Correlation-informed ordered dictionary learning for imaging in complex media

Chrysoula Tsogka

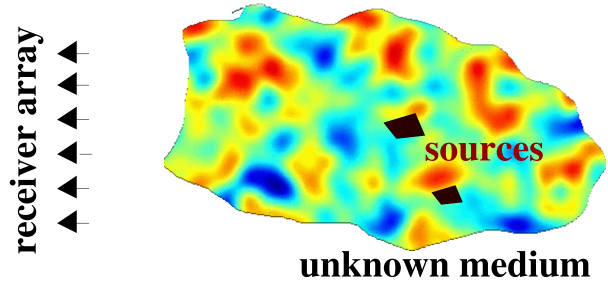
Collaborators: M. Moscoso (Univ. Carlos III de Madrid),
A. Novikov (Penn State),
G. Papanicolaou (Stanford University)

Department of Applied Mathematics, University of California, Merced
Support: AFOSR FA9550-21-1-0196

Table of contents

Inverse problems

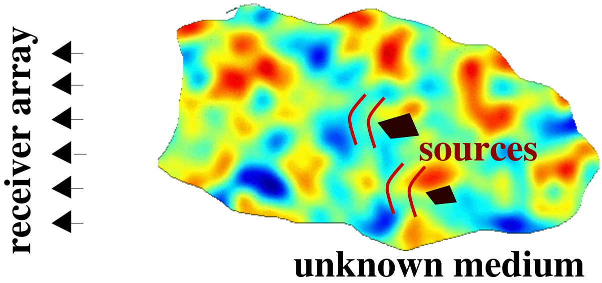
Wave propagation



- In this talk we consider the passive imaging or inverse source imaging problem

Inverse problems

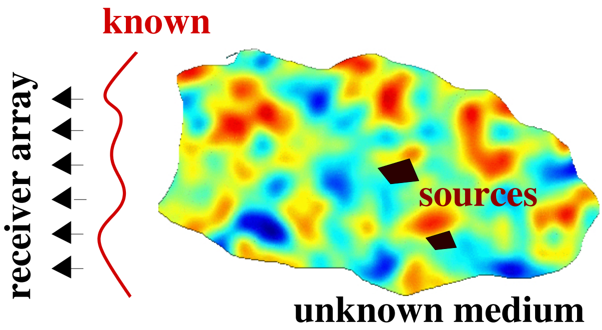
Wave propagation



- In this talk we consider the passive imaging or inverse source imaging problem
- We seek to locate sources inside an imaging window of interest

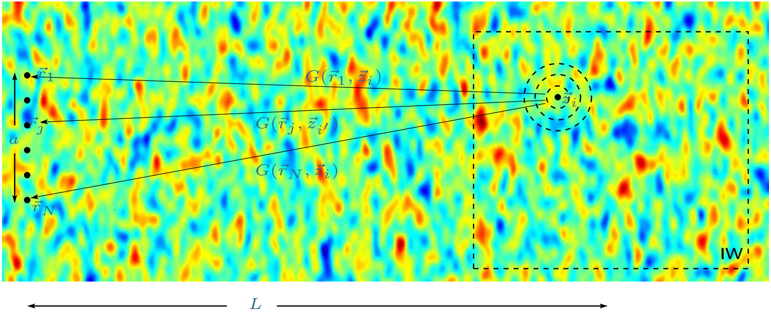
Inverse problems

Wave propagation



- In this talk we consider the passive imaging or inverse source imaging problem
- We seek to locate sources inside an imaging window of interest
- Using as data the field recorded on an array of receivers

Sensing matrix formalism



We introduce the $N \times K$ matrix

$$\boldsymbol{\mathcal{G}} = [\boldsymbol{g}(\vec{z}_1) \cdots \boldsymbol{g}(\vec{z}_K)]$$

where each column $\boldsymbol{g}(\vec{z}_i)$ is the Green's function vector

$$\boldsymbol{g}(\vec{z}_i) = [G(\vec{r}_1, \vec{z}_i), G(\vec{r}_2, \vec{z}_i), \dots G(\vec{r}_N, \vec{z}_i)]^T$$

corresponding to N measurements at all receivers and frequencies when in the random medium, there is a single source at position $\vec{z}_i$.

Wave propagation

The Green's function satisfies the wave equation

$$\Delta G(\vec{z},\vec{r};\omega) + \kappa^2\, n^2(\vec{z})\, G(\vec{z},\vec{r};\omega) = \delta(\vec{z}-\vec{r}),$$

where $\kappa = \omega/c_0$ is the wavenumber, c_0 is a constant reference wave speed. The random index of refraction is $n(\vec{z}) = c_0/c(\vec{z})$ with local wave speed $c(\vec{z})$.

