A Broader impacts and limitations.

In this section, we discuss the broader impacts and limitations of our work.

A.1 Broader impacts

Automated systems that use human feedback are being used in an increasing number of contexts, spanning everything from predicting user preferences to finetuning language models. It is important to ensure that such systems are as accurate as possible; this naturally requires humans to respond in an accurate and consistent manner. Using the perceptual adjustment query to collect data in such settings could lead to more expressive responses without heavy cognitive burdens on users. Furthermore, providing a user the additional context of a continuous spectrum of items may result in more self-consistent responses. The downstream effects of collecting more expressive and self-consistent human responses could lead to improved models or entirely new paradigms of model development for a myriad of problem settings. With these advantages come associated risks as well. Due to how expressive the responses to PAQs are, the effects of adversarial responses may be magnified. That is, if an adversary purposely chooses to respond in an antagonistic manner, models trained with PAQs may be trained poorly or in opposition to the stated goal. Mitigating such effects likely requires a holistic approach from both the query design and robust model design perspectives.

A.2 Limitations

From a data collection perspective, the perceptual adjustment query requires access to a continuous space where each point corresponds to an item. In many applications, assuming access to this continuous space is reasonable. For example, if we use PAQs to characterize color blindness, then a natural continuous space is the RGB color space. In general, we situate our data collection within the latent space of a generative model, such as a GAN. While GANs are capable of producing extremely high fidelity images, these images are not always free of semantically meaningful artifacts. Our query design and modeling assumptions do not explicitly consider the case where a portion of the continuous spectrum of items may be corrupted. Furthermore, because our work is an initial exploration into low-rank matrix estimation from inverted measurements, we have not considered scenarios such as unbounded noise or heavier-tailed sensing vectors, and we have not established information-theoretic lower bounds for the inverted measurement paradigm. We hope that further exploration of the inverted measurement paradigm will lead to a rich line of follow-up work.

B Preliminaries and Notation

In this section, we provide an overview of the key tools that are utilized in our proofs. We first introduce notation which is used throughout our proofs.

Notation. For two real numbers a and b, let $a \wedge b = \min\{a, b\}$. Given a vector $\boldsymbol{x} \in \mathbb{R}^d$, denote $\|\boldsymbol{x}\|_1$ and $\|\boldsymbol{x}\|_2$ as the ℓ_1 and ℓ_2 norm, respectively. Denote $\mathcal{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ to be the set of vectors with unit ℓ_2 norm. Given a matrix $\boldsymbol{A} \in \mathbb{R}^{d_1 \times d_2}$, denote $\|\boldsymbol{A}\|_F$, $\|\boldsymbol{A}\|_*$, and $\|\boldsymbol{A}\|_{\mathrm{op}}$ as the Frobenius norm, nuclear norm, and operator norm, respectively. Denote $\mathbb{S}^{d \times d} = \{\boldsymbol{A} \in \mathbb{R}^{d \times d} : \boldsymbol{A} = \boldsymbol{A}^\top\}$ to be the set of symmetric $d \times d$ matrices. Denote $\boldsymbol{A} \succeq \boldsymbol{0}$ to mean $\boldsymbol{A}$ is symmetric positive semi-definite. For $\boldsymbol{A} \succeq \boldsymbol{0}$, define the (pseudo-) inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{A}} = \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{y}$ and the associated (pseudo-) norm $\|\boldsymbol{x}\|_{\boldsymbol{A}} = \sqrt{\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x}}$. For matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{d_1 \times d_2}$, denote $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \operatorname{tr}\left(\boldsymbol{A}^\top \boldsymbol{B}\right)$ as the Frobenius inner product.

We use the notation $f(x) \lesssim g(x)$ to denote that there exists some universal positive constant $c > 0$, such that $f(x) \le c \cdot g(x)$, and use the notation $f(x) \gtrsim g(x)$ when $g(x) \lesssim f(x)$.

We define random matrices

$$\bar{\boldsymbol{A}} = \bar{\gamma}^2 \boldsymbol{a}\boldsymbol{a}^\top = \frac{y + \bar{\eta}}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \boldsymbol{a}\boldsymbol{a}^\top \tag{13}$$

and

$$\widetilde{\boldsymbol{A}} = \widetilde{\gamma}^2 \boldsymbol{a}\boldsymbol{a}^\top = \left(\frac{y + \bar{\eta}}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \wedge \tau\right) \boldsymbol{a}\boldsymbol{a}^\top \tag{14}$$

as the sensing matrix formed with the m-averaged responses $\bar{\gamma}$ and truncated responses $\widetilde{\gamma}$, respectively.

B.1 Inverted measurement sensing matrices result in estimation bias.

Recall from Equation (4) that the random sensing matrix $\boldsymbol{A}^{\mathsf{inv}}$ takes the form

$$\boldsymbol{A}^{\mathsf{inv}} = \frac{y+\eta}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \boldsymbol{a}\boldsymbol{a}^\top. \tag{15}$$

Standard trace regression analysis requires that the bias term $\mathbb{E}\left[\eta \boldsymbol{A}\right] = \mathbf{0}$, typically by assuming (at least) that η is zero-mean conditioned on the sensing matrix $\boldsymbol{A}$. The following lemma shows that the bias term associated with the inverted measurements sensing matrix $\boldsymbol{A}^{\mathsf{inv}}$ is nonzero, resulting in biased estimation

Lemma 1. *Let $\boldsymbol{A}^{\mathsf{inv}}$ be the random matrix defined in Eq. (4) and η be the measurement noise. Then,*

$$\mathbb{E}\left[\eta \boldsymbol{A}^{\mathsf{inv}}\right] \neq \mathbf{0}. \tag{16}$$

The proof of Lemma 1 is provided in Appendix B.6.1. As a result, utilizing established low-rank matrix estimators will result in biased estimation.

B.2 Sub-exponential random variables.

Our analysis will depend on sub-exponential random variables, a class of random variables with heavier tails than Gaussian. While many definitions of sub-exponential random variables exist (see, for example, [43, Chapter 2.7]), we will make use of one particular property, presented below.

If X is a sub-exponential random variable, then there exists some constant c (only dependent on the distribution underlying the random variable X) such that for all integers $p \geq 1$,

$$\left(\mathbb{E}|X|^p\right)^{1/p} \leq cp. \tag{17}$$

B.3 Bernstein's inequality.

A key ingredient in our proofs is the well-known Bernstein's inequality, which is a concentration inequality for sums of independent sub-exponential random variables.

Lemma 2 (Bernstein's inequality, adapted from [44, Theorem 2.10])**.** *Let $X_1, \ldots, X_n$ be independent real-valued random variables. Assume exist positive numbers u_1 and u_2 such that*

$$\mathbb{E}\left[X_i^2\right] \leq u_1 \quad \text{and} \quad \mathbb{E}\left[|X_i|^p\right] \leq \frac{p!}{2} u_1 u_2^{p-2} \text{ for all integers } p \geq 2, \tag{18a}$$

Then for all $t > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n \left(X_i - \mathbb{E}\left[X_i\right]\right)\right| \geq \sqrt{\frac{2u_1 t}{n}} + \frac{u_2 t}{n}\right) \leq 2\exp(-t). \tag{18b}$$

B.4 Moments of the ratios of quadratic forms.

Because the quadratic term $\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}$ appears in the denominator of our sensing matrices, our analysis depends on quantifying the moments of the ratios of quadratic forms. This is done in the following lemma.

Lemma 3. *There exists an absolute constant $c > 0$ such that the following is true. Let $\boldsymbol{a} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$, $\boldsymbol{\Sigma}^\star \in \mathbb{R}^{d\times d}$ be any PSD matrix with rank r, and $\boldsymbol{U} \in \mathbb{R}^{d\times d}$ be an arbitrary symmetric matrix.*

(a) Suppose that $r > 8$. Then we have

$$\mathbb{E}\left(\frac{1}{\boldsymbol{a}^T \boldsymbol{\Sigma}^\star \boldsymbol{a}}\right)^4 \leq \frac{c}{\sigma_r^4 r^4}.$$

(b) Suppose that $r > 2$. Then we have

$$\mathbb{E}\left(\frac{\boldsymbol{a}^\top \boldsymbol{U} \boldsymbol{a}}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}}\right) \leq \frac{c}{\sigma_r r}\|\boldsymbol{U}\|_*. \tag{19}$$

The proof of Lemma 3 is presented in Appendix B.6.2.

B.5 A fourth moment bound for $\bar{\gamma}^2$.

Throughout our analysis, we will utilize the fact that $\bar{\gamma}^2$ has a bounded fourth-moment. This bound is characterized in the following lemma.

Lemma 4. *Assume $r > 8$. Then the bound*

$$\mathbb{E}\left[\left(\bar{\gamma}^2\right)^4\right] \lesssim \left(\frac{y+\eta^{\uparrow}}{\sigma_r r}\right)^4 \tag{20}$$

holds, where σ_r is the smallest non-zero singular value of $\boldsymbol{\Sigma}^{\star}$.

The proof of Lemma 4 is presented in Appendix B.6.3. For notational simplicity of the proofs, we denote $M = c\left(\frac{y+\eta^{\uparrow}}{\sigma_r r}\right)^4$.

B.6 Proofs of preliminary lemmas

In this section, we present proofs for preliminary lemmas from Appendices B.1, B.4, and B.5.

B.6.1 Proof of Lemma 1

We show that $\mathbb{E}\left[\eta \boldsymbol{A}^{\mathsf{inv}}\right] \neq \mathbf{0}$. Using the independence of η and $\boldsymbol{a}$ and the assumption that η is zero mean, we have

$$\begin{aligned}
\mathbb{E}\left[\eta \boldsymbol{A}^{\mathsf{inv}}\right] &= \mathbb{E}\left[\frac{\eta(y+\eta)}{\boldsymbol{a}^{\top}\boldsymbol{\Sigma}^{\star}\boldsymbol{a}}\boldsymbol{a}\boldsymbol{a}^{\top}\right] && (21)\\
&= \mathbb{E}\left[\eta(y+\eta)\right]\mathbb{E}\left[\frac{1}{\boldsymbol{a}^{\top}\boldsymbol{\Sigma}^{\star}\boldsymbol{a}}\boldsymbol{a}\boldsymbol{a}^{\top}\right] && (22)\\
&= \nu_{\eta}^2\,\mathbb{E}\left[\frac{1}{\boldsymbol{a}^{\top}\boldsymbol{\Sigma}^{\star}\boldsymbol{a}}\boldsymbol{a}\boldsymbol{a}^{\top}\right]. && (23)
\end{aligned}$$

This expectation is non-zero, as the random matrix $\frac{1}{\boldsymbol{a}^{\top}\boldsymbol{\Sigma}^{\star}\boldsymbol{a}}\boldsymbol{a}\boldsymbol{a}^{\top}$ is symmetric-positive semidefinite. Therefore, we have $\mathbb{E}\left[\eta \boldsymbol{A}^{\mathsf{inv}}\right] \neq \mathbf{0}$, as desired.

B.6.2 Proof of Lemma 3

Without loss of generality, we assume that $\boldsymbol{\Sigma}^{\star}$ is diagonal for the remainder of this proof. To prove each part of Lemma 3, we utilize results on the moments of ratios of quadratic forms. For non-negative integers p and q, we first verify that the mixed moment $\mathbb{E}\left[\frac{(\boldsymbol{a}^{\top}\boldsymbol{U}\boldsymbol{a})^p}{(\boldsymbol{a}^{\top}\boldsymbol{\Sigma}^{\star}\boldsymbol{a})^q}\right]$ exists. By [45, Proposition 1], the mixed moment exists if $\frac{r}{2} > q$. This is assumed to be true for all parts of Lemma 3.

