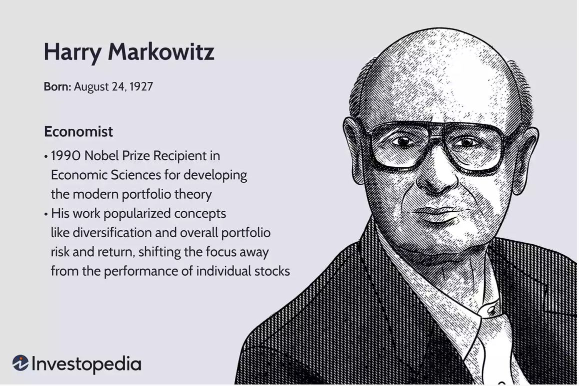
On the Markowitz Optimization Enigma

Alex Shkolnik (shkolnik@ucsb.edu)
(*joint with Hubeyb Gurdogan.*)

CDAR Risk Seminar.
UC Berkeley, CA.
September 19, 2023.

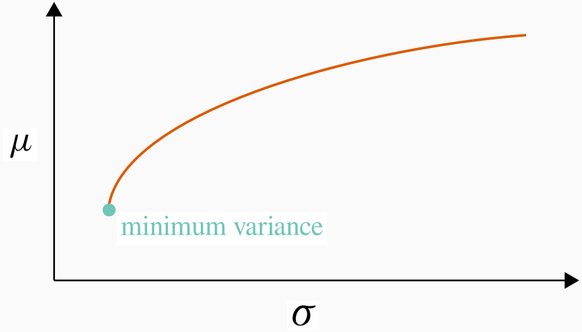
Department of Statistics & Applied Probability
University of California, Santa Barbara

In honor of Harry Markowitz, 1927–2023.



Markowitz's optimization enigma

Since Markowitz (1952), quantitative investors have constructed portfolios with mean-variance optimization.



- A simple quadratic program given a covariance matrix Σ.

The Markowitz enigma entails the observation that risk variance minimizers are, fundamentally, "estimation-error maximizers".

The Markowitz Optimization Enigma: Is 'Optimized' Optimal?

– (Michaud 1989).

The Markowitz enigma entails the observation that risk variance minimizers are, fundamentally, "estimation-error maximizers".

The Markowitz Optimization Enigma: Is 'Optimized' Optimal?
– (Michaud 1989).

A mathematical characterization (Goldberg et al. 2022).

Let Σ be a $p \times p$ covariance matrix.

$$\min_{w \in \mathbb{R}^p} \sigma^2$$
$$\sigma^2 = \langle w, \Sigma w \rangle$$
$$\langle w, \mathrm{e} \rangle = 1 \,.$$

- The vector $\mathrm{e} = (1, \ldots, 1) \in \mathbb{R}^p$.
- $\langle u, v \rangle = u^\top v$ and $|v| = \sqrt{\langle v, v \rangle}$.

Equity risk factors are high dimensional vectors.

$$\Sigma = BB^\top + \Omega$$

The columns of a $p \times q$ matrix B.

- Market
- Style
- Industry
- Climate, Tech innovation, etc

Systematic risk $BB^\top$ plus specific risk matrix Ω.

$$\Sigma = BB^\top + \Omega$$

- Specific risk may be diversified away naively,

$$\sup_p \rho(\Omega) = \sup_p \sup_{|x|=1} \langle x, \Omega x \rangle < \infty .$$

- The systematic risk may not! (i.e., eigenvalues of $BB^\top$ diverge).

Approximate factor model (Chamberlain & Rothschild 1983).

Assumption M. *$B = B_{p \times q}$ and $\Omega = \Omega_{p \times p}$ satisfy (a) and (b).*

(a) $0 < \liminf_p \inf_{|v|=1} \langle v, \Omega v \rangle < \limsup_p \sup_{|v|=1} \langle v, \Omega v \rangle < \infty$.

(b) $\lim_p (B^\top B)/p$ *exists as an invertible* $q \times q$ *matrix.*

We do not have access to Σ but to some estimated model,

$$\hat{\Sigma} = HH^\top + \hat{\omega}^2 \mathrm{I}.$$

- H is a $p \times q$ matrix estimating B.
- The simple estimate $\hat{\omega}^2 \mathrm{I}$ of Ω will suffice.

