James-Stein estimation of the first principal component

Alex Shkolnik*

August 31, 2021

This Version: September 4, 2021.†

Abstract

The Stein paradox has played an influential role in the field of high dimensional statistics. This result warns that the sample mean, classically regarded as the "usual estimator", may be suboptimal in high dimensions. The development of the James-Stein estimator, that addresses this paradox, has by now inspired a large literature on the theme of "shrinkage" in statistics. In this direction, we develop a James-Stein type estimator for the first principal component of a high dimension and low sample size data set. This estimator shrinks the usual estimator, an eigenvector of a sample covariance matrix under a spiked covariance model, and yields superior asymptotic guarantees. Our derivation draws a close connection to the original James-Stein formula so that the motivation and recipe for shrinkage is intuited in a natural way.

Keywords: Stein's paradox, James-Stein estimator, sample eigenvectors, PCA.

*Department of Statistics and Applied Probability, University of California, Santa Barbara, CA and Consortium for Data Analytics in Risk, University of California, Berkeley, CA. Email: shkolnik@ucsb.edu.

†I am indebted to Lisa Goldberg for conjecturing the result derived in this work.

1. Introduction. The Stein paradox has played an influential role in the field of high dimensional statistics. This result warns that the sample mean, classically regarded as the "usual estimator", may be suboptimal in high dimensions. In particular, Stein (1956) showed that the usual estimator of a location parameter $\theta \in \mathbb{R}^p$ from uncorrelated Gaussian observations becomes inadmissible when $p > 2$ under a mean-squared error criterion. That is, an estimator with a uniformly lower risk must exist. That estimator was established by James & Stein (1961) and eponymously named.

Among the numerous perspectives that motivate the James-Stein estimator,[1] the empirical Bayes perspective (see Efron & Morris (1975)) is particularly elegant. Letting η denote the sample mean computed with n measurements of an unknown $\theta \in \mathbb{R}$ and assuming an additive, normally distributed error w that has a zero mean and a variance ν^2 (e.g., $\nu = \delta/\sqrt{n}$ where each measurement has standard error δ), we write

$$\eta = \theta + w\,. \tag{1}$$

Taking a Gaussian prior on the unknown θ, that is independent of w, implies that

$$\mathrm{E}(\theta \,|\, \eta) = \mathrm{E}(\eta) + \left(1 - \frac{\nu^2}{\mathrm{Var}(\eta)}\right)(\eta - \mathrm{E}(\eta))\,, \tag{2}$$

the bivariate-normal conditional expectation formula. While, by definition of conditional expectation, $\mathrm{E}(\theta \,|\, \eta)$ is the best estimator of θ in the sense of mean-squared error, it cannot be implemented directly as the first two moments of η are unknown.[2] Stein's paradox now amounts to the fact that "good" substitutes for $\mathrm{E}(\eta)$ and $\mathrm{Var}(\eta)$ are available only in higher dimensions; precisely, when $\theta \in \mathbb{R}^p$ with $p > 2$.

Formula (2) extends easily to the multivariate[3] case and motivates the estimator

$$\eta(\mathrm{c}) = m + \mathrm{c}\,(\eta - m) \tag{3}$$

where m is an estimate (or guess) of the expected value of $\eta \in \mathbb{R}^p$ and $\mathrm{c} \in (0, 1)$ is a shrinkage parameter. In words, (3) attempts to center the entries of η, shrinks the resulting entries and recenters at m. Assuming ν is known and $p > 2$, setting

$$\mathrm{c} = 1 - \frac{\nu^2}{\mathrm{s}^2(\eta)}\left(\frac{p-2}{p}\right) \tag{4}$$

where $\mathrm{s}^2(\eta) = \sum_{i=1}^{p}(\eta_i - m_i)^2/p$ yields the James-Stein estimator. Remarkably, any fixed $m \in \mathbb{R}^p$ (e.g., Stein (1956) considers the origin)[4] results in an estimator (3) with a strictly smaller mean-squared error than η (Efron & Morris 1975, Section 1). While

[1] A few examples include the Galtonian regression perspective promoted by Stigler (1990), the purely frequentist development of the estimator in Gupta & Peña (1991) and the geometrical explanation in Brown & Zhao (2012) that builds on Stein's original heuristic argument (Stein 1956, Section 1).

[2] Often, ν is assumed to be known but estimates $\hat{\nu}$ can also be used as done in James & Stein (1961).

[3] See formulas in Anderson (2003, Section 2.5) which can be used to design James-Stein estimators of an unkown $\theta \in \mathbb{R}^p$ when w has a general covariance Σ as in Bock (1975).

[4] A natural choice for m takes the sample mean of η in each entry, referred to as shrinkage toward the "grand mean". However, this would require $p > 3$ (Efron & Morris 1975)).

three is provably the critical dimension[5], Stein (1956, Section 1) heuristically argued for higher performance in higher dimensions and for relaxing the normality.

The James-Stein estimator has inspired a rich literature on the theme of "shrinkage" in statistics. Just a small sampling of examples includes ridge regression (Hoerl & Kennard 1970), the LASSO (Tibshirani 1996), the Ledoit-Wolf covariance estimator (Ledoit & Wolf 2004) and the Elastic Net (Zou & Hastie 2005). Excellent textbook treatments of the ideas behind the Stein paradox and James-Stein shrinkage include Gruber (2017) and Fourdrinier, Strawderman & Wells (2018). In this paper, we leverage these ideas to develop and analyze a James-Stein estimator for the first principal component of a sample covariance matrix. The results again prove the efficacy of James-Stein estimation, and do so for one the cornerstone methods in high-dimensional statistics, principal component analysis (Jolliffe & Cadima 2016).

Consider a $p \times p$ sample covariance matrix S that is based on n observations of some random vector $y \in \mathbb{R}^p$. Without loss of generality, we write

$$\mathrm{S} = \mathscr{s}_p^2 hh^\top + \mathrm{G} \tag{5}$$

for $\mathrm{G} = \mathrm{S} - \mathscr{s}_p^2 hh^\top$ and h, the sample eigenvector with the largest eigenvalue, i.e.,

$$\mathrm{S}h = \mathscr{s}_p^2 h \quad \text{and} \quad \mathscr{s}_p^2 = \max_{|z|=1} \langle z, \mathrm{S}z \rangle . \tag{6}$$

By convention h has unit length and corresponds to a direction along which the variance of S (i.e., $\mathscr{s}_p^2$) is maximum (i.e., the first principal component). A substantial and rapidly growing literature exists to study the (p and/or n) asymptotic behavior of the eigenpair $(\mathscr{s}_p^2, h)$ and the remaining eigenstructure in order to quantify either the estimation or the empirical error. See Wang & Fan (2017) for recent results and a systematic discussion of this literature. The topic of shrinkage estimators arises naturally in this context and was raised by Stein (1986), who suggested improving the usual estimate S via eigenvalue shrinkage. Indeed, there is by now a large literature on estimators that adjust the eigenvalues of sample covariance matrices to improve their performance with respect to some loss function (Donoho, Gavish & Johnstone 2018).

