ON THE SENSITIVITY OF SPECTRAL INITIALIZATION FOR NOISY PHASE RETRIEVAL

Vincent Monardo and Yuejie Chi

Department of Electrical and Computer Engineering
Carnegie Mellon University
Emails: {vmonardo, yuejiec}@andrew.cmu.edu

ABSTRACT

The spectral method is an important approach for signal estimation that is often used as an initialization to iterative methods as well as a stand-alone estimator, where the signal is estimated by the top eigenvector of certain carefully-constructed data matrix. A recent line of work has characterized the asymptotic behavior of such data matrices used in spectral methods, which reveals an interesting phase transition phenomenon: there exists a critical sampling threshold below which the estimate of the spectral method is uninformative. Furthermore, optimal preprocessing functions are developed to minimize this critical sampling threshold. In particular, most of the existing work is focused on the noiseless phase retrieval problem. In this paper, our goal is to examine the sensitivity of such optimal preprocessing functions in noisy phase retrieval, when there is a mismatch between the noise model used in deriving the optimal preprocessing function and the actual noise model in practice. Our results provide important insights into the choice of preprocessing functions in spectral methods.

Index Terms— spectral method, phase retrieval, noise sensitivity

1. INTRODUCTION

Consider the problem of estimating an n-dimensional vector $\boldsymbol{x}^\natural \in \mathbb{C}^n$ from a set of m generalized linear measurements of the form

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

where $p(\cdot \mid \cdot)$ is a known conditional probability density that describes how the measurements are obtained, $\{\boldsymbol{a}_i\}_{i=1}^m$ is an ensemble of sensing vectors, and $\langle \cdot, \cdot \rangle$ is the inner product. The spectral method [1, 2] is a popular approach for estimating $\boldsymbol{x}^\natural$, where it is estimated via the top eigenvector (up to scaling) of a carefully constructed data matrix that is a sum of rank-one matrices $\boldsymbol{a}_i \boldsymbol{a}_i^H$, each weighted by the corresponding measurement y_i, or a function $\mathcal{T}(y_i)$ of it. The spectral method can either be used as a stand-alone estimator or as the initialization of a more sophisticated method. 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 [3–9] or provide an anchor vector to a convex linear program such as the Phasemax [10, 11].

Recently, a few works [12, 13] studied the asymptotic performance of the spectral method under the Gaussian design, where the sensing vectors are generated with i.i.d. standard complex Gaussian entries, in the regime where both m and n go to infinity with a fixed sampling ratio $\alpha = m/n$. It not only provides a precise characterization of the performance of the spectral method, but also reveals an interesting phase transition phenomenon: there exists a critical sampling ratio threshold such that below which the estimate of the spectral method is uninformative, i.e. it is orthogonal to the ground truth signal. Moreover, [13] provides formula for an optimal *preprocessing* function $\mathcal{T}(y_i)$ to minimize this critical sampling threshold.

The goal of this paper is to study the sensitivity of spectral initialization under model mismatch, when the actual signal model in practice is different from the one used to derive the optimal preprocessing function. This is of great relevance in practice, since typically the model is only imperfectly known, and may change during deployment. Therefore, it is necessary to see if the performance of the preprocessing functions are robust to model mismatch.

1.1. Our Contributions

In this work, we demonstrate the impact of model mismatch on the performance of preprocessing functions that are derived for specific noise models. In order to reach this goal, we consider the problem of recovering an n-dimensional complex signal from m quadratic measurements, known as *phase retrieval*, which is also studied in [12,13] for the noiseless setting. In contrast, we consider the noisy setting, where the measurements are corrupted by either additive white Gaussian noise (AWGN) or Poisson noise. We first derive optimal preprocessing functions using the formula provided in [13], and characterize how the critical sampling threshold varies as a function of the noise level. Furthermore, through both empirical experiments and theoretical analyses, we examine the performance of the preprocessing functions when the noise level of the measurements are different from the one used in the derivation, which suggest some preprocessing functions are more sensitive to model perturbations.

1.2. Related Works

Though we use phase retrieval as an example, the spectral method has been applied to many statistical estimation problems such as low-rank matrix estimation [14–16], blind deconvolution [7, 17], subspace estimation [18], to name a few. Many regularized variants of the spectral method have been proposed to improve its performance, which typically apply truncation or trimming to remove measurements that have high leverage [4, 19]. The asymptotic performance of the spectral method is analyzed first in [12] and then [13] for generalized linear models.

Before continuing, it is worth emphasizing the critical role of initialization in nonconvex statistical estimation. For several problems such as phase retrieval [20], low-rank matrix sensing [21] and completion [22], it is shown that with high probability, there are no spurious local minima in the landscape of the loss function except

*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.

strict saddle points. Therefore, gradient descent with random initialization converge to the global optima almost surely for such problems [23], however the iteration complexity can be very high. In contrast, an optimally designed spectral initialization can provably land in a local basin of attraction near the ground truth [3], leading to faster convergence.

1.3. Paper Organization and Notations

The rest of this paper is organized as follows. Section 2 presents the signal model, and provides key metrics and backgrounds. Section 3 presents the optimal preprocessing functions derived for different noise models in phase retrieval. Section 4 provides numerical experiments to study the noise sensitivity of different preprocessing functions. Finally, we conclude in Section 5.

Throughout the paper, we use boldfaced symbols to represent vectors and matrices. For any vector $\boldsymbol{v}$, we let $\|\boldsymbol{v}\|_2$ denote the ℓ_2 norm, and let $\boldsymbol{v}^H$ indicate the conjugate transpose. We use $\mathcal{Q}(x)$ to represent the tail distribution function of the standard normal distribution, i.e. $\mathcal{Q}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)dz$. Given two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^n x_i y_i^*$. We let $\mathbb{E}_x\{\cdot\}$ denote taking the expectation with respect to a random variable x.

2. BACKGROUNDS

In this section, we present the signal model for noisy phase retrieval, define key metrics and review the phase transition phenomenon as well as the design of optimal preprocessing functions in [12, 13] for the spectral method.

2.1. Signal Model and Important Definitions

We consider the problem of recovering an n-dimensional signal $\boldsymbol{x}^\natural \in \mathbb{C}^n$ from m intensity (quadratic) measurements in the presence of noise, where each measurement y_i is collected according to

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

The set of known sensing vectors $\boldsymbol{a}_i$ is independently drawn from a complex Gaussian distribution, namely $\boldsymbol{a}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n) + j\mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n)$, where $\boldsymbol{I}_n$ is the identity matrix of size n. Let $\alpha = m/n$ be defined as the sampling ratio. Furthermore, $p(\cdot)$ specifies some noise distribution such as additive white Gaussian noise (AWGN), or Poisson noise.