ASSUMPTION H. *$H = H_{p \times q}$ satisfy* $\limsup_p |\mathrm{e}_H| / |\mathrm{e}| < 1$ *and* $\lim_p (H^\top H)/p$ *exists as a $q \times q$ invertible matrix ($q \geq 1$ is fixed).*

Let $\hat{w} = \hat{w}(H, \nu)$ be the minimum variance portfolio w.r.t. $\hat{\Sigma}$. Let

$$\mathscr{V}_p = \sqrt{\langle \hat{w}, \Sigma \hat{w} \rangle}$$

which represents the *true risk* of the estimated portfolio $\hat{w}$.

What is the behaviour of $\mathscr{V}_p$ as p grows?

Which entries in $\hat{\Sigma}$ are the most responsible?

How does this compare to the *estimated risk* of the portfolio $\hat{w}$? i.e.,

$$\hat{\sigma}_p^2 = \langle \hat{w}, \hat{\Sigma} \hat{w} \rangle .$$

Write $f_p \asymp g_p$ for $c \leq f_g / g_p \leq C$ eventually for constants $c, C > 0$.

Proposition (Markowitz Enigma). *Suppose Assumptions M and H hold. Then for some vector function $\mathcal{E}_p(H)$ to be specified,*

$$\frac{\mathcal{V}_p^2}{\hat{\sigma}_p^2} \asymp 1 + p\,|\mathcal{E}_p(H)|^2 .$$

- The roots of $\mathcal{E}_p(\cdot)$ is not so easy to find.
- The ratio of the true risk to the actual risk diverged unless we do.

The hierarchy of challenges

The grand prize is awarded to the $H_\sharp$ for which $p\mathscr{E}_p^2(H_\sharp)$ vanishes (say, asymptotically or even p finite).

It is here that the quality of the estimate of Ω matters.

An estimate $H_\sharp$ that bounds $p\mathscr{E}_p^2(H_\sharp)$ would place second.

The third place is for estimators $H_\sharp$ for which we have,

$$\lim_{p} \mathscr{E}_{p}(H_\sharp) = 0_{q} \,.$$

This may seem insufficient but it has merit.

Whats the merit of 3rd place?

- *It may be all that can be proved but better in practice.*

But also ... write down the eigendecomposition,

$$BB^\top = \mathscr{B}\Lambda_p^2\mathscr{B}$$

and let $f_p \sim g_p$ denote $\lim_p f_p/g_p = 1$.

LEMMA. *Suppose Assumptions M and H hold. Then,* $\hat{\sigma}_p \asymp 1/\sqrt{p}$ *and*

$$\mathscr{V}_p \sim \left|(\Lambda_p/\sqrt{p})\mathscr{E}_p(H)\right|.$$

i.e. 3rd place gets us a riskless portfolio in the limit.

Today we will derive $H_\sharp$ which provably places 3rd in the hierarchy but in (Gaussian) simulation will exhibit all the properties of 2nd place.

Bad matrix

Proposition (Markowitz Enigma). *Suppose Assumptions M and H hold. Then for some vector function* $\mathscr{E}_p(H)$ *to be specified,*

$$\frac{\mathscr{V}_p^2}{\hat{\sigma}_p^2} \asymp 1 + p\,|\mathscr{E}_p(H)|^2\,.$$

What is the cause? Lets compare Σ and its estimate $\hat{\Sigma}$.

$$\Sigma = BB^\top + \Omega \qquad \text{vs} \qquad \hat{\Sigma} = HH^\top + \hat{\omega}^2 \mathrm{I}\,.$$

- The above gives away that the estmate $\hat{\omega}^2$ is not too important.
- The estimate H has $p \times q$ entries (which ones are bad?).

Recall the eigendecomposition $BB^\top = \mathscr{B}\Lambda_p^2\mathscr{B}$ in terms of which the optimization bias vector turns out to satisfy the relation,

$$\sqrt{p}\,\mathscr{E}_p(H) = \frac{\mathscr{B}^\top(\mathrm{e} - \mathrm{e}_H)}{\langle \mathrm{e}, \mathrm{e} - \mathrm{e}_H\rangle}$$

where e_H denotes a projection of e onto $\textsc{col}(H)$.