By [45, Proposition 2], we have the following expression for the mixed moment $\mathbb{E}\left[\frac{(\boldsymbol{a}^{\top}\boldsymbol{U}\boldsymbol{a})^p}{(\boldsymbol{a}^{\top}\boldsymbol{\Sigma}^{\star}\boldsymbol{a})^q}\right]$:

$$\mathbb{E}\left[\frac{(\boldsymbol{a}^{\top}\boldsymbol{U}\boldsymbol{a})^p}{(\boldsymbol{a}^{\top}\boldsymbol{\Sigma}^{\star}\boldsymbol{a})^q}\right] = \frac{1}{\Gamma(q)}\int_0^{\infty} t^{q-1}\left|\boldsymbol{\Delta}_t\right|\mathbb{E}\left[\left(\boldsymbol{a}^{\top}\boldsymbol{\Delta}_t\boldsymbol{U}\boldsymbol{\Delta}_t\boldsymbol{a}\right)^p\right]dt, \tag{24}$$

where $\boldsymbol{\Delta}_t = (\boldsymbol{I}_d + 2t\boldsymbol{\Sigma}^{\star})^{-1/2}$ and $|\boldsymbol{\Delta}_t|$ is the determinant of $\boldsymbol{\Delta}_t$. Our results will depend on characterizing $|\boldsymbol{\Delta}_t|$. We begin by noting that $\boldsymbol{\Delta}_t$ is a diagonal matrix of the following form

$$\boldsymbol{\Delta}_t = \begin{bmatrix} \frac{1}{(1+2t\sigma_1)^{1/2}} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{(1+2t\sigma_r)^{1/2}} & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix}. \tag{25}$$

It follows that the determinant $|\mathbf{\Delta}_t|$ can be written as the product $|\mathbf{\Delta}_t| = \prod_{j=1}^{r} \frac{1}{(1+2t\sigma_r)^{1/2}}$. Furthermore, this product can be bounded as follows:

$$\frac{1}{(1+2t\sigma_1)^{r/2}} \le |\mathbf{\Delta}_t| \le \frac{1}{(1+2t\sigma_j)^{r/2}}. \tag{26}$$

We are now ready to prove parts (a) and (b).

Part (a). This case corresponds to the case where $p = 0$ and $q = 4$. Using the integral expression (24) and upper bound on determinant (26), with these values of p and q, we have

$$\mathbb{E}\left[\left(\frac{1}{\boldsymbol{a}^\top \mathbf{\Sigma}^\star \boldsymbol{a}}\right)^4\right] = \frac{1}{\Gamma(4)} \int_0^\infty t^3 \, |\mathbf{\Delta}_t| \, dt \tag{27}$$

$$\le \frac{1}{\Gamma(4)} \int_0^\infty t^3 \frac{1}{(1+2t\sigma_r)^{r/2}} dt \tag{28}$$

Making the substitution $s = 1 + 2t\sigma_r$, we can evaluate the integral as follows.

$$\mathbb{E}\left[\left(\frac{1}{\boldsymbol{a}^\top \mathbf{\Sigma}^\star \boldsymbol{a}}\right)^4\right] \le \frac{1}{2\Gamma(4)\sigma_r} \int_1^\infty \left(\frac{s-1}{2\sigma_r}\right)^3 \frac{1}{s^{r/2}} ds \tag{29}$$

$$\lesssim \frac{1}{\sigma_r^4} \int_1^\infty \frac{(s-1)^3}{s^{r/2}} ds \tag{30}$$

$$= \frac{1}{\sigma_r^4} \int_1^\infty \left(\frac{s^3}{s^{r/2}} - 3\frac{s^2}{s^{r/2}} + 3\frac{s}{s^{r/2}} - \frac{1}{s^{r/2}}\right) ds \tag{31}$$

$$= \frac{1}{\sigma_r^4} \left(\frac{2}{r-8} - \frac{6}{r-6} + \frac{6}{r-4} - \frac{2}{r-2}\right). \tag{32}$$

Therefore, we have that there exists some absolute constant c such that

$$\mathbb{E}\left[\left(\frac{1}{\boldsymbol{a}^\top \mathbf{\Sigma}^\star \boldsymbol{a}}\right)^4\right] \le \frac{c}{\sigma_r^4 r^4}, \tag{33}$$

as desired.

Part (b). This case corresponds to the case where $p = q = 1$. We begin again with the integral expression (24) and upper bound on determinant (26):

$$\mathbb{E}\left[\left(\frac{\boldsymbol{a}^\top \boldsymbol{U} \boldsymbol{a}}{\boldsymbol{a}^\top \mathbf{\Sigma}^\star \boldsymbol{a}}\right)\right] = \frac{1}{\Gamma(1)} \int_0^\infty |\mathbf{\Delta}_t| \, \mathbb{E}\left[\boldsymbol{a}^\top \mathbf{\Delta}_t \boldsymbol{U} \mathbf{\Delta}_t \boldsymbol{a}\right] dt \tag{34}$$

$$\le \frac{1}{\Gamma(1)} \int_0^\infty \frac{1}{(1+2t\sigma_r)^{r/2}} \mathbb{E}\left[\boldsymbol{a}^\top \mathbf{\Delta}_t \boldsymbol{U} \mathbf{\Delta}_t \boldsymbol{a}\right] dt \tag{35}$$

We now bound the expectation term $\mathbb{E}\left[\boldsymbol{a}^\top \mathbf{\Delta}_t \boldsymbol{U} \mathbf{\Delta}_t \boldsymbol{a}\right]$. Note that for $\boldsymbol{a} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$, the expectation $\mathbb{E}\left[\boldsymbol{a}^\top \boldsymbol{B} \boldsymbol{a}\right] = \operatorname{tr}(\boldsymbol{B})$ for any symmetric matrix $\boldsymbol{B}$. Therefore, we have

$$\mathbb{E}\left[\boldsymbol{a}^\top \mathbf{\Delta}_t \boldsymbol{U} \mathbf{\Delta}_t \boldsymbol{a}\right] = \operatorname{tr}(\mathbf{\Delta}_t \boldsymbol{U} \mathbf{\Delta}_t) \tag{36}$$

$$\le \|\mathbf{\Delta}_t \boldsymbol{U} \mathbf{\Delta}_t\|_*. \tag{37}$$

Above, we have used the fact that $\operatorname{tr}(\boldsymbol{B}) \le \|\boldsymbol{B}\|_*$ for any symmetric matrix $\boldsymbol{B}$. By Hölder's inequality for Schatten-p norms, we have that $\|\mathbf{\Delta}_t \boldsymbol{U} \mathbf{\Delta}_t\|_* \le \|\mathbf{\Delta}_t\|_{\mathrm{op}}^2 \|\boldsymbol{U}\|_*$. Because $\mathbf{\Delta}_t$ is diagonal and the entries of $\mathbf{\Delta}_t$ are bounded between 0 and 1, we can bound the operator norm as $\|\mathbf{\Delta}_t\|_{\text{op}} \leq 1$. Therefore

$$\mathbb{E}\left[\boldsymbol{a}^\top \mathbf{\Delta}_t \boldsymbol{U} \mathbf{\Delta}_t \boldsymbol{a}\right] \leq \|\boldsymbol{U}\|_* \tag{38}$$

Substituting this upper bound for the expectation term into the integral, we obtain

$$\mathbb{E}\left[\frac{\boldsymbol{a}^\top \boldsymbol{U} \boldsymbol{a}}{\boldsymbol{a}^\top \mathbf{\Sigma}^\star \boldsymbol{a}}\right] \leq \|\boldsymbol{U}\|_* \int_0^\infty \frac{1}{(1+2t\sigma_j)^{r/2}} dt. \tag{39}$$

Evaluating this integral, we have for some absolute constant c,

$$\mathbb{E}\left[\frac{\boldsymbol{a}^\top \boldsymbol{U} \boldsymbol{a}}{\boldsymbol{a}^\top \mathbf{\Sigma}^\star \boldsymbol{a}}\right] \leq \frac{c}{\sigma_r r} \|\boldsymbol{U}\|_*, \tag{40}$$

as desired.

B.6.3 Proof of Lemma 4

By the bounded noise assumption, $y + \bar{\eta} \leq y + \eta^\uparrow$. Therefore, we have

$$\mathbb{E}\left[\left(\bar{\gamma}^2\right)^4\right] = \mathbb{E}\left[\left(\frac{y+\bar{\eta}}{\boldsymbol{a}^\top \mathbf{\Sigma}^\star \boldsymbol{a}}\right)^4\right] \tag{41}$$

$$\leq (y+\bar{\eta})^4 \cdot \mathbb{E}\left[\left(\frac{1}{\boldsymbol{a}^\top \mathbf{\Sigma}^\star \boldsymbol{a}}\right)^4\right]. \tag{42}$$

It therefore suffices to bound the fourth moment of $\frac{1}{\boldsymbol{a}^\top \mathbf{\Sigma}^\star \boldsymbol{a}}$, which is done in Lemma 3. Therefore,

$$\mathbb{E}\left[\left(\frac{1}{\boldsymbol{a}^\top \mathbf{\Sigma}^\star \boldsymbol{a}}\right)^4\right] \lesssim \left(\frac{1}{\sigma_r r}\right)^4, \tag{43}$$

as desired.

C Proof of Theorem 1

Our goal is to derive finite sample error bounds for the estimator in Equation (8). For our estimator, if the regularization parameter is set to be sufficiently large (which we will characterize later), then the error matrix is guaranteed to be in some *error set* $\mathcal{E}$. For rank r symmetric positive semidefinite matrices, the error set $\mathcal{E}$ can be characterized as [24]

$$\mathcal{E} = \left\{\boldsymbol{U} \in \mathbb{S}^{d\times d} : \|\boldsymbol{U}\|_* \leq 4\sqrt{2r}\|\boldsymbol{U}\|_F\right\}, \tag{44}$$

where recall that $\mathbb{S}^{d\times d}$ denotes the set of symmetric $d \times d$ matrices.

A key condition for estimation under these settings is to ensure that the shrunken sensing matrices satisfy a restricted strong convexity (RSC) condition over the error set $\mathcal{E}$. That is, we must show that there exists some positive constant κ such that

$$\frac{1}{n}\sum_{i=1}^n \langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{U}\rangle^2 \geq \kappa \|\boldsymbol{U}\|_F^2 \qquad \text{for all } \boldsymbol{U} \in \mathcal{E}. \tag{45}$$

We begin by stating a proposition that characterizes the deterministic upper bound on the estimation error.

Proposition 1 ([39, Theorem 1] with $q = 0$)**.** *Suppose that $\mathbf{\Sigma}^\star$ has rank r and the shrunken sensing matrices satisfy the restricted strong convexity condition with positive constant κ. Then if the regularization parameter satisfies*

$$\lambda_n \geq 2\left\|\frac{1}{n}\sum_{i=1}^n y\widetilde{\boldsymbol{A}}_i - \frac{1}{n}\sum_{i=1}^n \langle \widetilde{\boldsymbol{A}}_i, \mathbf{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}_i\right\|_{\text{op}} \tag{46a}$$

any optimal solution $\widehat{\mathbf{\Sigma}}$ of the optimization program (8) satisfies

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}^\star\|_F \leq \frac{32\sqrt{r}\lambda_n}{\kappa} \tag{46b}$$

This theorem is a special case of Theorem 1 in [39], which is in turn adapted from Theorem 1 in [24] (see [24] or [39] for the proof). Proposition 1 is a deterministic and nonasymptotic result and provides a roadmap for proving upper bounds. First, we show that the operator norm (46a) can be upper bounded with high probability, allowing us to set the regularization parameter λ_n accordingly. Second, we show that the RSC condition (45) is satisfied with high probability. We begin by bounding the operator norm (46a) in the following proposition.