The spectral method first constructs a data matrix

$$\boldsymbol{D} = \frac{1}{m} \sum_{i=1}^m \mathcal{T}(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^H, \tag{3}$$

where $\mathcal{T} : \mathbb{R} \longrightarrow \mathbb{R}$ is a deterministic function which we refer to as a *preprocessing function*. With spectral initialization, we take the top eigenvector $\hat{\boldsymbol{x}} \in \mathbb{C}^n$ of $\boldsymbol{D}$ as the initialization[1]. The intuition behind the spectral initialization is that, if we let $m \to \infty$, the top eigenvector of $\boldsymbol{D}$ perfectly recovers $\boldsymbol{x}^\natural$ up to scaling. In practice, it is desirable to carefully design the preprocessing function so that $\hat{\boldsymbol{x}}$ is as close as possible to $\boldsymbol{x}^\natural$ when it is calculated using a finite number of measurements. Thus, we wish to carefully design preprocessing functions that leverage our knowledge of the measurement model (2) in order to obtain a desirable initialization.

[1]The norm $\|\boldsymbol{x}^\natural\|_2$ can be estimated easily, for example using the average of the measurements, which is not the focus here.

We study the regime where $m \to \infty$ and $n \to \infty$, but their ratio tends to a positive constant, i.e. $m/n \to \alpha$. The two metrics we will consider to determine the success of a preprocessing function are the *cosine-squared similarity* between the ground truth vector $\boldsymbol{x}^\natural$ and the top eigenvector $\hat{\boldsymbol{x}}$ of the data matrix $\boldsymbol{D}$:

$$\rho(\hat{\boldsymbol{x}}, \boldsymbol{x}^\natural) = \frac{|\langle \hat{\boldsymbol{x}}, \boldsymbol{x}^\natural \rangle|^2}{\|\hat{\boldsymbol{x}}\|_2^2 \, \|\boldsymbol{x}^\natural\|_2^2}, \tag{4}$$

and the *sampling threshold*,

$$\alpha_u = \underset{\alpha^*}{\operatorname{argmin}} \left\{ \forall \alpha > \alpha^*, \ \liminf_{m \to \infty} \mathbb{E}_y \left\{ \rho(\hat{\boldsymbol{x}}, \boldsymbol{x}^\natural) \right\} > 0 \right\}, \tag{5}$$

the minimum sampling threshold required to have a non-zero cosine-squared similarity.

By analyzing the asymptotic characterization of the spectral method, one can derive the behavior of the cosine-squared similarity as a function of the sampling ratio. The cosine-squared similarity $0 \le \rho(\hat{\boldsymbol{x}}, \boldsymbol{x}^\natural) \le 1$ gives us a measure of how aligned our estimate is with the ground truth, by specifying the correlations between these two vectors. In [12], Lu and Li provided precise asymptotic predictions of the cosine-squared similarity as a function of the sampling threshold α and the distribution $p(\cdot)$. These predictions highlight the existence of a phase transition phenomenon, such that there exists a sampling threshold α_u that determines the effectiveness of the spectral method. When the sampling ratio α is below the threshold $\alpha < \alpha_u$, we have $\rho = 0$ and the spectral method is uninformative; and when $\alpha > \alpha_u$, we have ρ bounded away from 0, and the spectral estimate can be computed efficiently via the power method. The precise form of $\rho(\hat{\boldsymbol{x}}, \boldsymbol{x}^\natural)$ with respect to α is derived in [12] for generalized linear models including phase retrieval.

2.2. Optimal Preprocessing Functions

The results in [12] suggest that the performance of spectral initialization can be drastically different using different preprocessing functions. In [13], an *optimal* preprocessing function was proposed to optimize the sampling threshold, so that it obtains the minimum sampling threshold and the best cosine-squared similarity for all sampling ratios. In [**?**], Luo et. al. constructed an optimal design of spectral methods that is *uniformly optimal* for all sampling ratios. We present Theorem 1 from [**?**] below.

Theorem 1 (Theorem 1 from [**?**]). *Let $\boldsymbol{x}^\natural \in \mathbb{C}^n$, $\{\boldsymbol{a}_i\}_{i=1}^m$ be a known sensing vector ensemble, and $y_i \sim p(y \mid |\langle \boldsymbol{a}_i, \boldsymbol{x}^\natural \rangle|^2)$. 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*

$$\mathcal{T}(y) = 1 - \frac{\mathbb{E}_s\{p(y \mid |s|)\}}{\mathbb{E}_s\{|s|^2 p(y \mid |s|)\}}. \tag{6}$$

Furthermore, the sampling threshold can be derived as

$$\alpha_u = \left(\int_{\mathbb{R}} \frac{\mathbb{E}_s\left\{p(y \mid |s|)(1 - |s|^2)\right\}}{\mathbb{E}_s\left\{p(y \mid |s|)\right\}} \right)^{-1}. \tag{7}$$

This theorem is very useful for obtaining optimal preprocessing functions for a given measurement ensemble (2). In [13], an optimal preprocessing function was derived for the noiseless phase retrieval problem, given as $\mathcal{T}(y) = 1 - 1/y$, with a corresponding sampling threshold $\alpha_u = 1$.

3. OPTIMAL PREPROCESSING FUNCTIONS FOR NOISY PHASE RETRIEVAL

We proceed to present the optimal preprocessing functions for the cases where the measurements are contaminated by AWGN and Poisson noise in phase retrieval (2), as well as the corresponding theoretical sampling threshold, which were not considered in [13]. Due to space limits, all the proofs are omitted and can be found at the full version at [24]. In Propositions 1 and 2, the preprocessing functions and sampling thresholds are obtained using (6) and (7), respectively, by plugging in the relevant probability distribution and sensing vector model.

Proposition 1 (AWGN, Complex Gaussian Sensing Vectors)**.** *Consider the set up described in Theorem 1, where* $\boldsymbol{a}_i \overset{i.i.d.}{\sim} \mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n) + j\mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n)$ *and* $p(y \mid |\langle \boldsymbol{a}_i, \boldsymbol{x}^\natural\rangle|^2) \sim \mathcal{N}(|\langle \boldsymbol{a}_i, \boldsymbol{x}^\natural\rangle|^2, \sigma^2)$*. Then the optimal preprocessing function is given as*

$$\mathcal{T}_\sigma(y) = 1 - \left(y - \sigma^2 + \sqrt{\frac{\sigma^2}{2\pi}} \frac{\exp\left(-\frac{(y-\sigma^2)^2}{2\sigma^2}\right)}{\mathcal{Q}\left(-\frac{(y-\sigma^2)}{\sigma}\right)} \right)^{-1}, \quad (8)$$

and the corresponding sampling threshold is

$$\alpha_u = \left(1 - \sigma^2 - \sigma^4 + \frac{\sigma^3 e^{-\sigma^2/2}}{2\pi} \int_{-\infty}^{\infty} \frac{\exp\left(-\sigma u - u^2\right)}{\mathcal{Q}(-u)} du \right)^{-1}. \quad (9)$$

