Optimal Spectral Initialization for Signal Recovery with Applications to Phase Retrieval

Wangyu Luo, Wael Alghamdi and Yue M. Lu

Abstract—We present the optimal design of a spectral method widely used to initialize nonconvex optimization algorithms for solving phase retrieval and other signal recovery problems. Our work leverages recent results that provide an exact characterization of the performance of the spectral method in the high-dimensional limit. This characterization allows us to map the task of optimal design to a constrained optimization problem in a weighted L^2 function space. The latter has a closed-form solution. Interestingly, under a mild technical condition, our results show that there exists a fixed design that is uniformly optimal over all sampling ratios. Numerical simulations demonstrate the performance improvement brought by the proposed optimal design over existing constructions in the literature. In a recent work, Mondelli and Montanari have shown the existence of a weak recovery threshold below which the spectral method cannot provide useful estimates. Our results serve to complement that work by deriving the fundamental limit of the spectral method beyond the aforementioned threshold.

Index Terms—Spectral initialization, phase retrieval, signal estimation, nonconvex optimization, phase transition, optimal spectral methods

I. INTRODUCTION

An active line of recent work studies nonconvex optimization algorithms for solving the classical phase retrieval problem (see, *e.g.*, [1]–[8]). Compared to methods using convex relaxation [9]–[12], the nonconvex approaches tend to require much lower computational complexity and memory footprints. A key ingredient in many such algorithms is a simple yet highly effective spectral method [1], [3], [13]. It provides an initial estimate that is sufficiently close to the target signal. Starting from this "warm start", local search schemes such as gradient descent can then carry out further refinement to reach globally optimal solutions.

This paper studies the optimal design of the aforementioned spectral method. Throughout the paper, we consider the following sensing model. Let $\boldsymbol{\xi} \in \mathbb{C}^n$ denote the target signal we seek to estimate, and $\{\boldsymbol{a}_i \in \mathbb{C}^n\}_{1 \leq i \leq m}$ be a collection of sensing vectors. Given $s_i = \langle \boldsymbol{a}_i, \boldsymbol{\xi} \rangle$, the ith measurement $y_i \in \mathbb{R}$ is drawn independently from

$$y_i \sim p\big(y \,\big|\, |s_i|\big), \qquad 1 \leq i \leq m, \qquad (1)$$

where $p(\cdot \,|\, \cdot)$ is a conditional density function modeling the (potentially noisy) sensing process. Clearly, the phase information of s_i is missing, as y_i only depends on the magnitude

of s_i. The spectral method we study consists of two simple steps [1], [3], [13]. First, construct a data matrix as

$$\boldsymbol{D} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{i=1}^{m} \mathcal{T}(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^*, \qquad (2)$$

where $\mathcal{T} : \mathbb{R} \mapsto \mathbb{R}$ is a user-specified *preprocessing function*, and $\boldsymbol{a}_i^*$ denotes the conjugate transpose of $\boldsymbol{a}_i$. Second, we compute $\boldsymbol{x}_1$, an eigenvector associated with the largest eigenvalue of $\boldsymbol{D}$. The vector $\boldsymbol{x}_1$ is then our initial estimate of $\boldsymbol{\xi}$ (up to an unknown scalar).

The idea of this spectral method first appeared in the statistics literature under the name of principal Hessian directions [13]. In the context of phase retrieval, it was introduced by Netrapalli, Jain, and Sanghavi as an initialization step for their alternating minimization algorithm [1]. Finite sample performance analysis of the spectral method can be found in [1]–[3]. Under Gaussian design, the squared correlation between the eigenvector $\boldsymbol{x}_1$ and the target vector $\boldsymbol{\xi}$ is shown to approach 1 with high probability, when the number of samples m is sufficiently large with respect to the signal dimension n. In particular, by introducing a trimming step on the measurements (see (4) below), Chen and Candes [3] show that it suffices to have $m \geq c \cdot n$, where c is some sufficiently large constant.

In [14], Lu and Li presented an asymptotically exact characterization of the performance of the spectral methods. Specifically, under Gaussian design and when $m, n \to \infty$ at a fixed ratio $\alpha = m/n$, they show that the squared correlation between $\boldsymbol{\xi}$ and $\boldsymbol{x}_1$ converges in probability to a deterministic value, *i.e.*,

$$\frac{|\langle \boldsymbol{\xi}, \boldsymbol{x}_1 \rangle|^2}{\|\boldsymbol{\xi}\|^2 \|\boldsymbol{x}_1\|^2} \xrightarrow[n \to \infty]{\mathcal{P}} \rho(\alpha; \mathcal{T}(\cdot)). \qquad (3)$$

Moreover, explicit formulas are available to compute the limit value $\rho(\alpha; \mathcal{T}(\cdot))$. [See Section III-A for details.] The above asymptotic characterization was first derived for the real-valued case and under the assumption that $\mathcal{T}(y) \geq 0$ [14]. Then Mondelli and Montanari generalize the characterization to the complex-valued case in [15], where the assumption that $\mathcal{T}(y)$ be nonnegative is also shown to be unnecessary.

The performance of the spectral method depends heavily on the form of the preprocessing function $\mathcal{T}(\cdot)$ used in (2). (Accordingly, on the right-hand side of (3), our notation for the limit squared correlation $\rho(\alpha; \mathcal{T}(\cdot))$ makes its dependence on $\mathcal{T}(\cdot)$ explicit.) Several designs have been proposed in the literature, including the trimming scheme introduced in [3]:

$$\mathcal{T}_{\text{trim}}(y) = y \, \mathbb{1}_{\{|y| \leq a\}}, \qquad (4)$$

and the subset scheme proposed in [6]:

$$\mathcal{T}_{\text{subset}}(y) = \mathbb{1}_{\{|y| \geq b\}}. \tag{5}$$

In (4) and (5), $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function on a set, and a, b are some tuning parameters. See also [5] for yet another design that improves the robustness of the method in the presence of outliers.

While the existing designs in the literature are all based on sound intuitions (and ingenuity), they are not expected to be optimal. Equipped with the exact asymptotic characterizations obtained in [14], [15], we now have the luxury to ask the following question: given any specific sensing model in (1), what is the corresponding *optimal form* of the preprocessing function? Specifically, we consider the following optimal design problem:

$$\rho_{\text{optimal}}(\alpha) \overset{\text{def}}{=} \sup_{\mathcal{T}(\cdot) \in \mathcal{F}} \rho(\alpha; \mathcal{T}(\cdot)), \tag{6}$$

where $\rho(\alpha; \mathcal{T}(\cdot))$ is the limit squared correlation in (3), and $\mathcal{F}$ denotes a set of admissible functions from which we search for the optimal one. The exact definition of $\mathcal{F}$ will be given in (17) in Section II. It serves to restrict the search space to make sure that the asymptotic predictions obtained in [14], [15] are applicable. In what follows, we refer to $\rho_{\text{optimal}}(\alpha)$ as the *optimal performance curve*.

The first result addressing the optimal design problem was obtained by Mondelli and Montanari [15], who show that

$$\rho_{\text{optimal}}(\alpha) = 0, \qquad \text{for } \alpha \leq \alpha_{\text{weak}}, \tag{7}$$

where α_{weak} is a fundamental threshold (referred to as the *weak recovery threshold* in this paper). Given the definition of $\rho_{\text{optimal}}(\alpha)$, the result in (7) implies that, when the sampling ratio $\alpha \leq \alpha_{\text{weak}}$, the spectral method cannot provide any estimate that has nontrivial correlation with the target vector $\boldsymbol{\xi}$, no matter how one chooses the preprocessing function. Moreover, Mondelli and Montanari show that

$$\rho_{\text{optimal}}(\alpha) > 0, \qquad \text{for } \alpha > \alpha_{\text{weak}}. \tag{8}$$

They establish this by constructing a specific preprocessing function, denoted by $\mathcal{T}_{\text{MM}}(\cdot)$, such that

$$\rho(\alpha; \mathcal{T}_{\text{MM}}(\cdot)) > 0 \qquad \text{for } \alpha > \alpha_{\text{weak}}.$$

For any sensing model (1), explicit formulas are provided in [15] to compute α_{weak} and $\mathcal{T}_{\text{MM}}(\cdot)$. We defer such technical details to Section II [see (13) and (14)].

We note that, while the preprocessing functions $\mathcal{T}_{\text{MM}}(\cdot)$ serve to show that the optimal performance curve $\rho_{\text{optimal}}(\alpha)$ is strictly positive when $\alpha > \alpha_{\text{weak}}$, these functions do not solve the optimization problem in (6). Thus, important questions remain as to what $\rho_{\text{optimal}}(\alpha)$ should be beyond α_{weak} and whether there are optimal preprocessing functions that can potentially achieve this bound.