Let $\textsc{col}(A)$ denote the column span of the matrix A and let e_A denote the orthogonal projection of the vector e on $\textsc{col}(A)$, e.g.,

$$e_H = HH^\dagger e$$

where $A^\dagger = (A^\top A)^{-1} A^\top$, the Moore-Penrose of A of full column rank.

With this notation $AA^\dagger B$ projects a matrix B onto $\textsc{col}(A)$.

LEMMA. *For any invertible matrix K we have $e_H = e_{HK}$.*

Write down the following eigendecomposition,

$$HH^\top = \mathscr{H}\,\mathcal{S}_p^2\mathscr{H} = \sum_{(s^2,h)} s^2 hh^\top$$

and note that $\mathscr{H} = H\,\mathcal{S}_p^{-1}$.

Applying the lemma with $K = \mathcal{S}_p^{-1}$ and setting $z = \mathrm{e}/\sqrt{p}$,

$$\mathscr{E}_p(H) = \mathscr{E}_p(\mathscr{H}) = \frac{\mathscr{B}^\top z - (\mathscr{B}^\top\mathscr{H})\mathscr{H}^\top z}{1 - |\mathscr{H}^\top z|^2}\,.$$

The optimization bias vector function $\mathscr{E}_p(\cdot)$ is given by,

$$\mathscr{E}_p(H) = \mathscr{E}_p(\mathscr{H}) = \frac{\mathscr{B}^\top z - (\mathscr{B}^\top \mathscr{H})\mathscr{H}^\top z}{1 - |\mathscr{H}^\top z|^2}.$$

It is remarkable that this (and the true risk) depends (asymptotically) on just one of the three components that make up the model $\hat{\Sigma}$.

- $\mathscr{E}_p(H)$ *does not depend on the specific risk estimate* $\hat{\omega}^2$.
- $\mathscr{E}_p(H)$ *does not depend on lengths of the columns of* H.
- $\mathscr{E}_p(H)$ *is fully determined by* $\textsc{col}(H)$.

Bad matrix $\hat{\Sigma} = HH^\top + \hat{\omega}^2 \mathrm{I}$ mean "bad" directions for columns of H.

Game over for PCA

For $\mathrm{Y} = B\mathrm{X}^\top + \mathcal{E}$, a $p \times n$ matrix of returns to p securities,

$$\min_{\boldsymbol{H}} \left\| (\mathrm{Y} - HH^\dagger \mathrm{Y}) / \sqrt{n} \right\|_{\mathrm{F}}$$

attains a (nonunique) minimum at $H = B$ provided $\mathcal{E} = 0$.

PCA makes the solution unique by requiring orthonormal columns.
Corresponds to the q leading terms of the spectral decomposition of,

$$\mathrm{S} = \mathrm{Y}\mathrm{Y}^\top / n = \sum_{(\mathscr{s}^2, h)} \mathscr{s}^2 h h^\top$$

where the sum is over all the eigenvalue/eigenvector pairs $(\mathscr{s}^2, h)$.

Let $\mathscr{H}$ be the $p \times q$ matrix of the first q eigenvectors.

Ordering the eigenvalues of S as $\mathcal{s}^2_{1,p} \geq \mathcal{s}^2_{2,p} \geq \cdots \geq \mathcal{s}^2_{p,p}$ we have,

$$H = \mathscr{H}\mathcal{S}_p; \qquad H^\top H = \mathcal{S}^2_p,$$

where $\mathcal{S}^2_p$ is a diagonal matrix with (ordered) entries $\mathcal{s}^2_{1,p}, \ldots, \mathcal{s}^2_{q,p}$.

The average residual variance in the returns is nicely summarized by

$$\kappa^2_p = \frac{\sum_{j>q} \mathcal{s}^2_{j,p}}{n-q}.$$

The PCA model is $\hat{\Sigma} = HH^\top + \hat{\omega}^2\mathrm{I}$ where $\hat{\omega}^2 = n\kappa^2_p/p$ is an estimate with nice properties (estimates $\mathrm{Tr}(\Omega)/p$ consistently).

