On the Sensitivity of Spectral Initialization for Noisy Phase Retrieval

Vincent Monardo and Yuejie Chi
Department of Electrical and Computer Engineering
Carnegie Mellon University
Email: {vmonardo,yuejiec}@andrew.cmu.edu

Carnegie Mellon University

Estimation with spectral methods

- **Goal:** estimate $\boldsymbol{x}^\natural \in \mathbb{C}^n$ from m generalized linear measurements,

$$y_i \sim p(y \mid \langle \boldsymbol{a}_i, \boldsymbol{x}^\natural \rangle), \quad i = 1, 2, ..., m.$$

- The spectral method estimates $\boldsymbol{x}^\natural$ by taking the top eigenvector of a carefully constructed data matrix of the form:

$$\boldsymbol{D} = \frac{1}{m} \sum_{i=1}^{m} \mathcal{T}(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^H.$$

where $\mathcal{T}(\cdot)$ are preprocessing functions of the data.

- For example, for the celebrated phase retrieval problem, the spectral method can be used to initialize a nonconvex iterative method such as gradient descent or alternating minimization or provide an anchor vector to a convex linear program such as the Phasemax.

Key metrics for preprocessing functions

- Preprocessing functions can significantly improve spectral estimation when they are designed for a given observation model.
- We compare their performance using two key metrics:
  - *cosine-squared similarity*:

$$\rho(\hat{\boldsymbol{x}}, \boldsymbol{x}^\natural) = \frac{|\langle \hat{\boldsymbol{x}}, \boldsymbol{x}^\natural \rangle|^2}{\|\hat{\boldsymbol{x}}\|_2^2 \|\boldsymbol{x}^\natural\|_2^2}.$$

  - *sampling threshold*:

$$\alpha_u = \underset{\alpha^*}{\operatorname{argmin}} \left\{ \forall \alpha > \alpha^*, \liminf_{m \to \infty} \mathbb{E}_y \left\{ \rho(\hat{\boldsymbol{x}}, \boldsymbol{x}^\natural) \right\} > 0 \right\}.$$

The optimal preprocessing function minimizes the sampling threshold α_u and maximizes $\rho(\hat{\boldsymbol{x}}, \boldsymbol{x}^\natural)$ for a fixed sample complexity.

Optimal preprocessing functions

Theorem 1 ([1]). *Define $s = \langle \boldsymbol{a}_i, \boldsymbol{x}^\natural \rangle$. Then the optimal preprocessing function for a pair of sensing vectors and noise distribution is given by*

$$\mathcal{T}(y) = 1 - \frac{\mathbb{E}_s\{p(y \mid |s|)\}}{\mathbb{E}_s\{|s|^2 p(y \mid |s|)\}}. \tag{1}$$

Furthermore, the sampling threshold can be derived as

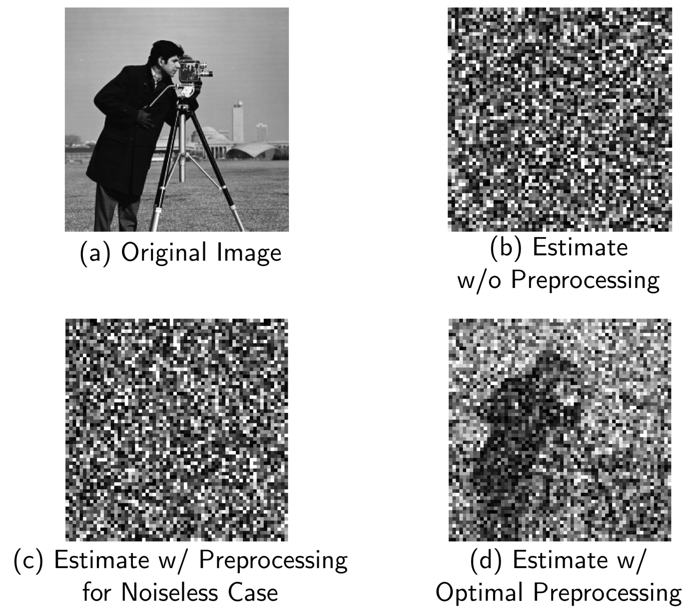
$$\alpha_u = \left(\int_{\mathbb{R}} \frac{\mathbb{E}_s \left\{ p(y \mid |s|)(1 - |s|^2) \right\}}{\mathbb{E}_s \left\{ p(y \mid |s|) \right\}} \right)^{-1}. \tag{2}$$

In the noiseless case, $\mathcal{T}(y) = 1 - 1/y$ and $\alpha_u = 1$.