Proposition 2. *Let $y^{\uparrow} = y + \eta^{\uparrow}$. Suppose that $\mathbf{\Sigma}^{\star}$ has rank r, with $r > 8$. Then there exists a positive absolute constant C such that the bound*

$$\left\| \frac{1}{n} \sum_{i=1}^{n} y \widetilde{\boldsymbol{A}}_i - \frac{1}{n} \sum_{i=1}^{n} \langle \widetilde{\boldsymbol{A}}_i, \mathbf{\Sigma}^{\star} \rangle \widetilde{\boldsymbol{A}}_i \right\|_{\mathrm{op}} \leq C y^{\uparrow} \left(\frac{y^{\uparrow}}{\sigma_r r} \sqrt{\frac{d}{n}} + \frac{d}{n} \tau + \left(\frac{y^{\uparrow}}{\sigma_r r} \right)^2 \frac{1}{\tau} + \frac{1}{\sigma_r r} \frac{\nu_\eta^2}{m} \right) \tag{47}$$

holds with probability at least $1 - 4\exp(-d)$.

The proof of Proposition 2 is provided in Appendix C.1. Next, we show that the RSC condition (45) is satisfied with high probability, as is done in the following proposition.

Proposition 3. *Let κ_y be the median of $y + \bar{\eta}$ and let $\mathcal{E}$ be the error set defined in Eq. (44). Suppose that the truncation threshold τ satisfies $\tau \geq \frac{\kappa_y}{\mathrm{tr}(\mathbf{\Sigma}^{\star})}$. Then, there exist positive absolute constants $\kappa_{\mathcal{L}}$, c, and C such that if the number of effective measurements satisfy*

$$n \geq Crd \tag{48a}$$

then we have

$$\frac{1}{n} \sum_{i=1}^{n} \langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{U} \rangle^2 \geq \kappa_{\mathcal{L}} \left(\frac{\kappa_y}{\mathrm{tr}\,(\mathbf{\Sigma}^{\star})} \right)^2 \|\boldsymbol{U}\|_F^2 \tag{48b}$$

simultaneously for all matrices $\boldsymbol{U} \in \mathcal{E}$ with probability greater than $1 - \exp(-cn)$.

The proof of Proposition 3 is provided in Appendix C.2. We now utilize the results of Propositions 1, 2 and 3 to derive our final error bounds. By Proposition 2, we know that the operator norm (46a) can be upper bounded with high probability. We set the regularization parameter λ_n to satisfy

$$\lambda_n \geq C_1 y^{\uparrow} \left(\frac{y^{\uparrow}}{\sigma_r r} \sqrt{\frac{d}{n}} + \frac{d}{n} \tau + \left(\frac{y^{\uparrow}}{\sigma_r r} \right)^2 \frac{1}{\tau} + \frac{1}{\sigma_r r} \frac{\nu_\eta^2}{m} \right). \tag{49}$$

for an appropriate constant C_1. Furthermore, by Proposition 3, we have that there exists some universal constant C_2 such that if the number of effective measurements satisfies $n \geq C_2 rd$, the RSC condition also holds for constant $\kappa = \kappa_{\mathcal{L}} \left(\frac{\kappa_y}{\mathrm{tr}(\mathbf{\Sigma}^{\star})} \right)^2$ with high probability. Taking a union bound, we have that Proposition 2 and Proposition 3 hold simultaneously with probability at least $1 - 4\exp(-d) - \exp(-cn)$. By Proposition 1, the bound

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}^{\star}\|_F \leq 32\sqrt{r} \frac{\lambda_n}{\kappa_{\mathcal{L}} \left(\frac{\kappa_y}{\mathrm{tr}(\mathbf{\Sigma}^{\star})} \right)^2} \tag{50}$$

$$\leq C \left(\frac{\mathrm{tr}\,(\mathbf{\Sigma}^{\star})}{\kappa_y} \right)^2 \sqrt{r} \lambda_n \tag{51}$$

holds with probability at least $1 - 4\exp(-d) - \exp(-cn)$, as desired. Above, we have defined $C = \frac{32}{\kappa_{\mathcal{L}}}$.

C.1 Proof of Proposition 2.

Our goal is to derive an upper bound on the operator norm

$$\left\| \frac{1}{n} \sum_{i=1}^{n} y \widetilde{\boldsymbol{A}}_i - \frac{1}{n} \sum_{i=1}^{n} \langle \widetilde{\boldsymbol{A}}_i, \mathbf{\Sigma}^{\star} \rangle \widetilde{\boldsymbol{A}}_i \right\|_{\mathrm{op}}. \tag{52}$$

Step 1: decompose the error into five terms. We begin by adding and subtracting multiple quantities, as done below.

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} y\widetilde{\boldsymbol{A}}_i - \frac{1}{n}\sum_{i=1}^{n}\langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}_i &= \frac{1}{n}\sum_{i=1}^{n} y\widetilde{\boldsymbol{A}}_i - \mathbb{E}\left[y\widetilde{\boldsymbol{A}}\right] + \mathbb{E}\left[y\widetilde{\boldsymbol{A}}\right] - \mathbb{E}\left[y\bar{\boldsymbol{A}}\right] \\ &\quad + \mathbb{E}\left[y\bar{\boldsymbol{A}}\right] - \mathbb{E}\left[\langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}\right] + \mathbb{E}\left[\langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}\right] - \frac{1}{n}\sum_{i=1}^{n}\langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}_i \end{aligned} \tag{53}$$

$$\begin{aligned}&\overset{(i)}{=} \frac{1}{n}\sum_{i=1}^{n} y\widetilde{\boldsymbol{A}}_i - \mathbb{E}\left[y\widetilde{\boldsymbol{A}}\right] + \mathbb{E}\left[y\widetilde{\boldsymbol{A}}\right] - \mathbb{E}\left[y\bar{\boldsymbol{A}}\right] \\ &\quad + \mathbb{E}\left[\langle \bar{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \bar{\boldsymbol{A}}\right] - \mathbb{E}\left[\langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}\right] - \mathbb{E}\left[\bar{\eta}\bar{\boldsymbol{A}}\right] \\ &\quad + \mathbb{E}\left[\langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}\right] - \frac{1}{n}\sum_{i=1}^{n}\langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}_i. \end{aligned} \tag{54}$$

Above, (i) follows from substituting in $\langle \bar{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle - \bar{\eta}$ for y for the $\mathbb{E}\left[y\bar{\boldsymbol{A}}\right]$ term. To obtain our final bound, we bound the following operator norms.

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n} y\widetilde{\boldsymbol{A}}_i - \frac{1}{n}\sum_{i=1}^{n}\langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}_i\right\|_{\mathrm{op}} &\le \underbrace{y\left\|\frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{A}}_i - \mathbb{E}\left[\widetilde{\boldsymbol{A}}\right]\right\|_{\mathrm{op}}}_{\text{Term 1}} \\ &\quad + \underbrace{y\left\|\mathbb{E}\left[\widetilde{\boldsymbol{A}}\right] - \mathbb{E}\left[\bar{\boldsymbol{A}}\right]\right\|_{\mathrm{op}}}_{\text{Term 2}} \\ &\quad + \underbrace{\left\|\mathbb{E}\left[\langle \bar{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \bar{\boldsymbol{A}}\right] - \mathbb{E}\left[\langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}\right]\right\|_{\mathrm{op}}}_{\text{Term 3}} \\ &\quad + \underbrace{\left\|\mathbb{E}\left[\langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}\right] - \frac{1}{n}\sum_{i=1}^{n}\langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}_i\right\|_{\mathrm{op}}}_{\text{Term 4}} \\ &\quad + \underbrace{\left\|\mathbb{E}\left[\bar{\eta}\bar{\boldsymbol{A}}\right]\right\|_{\mathrm{op}}}_{\text{Term 5}}. \end{aligned} \tag{55}$$

In the remaining proof, we bound the five terms in (55) individually. We first discuss the nature of the five terms.

- **Terms 1 and 4:** These two terms characterize the difference between the empirical mean of quantities involving $\widetilde{\boldsymbol{A}}$ and their true expectation. In the proof, we show that the empirical mean concentrates around the expectation with high probability (see Lemma 5 and Lemma 8).
- **Terms 2 and 3:** These two terms characterize the difference in expectation introduced by truncating $\bar{\boldsymbol{A}}$ to $\widetilde{\boldsymbol{A}}$. Hence, these two terms characterize biases that arise from truncation. In the proof, these two terms diminish as $\tau \to \infty$ (see Lemma 6 and Lemma 7). Note that setting τ to ∞ is equivalent to no thresholding, and in this case $\widetilde{\boldsymbol{A}}$ becomes identical to $\bar{\boldsymbol{A}}$, and both terms diminish.
- **Term 5:** Term 5 is a bias term that arises from the fact that the mean of the noise η conditioned on sensing matrix $\bar{\boldsymbol{A}}$ is non-zero: $\mathbb{E}\left[\bar{\eta}|\bar{\boldsymbol{A}}\right] \neq 0$. We will show that this bias scales like $\frac{1}{m}$, allowing us to set the averaging number m to obtain consistent estimation.

By setting the truncation threshold τ carefully, we can make the Term 3 and 4 biases the same order as Terms 1 and 4.

Step 2: bound the five terms individually. In what follows, we provide five lemmas to bound each of the five terms individually. In the proofs of the five lemmas, we rely on an upper bound on the fourth moment of the m-sample averaged measurements $\bar{\gamma}^2$. Recall from Lemma 4 in Appendix B.5 that we have $\mathbb{E}[(\bar{\gamma}^2)^4] \le M = c\left(\frac{y+\eta^{\uparrow}}{\sigma_r r}\right)^4$. We also rely heavily on the following truncation properties relating the m-sample averaged measurements $\bar{\gamma}^2$ and truncated measurements $\tilde{\gamma}^2$:

$$\tilde{\gamma}_i^2 \le \tau \tag{TP1}$$

$$\tilde{\gamma}_i^2 \le \bar{\gamma}_i^2 \tag{TP2}$$

$$\bar{\gamma}_i^2 - \tilde{\gamma}_i^2 = (\bar{\gamma}_i^2 - \tilde{\gamma}_i^2)\cdot \mathbf{1}\{\bar{\gamma}_i^2 \ge \tau\}. \tag{TP3}$$

The following lemma provides a bound for Term 1.

Lemma 5. *Let $\widetilde{\boldsymbol{A}}_1, \ldots, \widetilde{\boldsymbol{A}}_n$ be i.i.d copies of a random matrix $\widetilde{\boldsymbol{A}}$ as defined in Eq. (14). There exists a universal constant $c > 0$ such that the following is true. For any $t > 0$, we have*

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{A}}_i - \mathbb{E}\left[\widetilde{\boldsymbol{A}}_i\right] \right\|_{\mathrm{op}} \lesssim \sqrt{\frac{M^{1/2}t}{n}} + \frac{\tau t}{n} \tag{56}$$

with probability at least $1 - 2\cdot 9^d \cdot \exp(-t)$.

The proof of Lemma 5 is provided in Appendix D.1. The next lemma provides a deterministic upper bound for Term 2.