It is evident that both the optimal preprocessing function and the sampling threshold depends on the noise level σ^2 in a nonlinear manner. Fig. 1 shows the optimal preprocessing function $\mathcal{T}_\sigma(y)$ with respect to y for varied noise levels. For smaller measurements, i.e. $0 < y < 1$, the preprocessing function maps the measurements to a wide range of negative values; as the value of the measurement continues to increase $y_i > 1$, the preprocessing functions maps the measurements to the range $(0, 1)$. When there is no noise $\sigma^2 = 0$, this recovers the optimal preprocessing function introduced in [13], that is $\mathcal{T}_0(y) = 1 - 1/y$. At we increase the noise level, the dynamic range of the $\mathcal{T}_\sigma(y)$ also decreases.

Proposition 2 (Poisson Noise, Complex Gaussian Sensing)**.** *Consider the set up described in Theorem 1, where* $\boldsymbol{a}_i \overset{i.i.d.}{\sim} \mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n) + j\mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n)$ *and* $p(y \mid |\langle \boldsymbol{a}_i, \boldsymbol{x}^\natural\rangle|^2)$ *follows a Poisson distribution with rate* $|\langle \boldsymbol{a}_i, \boldsymbol{x}^\natural\rangle|^2$*. Then the optimal preprocessing function is given as*

$$\mathcal{T}(y) = \frac{y-1}{y+1}. \quad (10)$$

The sampling threshold is $\alpha_u = 2$*.*

Similar to the AWGN case, the optimal preprocessing function for Poisson noise maps smaller measurements to a wider range of values (in a relative sense), while mapping larger measurements to values approaching 1 as $y \to \infty$, similar to a simple truncation scheme. For example, $y \in \{0, 1, ..., 5\}$ maps to a range of $-1 \leq \mathcal{T}(y) \leq \frac{2}{3}$, while for $y_i \geq 6$, $\frac{5}{7} \leq \mathcal{T}(y) < 1$. Additionally, notice that the sampling threshold for Poisson noise is 2, in contrast to the sampling threshold in the AWGN case, which can approach 1 when the noise goes to zero. In comparison, to achieve the same sampling threshold, processing the Poisson noise is about the same as AWGN at $\sigma^2 \approx 1.34$.

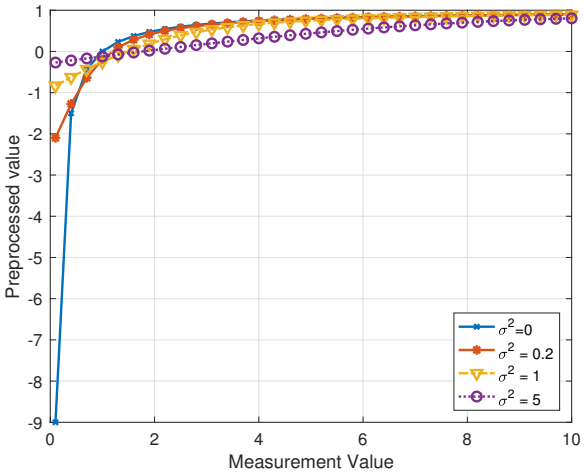

Fig. 1. The optimal preprocessing function $\mathcal{T}_\sigma(y)$ with respect to the measurement value y for noisy phase retrieval with AWGN under different noise levels $\sigma^2 = 0, 0.2, 1$, and 5.

4. SENSITIVITY STUDIES VIA NUMERICAL EXPERIMENTS

In this section, we provide numerical experiments that demonstrate the sensitivity to model mismatch of the preprocessing function performance, focusing on the cases of AWGN and Poisson noise for noisy phase retrieval. Specifically, imagine one applies the optimal preprocessing function designed for a postulated measurement model, while the actual measurement noise is different from the postulated one. This situation is highly relevant in practice, since we either do not have perfect knowledge of the model, or the model might change during measurements collection. We will examine the sensitivity of both the sampling threshold and the cosine-squared similarity under model mismatch.

For all experiments, the sensing vectors $\boldsymbol{a}_i$ are identically and independently distributed as complex Gaussian vectors, i.e. $\boldsymbol{a}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n) + j\mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n)$, for $i = 1, ..., m$. For the cosine-squared similarity curves, we refer to "trimming" as the preprocessing function where $\mathcal{T}(y) = \min(y, K)$ for some constant K. For testing the empirical performance of each preprocessing function, we run 8 Monte Carlo experiments, where the signal dimensions are fixed to $n = 1024$. We note that as the signal dimension increases, the empirical curve becomes closer to the theoretical curve. The theoretical curves for the cosine-squared similarity are obtained following the derivations from [12].

4.1. Sensitivity of Sampling Thresholds

We start by investigating the sampling threshold of the optimal preprocessing functions in AWGN for noisy phase retrieval. 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. Fig. 2 plots the sampling threshold of the preprocessing function with respect to the true noise level, when it is designed with respect to different postulated noise level $\sigma_f^2 = 0.2, 0.5, 1, 2, 3, 5$, and 7. In addition, Fig. 2 also plots the sampling threshold in (9) corresponds to the case when the postulated noise level in the preprocessing function matches with the true noise level,

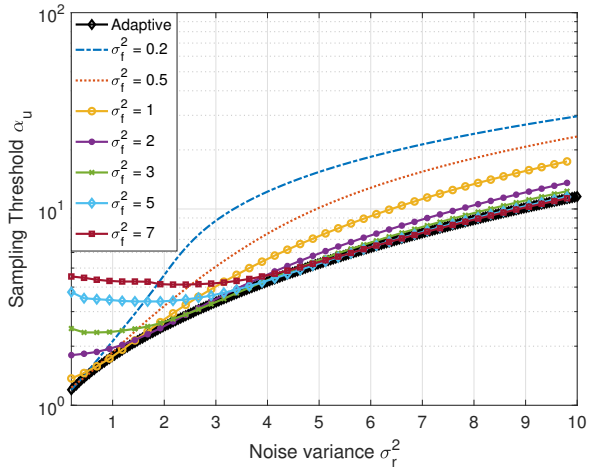

Fig. 2. The sampling threshold α_u with respect to the true AWGN noise level σ_r^2 for a preprocessing function designed to be optimal for a postulated noise level $\sigma_f^2 = 0.2, 0.5, 1, 2, 3, 5, 7$, respectively. The sampling threshold of the adaptive preprocessing function corresponds to (9) when $\sigma_f^2 = \sigma_r^2$.