In a homogeneous medium, $c(\vec{z}) \equiv c_0$ for any location $\vec{z}$ and, in this case, $G(\vec{z},\vec{r};\omega) = G_0(\vec{z},\vec{r};\omega)$, where

$$G_0(\vec{z},\vec{r};\omega) = \frac{\exp(i\,\kappa\,|\vec{z}-\vec{r}|)}{4\pi|\vec{z}-\vec{r}|}\,.$$

Table of contents

Wave propagation in random media

- We consider wave propagation for *long distances* in *weakly heterogeneous* media. In that case the perturbations induced by the medium inhomogeneities *accumulate* and have an *order one effect*.
- The random fluctuations of the wave speed are modeled as

$$\frac{1}{c^2(\vec{z})} = \frac{1}{c_0^2}\left(1 + \sigma\mu(\frac{\vec{z}}{\ell})\right).$$

- c_0 denotes the average speed
- σ denotes the strength of the fluctuations and ℓ is the correlation length
- $\mu(\cdot)$ is a zero mean stationary random process.
- We use the random travel time model which characterizes wave propagation in the high-frequency regime in random media with weak fluctuations $\sigma \ll 1$ and large correlation lengths ℓ compared to the wavelength λ.

Random travel time model

- The Green's function between two points $\vec{z}$ and $\vec{r}$ at a distance L from each other such that $L \gg \ell \gg \lambda$ is given by

$$G(\vec{z}, \vec{r}; \omega) = G_0(\vec{z}, \vec{r}; \omega) \exp\left[i\omega\nu(\vec{z}, \vec{r})\right].$$

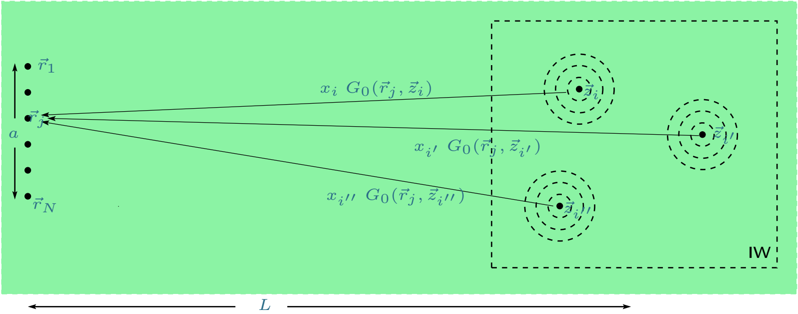
- G_0 is the Green's function in a homogeneous medium with speed c_0
- $\nu(\vec{z}, \vec{r})$ is the random function

$$\nu(\vec{z}, \vec{r}) = \frac{\sigma|\vec{z} - \vec{r}|}{2c_0} \int_0^1 ds\, \mu\left(\frac{\vec{r}}{\ell} + (\vec{z} - \vec{r})\frac{s}{\ell}\right),$$

which accounts for phase distortions induced by the medium's fluctuations.

- $\sigma_0 = \frac{\lambda_0}{\sqrt{\ell L}}$ is a characteristic strength of the medium's fluctuations for which the standard deviation of the random phase fluctuations is $O(1)$.
- We define the strength of the fluctuations in terms of σ_0 by introducing the dimensionless parameter $\tilde{\sigma} = \frac{\sigma}{\sigma_0}$.

Imaging in homogeneous media

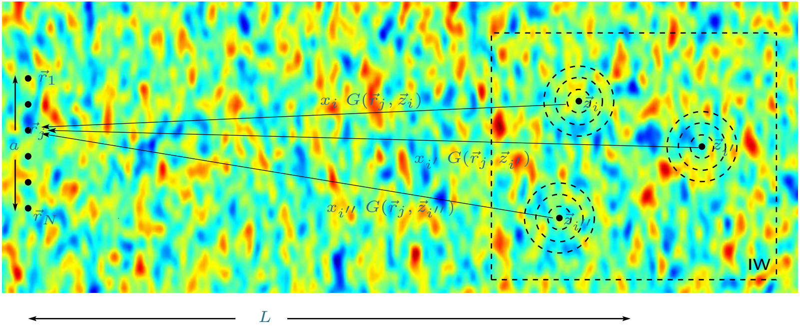


- In a homogeneous medium the data are

$$\boldsymbol{y} = \boldsymbol{\mathcal{G}}_0\, \boldsymbol{x}$$

- $\boldsymbol{x}$ is a vector whose jth component represents the complex amplitude of the source at location $\vec{z}_j$ in the image window, $j = 1, \ldots, K$.
- The sensing matrix $\boldsymbol{\mathcal{G}}_0$ is known. One needs regularization because usually $N \ll K$. ℓ_2 or ℓ_1 methods can be successfully used.
- Kirchhoff migration image is obtained by applying $\boldsymbol{\mathcal{G}}_0{}^*$ to the data *back-propagating the data in the homogeneous medium*