In this paper, we present a complete solution of the optimal design problem formulated in (6). Specifically, we provide an exact analytical expression for $\rho_{\text{optimal}}(\alpha)$ for all $\alpha > 0$ and for any sensing model. Moreover, under a mild technical condition (which is satisfied by many sensing models), we construct an optimal preprocessing function $\mathcal{T}_{\text{optimal}}(\cdot)$ that solves

the design problem. Somewhat surprising about the optimal solution is the fact that $\mathcal{T}_{\text{optimal}}(\cdot)$ does not depend on the sampling ratio α. In other words, the proposed $\mathcal{T}_{\text{optimal}}(\cdot)$ is *uniformly optimal* for all α. Finally, when the aforementioned technical condition does not hold, we show that the supremum in (6) cannot be achieved by any function in the admissible set $\mathcal{F}$. In this case, we construct a family of preprocessing functions $\mathcal{T}_\alpha^\varepsilon(\cdot)$ whose performance will approach $\rho_{\text{optimal}}(\alpha)$ as a parameter $\varepsilon \to 0$.

The rest of the paper is organized as follows. Our main results are stated as Theorem 1 in Section II. To illustrate these results, we present worked examples corresponding to two different sensing models. Numerical simulations demonstrate the performance improvements brought by the proposed optimal design over heuristic choices given in (4) and (5) as well as the function $\mathcal{T}_{\text{MM}}(\cdot)$ constructed in [15]. To set the stage for proving our results, Section III recalls the asymptotic characterization of the spectral method obtained in previous work [14], [15]. This characterization allows us to map the optimal design problem in (6) to a (constrained) optimization problem in a weighted L^2 function space. The proof of Theorem 1 is given in Section IV. Although we state and prove our results for the more general complex-valued case in this paper, the treatment of the real-valued case is the same, *mutatis mutandis*. See Remark 3 in Section II for an explanation of these changes.

II. MAIN RESULTS

We start by introducing two functions that will play central roles in our later technical discussions. Let

$$S \sim \mathcal{CN}(0, 1) \tag{9}$$

be a standard complex normal random variable. Define

$$\eta(y) \overset{\text{def}}{=} \mathbb{E}_S\big[p\big(y \,\big|\, |S|\big)\big] \tag{10}$$

and

$$\mu(y) \overset{\text{def}}{=} \mathbb{E}_S\big[|S|^2 \, p\big(y \,\big|\, |S|\big)\big], \tag{11}$$

where $p(\cdot|\cdot)$ is the conditional density function associated with the sensing model in (1). It is easy to verify that

$$\int \eta(y) \, \mathrm{d}y = \int \mu(y) \, \mathrm{d}y = 1. \tag{12}$$

The results of this paper as well as those of [15] can be conveniently stated in terms of these two functions. For example, the weak recovery threshold α_{weak} introduced in [15] can be written as

$$\alpha_{\text{weak}} \overset{\text{def}}{=} \Big[\int \frac{[\mu(y) - \eta(y)]^2}{\eta(y)} \, \mathrm{d}y \Big]^{-1}. \tag{13}$$

Moreover, the preprocessing function constructed in [15] is

$$\mathcal{T}_{\text{MM}}(y) \overset{\text{def}}{=} \frac{\sqrt{\alpha_{\text{weak}}}\, \mathcal{T}^*(y)}{\sqrt{\alpha} - (\sqrt{\alpha} - \sqrt{\alpha_{\text{weak}}})\mathcal{T}^*(y)}, \tag{14}$$

for $\alpha > \alpha_{\text{weak}}$, where

$$\mathcal{T}^*(y) \overset{\text{def}}{=} 1 - \eta(y)/\mu(y). \tag{15}$$

In Appendix A, we show that the integral on the right-hand side of (13) is always finite. Moreover, $\alpha_{\text{weak}} \geq 1$ under

any sensing model with complex-valued Gaussian design, and the lower bound is achieved by the case of noiseless phase retrieval, *i.e.*, $y_i = |\langle \boldsymbol{a}_i, \boldsymbol{\xi} \rangle|^2$. (For the real-valued case, the lower bound reduces to $\alpha_{\text{weak}} \geq 1/2$.)

A. Optimal Design

Our optimal design of the preprocessing function leverages the asymptotic characterizations given in [14], [15], which are derived under some technical assumptions on $\mathcal{T}(\cdot)$. Specifically, let Y be a random variable whose conditional distribution given S is

$$Y|S \sim p(y \mid |S|). \tag{16}$$

Let Γ_Y denote the support of the probability measure of Y. We shall assume that the preprocessing function $\mathcal{T}(y)$ belongs to the following *admissible set*:

$$\mathcal{F} \overset{\text{def}}{=} \left\{ \mathcal{T}(y) : 0 < \sup_{y \in \Gamma_Y} \mathcal{T}(y) < \infty \text{ and } \inf_{y \in \Gamma_Y} \mathcal{T}(y) > -\infty \right\}. \tag{17}$$

In words, we require that $\mathcal{T}(y)$ should have a bounded range and that the upper boundary of that range should be positive.

Theorem 1: Suppose that the target signal $\boldsymbol{\xi}$ is an arbitrary vector in $\mathbb{C}^n$ with $\|\boldsymbol{\xi}\| = 1$, and that the sensing vectors $\{\boldsymbol{a}_i\}_{1 \leq i \leq m}$ are drawn independently from the circularly-symmetric complex normal distribution, *i.e.*, $\boldsymbol{a}_i \sim_{\text{i.i.d.}} \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_n)$. As $m, n \to \infty$ with $m/n \to \alpha \in (0, \infty)$, the following hold with respect to the optimal design problem in (6):

1) For each $\alpha > \alpha_{\text{weak}}$, let β_α denote the unique positive solution of the equation $f(\beta) = 1/\alpha$, where

$$f(\beta) \overset{\text{def}}{=} \int_{\Gamma_Y} \frac{[\mu(y) - \eta(y)]^2}{\eta(y) + \mu(y)/\beta} \, \mathrm{d}y. \tag{18}$$

Then

$$\rho_{\text{optimal}}(\alpha) = \begin{cases} (1 + \beta_\alpha)^{-1}, & \text{for } \alpha > \alpha_{\text{weak}}, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

2) Assume $\alpha_{\text{weak}} < \infty$. If $\inf_{y \in \Gamma_Y} \frac{\mu(y)}{\eta(y)} > 0$, the optimal performance curve $\rho_{\text{optimal}}(\alpha)$ can be achieved by

$$\mathcal{T}_{\text{optimal}}(y) = 1 - \eta(y)/\mu(y) \in \mathcal{F}. \tag{20}$$

3) Assume $\alpha_{\text{weak}} < \infty$. If $\inf_{y \in \Gamma_Y} \frac{\mu(y)}{\eta(y)} = 0$, then $\rho_{\text{optimal}}(\alpha)$ cannot be achieved by any function $\mathcal{T}(\cdot) \in \mathcal{F}$. However, there exists a family of functions $\{\mathcal{T}_\alpha^\varepsilon(\cdot) \in \mathcal{F}\}_{0 < \varepsilon < 1}$ such that

$$\lim_{\varepsilon \to 0} \rho(\alpha; \mathcal{T}_\alpha^\varepsilon(\cdot)) = \rho_{\text{optimal}}(\alpha).$$

4) As an explicit construction of such a family, we can set $\mathcal{T}_\alpha^\varepsilon(y) = \frac{c_\alpha^\varepsilon(y)}{1 + c_\alpha^\varepsilon(y)}$, where

$$c_\alpha^\varepsilon(y) = \max \left\{ v_\alpha^\varepsilon \frac{\mu(y) - \eta(y)}{\eta(y) + \mu(y)/\beta_\alpha}, -1 + \varepsilon \right\}. \tag{21}$$

Here, β_α is the same constant as in (19), and $v_\alpha^\varepsilon \geq 1$ is a scalar that can be uniquely determined by the linear constraint $\int_{\Gamma_Y} c_\alpha^\varepsilon(y)[\mu(y) - \eta(y)] \, \mathrm{d}y = 1/\alpha$.

Remark 1: Theorem 1, whose proof is given in Section IV, provides a complete solution to the optimal design problem formulated in (6). As mentioned earlier, when $\inf_{y \in \Gamma_Y} \frac{\mu(y)}{\eta(y)} > 0$, the preprocessing function $\mathcal{T}_{\text{optimal}}(\cdot)$ given in (20) is uniformly optimal, as it does not depend on the sampling ratio α. We also note that there is a strong connection between $\mathcal{T}_{\text{optimal}}(\cdot)$ and the function $\mathcal{T}_{\text{MM}}(\cdot)$ designed in [15]. In fact, $\mathcal{T}_{\text{optimal}}(\cdot)$ is exactly equal to $\mathcal{T}^*(\cdot)$ in (15). Thus, from a practical perspective, our results have a simple message: instead of following the recipe of (14), which modifies the preprocessing function for different values of α, one should just use a fixed optimal function.

Remark 2: From the definitions given in (10) and (11), one can readily verify that $\mathcal{T}_{\text{optimal}}(y) = 1 - \eta(y)/\mu(y) = 1 - 1/\mathbb{E}_{S|Y}[|S|^2 \mid y]$. This connection between $\mathcal{T}_{\text{optimal}}(\cdot)$ and the scalar conditional mean estimator $\mathbb{E}_{S|Y}[|S|^2 \mid y]$ is very intriguing and might be worth further investigation.

