A Parallel Algorithm for Phase Retrieval with Dictionary Learning

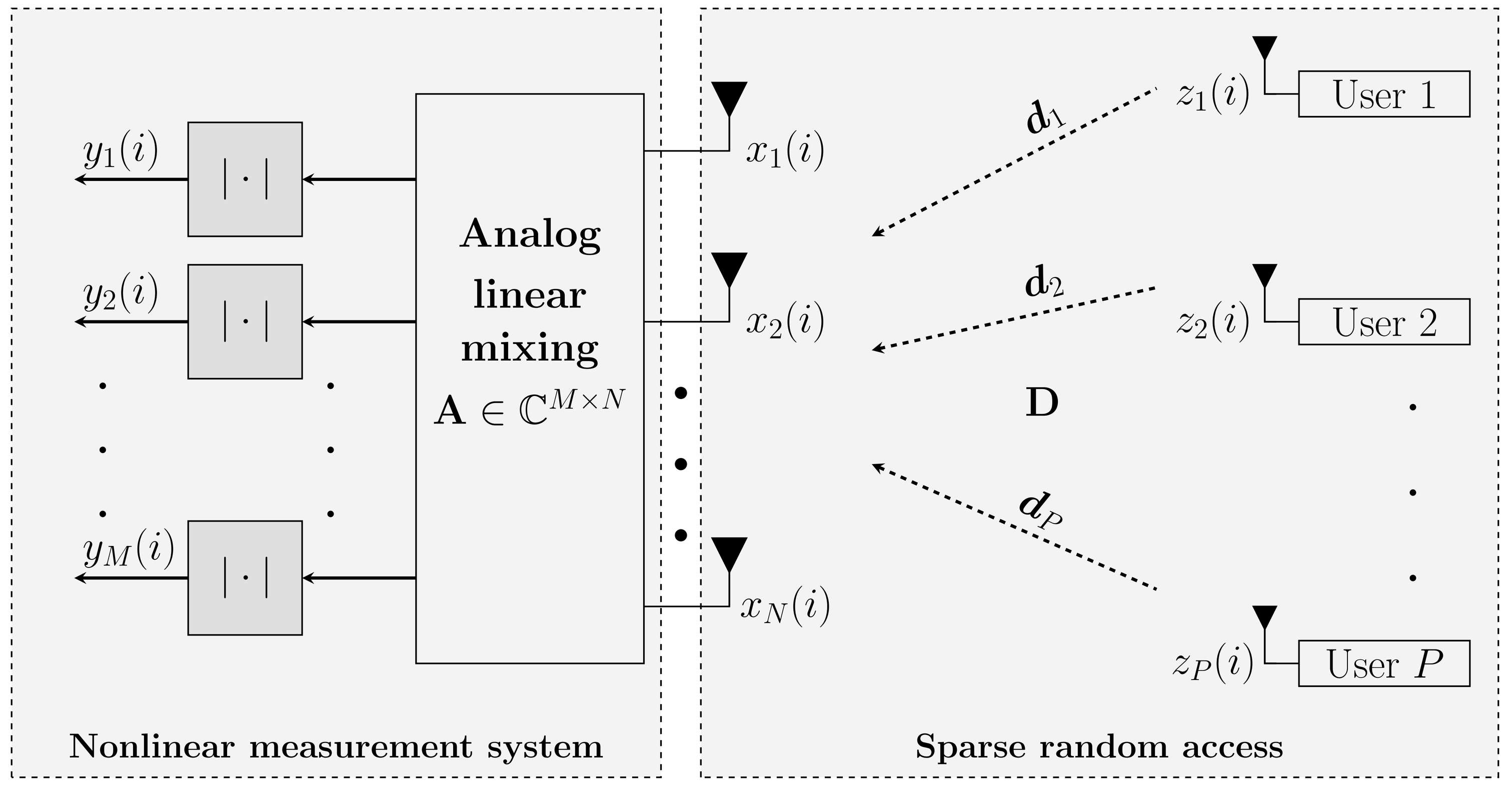
Tianyi Liu[1], Andreas M. Tillmann[2], Yang Yang[3], Yonina C. Eldar[4], Marius Pesavento[1]