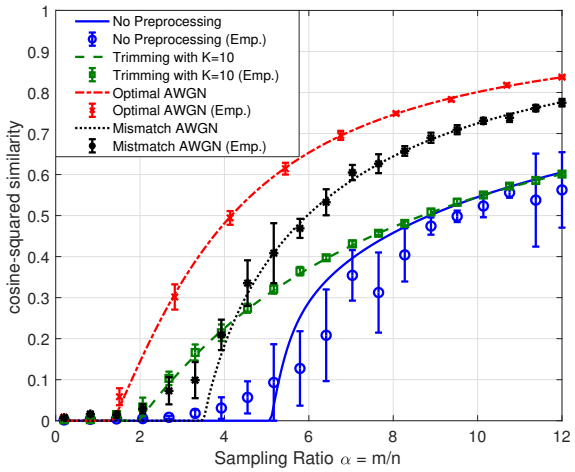

Fig. 3. The theoretical prediction and the empirical realizations of the cosine-squared similarity of each preprocessing function with respect to the sampling ratio under AWGN with $\sigma_r^2 = 0.5$. The Optimal AWGN curve postulates $\sigma_f^2 = \sigma_r^2 = 0.5$, whereas the Mismatch AWGN curve assume $\sigma_f^2 = 5$.

dubbed as the *adaptive* AWGN preprocessing function. Clearly, the adaptive AWGN preprocessing function where $\sigma_f^2 = \sigma_r^2$ serves as a lower bound of the minimal sampling threshold. It is interesting to observe that when the mismatch level is small, the deviation in the sampling threshold is also relatively small. However, the performance can be drastically worse when the mismatch level is high.

4.2. Sensitivity of Cosine-Squared Similarities

Next, we examine the sensitivity of cosine-squared similarities over a wide range of the sampling ratio in both AWGN and Poisson noise.

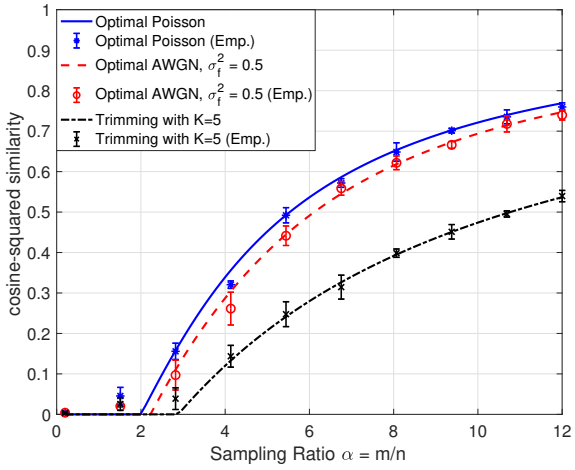

Fig. 4. The theoretical prediction and the empirical realizations of the cosine-squared similarity of each preprocessing function with respect to the sampling ratio under Poisson noise. The Optimal AWGN curve postulates $\sigma_f^2 = 0.5$ and applies (8).

In Fig. 3, we examine the empirical and theoretical performance of the optimal AWGN preprocessing function in the presence of AWGN with true noise level $\sigma_r^2 = 0.5$ as the sampling ratio increases. The AWGN preprocessing function where $\sigma_f^2 = \sigma_r^2$ outperforms the other in terms of achieving the theoretical sampling ratio threshold as well as the overall cosine-squared similarity performance. Compared with the mismatched AWGN preprocessing function where $\sigma_f^2 = 5$, the trimming preprocessing function obtains a smaller sampling ratio threshold, but the overall cosine-squared similarity does not increase at the same rate as the mismatched AWGN preprocessing function. Therefore, one may not determine the performance of the spectral method based only on the sampling threshold. The performance of the vanilla spectral initialization without preprocessing is much worse with more variability.

Lastly, we test the performance of preprocessing functions derived for different noise models in the presence of Poisson noise in Fig. 4. The optimal preprocessing function for Poisson noise obtains the optimal sampling ratio threshold and outperforms the other preprocessing functions also in terms of the cosine-squared similarity. The AWGN preprocessing function with $\sigma_f^2 = 0.5$ performs nearly as well as the sampling ratio increases, but does not obtain the optimal sampling ratio threshold.

5. CONCLUSIONS

In this paper we studied the sensitivity of optimal preprocessing functions in spectral methods when there is a mismatch between the theoretical model and the practical model. Using noisy phase retrieval as a case study, we derived the optimal preprocessing functions under both Gaussian noise and Poisson noise, and further compared their performances under model mismatch. Our study highlights the importance of considering model mismatch and designing robust preprocessing functions that provide desirable performance over a wide range of measurement distributions.

6. REFERENCES

[1] K.-C. Li, "On principal Hessian directions for data visualization and dimension reduction: Another application of stein's lemma," *Journal of the American Statistical Association*, vol. 87, no. 420, pp. 1025–1039, 1992.

[2] Y. Chi, Y. M. Lu, and Y. Chen, "Nonconvex optimization meets low-rank matrix factorization: An overview," *arXiv preprint arXiv:1809.09573*, 2018.

[3] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, pp. 1985–2007, April 2015.

[4] Y. Chen and E. J. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Comm. Pure Appl. Math.*, vol. 70, no. 5, pp. 822–883, 2017.

[5] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi, "A nonconvex approach for phase retrieval: Reshaped Wirtinger flow and incremental algorithms," *Journal of Machine Learning Research*, vol. 18, no. 141, pp. 1–35, 2017.

[6] H. Zhang, Y. Chi, and Y. Liang, "Provable non-convex phase retrieval with outliers: Median truncated Wirtinger flow," in *International conference on machine learning*, pp. 1022–1031, 2016.

[7] C. Ma, K. Wang, Y. Chi, and Y. Chen, "Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution," *arXiv preprint arXiv:1711.10467*, 2017.

[8] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," *Advances in Neural Information Processing Systems (NIPS)*, 2013.

[9] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Transactions on Information Theory*, 2017.

[10] T. Goldstein and C. Studer, "Phasemax: Convex phase retrieval via basis pursuit," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2675–2689, 2018.

[11] S. Bahmani and J. Romberg, "Phase retrieval meets statistical learning theory: A flexible convex relaxation," in *Artificial Intelligence and Statistics*, pp. 252–260, 2017.

[12] Y. M. Lu and G. Li, "Phase transitions of spectral initialization for high-dimensional nonconvex estimation," *arXiv preprint arXiv:1702.06435*, 2017.

[13] M. Mondelli and A. Montanari, "Fundamental limits of weak recovery with applications to phase retrieval," *arXiv preprint arXiv:1708.05932*, 2017.

[14] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Transactions on Information Theory*, vol. 56, pp. 2980 –2998, June 2010.