Imaging in random media



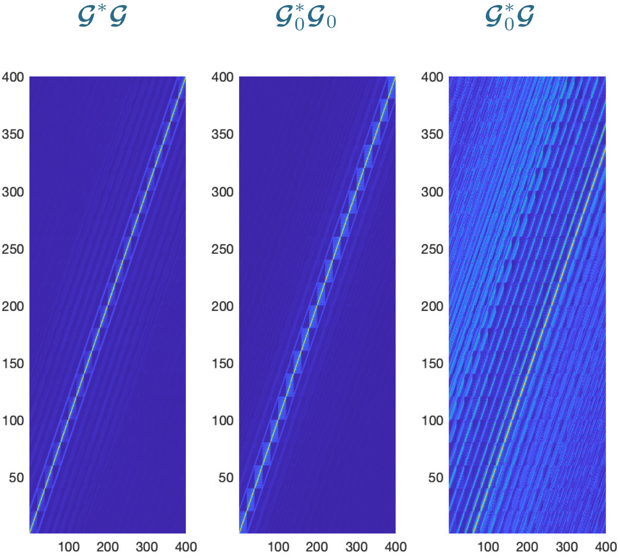
- In random media the data are

$$y = \mathcal{G}\, x\,,$$

- The measurement matrix $\mathcal{G}$ is unknown.
- An option is to approximate $\mathcal{G} \sim \mathcal{G}_0$, where $\mathcal{G}_0$ is the *known* sensing matrix in the homogeneous medium.
- Kirchhoff migration image is obtained by applying ${\mathcal{G}_0}^*$ to the data *back-propagating the data in the homogeneous medium*

Imaging in random media

The $\mathcal{G}^*\mathcal{G}$ matrix



Imaging in random media

Coherent Interferometry

- Our model for the sensing matrix is for a *fictitious medium* (homogeneous with no fluctuations) $\rightsquigarrow$ phases are not accounted for correctly $\rightsquigarrow$ *instability* in the reconstruction

Imaging in random media

Coherent Interferometry

- Our model for the sensing matrix is for a *fictitious medium* (homogeneous with no fluctuations) $\rightsquigarrow$ phases are not accounted for correctly $\rightsquigarrow$ *instability* in the reconstruction
- To stabilize the imaging process we form cross-correlations and restrict the cc-data to nearby frequency and receiver offsets : $|\omega - \omega'| \leq \Omega_d$ and $|\boldsymbol{r} - \boldsymbol{r}'| \leq X_d$.

Imaging in random media

Coherent Interferometry

- Our model for the sensing matrix is for a *fictitious medium* (homogeneous with no fluctuations) $\rightsquigarrow$ phases are not accounted for correctly $\rightsquigarrow$ *instability* in the reconstruction
- To stabilize the imaging process we form cross-correlations and restrict the cc-data to nearby frequency and receiver offsets : $|\omega - \omega'| \leq \Omega_d$ and $|\boldsymbol{r} - \boldsymbol{r}'| \leq X_d$.
- Statistical stability means the variance of the image is small compared to its mean square, with respect to the realizations of the random medium. It comes at the cost of loss in resolution :

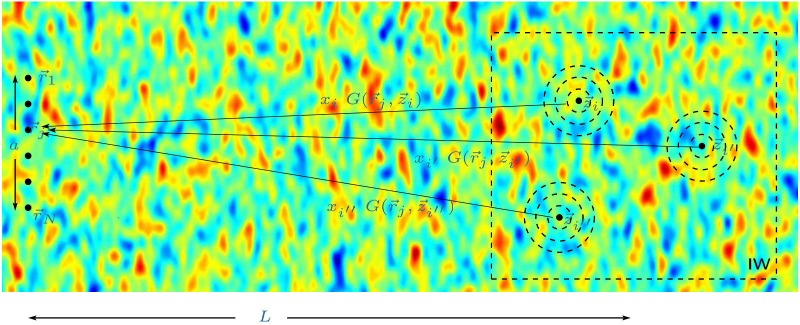
  cross-range : $\lambda L/a \rightsquigarrow \lambda L/X_d$
  range (depth) : $c_0/B \rightsquigarrow c_0/\Omega_d$

Statistical Stability

Statistical stability for cross-correlation data has been studied in the framework of Coherent Interferometry (CINT) which amounts to *back propagating cc-data*, rather than the raw data.