Remark 3 (The real-valued case): The results of Theorem 1 can be directly applied to the real-valued case, after we make the following changes: (1) In the definitions of $\eta(y)$ and $\mu(y)$ in (10) and (11), the random variable S is now drawn from $\mathcal{N}(0, 1)$ instead of $\mathcal{CN}(0, 1)$; (2) In the statement of Theorem 1, we shall assume $\boldsymbol{\xi} \in \mathbb{R}^n$ and that the sensing vectors $\boldsymbol{a}_i \sim_{\text{i.i.d.}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$.

B. Worked Examples

To show how the results stated in Theorem 1 can be applied in practice, we present two worked examples corresponding to two different sensing models.

Example 1 (Poisson measurements): Here we consider the Poisson model, where

$$y_i \sim \text{Poisson}\left(\kappa \cdot |\langle \boldsymbol{a}_i, \boldsymbol{\xi} \rangle|^2\right),$$

and $\kappa > 0$ is an additional parameter indicating the signal-to-noise ratio in the sensing process. Note that the measurements $\{y_i\}$ here are nonnegative integers instead of continuous variables. Theorem 1 still applies. We just need to understand the integration in (18) as summations.

Let $Z = |S|^2$, with S defined as in (9). It is well-known that Z follows the exponential distribution with parameter 1. Using this property, we can compute the function in (10) as

$$\eta(y) = \mathbb{E}_Z \left[\frac{e^{-\kappa Z}(\kappa Z)^y}{y!} \right]$$
$$= \frac{\kappa^y}{y!} \int_0^\infty e^{-\kappa z} z^y e^{-z} \, \mathrm{d}z = \frac{\kappa^y}{(\kappa + 1)^{y+1}}. \tag{22}$$

Similarly, the function in (11) becomes

$$\mu(y) = \frac{\kappa^y(y + 1)}{(\kappa + 1)^{y+2}}. \tag{23}$$

Since y only takes nonnegative integer values,

$$\inf_y \frac{\mu(y)}{\eta(y)} = \inf_y \frac{y + 1}{\kappa + 1} = \frac{1}{\kappa + 1} > 0.$$

It then follows from Theorem 1 that there exists a uniformly optimal preprocessing function, which in our case is

$$\mathcal{T}_{\text{optimal}}(y) = 1 - \frac{\kappa + 1}{y + 1} = \frac{y - \kappa}{y + 1}. \tag{24}$$

Substituting (22) and (23) into (13), we get

$$\alpha_{\text{weak}} = \left[\sum_{y=0}^{\infty} \frac{[\mu(y) - \eta(y)]^2}{\eta(y)} \right]^{-1}$$

$$= \left[\sum_{y=0}^{\infty} \frac{\kappa^y (y - \kappa)^2}{(\kappa + 1)^{y+3}} \right]^{-1} = 1 + 1/\kappa.$$

Finally, the function $f(\beta)$ in (18) can be calculated as

$$f(\beta) = \frac{\beta}{(\kappa + 1)^2} \sum_{y=0}^{\infty} \left(\frac{\kappa}{\kappa + 1} \right)^y \frac{(y - \kappa)^2}{y + \beta(\kappa + 1) + 1}$$

$$= C_{\beta,\kappa} \int_0^{\frac{\kappa}{\kappa+1}} \frac{x^{\beta(\kappa+1)}}{1 - x} \, dx - \beta(\beta + 1), \tag{25}$$

where $C_{\beta,\kappa} = \beta(\beta+1)^2(1 + \frac{1}{\kappa})^{\beta(\kappa+1)+1}$. In reaching (25) we have used the identities that

$$\int_0^v \frac{x^u}{1 - x} \, dx = \sum_{y=0}^{\infty} \int_0^v x^{u+y} \, dx = \sum_{y=0}^{\infty} \frac{v^{y+u+1}}{y + u + 1},$$

for any $u > 0$ and $0 < v < 1$. In our case, we choose $u = \beta(\kappa + 1)$ and $v = \kappa/(\kappa + 1)$.

Figure 1 shows the function $f(\beta)$ for $\kappa = 5$. From its definition in (18), it is easy to verify that $f(\beta)$ is strictly increasing, and that $\lim_{\beta \to 0^+} f(\beta) = 0$ and $\lim_{\beta \to \infty} f(\beta) = 1/\alpha_{\text{weak}}$. It follows that, for each $\alpha > \alpha_{\text{weak}}$, there is a unique $\beta_\alpha > 0$ satisfying the equation $f(\beta_\alpha) = 1/\alpha$. Applying Theorem 1, the optimal performance curve is simply $\rho_{\text{optimal}}(\alpha) = (1 + \beta_\alpha)^{-1}$.

In Figure 2, we compare the proposed optimal preprocessing function in (24) against the the trimming scheme in (4), the subset scheme in (5), as well as $\mathcal{T}_{\text{MM}}(\cdot)$ in (14). In our experiments, the signal dimension is set to $n = 4096$ and $\kappa = 5$. For each given α, we set the parameter a in (4) to be the optimal integer choice within $\{1, 2, \ldots, 50\}$. For (5), its parameter b is tuned in the same way. For our optimal design, its theoretical curve is given by $\rho_{\text{optimal}}(\alpha)$; for the other three functions, we use the asymptotic predictions to be detailed in Section III-A to evaluate their theoretical curves. Simulations results in the figure show the averages over 16 independent trials, with the error bars indicating ± 1 standard deviation. The figure clearly demonstrates the improvement brought by the optimal design. In particular, we can see that the optimal preprocessing function (24) achieves the upper bound $\rho_{\text{optimal}}(\alpha)$. Its performance dominates that of (4), (5), and (14) uniformly over all α.

Example 2 (The Gaussian channel): In the second example, we consider a sensing model with Gaussian noise:

$$y_i = \max\left\{ \left| \langle \boldsymbol{a}_i, \boldsymbol{\xi} \rangle \right|^2 + w_i, \; 0 \right\},$$

where $w_i \sim_{\text{i.i.d.}} \mathcal{N}(0, \sigma^2)$. When $\sigma^2 > 0$, we have

$$\eta(y) = \begin{cases} \exp\left(\frac{\sigma^2}{2} - y \right) \Phi\left(\frac{y}{\sigma} - \sigma \right), & \text{if } y > 0, \\ \eta_0 \, \delta(y), & \text{if } y = 0, \end{cases}$$

and

$$\mu(y) = \begin{cases} (y - \sigma^2)\eta(y) + \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2} \right), & \text{if } y > 0, \\ \mu_0 \, \delta(y), & \text{if } y = 0, \end{cases}$$
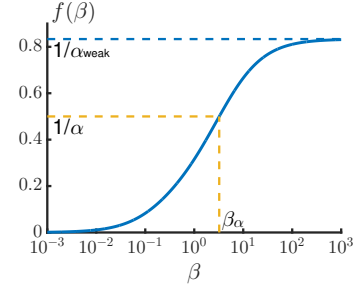

Fig. 1. The function $f(\beta)$ in (25) for $\kappa = 5$. Here, $\lim_{\beta \to \infty} f(\beta) = 1/\alpha_{\text{weak}} = 5/6$. For any $\alpha > \alpha_{\text{weak}}$, there is a unique positive solution β_α to the equation $f(\beta) = 1/\alpha$.
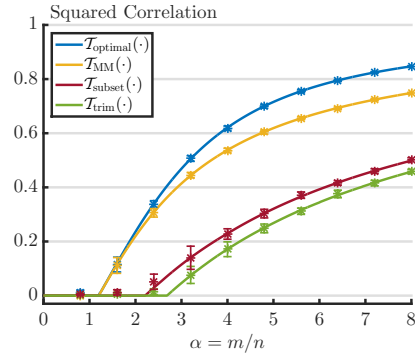

Fig. 2. Analytical predictions and numerical simulations for the Poisson channel with different preprocessing functions. $\mathcal{T}_{\text{optimal}}(\cdot)$ denotes the optimal function given in (24). It is compared against the trimming scheme in (4), the subset scheme in (5), as well as $\mathcal{T}_{\text{MM}}(\cdot)$ in (14). Numerical results are averaged over 16 independent trials.

where $\Phi(\cdot)$ is the CDF of the standard normal distribution, $\delta(\cdot)$ denotes the Dirac delta function, and η_0, μ_0 are two numerical constants defined as

$$\eta_0 \stackrel{\text{def}}{=} \int_0^\infty \Phi\left(-z/\sigma \right) e^{-z} \, dz = \frac{1}{2} - e^{\sigma^2/2} \Phi(-\sigma) \tag{26}$$

and

$$\mu_0 \stackrel{\text{def}}{=} \int_0^\infty \Phi\left(-z/\sigma \right) z e^{-z} \, dz \tag{27}$$

$$= \frac{1}{2} + (\sigma^2 - 1) e^{\sigma^2/2} \Phi(-\sigma) - \sigma/\sqrt{2\pi}$$

respectively. In Appendix B, we show that

$$\inf_{y \geq 0} \frac{\mu(y)}{\eta(y)} > 0. \tag{28}$$

It then follows from Theorem 1 that the optimal performance curve $\rho(\alpha; \mathcal{T}(\cdot))$ can be achieved by the following uniformly optimal preprocessing function

$$\mathcal{T}_{\text{optimal}}(y) = \begin{cases} 1 - \left[y - \sigma^2 + \frac{\sigma \Phi'(y/\sigma - \sigma)}{\Phi(y/\sigma - \sigma)} \right]^{-1}, & \text{if } y > 0, \\ 1 - \frac{\eta_0}{\mu_0}, & \text{if } y = 0, \end{cases} \tag{29}$$

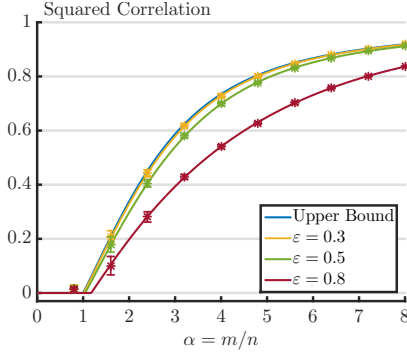where $\Phi'(\cdot)$ denotes the PDF of the standard normal distribution.