In this paper, we develop and analyze a James-Stein estimator for the first principal component of a high-dimension and low-sample (HDLS) data set.[6] The recipe for the estimator begins with h and $\mathscr{s}_p^2$ in (6) and the next $(q-1)$ largest sample eigenvalues $\mathscr{s}_{p-1}^2, \ldots, \mathscr{s}_{p-q+1}^2$ ($\min(n, p) > q$) corresponding to a model with q spikes.[7]

Step 1. Set $\eta = \mathscr{s}_p h$, compute the sample statistics $m(\eta) = \sum_{i=1}^p \eta_i / p$ and $\mathrm{s}^2(\eta) = \sum_{i=1}^p (\eta_i - m(\eta))^2 / p$, and define

$$\mathrm{c} = 1 - \frac{\hat{\mathfrak{v}}^2}{\mathrm{s}^2(\eta)} \quad \text{where} \quad \hat{\mathfrak{v}}^2 = \left(\frac{\mathrm{tr}(\mathrm{S}) - (\mathscr{s}_p^2 + \cdots + \mathscr{s}_{p-q+1}^2)}{\min(n, p) - q} \right) / p .$$

[5] The theme of a critical dimension is encountered frequently in statistics and probability. Brown (1971), for example, derives a close mathematical relationship between the admissibility of the James-Stein estimator and the transience of the Brownian motion in $\mathbb{R}^p$, which also requires $p > 2$.

[6] The HDLS framework, as introduced in Hall, Marron & Neeman (2005) and Ahn, Marron, Muller & Chi (2007), is increasingly relevant for data science (Aoshima, Shen, Shen, Yata, Zhou & Marron 2018).

[7] Roughly speaking q is the number of factors (or spikes) in the data, after which a sufficiently large eigengap (between the qth and the next eigenvalue) is observed (see Fan, Guo & Zheng (2020)).

Step 2. Return the estimator (corrected principal component)[8]

$$h^{\mathrm{JS}} = \frac{1}{\sqrt{p}} \left(\frac{m(\eta) + \mathrm{c}\,(\eta - m(\eta))}{\sqrt{m^2(\eta) + \mathrm{c}^2\,\mathrm{s}^2(\eta)}} \right).$$

The vector h^{JS} is the James-Stein estimator of the first principal component of the data. The numerator contains the shrinkage formula (3) while the divisor normalizes the shrunk vector to a unit length (by convention). The relationship to the shrinkage parameter c in (4) is evident by treating p as large. The estimate $\hat{\mathrm{v}}^2$ corresponds to the bulk of the eigenvalue spectrum, and may be viewed the "noise" in the context of a signal-to-noise ratio that plays a prominent role of the results in Sections 2 & 3.

It is reasonable to suspect that a James-Stein type shrinkage of the principal component h, a high dimensional vector, could improve either the convergence rate or accuracy of the limit in some appropriate asymptotic regime. However, the standard orthonormal transformation and the eigengap partition of the sample eigenvectors, that is typically leveraged by their asymptotic analyses (e.g., Paul (2007), Shen, Shen, Zhu & Marron (2016) and Wang & Fan (2017)), can obscure the systematic nature of the sample bias. As sensibly pointed out by Wang & Fan (2017) in reference to the partition of the sample eigenvector, the "*two parts intertwine in such a way that correction for the biases of estimating eigenvectors is almost impossible.*" However, in the original (untransformed) coordinate system and the HDLS asymptotic regime, the bias can in fact be identified, characterized and (partially) corrected. This program was carried out in Goldberg, Papanicalaou & Shkolnik (2021), who adopt a factor model in an HDLS regime and utilize a portfolio theory application to motivate their analysis.

The main results of this paper rederive the adjustment of Goldberg et al. (2021) but within a James-Stein type framework. In particular, we establish identity (1) in which $\eta = \mathfrak{z}_p h$ and θ is related to the associated population eigenvector. From here, the James-Stein shrinkage acts on the perturbation w so that the estimator h^{JS} outperforms h on the mean-squared error and angle metrics as $p \uparrow \infty$. The theoretical guarantees provided here are new and their proofs rely on a different set of mathematical tools than Goldberg et al. (2021). In particular, the new approach leverages Weyl's inequality and the Davis-Kahan theorem from matrix perturbation theory to give simpler proofs and potentially expand the scope of applicability of the resulting estimator. The HDLS regime, in which the number of variables p grows to infinity, and the number of observations n to be fixed, plays a crucial role in the analysis.

The paper is organized as follows. Section 2 defines the spiked covariance model underlying our results. Section 3 develops the James-Stein estimator h^{JS} and Section 4 proves the theoretical guarantees for this estimator. Appendix A contains proofs of the auxiliary results. The following notation is used throughout. Let $\langle u, v \rangle$ denote the standard inner product of $u, v \in \mathbb{R}^d$ so that $|u| = \sqrt{\langle u, u \rangle}$ and $m(u) = \langle u, \mathrm{e} \rangle / d$ where $\mathrm{e} = (1, \ldots, 1)^\top$ are the length and mean. Set $\mathrm{s}^2(u) = |u - m(u)|^2/d$ and $\mathrm{cov}(u, v) = \langle u - m(u), v - m(v) \rangle / d$ (see the notation of footnote 8). We use a subscript $1 \leq p \leq \infty$ to highlight the dependence on p of various quantities, e.g., $m(\eta) = m_p(\eta)$ for $\eta \in \mathbb{R}^p$ and $m_\infty(\eta)$ is the limit $\lim_{p \uparrow \infty} m_p(\eta)$ when it exists.

[8] With a slight abuse of the notation, $u - x = (u_1 - x, \ldots, u_p - x)$ for $u \in \mathbb{R}^p$ and $x \in \mathbb{R}$.

2. A scarcely sampled spiked model. We use a spiked covariance model, borrowed from the HDLS literature. We also restrict ourselves to a single unbounded spike in the "boundary case" (see Jung, Sen & Marron (2012)) wherein the largest eigenvalue of the covariance matrix grows linearly in p. In particular, consider a mean-zero (w.l.o.g.) random vector $y \in \mathbb{R}^p$ with a $p \times p$ covariance matrix $\Sigma = \mathrm{Var}(y)$ and let,

$$\Sigma = \Gamma + \beta\beta^\top \tag{7}$$

for a symmetric, positive-semidefinite $p \times p$ matrix Γ and vector $\beta \in \mathbb{R}^p$. The following affirms that $b = \beta/|\beta|$ is an eigenvector of Σ with eigenvalue $\langle \beta, \beta \rangle$.

Assumption 2.1 (w.l.o.g.). $\Gamma b = 0$ *and* $m_p(b) \geq 0$ *for any* p.

To state our additional assumptions on the model, we project the data vector $y \in \mathbb{R}^p$ onto the eigenvectors of Σ. More precisely, define

$$\psi_p = \langle \beta, y \rangle / \langle \beta, \beta \rangle . \tag{8}$$

It is immediate that $\mathrm{E}(\psi_p) = 0$ and $\mathrm{Var}(\psi_p) = 1$. For every eigenvalue α_i of Γ, let

$$\phi_i = \langle \gamma, y \rangle \quad \text{where} \quad \gamma \in \mathbb{R}^p \ : \ \Gamma\gamma = \alpha_i \gamma . \tag{9}$$

Note, $\mathrm{E}(\phi_i) = 0$ and $\mathrm{Var}(\phi_i) = \alpha_i$. As the dimension p grows we obtain a sequence $\{\phi_i\}_{i \geq 1}$. As a technical remark, $\phi_i = \phi_{i,p}$ and $\alpha_i = \alpha_{i,p}$ depend on p. We consider a sequence of models (7) constructed from sequences $\{\beta_i\}_{i \geq 1}$ and $\{\Gamma_p\}$.

