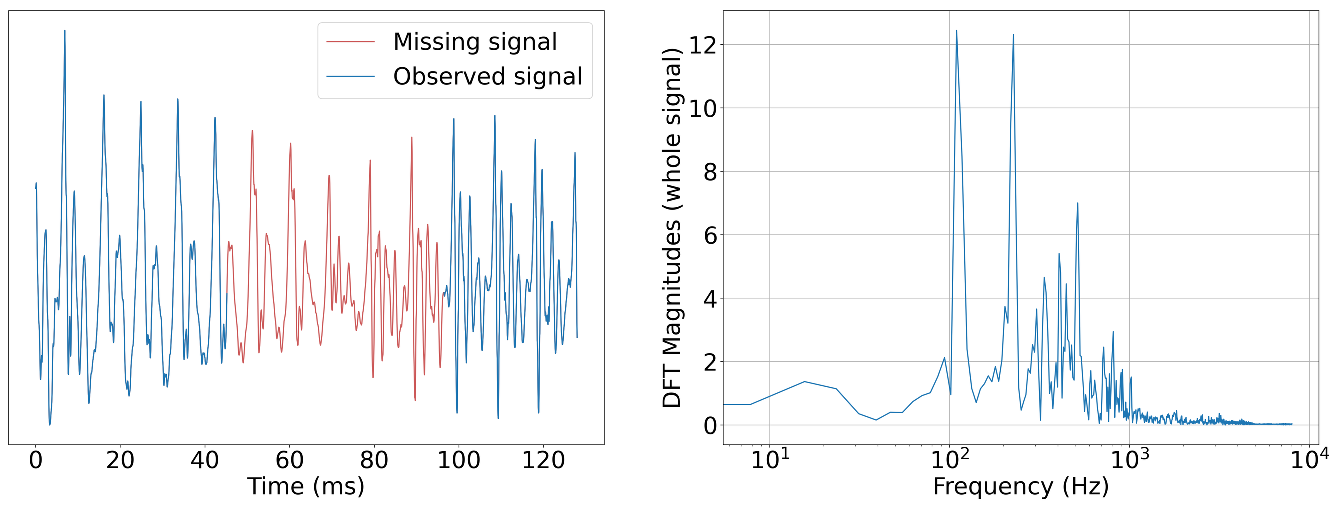
From audio inpainting to phase retrieval

Louis Bahrman, Marina Krémé, Paul Magron, Antoine Deleforge
Université de Lorraine, CNRS, Inria, Loria, F-54000 Nancy, France

OVERVIEW



Goal: reconstructing the missing samples $\boldsymbol{x}_{\bar{v}} \in \mathbb{R}^d$ of a signal $\boldsymbol{x}^\star \in \mathbb{R}^L$ from its discrete Fourier transform (DFT) magnitudes $\boldsymbol{b} \in \mathbb{R}_+^L$ and some known samples $\boldsymbol{x}_v = \boldsymbol{y} \in \mathbb{R}^{L-d}$.

Contributions

- A problem formulation that highlights the connection to phase retrieval
- A Gerchberg-Saxton (GS)-like alternating projection algorithm
- Investigation of a suitable initialization

PROPOSED METHOD

Problem setting

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}\in\mathbb{R}^L}{\text{argmin}}\, \| |\boldsymbol{\Phi}\boldsymbol{x}| - \boldsymbol{b}\|^2 \quad \text{s.t.} \quad \boldsymbol{x}_v = \boldsymbol{y}$$

where $\boldsymbol{\Phi} \in \mathbb{C}^{L\times L}$ is the DFT matrix.

Connection to phase retrieval

- Consider the auxiliary phase variable: $\boldsymbol{u} \in \mathbb{C}^L$ with $|\boldsymbol{u}| = 1$.
- The inpainting problem can be rewritten as [3]:

$$\boldsymbol{x}^*, \boldsymbol{u}^* = \underset{\boldsymbol{x}\in\mathbb{R}^L, \boldsymbol{u}\in\mathbb{C}^L}{\text{argmin}}\, \|\boldsymbol{\Phi}\boldsymbol{x} - \text{diag}(\boldsymbol{b})\boldsymbol{u}\|^2 \quad \text{s.t.} \quad |\boldsymbol{u}| = 1 \quad \text{and} \quad \boldsymbol{x}_v = \boldsymbol{y}$$