1. Simulate returns Y.
2. Assemble a PCA model $\hat{\Sigma}$.
3. Compute the portfolio $\hat{w}$.
   (minimizes estimated risk $\hat{\sigma}^2 = \langle \hat{w}, \hat{\Sigma}\hat{w} \rangle$ s.t. $\langle \hat{w}, \mathrm{e} \rangle = 1$).
4. Since we simulated the returns we use access to Σ to compute,

$$\frac{\mathscr{V}_p}{\hat{\sigma}_p} = \sqrt{\frac{\langle \hat{w}, \Sigma \hat{w} \rangle}{\langle \hat{w}, \hat{\Sigma} \hat{w} \rangle}} \qquad \left(= \frac{\text{true risk}}{\text{estimated risk}} \right).$$

We adopt a stylized Barra model for the factor risk matrix B.

- *First column of* B *is the market risk.*
- *Next three columns of* B *are style factors (value, momentum, etc).*
- *Next four columns of* B *are industry membership factors.*

A $q = 8$ factor model from which we simulate Gaussian returns.

The number of observations is $n = 60$ (held fixed).

p	$\hat{\sigma}_p$ (estimated risk)	$\mathscr{V}_p$ (true risk)	Ratio
500	5.30	13.88	2.64
1000	3.83	13.55	3.55
2000	2.65	12.81	4.85
4000	1.84	12.33	6.71
8000	1.32	12.45	9.46
16000	0.94	12.32	13.24

Sneak peak at the mystery estimator $\mathscr{H}_\sharp$.

p	$\hat{\sigma}_p$ (estimated risk)	$\mathscr{V}_p$ (true risk)	Ratio
500	13.69	13.13	0.96
1000	11.13	11.14	1.01
2000	8.36	8.66	1.04
4000	6.14	6.61	1.08
8000	4.54	4.82	1.06
16000	3.19	3.43	1.07

After two years of work, letting $\Psi^2 = \mathrm{I} - \kappa_p^2 \mathcal{S}_p^{-2}$ and

$$\theta = \frac{(\Psi^{-2} - \mathrm{I})\mathscr{H}^\top z}{|z - z_{\mathscr{H}}|} \in \mathbb{R}^q , \qquad a_\pm = 1 \pm \frac{1}{\sqrt{1 + |\theta|^2}} .$$

we arrive at the following for the corrected principal components,

$$\mathscr{H}_\sharp = \mathscr{H}\left(\mathrm{I} - a_+ v v^\top\right) - \sqrt{a_+ a_-}\, \frac{z - z_{\mathscr{H}}}{|z - z_{\mathscr{H}}|} v ; \qquad \left(v = \frac{\theta}{|\theta|}\right)$$

which is a $p \times q$ matrix of orthonormal columns (i.e. ${\mathscr{H}_\sharp}^\top \mathscr{H}_\sharp = \mathrm{I}$).

Theorem (Gurdogan & Shkolnik 2023+).

Suppose Assumption A (next slide) holds. Then, almost surely,

$$\lim_{\boldsymbol{p}} \mathscr{E}_{\boldsymbol{p}}(\mathscr{H}_{\sharp}) = 0_{\boldsymbol{q}}\,.$$

Recall $\mathrm{Y} = B\mathrm{X}^\top + \mathcal{E}$.

Assumption A. Assumption M on the matrices B and Ω holds. The following also hold with limits interpreted in the almost sure sense.

(a) Only Y is observed (the variables X, $\mathcal{E}$ are latent).

(b) The true number of factors q is known and $n > q$ is fixed.

(c) $\mathrm{X}^\top\mathrm{X}$ is $(q \times q)$ invertible (and does not depend on p).

(d) $\lim_p p^{-1}(\mathcal{E}^\top\mathcal{E}) = \omega^2\mathrm{I}$ for $\omega^2 = \lim_p \mathrm{Tr}(\Omega)/p \in (0, \infty)$.

(e) $\lim_p p^{-1}\|\mathcal{E}^\top B\| = 0$ for some matrix norm $\|\cdot\|$.