Assumption 2.2. *For constants* $\mu \in \mathbb{R}$ *and* $\sigma, \delta \in (0, \infty)$ *as* $p \uparrow \infty$ *we have:*

(i) $m_\infty(\beta) = \mu$ *and* $\mathrm{s}^2_\infty(\beta) = \sigma^2$.

(ii) $\psi_\infty = \lim_{p \uparrow \infty} \psi_p$ *exists as a* $\mathbb{R}$*-valued random variable almost surely.*

(iii) $m_\infty(\phi) = m_\infty(\varphi) = 0$ *almost surely for* $\{\varphi_i\}_{i \geq 1}$ *with* $\varphi_i = \varphi_{i,p} = \phi_i$ $pm(\gamma^i)$.

(iv) $\mathrm{s}^2_\infty(\phi) = m_\infty(\alpha) = \delta^2$ *almost surely.*

Condition (i) imposes regularity on the sequence $\{\beta_i\}_{i \geq 1}$ and implies that the largest eigenvalue of Σ (i.e., $\langle \beta, \beta \rangle$) grows linearly with p. The random variable ψ_∞ in (ii) is closely related to a principal component score (in the limit $p \uparrow \infty$), and it captures the randomness along the first (population) principal component. Conditions (iii) and (iv) are related to certain requirements on a measure of sphericity of the model (e.g., $\mathrm{tr}(\Sigma)^2/(\mathrm{tr}(\Sigma^2)p)$) and summarize the conclusions of Jung & Marron (2009, Theorem 1). In particular, (iii) may be viewed as laws of large numbers for $\{\phi_i\}_{i \geq 1}$ and $\{\varphi_i\}_{i \geq 1}$ and suggests $pm(\gamma^i) \asymp 1$ whereas (i) implies the first principal component has $\sqrt{p}m(b) \asymp 1$. This highlights that the spike eigenvector b differs from those of Γ both in terms of the magnitude of its eigenvalue as well as the structure of the vector itself. According to (iv), the average eigenvalue of Γ (unlike the spike eigenvalue $\langle \beta, \beta \rangle$) is bounded in p, and per (iii), the eigenvectors γ^i do not have entries biased towards a nonzero mean $\sqrt{p}m(\gamma^i)$, unlike b (an exception is $\mu = 0$, which is the case when the James-Stein estimator will turn out to be ineffective; see Remark 2.8).

The following assumption is a standard one in statistical data analysis but may be (and often is) relaxed in the HDLS setup (e.g., Jung et al. (2012)).

Assumption 2.3. *There are a fixed $n \geq 2$ i.i.d. observations of $y \in \mathbb{R}^p$.*

Our forthcoming results hold even when only 2 observations are available, hence, a scarcely sampled model. Let Y be the $p \times n$ data matrix with the kth column containing the kth observation of y, and define the sample covariance matrix S by

$$\mathrm{S} = \mathrm{Y}\mathrm{Y}^\top / n \,. \tag{10}$$

We let $h \in \mathbb{R}^p$ denote the eigenvector of S with the largest eigenvalue $\mathfrak{s}_p^2$ (see (6)). It is unique only up to sign (and $|h| = 1$), motivating the following (c.f. Assumption 2.1).

Assumption 2.4 (w.l.o.g.). *$m_p(h) \geq 0$ for any p.*

We write $\mathrm{S} = \mathrm{G} + \eta\eta^\top$ in analogy to (7) (taking $\mathrm{G} = \mathrm{S} - \eta\eta^\top$) and set

$$\eta = \mathfrak{s}_p h \,. \tag{11}$$

Next, define the following measure of finite-sample distortion,

$$\chi_n^2 = \langle \mathcal{X}, \mathcal{X} \rangle / n \quad \text{and} \quad \mathcal{X} = \lim_{p \uparrow \infty} \frac{\mathrm{Y}^\top \beta}{\langle \beta, \beta \rangle} \,. \tag{12}$$

The latter limit exists by Assumption 2.2 while Assumption 2.3 implies that $\mathcal{X} \in \mathbb{R}^n$ has i.i.d. entries (distributed as ψ_∞). Consequently, we have $\chi_n^2 \to 1$ as $n \uparrow \infty$.

We can measure the error in any estimator η of θ by the mean-squared error, as would be consistent with the James-Stein framework.

$$\mathrm{MSE}_p(\eta \,|\, \theta) = \langle \eta - \theta, \eta - \theta \rangle / p \tag{13}$$

Proposition 2.5. *Let $\theta = \chi_n \beta$ and suppose Assumptions 2.1–2.4 hold. Then,*

$$\mathrm{MSE}_\infty(\eta \,|\, \theta) = \frac{\delta^2}{n} \,. \tag{14}$$

Remark 2.6. *If $\theta = \beta$ then the right side would be multiplied by the factor, $1 + \frac{\mathrm{SNR}^2}{\mathrm{r}_\infty^2} \left(\frac{\chi_n - 1}{\chi_n} \right)^2$ where we define a signal-to-noise ratio* SNR *and signal-incoherence* r_∞ *as*

$$\mathrm{SNR} = \left(\frac{\sigma}{\delta} \right) \chi_n \sqrt{n} \quad \textit{and} \quad \mathrm{r}_\infty = \frac{1}{\sqrt{1 + (\mu/\sigma)^2}} \,. \tag{15}$$

For SNR*, we regard $\sigma \chi_n$ as a distorted signal, and $\delta / \sqrt{n}$ as noise that vanishes as n grows (also, $\sigma \chi_n \to \sigma$). The signal-incoherence r_∞ is the limit of $\mathrm{r}_p = \mathrm{r}_p(\beta) = \mathrm{s}_p(\beta)/|\beta|$ (per Assumption 2.2) determined by the signal-to-noise ratio μ/σ of the vector β. A large value of r_∞ corresponds to more variation, or "incoherence" in the entries $\{\beta_i\}_{i \geq 1}$*

A more standard way to evaluate the goodness of a sample eigenvector h is via its angle away from its population counterpart $b = \beta/|\beta|$. To this end, let

$$\mathrm{SPH}_p(h \,|\, b) = \mathrm{SPH}_p(\eta \,|\, \theta) = \sin^2 \left(\arccos \frac{\langle \eta, \theta \rangle}{|\eta||\theta|} \right). \tag{16}$$

Proposition 2.7. *Let $\theta = \chi_n \beta$ and suppose Assumptions 2.1–2.4 hold. Then,*

$$\mathrm{SPH}_\infty(\eta \,|\, \theta) = \frac{\mathrm{r}_\infty^2}{\mathrm{r}_\infty^2 + \mathrm{SNR}^2} \,. \tag{17}$$

Remark 2.8. *Clearly, (17) also holds with $\theta = \beta$.*

3. James-Stein estimation of sample eigenvectors. Having established that the sample eigenvector corresponding to the spike (i.e., the largest eigenvalue) carries finite sample error, it is natural to ask whether James-Stein shrinkage can improve this "usual" estimator. The key to this question is the (to be established) identity

$$\eta = \theta + w \quad \text{and} \quad \theta = \chi_n \beta \tag{18}$$

for a random vector $w \in \mathbb{R}^p$ specified in (27) of Section 4.1. As in definition (11), we have $\eta = \mathfrak{z}_p h$ where h is the sample eigenvector with the largest eigenvalue, $\mathfrak{z}_p^2$. The perturbation w of θ turns out to be such that the shrinkage of η is effective. We remark that the recipe of Section 1 extends our derivation of the James-Stein estimator below to the case of multiple (there, q) spikes in a natural way. This extension is effective because the eigenvectors corresponding to the spikes are mutually orthogonal, but it is suboptimal. An optimal estimator in the multi-spiked setup is left for future work.