- Decompose $\boldsymbol{\Phi}$ and $\boldsymbol{x}$ as:

$$\boldsymbol{\Phi}\mathbf{x} = \begin{bmatrix} \boldsymbol{\Phi}_v & \boldsymbol{\Phi}_{\bar{v}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_v \\ \mathbf{x}_{\bar{v}} \end{bmatrix}$$

- Since $\boldsymbol{\Phi}$ is a unitary matrix (with Hermitian transpose $\boldsymbol{\Phi}^{\mathsf{H}}$), the problem rewrites:

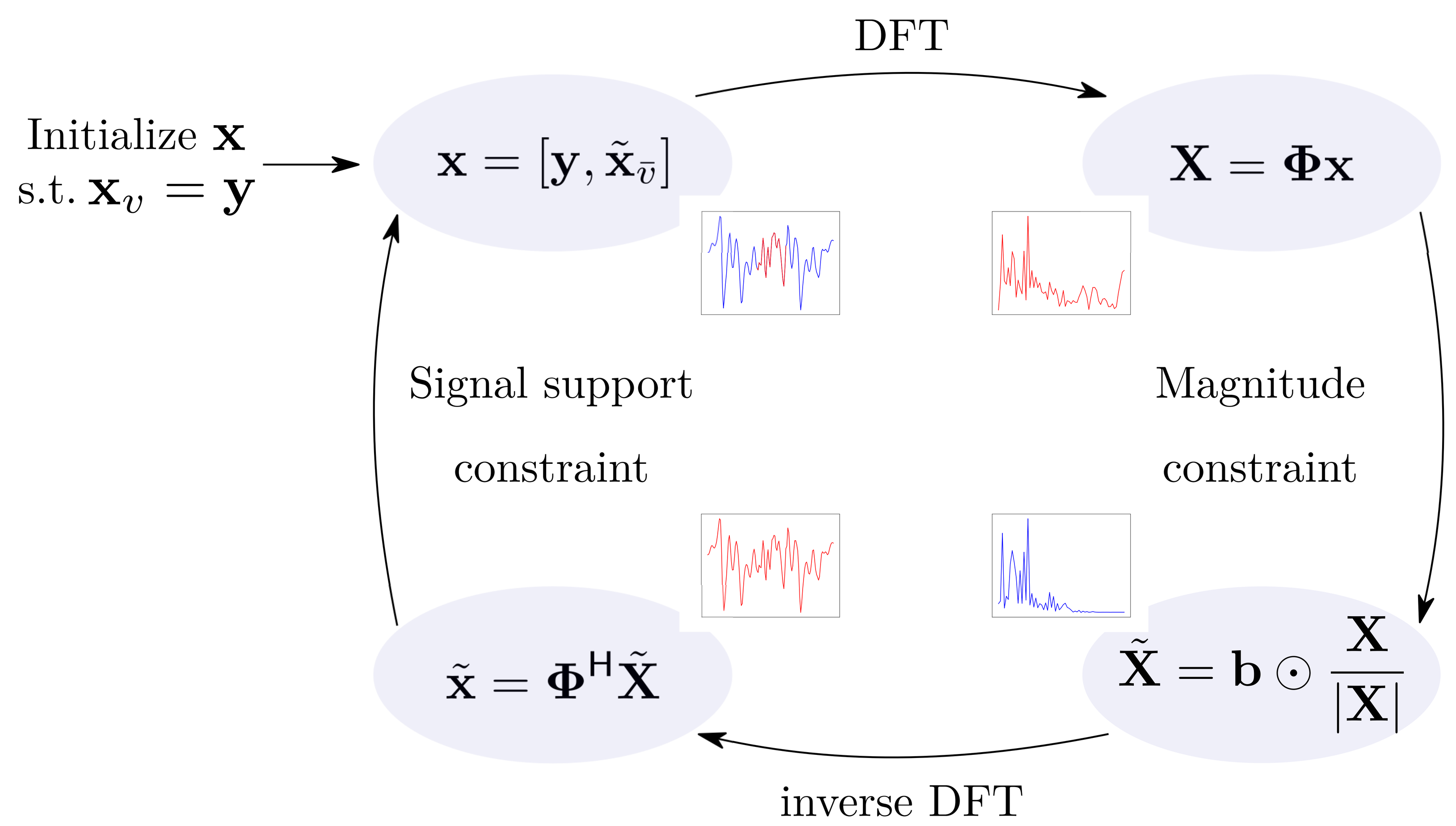
$$\boldsymbol{x}_{\bar{v}}^*, \boldsymbol{u}^* = \underset{\boldsymbol{x}_{\bar{v}}\in\mathbb{R}^d, \boldsymbol{u}\in\mathbb{C}^L}{\text{argmin}}\, \|\boldsymbol{y} - \boldsymbol{\Phi}_v^{\mathsf{H}}\text{diag}(\boldsymbol{b})\boldsymbol{u}\|^2 + \|\boldsymbol{x}_{\bar{v}} - \boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}}\text{diag}(\boldsymbol{b})\boldsymbol{u}\|^2 \quad \text{s.t.} \quad |\boldsymbol{u}| = 1$$