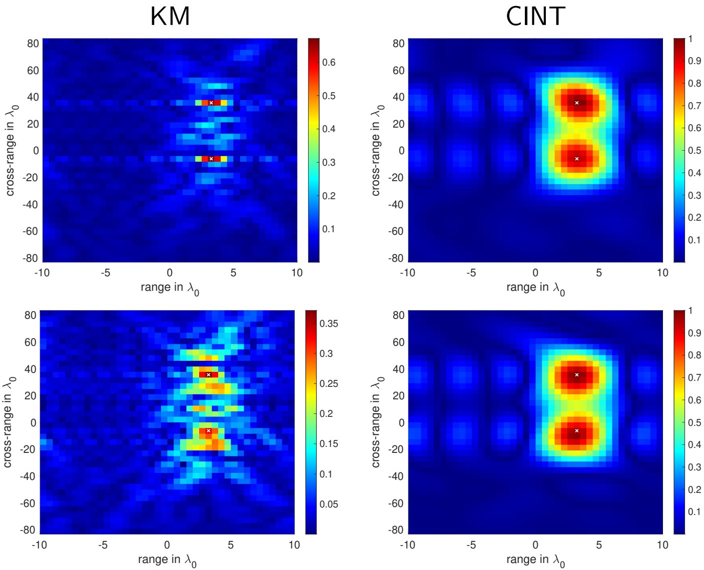
- L. Borcea, G. Papanicolaou and C.T, *Interferometric array imaging in clutter*, Inverse Problems, **21** (2005), pp. 1419-1460.
- L. Borcea, G. Papanicolaou and C.T., *Adaptive interferometric imaging in clutter and optimal illumination*, Inverse Problems, **22** (2006), pp. 1405–1436.
- L. Borcea, G. Papanicolaou and C.T., *Coherent interferometric imaging in clutter*, Geophysics, **71** (2006), pp. SI165-SI175
- L. Borcea, G. Papanicolaou and C.T., *Optimal illumination and waveform design for imaging in random media*, J. Acoust. Soc. Am., **122** (2007), pp. 3507-3519.
- L. Borcea, J. Garnier, G. Papanicolaou and C.T., *Enhanced statistical stability in coherent interferometric imaging*, Inverse Problems, **27** (2011), p. 085003.

Numerical setup for imaging in random media



- The IW is far from the array at distance $L = 100\ell$.
- Correlation length is $\ell = 100\lambda$, strength of fluctuations is $\tilde{\sigma} = 0.4$.
- The size of the array is $a = 24\ell$.
- The decoherence parameters are selected to be optimal for CINT $X_d = 4\ell$, $\Omega_d = B/2$.

Imaging results random medium



Imaging results random medium

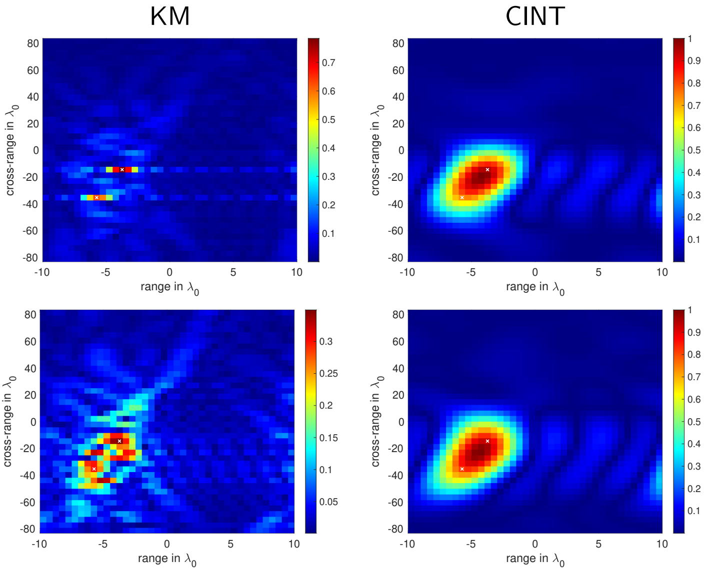


Table of contents

Challenge

- Our objective is to develop a *statistically stable* imaging method in *random* media without *loss in resolution*.
- We assume that we have access to *multiple diverse measurements*

$$y_m = \boldsymbol{\mathcal{G}}\, x_m,\ m = 1, 2, \ldots, M,\ \ M \geqslant CK \ln K$$