[15] S. Tu, R. Boczar, M. Simchowitz, M. Soltanolkotabi, and B. Recht, "Low-rank solutions of linear matrix equations via Procrustes flow," in *Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48*, pp. 964–973, JMLR. org, 2016.

[16] Y. Chi and Y. M. Lu, "Kaczmarz method for solving quadratic equations," *IEEE Signal Processing Letters*, vol. 23, no. 9, pp. 1183–1187, 2016.

[17] X. Li, S. Ling, T. Strohmer, and K. Wei, "Rapid, robust, and reliable blind deconvolution via nonconvex optimization," *CoRR*, vol. abs/1606.04933, 2016.

[18] Y. Chi and H. Fu, "Subspace learning from bits," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4429–4442.

[19] P. Chen, A. Fannjiang, and G.-R. Liu, "Phase retrieval with one or two diffraction patterns by alternating projections with the null initialization," *Journal of Fourier Analysis and Applications*, pp. 1–40, 2015.

[20] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retrieval," in *Information Theory (ISIT), 2016 IEEE International Symposium on*, pp. 2379–2383, IEEE, 2016.

[21] S. Bhojanapalli, B. Neyshabur, and N. Srebro, "Global optimality of local search for low rank matrix recovery," in *Advances in Neural Information Processing Systems*, pp. 3873–3881, 2016.

[22] R. Ge, J. D. Lee, and T. Ma, "Matrix completion has no spurious local minimum," in *Advances in Neural Information Processing Systems*, pp. 2973–2981, 2016.

[23] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht, "Gradient descent converges to minimizers," *arXiv preprint arXiv:1602.04915*, 2016.

[24] A. W. Luo, Wangyu and Y. M. Lu, "Optimal spectral initialization for signal recovery with applications to phase retrieval," *arXiv preprint arXiv:1811.04420v1*, 2018.

[25] V. Monardo and Y. Chi, "On the sensitivity of spectral initialization for noisy phase retrieval," 2018. `https://users.ece.cmu.edu/~yuejiec/papers/spectralinit.pdf`.

In the appendix, we provide the derivations of the optimal preprocessing functions in AWGN and Poisson noise using the formula in [13]. For the derivations, suppose that $\boldsymbol{x} \in \mathbb{C}^n$ is uniformly distributed over the n-dimensional complex sphere with radius $\sqrt{n}$, i.e., $\boldsymbol{x} \sim \text{Unif}(\sqrt{n}\mathbb{S}_{\mathbb{C}}^{n-1})$, and $\{\boldsymbol{a}_i\}_{i=1}^m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \frac{1}{2n}\boldsymbol{I}_n) + j\mathcal{N}(\boldsymbol{0}, \frac{1}{2n}\boldsymbol{I}_n)$. Let $s_i = \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle$, the optimal preprocessing function can be derived as

$$\mathcal{T}(y) = 1 - \frac{\mathbb{E}_s\{p(y \mid |s|)\}}{\mathbb{E}_s\{|s|^2 p(y \mid |s|)\}}. \tag{11}$$

Furthermore, the sampling threshold can be derived as

$$\alpha_u = \left(\int_{\mathbb{R}} \frac{\mathbb{E}_s\left\{p(y \mid |s|)(1-|s|^2)\right\}}{\mathbb{E}_s\left\{p(y \mid |s|)\right\}} \right)^{-1}. \tag{12}$$

A. DERIVATIONS FOR THE AWGN CASE

In the presence of AWGN, the measurement y_i follows the distribution $y_i \sim p(y \mid |s_i|)$, where

$$p(y \mid |s_i|) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y - |s_i|^2)^2}{2\sigma^2} \right). \tag{13}$$

Furthermore, we can rewrite $s_i = \text{Re}(s_i) + j\,\text{Im}(s_i)$, where $(\text{Re}(s_i), \text{Im}(s_i)) \sim \mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_2)$. Let $R = |s_i| = \sqrt{\text{Re}(s_i)^2 + \text{Im}(s_i)^2}$. Then R follows a Rayleigh distribution with scale parameter $\frac{1}{\sqrt{2}}$.

A.1. Optimal Preprocessing Function for AWGN

By the calculations in A.1.1, we have

$$\mathbb{E}_s\{p(y \mid |s|)\} = \exp\left(-\frac{(2y - \sigma^2)}{2} \right) \mathcal{Q}(-(y - \sigma^2)/\sigma). \tag{14}$$

By the calculations in A.1.2, we have

$$\mathbb{E}_s\{|s|^2 p(y \mid |s|)\} = (y - \sigma^2) \exp\left(-\frac{(2y - \sigma^2)}{2} \right) \mathcal{Q}(-(y - \sigma^2)/\sigma) + \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2} \right). \tag{15}$$

Therefore, plugging (14) and (15) into (11), we obtain the optimal preprocessing function as

$$\begin{aligned} \mathcal{T}(y) = 1 - \frac{\mathbb{E}_s\{p(y \mid |s|)\}}{\mathbb{E}_s\{|s|^2 p(y \mid |s|)\}} &= 1 - \frac{\exp\left(-\frac{(2y-\sigma^2)}{2} \right) \mathcal{Q}(-(y-\sigma^2)/\sigma)}{(y-\sigma^2)\exp\left(-\frac{(2y-\sigma^2)}{2} \right) \mathcal{Q}(-(y-\sigma^2)/\sigma) + \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2} \right)} \\ &= 1 - \left(y - \sigma^2 + \sqrt{\frac{\sigma^2}{2\pi}} \frac{\exp\left(-\frac{(y-\sigma^2)^2}{2\sigma^2} \right)}{\mathcal{Q}\left(-\frac{(y-\sigma^2)}{\sigma} \right)} \right)^{-1}. \end{aligned} \tag{16}$$

A.1.1. Calculations for $\mathbb{E}_s\{p(y \mid |s|)\}$ in AWGN

We have

$$\begin{aligned}
\mathbb{E}_s\{p(y \mid |s|)\} &= \mathbb{E}_R\{p(y \mid R)\} \\
&= \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-z)^2}{2\sigma^2}\right) \exp(-z) dz \\
&= \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2 - 2yz + z^2}{2\sigma^2}\right) \exp(-z) dz \\
&= \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2 - 2yz + z^2 + 2\sigma^2 z}{2\sigma^2}\right) dz \\
&= \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2 - (2y - 2\sigma^2)z + z^2}{2\sigma^2}\right) dz \\
&= \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z - (y-\sigma^2))^2 + \sigma^2(2y - \sigma^2)}{2\sigma^2}\right) dz \\
&= \exp\left(-\frac{(2y-\sigma^2)}{2}\right) \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z-(y-\sigma^2))^2}{2\sigma^2}\right) dz \\
&= \exp\left(-\frac{(2y-\sigma^2)}{2}\right) \int_{-(y-\sigma^2)}^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right) dz \\
&= \exp\left(-\frac{(2y-\sigma^2)}{2}\right) \mathcal{Q}(-(y-\sigma^2)/\sigma).
\end{aligned}$$