Alternating projections algorithm

- Given $\boldsymbol{u}$, the solution is $\boldsymbol{x}_{\bar{v}} = \boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}}\text{diag}(\boldsymbol{b})\boldsymbol{u}$.
- To find $\boldsymbol{u}$, we consider:

$$\boldsymbol{u}^* = \underset{\boldsymbol{u}\in\mathbb{C}^L}{\text{argmin}}\, \|\boldsymbol{x} - \boldsymbol{\Phi}^{\mathsf{H}}\text{diag}(\boldsymbol{b})\boldsymbol{u}\|^2 \quad \text{s.t.} \quad |\boldsymbol{u}| = 1$$

which leads to $\boldsymbol{u} = phase(\boldsymbol{\Phi}\boldsymbol{x}) = \dfrac{\boldsymbol{\Phi}\boldsymbol{x}}{|\boldsymbol{\Phi}\boldsymbol{x}|}$.



PRELIMINARY NUMERICAL SIMULATIONS

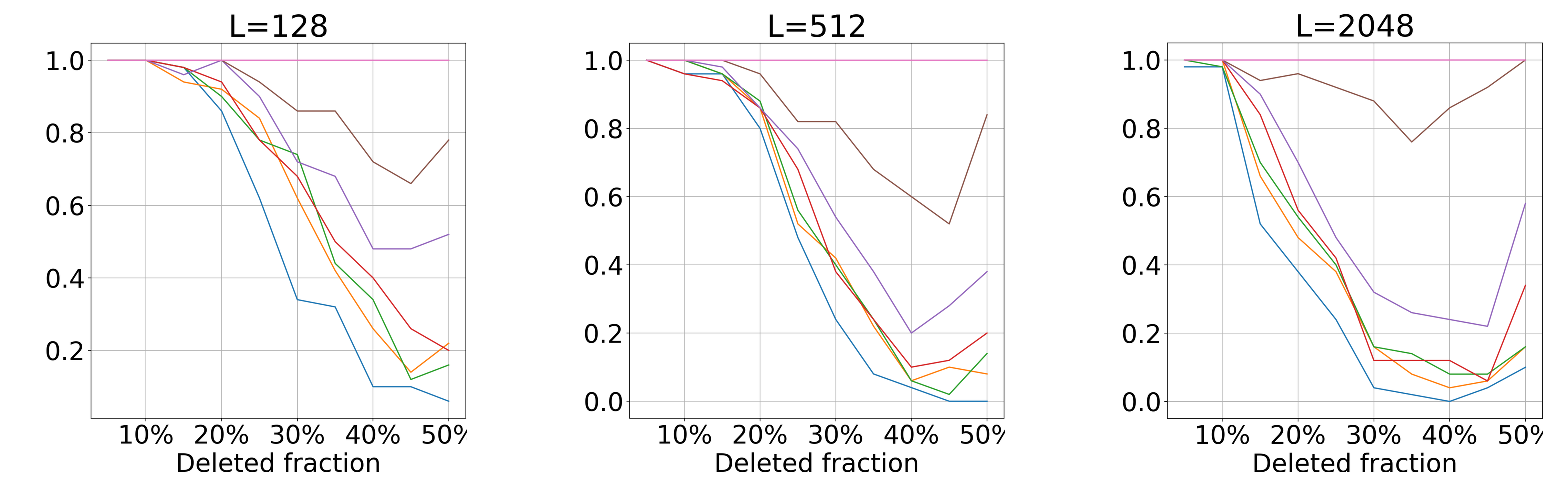
Data:

- 100 speech excerpts from the Librispeech dataset, sampled at 16 kHz.
- Signal length: $L \in \{128, 512, 2048\}$ samples.
- Fraction of deleted signal: from 5% to 50%.

