Improving quadratic optimization inputs with concentration of measure

Lisa Goldberg(lrg@berkeley.edu)

Four University Seminar

6 September 2024

You are a junior research assistant to a trader, who asks you to quickly estimate the risk of a minimum variance portfolio in the S&P 500 with a year's worth of daily data

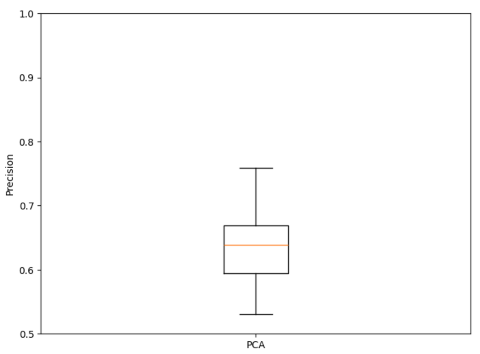
$$\min_{w \in \mathbb{R}^p} w^\top \Sigma w$$
$$w^\top \mathbf{1} = 1$$

$$w_* = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}}$$

$$V^2(w_*) = \frac{1}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}}$$

How good is your estimate?

The estimate is imprecise



Precision

$$\mathcal{P} = \frac{\text{estimated volatility}}{\text{true volatility}}$$

Data are generated from a one factor model, $r = \beta f + \epsilon$. Factor returns f are drawn independently from $N(0, 0.16^2)$ and specific returns are drawn independently from $N(0, 0.50)$. While values of β are parameters to be estimated, they are drawn from $N(1, 0.5)$. Boxplot generated with 200 simulated paths. Analysis and graphics by Rahul Vinoth.

The Bianchi experiment

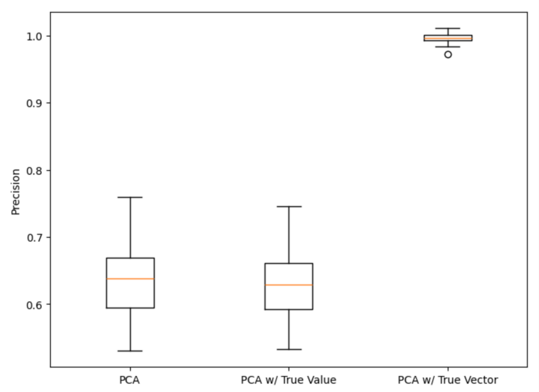
In simulation:

Replace the estimated eigenvalue with the truth, leaving the corrupted eigenvector in place

Switch the roles of eigenvalue and eigenvector in the previous step

Data are generated from a one factor model, $r = \beta f + \epsilon$. Factor returns f are drawn independently from $N(0, 0.16^2)$ and specific returns are drawn independently from $N(0, 0.50)$. While values of β are parameters to be estimated, they are drawn from $N(1, 0.5)$. Boxplot generated with 200 simulated paths. Analysis and graphics by Rahul Vinoth.

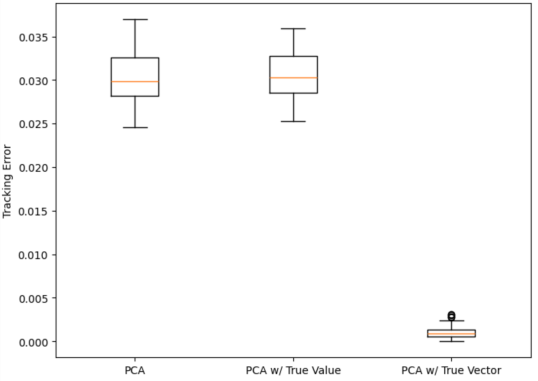
It appears to be errors in eigenvectors that lower precision



As we improve precision by correcting eigenvectors, we also move the optimized portfolio closer to the optimum

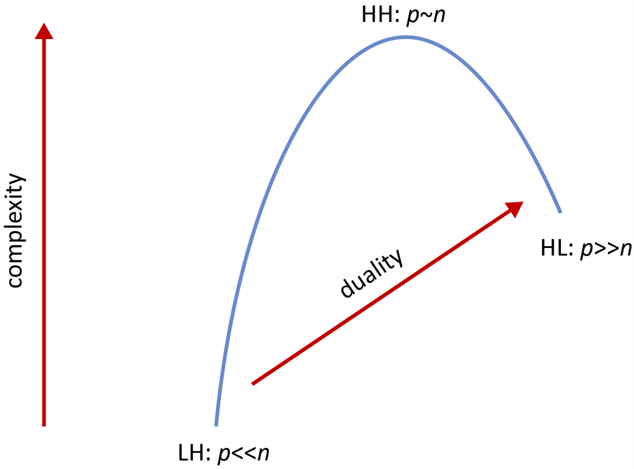
Tracking error, the workhorse of financial services, is the distance between two portfolios

$$TE^2 = (w_* - \mathbf{w})^\top \mathbf{\Sigma} (w_* - \mathbf{w})$$



Data are generated from a one factor model, $r = \beta f + \epsilon$. Factor returns f are drawn independently from $N(0, 0.16^2)$ and specific returns are drawn independently from $N(0, 0.50^2)$. While values of β are parameters to be estimated, they are drawn from $N(1, 0.5)$. Analysis and graphics by Rahul Vinoth.

We work in the high dimension low sample size (HL) regime of random matrix theory, where we benefit from duality and concentration of measure, a blessing of dimensionality



source: Goldberg & Kercheval (2023)

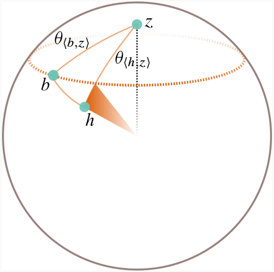
Concentration of measure implies that when $p >> n$, the sample leading eigenvector h is bound to be further away from an anchor point z than the population leading eigenvector b

In a suitable spiked factor model, almost surely as $p \to \infty$,

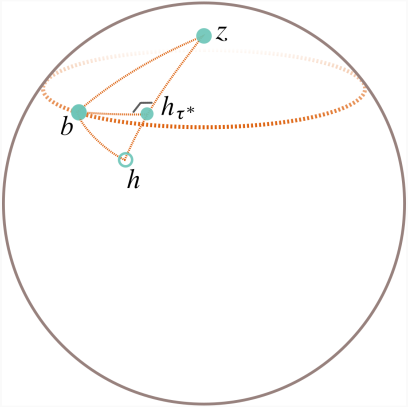
$$\langle h, z \rangle = \langle h, b \rangle \langle b, z \rangle$$

with

$$\langle h, b \rangle = \frac{\lambda^2 - \ell^2}{\lambda^2}$$



Applying the right amount of shrinkage of h toward the anchor point z yields a better estimate of b



source: Goldberg, Papanicalaou & Shkolnik (2022)

Points to remember

From the Bianchi experiment, we learn that errors in the estimated leading eigenvector appear to corrupt optimization. Eigenvalue errors appear to matter less.

Tracking error is a useful, interpretable measure of distance between two portfolios.

In the HL regime, concentration of measure implies that with high probability, an estimated eigenvector will be further from an external point than the true eigenvector.

data-driven correction of eigenvector biases in the HL regime

References

Goldberg, L. R. & Kercheval, A. N. (2023), 'James Stein for the leading eigenvector', *Proceedings of the National Academy of Sciences* **120**.

Goldberg, L. R., Papanicalaou, A. & Shkolnik, A. (2022), 'The dispersion bias', *SIAM Journal on Financial Mathematics* **13**(2), 521–550.