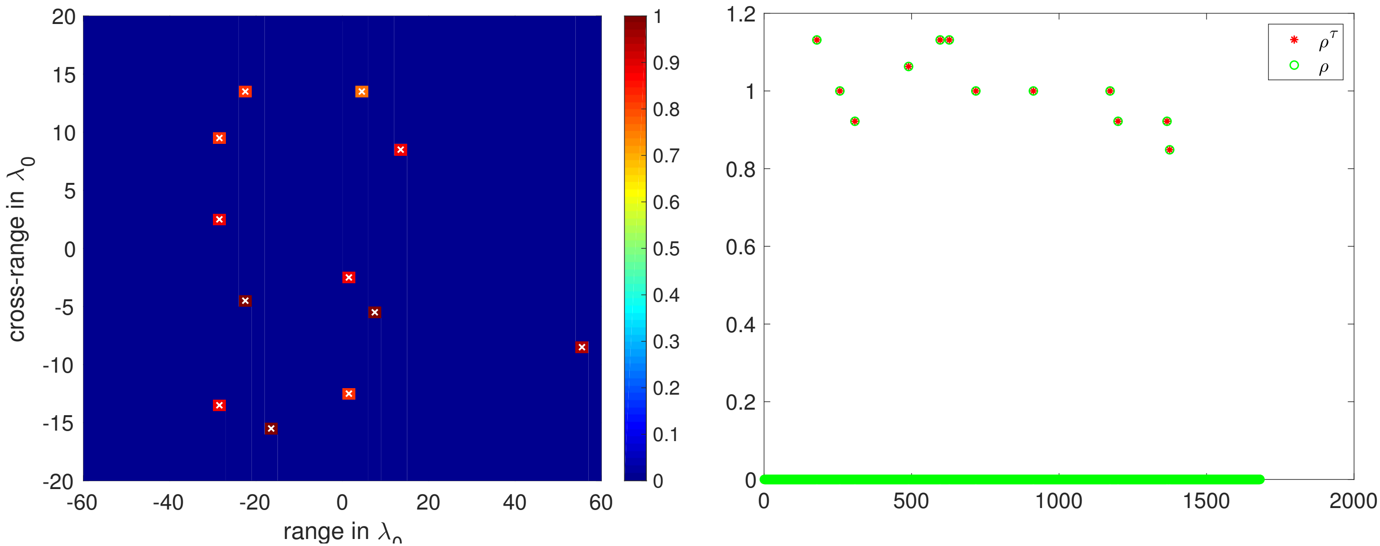


Imaging setup



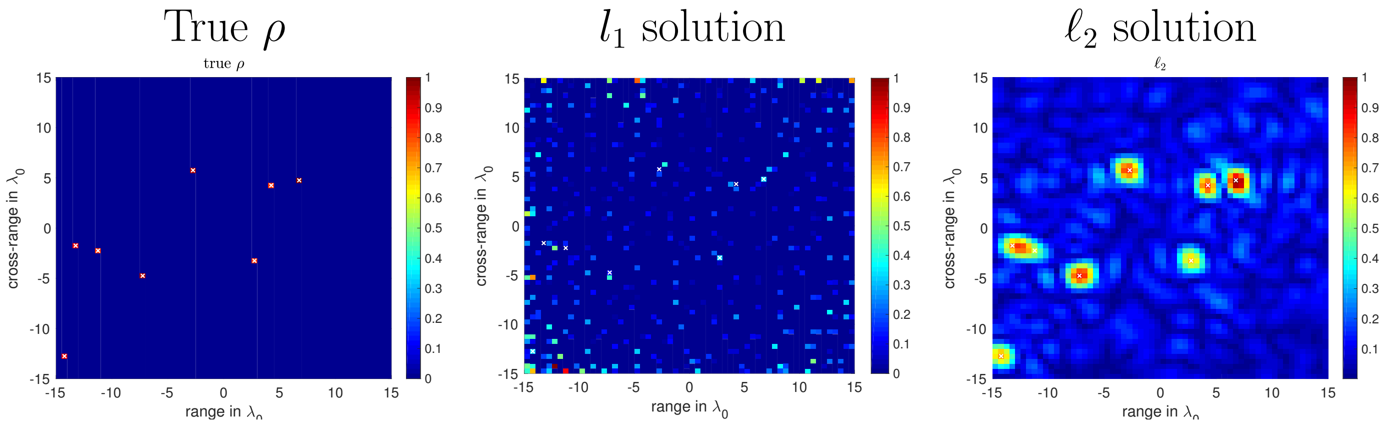
An signal is emitted from x_s at the array of N transducers, it illuminates the scatterers in the image window (IW). The point scatterers are at y_j, and now they can be viewed as the sources of the signal. They send the scattered signal back to the array. There are K pixels in the IW. The number of scatters is $M < N$, and $N < K$, typically. The map $A\rho \Rightarrow b$ in the *paraxial approximation* is (up to a constant) the (partial) Fourier transform.