Fig. 3. Analytical predictions and numerical simulations for the noiseless observation model. The blue curve corresponds to the theoretical upper bound that no preprocessing function in the admissible set $\mathcal{F}$ can achieve. This upper bound can be approached by a family of preprocessing functions $\mathcal{T}_\alpha^\varepsilon(\cdot)$, as $\varepsilon \to 0$. Numerical results (for $n = 4096$) are averaged over 16 independent trials, with the error bars indicating ± 1 standard deviation.

Next, we consider the noiseless case, *i.e.*, $\sigma^2 = 0$ and $y_i = \left|\langle \boldsymbol{a}_i, \boldsymbol{\xi}\rangle\right|^2$. Here, the two functions $\eta(y)$ and $\mu(y)$ take much simpler forms:

$$\eta(y) = e^{-y} \quad \text{and} \quad \mu(y) = ye^{-y}, \tag{30}$$

but the challenge arises from the fact that, in this case,

$$\inf_{y \geq 0} \frac{\mu(y)}{\eta(y)} = \frac{\mu(0)}{\eta(0)} = 0. \tag{31}$$

As stated in Theorem 1, under (31), the optimal performance curve $\rho(\alpha; \mathcal{T}(\cdot))$ cannot be achieved by any admissible function $\mathcal{T}(\cdot) \in \mathcal{F}$. In this case, we can construct a family of admissible preprocessing functions whose performance arbitrarily approaches $\rho(\alpha; \mathcal{T}(\cdot))$. For any $\alpha > \alpha_{\text{weak}}$, these functions take the form of

$$\mathcal{T}_\alpha^\varepsilon(y) = \frac{c_\alpha^\varepsilon(y)}{1 + c_\alpha^\varepsilon(y)}, \tag{32}$$

where $0 < \varepsilon < 1$ is a parameter. Substituting (30) into (21), we have

$$c_\alpha^\varepsilon(y) = \max\left\{v_\alpha^\varepsilon \frac{\beta_\alpha(y-1)}{\beta_\alpha + y}, -1 + \varepsilon\right\},$$

and v_α^ε is a positive constant that can be uniquely determined by the following equation:

$$\int_0^\infty c_\alpha^\varepsilon(y)(y-1)e^{-y}\,\mathrm{d}y = 1/\alpha.$$

We show in Section IV-C that, as $\varepsilon \to 0$, the performance of $\mathcal{T}_\alpha^\varepsilon(\cdot)$ will converge to the optimal performance curve $\rho_{\text{optimal}}(\alpha)$. This is demonstrated in Figure 3, where we compare the performance curves of the preprocessing functions (32) for three different values of ε against the theoretical upper bound $\rho_{\text{optimal}}(\alpha)$. We see that, when $\varepsilon = 0.3$, the achievable curve is already very close to $\rho_{\text{optimal}}(\alpha)$.

III. TECHNICAL BACKGROUND

A. Asymptotic Characterizations of the Spectral Method

To set the stage for proving our main results on optimal design, we first review the precise asymptotic characterizations of the spectral method obtained in previous work [14], [15].

Let S and Y be the random variables defined in (9) and (16), respectively. Recall the admissible set $\mathcal{F}$ defined in (17). For any preprocessing function $\mathcal{T}(\cdot) \in \mathcal{F}$, the support of the probability measure of the random variable $\mathcal{T}(Y)$ is bounded. Let τ denote the upper boundary of the support, *i.e.*,

$$\tau \stackrel{\text{def}}{=} \sup_{y \in \Gamma_Y} \mathcal{T}(y) < \infty.$$

We consider two functions

$$\phi(\lambda) \stackrel{\text{def}}{=} \lambda \, \mathbb{E}_{S,Y}\left[\frac{\mathcal{T}(Y)\,|S|^2}{\lambda - \mathcal{T}(Y)}\right] \tag{33}$$

and

$$\psi_\alpha(\lambda) \stackrel{\text{def}}{=} \lambda/\alpha + \lambda \, \mathbb{E}_Y\left[\frac{\mathcal{T}(Y)}{\lambda - \mathcal{T}(Y)}\right], \tag{34}$$

both defined on the open interval (τ, ∞). Within their domains, it is easy to check that both functions are convex and that $\phi(\lambda)$ is strictly decreasing. Consequently, if the following conditions

$$\psi_\alpha(\lambda^*) = \phi(\lambda^*) \tag{35}$$

and

$$\psi_\alpha'(\lambda^*) > 0, \tag{36}$$

hold for some $\lambda^* \in (\tau, \infty)$, then that λ^* must be unique.

Theorem 2 (Asymptotic characterization [14], [15]): Let the target signal $\boldsymbol{\xi}$ be an arbitrary vector in $\mathbb{C}^n$ with $\|\boldsymbol{\xi}\| = 1$. Assume that the preprocessing function $\mathcal{T}(\cdot) \in \mathcal{F}$, and that the sensing vectors $\boldsymbol{a}_i \sim_{\text{i.i.d.}} \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_n)$. As $m, n \to \infty$ with $m/n \to \alpha \in (0, \infty)$, we have

$$\frac{\left|\langle \boldsymbol{\xi}, \boldsymbol{x}_1\rangle\right|^2}{\|\boldsymbol{\xi}\|^2 \|\boldsymbol{x}_1\|^2} \xrightarrow{\mathcal{P}} \rho(\alpha; \mathcal{T}(\cdot)),$$

where the limit value on the right-hand side is

$$\rho(\alpha; \mathcal{T}(\cdot)) \stackrel{\text{def}}{=} \begin{cases} \frac{\psi_\alpha'(\lambda^*)}{\psi_\alpha'(\lambda^*) - \phi'(\lambda^*)}, & \text{if (35)(36) hold for } \lambda^* > \tau, \\ 0, & \text{otherwise.} \end{cases} \tag{37}$$

The above theorem shows that, in the high-dimensional limit, the squared correlation between $\boldsymbol{\xi}$ and the estimate $\boldsymbol{x}_1$ converges in probability to a deterministic value $\rho(\alpha; \mathcal{T}(\cdot))$, which can be exactly computed as in (37). Moreover, this asymptotic prediction exhibits a phase transition phenomenon: $\rho(\alpha; \mathcal{T}(\cdot))$ is nonzero *if and only if* (35) and (36) hold for some $\lambda^* > \tau$.

Remark 4: The asymptotic prediction stated in Theorem 2 was first obtained in [14] for the case of real-valued sensing vectors. In that setting, the random variable S in (33) and (34) should be drawn from $\mathcal{N}(0, 1)$. Additionally, [14] makes an assumption that $\mathcal{T}(y) \geq 0$. Later, Mondelli and Montanari show that the same characterization holds for the complex-valued case. They also removed the restriction that $\mathcal{T}(y)$ be positive by generalizing a result for spiked random positive semidefinite matrices [16] to spiked Hermitian matrices. In [14], [15], there is also a technical assumption that, as λ approaches τ from the right,

$$\lim_{\lambda \to \tau^+} \mathbb{E}_Y\left[\frac{\mathcal{T}(Y)}{(\lambda - \mathcal{T}(Y))^2}\right] = \lim_{\lambda \to \tau^+} \mathbb{E}_{S,Y}\left[\frac{\mathcal{T}(Y)\,|S|^2}{\lambda - \mathcal{T}(Y)}\right] = \infty. \tag{38}$$

However, a close inspection of the arguments in [14] (especially those in Propositions 2, 3, and 4 there) will show that the first equality in (38) is used to guarantee that $\phi(\lambda)$ intersects with a modified version of $\psi_\alpha(\lambda)$ at a point $\lambda^* > \tau$ whereas the second equality in (38) is needed in the proof of an unrelated result. Neither condition will be necessary here, if we present the limit value $\rho(\alpha; \mathcal{T}(\cdot))$ in the form of (37).