TECHNISCHE UNIVERSITÄT DARMSTADT

Introduction

- Proposal of a new formulation for Phase Retrieval with Dictionary Learning with reduced variable size and number of regularization parameters
- Proposal of a algorithm that can be parallelized for the new formulation

Multi-antenna Random Access Network



Signal model: $\mathbf{Y} = |\mathbf{ADZ}| + \mathbf{N}$

- $\mathbf{Z} \in \mathbb{C}^{P\times I}$: Sparse transmitted signal matrix
- $\mathbf{D} \in \mathbb{C}^{N\times P}$: Spatial signature matrix
- $\mathbf{X} \in \mathbb{C}^{N\times I}$: Received signal matrix
- $\mathbf{Y} \in \mathbb{R}_+^{M\times I}$: Magnitude-only measurements
- $\mathbf{A} \in \mathbb{C}^{M\times N}$: Analog linear mixing network
- $\mathbf{N} \in \mathbb{R}^{M\times I}$: Sensor noise matrix

Problem Formulations

Objective: Jointly estimate the spatial signature $\mathbf{D}$ and the sparse transmitted signal $\mathbf{Z}$ from the magnitude-only measurements $\mathbf{Y}$

Proposed ℓ_1-regularized nonlinear least-squares (LS) formulation

$$\min_{\mathbf{D},\mathbf{Z}} \quad h(\mathbf{D},\mathbf{Z}) = \underbrace{\frac{1}{2}\|\mathbf{Y} - |\mathbf{ADZ}|\|_\mathsf{F}^2}_{f(\mathbf{D},\mathbf{Z})\text{: loss}} + \underbrace{\lambda\|\mathbf{Z}\|_{1,1}}_{g(\mathbf{Z})\text{: regularization}}$$

$$\text{s.t.} \quad \mathbf{D} \in \mathcal{D} = \{\mathbf{D} \mid \|\mathbf{d}_p\|_2 \le 1, p = 1,\dots,P\}$$

- $\|\mathbf{Z}\|_{1,1} = \sum_{p=1}^P \sum_{i=1}^I |z_{pi}|$
- $\lambda > 0$: regularization parameter
- Restrict $\mathbf{D}$ to avoid scaling ambiguities
- Loss f is nonsmooth and nonconvex

Alternative formulation (SC-PRIME) [1]

$$\min_{\mathbf{X},\mathbf{D}\in\mathcal{D},\mathbf{Z}} \frac{1}{2}\|\mathbf{Y} - |\mathbf{AX}|\|_\mathsf{F}^2 + \frac{\mu}{2}\|\mathbf{X} - \mathbf{DZ}\|_\mathsf{F}^2 + \rho\|\mathbf{Z}\|_{1,1}$$

- Nonconvex and nonsmooth
- Additional auxiliary variable $\mathbf{X}$
- $\mu, \rho > 0$: regularization parameters

Proposed Algorithm

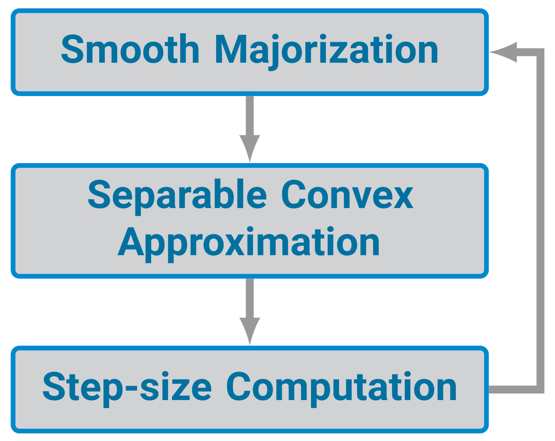


Fig. 1: Block Successive Convex Approximation (BSCA) framework

Smooth Majorization
Construct a smooth function $\bar{f}^{(t)}$ for function f at the iteration point $(\mathbf{D}^{(t)},\mathbf{Z}^{(t)})$ such that

$$\bar{f}^{(t)}(\mathbf{D}^{(t)},\mathbf{Z}^{(t)}) = f(\mathbf{D}^{(t)},\mathbf{Z}^{(t)})$$
$$\bar{f}^{(t)}(\mathbf{D},\mathbf{Z}) \ge f(\mathbf{D},\mathbf{Z}) \text{ for all } (\mathbf{D},\mathbf{Z})$$