Imaging of sparse scenes



Left: the true image. I show 2-dimensional images for simplicity. Right: the recovered solution vector is plotted with red stars and the true solution vector of $A\rho = b$ with green circles.

Noise, l_1 versus l_2 regularization



l_1-methods are unstable to noise., l_2-methods loose resolution.

l_1-regularization and Lasso

Noiseless. Want to solve a sparsity promoting optimization

$$\rho = \arg\min \|\tilde{\rho}\|_0, \text{ subject to } A\tilde{\rho} = b$$

where $\|\rho\|_0 = \{\#\rho_i \neq 0\}$. It is expensive, so we solve

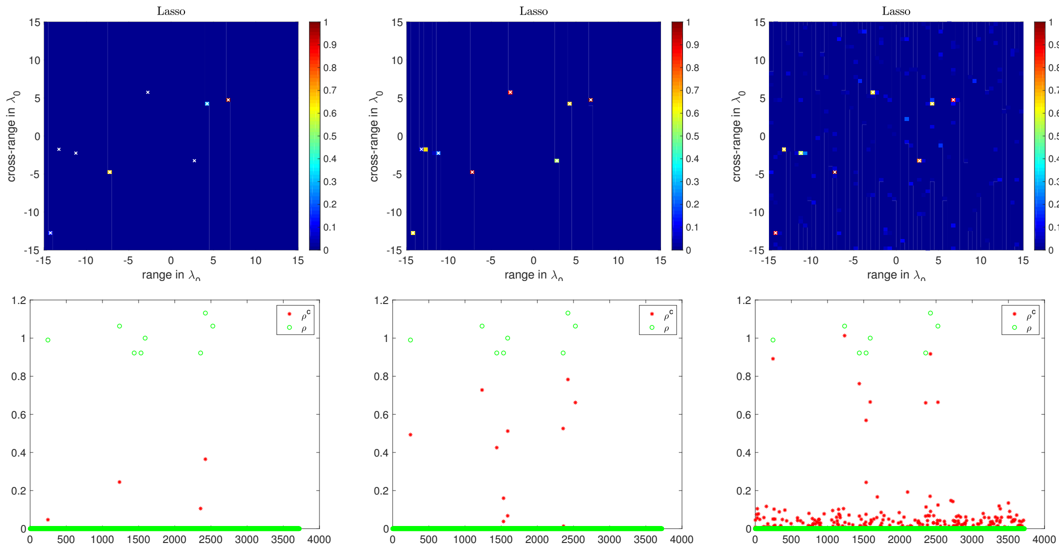
$$\rho = \arg\min \|\tilde{\rho}\|_1, \text{ subject to } A\tilde{\rho} = b$$

where $\|\rho\|_1 = \sum_i |\rho_i|$.
Noisy case, Lasso. R. Tibshirani '96, Chen & D.Donoho '94, F.Santosa & W.Symes '86

$$\rho = \arg\min \left(\lambda \|\tilde{\rho}\|_1 + \frac{\|A\tilde{\rho} - b\|_2^2}{2} \right)$$

where $\|\rho\|_2 = \sqrt{\sum_i |\rho_i|^2}$ and λ is a tuning parameter.