3.1. The JS estimator. Equation (18) establishes a relationship between the sample and population eigenvectors that suggests a James-Stein estimator may be derived. An informal derivation proceeds as follows. Consider the shrinkage parameter

$$\mathrm{c} = 1 - \frac{\hat{\mathfrak{v}}^2}{\mathrm{s}^2(\eta)} \tag{19}$$

based on (4) with $\hat{\mathfrak{v}}$, an estimate of the "noise". It is reasonable to assign the latter to be the average of the non-spiked, non-zero eigenvalues of S. That is,

$$\hat{\mathfrak{v}}^2 = \left(\frac{\mathrm{tr}(\mathrm{S}) - \mathfrak{z}_p^2}{n-1} \right) / p \qquad (p \geq n), \tag{20}$$

where the scaling by p turns out to be necessary due to the counterintuitive behavior of the HDLS asymptotics. When $p < n$ the divisor $n-1$ must be replaced by $p-1$.

This paves the way for the James-Stein sample eigenvector estimate,

$$\eta^{\mathrm{JS}} = m(\eta) + \mathrm{c}\,(\eta - m(\eta))$$

of the unnormalized eigenvector and by convention, we take unit length version,

$$h^{\mathrm{JS}} = \frac{\eta^{\mathrm{JS}}}{|\eta^{\mathrm{JS}}|} = \frac{1}{\sqrt{p}} \left(\frac{m(h) + \mathrm{c}\,(h - m(h))}{\sqrt{m^2(h) + \mathrm{c}^2\,\mathrm{s}^2(h)}} \right) \tag{21}$$

as the James-Stein estimator of the population eigenvector $b = \beta/|\beta|$.

The following James-Stein type theorems characterize the improvement due to shrinkage in the original mean-squared sense as well as in the angle metrics.

Theorem 3.1. *Suppose Assumptions 2.1–2.4 hold. Then, almost surely,*

$$\mathrm{MSE}_\infty(\eta^{\mathrm{JS}} \,|\, \theta) = \mathrm{c}_\infty \,\mathrm{MSE}_\infty(\eta \,|\, \theta)$$

where $\mathrm{c}_\infty \in (0,1)$ *is the limit of* $\mathrm{c}_p = \mathrm{c}$ *in* (19) *with* SNR *defined in* (15) *and*

$$\mathrm{c}_\infty = \frac{\mathrm{SNR}^2}{1 + \mathrm{SNR}^2}. \tag{22}$$

Theorem 3.2. *Suppose Assumptions 2.1–2.4 hold. Then, almost surely,*

$$\mathrm{SPH}_\infty(\eta^{\mathrm{JS}} \,|\, \theta) = \mathrm{SPH}_\infty(h^{\mathrm{JS}} \,|\, b) = \mathrm{d}_\infty \,\mathrm{SPH}_\infty(h \,|\, b)$$

where $\mathrm{d}_\infty \in [\mathrm{c}_\infty, 1]$ *where* c_∞ *is in* (22) *and with* SNR *and* r_∞ *in* (15)*, we have*

$$\mathrm{d}_\infty = \mathrm{c}_\infty + \frac{\mathrm{r}_\infty^2}{1 + \mathrm{SNR}^2}. \tag{23}$$

Related results may be found in Goldberg, Papanicolaou, Shkolnik & Ulucam (2020) and Goldberg et al. (2021) but the metrics there are motivated by solutions of certain quadratic programs that are useful in finance and portfolio theory.

Remark 3.3. *Note,* $\mathrm{d}_\infty = 1$ *(i.e., no improvement in angle) if and only if* $\mu = 0$*.*

3.2. The geometry of Stein's paradox. We shed insight into the James-Stein estimator in (21) by deriving general conditions under which Theorems 3.1 and 3.2 hold. Our analysis adopts the pure frequentist perspective in Gupta & Peña (1991) and supplements it by illustrating the Euclidean and the spherical geometry of the estimator. The two geometries reflect the definitions of the error (MSE and SPH) in the two theorems.

Lemma 3.4. *Let* $\eta = \theta + w$ *for* $\theta, w \in \mathbb{R}^p$*. Then, the solutions of the optimizations* $\min_{c\in\mathbb{R}} \mathrm{MSE}\,(\eta(c) \,|\, \theta)$ *and* $\min_{c\in\mathbb{R}} \mathrm{SPH}\,(\eta(c) \,|\, \theta)$ *(see* (13) *and* (16)*) are given by*

$$\mathrm{c}^{\mathrm{MSE}} = \frac{\mathrm{cov}(\theta, \eta)}{\mathrm{s}^2(\eta)} \quad \textit{and} \quad \mathrm{c}^{\mathrm{SPH}} = \frac{m(\eta)}{m(\theta)}\,\mathrm{c}^{\mathrm{MSE}}. \tag{24}$$

The next assumptions may be viewed as laws of large numbers in the random setting or regularity conditions in a deterministic one. They concern the sequences $\{\theta_i\}_{i=1}^\infty$ and $\{w_i\}_{i=1}^\infty$, and allow for dependence on p (i.e., $\theta_i = \theta_i^{(p)}$ and $w_i = w_i^{(p)}$).

Assumption 3.5. *For constants* $m \in \mathbb{R}$ *and* $\nu, \xi \in (0, \infty)$ *as* $p \uparrow \infty$*, we have:*

(i) $m_\infty(\theta) = m$ *and* $\mathrm{s}^2_\infty(\theta) = \xi^2$*,*

(ii) $m_\infty(w) = 0$ *and* $\mathrm{s}^2_\infty(w) = \nu^2$*,*

(iii) $\mathrm{cov}_\infty(\theta, w) = 0$*,*

(iv) *there exists an estimator* $\hat{\nu} = \hat{\nu}_p$ *for each* p *with* $\hat{\nu}_\infty = \nu$*.*

The following identities follow by direct calculation.

Lemma 3.6. *Suppose* $\{\theta_i\}$ *and* $\{w_i\}$ *satisfy Assumption 3.5 and* $\eta_i = \theta_i + w_i$*. Then (almost surely),* $m_\infty(\eta) = m$*,* $\mathrm{cov}_\infty(\eta, \theta) = \xi^2$ *and* $\mathrm{s}^2_\infty(\eta) = \xi^2 + \nu^2$*.*

We define the signal-to-noise ratio SNR and the signal-incoherence r_∞ as

$$\mathrm{SNR} = \frac{\xi}{\nu} \quad \text{and} \quad \mathrm{r}_\infty = \frac{1}{\sqrt{1 + (m/\xi)^2}}, \tag{25}$$

which are compatible with (15) upon taking $\theta = \chi_n \beta$ and $\nu = \delta/\sqrt{n}$. The following result establishes the conclusions of Theorems 3.1 and 3.2 in our abstract setting.

Proposition 3.7. *Let* $\eta = \theta + w$ *where* $\theta, w \in \mathbb{R}^p$ *and an estimator* $\hat{\nu}$ *satisfy Assumption 3.5. Then, for* c_∞ *and* d_∞ *defined in (22) and (23) but with* SNR *and* r_∞ *in (25) the estimate* $\eta(\mathrm{c}) = \eta + \mathrm{c}\,(\eta - m(\eta))$ *with parameter* $\mathrm{c} = 1 - \frac{\hat{\nu}^2}{\mathrm{s}^2(\eta)}$ *satisfies*

$$\mathrm{MSE}_\infty(\eta(\mathrm{c})\,|\,\theta) = \mathrm{c}_\infty \mathrm{MSE}_\infty(\eta\,|\,\theta) \;\; \textit{and} \;\; \mathrm{SPH}_\infty(\eta(\mathrm{c})\,|\,\theta) = \mathrm{d}_\infty \mathrm{SPH}_\infty(\eta\,|\,\theta).$$

Moreover, the optimal parameters $\mathrm{c}^{\mathrm{MSE}}$ *and* $\mathrm{c}^{\mathrm{SPH}}$ *in (24) converge as* $p \uparrow \infty$ *to* c_∞.

