Discussion: "Vintage Factor Analysis with Varimax Performs Statistical Inference" by Rohe and Zeng

Yunxiao Chen[1] and Gongjun Xu[2]

London School of Economics and Political Science[1] and University of Michigan[2]

Y.Chen186@lse.ac.uk[1] and gongjun@umich.edu[2]

We congratulate Rohe and Zeng for this insightful paper that elegantly connects psychometric methods and statistical and machine learning applications. We would like to mention several lines of related research. First, the problem is closely related to the latent variable selection problems (e.g., Chen et al., 2015; Xu and Shang, 2018), where regularised estimation procedures are proposed to learn sparse loading structures. In fact, the vsp procedure can be viewed as the limiting case of a regularised estimation procedure, in the sense that $\hat{U}$ from Algorithm vsp is the limit of

$$\hat{U}^{\lambda} = \arg\min_{U} \left(\min_{D,V} \|\tilde{A} - UDV^T\|_F^2 - \lambda \left(\sum_{l=1}^{k} \frac{1}{n} \sum_{i=1}^{n} \left([U]_{il}^4 - \left(\frac{1}{n} \sum_{q=1}^{n} [U]_{ql}^2 \right)^2 \right) \right) \right), \lambda > 0$$

when λ goes to zero if the solution path is smooth, where U and V satisfy the same constraints as in SVD while D is allowed to be non-diagonal (Chen and Rohe, 2020). Note that the regularisation is used to learn the sparse loading structure rather than to avoid over-fitting, and thus, it does not require the tuning parameter to depend on the noise level. We agree that rotation is more convenient under many models, but the regularised estimation approach might be more general for some more complex latent variable models.

Second, various rotation methods have been proposed in the psychometric literature to find simple and scientifically meaningful factor loading structures. For example, consider the L_1 criterion that minimises the objective function

$$c(R, U) = \sum_{i=1}^{n} \sum_{l=1}^{k} |[UR]_{il}|.$$

This criterion is closely related to L_1 regularisation and ensures statistical consistency under suitable conditions (Jennrich, 2004, 2006; Liu et al., 2022). We run a small simulation study to compare the Varimax and L_1 rotations, where data are generated from the current factor model. Two settings are used to generate Z – one sparse setting where $P([Z]_{ij} = 0) = 0.5$ and one dense setting where $[Z]_{ij}$ follows a heavy-tail distribution. The L_1 rotation method only replaces $v(R, \hat{U})$ in step 3 of the vsp algorithm with $c(R, \hat{U})$. The mean squared errors for the estimation of Z are given in Figure 1, where the L_1 rotation performs better under the sparse setting while the vsp outperforms under the dense setting.

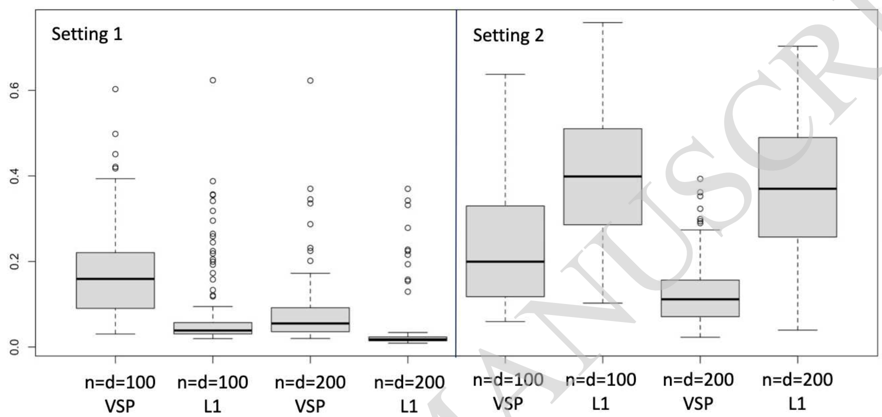

Figure 1: Box plots of mean squared errors $\|\hat{Z} - ZP_n\|_F^2/(nk)$ from 100 simulations. The four box plots corresponds to the combinations of two settings ($k = 5, n = d = 100, 200$) and two rotation methods (Varimax and L_1). In the simulations, we generate $A = ZY^T + W$, where $[A]_{ij}$ and $[W]_{ij}$ are independent standard normal variables. Under Setting 1, $[Z]_{ij} = \sqrt{2}[C]_{ij}[S]_{ij}$, where $[C]_{ij}$ are independent standard normal random variables and $[S]_{ij}$ are independent Bernoulli random variables with success probability 0.5. Under Setting 2, $[Z]_{ij} = [T]_{ij}/\sqrt{5/3}$, where $[T]_{ij}$ follows a t distribution with 5 degrees of freedom. R code for the simulation can be found on `https://stats.lse.ac.uk/cheny185/L1_rotation.R`.

Finally, another interesting extension of the current work is to non-linear factor models that assume $\mathbb{E}([A]_{ij}|Z, B, Y) = f([ZBY^T]_{ij})$, for some known smooth and strictly monotone non-linear function f (e.g., logistic function for binary data). Due to the non-linear transformation, the current vsp procedure does not directly apply. One solution is to first apply the universal singular value thresholding procedure (Chatterjee, 2015) to A to estimate $(f([ZBY^T]_{ij}))_{n\times d}$, which yields an estimate of ZBY^T through element-wise f^{-1} transformations; see Zhang et al. (2020) for more details and the related consistency theory. Then, one can learn Z by steps 2 and 3 of Algorithm vsp.

References

Chatterjee, S. (2015). Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, 43(1):177–214.

Chen, F. and Rohe, K. (2020). A new basis for sparse principal component analysis. *arXiv preprint arXiv:2007.00596*.

Chen, Y., Liu, J., Xu, G., and Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, 110(510):850–866.

Jennrich, R. I. (2004). Rotation to simple loadings using component loss functions: The orthogonal case. *Psychometrika*, 69(2):257–273.

Jennrich, R. I. (2006). Rotation to simple loadings using component loss functions: The oblique case. *Psychometrika*, 71(1):173–191.

Liu, X., Wallin, G., Chen, Y., and Moustaki, I. (2022). Rotation to sparse loadings using L^p losses and related inference problems. *arXiv preprint arXiv:2206.02263*.

Xu, G. and Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association*, 113(523):1284–1295.

Zhang, H., Chen, Y., and Li, X. (2020). A note on exploratory item factor analysis by singular value decomposition. *Psychometrika*, 85(2):358–372.