Lemma 6. *Let $\bar{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{A}}$ be the random matrices defined in Eq. (13) and Eq. (14), respectively. Then the bound*

$$\left\| \mathbb{E}\left[\widetilde{\boldsymbol{A}}\right] - \mathbb{E}\left[\bar{\boldsymbol{A}}\right] \right\|_{\mathrm{op}} \lesssim \frac{M^{1/2}}{\tau} \tag{57}$$

holds.

The proof of Lemma 6 is provided in Appendix D.2. The following lemma provides a deterministic upper bound for Term 3.

Lemma 7. *Let $\bar{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{A}}$ be the random matrices defined in Eq. (13) and Eq. (14), respectively. Then the bound*

$$\left\| \mathbb{E}\left[\langle \bar{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \bar{\boldsymbol{A}}\right] - \mathbb{E}\left[\langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}\right] \right\|_{\mathrm{op}} \lesssim \frac{(y+\eta^{\uparrow})M^{1/2}}{\tau} \tag{58}$$

holds.

The proof of Lemma 7 is provided in Appendix D.3. The following lemma provides a bound for Term 4.

Lemma 8. *Let $\widetilde{\boldsymbol{A}}_1, \ldots, \widetilde{\boldsymbol{A}}_n$ be i.i.d copies of a random matrix $\widetilde{\boldsymbol{A}}$ as defined in Eq. (14). There exists a universal constant $c > 0$ such that the following is true. For any $t > 0$, we have*

$$\left\| \mathbb{E}\left[\langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}\right] - \frac{1}{n}\sum_{i=1}^{n}\langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}_i \right\|_{\mathrm{op}} \lesssim (y+\eta^{\uparrow})\left(\sqrt{\frac{M^{1/2}t}{n}} + \frac{\tau t}{n}\right) \tag{59}$$

with probability at least $1 - 2\cdot 9^d \cdot \exp(-t)$.

The proof of Lemma 8 is provided in Appendix D.4. We note that Terms 2 and 3 are bias that result from shrinkage, but crucially are inversely dependent on the shrinkage threshold τ. This fact allows us to set τ so that the order of Terms 2 and 3 match the order of Terms 1 and 4. In particular, with the choice of $\tau = M^{1/4}\sqrt{\frac{n}{d}}$, all terms are of order $M^{1/4}\sqrt{\frac{d}{n}}$.

The final lemma bounds Term 5, which is a bias that arises from the dependence of the sensing matrix $\bar{\boldsymbol{A}}$ on the noise η.

Lemma 9. *Let $\bar{\boldsymbol{A}}$ be the random matrix defined in Eq. (13). Suppose that $\boldsymbol{\Sigma}^\star$ has rank r with $r > 2$. Then we have*

$$\mathbb{E}\left[\left\|\bar{\eta}\bar{\boldsymbol{A}}\right\|_{\mathrm{op}}\right] \lesssim \frac{1}{\sigma_r r}\frac{\nu_\eta^2}{m}. \tag{60}$$

The proof of Lemma 9 is provided in Appendix D.5. We note that the bias scales with the variance of the m-sample averaged noise $\bar{\eta}$, which scales inversely with m.

Step 3: combine the five terms. We set $t = (\log 9 + 1)d$ and denote $y^\uparrow = y + \eta^\uparrow$. Utilizing Lemmas 5–9, we arrive at an upper bound for the operator norm. We have that with probability at least $1 - 4\exp(-d)$,

$$\left\| \frac{1}{n}\sum_{i=1}^n y\widetilde{\boldsymbol{A}}_i - \frac{1}{n}\sum_{i=1}^n \langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}_i \right\|_{\mathrm{op}} \lesssim (y^\uparrow + 1)\left(\sqrt{\frac{M^{1/2}d}{n}} + \frac{d}{n}\tau + \frac{M^{1/2}}{\tau}\right) + \frac{1}{\sigma_r r}\frac{\nu_\eta^2}{m} \tag{61}$$

$$\overset{(i)}{\lesssim} y^\uparrow \left(\frac{y^\uparrow}{\sigma_r r}\sqrt{\frac{d}{n}} + \frac{d}{n}\tau + \left(\frac{y^\uparrow}{\sigma_r r}\right)^2 \frac{1}{\tau} + \frac{1}{\sigma_r r}\frac{\nu_\eta^2}{m}\right) \tag{62}$$

as desired. Above, (i) follows from substituting in the expression for M from Lemma 4.

C.2 Proof of Proposition 3

Our objective is to show that the there exists some constant κ such that the RSC condition

$$\frac{1}{n}\sum_{i=1}^n \langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{U}\rangle^2 \geq \kappa \|\boldsymbol{U}\|_F^2 \tag{63}$$

holds uniformly for all matrices $\boldsymbol{U}$ in the error set

$$\mathcal{E} = \left\{ \boldsymbol{U} \in \mathbb{S}^{d\times d} : \|\boldsymbol{U}\|_* \leq 4\sqrt{2r}\|\boldsymbol{U}\|_F \right\}. \tag{64}$$

Recall from the definition of $\widetilde{\boldsymbol{A}}$ that

$$\widetilde{\boldsymbol{A}}_i = \widetilde{\gamma}_i^2 \boldsymbol{a}_i\boldsymbol{a}_i^\top \tag{65}$$

$$= \left(\frac{y + \bar{\eta}_i}{\boldsymbol{a}_i^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}_i} \wedge \tau\right)\boldsymbol{a}_i\boldsymbol{a}_i^\top \tag{66}$$

so we have

$$\langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{U}\rangle^2 = \left(\frac{y + \bar{\eta}_i}{\boldsymbol{a}_i^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}_i} \wedge \tau\right)^4 \left(\boldsymbol{a}_i^\top \boldsymbol{U}\boldsymbol{a}_i\right)^2.$$

This implies that $\sum_{i=1}^n \langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{U}\rangle^2$ is nondecreasing in τ when $\tau > 0$. As a result, defining the random matrix

$$\widetilde{\boldsymbol{A}}^{\tau'} = \left(\frac{y + \bar{\eta}}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \wedge \tau'\right)\boldsymbol{a}\boldsymbol{a}^\top, \tag{67}$$

we have that the following lower bound holds for any $\tau' \leq \tau$.

$$\frac{1}{n}\sum_{i=1}^n \langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{U}\rangle^2 \geq \frac{1}{n}\sum_{i=1}^n \langle \widetilde{\boldsymbol{A}}_i^{\tau'}, \boldsymbol{U}\rangle^2. \tag{68}$$

Above $\widetilde{\boldsymbol{A}}_1^{\tau'}, \ldots, \widetilde{\boldsymbol{A}}_n^{\tau'}$ are i.i.d copies of the random matrix (67). We will lower bound $\frac{1}{n}\sum_{i=1}^n \langle \widetilde{\boldsymbol{A}}_i^{\tau'}, \boldsymbol{U}\rangle^2$ for an appropriate value of τ', which we will set later. To proceed, we will use a small-ball argument [46, 47] based on the following lemma.

Lemma 10 ([47, Proposition 5.1], adapted to our notation)**.** *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^{d\times d}$ be i.i.d. copies of a random matrix $\boldsymbol{X} \in \mathbb{R}^{d\times d}$. Let $E \subset \mathbb{R}^{d\times d}$. Let $\xi, Q > 0$ be such that for every $\boldsymbol{U} \in E$,*

$$\mathbb{P}\left(|\langle \boldsymbol{X}, \boldsymbol{U}\rangle| \geq 2\xi\right) \geq Q. \tag{69}$$

Furthermore, denote the Rademacher width as

$$W = \mathbb{E}\left[\sup_{\boldsymbol{U}\in E} \frac{1}{n}\sum_{i=1}^n \varepsilon_i \langle \boldsymbol{X}_i, \boldsymbol{U}\rangle\right],$$

where $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. Rademacher random variables independent of the $\boldsymbol{X}_i$'s. Then, for any $t > 0$, with probability at least $1 - \exp\left(-\frac{nt^2}{2}\right)$,

$$\inf_{\boldsymbol{U} \in E} \left(\frac{1}{n} \sum_{i=1}^{n} \langle \boldsymbol{X}_i, \boldsymbol{U} \rangle^2 \right)^{1/2} \geq \xi(Q - t) - 2W.$$

We apply Lemma 10 with $\boldsymbol{X}_i = \widetilde{\boldsymbol{A}}_i^{\tau'}$ and with set E as

$$E = \mathcal{E} \cap \{\boldsymbol{U} \in \mathbb{R}^{d \times d} : \|\boldsymbol{U}\|_F = 1\} \tag{70}$$

$$= \{\boldsymbol{U} \in \mathbb{S}^{d \times d} : \|\boldsymbol{U}\|_F = 1, \|\boldsymbol{U}\|_* \leq 4\sqrt{2r}\} \tag{71}$$

The rest of the proof is comprised of two key steps. To invoke Lemma 10, the first step establishes the inequality (69) by lower bounding Q. The second step upper bounds the Rademacher width W. The following lemma provides the lower bound on Q.

Lemma 11. *Consider any $\tau' \in (0, \tau)$. There exist absolute constants $c_1, c_2 > 0$ such that for every $\boldsymbol{U} \in E$, we have*

$$\mathbb{P}\left(\left| \langle \widetilde{\boldsymbol{A}}^{\tau'}, \boldsymbol{U} \rangle \right| \geq c_1 \left(\frac{\kappa_y}{\operatorname{tr}(\boldsymbol{\Sigma}^\star)} \wedge \tau' \right) \right) \geq c_2. \tag{72}$$

The proof of Lemma 11 is presented in Appendix E.1. We now turn to the second step of the proof, which is bounding the Rademacher width W. The next lemma characterizes this width.

Lemma 12. *Consider any $\tau' \in (0, \tau)$. Let $\widetilde{\boldsymbol{A}}_1^{\tau'}, \dots, \widetilde{\boldsymbol{A}}_n^{\tau'} \in \mathbb{R}^{d \times d}$ be i.i.d. copies of the random matrix $\widetilde{\boldsymbol{A}}^{\tau'} \in \mathbb{R}^{d \times d}$ defined in Equation (67). Let E be the set defined in Equation (71). Then, there exists some absolute constants c_1 and c_2 such that if*

$$n \geq c_1 d \tag{73a}$$

the bound

$$\mathbb{E}\left[\sup_{\boldsymbol{U} \in E} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \langle \widetilde{\boldsymbol{A}}_i^{\tau'}, \boldsymbol{U} \rangle \right] \leq c_2 \tau' \sqrt{\frac{rd}{n}} \tag{73b}$$

holds.