Tuning λ in Lasso



LASSO results with $\lambda = 1$, $\lambda = 0.5$ (optimal) and $\lambda = 0.1$.

l_1 & LASSO

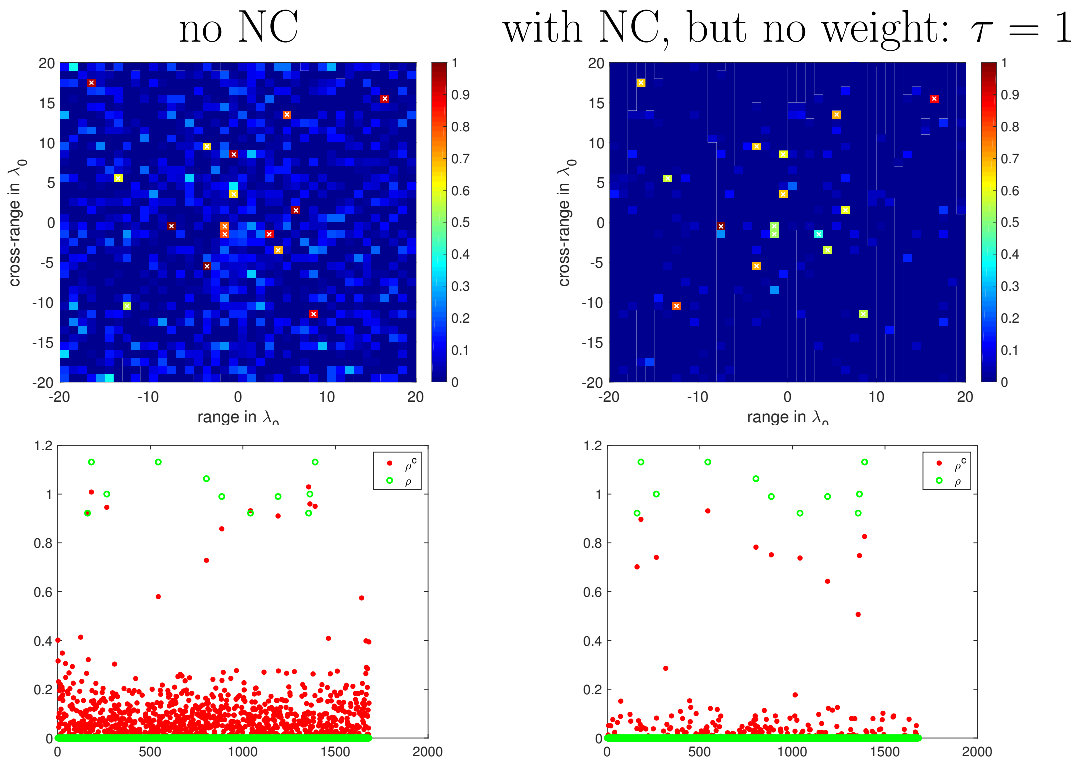
LASSO: finds a *sparse* approximate solution $A\rho \approx b$ if the tuning parameter λ is chosen correctly.

l_1: finds a *sparse* solution only if noise $e = 0$. No tuning parameters. If e is small, the sparse approximate solution can be found by thresholding. Thresholding has to be tuned.