Separable Convex Approximation
Design a separable convex approximation $\tilde{f}^{(t)}$ for $\bar{f}^{(t)}$ such that $(\widetilde{\mathbf{D}}^{(t)} - \mathbf{D}^{(t)}, \widetilde{\mathbf{Z}}^{(t)} - \mathbf{Z}^{(t)})$ is a descent direction of $\bar{f}^{(t)} + g$, where

$$(\widetilde{\mathbf{D}}^{(t)},\widetilde{\mathbf{Z}}^{(t)}) = \arg\min_{\mathbf{D}\in\mathcal{D},\mathbf{Z}} \tilde{f}^{(t)}(\mathbf{D},\mathbf{Z}) + g(\mathbf{Z})$$

Step-size Computation
Jointly update along the descent direction with a step-size obtained by line search on $\bar{f}^{(t)} + g$

Separable Convex Approximation

- "Best-response" approximation

$$(\widetilde{\mathbf{D}}^{(t)},\widetilde{\mathbf{Z}}^{(t)}) = \arg\min_{\mathbf{D}\in\mathcal{D},\mathbf{Z}} \underbrace{\sum_{p=1}^P \bar{f}^{(t)}(\mathbf{d}_p,\mathbf{D}_{-p}^{(t)},\mathbf{Z}^{(t)}) + \sum_{i=1}^I\sum_{p=1}^P \bar{f}^{(t)}(z_{pi},\mathbf{D}^{(t)},\mathbf{Z}_{-pi}^{(t)})}_{\tilde{f}^{(t)}(\mathbf{D},\mathbf{Z})} + \underbrace{\lambda\|\mathbf{Z}\|_{1,1}}_{g(\mathbf{Z})}$$

 - $\mathbf{D}_{-p}$ is the $N\times(P-1)$ matrix obtained by removing column $\mathbf{d}_p$ from $\mathbf{D}$
 - $\mathbf{Z}_{-pi}$ is the collection of all entries of $\mathbf{Z}$ except z_{pi}
- Decomposed into $P + (P\times I)$ independent subproblems: for $p = 1,\dots,P, i = 1,\dots,I$,

$$\tilde{\mathbf{d}}_p^{(t)} = \arg\min_{\mathbf{d}_p} \bar{f}^{(t)}(\mathbf{d}_p,\mathbf{D}_{-p}^{(t)},\mathbf{Z}^{(t)}) \quad \text{s.t. } \|\mathbf{d}_p\|_2 \le 1, \quad \leftarrow \quad \ell_2\text{-norm constrained LS}$$

$$\tilde{z}_{pi}^{(t)} = \arg\min_{z_{pi}} \bar{f}^{(t)}(z_{pi},\mathbf{D}^{(t)},\mathbf{Z}_{-pi}^{(t)}) + \lambda|z_{pi}|, \quad \leftarrow \quad \text{Single-variate LASSO}$$

- For the ℓ_2-norm constrained LS, by Karush-Kuhn-Tucher (KKT) conditions, the primal optimal solution:

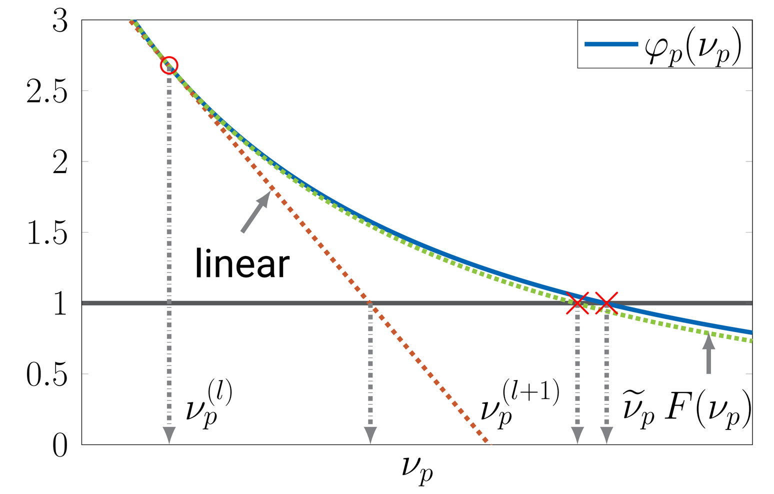
$$\tilde{\mathbf{d}}_p^{(t)} = \left(\mathbf{H}_p^\mathsf{H}\mathbf{H}_p + \tilde{\nu}_p^{(t)}\mathbf{I}_K\right)^{-1}\mathbf{H}_p^\mathsf{H}\mathbf{c}_p$$