(f) $\limsup_p |\mathrm{e}_B|/|\mathrm{e}| < 1$ where $\mathrm{e}_B = BB^\dagger\mathrm{e}$.

The GPS program

Goldberg et al. (2022) solve this problem when $q = 1$.

- $B = \beta \in \mathbb{R}^p$ and $b = \frac{\beta}{|\beta|}$.
- $H = \beta \in \mathbb{R}^p$ and $h = \frac{\eta}{|\eta|}$.
- $z = \frac{\mathrm{e}}{|\mathrm{e}|} = (1, \ldots, 1)/\sqrt{p}$.

$$\mathscr{E}_p(h) = \frac{\langle b, z \rangle - \langle h, b \rangle \langle h, z \rangle}{1 - \langle h, z \rangle^2}$$

The unknown quantities $\langle h, b \rangle$ and $\langle b, z \rangle$ admit estimates which are functions of the sample eigenvalues (Goldberg et al. 2022, Theorem 3.1).

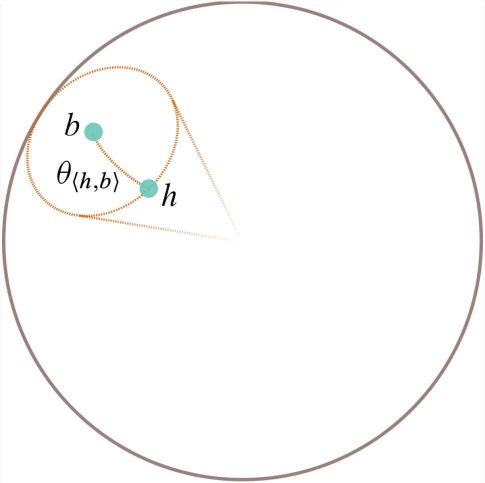
W.l.o.g. it is assumed that $\langle h, z \rangle, \langle b, z \rangle \geq 0$.

$$\psi^2 = 1 - \left(\frac{\sum_{j>1} s_{j,p}^2}{s_{1,p}^2 (n-1)} \right)$$

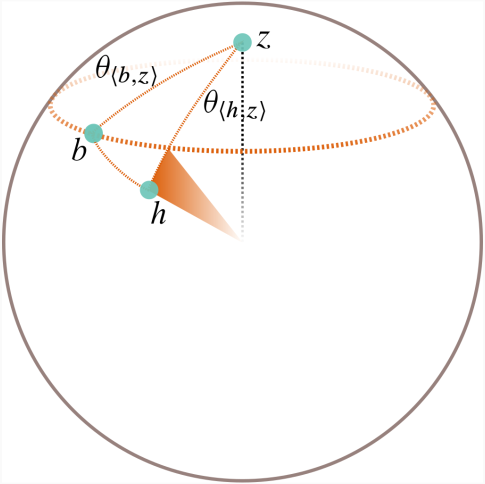
THEOREM (Goldberg et al. 2022, Theorem 3.1). Under mild assumptions (relaxed further in Assumption A) almost surely as $p \to \infty$,

$$\langle h, b \rangle \sim \psi \qquad \langle h, z \rangle \sim \langle h, b \rangle \langle b, z \rangle .$$

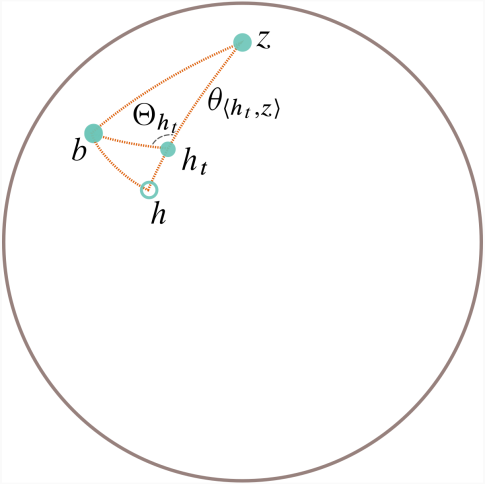
$\langle h, b \rangle \sim \psi$



$\langle h, z \rangle \sim \psi \langle b, z \rangle$



$$h_t = \frac{h+tz}{|h+tz|}$$



A lengthy but straightforward calculation with $h_t = \frac{h+tz}{|h+tz|}$ yields,