B. Reformulations of the Optimal Design Problem

The optimal design problem formulated in (6) seeks to find the supremum of $\rho(\alpha; \mathcal{T}(\cdot))$ over all admissible preprocessing functions in the set $\mathcal{F}$. Using the asymptotic characterizations of $\rho(\alpha; \mathcal{T}(\cdot))$ given in (37), we can convert the problem to

$$\sup_{\mathcal{T}(\cdot) \in \mathcal{F}} \frac{\psi_\alpha'(\lambda^*)}{\psi_\alpha'(\lambda^*) - \phi'(\lambda^*)}$$
$$\text{s.t. } \psi_\alpha(\lambda^*) = \phi(\lambda^*) \text{ and } \psi_\alpha'(\lambda^*) > 0 \text{ for some } \lambda^* > \tau.$$

For a given $\mathcal{T}(\cdot)$ and a given α, it is possible that there is no $\lambda^* > \tau$ satisfying the equality and inequality constraints in the above optimization problem. In that case, the value of the objective function is understood to be equal to 0.

Finding the supremum of $\frac{\psi_\alpha'(\lambda^*)}{\psi_\alpha'(\lambda^*) - \phi'(\lambda^*)}$ is equivalent to finding the infimum of $\frac{-\phi'(\lambda^*)}{\psi_\alpha'(\lambda^*)}$, where the numerator $-\phi'(\lambda^*)$ is positive due to the monotonicity of $\phi(\lambda)$. We also rewrite the expectations in (33) and (34) in terms of $\eta(y)$ and $\mu(y)$ defined in (10) and (11), as

$$\phi(\lambda) = \lambda \int_{\Gamma_Y} \frac{\mathcal{T}(y)}{\lambda - \mathcal{T}(y)} \mu(y) \, \mathrm{d}y$$

$$\psi_\alpha(\lambda) = \lambda/\alpha + \lambda \int_{\Gamma_Y} \frac{\mathcal{T}(y)}{\lambda - \mathcal{T}(y)} \eta(y) \, \mathrm{d}y.$$

Taking derivatives then gives us

$$\inf_{\mathcal{T}(\cdot) \in \mathcal{F}} \frac{\int_{\Gamma_Y} \left[\frac{\mathcal{T}(y)}{\lambda^* - \mathcal{T}(y)}\right]^2 \mu(y) \, \mathrm{d}y}{1/\alpha - \int_{\Gamma_Y} \left[\frac{\mathcal{T}(y)}{\lambda^* - \mathcal{T}(y)}\right]^2 \eta(y) \, \mathrm{d}y} \tag{39}$$

$$\text{s.t. } \int_{\Gamma_Y} \frac{\mathcal{T}(y)}{\lambda^* - \mathcal{T}(y)} [\mu(y) - \eta(y)] \, \mathrm{d}y = 1/\alpha$$

$$\text{and } \int_{\Gamma_Y} \left[\frac{\mathcal{T}(y)}{\lambda^* - \mathcal{T}(y)}\right]^2 \eta(y) \, \mathrm{d}y < 1/\alpha \text{ for some } \lambda^* > \tau.$$

The problem in (39) still appears unwieldy. To further simplify it, we observe that the objective function of (39) as well as the equality and inequality constraints are all scale invariant since they are related to $\mathcal{T}(y)$ only through the ratio $\frac{\mathcal{T}(y)}{\lambda^* - \mathcal{T}(y)}$. Thus, if $(\mathcal{T}(y), \lambda^*)$ is a feasible solution satisfying the constraints, so will be $(a\mathcal{T}(y), a\lambda^*)$ for any constant $a > 0$. Meanwhile, the value of the objective function remains unchanged. Exploiting this invariance, we can always assume $\lambda^* = 1$, without loss of generality. This allows us to introduce a change of variables

$$c(y) \stackrel{\text{def}}{=} \frac{\mathcal{T}(y)}{1 - \mathcal{T}(y)}, \tag{40}$$

and simplify (39) as

$$(\mathrm{P}_0) \quad V_0 \stackrel{\text{def}}{=} \inf_{c(\cdot) \in \mathcal{F}_c} \frac{\int_{\Gamma_Y} [c(y)]^2 \mu(y) \, \mathrm{d}y}{1/\alpha - \int_{\Gamma_Y} [c(y)]^2 \eta(y) \, \mathrm{d}y} \tag{41}$$

$$\text{s.t. } \int_{\Gamma_Y} c(y)[\mu(y) - \eta(y)] \, \mathrm{d}y = 1/\alpha \tag{42}$$

$$\text{and } \int_{\Gamma_Y} [c(y)]^2 \eta(y) \, \mathrm{d}y < 1/\alpha, \tag{43}$$

where $\mathcal{F}_c$ denotes the admissible set for $c(y)$, defined as

$$\mathcal{F}_c \stackrel{\text{def}}{=} \left\{ c(\cdot) : 0 < \sup_{y \in \Gamma_Y} c(y) < \infty \text{ and } \inf_{y \in \Gamma_Y} c(y) > -1 \right\}. \tag{44}$$

Note that the mapping (40), or equivalently, $\mathcal{T}(y) = \frac{c(y)}{1 + c(y)}$, provides a one-to-one correspondance between $\mathcal{F}_c$ and

$$\mathcal{F} \cap \left\{ \mathcal{T}(\cdot) : \sup_{y \in \Gamma_Y} \mathcal{T}(y) < 1 \right\}.$$

The additional constraint that $\sup_{y \in \Gamma_Y} \mathcal{T}(y) < 1$ is both necessary and sufficient for our purpose, as we have fixed $\lambda^* = 1$. Moreover, we adopt the following notational convention: if there is no feasible $c(y) \in \mathcal{F}_c$ satisfying the constraints (42) and (43), the value of the objective function $V_0 = +\infty$.

It will be more convenient to study

$$(\mathrm{P}_1) \quad V_1 \stackrel{\text{def}}{=} \inf_{c(\cdot) \in \mathcal{H}} \frac{\int_{\Gamma_Y} [c(y)]^2 \mu(y) \, \mathrm{d}y}{1/\alpha - \int_{\Gamma_Y} [c(y)]^2 \eta(y) \, \mathrm{d}y} \tag{45}$$

$$\text{s.t. (42) and (43) hold,} \tag{46}$$

where we simply relax $\mathcal{F}_c$ in (44) to a larger set

$$\mathcal{H} \stackrel{\text{def}}{=} \left\{ c(\cdot) : \int_{\Gamma_Y} [c(y)]^2 (\eta(y) + \mu(y)) \, \mathrm{d}y < \infty \right\}. \tag{47}$$

We note that any function in $\mathcal{F}_c$ is finitely bounded. It then follows from (12) that the function must belong to $\mathcal{H}$. In the next section, we will present a closed-form solution to (P_1). It forms the foundation of our proof of Theorem 1.

IV. Proof of Theorem 1

In this section, we prove Theorem 1 in three steps. First, we show in Section IV-A that the right-hand side of (19) is an upper bound for the optimal performance curve $\rho_{\mathrm{optimal}}(\alpha)$. To establish equality, we consider two cases, depending on the value of $\inf_y \mu(y)/\eta(y)$. When $\inf_y \mu(y)/\eta(y) > 0$, we show in Section IV-B that the aforementioned upper bound can be achieved by the uniformly optimal solution given in (20), and that this optimal solution belongs to the feasible set $\mathcal{F}$. The remaining case, when $\inf_y \mu(y)/\eta(y) = 0$, is considered in Section IV-C, where we construct a family of admissible functions $\{\mathcal{T}_\alpha^\varepsilon(\cdot)\}_{0 < \varepsilon < 1}$ and show that their performance curves approach the upper bound.

A. An Upper Bound for $\rho_{\mathrm{optimal}}(\alpha)$

Following the discussions in Section III-B, we know that

$$\rho_{\mathrm{optimal}}(\alpha) = \frac{1}{1 + V_0} \leq \frac{1}{1 + V_1}, \tag{48}$$

where V_0 and V_1 are the optimal values of (P_0) and (P_1), respectively. That $V_1 \leq V_0$ is due to the fact that the set $\mathcal{F}_c$

in (P_0) is a subset of $\mathcal{H}$ in (P_1). In what follows, we present a solution of (P_1).

For each $\beta > 0$, the (sublevel set) condition

$$\frac{\int_{\Gamma_Y} [c(y)]^2 \mu(y)\, dy}{1/\alpha - \int_{\Gamma_Y} [c(y)]^2 \eta(y)\, dy} \le \beta$$

is equivalent to $\int_{\Gamma_Y} [c(y)]^2 [\mu(y)/\beta + \eta(y)]\, dy \le 1/\alpha$. To lighten the notation, we define two functionals

$$\mathcal{Q}(c(\cdot), \beta) \overset{\text{def}}{=} \int_{\Gamma_Y} [c(y)]^2 [\mu(y)/\beta + \eta(y)]\, dy, \qquad (49)$$

and

$$\mathcal{L}(c(\cdot)) \overset{\text{def}}{=} \int_{\Gamma_Y} c(y) [\mu(y) - \eta(y)]\, dy. \qquad (50)$$

It is easy to see that (P_1) is equivalent to

$$V_1 = \inf_{\Omega_\beta \text{ is nonempty}} \beta, \qquad (51)$$

where

$$\Omega_\beta \overset{\text{def}}{=} \left\{ c(\cdot) \in \mathcal{H} : \mathcal{Q}(c(\cdot), \beta) \le \tfrac{1}{\alpha} \text{ and } \mathcal{L}(c(\cdot)) = \tfrac{1}{\alpha} \right\}. \quad (52)$$

Note that we can omit the quadratic constraint in (43) as it is implied by $\mathcal{Q}(c(\cdot), \beta) \le \tfrac{1}{\alpha}$.