A.1.2. Calculations for $\mathbb{E}_s\{|s|^2 p(y \mid |s|)\}$ in AWGN

Similarly,

$$\begin{aligned}
\mathbb{E}_s\{|s|^2 p(y \mid |s|)\} &= \int_0^\infty \frac{z}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-z)^2}{2\sigma^2}\right) \exp(-z) dz \\
&= \exp\left(-\frac{(2y-\sigma^2)}{2}\right) \int_0^\infty \frac{z}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z-(y-\sigma^2))^2}{2\sigma^2}\right) dz \\
&= \exp\left(-\frac{(2y-\sigma^2)}{2}\right) \int_{-(y-\sigma^2)}^\infty \frac{z + (y-\sigma^2)}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right) dz \\
&= \exp\left(-\frac{(2y-\sigma^2)}{2}\right) \left[\int_{-(y-\sigma^2)}^\infty \frac{z}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right) dz + \int_{-(y-\sigma^2)}^\infty \frac{(y-\sigma^2)}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right) dz\right] \\
&= \exp\left(-\frac{(2y-\sigma^2)}{2}\right) \int_{-(y-\sigma^2)}^\infty \frac{z}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right) dz + (y-\sigma^2) \exp\left(-\frac{(2y-\sigma^2)}{2}\right) \mathcal{Q}(-(y-\sigma^2)/\sigma).
\end{aligned}$$

Note that the first term can be further simplified as

$$\begin{aligned}
\int_{-(y-\sigma^2)}^\infty \frac{z}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right) dz &= \frac{1}{\sqrt{\pi}} \int_{-(y-\sigma^2)}^\infty \frac{z}{\sigma\sqrt{2}} \exp\left(-\left(\frac{z}{\sigma\sqrt{2}}\right)^2\right) dz \\
&= \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\frac{(y-\sigma^2)}{\sigma\sqrt{2}}}^\infty u \exp\left(-u^2\right) du \\
&= \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \left[-\frac{1}{2}\exp(-u^2)\right]_{-\frac{(y-\sigma^2)}{\sigma\sqrt{2}}}^\infty \\
&= \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{(y-\sigma^2)^2}{2\sigma^2}\right),
\end{aligned}$$

leading to the expression in (15).

A.2. Sampling Ratio Threshold for AWGN

Plugging (14) and (15) into (12), we tackle the calculation of the sampling threshold as

$$\begin{aligned}
\alpha_u^{-1} &= \int_{\mathbb{R}} \frac{\left(\mathbb{E}_s\{p(y \mid |s|)(|s|^2-1)\}\right)^2}{\mathbb{E}_s\{p(y \mid |s|)\}} dy \\
&= \int_{\mathbb{R}} \frac{\left[\mathbb{E}_s\{|s|^2 p(y \mid |s|)\} - \mathbb{E}_s\{p(y \mid |s|)\}\right]^2}{\mathbb{E}_s\{p(y \mid |s|)\}} dy \\
&= \int_{\mathbb{R}} \frac{\left[(y-\sigma^2)\exp\left(-\frac{(2y-\sigma^2)}{2}\right)\mathcal{Q}(-(y-\sigma^2)/\sigma) + \frac{\sigma}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\sigma^2}\right) - \exp\left(-\frac{(2y-\sigma^2)}{2}\right)\mathcal{Q}(-(y-\sigma^2)/\sigma)\right]^2}{\exp\left(-\frac{(2y-\sigma^2)}{2}\right)\mathcal{Q}(-(y-\sigma^2)/\sigma)} dy \\
&= \int_{\mathbb{R}} \frac{\left[(y-\sigma^2-1)\exp\left(-\frac{(2y-\sigma^2)}{2}\right)\mathcal{Q}(-(y-\sigma^2)/\sigma) + \frac{\sigma}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\sigma^2}\right)\right]^2}{\exp\left(-\frac{(2y-\sigma^2)}{2}\right)\mathcal{Q}(-(y-\sigma^2)/\sigma)} dy \\
&= \underbrace{\int_{\mathbb{R}} (y-\sigma^2-1)^2 \exp\left(-\frac{(2y-\sigma^2)}{2}\right)\mathcal{Q}(-(y-\sigma^2)/\sigma) dy}_{\mathfrak{T}_1} \qquad (17) \\
&+ \underbrace{\int_{\mathbb{R}} \frac{2\sigma(y-\sigma^2-1)}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\sigma^2}\right) dy}_{\mathfrak{T}_2} \qquad (18) \\
&+ \underbrace{\frac{\sigma^2}{2\pi}\int_{\mathbb{R}} \frac{\exp(-\frac{y^2}{2\sigma^2})\exp(-\frac{(y-\sigma^2)^2}{2\sigma^2})}{\mathcal{Q}(-(y-\sigma^2)/\sigma)} dy}_{\mathfrak{T}_3}. \qquad (19)
\end{aligned}$$

We separately handle the calculation of the three terms, $\mathfrak{T}_1$, $\mathfrak{T}_2$, and $\mathfrak{T}_3$ in equations (17), (18), and (19), respectively.

A.2.1. Calculation of $\mathfrak{T}_1$

$$\begin{aligned}
\mathfrak{T}_1 &= \int_{\mathbb{R}} (y-\sigma^2-1)^2 \exp\left(-\frac{(2y-\sigma^2)}{2}\right)\mathcal{Q}(-(y-\sigma^2)/\sigma) dy \\
&= \int_{\mathbb{R}} (y-\sigma^2-1)^2 \exp\left(-\frac{(2y-\sigma^2)}{2}\right)\left[\int_{-(y-\sigma^2)}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{z^2}{2\sigma^2}\right) dz\right] dy \\
&= \int_{\mathbb{R}} \left[\int_{-(y-\sigma^2)}^{\infty} (y-\sigma^2-1)^2 \exp\left(-\frac{(2y-\sigma^2)}{2}\right)\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{z^2}{2\sigma^2}\right) dz\right] dy \\
&= \int_{-\infty}^{\infty} \left[\int_{-(z-\sigma^2)}^{\infty} (y-\sigma^2-1)^2 \exp\left(-\frac{(2y-\sigma^2)}{2}\right)\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{z^2}{2\sigma^2}\right) dy\right] dz \\
&= \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{\sigma^2}{2}\right)\int_{-\infty}^{\infty}\exp\left(-\frac{z^2}{2\sigma^2}\right)\underbrace{\left[\int_{-(z-\sigma^2)}^{\infty}(y-\sigma^2-1)^2\exp(-y)dy\right]}_{\mathfrak{S}} dz.
\end{aligned}$$