Figure 1 illustrates the geometry of the estimator $\eta(\mathrm{c})$ of the vector θ.

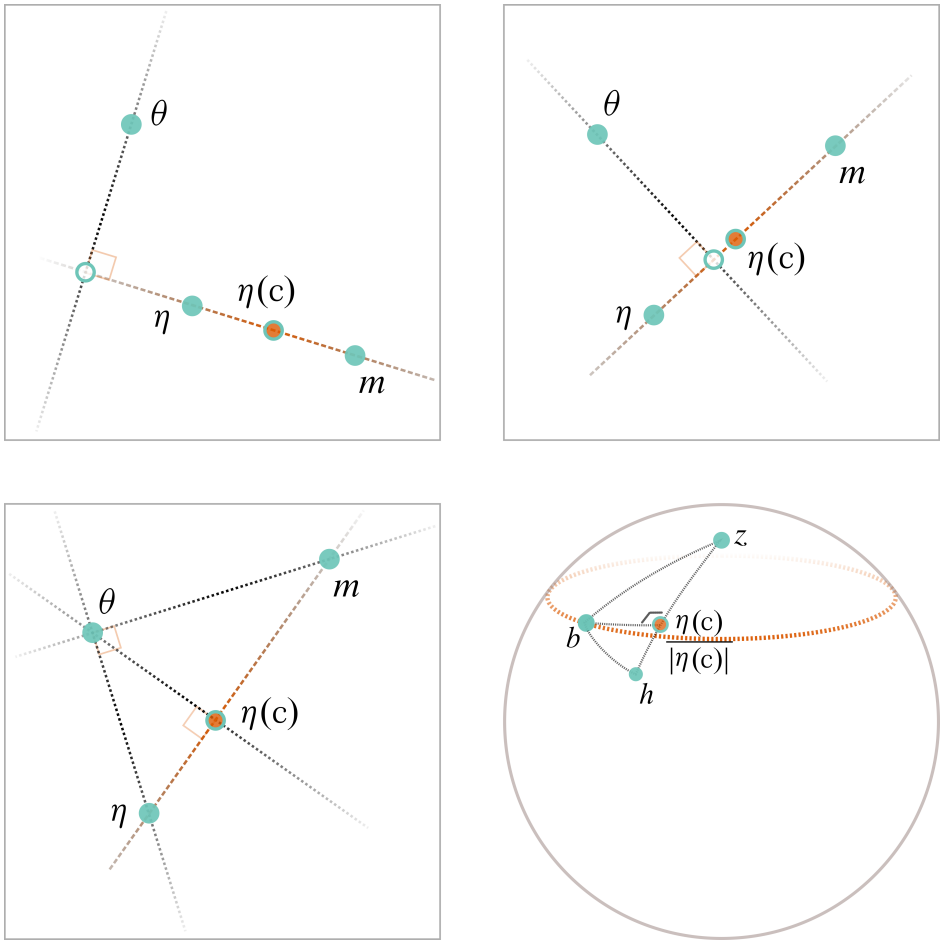

Figure 1. *Illustration of the estimator* $\eta(\mathrm{c})$ *in low* (top left)*, high* (top right)*, and limiting* (bottom left) *dimensions, relative to the shrinkage target* $m \in \mathbb{R}^p$*, the vector with all entries equal to* $m(\eta)$*. The open circle marks the estimator with the optimal shrinkage parameter* $\mathrm{c}^{\mathrm{MSE}}$*. In Euclidean geometry, the estimator* $\eta(\mathrm{c})$ *is located anywhere on the ray originating at the observation* η *and passing through* m*, because* $\mathrm{c} \leq 1$ *while* $\eta(1) = \eta$ *and* $\eta(0) = m$*. The spherical geometry* (bottom right) *presents the limiting* ($p = \infty$) *analog of the illustration in the bottom left. There,* $b = \theta/|\theta|$*,* $h = \eta/|\eta|$ *and* $z = m/|m|$ *and the contour describes all vectors with mean entry* $m(b)$*, which highlights the difference between the two geometries.*

4. Proofs the main results. We proceed by the following three main steps.

(1) We establish the key identity $\eta = \theta + w$ with $\theta = \chi_n \beta$ per (18) in Section 4.1.

(2) We derive the convergence properties in the HDLS regime of the eigenvalues and eigenvectors of S under our spiked covariance model setting in Section 4.2.

(3) We verify that θ, w in (1) and $\hat{v}$ in (20) satisfy the conditions of Assumption 3.5, which leads to the guarantees for James-Stein shrinkage in Proposition 3.7.

Theorems 3.1 and 3.2 and then corollaries of Proposition 3.7, which is proved in Appendix A. We will make use of two classic results in matrix perturbation theory.

Theorem (Weyl). *Let* A *and* $(\mathrm{A} + \Delta)$ *be (real) symmetric* $n \times n$ *matrices with eigenvalues* $\alpha_1 \geq \cdots \geq \alpha_n$ *and* $\zeta_1 \geq \cdots \geq \zeta_n$ *respectively. Then,*

$$\max_{1 \leq j \leq n} |\alpha_j - \zeta_j| \leq |\Delta|.$$

For a proof see Horn & Johnson (2013) (also Weyl (1912)).

Theorem (Davis-Kahan). *Let* A *and* $(\mathrm{A} + \Delta)$ *be (real) symmetric* $n \times n$ *matrices with* $\mathrm{A}a^j = \alpha_j a^j$ *and* $(\mathrm{A} + \Delta)b^j = \beta_j b^j$ *for eigenvectors* $a^j, b^j \in \mathbb{R}^n$ *and eigenvalues* $\alpha_j, \beta_j \in \mathbb{R}$. *Suppose* $\alpha_1 \geq \cdots \geq \alpha_n$ *and* $\beta_1 \geq \cdots \geq \beta_n$ *with the convention* $\alpha_0 = \infty = -\alpha_{n+1}$ *and assume* $\gamma_j = \min\{\alpha_{j-1} - \alpha_j, \alpha_j - \alpha_{j+1}\} > 0$. *Then,*

$$|a^j - b^j| \leq \frac{3}{\gamma_j}|\Delta| \qquad \textit{provided (w.l.o.g.)} \ \langle a^j, b^j \rangle \geq 0.$$

This result is proved in Yu, Wang & Samworth (2015, Corollary 1).

4.1. Establishing the key identity. A key tool for random matrix theory in the HDLS regime is the dual sample covariance matrix. This is $n \times n$ matrix ($n \geq 2$ is fixed),

$$\mathrm{L} = \mathrm{Y}^\top \mathrm{Y}/p. \tag{26}$$

The next result is well known and relates the spectra of $\mathrm{S} = \mathrm{Y}\mathrm{Y}^\top/n$ and L.