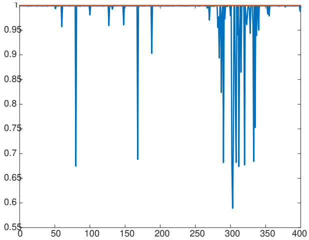
- Each $\boldsymbol{x}_m$ is s-sparse and "random" in $\mathbb{C}^N$.
- Given many $\boldsymbol{y}_m$ can we reconstruct the columns of $\boldsymbol{\mathcal{G}}$?
- This is precisely the objective of *Sparse Dictionary Learning*.

1st ingredient : Sparse Dictionary Learning

- *Sparse Dictionary Learning* aims at finding a set of basic elements, called a dictionary, and a sparse representation of the input data in the form of a linear combination of these basic elements.

1st ingredient : Sparse Dictionary Learning

- *Sparse Dictionary Learning* aims at finding a set of basic elements, called a dictionary, and a sparse representation of the input data in the form of a linear combination of these basic elements.
- Given many $y_m = \mathcal{G} x_m$ we can reconstruct the columns of $\mathcal{G}$ and find the coefficients x_m.



Maximum correlation between each estimated column and the true ones in the sensing matrix. In red the columns of the matrix are more incoherent. In blue the columns of the matrix are more coherent.

1st ingredient : Sparse Dictionary Learning

- DL works well when the columns of $\mathcal{G}$ are incoherent.
- Reconstructing the columns of $\mathcal{G}$ is *not enough* for imaging. The columns of $\mathcal{G}$ are *unordered*.
- We need to associate each column of $\mathcal{G}$ to its corresponding point in the image window.

2nd ingredient : Multidimensional Scaling

- Think of points in the IW as vertices $\vec{z}_i$ of a graph G.
- If all the Euclidean distances $D = (d_{ij})$ between $\vec{z}_i$ and $\vec{z}_j$ are known, the classical MDS algorithm (Torgerson, 1952) recovers the points (up to rotation and translation).

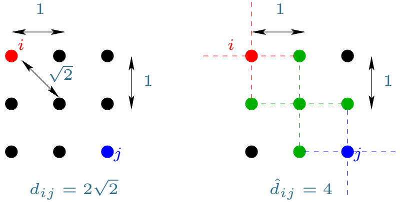
$$-\frac{1}{2}LDL = LZZ^T L\,, Z = [\vec{z}_1, \vec{z}_2, \dots, \vec{z}_K]^T$$

where $L = \mathbf{I}_K - \mathbb{1}_K \mathbb{1}_K^T/K$ is a centering matrix, with $\mathbf{I}_K$ the $K \times K$ identity matrix, and $\mathbb{1}_K$ the column vector of all ones.

- This means that the matrices ZZ^T and $-D/2$ are equal when the center of mass of the configuration is moved to zero.
- Z can be estimated from the SVD of $-\frac{1}{2}LDL$
- *We don't know* D

2nd ingredient : Multidimensional Scaling

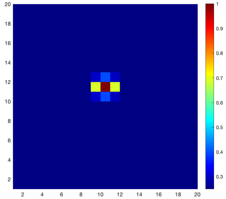
- What if we use another distance ?
- Take a graph $G = (V, E)$, where edges connect the nearest vertices. Define the geodesic graph distance between all vertices. That is, define a proxy distance between vertices i and j, denoted $\hat{d}_{ij}$, to be equal to the number of edges in the shortest path connecting i and j.



- Apply the classical MDS to this proxy metric. Compute SVD of $-\frac{1}{2}L\hat{D}L$ This is MDS-MAP (Y. Shang et al, 2003).
- How can we find the nearest vertices of each column of $\mathcal{G}$?