The proof of Lemma 12 is presented in Appendix E.2. Invoking Lemma 11 and Lemma 12, we have that for some constant c_4, as long as $n \geq c_4 d$, the bound

$$\begin{aligned}\inf_{\boldsymbol{U} \in E} \left(\frac{1}{n} \sum_{i=1}^{n} \langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{U} \rangle^2 \right)^{1/2} &\geq \inf_{\boldsymbol{U} \in E} \left(\frac{1}{n} \sum_{i=1}^{n} \langle \widetilde{\boldsymbol{A}}_i^{\tau'}, \boldsymbol{U} \rangle^2 \right)^{1/2} \\ &\geq c_1' \left(\frac{\kappa_y}{\operatorname{tr}(\boldsymbol{\Sigma}^\star)} \wedge \tau' \right) (c_2 - t) - c_3 \tau' \sqrt{\frac{rd}{n}} \end{aligned} \tag{74}$$

with probability at least $1 - \exp\left(-\frac{nt^2}{2}\right)$. We set $\tau' = \frac{\kappa_y}{\operatorname{tr}(\boldsymbol{\Sigma}^\star)}$, where κ_y is the median of the random quantity $y + \bar{\eta}$. By the assumption $\tau \geq \frac{\kappa_y}{\operatorname{tr}(\boldsymbol{\Sigma}^\star)}$, this choice of τ' satisfies $\tau' \leq \tau$. Setting $t = \frac{c_2}{2}$, we have for some constant c, that with probability at least $1 - \exp(-cn)$,

$$\inf_{\boldsymbol{U} \in E} \frac{1}{n} \left(\sum_{i=1}^{n} \langle \widetilde{\boldsymbol{A}}_i^{\tau'}, \boldsymbol{U} \rangle^2 \right)^{1/2} \geq \frac{c_1' c_2}{2} \frac{\kappa_y}{\operatorname{tr}(\boldsymbol{\Sigma}^\star)} - c_3 \frac{\kappa_y}{\operatorname{tr}(\boldsymbol{\Sigma}^\star)} \sqrt{\frac{rd}{n}}. \tag{75}$$

Therefore, if $n \geq \max\left\{ \left(\frac{4c_3}{c_1' c_2} \right)^2, c_4 \right\} rd$, we have

$$\inf_{\boldsymbol{U} \in \mathcal{E}} \frac{1}{n} \sum_{i=1}^{n} \langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{U} \rangle^2 \geq \left(\frac{c_1' c_2}{4} \frac{\kappa_y}{\operatorname{tr}(\boldsymbol{\Sigma}^\star)} \right)^2 \|\boldsymbol{U}\|_F^2 \tag{76}$$

with probability at least $1 - \exp(-cn)$. We conclude by setting $\kappa_{\mathcal{L}} = \left(\frac{c_1' c_2}{4} \right)^2$ and $C = \max\left\{ \left(\frac{4c_3}{c_1' c_2} \right)^2, c_4 \right\}$.

D Proof of supporting lemmas for Proposition 2

In this section, we prove the supporting lemmas for Proposition 2.

D.1 Proof of Lemma 5.

Let $\mathcal{S}_{\frac{1}{4}} \subseteq \mathcal{S}^{d-1}$ be a $\frac{1}{4}$-covering of unit-norm d-dimensional vectors. By a covering argument [43, Exercise 4.4.3], for any symmetric matrix $\boldsymbol{U} \in \mathbb{R}^{d\times d}$, its operator norm is bounded by $\|\boldsymbol{U}\|_{\text{op}} \leq 2\sup_{\boldsymbol{v}\in\mathcal{S}_{\frac{1}{4}}} \left|\boldsymbol{v}^\top \boldsymbol{U}\boldsymbol{v}\right|$. Hence, we have

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{A}}_i - \mathbb{E}\left[\widetilde{\boldsymbol{A}}\right]\right\|_{\text{op}} &\leq 2\sup_{\boldsymbol{v}\in\mathcal{S}_{\frac{1}{4}}} \left|\boldsymbol{v}^\top\left(\frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{A}}_i - \mathbb{E}\left[\widetilde{\boldsymbol{A}}\right]\right)\boldsymbol{v}\right| \\ &= 2\sup_{\boldsymbol{v}\in\mathcal{S}_{\frac{1}{4}}} \left|\frac{1}{n}\sum_{i=1}^n \boldsymbol{v}^\top\widetilde{\boldsymbol{A}}_i\boldsymbol{v} - \mathbb{E}\left[\boldsymbol{v}^\top\widetilde{\boldsymbol{A}}\boldsymbol{v}\right]\right|. \end{aligned} \tag{77}$$

We now apply Bernstein's inequality to bound (77). We first assume the Bernstein condition holds with $u_1 = c_1 M^{\frac{1}{2}}$ and $u_2 = c_2\tau$ for some universal positive constants c_1, c_2. Namely, for each integer $p \geq 2$, we show that for any unit vector $\boldsymbol{v} \in \mathbb{R}^d$,

$$\mathbb{E}\left[\left|\boldsymbol{v}^\top\widetilde{\boldsymbol{A}}\boldsymbol{v}\right|^p\right] \leq \frac{p!}{2}u_1 u_2^{p-2}. \tag{78}$$

We first provide the rest of the proof assuming that (78) holds, followed by proving (78). By Bernstein's inequality (see Lemma 2), under condition (78), we have that for any unit vector $\boldsymbol{v} \in \mathbb{R}^d$ and any $t > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n \boldsymbol{v}^\top\widetilde{\boldsymbol{A}}_i\boldsymbol{v} - \mathbb{E}\left[\boldsymbol{v}^\top\widetilde{\boldsymbol{A}}\boldsymbol{v}\right]\right| \geq 2\left(\sqrt{\frac{u_1 M^{1/2} t}{n}} + \frac{u_2\tau t}{n}\right)\right) \leq 2\exp\left(-t\right). \tag{79}$$

By Vershynin [43, Corollary 4.2.13], the cardinality of the covering set $\mathcal{S}_{\frac{1}{4}}$ is bounded above by 9^d. Therefore, by a union bound,

$$\mathbb{P}\left(\sup_{\boldsymbol{v}\in\mathcal{S}_{\frac{1}{4}}}\left|\frac{1}{n}\sum_{i=1}^n \boldsymbol{v}^\top\widetilde{\boldsymbol{A}}_i\boldsymbol{v} - \mathbb{E}\left[\boldsymbol{v}^\top\widetilde{\boldsymbol{A}}\boldsymbol{v}\right]\right| \geq 2\left(\sqrt{\frac{u_1 M^{1/2} t}{n}} + \frac{u_2\tau t}{n}\right)\right) \leq 2\cdot 9^d\cdot\exp\left(-t\right). \tag{80}$$

Combining (77) and (80), for any $t > 0$, we have

$$\left\|\frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{A}}_i - \mathbb{E}\left[\widetilde{\boldsymbol{A}}\right]\right\|_{\text{op}} \lesssim \sqrt{\frac{M^{1/2}t}{n}} + \frac{\tau t}{n} \tag{81}$$

with probability at least $1 - 2\cdot 9^d\cdot\exp\left(-t\right)$, as desired. It remains to prove the Bernstein condition (78).

Proving the Bernstein condition (78) holds. We fix any unit vector $\boldsymbol{v} \in \mathbb{R}^d$. Plugging in $\widetilde{\boldsymbol{A}} = \widetilde{\gamma}^2\boldsymbol{a}\boldsymbol{a}^\top$, we have $\boldsymbol{v}^\top\widetilde{\boldsymbol{A}}\boldsymbol{v} = \widetilde{\gamma}^2(\boldsymbol{v}^\top\boldsymbol{a})^2$. Since the random variable $\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$, and $\boldsymbol{v}$ is a unit vector, it follows that $\boldsymbol{v}^\top\boldsymbol{a} \sim \mathcal{N}(0,1)$. Denote by $G \sim \mathcal{N}(0,1)$ a standard normal random variable. For any integer $p \geq 2$, we have

$$\begin{aligned}\mathbb{E}[\left|\boldsymbol{v}^\top\widetilde{\boldsymbol{A}}\boldsymbol{v}\right|^p] = \mathbb{E}\left[(\widetilde{\gamma}^2 G^2)^p\right] &\overset{\text{(i)}}{\leq} \tau^{p-2}\mathbb{E}\left[\left(\widetilde{\gamma}^2\right)^2 G^{2p}\right] \\ &\overset{\text{(ii)}}{\leq} \tau^{p-2}\cdot\mathbb{E}\left[\left(\bar{\gamma}^2\right)^2 G^{2p}\right] \\ &\overset{\text{(iii)}}{\leq} \tau^{p-2}\left(\mathbb{E}\left[\left(\bar{\gamma}^2\right)^4\right]\cdot\mathbb{E}\left[G^{4p}\right]\right)^{1/2} \\ &\overset{\text{(iv)}}{\leq} \tau^{p-2}\left(M\cdot\mathbb{E}\left[G^{4p}\right]\right)^{1/2}, \end{aligned} \tag{82}$$

where steps (i) and (ii) follow from TP1 and TP2, respectively, step (iii) follows from Cauchy–Schwarz, and step (iv) follows from the upper bound on the fourth moment of $\bar{\gamma}^2$. Note that G^2 follows a Chi-Square distribution with 1 degree of freedom, and hence sub-exponential. Recall from (17) in Appendix B.2 that there exists some constant $c > 0$ such that we have $\left(\mathbb{E}\left[(G^2)^p\right]\right)^{1/p} \leq cp$ for all $p \geq 1$. Hence, we have

$$\begin{aligned}\left(\mathbb{E}\left[G^{4p}\right]\right)^{1/2} &\leq (2cp)^p = (2ec)^p \cdot (\frac{p}{e})^p \\ &\overset{\text{(i)}}{<} p! \cdot (2ec)^p \end{aligned} \tag{83}$$

where step (i) is true by Stirling's inequality that for all $p \geq 1$,

$$p! > \sqrt{2\pi p}\left(\frac{p}{e}\right)^p e^{\frac{1}{12p+1}} > (\frac{p}{e})^p.$$

Plugging (83) to (82) and rearranging terms completes the proof of Bernstein condition (78).

D.2 Proof of Lemma 6.

We first begin by showing that $\mathbb{E}\left[\bar{\boldsymbol{A}}\right] - \mathbb{E}\left[\widetilde{\boldsymbol{A}}\right] \succeq \boldsymbol{0}$. Substituting in the definitions of $\bar{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{A}}$, we have $\mathbb{E}\left[\bar{\boldsymbol{A}}\right] - \mathbb{E}\left[\widetilde{\boldsymbol{A}}\right] = \mathbb{E}\left[(\bar{\gamma}^2 - \widetilde{\gamma}^2)\boldsymbol{a}\boldsymbol{a}^\top\right]$. By TP2, we have $\bar{\gamma}^2 \geq \widetilde{\gamma}^2$, meaning that $\bar{\gamma}^2 - \widetilde{\gamma}^2$ is non-negative. The expectation of a non-negative quantity times an outer product is symmetric positive semidefinite. Therefore, we can write the operator norm as

$$\left\|\mathbb{E}\left[\widetilde{\boldsymbol{A}}\right] - \mathbb{E}\left[\bar{\boldsymbol{A}}\right]\right\|_{\text{op}} = \sup_{\boldsymbol{v}\in\mathcal{S}^{d-1}} \boldsymbol{v}^\top\left(\mathbb{E}\left[\bar{\boldsymbol{A}}\right] - \mathbb{E}\left[\widetilde{\boldsymbol{A}}\right]\right)\boldsymbol{v}. \tag{84}$$

We now show that there exists a uniform upper bound on the quantity

$$\boldsymbol{v}^\top\left(\mathbb{E}\left[\bar{\boldsymbol{A}}\right] - \mathbb{E}\left[\widetilde{\boldsymbol{A}}\right]\right)\boldsymbol{v} \tag{85}$$

for a unit-norm d-dimensional vector $\boldsymbol{v}$, therefore bounding the operator norm. We again note $(\boldsymbol{v}^\top\boldsymbol{a}) \sim \mathcal{N}(0,1)$ and denote $G \sim \mathcal{N}(0,1)$. Then we rewrite the quantity (85) by substituting in the expression for sensing matrices $\bar{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{A}}$, as follows.

$$\begin{aligned}\boldsymbol{v}^\top\left(\mathbb{E}\left[\bar{\boldsymbol{A}}\right] - \mathbb{E}\left[\widetilde{\boldsymbol{A}}\right]\right)\boldsymbol{v} &= \mathbb{E}\left[\boldsymbol{v}^\top\left(\bar{\boldsymbol{A}} - \widetilde{\boldsymbol{A}}\right)\boldsymbol{v}\right] && (86)\\ &= \mathbb{E}\left[\boldsymbol{v}^\top\left(\bar{\gamma}^2\boldsymbol{a}\boldsymbol{a}^\top - \widetilde{\gamma}^2\boldsymbol{a}\boldsymbol{a}^\top\right)\boldsymbol{v}\right] && (87)\\ &= \mathbb{E}\left[\left(\bar{\gamma}^2 - \widetilde{\gamma}^2\right)G^2\right]. && (88)\end{aligned}$$