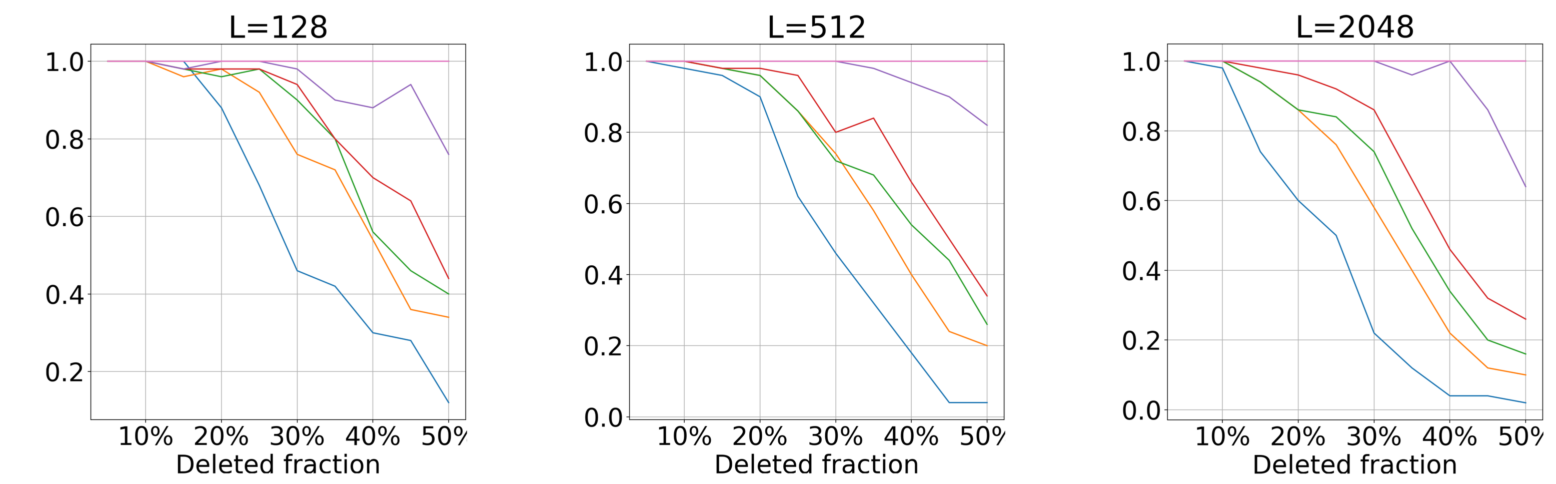
Initialization: Ground truth $\boldsymbol{x}^\star$ + white noise at various SNRs for the missing samples.

Results:

- Probability of reaching $\dfrac{\| |\boldsymbol{\Phi}\boldsymbol{x}^*| - \boldsymbol{b}\|^2}{d^2} \leq 10^{-8}$:



- Probability of reaching $\geq$ 20 dB of output SNR:



Legend: Initial SNR (dB): $-\infty$ -3 0 3 10 20 50

- Comments:
  - High probability to reach the true signal when the initial SNR is larger than 10 dB.
  - $\dfrac{\| |\boldsymbol{\Phi}\boldsymbol{x}^*| - \boldsymbol{b}\|^2}{d^2} \leq 10^{-8}$ implies almost surely that the output SNR is larger than 20 dB.
  - Beyond 40% of missing samples, the search space becomes very large: multiple solutions seem valid according to the first metric, whereas the output SNR remains low.