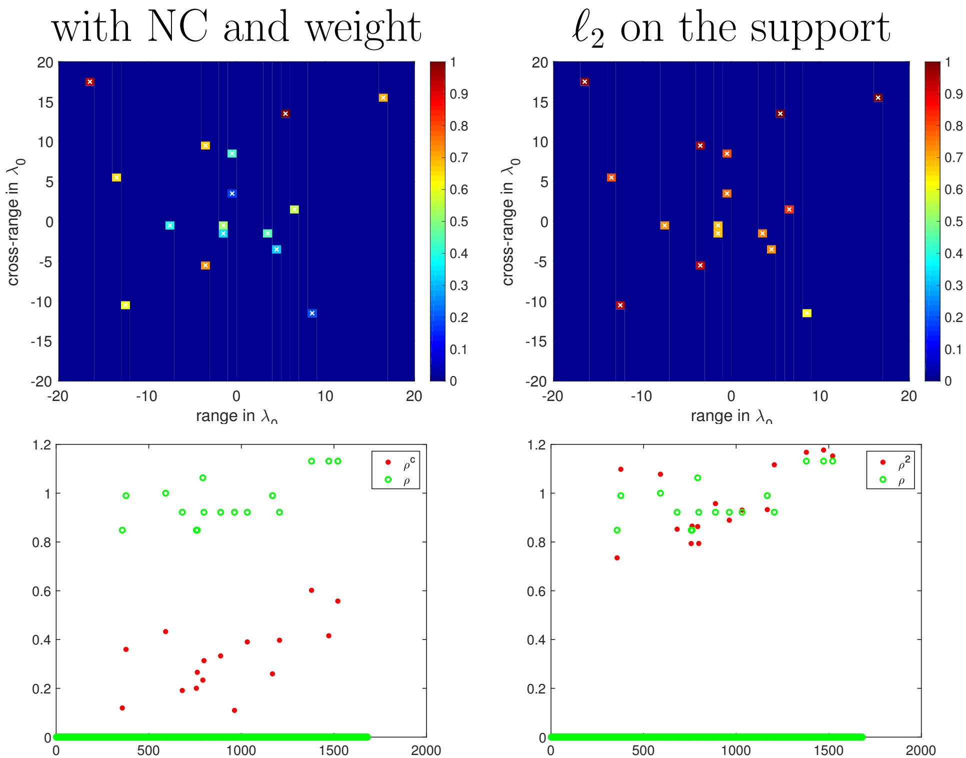
We propose to solve

$$(\rho_\tau, \eta) = \arg\min \left(\tau \|\rho\|_1 + \|\eta\|_1 \right), \text{ subject to } A\rho + C\eta = b$$

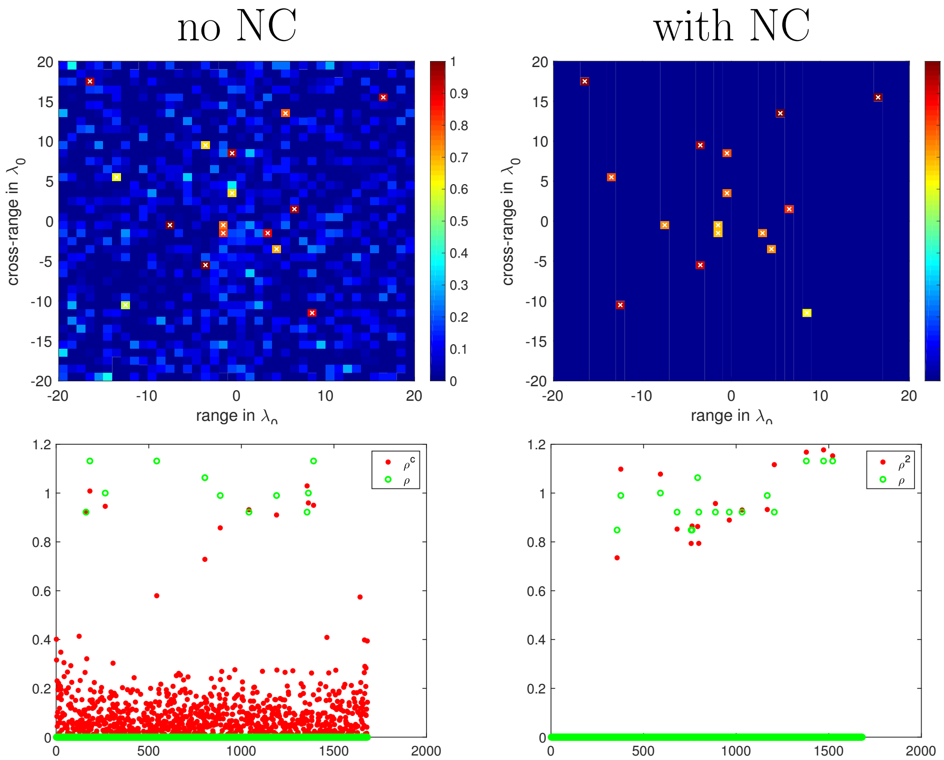
where C is the *noise collector* matrix and τ is the weight of the noise collector, and $b = b_0 + e$.
We can prove $\rho \approx \rho_\tau$ and $C\eta \approx e$ if τ is chosen correctly.
We can choose $\tau = O(\sqrt{\ln N})$ for *any* level of noise, before de-noising.



The top row - the images, the bottom row - the solution vector with red stars and the true solution vector with green circles.