We continue from here by utilizing properties of the shrunken measurements, as follows.

$$\begin{aligned}\mathbb{E}\left[\left(\bar{\gamma}^2 - \widetilde{\gamma}^2\right)G^2\right] &\overset{\text{(i)}}{=} y\,\mathbb{E}\left[\left(\bar{\gamma}^2 - \widetilde{\gamma}^2\right)G^2\mathbf{1}\{\bar{\gamma}^2 \geq \tau\}\right] && (89)\\ &\leq \mathbb{E}\left[\bar{\gamma}^2G^2\mathbf{1}\{\bar{\gamma}^2 \geq \tau\}\right] && (90)\\ &\overset{\text{(ii)}}{\leq} \left(\mathbb{E}\left[(\bar{\gamma}^2G^2)^2\right]\cdot\mathbb{E}\left[\mathbf{1}\{\bar{\gamma}^2 \geq \tau\}\right]\right)^{1/2} && (91)\\ &\overset{\text{(iii)}}{\leq} \left(\mathbb{E}\left[|\bar{\gamma}^2|^4\right]\cdot\mathbb{E}\left[|G^2|^4\right]\right)^{1/4}\left(\mathbb{P}\left(\bar{\gamma}^2 \geq \tau\right)\right)^{1/2}, && (92)\end{aligned}$$

where step (i) follows from TP3, and (ii) and (iii) follow from Cauchy–Schwarz. We proceed by bounding each of the above terms separately. First, recall from Lemma 4 in Appendix B.5 that the fourth moment $\mathbb{E}\left[|\bar{\gamma}^2|^4\right]$ is bounded above by M. Second, G^2 is a sub-exponential random variable. By Appendix B.2, we have that $\mathbb{E}\left[|G^2|^4\right]^{1/4} \leq c$ for some constant c. It remains to bound $\left(\mathbb{P}\left(\bar{\gamma}^2 \geq \tau\right)\right)^{1/2}$, which we do below.

$$\begin{aligned}\left(\mathbb{P}\left(\bar{\gamma}^2 \geq \tau\right)\right)^{1/2} &\overset{(iv)}{\leq} \left(\frac{\mathbb{E}\left[|\bar{\gamma}^2|^2\right]}{\tau^2}\right)^{1/2} && (93)\\ &\overset{(v)}{\leq} \frac{\left(\mathbb{E}\left[|\bar{\gamma}^2|^4\right]\right)^{1/4}}{\tau} && (94)\\ &\overset{(vi)}{\leq} \frac{M^{1/4}}{\tau}. && (95)\end{aligned}$$

Above, (iv) follows from Markov's inequality, (v) follows from Cauchy–Schwarz, and (vi) follows from the fourth moment bound on the averaged scaling $\bar{\gamma}^2$. Putting everything together, we have that the bound

$$\boldsymbol{v}^\top \left(\mathbb{E}\left[\bar{\boldsymbol{A}}\right] - \mathbb{E}\left[\widetilde{\boldsymbol{A}}\right]\right) \boldsymbol{v} \lesssim \frac{M^{1/2}}{\tau} \tag{96}$$

holds uniformly for all vectors $\boldsymbol{v} \in \mathcal{S}^{d-1}$. Therefore,

$$\left\|\mathbb{E}\left[\widetilde{\boldsymbol{A}}\right] - \mathbb{E}\left[\bar{\boldsymbol{A}}\right]\right\|_{\mathrm{op}} \lesssim \frac{M^{1/2}}{\tau}, \tag{97}$$

as desired

D.3 Proof of Lemma 7.

We begin by substituting in the definition $\bar{\boldsymbol{A}} = \bar{\gamma}^2 \boldsymbol{a}\boldsymbol{a}^\top$, the matrix $\langle \bar{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star \rangle \bar{\boldsymbol{A}}$ can be written as $\bar{\gamma}^4 \boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}$. Similarly, we can re-write the matrix $\langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star \rangle \widetilde{\boldsymbol{A}}$ as $\widetilde{\gamma}^4 \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right) \boldsymbol{a}\boldsymbol{a}^\top$. Therefore, our goal is to bound the operator norm

$$\left\|\left(\bar{\gamma}^4 - \widetilde{\gamma}^4\right) \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right) \boldsymbol{a}\boldsymbol{a}^\top\right\|_{\mathrm{op}} \tag{98}$$

We note that the matrix $\left(\bar{\gamma}^4 - \widetilde{\gamma}^4\right) \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right) \boldsymbol{a}\boldsymbol{a}^\top$ is symmetric positive semidefinite, as it is the product of a non-negative scalar $\left(\bar{\gamma}^4 - \widetilde{\gamma}^4\right) \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right)$ and an outer product. Similar to the proof of Lemma 6, we now show a uniform upper bound on the quantity $\boldsymbol{v}^\top \mathbb{E}\left[\left(\bar{\gamma}^4 - \widetilde{\gamma}^4\right) \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right) \boldsymbol{a}\boldsymbol{a}^\top\right] \boldsymbol{v}$ for arbitrary unit-norm d-dimensional vector $\boldsymbol{v}$.

Again, note that $\boldsymbol{v}^\top \boldsymbol{a} \sim \mathcal{N}(0,1)$ and denote $G \sim \mathcal{N}(0,1)$. We begin by substituting in the expressions for G.

$$\begin{aligned}
\boldsymbol{v}^\top \mathbb{E}\left[\left(\bar{\gamma}^4 - \widetilde{\gamma}^4\right) \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right) \boldsymbol{a}\boldsymbol{a}^\top\right] \boldsymbol{v} &= \mathbb{E}\left[\left(\bar{\gamma}^4 - \widetilde{\gamma}^4\right) \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right) \boldsymbol{v}^\top \boldsymbol{a}\boldsymbol{a}^\top \boldsymbol{v}\right] && (99)\\
&= \mathbb{E}\left[\left(\bar{\gamma}^4 - \widetilde{\gamma}^4\right) \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right) G^2\right] && (100)
\end{aligned}$$

Next, we can proceed by manipulating the $\bar{\gamma}^4 - \widetilde{\gamma}^4$ term to remove the term $\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}$, as follows.

$$\begin{aligned}
\mathbb{E}\left[\left(\bar{\gamma}^4 - \widetilde{\gamma}^4\right) \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right) G^2 \mathbf{1}\{\gamma^2 \geq \tau\}\right] &= \mathbb{E}\left[\left(\bar{\gamma}^2 + \widetilde{\gamma}^2\right) \left(\bar{\gamma}^2 - \widetilde{\gamma}^2\right) \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right) G^2\right] && (101)\\
&\overset{(i)}{\leq} \mathbb{E}\left[2\bar{\gamma}^2 \left(\bar{\gamma}^2 - \widetilde{\gamma}^2\right) \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right) G^2\right] && (102)\\
&\overset{(ii)}{=} 2\mathbb{E}\left[(y + \bar{\eta}) \left(\bar{\gamma}^2 - \widetilde{\gamma}^2\right) G^2\right] && (103)\\
&\overset{(iii)}{\leq} 2(y + \eta^\uparrow)\mathbb{E}\left[\left(\bar{\gamma}^2 - \widetilde{\gamma}^2\right) G^2 \mathbf{1}\{\gamma^2 \geq \tau\}\right] && (104)
\end{aligned}$$

Above, (i) follows from TP2, (ii) follows from the definition $\bar{\gamma}^2 = \frac{y + \bar{\eta}}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}}$, and (iii) follows from TP3 and the upper bound on noise η.

The rest of the proof follows the exact steps of the proof of Lemma 6, provided in Section D.2. Therefore, we have the bound

$$\left\|\mathbb{E}\left[\left(\bar{\gamma}^4 - \widetilde{\gamma}^4\right) \left(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right) \boldsymbol{a}\boldsymbol{a}^\top\right]\right\|_{\mathrm{op}} \lesssim \frac{(y + \eta^\uparrow) M^{1/2}}{\tau}, \tag{105}$$

as desired.

D.4 Proof of Lemma 8

The proof follows the steps as the proof of Lemma 5, and we explain the difference where we now provide a Bernstein condition with $u_1 = c_1(y + \eta^\uparrow)^2$ and $u_2 = c_2(y + \eta^\uparrow)\tau$. Namely, for every integer $p \geq 2$, we have (cf. 78)

$$\mathbb{E}\left[\left|\boldsymbol{v}^\top \langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star \rangle \widetilde{\boldsymbol{A}} \boldsymbol{v}\right|^p\right] \leq \frac{p!}{2} u_1 u_2^{p-2}. \tag{106}$$

Plugging in $\widetilde{\boldsymbol{A}} = \widetilde{\gamma}^2 \boldsymbol{a}\boldsymbol{a}^\top$, we have

$$\begin{aligned}
\mathbb{E}\left[\left|\boldsymbol{v}^\top \langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}\boldsymbol{v}\right|^p\right] &= \mathbb{E}\left[\left(\widetilde{\gamma}^2 \boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right)^p \cdot \left|\boldsymbol{v}^\top \widetilde{\boldsymbol{A}}\boldsymbol{v}\right|^p\right] \\
&\overset{\text{(i)}}{\leq} \mathbb{E}\left[\left(\bar{\gamma}^2 \boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}\right)^p \cdot \left|\boldsymbol{v}^\top \widetilde{\boldsymbol{A}}\boldsymbol{v}\right|^p\right] && (107) \\
&\overset{\text{(ii)}}{=} \mathbb{E}\left[(y+\bar{\eta})^p \cdot \left|\boldsymbol{v}^\top \widetilde{\boldsymbol{A}}\boldsymbol{v}\right|^p\right] && (108) \\
&\overset{\text{(iii)}}{\leq} (y+\eta^\uparrow)^p \cdot \mathbb{E}\left[\left|\boldsymbol{v}^\top \widetilde{\boldsymbol{A}}\boldsymbol{v}\right|^p\right]. && (109)
\end{aligned}$$

Plugging in (78) from Lemma 5 to bound the term $\mathbb{E}\left[\left|\boldsymbol{v}^\top \widetilde{\boldsymbol{A}}\boldsymbol{v}\right|^p\right]$ in (109) completes the proof of the Bernstein condition (106).

Above, (i) follows from TP2, (ii) follows from the definition $\bar{\gamma}^2 = \frac{y+\bar{\eta}}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}}$, and (iii) follows from the upper bound on the noise η. The rest of the proof follows in the same manner as the proof of Lemma 5, as presented in Section D.1, with an additional factor of $y + \eta^\uparrow$. Therefore, like in Section D.1, the bound

$$\left\|\mathbb{E}\left[\langle \widetilde{\boldsymbol{A}}, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}\right] - \frac{1}{n}\sum_{i=1}^n \langle \widetilde{\boldsymbol{A}}_i, \boldsymbol{\Sigma}^\star\rangle \widetilde{\boldsymbol{A}}_i\right\|_{\text{op}} \lesssim (y+\eta^\uparrow)\left(\sqrt{\frac{M^{1/2}t}{n}} + \frac{\tau t}{n}\right) \tag{110}$$

holds with probability greater than $1 - 2 \cdot 9^d \cdot \exp(-t)$, as desired.