ONGOING WORK

- Theoretical study on the **uniqueness** and the **stability of the solution**.
- **Initialization strategy** by adapting the **spectral method** [2]:
  - Rewrite the problem as:

$$\begin{bmatrix} \boldsymbol{x}_{\bar{v}}^* \\ \cdot \end{bmatrix} = \underset{\tilde{\boldsymbol{x}}_{\bar{v}}\in\mathbb{R}^d}{\text{argmin}} \left\| \left| \widetilde{\boldsymbol{\Phi}} \begin{bmatrix} \boldsymbol{x}_{\bar{v}} \\ 1 \end{bmatrix} \right| - \boldsymbol{b} \right\|^2 \quad \text{with} \quad \widetilde{\boldsymbol{\Phi}} = [\boldsymbol{\Phi}_{\bar{v}}; \boldsymbol{\Phi}_v\boldsymbol{x}_v] \in \mathbb{C}^{L\times(d+1)}$$

  - Let $\widetilde{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{x}_{\bar{v}} \\ 1 \end{bmatrix}$. The above problem is equivalent to:

$$\widetilde{\boldsymbol{x}}^* = \underset{\widetilde{\boldsymbol{x}}\in\mathbb{R}^{d+1}}{\text{argmin}} \left\| |\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{x}}| - \boldsymbol{b} \right\|^2 \quad \text{s.t.} \quad \widetilde{\boldsymbol{x}}_{d+1} = 1$$

  - Without constraint, this problem boils down to phase retrieval $\rightarrow$ initialization with the spectral method:

$$\widetilde{\boldsymbol{x}}^{\text{sp}} \leftarrow \text{GEV}\left(\frac{1}{L}\sum_{k=1}^{L} b_k \boldsymbol{e}_k^{\mathsf{H}} \widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Phi}}^{\mathsf{H}} \boldsymbol{e}_k\right)$$

    (GEV: Eigenvector with the highest eigenvalue)
  - Considering the constraint, we seek the following initialization:

$$\boldsymbol{x}_{\bar{v}}^{(0)} = \underset{\boldsymbol{x}_{\bar{v}}\in\mathbb{R}^d}{\text{argmin}} \left\| \widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{x}}^{\text{sp}} - (\boldsymbol{\Phi}_v\boldsymbol{x}_v + \boldsymbol{\Phi}_{\bar{v}}\boldsymbol{x}_{\bar{v}}) \right\|^2$$

    which yields:

$$\boldsymbol{x}_{\bar{v}}^{(0)} = \boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}}\left(\boldsymbol{\Phi}_{\bar{v}}\widetilde{\boldsymbol{x}}_{:d}^{\text{sp}} + (\widetilde{\boldsymbol{x}}_{d+1}^{\text{sp}} - 1)\Phi_v x_v\right)$$

- **Lifting-based methods** [1] to find $\boldsymbol{u}$

$$\min_{\widetilde{\boldsymbol{U}}\succeq 0} \text{trace}(\boldsymbol{C}\widetilde{\boldsymbol{U}}) \quad \text{s.t diag}(\widetilde{\boldsymbol{U}}) = 1,$$

where $\boldsymbol{C} = \boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}$ with $\boldsymbol{A} = [\boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}}\text{diag}(\boldsymbol{b}), -\boldsymbol{x}_v]$ and $\widetilde{\boldsymbol{U}} = \tilde{\boldsymbol{u}}^{\mathsf{H}}\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}} = [\boldsymbol{u}, 1]^{\mathsf{T}}$

- Extension to **noisy observations** i.e. $|\boldsymbol{\Phi}\boldsymbol{x}| = \boldsymbol{b} + \eta$ where η is some noise.

REFERENCES

[1] Emmanuel J. Candès, Thomas Strohmer, and Vladislav Voroninski. "PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming". In: *CoRR* abs/1109.4499 (2011).
[2] Emmanuel J. Candès, Xiaodong Li, and Mahdi Soltanolkotabi. "Phase Retrieval via Wirtinger Flow: Theory and Algorithms". In: *Transactions on Information Theory* 61.4 (2015), pp. 1985–2007.
[3] Irène Waldspurger, Alexandre d'Aspremont, and Stéphane Mallat. "Phase recovery, maxcut and complex semidefinite programming". In: *Math. Progr. (Ser. A)* 149.1-2 (Feb. 2015), pp. 47–81.