The top row - the images, the bottom row - the solution vector with red stars and the true solution vector with green circles.



no NC
with NC
cross-range in λ_0
range in λ_0
ρ^c
ρ
ρ^2

Design of the Noise Collector

(i) Columns of NC should be sufficiently orthogonal to the columns of A, so it does not absorb signals with meaningful information.

(ii) Columns of NC 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 NC should grow slower than exponential with N, otherwise the method is impractical.

(iv) Deterministic approach. If we fill up C imposing

$$|\vec{a}_i \cdot \vec{c}_j| < \frac{\alpha}{\sqrt{N}} \ \forall i, j \ , \text{ and } |\vec{c}_i \cdot \vec{c}_j| < \frac{\alpha}{\sqrt{N}} \ \forall i \neq j, \tag{1}$$

then the Kabatjanskii-Levenstein inequality implies that the number Σ of columns in C grows at most polynomially: $N^{\alpha} \leqslant \Sigma \leqslant N^{\alpha^2}$.

Probabilistic approach to design of NC

If the columns of C are drawn at random independently. then the dot product of any two random unit vectors is still typically of order $1/\sqrt{N}$. We have an asymptotically negligible event that our noise collector is bad The decoherence constraint is weakened by a logarithmic factor.

Lemma: Choose $\beta > 1$, and pick $\Sigma = N^\beta$ vectors $\vec{c}_i$ at random and independently on $\mathbb{S}^{N-1}$. Then, for any $\kappa > 0$ there are constants $c_0(\kappa, \beta)$ and $\alpha > 1/2$, such that (i)

$$|\vec{a}_i \cdot \vec{c}_j| < c_0 \sqrt{\ln N}/\sqrt{N} \quad \text{for all } i, j, \tag{2}$$

and (ii) for any $\vec{e} \in \mathbb{S}^{N-1}$ there exists at least one $\vec{c}_j$, so

$$|\vec{e} \cdot \vec{c}_j| > \alpha/\sqrt{N}, \tag{3}$$

with large probability $1 - 1/N^\kappa$. In addition the condition number of $[A|C]$ is $O(1)$.

False Discovery Rate is zero

Theorem 1: (No phantom signal) Suppose there is no signal: $\rho = 0$ and $e/\|e\|_{l_2}$ is uniformly distributed on the unit sphere. For any $\kappa > 0$ we can construct the noise collector and choose weight τ so that $\rho_\tau = 0$ with probability $1 - 1/N^\kappa$.

Theorem 2: Let ρ be an M-sparse solution of $A\rho = b_0$. If the columns of A are decoherent: $|a_i \cdot a_j| \leqslant \frac{1}{3M}$, then $\text{supp}(\rho_\tau) \subseteq \text{supp}(\rho)$ with probability $1 - 1/N^\kappa$.

Supports of ρ and $\rho_{\mathcal{T}}$ agree

Theorem 3: Suppose r is the magnitude of smallest non-zero entry of ρ. If $\|e\|_{l_2}/\|b_0\|_{l_2} \leqslant c_2\sqrt{\ln N}$, $c_2 = c_2(\kappa, \beta, r, M)$, then $\mathrm{supp}(\rho_\tau) = \mathrm{supp}(\rho)$, with probability $1 - 1/N^\kappa$.

Theorem 4: (Exact Recovery): If there is no noise $e = 0$. Then $\rho_\tau = \rho$ with probability $1 - 1/N^\kappa$.