Lemma 4.1. *Let* $\mathrm{L}u = \ell^2 u$ *where* $\ell^2 \in (0, \infty)$ *and* $u \in \mathbb{R}^n$. *Then,* $\mathrm{S}v = \imath^2 v$ *where* $v = \mathrm{Y}u/(\sqrt{p}\ell)$ *and* $\imath^2 = \ell^2 p/n$. *Conversely, let* $\mathrm{S}v = \imath^2 v$ *where* $\imath^2 \in (0, \infty)$ *and* $v \in \mathbb{R}^p$. *Then,* $\mathrm{L}u = \ell^2 u$ *where* $u = \mathrm{Y}^\top v/(\sqrt{n}\imath)$ *and* $\ell^2 = \imath^2 n/p$.

Proof. Multiplying the identity $\mathrm{L}u = \ell^2 u$ by Y from both sides, we obtain

$$\mathrm{YL}u = \ell^2 \mathrm{Y}u \quad \Rightarrow \quad \mathrm{SY}u = \Big(\frac{\ell^2 p}{n}\Big)\mathrm{Y}u.$$

Note, $(\mathrm{Y}u)^\top(\mathrm{Y}u) = (u^\top \mathrm{L}u)p = \ell^2 p$, so $v = (\mathrm{Y}u)/(\sqrt{p}\ell)$ has unit length. Dividing by $\sqrt{p}\ell$ yields $\mathrm{S}v = \imath^2 v$ as required. The converse has an identical argument. ∎

The spike model $\Sigma = \Gamma + \beta\beta^\top$ has a full basis of eigenvectors given by $b = \beta/|\beta|$ and $\{\gamma^i\}_{i=1}^{p-1}$, the latter corresponding to the nonzero eigenvalues $\{\alpha\}_{i=1}^{p-1}$ of Γ. Thus,

$$y = \langle b, y\rangle b + \textstyle\sum_{i=1}^{p}\langle \gamma^i, y\rangle \gamma^i = \beta\psi + \epsilon$$

where $\epsilon = \sum_{i=1}^{p}\phi_i\gamma^i \in \mathbb{R}^p$ and $\psi = \psi_p = \langle\beta, y\rangle/\langle\beta,\beta\rangle$ as in (8). Consequently, letting Y denote the $p \times n$ matrix of i.i.d. observations of $y \in \mathbb{R}^p$, we have

$$\mathrm{Y} = \beta\mathrm{X}^\top + \mathcal{E}\,.$$

where $\mathrm{X} = \mathrm{Y}^\top\beta/\langle\beta,\beta\rangle \in \mathbb{R}^n$ consists of i.i.d. observations of ψ and $\mathcal{E}$ is a $p \times n$ matrix with i.i.d. columns consisting of the observations of ϵ as defined above.

By orthogonality, we obtain that $\mathrm{L} = \mathrm{Y}^\top\mathrm{Y}/p = (\langle\beta,\beta\rangle/p)\,\mathrm{X}\mathrm{X}^\top + \mathcal{E}\mathcal{E}^\top/p$. Let x_p be the eigenvector of L with the largest eigenvalue, ℓ_p^2, and $x_\infty = \mathcal{X}/\chi_n \in \mathbb{R}^n$ (the unit length normalization of $\mathcal{X}$), where $\mathcal{X}$ and χ_n are defined in (12). By Lemma 4.1,

$$\begin{aligned} h = \frac{\mathrm{Y}x_p}{\sqrt{p}\ell_p} &= \frac{(\beta\mathrm{X}^\top + \mathcal{E})x}{\sqrt{p}\ell_p} = \beta\,\frac{\chi_n\langle x_p, x_\infty\rangle}{\sqrt{p/n}\ell_p} + \frac{\mathcal{E}x_p}{\sqrt{p}\ell_p} \\ &= \beta\Big(\frac{\chi_n}{\mathfrak{s}_p}\Big) + \beta\Big(\frac{\chi_n}{\mathfrak{s}_p}\Big)\langle x_p, x_p - x_\infty\rangle + \frac{\mathcal{E}x_p}{\sqrt{n}\mathfrak{s}_p} \end{aligned}$$

We deduce that $\eta = \mathfrak{s}_p h = \chi_n\beta + w$ as required by (18) where

$$w = \chi_n\beta\,\langle x_p, x_p - x_\infty\rangle + \frac{\mathcal{E}x_p}{\sqrt{n}}\,. \tag{27}$$

4.2. Convergence of the eigen- values/vectors. It is not difficult to establish that limit of $\mathrm{L} = \mathrm{L}^{(p)}$ as $p \uparrow \infty$ (in any norm on $\mathbb{R}^{n\times n}$) takes the following form

$$\mathrm{L}^{(\infty)} = (\sigma^2 + \mu^2)\,(n\chi_n^2)\,x_\infty x_\infty^\top + \delta^2\mathrm{I}.$$

The first term is the limit of $(\langle\beta,\beta\rangle/p)\,\mathrm{X}\mathrm{X}^\top$ under Assumption 2.2 (note that $\langle\beta,\beta\rangle = \mathrm{s}^2(\beta) + m^2(\beta)$) and the definitions of X and x_∞ above. By Assumption 2.3, the columns of $\mathcal{E}$ are i.i.d. copies of ϵ with $\mathrm{E}(\epsilon) = 0_p$ by definition, the strong law of large numbers, confirms the that the off-diagonal entries of the second term are zero. That δ^2 determines all the diagonal entries is again a consequence of Assumption 2.2. This entails proving that $\langle\epsilon,\epsilon\rangle/p$ converges almost surely to δ^2 as is done in Section 4.3 item (ii).

The matrix $\mathrm{L}^{(\infty)}$ has an easily described spectrum. Its largest eigenvalue is given by $\ell_\infty^2 = (\sigma^2 + \mu^2)\,(n\chi_n^2) + \delta^2$ and has the eigenvector x_∞. All remaining eigenvalues equal δ^2. Since $\mathrm{L}^{(p)}$ converges (in any norm) to $\mathrm{L}^{(\infty)}$, the largest eigenvalue ℓ_p^2 converges to ℓ_∞^2 almost surely. All the remaining eigenvalues converge to δ^2.

By Weyl's inequality, setting $\mathrm{A} = \Omega$ and $\mathrm{B} = (\mathrm{L} - \Omega)$ so that $\mathrm{A} + \mathrm{B} = \mathrm{L}$, we have

$$|\ell_{p-i+1}^2 - ((\sigma^2 + \mu^2)\,(n\chi_n^2)1_{\{i=1\}} + \delta^2)| \le |\mathrm{L}^{(p)} - \mathrm{L}^{(\infty)}| \qquad 1 \le i \le n\,. \tag{28}$$

We immediately deduce (by Lemma 4.1) the following result, variants of which appear in Jung et al. (2012), Shen et al. (2016) and Goldberg et al. (2021). Let $\mathfrak{s}_{p-i+1}^2$ denote the ith largest eigenvalue of S (for $i > \min(p,n)$ all eigenvalues are zero).

Proposition 4.2. *Fix $n \geq 2$ and suppose Assumptions 2.1, 2.2 and 2.3 hold. Then, $\lim_{p\uparrow\infty} \jmath^2_{p-i+1}/p = 1_{\{i=1\}}\chi_n^2(\sigma^2+\mu^2) + \delta^2/n$ almost surely for fixed $1 \leq i \leq n$.*

Next, by the Davis-Kahan theorem with $\mathrm{L}^{(p)} = \mathrm{L}^{\infty} + \Delta$ and $\Delta = \mathrm{L}^{(p)} - \mathrm{L}^{(\infty)}$, for x_p and x_∞ the eigenvectors of $\mathrm{L}^{(p)}$ and $\mathrm{L}^{(\infty)}$ with largest eigenvalues respectively,

$$|x_p - x_\infty| \leq \frac{3}{\delta^2}\,|\mathrm{L}^{(p)} - \mathrm{L}^{(\infty)}|. \tag{29}$$

Note the condition $\langle x_p, x_\infty\rangle \geq 0$ is without loss of generality as the orientation of the eigenvectors is always arbitrary. The following result follows immediately.