- The dual optimal solution $\tilde{\nu}_p^{(t)}$ is obtained by solving the rational equation:

$$\underbrace{\sum_{i=1}^r \frac{|\tilde{c}_{ip}|^2}{(\sigma_i^2 + \nu_p)^2}}_{\varphi_p(\nu_p)} = 1, \quad \text{for } \nu_p > 0$$

- Successively approximate φ_p with a simple rational function [2]

$$F(\nu_p;\alpha,\beta) = \frac{\alpha}{(\beta - \nu_p)^2}$$



Smooth Majorization

- Majorization for the loss function f at the current point $(\mathbf{D}^{(t)},\mathbf{Z}^{(t)})$:

$$f(\mathbf{D},\mathbf{Z}) = \frac{1}{2}(\|\mathbf{Y}\|_\mathsf{F}^2 + \|\mathbf{ADZ}\|_\mathsf{F}^2) - \mathrm{tr}\{\mathbf{Y}^\mathsf{H}|\mathbf{ADZ}|\}$$
$$\le \frac{1}{2}(\|\mathbf{Y}\|_\mathsf{F}^2 + \|\mathbf{ADZ}\|_\mathsf{F}^2) - \mathrm{tr}\left\{\mathbf{Y}^\mathsf{H}\Re\left[(\mathbf{ADZ}) \odot e^{-\mathrm{j}\arg(\mathbf{AD}^{(t)}\mathbf{Z}^{(t)})}\right]\right\} = \underbrace{\frac{1}{2}\left\|\mathbf{Y}^{(t)} - \mathbf{ADZ}\right\|_\mathsf{F}^2}_{\bar{f}^{(t)}(\mathbf{D},\mathbf{Z})}$$

with $\mathbf{Y}^{(t)} = \mathbf{Y} \odot e^{\mathrm{j}\arg(\mathbf{AD}^{(t)}\mathbf{Z}^{(t)})}$

- Majorizing function $\bar{f}^{(t)}$ is smooth but nonconvex due to bilinear terms $\mathbf{DZ}$

Simulation Results

- As a comparison, we apply the BSCA framwork to solve the SC-PRIME formulation
- Accuracy evaluation metrics:
 - Minimum normalized squared error (MNSE)
 - F-measure is used to evaluate the accuracy of sparsity pattern of $\mathbf{Z}$
- Number of receive antennas $N = 16$; number of time-slots $I = 256$; SNR $= 15$ dB
- Gaussian mixing $\mathbf{A}$ with oversampling rate $M_1/N = 4$
- A debiasing step is performed for both algorithms