Failure to recover

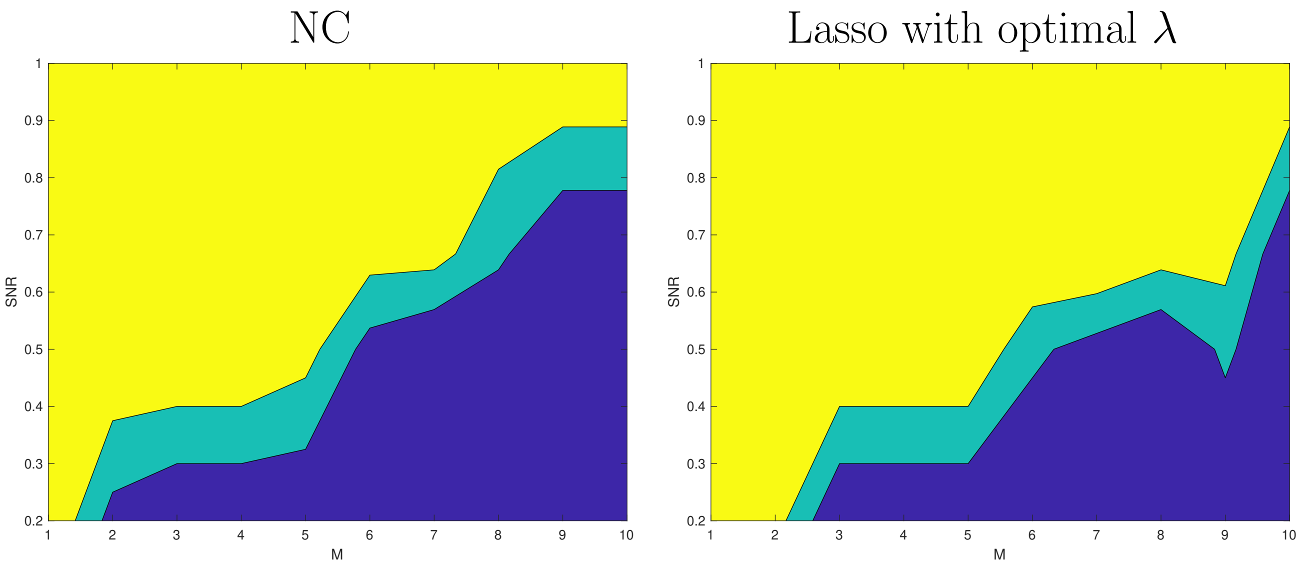
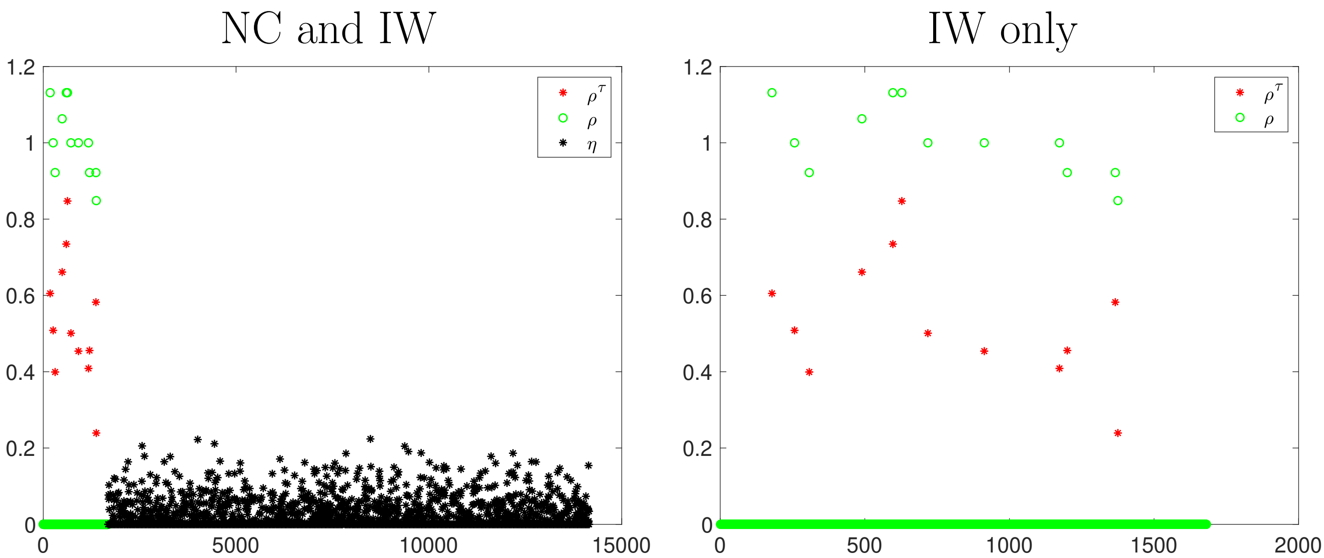


Image Window and Noise Collector



Coefficients of the solution, 0dB, $N = 625$, $K = 1000$, $\Sigma = 10000$.