D.5 Proof of Lemma 9.

Substituting in $\bar{\boldsymbol{A}} = \bar{\gamma}^2 \boldsymbol{a}\boldsymbol{a}^\top = \frac{y+\bar{\eta}}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \boldsymbol{a}\boldsymbol{a}^\top$, we have

$$\begin{aligned}
\left\|\mathbb{E}\left[\bar{\eta}\bar{\boldsymbol{A}}\right]\right\|_{\text{op}} &= \left\|\mathbb{E}\left[\bar{\eta}(y+\bar{\eta})\frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}}\right]\right\|_{\text{op}} \\
&= \left\|\mathbb{E}\left[\bar{\eta}(y+\bar{\eta})\right] \cdot \mathbb{E}\left[\frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}}\right]\right\|_{\text{op}} \\
&= \frac{\sigma_\eta^2}{m}\left\|\mathbb{E}\left[\frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}}\right]\right\|_{\text{op}}. && (111)
\end{aligned}$$

To bound the operator norm term in (111), recall from Lemma 3(b) in Appendix B.4 that for any matrix $\boldsymbol{U}$, we have

$$\mathbb{E}\left(\frac{\boldsymbol{a}^\top \boldsymbol{U}\boldsymbol{a}}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}}\right) \lesssim \frac{1}{\sigma_r r}\|\boldsymbol{U}\|_*. \tag{112}$$

Note that $\frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}}$ is symmetric positive semidefinite, so we have

$$\begin{aligned}
\left\|\mathbb{E}\left[\frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}}\right]\right\|_{\text{op}} &= \sup_{\boldsymbol{v}\in\mathcal{S}^{d-1}}\left|\boldsymbol{v}^\top \mathbb{E}\left[\frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}}\right]\boldsymbol{v}\right| \\
&= \sup_{\boldsymbol{v}\in\mathcal{S}^{d-1}} \mathbb{E}\left[\frac{\boldsymbol{a}^\top(\boldsymbol{v}\boldsymbol{v}^\top)\boldsymbol{a}}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}}\right] \\
&\overset{\text{(i)}}{\lesssim} \frac{1}{\sigma_r r}\sup_{\boldsymbol{v}\in\mathcal{S}^{d-1}}\|\boldsymbol{v}\boldsymbol{v}^\top\|_* \\
&\overset{\text{(ii)}}{=} \frac{1}{\sigma_r r}, && (113)
\end{aligned}$$

where step (i) is true by plugging in (112), and step (ii) is true because $\|\boldsymbol{v}\boldsymbol{v}^\top\|_* = 1$ for any unit norm vector $\boldsymbol{v}$. Plugging (113) back to (111), we have

$$\left\|\mathbb{E}\left[\bar{\eta}\bar{\boldsymbol{A}}\right]\right\|_{\text{op}} \lesssim \frac{1}{\sigma_r r} \cdot \frac{\nu_\eta^2}{m}, \tag{114}$$

as desired.

E Proof of supporting lemmas for Proposition 3

In this section, we prove the supporting lemmas for Proposition 3.

E.1 Proof of Lemma 11

For the proof, we first fix any $\boldsymbol{U} \in \mathcal{E} \cap \{\boldsymbol{U} \in \mathbb{S}^{d\times d} : \|\boldsymbol{U}\|_F = 1\}$. Let κ_y be the median of $y + \bar{\eta}$ and let $\mathcal{G}$ be the event that $y + \eta \geq \kappa_y$, which occurs with probability $\frac{1}{2}$. For any $\xi > 0$, because the averaged noise $\bar{\eta}$ and sensing vector $\boldsymbol{a}$ are independent,

$$\mathbb{P}\left(\left|\langle \widetilde{\boldsymbol{A}}^{\tau'}, \boldsymbol{U}\rangle\right| \geq \xi\right) = \mathbb{P}\left(\left(\frac{y+\eta}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \wedge \tau'\right) \left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right| \geq \xi\right) \tag{115}$$

$$= \mathbb{P}\left(\left(\frac{y+\eta}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \wedge \tau'\right) \left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right| \geq \xi \middle| \mathcal{G}\right) \mathbb{P}(\mathcal{G}) \tag{116}$$

$$= \frac{1}{2}\mathbb{P}\left(\left(\frac{y+\eta}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \wedge \tau'\right) \left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right| \geq \xi \middle| \mathcal{G}\right) \tag{117}$$

$$\geq \frac{1}{2}\mathbb{P}\left(\left(\frac{\kappa_y}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \wedge \tau'\right) \left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right| \geq \xi\right) \tag{118}$$

We proceed by bounding the terms in (118) separately.

Lower bound on $\mathbb{P}\left(\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right| \geq c_1\right)$. We use the approach from [32, Section 4.1]. By Paley-Zygmund inequality,

$$\mathbb{P}\left(\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right|^2 \geq \frac{1}{2}\mathbb{E}\left[\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right|^2\right]\right) \geq \frac{1}{4}\frac{\left(\mathbb{E}\left[\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right|^2\right]\right)^2}{\mathbb{E}\left[\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right|^4\right]} \tag{119}$$

As noted in [32, Section 4.1], there exists some constant c_2' such that for any matrix $\boldsymbol{U}$ with unit Frobenius norm,

$$\mathbb{E}\left[\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right|^2\right] \geq 1 \quad \text{and} \quad \mathbb{E}\left[\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right|^4\right] \leq c_2'\left(\mathbb{E}\left[\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right|^2\right]\right)^2. \tag{120}$$

Note that by the definition, every matrix $\boldsymbol{U} \in E$ has unit Frobenius norm. Utilizing Paley-Zygmund (119) and the bounds on the second and fourth moment of $\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle$ (120), there exist positive constants c_1 and c_2 such that

$$\mathbb{P}\left(\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right| \geq c_1\right) \geq c_2. \tag{121}$$

Upper bound on $\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}$. By Hanson-Wright inequality [48, Theorem 1.1], there exist some positive absolute constants c and c_3' such that for any $t > 0$, we have

$$\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a} \leq c_3'\left(\operatorname{tr}(\boldsymbol{\Sigma}^\star) + \|\boldsymbol{\Sigma}^\star\|_F\sqrt{t} + \|\boldsymbol{\Sigma}^\star\|_{\mathrm{op}}\, t\right) \tag{122}$$

with probability at least $1 - 2\exp(-ct)$. Set t to be a constant such that $2\exp(-ct) = \frac{c_2}{2}$ and note that for symmetric positive semidefinite matrix $\boldsymbol{\Sigma}^\star$, the bounds $\|\boldsymbol{\Sigma}^\star\|_F \leq \operatorname{tr}(\boldsymbol{\Sigma}^\star)$ and $\|\boldsymbol{\Sigma}^\star\|_{\mathrm{op}} \leq \operatorname{tr}(\boldsymbol{\Sigma}^\star)$ hold. As a result, we have that there exists some constant c_3 such that

$$\mathbb{P}\Big(\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a} \leq c_3 \operatorname{tr}(\boldsymbol{\Sigma}^\star)\Big) \geq 1 - \frac{c_2}{2}. \tag{123}$$

By a union bound of (121) and (123), we have

$$\begin{aligned}
&\mathbb{P}\left(\left(\frac{\kappa_y}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \wedge \tau'\right)\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right| \geq c_1\left(\frac{\kappa_y}{c_3 \operatorname{tr}(\boldsymbol{\Sigma}^\star)} \wedge \tau'\right)\right) \\
&\geq \mathbb{P}\left(\frac{\kappa_y}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \wedge \tau' \geq \frac{\kappa_y}{c_3 \operatorname{tr}(\boldsymbol{\Sigma}^\star)} \wedge \tau'\right) + \mathbb{P}\left(\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right| \geq c_1\right) - 1 \\
&\geq \mathbb{P}\left(\frac{\kappa_y}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}} \geq \frac{\kappa_y}{c_3 \operatorname{tr}(\boldsymbol{\Sigma}^\star)}\right) + \mathbb{P}\left(\left|\langle \boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{U}\rangle\right| \geq c_1\right) - 1 \geq \frac{c_2}{2}
\end{aligned} \tag{124}$$

Redefining constants c_1 and c_2 appropriately, we have

$$\mathbb{P}\left(\left|\langle \widetilde{\boldsymbol{A}}^{\tau'}, \boldsymbol{U}\rangle\right| \geq c_1 \left(\frac{\kappa_y}{\operatorname{tr}(\boldsymbol{\Sigma}^\star)} \wedge \tau'\right)\right) \geq c_2, \tag{125}$$

as desired.

E.2 Proof of Lemma 12

We begin by noting that for any matrix $\boldsymbol{U} \in E$,

$$\begin{aligned}\mathbb{E}\left[\sup_{\boldsymbol{U}\in E} \frac{1}{n}\sum_{i=1}^n \varepsilon_i \langle \widetilde{\boldsymbol{A}}_i^{\tau'}, \boldsymbol{U}\rangle\right] &\overset{\text{(i)}}{\leq} \mathbb{E}\left[\sup_{\boldsymbol{U}\in E} \left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i \widetilde{\boldsymbol{A}}_i^{\tau'}\right\|_{\text{op}} \|\boldsymbol{U}\|_*\right] \\ &\overset{\text{(ii)}}{\leq} 4\sqrt{2r}\,\mathbb{E}\left[\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i \widetilde{\boldsymbol{A}}_i^{\tau'}\right\|_{\text{op}}\right],\end{aligned} \tag{126}$$

where step (i) follows from Hölder's inequality, and step (ii) follows from the definition of the set E. It remains of the proof to bound the expected operator norm in (126). We do this with a trivial modification of the approaches in [49, Section 5.4.1], [47, Section 8.6], [32, Section 4.1] to accommodate the bounded term $\left(\frac{y+\bar{\eta}_i}{\boldsymbol{a}_i^\top \boldsymbol{\Sigma}^\star \boldsymbol{a}_i} \wedge \tau'\right)$ that appears in each of the matrices $\widetilde{\boldsymbol{A}}_i^{\tau'}$. As a result, there exist universal constants c_1 and c_2 such that if n satisfies $n \geq c_2 d$, then the bound

$$\mathbb{E}\left[\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i \widetilde{\boldsymbol{A}}_i^{\tau'}\right\|_{\text{op}}\right] \leq c_1 \tau' \sqrt{\frac{d}{n}} \tag{127}$$

holds. We conclude by re-defining c_1 appropriately.

F Proof of Corrollary 1

The proof consists of two steps. We first verify that the choices of the averaging parameter m and truncation threshold τ as

$$m = \left\lfloor \left(\frac{(\nu_\eta^2)^2 N}{d}\right)^{1/3} \right\rfloor \quad \text{and} \quad \tau = \frac{y^\uparrow}{\sigma_r r}\sqrt{\frac{N}{md}}, \tag{128}$$

satisfy the assumptions $n \gtrsim rd$ and $\tau \geq \frac{\kappa_y}{\operatorname{tr}(\boldsymbol{\Sigma}^\star)}$. We then invoke Theorem 1.

Verifying the condition on n. We have

$$\begin{aligned} n &= \frac{N}{m} \\ &\overset{\text{(i)}}{=} N\left(\frac{(\nu_\eta^2)^2 N}{d}\right)^{-1/3} \\ &= \left(N^2 \frac{d}{(\nu_\eta^2)^2}\right)^{1/3} && (129) \\ &\overset{\text{(ii)}}{\gtrsim} \left((\nu_\eta^2)^2 r^3 d^2 \frac{d}{(\nu_\eta^2)^2}\right)^{1/3} && (130) \\ &= rd, && (131)\end{aligned}$$

where step (i) is true by plugging in the choice of m from (128), and step (ii) is true by plugging in the assumption $N \gtrsim \nu_\eta^2\, r^{3/2} d$. Thus the condition $n \gtrsim rd$ of Theorem 1 is satisfied.