Proposition 4.3. *Fix $n \geq 2$ and suppose that Assumptions 2.1, 2.2 and 2.3 hold. Then, we have $|x_p - x_\infty| \to 0$ as $p \uparrow \infty$ almost surely.*

4.3. Verifying Assumption 3.5. We make use of Assumptions 2.1–2.2. There are four items to verify for $\theta = \chi_n\beta$ and w in (27). Take, $m = \chi_n\mu$, $\xi = \chi_n\sigma$ and $\nu = \delta/\sqrt{n}$.

(i) We have $m_\infty(\theta) = \chi_n\mu$ and $\mathrm{s}^2_\infty(\theta) = \chi_n^2\sigma^2$ by Assumption 2.2 part (i).

(ii) To see that $m_\infty(w) = 0$, we compute $m(w)$ using (27) which gives

$$m(w) = \chi_n m(\beta)\langle x_p, x_p - x_\infty\rangle + m(\mathcal{E}x_p)/\sqrt{n}$$

The first term vanishes by Proposition 4.3 since $|\langle x_p, x_p - x_\infty\rangle| \leq |x_\infty - x_p|$ and $m_\infty(\beta)$ and χ_n are finite almost surely by Assumption 2.2 part (ii). The second term vanishes because $m(\mathcal{E}x_p)$ may be written as a linear combination of a fixed n realizations of $m(\epsilon)$ with coefficients being the entries of x_p, and $|x_\infty|$ is bounded. To this end, $m(\epsilon) = m_p(\epsilon) = \sum_{i=1}^p \phi_i m(\gamma^i) = m_p(\varphi)$ which tends to zero as $p \uparrow \infty$ (almost surely) by Assumption 2.2 part (iii).

Similarly, to verify that $\mathrm{s}^2_\infty(w) = \nu$, we use (27) again to calculate that

$$\mathrm{s}^2(w) = \chi_n^2(\langle\beta,\beta\rangle/p)\langle x_p, x_p - x_\infty\rangle^2 + \langle \mathcal{E}x_p, \mathcal{E}x_p\rangle/(pn) - m^2(w)$$

where we have used the fact that $\langle\beta, \mathcal{E}\rangle = 0_n$ by Assumption 2.1. Since χ_n and the limit $\chi_n^2(\sigma^2+\mu^2)$ of $\langle\beta,\beta\rangle/p$, as above, by Proposition 4.3 the first term tends to zero as $p \uparrow \infty$. For the second term, we note that $\langle \mathcal{E}x_p, \mathcal{E}x_p\rangle$ may be written as convex combination of a fixed n realizations of $\langle\epsilon,\epsilon\rangle$ with coefficients being the entries of x_p squared, and $|x_p|^2 = |x_\infty|^2 = 1$. We have $\langle\epsilon,\epsilon\rangle/p = \sum_{i=1}^p \phi_i^2/p = \mathrm{s}^2(\phi) + m^2(\phi)$ as the $\{\gamma^i\}_{i=1}^p$ are orthonormal. By Assumption 2.2 parts (iii) and (iv), we have $\mathrm{s}^2_\infty(\phi) = \delta^2$ and $m_\infty(\phi) = 0$. It follows that the second term converges to $\delta^2|x_\infty|^2/n = \delta^2/n = \nu^2$. The last term tends to zero since $m_\infty(w) = 0$ as above, and the claim now follows.

(iii) We have $\mathrm{cov}(\theta, w) = \langle\theta, w\rangle/p - m(\theta)m(w)$ and since $m_\infty(\theta)$ is finite, the second term vanishes as $m_\infty(w)$ as above. Again by (27) and since $\langle\beta,\mathcal{E}\rangle = 0_n$,

$$\langle\theta, w\rangle/p = \chi_n^2(\langle\beta,\beta\rangle/p)\,\langle x_p, x_p - x_\infty\rangle$$

which vanishes in the limit by the same arguments as in (ii) above.

(iv) To see that $\hat\nu = \hat\nu_p$ in (20) is an asymptotically exact estimate of $\nu = \delta/\sqrt{n}$, we use Proposition 4.2. Since $(\mathrm{tr}(\mathrm{S}) - \jmath_p^2)/p = \sum_{i=2}^n \jmath^2_{p-i+1}/p$, under the hypotheses of Proposition 4.2 converges almost surely to $\delta^2(n-1)/n$, $\hat\nu_\infty = \nu$.

A. Auxiliary proofs. As shown in Section 4.3, the hypotheses of Proposition 2.5 and Proposition 2.7 (i.e., Assumptions 2.1 –2.4) guarantee that the conditions on $\{\theta_i\}_{i\geq 1}$ and $\{w_i\}_{i\geq 1}$ in Assumption 3.5 are satisfied. Consequently the proofs of these two results, as well as that of Proposition 3.7 which requires Assumption 3.5 directly, reduce to the calculations below. The proof of Lemma 3.6 is omitted as it is elementary and that of Lemma 3.4 is a direct consequence of some of the expressions below.

For any $\mathrm{c}\in\mathbb{R}$ and $\eta(\mathrm{c}) = m(\eta) + \mathrm{c}(\eta - m(\eta))$, by direct calculation

$$\mathrm{MSE}(\eta(\mathrm{c})\,|\,\theta) = (m(\eta) - m(\theta))^2 + \mathrm{s}^2(\theta) + \mathrm{c}^2\mathrm{s}^2(\eta) - 2\mathrm{c}\,\mathrm{ccov}(\eta,\theta)\,.$$

When $\mathrm{c} = 1$ for which $\eta(1) = \eta$ and applying Lemma 3.6 yields

$$\mathrm{MSE}_\infty(\eta\,|\,\theta) = \xi^2 + (\xi^2 + \nu^2) - 2\xi^2 = \nu^2 = \delta^2/n$$

which proves Proposition 2.5. The sine of the angle squared metric is computed as

$$\mathrm{SPH}(\eta(\mathrm{c})\,|\,\theta) = 1 - \left(\frac{\langle\eta(\mathrm{c}),\theta\rangle}{|\eta(\mathrm{c})||\theta|}\right)^2 = 1 - \frac{\big(m(\eta)m(\theta) + \mathrm{ccov}(\eta,\theta)\big)^2}{(m^2(\eta) + \mathrm{c}^2\mathrm{s}^2(\eta))(m^2(\theta) + \mathrm{s}^2(\theta))}\,.$$

Using the raw estimate $\eta = \eta(1)$ for which $\mathrm{c} = 1$ we deduce by Lemma 3.6 that

$$\mathrm{SPH}_\infty(\eta\,|\,\theta) = 1 - \frac{m^2 + \xi^2}{m^2 + \xi^2 + \nu^2} = \frac{\nu^2}{\xi^2 + m^2 + \nu^2} = \frac{\mathrm{r}_\infty^2}{\mathrm{SNR}^2 + \mathrm{r}_\infty^2}$$

which establishes Proposition 2.7 with SNR and r_∞ in (15) (c.f. (25)).