GeLMA with Fast Noise Collector

Require: Set $\rho = 0$, $z = 0$. $\eta = 0$.

Pick $\beta = \beta(A, C)$, and $\tau = .8\sqrt{\ln N}$.

repeat

$r = b - A\rho - C\eta$

$z \Leftarrow z + \beta r$

$\rho \Leftarrow \mathcal{S}_{\tau\beta}(\rho + \beta A^*(z + r))$

$\eta \Leftarrow \mathcal{S}_{\beta}(\eta + \beta C^*(z + r))$

until Convergence

Calibrate τ so that FDR is zero when $b = e$ (only noise). "No phantom signal" criterion.

Fast Noise Collector C is several random circular matrices. Then it is cheap to store C, and we can use FFT for matrix-vector multiplication.

The soft shrinkage-thresholding operator

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

Geometric interpretation of $\vec{z}$

$$\vec{a}_i \cdot \vec{z} = \tau \operatorname{sign}(\rho_i), \text{ if } \rho_i \neq 0, \text{ and } |\vec{a}_i \cdot \vec{z}| \leqslant \tau \text{ if } \rho_i = 0.$$

Assume both $\vec{a}_i$ and $-\vec{a}_i$ are columns of A, and all $\rho_i \geqslant 0$ then

$$H_A = \{x \in \mathbb{R}^N : x = \sum_i \alpha_i \vec{a}_i, \sum_i \alpha_i \leqslant 1, \alpha_i \geqslant 0\}$$

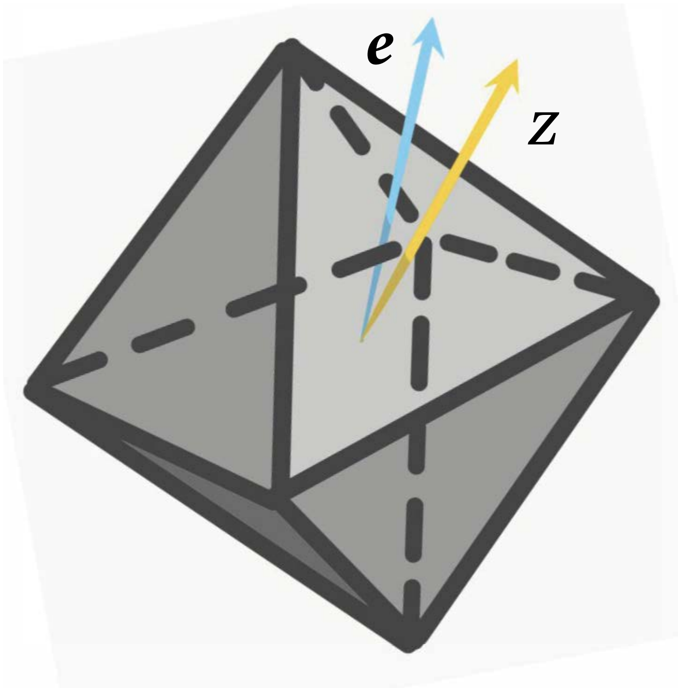
Suppose Λ is the support of ρ, typically (for non-sparse ρ) $|\Lambda| = N$. Then, the simplex

$$\left\{ \vec{x} \in \mathbb{R}^N \,\middle|\, \vec{x} = \sum_{i \in \Lambda} \alpha_i \vec{a}_i, \sum_{i \in \Lambda} \alpha_i = 1, \alpha_i \geqslant 0 \right\}$$

has the unique normal vector $\vec{n}$, which is collinear to $\vec{z}$ because

$$1 = \vec{z} \cdot \vec{a}_i = \frac{\vec{z} \cdot \vec{b}}{||\vec{b}||_A}, \ \forall i \in \Lambda, \text{ and } \vec{z} \cdot \vec{a}_j < 1, \forall j \notin \Lambda. \tag{4}$$

Geometric interpretation of z



If A is the identity matrix then $\Phi(e) = (\text{sign}(e_1), \ldots, \text{sign}(e_N))$.