Lemma 1: For any $\beta > 0$,

$$\frac{1}{\alpha^2 f(\beta)} = \min_{c(\cdot) \in \mathcal{H}} \mathcal{Q}(c(\cdot), \beta)$$
$$\text{s.t. } \mathcal{L}(c(\cdot)) = 1/\alpha \qquad (53)$$

where $f(\beta)$ is the function defined in (18). Moreover, the optimal solution to (53) is given by

$$c^*(y) = \frac{\mu(y) - \eta(y)}{[\alpha f(\beta)](\eta(y) + \mu(y)/\beta)}. \qquad (54)$$

Proof: Since $\eta(y) + \mu(y)/\beta$ is a nonnegative function, we can define a *weighted L^2* function space, where the inner product between two functions $f_1(y)$ and $f_2(y)$ is

$$\langle f_1(y), f_2(y) \rangle_\beta \overset{\text{def}}{=} \int_{\Gamma_Y} f_1(y) f_2(y) [\eta(y) + \mu(y)/\beta]\, dy.$$

The optimization problem (53) then becomes that of finding a minimum norm solution on a linear variety, *i.e.*,

$$\min_{c(\cdot)} \langle c(y), c(y) \rangle_\beta$$
$$\text{s.t. } \langle c(y), \frac{\mu(y) - \eta(y)}{\eta(y) + \mu(y)/\beta} \rangle_\beta = 1/\alpha.$$

The optimal solution is well-known (see, *e.g.*, [17]). It should take the form of

$$c^*(y) = v \frac{\mu(y) - \eta(y)}{\eta(y) + \mu(y)/\beta},$$

where the scaling constant v is determined by the constraint $\int_{\Gamma_Y} c^*(y)[\mu(y) - \eta(y)]\, dy = 1/\alpha$. Solving this equation gives $v = 1/[\alpha f(\beta)]$ and thus (54), the squared norm of which gives us the left-hand side of (53). ∎

Applying Lemma 1, we know that the set Ω_β in (52) is nonempty if and only if

$$f(\beta) \ge 1/\alpha. \qquad (55)$$

From its definition given in (18), $f(\beta)$ is a strictly increasing function, with

$$\lim_{\beta \to 0} f(\beta) = 0 \quad \text{and} \quad \lim_{\beta \to \infty} f(\beta) = 1/\alpha_{\text{weak}}.$$

Consequently, for $\alpha \le \alpha_{\text{weak}}$, the condition (55) cannot be satisfied by any finite β. In this case, the optimal value of (51) is $V_1 = \infty$. Using (48), we conclude that

$$\rho_{\text{optimal}}(\alpha) = 0 \quad \text{for } \alpha \le \alpha_{\text{weak}},$$

recovering the results previously obtained in [15].

When $\alpha > \alpha_{\text{weak}}$, there is a unique $\beta_\alpha > 0$ such that $f(\beta_\alpha) = 1/\alpha$. It follows from the monotonicity of $f(\beta)$ that $V_1 = \beta_\alpha$. Substituting this into (48), we get

$$\rho_{\text{optimal}}(\alpha) \le (1 + \beta_\alpha)^{-1} \quad \text{for } \alpha > \alpha_{\text{weak}}. \qquad (56)$$

In the next two subsections, we show that this inequality is in fact an equality.

B. Uniformly Optimal Preprocessing Function

First consider the case when $\inf_y \mu(y)/\eta(y) > 0$. In all of our subsequent discussions, we shall also assume that $\alpha_{\text{weak}} < \infty$. For each $\alpha > \alpha_{\text{weak}}$, we know from Lemma 1 that the upper bound in (56) is achieved by

$$c^*(y) = \frac{\mu(y) - \eta(y)}{[\alpha f(\beta_\alpha)](\eta(y) + \mu(y)/\beta_\alpha)} = \frac{\mu(y) - \eta(y)}{\eta(y) + \mu(y)/\beta_\alpha}, \qquad (57)$$

where the second equality is due to the fact that $f(\beta_\alpha) = 1/\alpha$. It is easy to verify that

$$-1 < \inf_{y \in \Gamma_Y} c^*(y) \le \sup_{y \in \Gamma_Y} c^*(y) \le \beta_\alpha. \qquad (58)$$

From (40), the corresponding preprocessing function is

$$\mathcal{T}(y) = \frac{c^*(y)}{1 + c^*(y)} = \frac{1}{1 + 1/\beta_\alpha}\left[1 - \frac{\eta(y)}{\mu(y)}\right].$$

As mentioned in Section III-B, the performance of the spectral algorithm is scale invariant. So a scaled version

$$\mathcal{T}_{\text{optimal}}(y) = 1 - \frac{\eta(y)}{\mu(y)} \qquad (59)$$

can achieve the same performance. Next, we show that $\mathcal{T}_{\text{optimal}}(y) \in \mathcal{F}$. This would then imply that $\rho_{\text{optimal}}(\alpha) \ge (1 + \beta_\alpha)^{-1}$, which, together with (56), gives us (19).

Since $\eta(y) \ge 0$ and $\mu(y) \ge 0$,

$$\tau = \sup_{y \in \Gamma_Y} \mathcal{T}_{\text{optimal}}(y) \le 1.$$

Under the condition $\inf_y \mu(y)/\eta(y) > 0$, we also have $\inf_{y \in \Gamma_Y} \mathcal{T}_{\text{optimal}}(y) > -\infty$. What remains to be shown is that $\tau > 0$. To that end, we note that, under the assumption $\alpha_{\text{weak}} < \infty$, there must exist y for which $\mu(y) > \eta(y)$. (See our discussions at the end of Appendix A.) This then guarantees that $\tau > 0$.

C. Truncated Preprocessing Functions

In this section, we consider the case when

$$\inf_y \mu(y)/\eta(y) = 0. \tag{60}$$

Under (60), the function $\mathcal{T}_{\text{optimal}}(\cdot)$ in (59) is not lower bounded, and thus it is not in the admissible set $\mathcal{F}$. By the uniqueness of (P_1) and since $\mathcal{F}$ is a subset of $\mathcal{H}$ in (47), we can conclude that the right-hand side of (19) cannot be achieved by any function in $\mathcal{F}$ in this case.

Next, we show that the family of preprocessing functions $\{\mathcal{T}_\alpha^\varepsilon(\cdot) \in \mathcal{F}\}_{0 < \varepsilon < 1}$ defined in (21) can approach the optimal performance curve, *i.e.*,

$$\lim_{\varepsilon \to 0} \rho(\alpha; \mathcal{T}_\alpha^\varepsilon(\cdot)) = \rho_{\text{optimal}}(\alpha) \tag{61}$$

for all $\alpha > \alpha_{\text{weak}}$.

We start by showing that the scaler v_α^ε in (21) can indeed be uniquely determined by the linear constraint $\mathcal{L}(c_\alpha^\varepsilon(\cdot)) = 1/\alpha$. To that end, we define

$$c_v(y) = \max\left\{ v \frac{\mu(y) - \eta(y)}{\eta(y) + \mu(y)/\beta_\alpha}, -1 + \varepsilon \right\}, \tag{62}$$

with $v \geq 0$ being a varying parameter, and examine

$$h(v) \stackrel{\text{def}}{=} \int_{\Gamma_Y} c_v(y)[\mu(y) - \eta(y)]\,\mathrm{d}y \tag{63}$$

$$= \int_{\Gamma_Y} \max\left\{ v \frac{\mu(y) - \eta(y)}{\eta(y) + \mu(y)/\beta_\alpha}, -1 + \varepsilon \right\} [\mu(y) - \eta(y)]\,\mathrm{d}y.$$

We note that $-1 + \varepsilon < 0$ and $c_v(y)[\mu(y) - \eta(y)] \geq 0$ for all $y \in \Gamma_Y$. A moment of thought will convince us that $h(v)$ is an increasing function of v for $v \geq 0$. Moreover, $h(0) = 0$. To study the limit of the function as $v \to \infty$, we denote by Γ_Y^+ the subset over which $\mu(y) > \eta(y)$. As shown in Appendix A, when $\alpha_{\text{weak}} < \infty$, the subset Γ_Y^+ is nonempty. We have

$$h(v) \geq \int_{\Gamma_Y^+} c_v(y)[\mu(y) - \eta(y)]\,\mathrm{d}y$$

$$= v \int_{\Gamma_Y^+} \frac{[\mu(y) - \eta(y)]^2}{\eta(y) + \mu(y)/\beta_\alpha}\,\mathrm{d}y,$$

which tends to ∞ as $v \to \infty$. It then follows that there exists a unique positive solution v_α^ε to the equation $h(v) = 1/\alpha$.