Gaussian and Poisson noise

We derive optimal preprocessing functions tailored to Gaussian noise and to Poisson noise:

- Gaussian Noise: $\mathcal{T}_\sigma(y) = 1 - \left(y - \sigma^2 + \sqrt{\frac{\sigma^2}{2\pi}} \frac{\exp\left(-\frac{(y-\sigma^2)^2}{2\sigma^2}\right)}{\mathcal{Q}\left(-\frac{(y-\sigma^2)}{\sigma}\right)} \right)^{-1}.$
  - Sampling threshold: $\alpha_u = \left(1 - \sigma^2 - \sigma^4 + \frac{\sigma^3 e^{-\sigma^2/2}}{2\pi} \int_{-\infty}^{\infty} \frac{\exp\left(-\sigma u - u^2\right)}{\mathcal{Q}(-u)} du \right)^{-1}$
- Poisson Noise: $\mathcal{T}(y) = \frac{y-1}{y+1}$, sampling threshold: $\alpha_u = 2$.
- Preprocessing functions are model-dependent; we investigate their performance in the regime where there is a mismatch between the hypothesized noise model and the true noise model.



- Example: With additive white Gaussian noise $\sigma^2 = 1.0, n = 64 \times 64$, and $\alpha = 5$, we compare no preprocessing, preprocessing for the noiseless case, and optimal preprocessing.

Sensitivity of sampling thresholds

We assume that a fixed preprocessing function is used to process the measurements, designed for a postulated noise level σ_f^2, while the true noise level is set at σ_r^2, and pinpoint the sampling threshold.

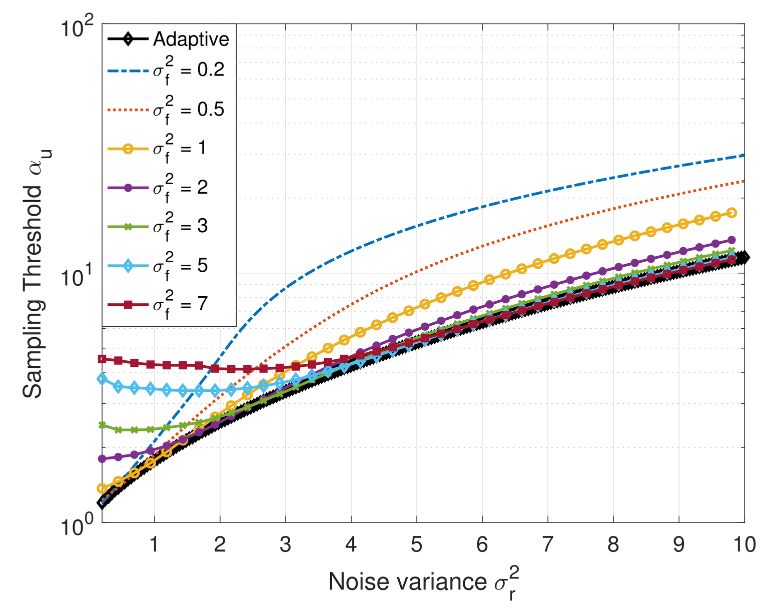


Figure 1: The sampling threshold α_u with respect to the true AWGN noise level σ_r^2 for a preprocessing function designed to be optimal for a postulated noise level $\sigma_f^2 = 0.2, 0.5, 1, 2, 3, 5, 7$, respectively.

Sensitivity of cosine-squared similarities

We theoretically and emprically demonstrate the cosine-squared similarity of various preprocessing functions over a wide range of sampling ratios.