Duality

Given A we have H_A define $\vec{z} = \Phi_A(\vec{b})$. Let

$$Z = \{ \text{ all } \vec{z} = \Phi_A(\vec{e}) \text{ for some } \vec{e} \in \mathbb{S}^{N-1} \}$$

Then

$$\max_{z \in Z} z \cdot b = \min(\|\rho\|_1, \text{ subject to } A\rho = b),$$

and

$$\Phi_A(\vec{b}) = \arg\max_{z \in Z} z \cdot b.$$

Proof of Theorem 1

Solve

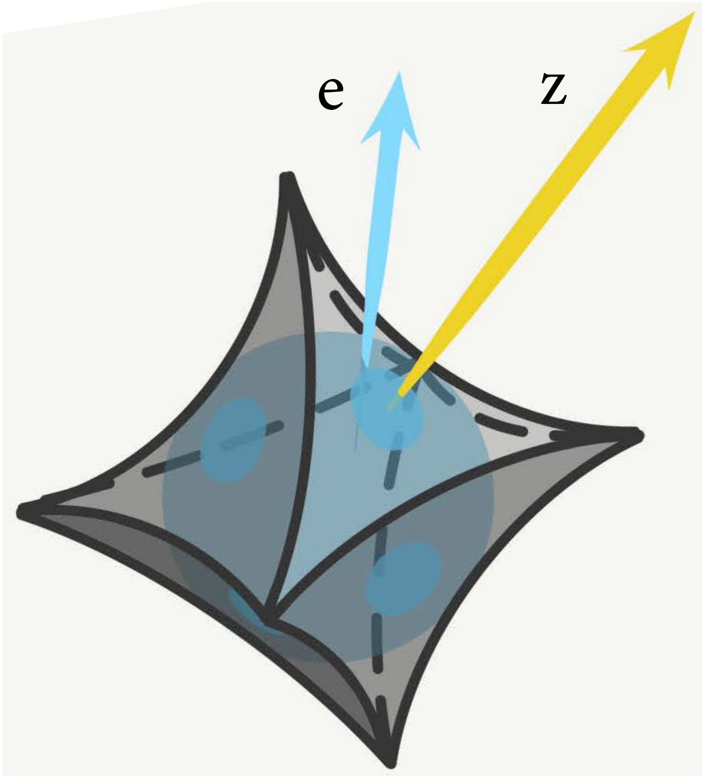
$$(\rho_\tau, \eta) = \arg\min \left(\tau\|\rho\|_1 + \|\eta\|_1\right), \text{ subject to } A\rho + C\eta = b$$

Theorem 1:(No phantom signal) Suppose there is no signal: $\rho = 0$ and $e/\|e\|_{l_2}$ is uniformly distributed on the unit sphere. For any $\kappa > 0$ we can construct the noise collector and choose weight τ so that $\rho_\tau = 0$ with probability $1 - 1/N^\kappa$.

Proof: Instead of Φ_A, consider $\Phi_C : e \to z$, where z is dual certificate of optimality of η. We want to show $\Phi_{[\tau A|C]} : e \to z$. It means that z is the also dual certificate of optimality of $(0, \eta)$, i.e.

$$|\vec{z} \cdot \vec{a}_j| < \tau, \forall j$$

Map $\Phi_C : e \to z$



Random vectors on $\mathbb{S}^{N-1}$

Everything is rotation invariant in $\Phi_C : e \to z$. Thus $n = z/\|z\|_2$ is uniformly distributed on $\mathbb{S}^{N-1}$.
$\|z\| = O(\sqrt{N})$, because l_1-balls are probabilistically l_2-small.
Coordinates of s uniformly distributed vector on $\mathbb{S}^{N-1}$ have i.i.d. Gaussian distribution $N(0, 1/\sqrt{N})$. The event $|\vec{z} \cdot \vec{a}_j| < \tau, \forall j$ does not happen with probability

$$\mathbb{P}\left(\max_j |\vec{z} \cdot \vec{a}_j| \geqslant \tau\right) \leqslant N^{\beta}\mathbb{P}\left(|\vec{n} \cdot \vec{a}_1| \geqslant \tau/\sqrt{N}\right)$$

$$\leqslant 2N^{\beta}\mathrm{e}^{-c\tau^2} \leqslant 1/N^{\kappa}, \text{ if } \tau = O(\sqrt{\ln N}).$$