To establish (61), we recall our reformulations of the optimal design problem presented in Section III-B. Given the equivalence of the optimal design problem and the optimization problem (P_0) in (41), our tasks boil down to showing that (1) $c_\alpha^\varepsilon(y) \in \mathcal{F}_c$ and (2) for each $\alpha > \alpha_{\text{weak}}$,

$$\lim_{\varepsilon \to 0} \int_{\Gamma_Y} [c_\alpha^\varepsilon(y)]^2 \eta(y)\,\mathrm{d}y = \int_{\Gamma_Y} [c^*(y)]^2 \eta(y)\,\mathrm{d}y, \tag{64}$$

$$\lim_{\varepsilon \to 0} \int_{\Gamma_Y} [c_\alpha^\varepsilon(y)]^2 \mu(y)\,\mathrm{d}y = \int_{\Gamma_Y} [c^*(y)]^2 \mu(y)\,\mathrm{d}y, \tag{65}$$

where $c^*(y)$ is the optimal solution given in (57). Note that the linear constraint (42) in (P_0) is always satisfied, due to the way we set the scalar v_α^ε in (21). The quadratic constraint (43) will also be satisfied for all sufficiently small ε, given the convergence in (64) and the fact that we have $\int_{\Gamma_Y} [c^*(y)]^2 \eta(y)\,\mathrm{d}y < 1/\alpha$.

By its definition in (21), it is easy to see that

$$-1 + \varepsilon \leq \inf_{y \in \Gamma_Y} c_\alpha^\varepsilon(y) \leq \sup_{y \in \Gamma_Y} c_\alpha^\varepsilon(y) \leq v_\alpha^\varepsilon \beta_\alpha. \tag{66}$$

Moreover, since $\mu(y) > \eta(y)$ over a nonempty subset of Γ_Y, we have $\sup_{y \in \Gamma_Y} c_\alpha^\varepsilon(y) > 0$. Thus, we can verify that $c_\alpha^\varepsilon(y) \in \mathcal{F}_c$.

Next, we show that $v_\alpha^\varepsilon \to 1$ as $\varepsilon \to 0$. Recall that

$$\mathcal{L}(c_\alpha^\varepsilon(\cdot)) = \mathcal{L}(c^*(\cdot)) = 1/\alpha. \tag{67}$$

The latter equality implies that

$$\int_{\Gamma_Y} \max\left\{ c^*(y), -1 + \varepsilon \right\} [\mu(y) - \eta(y)]\,\mathrm{d}y \leq 1/\alpha,$$

which, using the notation introduced in (62) and (63), can be written as $h(1) \leq 1/\alpha$. Meanwhile, the first equality in (67) can be written as $h(v_\alpha^\varepsilon) = 1/\alpha$. Since $h(v)$ is an increasing function of v, we must have $v_\alpha^\varepsilon \geq 1$.

Define two sets Γ_1 and Γ_2, with $\Gamma_1 \cup \Gamma_2 = \Gamma_Y$, such that $c_\alpha^\varepsilon(y) = -1 + \varepsilon$ for $y \in \Gamma_1$ and $c_\alpha^\varepsilon(y) = v_\alpha^\varepsilon \frac{\mu(y) - \eta(y)}{\eta(y) + \mu(y)/\beta_\alpha}$ for $y \in \Gamma_2$. Similarly, define a subset Γ_1' such that $c^*(y) \leq -1 + \varepsilon$ for $y \in \Gamma_1'$. We can easily verify that $\Gamma_1' \subset \Gamma_1$, since $c_\alpha^\varepsilon(y) = \max\left\{ v_\alpha^\varepsilon c^*(y), -1 + \varepsilon \right\}$ and $v_\alpha^\varepsilon \geq 1$. It follows from (67) that

$$0 = \mathcal{L}(c^*(\cdot)) - \mathcal{L}(c_\alpha^\varepsilon(\cdot))$$

$$= \int_{\Gamma_1'} \left[c^*(y) - (-1 + \varepsilon) \right] \left(\mu(y) - \eta(y) \right)\,\mathrm{d}y$$

$$+ \int_{\Gamma_1 \setminus \Gamma_1'} \left[c^*(y) - (-1 + \varepsilon) \right] \left(\mu(y) - \eta(y) \right)\,\mathrm{d}y$$

$$- (v_\alpha^\varepsilon - 1) \int_{\Gamma_2} \frac{\mu(y) - \eta(y)}{\eta(y) + \mu(y)/\beta_\alpha} \left(\mu(y) - \eta(y) \right)\,\mathrm{d}y. \tag{68}$$

For $y \in \Gamma_1 \setminus \Gamma_1'$, we know $\mu(y) - \eta(y) < 0$ and $c^*(y) \geq -1 + \varepsilon$. Thus,

$$\int_{\Gamma_1 \setminus \Gamma_1'} \left[c^*(y) - (-1 + \varepsilon) \right] \left(\mu(y) - \eta(y) \right)\,\mathrm{d}y \leq 0. \tag{69}$$

Also for $y \in \Gamma_1'$, we have $-1 \leq c^*(y) \leq (-1 + \varepsilon)$, so

$$\left| \int_{\Gamma_1'} \left[c^*(y) - (-1 + \varepsilon) \right] \left(\mu(y) - \eta(y) \right)\,\mathrm{d}y \right|$$

$$\leq \varepsilon \int_{\Gamma_Y} [\mu(y) + \eta(y)]\,\mathrm{d}y = 2\varepsilon. \tag{70}$$

Substituting (69) and (70) into (68), we get

$$(v_\alpha^\varepsilon - 1) \int_{\Gamma_2} \frac{[\mu(y) - \eta(y)]^2}{\eta(y) + \mu(y)/\beta_\alpha}\,\mathrm{d}y \leq 2\varepsilon.$$

Let Γ_Y^+ be the subset such that $\mu(y) > \eta(y)$ for $y \in \Gamma_Y^+$. We must have $\Gamma_Y^+ \subset \Gamma_2$. It follows that

$$1 \leq v_\alpha^\varepsilon \leq 1 + 2\varepsilon \left(\int_{\Gamma_Y^+} \frac{[\mu(y) - \eta(y)]^2}{\eta(y) + \mu(y)/\beta_\alpha}\,\mathrm{d}y \right)^{-1}$$

and thus $v_\alpha^\varepsilon \to 1$ as $\varepsilon \to 0$.

What remain to be shown are (64) and (65). The proofs for the two cases are essentially identical, so we only show how to establish (64), as follows:

$$
\left| \int_{\Gamma_Y} ([c_\alpha^\varepsilon(y)]^2 - [c^*(y)]^2)\eta(y)\,\mathrm{d}y \right|
$$

$$
\leq \int_{\Gamma_Y} \left| c_\alpha^\varepsilon(y) - c^*(y) \right| \left(\left| c_\alpha^\varepsilon(y) \right| + \left| c^*(y) \right| \right)\eta(y)\,\mathrm{d}y
$$

$$
\overset{(a)}{\leq} C_\varepsilon \int_{\Gamma_Y} \left| c_\alpha^\varepsilon(y) - c^*(y) \right| \eta(y)\,\mathrm{d}y
$$

$$
\leq C_\varepsilon \int_{\Gamma_Y} \left| c_\alpha^\varepsilon(y) - \max\left\{ c^*(y), -1+\varepsilon \right\} \right| \eta(y)\,\mathrm{d}y
$$

$$
\qquad + C_\varepsilon \int_{\Gamma_Y} \left| \max\left\{ c^*(y), -1+\varepsilon \right\} - c^*(y) \right| \eta(y)\,\mathrm{d}y
$$

$$
\overset{(b)}{\leq} C_\varepsilon |v_\alpha^\epsilon - 1| \int_{\Gamma_Y} \left| c^*(y) \right| \eta(y)\,\mathrm{d}y + C_\varepsilon \cdot \varepsilon \int_{\Gamma_Y} \eta(y)\,\mathrm{d}y
$$

$$
\overset{(c)}{\leq} C_\varepsilon |v_\alpha^\epsilon - 1| (1 + \beta_\alpha) + C_\varepsilon \cdot \varepsilon.
$$

Here, to obtain (a), we have used the boundedness of $c_\alpha^\varepsilon(y)$ as given in (66) and that of $c^*(y)$ as given in (58). Consequently, it is sufficient to set the constant to be

$$
C_\varepsilon = 2 + (v_\alpha^\varepsilon + 1)\beta_\alpha.
$$

To reach (b), we have used the following properties:

$$
\left| \max\left\{ x, -1+\varepsilon \right\} - \max\left\{ y, -1+\varepsilon \right\} \right| \leq |x - y|
$$

and

$$
\left| \max\left\{ x, -1+\varepsilon \right\} - x \right| \leq \varepsilon
$$

for any $x, y \geq -1$. Finally, the inequality (c) follows from the boundedness of $\left| c^*(y) \right|$ and the fact that $\int_{\Gamma_Y} \eta(y)\,\mathrm{d}y = 1$. Since $v_\alpha^\varepsilon \to 1$ as $\varepsilon \to 0$, we have (64).