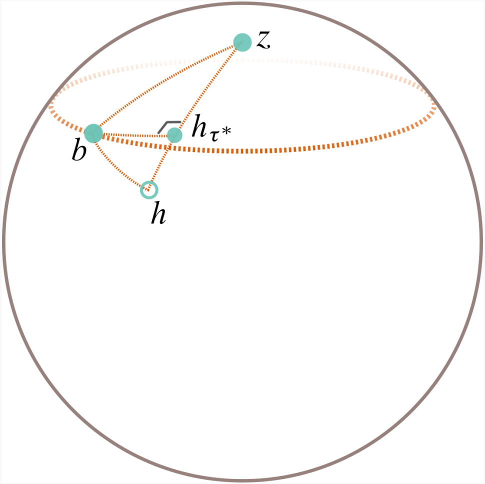
$$\mathscr{E}_p(h_t) = \mathscr{E}_p(h) - t\,\frac{\langle h, b\rangle - \langle b, z\rangle\langle h, z\rangle}{1 - \langle h, z\rangle^2}$$

revealing a root of the optimization bias at $t = \tau_*$,

$$\tau^* = \frac{\langle b, z\rangle - \langle h, b\rangle\langle b, z\rangle}{\langle h, b\rangle - \langle b, z\rangle\langle h, z\rangle}$$

with estimates for the unknowns in (Goldberg et al. 2022, Theorem 3.1).

$$\tau = \frac{\langle h, z\rangle(1 - \psi^2)}{\psi^2 - \langle h, z\rangle^2} \sim \tau^* \quad (p \uparrow \infty)\,.$$


z
h_{τ^*}
b
h

$$\mathscr{E}_p(\mathscr{H}) = \frac{\mathscr{B}^\top z - (\mathscr{B}^\top \mathscr{H})\mathscr{H}^\top z}{1 - |\mathscr{H}^\top z|^2}.$$

Do all factors matter?

Running the GPS recipe and pretend $8 = 1$ is not enough.

p	$\hat{\sigma}_p$ (estimated risk)	$\mathscr{V}_p$ (true risk)	Ratio
500	6.98	13.77	2.03
1000	5.34	13.08	2.54
2000	4.62	11.27	2.66
4000	5.14	8.04	1.60
8000	1.71	11.53	7.15
16000	1.44	10.67	8.08

More angles more problems

$F_p \sim G_p$ denotes that every $\lim_p (F_p)_{jk}/(G_p)_{jk} = 1$ (almost surely).

Theorem (Gurdogan & Shkolnik 2023+).

There does not exist a function $f : \mathbb{R}^{p\times n} \to \mathbb{R}^{q\times q}$ *with* $q \geq 2$ *for which,*

$$\mathscr{B}^\top \mathscr{H} \sim f(\mathrm{Y})$$

without "very strong assumptions" (e.g., $\mathrm{X}^\top \mathrm{X} = \mathrm{I}$*).*

- This does not mean we do not have limit theorems for the anlges between the sample and population principal component angles.

The is no way of *estimating* the optimization bias vector.

It is not clear how to find roots of $\mathscr{E}_{\boldsymbol{p}}(\cdot)$ for $q \geq 2$.

$$\mathscr{E}_{\boldsymbol{p}}(H) = \frac{\mathscr{B}^\top(\mathrm{e} - \mathrm{e}_{\boldsymbol{H}})}{\langle z, \mathrm{e} - \mathrm{e}_{\boldsymbol{H}}\rangle} = \frac{\mathscr{B}^\top z - (\mathscr{B}^\top \mathscr{H})\mathscr{H}^\top z}{1 - |\mathscr{H}^\top z|^2}.$$

The last equality only holds for $\mathscr{H}$ with orthonormal columns.

Even if we do find a root, it is likely to depend on the unknown $\mathscr{B}$ leading us back to the first problem. Let's take problem two first.