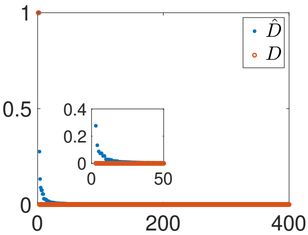
3rd ingredient : Cross-correlations

$\mathcal{G}^* \boldsymbol{g}(\vec{z}_i)$



A three-step algorithm

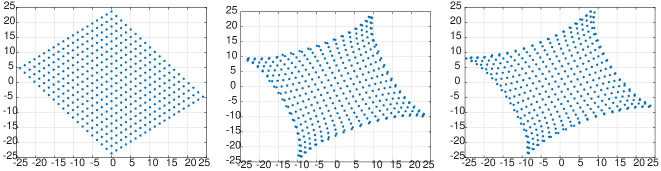
- Given many $y_m = \mathcal{G} x_m$, $m = 1, 2, \ldots, M$ use Dictionary Learning to reconstruct the *unordered* columns of $\mathcal{G}$.
- Construct a graph $G = (V, E)$, where vertices are the *unordered* columns of $\mathcal{G}$, and they are connected by edges if cross-correlations are close to 1.
- Compute the geodesic graph distance on the graph $G = (V, E)$ and apply the MDS algorithm to this proxy distance.



SVD of D and $\hat{D}$

Grid reconstruction

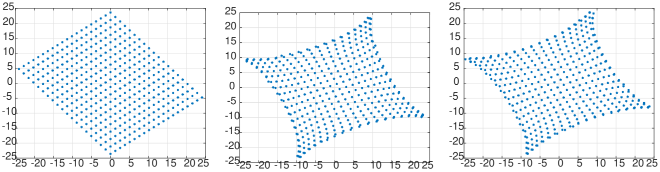
2nd step : MDS



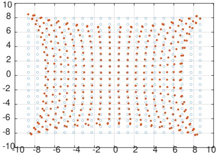
From left to right : Grid reconstruction using MDS with true Euclidean distances ; Grid reconstruction using the MDS-MAP algorithm with geodesic graph distances computed on the graph obtained from the true 4 nearest neighbors ; 4 nearest neighbors from cross-correlations are used.

Grid reconstruction

2nd step : MDS

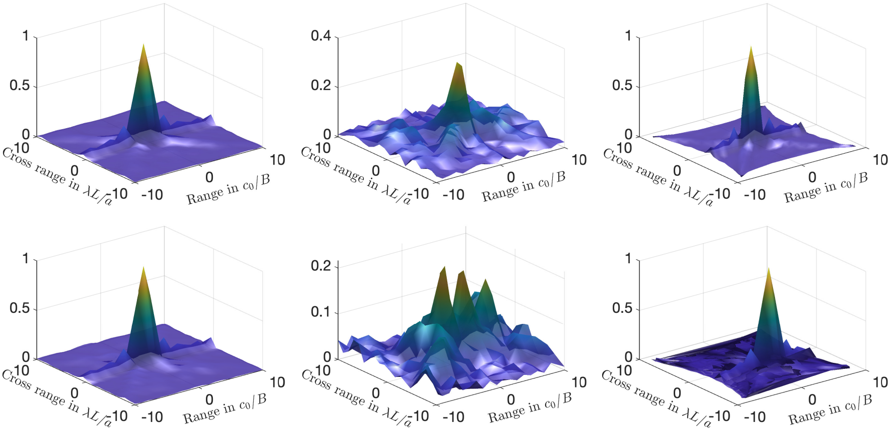


From left to right : Grid reconstruction using MDS with true Euclidean distances ; Grid reconstruction using the MDS-MAP algorithm with geodesic graph distances computed on the graph obtained from the true 4 nearest neighbors ; 4 nearest neighbors from cross-correlations are used.



Using 3 points as anchors we can estimate the scaling and the rotation needed to recover the absolute grid positions

Imaging results



From left to right, image formed using the true random Green's functions, the homogeneous Green's functions, and the recovered ones with the proposed method. Top : $\tilde{\sigma} = 0.6$. Bottom : $\tilde{\sigma} = 0.8$

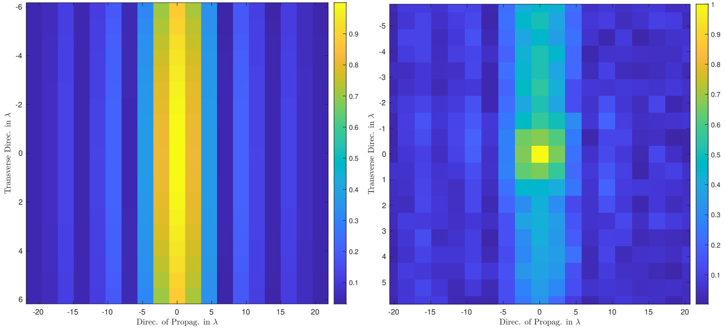
Table of contents

Time reversal and super-resolution in random media

- Time reversal has many important applications in several areas such as underwater sound, ultra-sound, imaging, and communications.
- Important properties of time reversal in multiple scattering media : super-resolution and statistical stability !
- References (∞) : M. Fink et al., G. Papanicolaou et al.