Verifying the condition on τ. For the term $\sqrt{\frac{N}{dm}}$ in the expression of τ in (128), note that, by the previous point, $\frac{N}{m} = n \gtrsim rd$ (with a constant that, WLOG and by necessity, is greater than 1). Thus $\sqrt{\frac{N}{dm}} \geq \sqrt{r} > 1$. Therefore, it suffices to verify that

$$\frac{y^{\uparrow}}{\sigma_r r} \geq \frac{\kappa_y}{\operatorname{tr}(\mathbf{\Sigma}^{\star})}. \tag{132}$$

By definition, we have $y^{\uparrow} \geq \kappa_y$. Furthermore, since $\mathbf{\Sigma}^{\star}$ is symmetric positive semidefinite, its eigenvalues are all non-negative and are the same as singular values, and hence $\sigma_r r \leq \operatorname{tr}(\mathbf{\Sigma}^{\star})$. Therefore, we have (132) holds, verifying the condition on τ.

Invoking Theorem 1. By setting λ_n to its lower bound in (9) and substituting in $n = N/m$ and our choice of τ from (128), we have

$$\lambda_n = C_1 \frac{(y^{\uparrow})^2}{\sigma_r r} \left(\sqrt{\frac{md}{N}} + \frac{\nu_\eta^2}{m} \right). \tag{133}$$

Substituting in our choice of m from (128), we have

$$\lambda_n = C_1 \frac{(y^{\uparrow})^2}{\sigma_r r} \left(\frac{\nu_\eta^2 d}{N} \right)^{1/3}. \tag{134}$$

Substituting this expression for λ_n into the error bound (10) and absorbing C_1 into the constant C, we have

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}^{\star}\|_F \leq C \left(\frac{\operatorname{tr}(\mathbf{\Sigma}^{\star})^2}{\sigma_r r} \right) \left(\frac{y^{\uparrow}}{\kappa_y} \right)^2 \sqrt{r} \left(\frac{\nu_\eta^2 d}{N} \right)^{1/3}. \tag{135}$$

Using the fact that $\operatorname{tr}(\mathbf{\Sigma}^{\star}) \leq \sigma_1 r$, we have

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}^{\star}\|_F \leq C \left(\frac{\sigma_1^2}{\sigma_r} \right) \left(\frac{y^{\uparrow}}{\kappa_y} \right)^2 r^{3/2} \left(\frac{\nu_\eta^2 d}{N} \right)^{1/3}, \tag{136}$$

as desired.

G Choice of value y

In this section, we discuss the scale-invariance of the learning from PAQs problem in more detail. Under the Mahalanobis model for human perception, there exists some ground truth metric—parameterized by $\mathbf{\Sigma}^{\star}$—that governs perception. Associated with $\mathbf{\Sigma}^{\star}$, is a (squared) distance $y_{\star}$ such that for any two items $\boldsymbol{x}$ and $\boldsymbol{x}' \in \mathbb{R}^d$, $\boldsymbol{x}, \boldsymbol{x}'$ are perceived to be similar if $\|\boldsymbol{x} - \boldsymbol{x}'\|^2_{\mathbf{\Sigma}^{\star}} < y_{\star}$ and dissimilar if $\|\boldsymbol{x} - \boldsymbol{x}'\|^2_{\mathbf{\Sigma}^{\star}} > y_{\star}$.

Our two-stage estimator for learning with PAQs assumes that the value of $y_{\star}$ is known, which practitioners are unlikely to know *a priori*. However, this is not an issue in practice due to *scale-invariance*. That is, for any constant $c > 0$, if we use $y = c y_{\star}$ in our estimation procedure, we will recover a scaled metric $c\mathbf{\Sigma}^{\star}$. Therefore, by to the scale-invariance of the problem, we may set y to any positive value without loss of generality. In the main paper, for ease of exposition, we assume that $\mathbf{\Sigma}^{\star}$ is the metric associated with the user's choice for y, and derive estimation error bounds for this metric.

References

[1] William L. Rankin and Joel W. Grube. A comparison of ranking and rating procedures for value system measurement. *European Journal of Social Psychology*, 10(3):233–246, 1980.

[2] Anne-Wil Harzing, Joyce Baldueza, Wilhelm Barner-Rasmussen, Cordula Barzantny, Anne Canabal, Anabella Davila, Alvaro Espejo, Rita Ferreira, Axele Giroud, Kathrin Koester, Yung-Kuei Liang, Audra Mockaitis, Michael J. Morley, Barbara Myloni, Joseph O.T. Odusanya, Sharon Leiba O'Sullivan, Ananda Kumar Palaniappan, Paulo Prochno, Srabani Roy Choudhury, Ayse Saka-Helmhout, Sununta Siengthai, Linda Viswat, Ayda Uzuncarsili Soydas, and Lena Zander. Rating versus ranking: What is the best way to reduce response and language bias in cross-national research? *International Business Review*, 18(4):417–432, 2009. ISSN 0969-5931.

[3] Georgios N. Yannakakis and John Hallam. Ranking vs. preference: A comparative study of self-reporting. In Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin, editors, *Affective Computing and Intelligent Interaction*, pages 437–446, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-24600-5.

[4] Nihar B. Shah, Joseph K Bradley, Abhay Parekh, Martin Wainwright, and Kannan Ramchandran. A case for ordinal peer-evaluation in MOOCs. In *NIPS Workshop on Data Driven Education*, 2013.

[5] Karthik Raman and Thorsten Joachims. Methods for ordinal peer grading. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, page 1037–1046, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450329569.

[6] Richard D. Goffin and James M. Olson. Is it all relative? Comparative judgments and the possible improvement of self-ratings and ratings of others. *Perspectives on Psychological Science*, 6(1):48–60, 2011.

[7] Nihar B. Shah, Sivaraman Balakrishnan, Joseph Bradley, Abhay Parekh, Kannan Ramch, ran, and Martin J. Wainwright. Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence. *Journal of Machine Learning Research*, 17(58):1–47, 2016.

[8] Jingyan Wang and Nihar B. Shah. Your 2 is my 1, your 3 is my 9: Handling arbitrary miscalibrations in ratings. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '19, page 864–872, Richland, SC, 2019. International Foundation for Autonomous Agents and Multiagent Systems.

[9] Dale Griffin and Lyle Brenner. *Perspectives on Probability Judgment Calibration*, chapter 9. Wiley-Blackwell, 2008. ISBN 9780470752937.

[10] Polina Harik, Brian Clauser, Irina Grabovsky, Ronald Nungester, David Swanson, and Ratna Nandakumar. An examination of rater drift within a generalizability theory framework. *Journal of Educational Measurement*, 46:43–58, 2009.

[11] Carol M. Myford and Edward W. Wolfe. Monitoring rater performance over time: A framework for detecting differential accuracy and differential scale category use. *Journal of Educational Measurement*, 46(4):371–389, 2009.

[12] Gregory Canal, Stefano Fenu, and Christopher Rozell. Active ordinal querying for tuplewise similarity learning. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI)*, pages 3332–3340. AAAI Press, 2020.

[13] Aurélien Bellet, Amaury Habrard, and Marc Sebban. Metric learning synthesis lectures on artificial intelligence and machine learning. *Morgan & Claypool Publishers, San Rafael*, 1:2, 2015.

[14] Yiming Ying, Kaizhu Huang, and Colin Campbell. Sparse metric learning via smooth optimization. *Advances in neural information processing systems*, 22, 2009.

[15] Wei Bian and Dacheng Tao. Constrained empirical risk minimization framework for distance metric learning. *IEEE transactions on neural networks and learning systems*, 23(8):1194–1205, 2012.

[16] Zheng-Chu Guo and Yiming Ying. Guaranteed classification via regularized similarity learning. *Neural computation*, 26(3):497–522, 2014.

[17] Aurélien Bellet and Amaury Habrard. Robustness and generalization for metric learning. *Neurocomputing*, 151:259–267, 2015.

[18] Blake Mason, Lalit Jain, and Robert Nowak. Learning low-dimensional metrics. *Advances in neural information processing systems*, 30, 2017.

[19] Austin Xu and Mark Davenport. Simultaneous preference and metric learning from paired comparisons. *Advances in Neural Information Processing Systems*, 33, 2020.

[20] Gregory Canal, Blake Mason, Ramya Korlakai Vinayak, and Robert D Nowak. One for all: Simultaneous metric and preference learning over multiple users. In *Advances in Neural Information Processing Systems*, 2022.

[21] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. 2014.

[22] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.

[23] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[24] Sahand Negahban and Martin J Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, 39(2):1069–1097, 2011.

[25] A Tsybakov and A Rohde. Estimation of high-dimensional low-rank matrices. *Annals of Statistics*, 39(2):887–930, 2011.

[26] Emmanuel J Candes and Yaniv Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory*, 57(4):2342–2359, 2011.

[27] Sahand N Negahban, Pradeep Ravikumar, Martin J Wainwright, and Bin Yu. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. 2012.

[28] T Tony Cai and Anru Zhang. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *IEEE transactions on information theory*, 60(1):122–132, 2013.

[29] Mark A Davenport and Justin Romberg. An overview of low-rank matrix recovery from incomplete observations. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):608–622, 2016.

[30] T Tony Cai and Anru Zhang. Rop: Matrix recovery via rank-one projections. *The Annals of Statistics*, 43(1):102–138, 2015.

[31] Yuxin Chen, Yuejie Chi, and Andrea J Goldsmith. Exact and stable covariance estimation from quadratic sampling via convex programming. *IEEE Transactions on Information Theory*, 61(7): 4034–4059, 2015.

[32] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017.

[33] Andrew D McRae, Justin Romberg, and Mark A Davenport. Optimal convex lifted sparse phase retrieval and pca with an atomic matrix norm regularizer. *IEEE Transactions on Information Theory*, 2022.

[34] Po-Ling Loh. Statistical consistency and asymptotic normality for high-dimensional robust m-estimators. 2017.

[35] Jianqing Fan, Quefeng Li, and Yuyan Wang. Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 79(1):247, 2017.

[36] Arkadij Semenovič Nemirovskij and David Borisovich Yudin. Problem complexity and method efficiency in optimization. 1983.

[37] Stanislav Minsker. Geometric median and robust estimation in banach spaces. 2015.

[38] Daniel Hsu and Sivan Sabato. Loss minimization and parameter estimation with heavy tails. *The Journal of Machine Learning Research*, 17(1):543–582, 2016.

[39] Jianqing Fan, Weichen Wang, and Ziwei Zhu. A shrinkage principle for heavy-tailed data: High-dimensional robust low-rank matrix recovery. *Annals of statistics*, 49(3):1239, 2021.

[40] Sahand Negahban and Martin J Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *The Journal of Machine Learning Research*, 13(1): 1665–1697, 2012.

[41] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[42] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60, 2018.

[43] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[44] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[45] Yong Bao and Raymond Kan. On the moments of ratios of quadratic forms in normal random variables. *Journal of Multivariate Analysis*, 117:229–245, 2013.

[46] Shahar Mendelson. Learning without concentration. *Journal of the ACM (JACM)*, 62(3):1–25, 2015.

[47] Joel A Tropp. Convex recovery of a structured signal from independent random linear measurements. In *Sampling Theory, a Renaissance*, pages 67–101. Springer, 2015.

[48] Mark Rudelson and Roman Vershynin. Hanson-wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18:1–9, 2013.

[49] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.