Navigating problem #2

We are going to expand $\textsc{col}(H)$ by e and define.

$$\mathscr{H}_z = \begin{pmatrix} \mathscr{H} & \dfrac{z - z_{\mathscr{H}}}{|z - z_{\mathscr{H}}|} \end{pmatrix}.$$

We try to see if a linear transformation can hit the root of $\mathscr{E}_{\boldsymbol{p}}(\cdot)$.

$$T \mapsto \mathscr{H}_z T, \qquad (T^\top T \in \mathbb{R}^{q \times q} \text{ invertible}).$$

This leads to the following transformation of the optimization bias.

$$\mathscr{E}_p(\mathscr{H}_z T) = \frac{\mathscr{B}^\top z - \mathscr{B}^\top \mathscr{H}_z T T^\dagger \mathscr{H}_z^\top z}{1 - |z_{\mathscr{H}_z T}|^2}$$

Slick observation, $T^\top T T^\dagger = T^\top$ (i.e., $T^\dagger = (T^\top T)^{-1} T^\top$).

This suggests we set $T = T_* = \mathscr{H}_z^\top \mathscr{B}$ above.

Provided $T_*^\top T_*$ is invertible and $|z_{\mathscr{H}_z T_*}| < 1$, for $T_* = \mathscr{H}_z^\top \mathscr{B}$,

$$\mathscr{E}_p(\mathscr{H}_z T_*) = \frac{\mathscr{B}^\top z - T_*^\top \mathscr{H}_z^\top z}{1 - |z_{\mathscr{H}_z T_*}|^2} = \frac{\mathscr{B}^\top z - \mathscr{B}^\top \mathscr{H}_z \mathscr{H}_z^\top z}{1 - |z_{\mathscr{H}_z T_*}|^2} = 0\,.$$

But expected, our impossibility result says T_* cannot be estimated.

Navigating problem #1

LEMMA. *For any invertible matrix K we have* $\mathrm{e}_{H} = \mathrm{e}_{HK}$.

As a corollary, for any invertible matrix K, we also have

$$\mathscr{E}(\mathscr{H}_z T_*) = \mathscr{E}(\mathscr{H}_z \mathscr{H}_z^\top \mathscr{B}) = \mathscr{E}(\mathscr{H}_z \mathscr{H}_z^\top \mathscr{B} K)$$

A good choice turns out to be $K = \mathscr{B}^\top \mathscr{H}$ because ...

Recall $\Psi^2 = \mathrm{I} - \kappa_p^2 \, \mathcal{S}_p^{-2}$.

Theorem (Gurdogan & Shkolnik 2023+).

Suppose Assumption A holds. Then, almost surely,

$$\mathscr{H}^\top \mathscr{B}\mathscr{B}^\top \mathscr{H} \sim \Psi^2 \quad \textit{and} \quad \mathscr{H}^\top z \sim (\mathscr{H}^\top \mathscr{B})\mathscr{B}^\top z\,.$$

Moreover, every diagonal entry of Ψ *is eventually in* $(0, 1)$ *wp1.*

Recall $\Psi^2 = \mathrm{I} - \kappa_p^2 \mathcal{S}_p^{-2}$.

THEOREM (Gurdogan & Shkolnik 2023+).

Suppose Assumption A holds. Then, almost surely,

$$\mathscr{H}^\top \mathscr{B}\mathscr{B}^\top \mathscr{H} \sim \Psi^2 \quad \textit{and} \quad \mathscr{H}^\top z \sim (\mathscr{H}^\top \mathscr{B})\mathscr{B}^\top z\,.$$

Moreover, every diagonal entry of Ψ *is eventually in* $(0, 1)$ *wp1.*

cf. (Goldberg et al. 2022, Theorem 3.1)

$$\langle h, b\rangle \sim \psi \qquad \text{and} \qquad \langle h, z\rangle \sim \psi\,\langle b, z\rangle$$

Recap.