Time reversal and super-resolution in random media

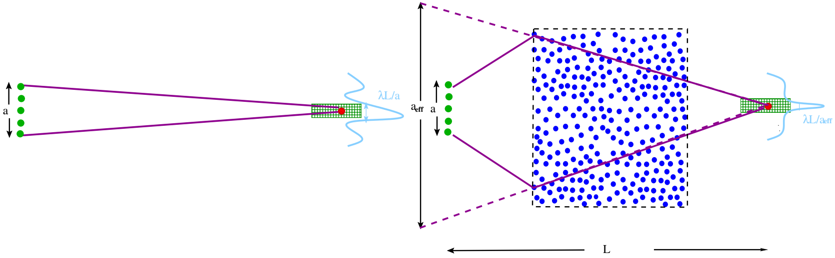
- Time reversal has many important applications in several areas such as underwater sound, ultra-sound, imaging, and communications.
- Important properties of time reversal in multiple scattering media : super-resolution and statistical stability !
- References (∞) : M. Fink et al., G. Papanicolaou et al.



TR refocusing in homogeneous (left) and random (right) media.
We observe super-resolution in the cross-range direction.

Time reversal and super-resolution in random media

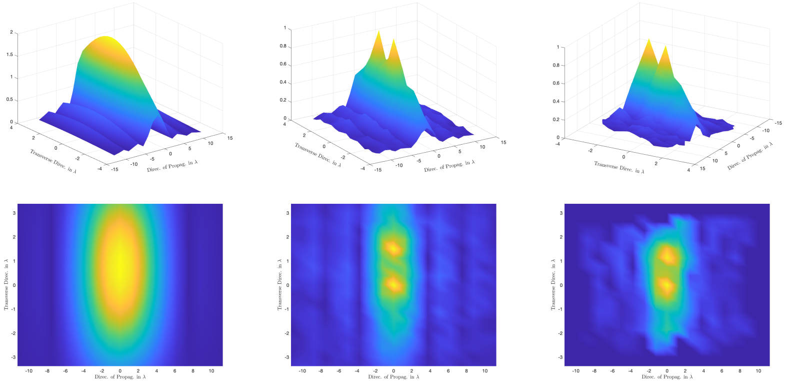
- Time reversal has many important applications in several areas such as underwater sound, ultra-sound, imaging, and communications.
- Important properties of time reversal in multiple scattering media : super-resolution and statistical stability !
- References (∞) : M. Fink et al., G. Papanicolaou et al.



Discrete random medium, use Foldy Lax

Imaging results in strongly scattering media

Super-resolution allows us to image nearby targets



From left to right, image in a homogeneous medium, image formed using the true random Green's functions, image formed using the recovered Green's functions with the proposed method.

Table of contents

Conclusions

- We propose a method that achieves statistically stable imaging in random media & benefits from multiple scattering allowing for super-resolution
- The method is applicable if we have multiple diverse measurements.
- The three key ingredients are
  1. Given many $\boldsymbol{y}_{\boldsymbol{m}} = \boldsymbol{\mathcal{G}}\boldsymbol{x}_{\boldsymbol{m}}$, $m = 1, 2, \ldots, M$ use Dictionary Learning to reconstruct the *unordered* columns of $\boldsymbol{\mathcal{G}}$.
  2. Construct a graph $G = (V, E)$, where vertices are the *unordered* columns of $\boldsymbol{\mathcal{G}}$, and edges are found from cross-correlations.
  3. Define the geodesic graph distance $G = (V, E)$ and apply the MDS to this proxy metric.

M. Moscoso, A. Novikov, G. Papanicolaou and C. T., *Correlation-informed ordered dictionary learning for imaging in complex media*, PNAS 121 (11) e2314697121 (2024) https://doi.org/10.1073/pnas.2314697121

Two delicate aspects

DL and coherence of the columns of $\mathcal{G}$

- *Dictionary Learning needs the columns of $\mathcal{G}$ to be incoherent but if columns are incoherent we cannot find neighbors and construct the graph needed for MDS-MAP*.

Two delicate aspects

DL and coherence of the columns of $\mathcal{G}$

- *Dictionary Learning needs the columns of $\mathcal{G}$ to be incoherent but if columns are incoherent we cannot find neighbors and construct the graph needed for MDS-MAP*.
- Subsampling and coherence : coherence increases when we decrease array size (and bandwidth).