Note that minimizing the expressions for $\mathrm{MSE}(\eta(\mathrm{c})\,|\,\theta)$ and $\mathrm{SPH}(\eta(\mathrm{c})\,|\,\theta)$ above over $\mathrm{c}\in\mathbb{R}$ yields the $\mathrm{c}^{\mathrm{MSE}}$ and $\mathrm{c}^{\mathrm{SPH}}$ in (24) proving Lemma 3.4. The limits as $p\uparrow\infty$ of these quantities is easily verified as $\mathrm{c}_\infty = \frac{\mathrm{SNR}^2}{1+\mathrm{SNR}^2}$ in (22) using Lemma 3.6. This establishes the last part of Proposition 3.7. To prove the first part, we again apply Lemma 3.6 to deduce that $\mathrm{c} = \mathrm{c}_p = 1 - \frac{\hat{\mathrm{v}}_p}{\mathrm{s}_p^2(\eta)} \sim 1 - \frac{\nu^2}{\xi^2+\nu^2} = \mathrm{c}_\infty$ as above. Also,

$$\begin{aligned}
\mathrm{MSE}_\infty(\eta(\mathrm{c})\,|\,\theta) &= \xi^2 + \mathrm{c}_\infty^2(\xi^2 + \nu^2) - 2\mathrm{c}_\infty\xi^2\\
&= \xi^2 + \mathrm{c}_\infty\xi^2 - 2\mathrm{c}_\infty\xi^2\\
&= \xi^2(1 - \mathrm{c}_\infty)\\
&= \frac{\xi^2}{1 + \mathrm{SNR}^2} = \mathrm{c}_\infty\nu^2 = \mathrm{c}_\infty\mathrm{MSE}_\infty(\eta\,|\,\theta)
\end{aligned}$$

as required. Similarly, by Lemma 3.6, we derive that

$$\begin{aligned}
\mathrm{SPH}_\infty(\eta(\mathrm{c})\,|\,\theta) &= 1 - \frac{(m^2 + \xi^2\mathrm{c}_\infty)^2}{(m^2 + \mathrm{c}_\infty^2(\xi^2 + \nu^2))(\mu^2 + \sigma^2)} = (1 - \mathrm{c}_\infty)\mathrm{r}_\infty^2\\
&= (1 - \mathrm{c}_\infty)(\mathrm{SNR}^2 + \mathrm{r}_\infty^2)\mathrm{SPH}_\infty(\eta\,|\,\theta)\\
&= \left(\frac{\mathrm{SNR}^2 + \mathrm{r}_\infty^2}{1 + \mathrm{SNR}^2}\right)\mathrm{SPH}_\infty(\eta\,|\,\theta) = \mathrm{d}_\infty\mathrm{SPH}_\infty(\eta\,|\,\theta)
\end{aligned}$$

for d_∞ as in (23). This concludes the proof of Proposition 3.7.

References.

Ahn, J., Marron, J., Muller, K. M. & Chi, Y.-Y. (2007), 'The high-dimension, low-sample-size geometric representation holds under mild conditions', *Biometrika* **94**(3), 760–766.

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, Inc.

Aoshima, M., Shen, D., Shen, H., Yata, K., Zhou, Y.-H. & Marron, J. (2018), 'A survey of high dimension low sample size asymptotics', *Australian & New Zealand journal of statistics* **60**(1), 4–19.

Bock, M. E. (1975), 'Minimax estimators of the mean of a multivariate normal distribution', *The Annals of Statistics* pp. 209–218.

Brown, L. D. (1971), 'Admissible estimators, recurrent diffusions, and insoluble boundary value problems', *The Annals of Mathematical Statistics* **42**(3), 855–903.

Brown, L. D. & Zhao, L. H. (2012), 'A geometrical explanation of Stein shrinkage', *Statistical Science* pp. 24–30.

Donoho, D. L., Gavish, M. & Johnstone, I. M. (2018), 'Optimal shrinkage of eigenvalues in the spiked covariance model', *Annals of statistics* **46**(4), 1742.

Efron, B. & Morris, C. (1975), 'Data analysis using stein's estimator and its generalizations', *Journal of the American Statistical Association* **70**(350), 311–319.

Fan, J., Guo, J. & Zheng, S. (2020), 'Estimating number of factors by adjusted eigenvalues thresholding', *Journal of the American Statistical Association* pp. 1–10.

Fourdrinier, D., Strawderman, W. E. & Wells, M. T. (2018), *Shrinkage estimation*, Springer.

Goldberg, L., Papanicalaou, A. & Shkolnik, A. (2021), Dispersion bias. Working paper.

Goldberg, L. R., Papanicolaou, A., Shkolnik, A. & Ulucam, S. (2020), 'Better betas', *The Journal of Portfolio Management* **47**(1), 119–136.

Gruber, M. H. (2017), *Improving efficiency by shrinkage: the James–Stein and ridge regression estimators*, Routledge.

Gupta, A. K. & Peña, E. A. (1991), 'A simple motivation for James-Stein estimators', *Statistics & probability letters* **12**(4), 337–340.

Hall, P., Marron, J. S. & Neeman, A. (2005), 'Geometric representation of high dimension, low sample size data', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(3), 427–444.

Hoerl, A. E. & Kennard, R. W. (1970), 'Ridge regression: Biased estimation for nonorthogonal problems', *Technometrics* **12**(1), 55–67.

Horn, R. A. & Johnson, C. R. (2013), *Matrix analysis*, Cambridge university press.

James, W. & Stein, C. (1961), Estimation with quadratic loss, *in* 'Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics', The Regents of the University of California.

Jolliffe, I. T. & Cadima, J. (2016), 'Principal Component Analysis: a review and recent developments', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **374**(2065), 20150202.

Jung, S. & Marron, J. S. (2009), 'PCA consistency in high dimension, low sample size context', *The Annals of Statistics* **37**(6B), 4104–4130.

Jung, S., Sen, A. & Marron, J. (2012), 'Boundary behavior in high dimension, low sample size asymptotics of PCA', *Journal of Multivariate Analysis* **109**, 190–203.

Ledoit, O. & Wolf, M. (2004), 'A well-conditioned estimator for large-dimensional covariance matrices', *Journal of multivariate analysis* **88**(2), 365–411.

Paul, D. (2007), 'Asymptotics of sample eigenstructure for a large dimensional spiked covariance model', *Statistica Sinica* pp. 1617–1642.

Shen, D., Shen, H., Zhu, H. & Marron, J. (2016), 'The statistics and mathematics of high dimension low sample size asymptotics', *Statistica Sinica* **26**(4), 1747.

Stein, C. (1956), Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *in* 'Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics', The Regents of the University of California.

Stein, C. (1986), 'Lectures on the theory of estimation of many parameters', *Journal of Soviet Mathematics* **34**(1), 1373–1403.

Stigler, S. M. (1990), 'The 1988 Neyman memorial lecture: a Galtonian perspective on shrinkage estimators', *Statistical Science* pp. 147–155.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.

Wang, W. & Fan, J. (2017), 'Asymptotics of empirical eigenstructure for high dimensional spiked covariance', *Annals of statistics* **45**(3), 1342.

Weyl, H. (1912), 'Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung)', *Mathematische Annalen* **71**(4), 441–479.

Yu, Y., Wang, T. & Samworth, R. J. (2015), 'A useful variant of the Davis–Kahan theorem for statisticians', *Biometrika* **102**(2), 315–323.

Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the royal statistical society: series B (statistical methodology)* **67**(2), 301–320.