- We found that $\mathscr{E}_p(\mathscr{H}_z T_*) = 0$ for $T_* = \mathscr{H}_z^\top \mathscr{B}$.
- The key lemma supplied that $\mathscr{E}_p(H) = \mathscr{E}_p(HK)$ for invertible K.
- Choosing $K = \mathscr{B}^\top \mathscr{H}$ leads to $T_{**} = T_* \mathscr{B}^\top \mathscr{H}$ with

$$T_{**} = \mathscr{H}_z \mathscr{B}\mathscr{B}^\top \mathscr{H}$$

  because we can estimate angles after projecting $\mathscr{H}$ onto $\textsc{col}(\mathscr{B})$.

$$T_\sharp \sim T_{**}\,.$$

- We truly take $K = \mathscr{B}^\top \mathscr{H} R$ for R that makes the columns orthonormal (i.e. q corrected principal components)

$$\mathscr{H}_\sharp = \mathscr{H}_z T_\sharp R\,.$$

Theorem (Gurdogan & Shkolnik 2023+).

Suppose Assumption A (next slide) holds. Then, almost surely,

$$\lim_p \mathscr{E}_p(\mathscr{H}_\sharp) = 0_q\,.$$

$$\mathscr{H}_\sharp = \mathscr{H}\left(\mathrm{I} - a_+ vv^\top\right) - \sqrt{a_+ a_-}\,\frac{z - z_{\mathscr{H}}}{|z - z_{\mathscr{H}}|}\,v\,; \qquad \left(v = \frac{\theta}{|\theta|}\right)$$

where $\mathscr{H}_\sharp^\top \mathscr{H}_\sharp = \mathrm{I}$ and the $\theta, a_\pm$ are given by

$$\theta = \frac{(\Psi^{-2} - \mathrm{I})\mathscr{H}^\top z}{|z - z_{\mathscr{H}}|} \in \mathbb{R}^q\,, \qquad a_\pm = 1 \pm \frac{1}{\sqrt{1 + |\theta|^2}}\,.$$

The optimization bias of PCA

Recall $\Psi^2 = \mathrm{I} - \kappa_p^2 \, \mathcal{S}_p^{-2}$.

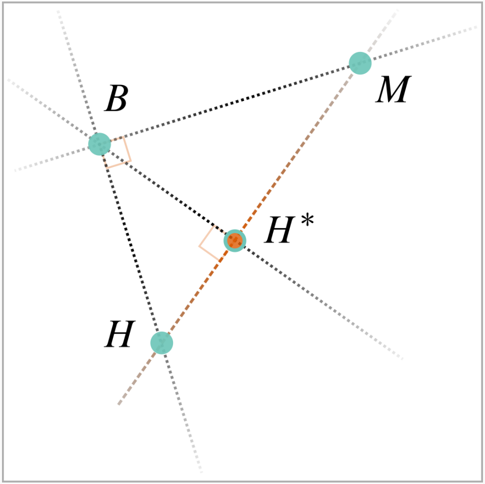
Theorem (Gurdogan & Shkolnik 2023+).

Suppose Assumption A holds. Then, almost surely,

$$|\mathcal{E}_p(\mathcal{H})| \sim \frac{\sqrt{\langle z, \mathcal{H} \, \Phi \, \mathcal{H}^\top z \rangle}}{1 - |\mathcal{H}^\top z|^2}$$

where $\Phi = 2(\Psi^{-2} - 1)(1 - \Psi^2)$ *and* $|\mathcal{E}_p(H)|$ *is eventually in* $(0, \infty)$.

The James-Stein connection


B
M
H*
H

The James-Stein estimator for PCA with shrinkage space $\textsc{col}(M)$.

$$\begin{aligned} H^{\mathrm{JS}} &= H\mathrm{C} + M(\mathrm{I} - \mathrm{C}) \\ M &= \mathscr{M}\mathscr{M}^\top \mathscr{H} \\ \mathrm{C} &= \mathrm{I} - \left((H - M)^\top (H - M)\right)^{-1} (\mathrm{I} - \Psi^2) \end{aligned}$$

References.

Chamberlain, G. & Rothschild, M. (1983), 'Arbitrage, factor structure, and mean-variance analysis on large asset markets', *Econometrica: Journal of the Econometric Society* pp. 1281–1304.

Goldberg, L., Papanicalaou, A. & Shkolnik, A. (2022), Dispersion bias.

Markowitz, H. (1952), 'Portfolio selection', *The Journal of Finance* 7(1), 77–91.

Michaud, R. O. (1989), 'The markowitz optimization enigma: Is 'optimized'optimal?', *Financial analysts journal* 45(1), 31–42.