We handle the integral within the brackets, $\mathfrak{S}$, first.

$$\begin{aligned}
\mathfrak{S} &= \left[\int_{-(z-\sigma^2)}^{\infty}(y-\sigma^2-1)^2\exp(-y)dy\right] \\
&= \int_{-(z-\sigma^2)}^{\infty}(y^2 - y(2\sigma^2+2) + (2\sigma^2+\sigma^4+1))\exp(-y)dy \\
&= \underbrace{\int_{-(z-\sigma^2)}^{\infty}(y^2)\exp(-y)dy}_{\mathfrak{S}_1} \underbrace{- \int_{-(z-\sigma^2)}^{\infty} y(2\sigma^2+2)\exp(-y)dy}_{\mathfrak{S}_2} + \underbrace{\int_{-(z-\sigma^2)}^{\infty}(2\sigma^2+\sigma^4+1)\exp(-y)dy}_{\mathfrak{S}_3}.
\end{aligned}$$

The first of these integrals, $\mathfrak{S}_1$, is

$$\begin{aligned}\mathfrak{S}_1 &= \int_{-(z-\sigma^2)}^{\infty} (y^2)\exp(-y)dy = -(y^2+2y+2)\exp(-y)\big|_{-(z-\sigma^2)}^{\infty} \\ &= [(z-\sigma^2)^2 - 2(z-\sigma^2)+2]\exp(z-\sigma^2) \\ &= [(z-\sigma^2-2)(z-\sigma^2)+2]\exp(z-\sigma^2) \\ &= [z^2 - z\sigma^2 - 2z - z\sigma^2 + \sigma^4 + 2\sigma^2 + 2]\exp(z-\sigma^2) \\ &= [z^2 - (2\sigma^2+2)z + \sigma^4 + 2\sigma^2 + 2]\exp(z-\sigma^2).\end{aligned}$$

The second of these integrals, $\mathfrak{S}_2$, is

$$\begin{aligned}\mathfrak{S}_2 &= -\int_{-(z-\sigma^2)}^{\infty} y(2\sigma^2+2)\exp(-y)dy \\ &= (2\sigma^2+2)\ (y+1)\exp(-y)|_{-(z-\sigma^2)}^{\infty} \\ &= -(2\sigma^2+2)(-(z-\sigma^2)+1)\exp(z-\sigma^2) \\ &= (2\sigma^2+2)(z-\sigma^2-1)\exp(z-\sigma^2) \\ &= (2\sigma^2 z + 2z - 2\sigma^4 - 2\sigma^2 - 2\sigma^2 - 2)\exp(z-\sigma^2) \\ &= [z(2\sigma^2+2) - 2\sigma^4 - 4\sigma^2 - 2]\exp(z-\sigma^2).\end{aligned}$$

The third of these integrals, $\mathfrak{S}_3$, is

$$\begin{aligned}\mathfrak{S}_3 &= \int_{-(z-\sigma^2)}^{\infty} (2\sigma^2+\sigma^4+1)\exp(-y)dy \\ &= -(2\sigma^2+\sigma^4+1)\ \exp(-y)|_{-(z-\sigma^2)}^{\infty} \\ &= (2\sigma^2+\sigma^4+1)\exp(z-\sigma^2)\end{aligned}$$

Therefore,

$$\begin{aligned}\mathfrak{S} &= \int_{-(z-\sigma^2)}^{\infty} (y^2)\exp(-y)dy - \int_{-(z-\sigma^2)}^{\infty} y(2\sigma^2+2)\exp(-y)dy + \int_{-(z-\sigma^2)}^{\infty} (2\sigma^2+\sigma^4+1)\exp(-y)dy \\ &= [z^2 - (2\sigma^2+2)z + \sigma^4 + 2\sigma^2 + 2]\exp(z-\sigma^2) + [z(2\sigma^2+2) - 2\sigma^4 - 4\sigma^2 - 2]\exp(z-\sigma^2) \\ &+ (2\sigma^2+\sigma^4+1)\exp(z-\sigma^2) \\ &= (z^2+1)\exp(z-\sigma^2)\end{aligned}$$

Using this result above, we substitute $\mathfrak{S}$ back into the equation for $\mathfrak{T}_1$, and obtain

$$\begin{aligned}\mathfrak{T}_1 &= \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{\sigma^2}{2}\right)\int_{-\infty}^{\infty}\exp\left(-\frac{z^2}{2\sigma^2}\right)(z^2+1)\exp(z-\sigma^2)dz \\ &= \exp\left(\frac{\sigma^2}{2}\right)\int_{-\infty}^{\infty}\frac{z^2}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(z^2-2\sigma^2 z+2\sigma^4)}{2\sigma^2}\right)dz \\ &+ \exp\left(\frac{\sigma^2}{2}\right)\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(z^2-2\sigma^2 z+2\sigma^4)}{2\sigma^2}\right)dz \\ &= \int_{-\infty}^{\infty}\frac{z^2}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(z-\sigma^2)^2}{2\sigma^2}\right)dz + \int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(z-\sigma^2)^2}{2\sigma^2}\right)dz \\ &= \sigma^4+\sigma^2+1.\end{aligned}$$

A.2.2. Calculation of $\mathfrak{T}_2$

$$\begin{aligned}\mathfrak{T}_2 &= \int_{\mathbb{R}}\frac{2\sigma(y-\sigma^2-1)}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\sigma^2}\right)dy \\ &= 2\sigma^2\int_{\mathbb{R}}\frac{y}{\sigma\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\sigma^2}\right)dy - \int_{\mathbb{R}}\frac{2\sigma^4+2\sigma^2}{\sigma\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\sigma^2}\right)dy \\ &= -2\sigma^4 - 2\sigma^2.\end{aligned}$$

A.2.3. Calculation of $\mathfrak{T}_3$

$$\begin{aligned}\mathfrak{T}_3 &= \frac{\sigma^2}{2\pi}\int_{\mathbb{R}} \frac{\exp(-\frac{y^2}{2\sigma^2})\exp(-\frac{(y-\sigma^2)^2}{2\sigma^2})}{\mathcal{Q}(-(y-\sigma^2)/\sigma)}dy \\ &= \frac{\sigma^3}{2\pi}\int_{-\infty}^{\infty} \frac{\exp\left(-\frac{(u+\sigma)^2}{2}\right)\exp\left(-\frac{u^2}{2}\right)}{\mathcal{Q}(-u)}du \\ &= \frac{\sigma^3\exp\left(-\sigma^2/2\right)}{2\pi}\int_{-\infty}^{\infty}\frac{\exp\left(-\sigma u\right)\exp\left(-u^2\right)}{\mathcal{Q}(-u)}du.\end{aligned}$$