APPENDIX

A. Several properties of α_{weak}

In this appendix, we show that the integral on the right-hand side of (13) is always finite. In particular, we establish the following fundamental lower bound on the weak recovery threshold.

Proposition 1: For any sensing model given in (1), we have

$$
\alpha_{\mathrm{weak}} \geq 1,
$$

where the lower bound is achieved when $y_i = \left| \langle \boldsymbol{a}_i, \boldsymbol{\xi} \rangle \right|^2$.

Proof: Using the definition in (13), we have

$$
\alpha_{\mathrm{weak}}^{-1} = \int_{\Gamma_Y} \frac{(\mu(y) - \eta(y))^2}{\eta(y)}\,\mathrm{d}y
$$

$$
= \int_{\Gamma_Y} \frac{[\mu(y)]^2}{\eta(y)}\,\mathrm{d}y - 2\int_{\Gamma_Y} \mu(y)\,\mathrm{d}y + \int_{\Gamma_Y} \eta(y)\,\mathrm{d}y
$$

$$
= \int_{\Gamma_Y} \frac{[\mu(y)]^2}{\eta(y)}\,\mathrm{d}y - 1, \tag{71}
$$

where the last equality is due to (12). Thus, we just need to show that $\int_{\Gamma_Y} \frac{[\mu(y)]^2}{\eta(y)}\,\mathrm{d}y \leq 2$.

Let $Z = |S|$, where $S \sim \mathcal{CN}(0,1)$. Let $g(z)$ denote the density function of Z. We have

$$
[\mu(y)]^2 = \left[\int_0^\infty z^2\, p(y \mid z) g(z)\,\mathrm{d}z \right]^2
$$

$$
\leq \left[\int_0^\infty z^4\, p(y \mid z) g(z)\,\mathrm{d}z \right] \left[\int_0^\infty p(y \mid z) g(z)\,\mathrm{d}z \right]
$$

$$
= \left[\int_0^\infty z^4\, p(y \mid z) g(z)\,\mathrm{d}z \right] \eta(y),
$$

where the bound is due to the Cauchy-Schwarz inequality. Using this upper bound, we have

$$
\int_{\Gamma_Y} \frac{[\mu(y)]^2}{\eta(y)}\,\mathrm{d}y \leq \int_0^\infty z^4 g(z) \int_{\Gamma_Y} p(y \mid z)\,\mathrm{d}y\,\mathrm{d}z
$$

$$
= \int_0^\infty z^4 g(z)\,\mathrm{d}z
$$

$$
= \mathbb{E}[Z^4] = 2. \tag{72}
$$

For the noiseless channel described in Section II-B, we have $\eta(y) = e^{-y}$ and $\mu(y) = y e^{-y}$ for $y \geq 0$. So $\int_{\Gamma_Y} \frac{[\mu(y)]^2}{\eta(y)}\,\mathrm{d}y = \int_0^\infty y^2 e^{-y} = 2$, which achieves the bound in (72). $\blacksquare$

Remark 5: When the sensing vectors are real-valued Gaussian random vectors, the lower bound for the weak recovery threshold reduces to $\alpha_{\mathrm{weak}} \geq 1/2$. We can use the same proof as above, with only one change: In (72), the random variable $Z = |S|$ with $S \sim \mathcal{N}(0,1)$. Thus, $\mathbb{E}[Z^4] = 3$.

In our proof of Theorem 1, we will also use the following property: Let $\Gamma_Y^+ \overset{\mathrm{def}}{=} \{ y : \mu(y) > \eta(y) \}$. If $\alpha_{\mathrm{weak}} < \infty$, then the subset Γ_Y^+ is nonempty, with $\int_{\Gamma_Y^+} \eta(y)\,\mathrm{d}y > 0$. To establish this, we recall from (71) that $\alpha_{\mathrm{weak}} < \infty$ if and only if

$$
1 < \int_{\Gamma_Y} [\mu(y)/\eta(y)]^2 \eta(y)\,\mathrm{d}y
$$

$$
= \int_{\Gamma_Y^+} [\mu(y)/\eta(y)]^2 \eta(y)\,\mathrm{d}y + \int_{\Gamma_Y \backslash \Gamma_Y^+} [\mu(y)/\eta(y)]^2 \eta(y)\,\mathrm{d}y
$$

$$
\leq \int_{\Gamma_Y^+} [\mu(y)/\eta(y)]^2 \eta(y)\,\mathrm{d}y + \int_{\Gamma_Y \backslash \Gamma_Y^+} \eta(y)\,\mathrm{d}y \tag{73}
$$

$$
\leq \int_{\Gamma_Y^+} [\mu(y)/\eta(y)]^2 \eta(y)\,\mathrm{d}y + 1, \tag{74}
$$

where (73) is due to the fact that $\mu(y)/\eta(y) \leq 1$ for $y \in \Gamma_Y \backslash \Gamma_Y^+$, and (74) is due to (12). Using (74), we can conclude that $\int_{\Gamma_Y^+} [\mu(y)/\eta(y)]^2 \eta(y)\,\mathrm{d}y > 0$ and thus $\int_{\Gamma_Y^+} \eta(y)\,\mathrm{d}y > 0$.

B. Proof of (28)

We note that

$$
\inf_{y \geq 0} \frac{\mu(y)}{\eta(y)} = \min\left\{ \inf_{y > 0} \frac{\mu(y)}{\eta(y)}, \frac{\mu_0}{\eta_0} \right\}.
$$

By construction, η_0, μ_0 as defined in (26) and (27) are both positive. Thus, to show (28), we just need to prove that $\inf_{y>0} \frac{\mu(y)}{\eta(y)} > 0$. To that end, we note that, for $y > 0$,

$$
\frac{\mu(y)}{\eta(y)} = \sigma h(y/\sigma - \sigma), \tag{75}
$$

where $h(x) \overset{\mathrm{def}}{=} x + \frac{\Phi'(x)}{\Phi(x)}$, with $\Phi(x)$ and $\Phi'(x)$ denoting the CDF and the PDF of the standard normal distribution,

respectively. The function $h(x)$ is related to the inverse Mill's ratio. It is a strictly increasing function, and $h(x) > 0$ for all x. See, *e.g.*, [18] for a proof. It follows that

$$\inf_{y>0} \frac{\mu(y)}{\eta(y)} = \lim_{y\to0^+} \frac{\mu(y)}{\eta(y)} = \sigma h(-\sigma) > 0.$$

REFERENCES

[1] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Advances in Neural Information Processing Systems*, 2013, pp. 2796–2804.

[2] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *Information Theory, IEEE Transactions on*, vol. 61, no. 4, pp. 1985–2007, 2015.

[3] Y. Chen and E. J. Candes, "Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems," in *Advances in Neural Information Processing Systems*, 2015, pp. 739–747.

[4] G. Li, Y. Gu, and Y. M. Lu, "Phase retrieval using iterative projections: Dynamics in the large systems limit," in *Proc. Allerton Conference on Communication, Control and Computing.*, Monticello, IL, Oct 2015.

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

[6] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving Systems of Random Quadratic Equations via Truncated Amplitude Flow," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 773–794, Feb 2018.

[7] Y. Chi and Y. M. Lu, "Kaczmarz method for solving quadratic equations," *IEEE Signal Process. Lett.*, vol. 23, no. 9, 2016.

[8] J. Ma, J. Xu, and A. Maleki, "Optimization-based amp for phase retrieval: The impact of initialization and ℓ_2-regularization," *arXiv preprint arXiv:1801.01170*, 2018.

[9] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[10] E. J. Candes and X. Li, "Solving quadratic equations via PhaseLift when there are about as many equations as unknowns," *Foundations of Computational Mathematics*, vol. 14, no. 5, pp. 1017–1026, 2014.

[11] K. Jaganathan, S. Oymak, and B. Hassibi, "Sparse phase retrieval: Convex algorithms and limitations," in *IEEE International Symposium on Information Theory*, 2013, pp. 1022–1026.

[12] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1-2, pp. 47–81, 2015.

[13] K.-C. Li, "On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma," *J. Am. Stat. Assoc*, vol. 87, no. 420, pp. 1025–1039, 1992.

[14] Y. M. Lu and G. Li, "Phase transitions of spectral initialization for high-dimensional nonconvex estimation," *Information and Inference*, 2018. [Online]. Available: https://arxiv.org/abs/1702.06435

[15] M. Mondelli and A. Montanari, "Fundamental limits of weak recovery with applications to phase retrieval," *arXiv:1708.05932*, 2018.

[16] Z. D. Bai and J. Yao, "On sample eigenvalues in a generalized spiked population model," *Journal of Multivariate Analysis*, vol. 106, no. 167–177, 2012.

[17] D. G. Luenberger, *Optimization by vector space methods.* John Wiley & Sons, 1997.

[18] M. R. Sampford, "Some inequalities on Mill's ratio and related functions," *The Annals of Mathematical Statistics*, vol. 24, no. 1, pp. 130–132, 1953.