Two delicate aspects

DL and coherence of the columns of $\mathcal{G}$

- *Dictionary Learning needs the columns of* $\mathcal{G}$ *to be incoherent but if columns are incoherent we cannot find neighbors and construct the graph needed for MDS-MAP*.
- Subsampling and coherence : coherence increases when we decrease array size (and bandwidth).
- We can use sub-sets of the data in the second step to increase the coherence and find neighbours

Two delicate aspects

Dictionary Learning and its initialization

- The DL problem is : find the dictionary $\mathcal{G}$ and the coeffs $\boldsymbol{x}_m$,

$$\min_{\mathcal{G},\mathbf{X}} \|\mathcal{G}\mathbf{X} - \mathbf{Y}\|_F^2, \ \text{s.t.} \|\boldsymbol{x}_m\|_0 \leqslant s, \ m = 1, \ldots, M.$$

where $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M] \in \mathbb{C}^{K \times M}$ is the coeffs matrix and $\mathbf{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_M] \in \mathbb{C}^{N \times M}$ the data matrix.

Two delicate aspects

Dictionary Learning and its initialization

- The DL problem is : find the dictionary $\mathcal{G}$ and the coeffs $\boldsymbol{x}_m$,

$$\min_{\mathcal{G},\mathbf{X}} \|\mathcal{G}\mathbf{X} - \mathbf{Y}\|_F^2, \text{ s.t.} \|\boldsymbol{x}_m\|_0 \leqslant s, \; m = 1, \ldots, M.$$

  where $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M] \in \mathbb{C}^{K \times M}$ is the coeffs matrix and $\mathbf{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_M] \in \mathbb{C}^{N \times M}$ the data matrix.
- We solve this using alternating minimization

Step I $\min \|\mathbf{X}\|_1 \;$ subject to $\mathcal{G}\mathbf{X} = \mathbf{Y}$

Step II $\min_{\mathcal{G}} \; \|\mathcal{G}\mathbf{X} - \mathbf{Y}\|_F^2$.

Two delicate aspects

Dictionary Learning and its initialization

- The DL problem is : find the dictionary $\boldsymbol{\mathcal{G}}$ and the coeffs $\boldsymbol{x}_m$,

$$\min_{\boldsymbol{\mathcal{G}},\mathbf{X}} \|\boldsymbol{\mathcal{G}}\mathbf{X} - \mathbf{Y}\|_F^2, \text{ s.t.} \|\boldsymbol{x}_m\|_0 \leqslant s, \; m = 1, \ldots, M.$$

where $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M] \in \mathbb{C}^{K\times M}$ is the coeffs matrix and $\mathbf{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_M] \in \mathbb{C}^{N\times M}$ the data matrix.

- We solve this using alternating minimization

Step I $\min \|\mathbf{X}\|_1 \;\; \text{subject to } \boldsymbol{\mathcal{G}}\mathbf{X} = \mathbf{Y}$

Step II $\min_{\boldsymbol{\mathcal{G}}} \;\; \|\boldsymbol{\mathcal{G}}\mathbf{X} - \mathbf{Y}\|_F^2 \;\; .$

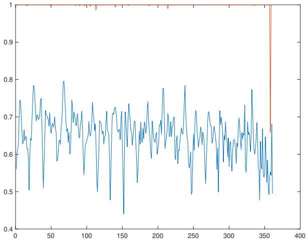
- Alternating minimization is non-convex. It converges if the initial guess $\boldsymbol{\mathcal{G}}$ is close to the true solution.

A probabilistic method for initialization

Compute the (projections of) *correlation-weighted covariance estimators*

$$\Sigma_j^{\mathrm{proj}} := \Sigma_j - \frac{\langle \Sigma_j, \Sigma \rangle_F}{\|\Sigma\|_F^2}\Sigma, \Sigma = \frac{1}{M}\sum_{m=1}^{M} y_m y_m^*, \Sigma_j := \frac{1}{M}\sum_{m=1}^{M} |\langle y_m, y_j \rangle|^2 y_m y_m^*.$$

Theorem (A. Novikov, S.White 2023) Suppose a measurement y_j comes from sensing at most s points, $s = o(N)$, and the columns of the sensing matrix $\mathcal{G}$ are incoherent. The s-dimensional subspace spanned by the principal eigenvectors of Σ_j^{proj} is close to the s-dimensional subspace spanned by the Green function vectors of these points, as $N \gg 1$ and $M \gg 1$.



JO des Poèts