Putting it all together, we finally obtain

$$\begin{aligned}\alpha_u &= \left(\mathfrak{T}_1+\mathfrak{T}_2+\mathfrak{T}_3\right)^{-1} \\ &= \left(1-\sigma^2-\sigma^4+\frac{\sigma^3\exp\left(-\sigma^2/2\right)}{2\pi}\int_{-\infty}^{\infty}\frac{\exp\left(-\sigma u\right)\exp\left(-u^2\right)}{\mathcal{Q}(-u)}du\right)^{-1}.\end{aligned}$$

B. DERIVATIONS FOR THE POISSON CASE

Let $s_i = \langle \boldsymbol{x}, \boldsymbol{a}_i\rangle$. In the presence of Poisson noise with rate $|s_i|^2$, the measurement y_i follows the distribution $y_i \sim p(y \mid |s_i|)$, where

$$p(y \mid |s_i|) = \frac{|s_i|^{2y}e^{-|s_i|^2}}{y!}. \tag{20}$$

B.1. Optimal Preprocessing Function for Poisson noise

First, since $|s|$ follows a Rayleigh distribution with scale parameter $\frac{1}{\sqrt{2}}$, we have

$$\begin{aligned}\mathbb{E}_s\{p(y \mid |s|)\} &= \frac{1}{y!}\int_0^{+\infty} z^y\exp(-2z)dz \\ &= \frac{1}{y!}\int_0^{+\infty}\frac{1}{2}(\frac{1}{2}x)^y\exp(-x)dx \\ &= \frac{1}{y!}\frac{1}{2^{y+1}}\int_0^{\infty}x^y\exp(-x)dx \\ &= \frac{1}{y!}\frac{1}{2^{y+1}}\Gamma(y+1).\end{aligned} \tag{21}$$

Furthermore,

$$\begin{aligned}\mathbb{E}_s\{|s|^2p(y \mid |s|)\} &= \int_0^{+\infty}\frac{z^{y+1}}{y!}\exp(-2z)dz \\ &= \frac{1}{y!}\int_0^{+\infty} z^{y+1}\exp(-2z)dz \\ &= \frac{1}{y!}\int_0^{+\infty}\frac{1}{2}(\frac{1}{2}x)^{y+1}\exp(-x)dx \\ &= \frac{1}{y!}\frac{1}{2^{y+2}}\int_0^{\infty}x^{y+1}\exp(-x)dx \\ &= \frac{1}{y!}\frac{1}{2^{y+2}}\Gamma(y+2) \\ &= \frac{y+1}{y!2^{y+2}}\Gamma(y+1)\end{aligned} \tag{22}$$

Then, plugging (21) and (22) into (11), we can obtain

$$\mathcal{T}(y) = 1-\frac{\frac{1}{y!}\frac{1}{2^{y+1}}\Gamma(y+1)}{\frac{y+1}{y!2^{y+2}}\Gamma(y+1)} = 1-\frac{2}{y+1} = \frac{y-1}{y+1}. \tag{23}$$

B.2. Optimal Sampling Threshold for Poisson noise

Plugging (21) and (22) into (12), we have that

$$\alpha_u^{-1} = \sum_{y\in\mathbb{R}} \frac{\left[\frac{y+1}{y!2^{y+2}}\Gamma(y+1) - \frac{1}{y!}\frac{1}{2^{y+1}}\Gamma(y+1)\right]^2}{\frac{1}{y!}\frac{1}{2^{y+1}}\Gamma(y+1)}, \tag{24}$$

which can be simplified as

$$\begin{aligned}
\sum_{y\in\mathbb{R}} \frac{\left[\frac{y+1}{y!2^{y+2}}\Gamma(y+1) - \frac{1}{y!}\frac{1}{2^{y+1}}\Gamma(y+1)\right]^2}{\frac{1}{y!}\frac{1}{2^{y+1}}\Gamma(y+1)} &= \sum_{y\in\mathbb{R}} \frac{\left[\frac{y+1}{y!2^{y+2}}\Gamma(y+1) - \frac{1}{y!}\frac{2}{2^{y+2}}\Gamma(y+1)\right]^2}{\frac{1}{y!}\frac{1}{2^{y+1}}\Gamma(y+1)} \\
&= \sum_{y\in\mathbb{R}} \frac{\left[\frac{y-1}{y!2^{y+2}}\Gamma(y+1)\right]^2}{\frac{1}{y!}\frac{1}{2^{y+1}}\Gamma(y+1)} \\
&= \sum_{y\in\mathbb{R}} \frac{\left[\frac{y-1}{y!2^{y+2}}\right]^2\Gamma(y+1)}{\frac{1}{y!}\frac{1}{2^{y+1}}} \\
&= \sum_{y\in\mathbb{R}} \frac{(y-1)^2}{y!2^{2y+3}}\Gamma(y+1) \\
&= \frac{1}{8}\sum_{y\in\mathbb{R}} \frac{y^2-2y+1}{y!2^y}\left[\int_0^{+\infty} x^y\exp(-x)dx\right] \\
&= \frac{1}{8}\int_0^{+\infty}\sum_{y\in\mathbb{R}}\left[\left(\frac{x}{2}\right)^y\frac{\exp\left(-\frac{x}{2}\right)}{y!}(y^2-2y+1)\right]\exp\left(-\frac{x}{2}\right)dx \\
&= \frac{1}{8}\int_0^{+\infty}\mathbb{E}_{y\sim\text{Poisson}(\frac{x}{2})}\left[y^2-2y+1\right]\exp\left(-\frac{x}{2}\right)dx \\
&= \frac{1}{8}\int_0^{+\infty}\left[\left(\frac{x}{2}\right)+\left(\frac{x}{2}\right)^2-2\left(\frac{x}{2}\right)+1\right]\exp\left(-\frac{x}{2}\right)dx \\
&= \frac{1}{8}\int_0^{+\infty}\left(\frac{x}{2}\right)^2\exp\left(-\frac{x}{2}\right)dx - \frac{1}{8}\int_0^{+\infty}\left(\frac{x}{2}\right)\exp\left(-\frac{x}{2}\right)dx + \frac{1}{8}\int_0^{+\infty}\exp\left(-\frac{x}{2}\right)dx \\
&= \frac{1}{4}\int_0^{+\infty}u^2\exp(-u)du - \frac{1}{4}\int_0^{+\infty}u\exp(-u)du + \frac{1}{4}\int_0^{+\infty}\exp(-u)du \\
&= \frac{1}{4}[2-1+1] = \frac{1}{2}
\end{aligned} \tag{25}$$

Therefore, $\alpha_u = 2$.