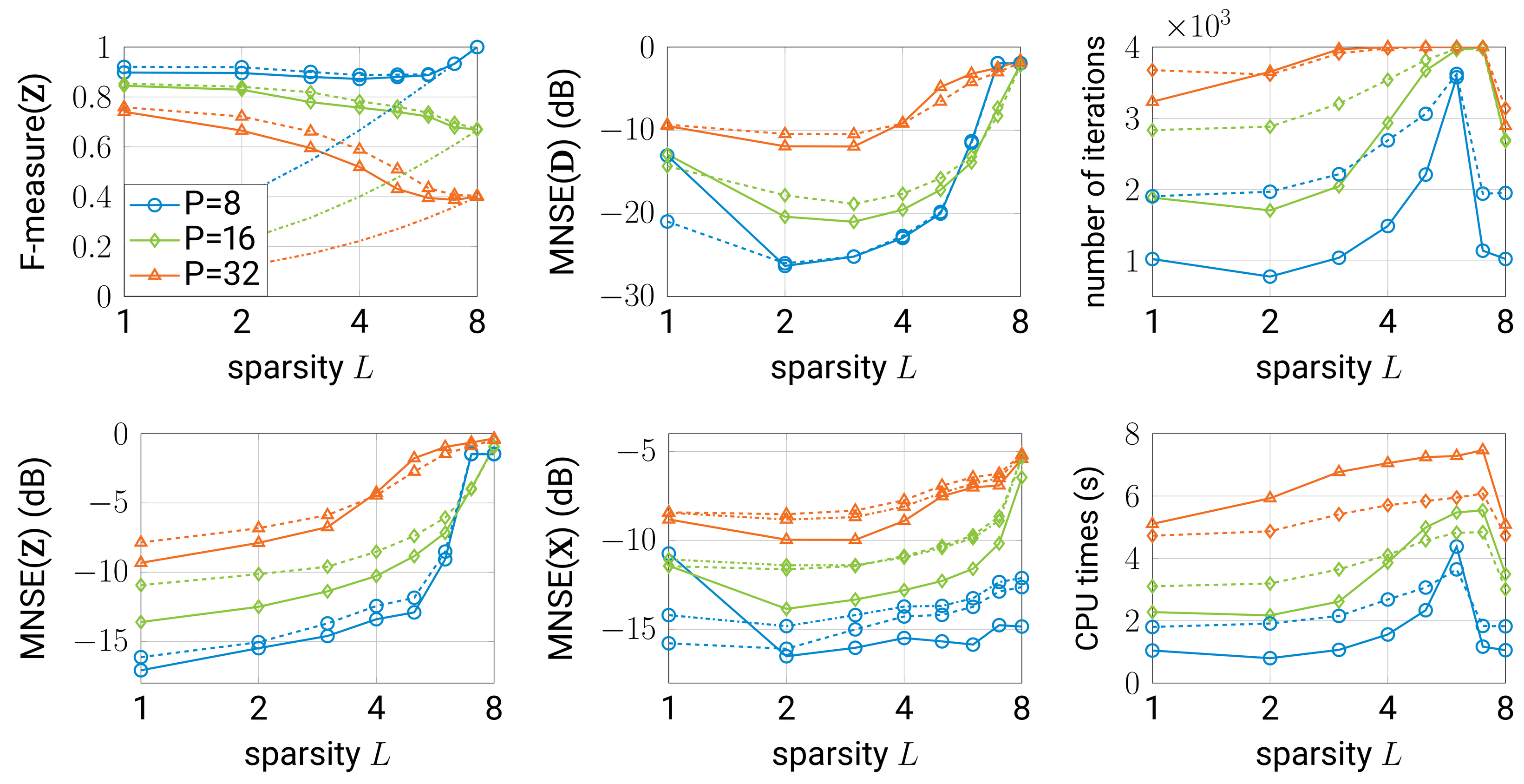


Fig. 2: Proposed (solid line) vs. SC-PRIME (dashed line)

Future Work

- Conditions on parameters for guaranteed recovery
- Initialization strategy for fast convergence to global optimum

Majorization Technique
For any $x \in \mathbb{C}$ and $\phi \in [0, 2\pi)$

$$|x| = |x \cdot e^{\mathrm{j}\phi}| \ge \Re\{x \cdot e^{\mathrm{j}\phi}\},$$

and the equality holds for $\phi = -\arg(x)$

[1] Communication Systems Group, Technische Universität Darmstadt, 64283 Darmstadt, Germany
[2] Institute for Mathematical Optimization, TU Braunschweig, 38106 Braunschweig, Germany
[3] Fraunhofer ITWM, 67663 Kaiserslautern, Germany
[4] Faculty of Math & CS, Weizmann Institute of Science, Rehovot 7610001, Israel

This work was supported by the EXPRESS project within the DFG priority program CoSIP (DFG-SPP 1798)

References:
[1] T. Qiu and D. P. Palomar, "Undersampled Sparse Phase Retrieval via Majorization–Minimization," IEEE Transactions on Signal Processing, vol. 65, no. 22, pp. 5957–5969, 2017
[2] J. R. Bunch, C. P. Nielsen, and D. C. Sorensen, "Rank-One Modification of the Symmetric Eigenproblem," Numer. Math, vol. 31, pp. 31–48, 1978.