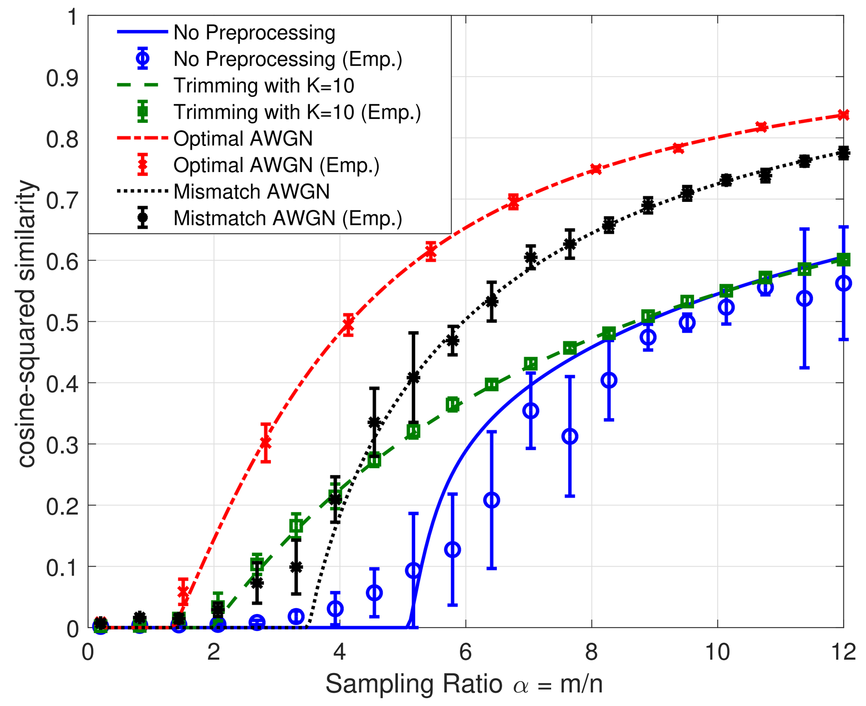


Figure 2: The theoretical prediction and the empirical realizations of the cosine-squared similarity of each preprocessing function with respect to the sampling ratio under AWGN with $\sigma_r^2 = 0.5$.

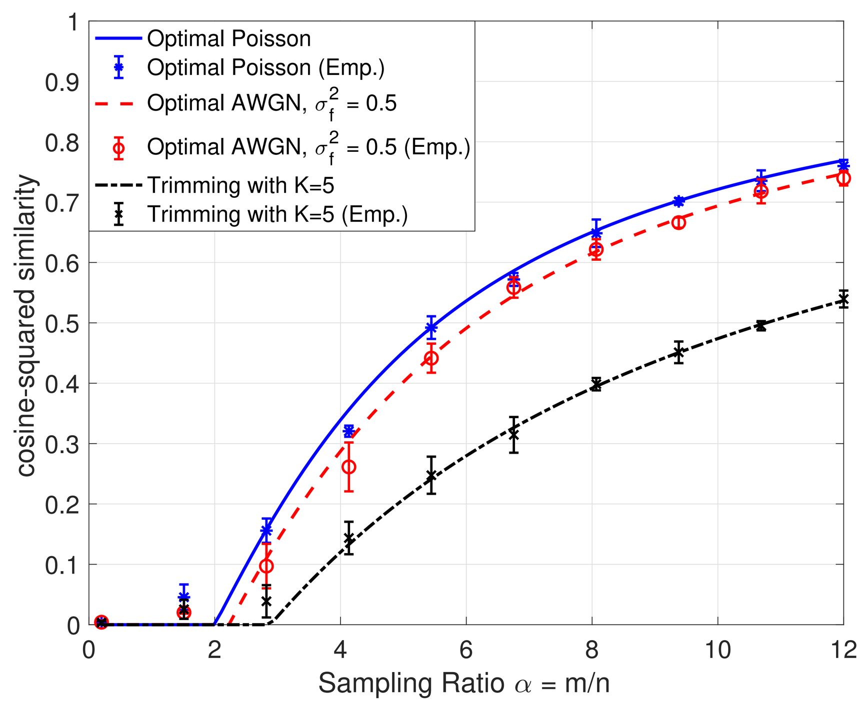


Figure 3: The theoretical prediction and the empirical realizations of the cosine-squared similarity of each preprocessing function with respect to the sampling ratio under Poisson noise.

References

[1] W. Luo, W. Alghamdi, and Y. M. Lu. Optimal spectral initialization for signal recovery with applications to phase retrieval. 2018.

[2] M. Mondelli and A. Montanari. Fundamental limits of weak recovery with applications to phase retrieval. 2017.

Acknowledgements

This work was supported in part by ONR under the grant N00014-18-1-2142, by ARO under the grant W911NF-18- 1-0303, and by NSF under the grants CAREER ECCS-1818571, ECCS-1833553 and CCF-1806154.