importers. Analyses of such data are often based on matrix factorisation models like the one presented by Rohe and Zeng:

$$A \approx ZBY^\top$$

$$a_{i,j} = z_i^\top B y_j,$$

where A is the data matrix, the values of $z_1, \ldots, z_n$ represent heterogeneity along the rows, and $y_1, \ldots, y_d$ represent heterogeneity along the columns. As with the identification of the loading matrix in factor analysis, the conundrum of matrix-variate data analysis is that there are infinitely many matrix factorizations that give the same low-rank least-squares approximation to A. How to select from among them? One approach is to abandon least-squares and instead infer the correct factorisation based on specific distributional assumptions about the latent row and column factors, often using a parametric statistical model. Such approaches can incorporate subject-matter knowledge about the factors into the estimation procedure, but are typically very computationally intensive. Alternatively, standard matrix factorisation methods, such as the singular value decomposition, are relatively inexpensive computationally, but they select a factorisation using arbitrary identifiability constraints that are not derived from the data.

What has been needed is a data analysis method that is computationally inexpensive, but also identifies the factors using information from the data. I propose a vote of thanks to Rohe and Zeng for providing just such a method. As they show in their article, the classic VARIMAX criterion, applied to rows or columns of a data matrix, can identify rotations that recover non-Gaussian latent factors. Their results unify several multivariate statistical methods, and highlight that much of what might be thought of as multivariate data analysis should really be considered as matrix-variate data analysis.

Conflict of interest: None declared.

References

Kaiser H. F. (1958). The VARIMAX criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200. https://doi.org/10.1007/BF02289233

https://doi.org/10.1093/jrsssb/qkad030
Advance access publication 4 April 2023

Seconder of the vote of thanks to Rohe & Zeng and contribution to the Discussion of 'Vintage Factor Analysis with Varimax Performs Statistical Inference'

Marianna Pensky

Department of Mathematics, University of Central Florida, Orlando, United States

Address for correspondence: Marianna Pensky, Department of Mathematics, University of Central Florida, Orlando, FL 32817, USA. Email: marianna.pensky@ucf.edu

We would like to congratulate the authors on publication of a truly seminal paper. Indeed, they managed to accomplish a rare and extremely valuable task: take a technique, Varimax, that has been used for half a century for generating sparse PCA, provide conditions for its applicability

Table 1. Δ_Z for the spectral clustering in Lei and Rinaldo (2015), vsp and adjusted vsp, averaged over 1,000 runs (standard deviations in parentheses)

Estimation in the stochastic block model					
n	a	w	Clustering	vsp	Adjusted vsp
100	0.5	0.6	0.1840 (0.0554)	0.2538 (0.0202)	0.1834 (0.0577)
200	0.5	0.6	0.0396 (0.0421)	0.1776 (0.0095)	0.0404 (0.0428)
300	0.5	0.6	0.0052 (0.0170)	0.1448 (0.0065)	0.0052 (0.0170)
400	0.5	0.6	0.0004 (0.0047)	0.1248 (0.0045)	0.0004 (0.0047)
500	0.5	0.6	0.0000 (0.0014)	0.1118 (0.0037)	0.0000 (0.0000)
100	0.5	0.8	0.5904 (0.0723)	0.5732 (0.0775)	0.5893 (0.0732)
200	0.5	0.8	0.3940 (0.0449)	0.3805 (0.0316)	0.3926 (0.0447)
300	0.5	0.8	0.2772 (0.0321)	0.3012 (0.0155)	0.2765 (0.0321)
400	0.5	0.8	0.2004 (0.0257)	0.2573 (0.0105)	0.2002 (0.0258)
500	0.5	0.8	0.1476 (0.0224)	0.2285 (0.0077)	0.1479 (0.0225)
100	0.25	0.6	0.4877 (0.0787)	0.4736 (0.0748)	0.4848 (0.0795)
200	0.25	0.6	0.2678 (0.0383)	0.3041 (0.0184)	0.2667 (0.0383)
300	0.25	0.6	0.1600 (0.0301)	0.2431 (0.0106)	0.1603 (0.0306)
400	0.25	0.6	0.0989 (0.0275)	0.2097 (0.0076)	0.0996 (0.0271)
500	0.25	0.6	0.0600 (0.0275)	0.1868 (0.0058)	0.0601 (0.0274)
100	0.25	0.8	0.6633 (0.0335)	0.6655 (0.0377)	0.6626 (0.0338)
200	0.25	0.8	0.6502 (0.0396)	0.6421 (0.0461)	0.6502 (0.0393)
300	0.25	0.8	0.6010 (0.0562)	0.5824 (0.0632)	0.6009 (0.0564)
400	0.25	0.8	0.5112 (0.0526)	0.4858 (0.0528)	0.5112 (0.0527)
500	0.25	0.8	0.4345 (0.0319)	0.4138 (0.0243)	0.4340 (0.0318)

and produce the error bounds. They name this new version Vintage Sparse PCA (vsp). In particular, if $X = ZBY^T$, where components of $Z = \{Z_{i,j}\}$ and $Y = \{Y_{i,j}\}$ are independent zero mean unit variance leptokurtic random variables, and rows of matrices Z and Y are identically distributed, then matrices Z and Y are identifiable, and Varimax allows one to do this. The paper provides very elegant arguments why kurtosis $\kappa > 3$ leads to identifiability of matrices Z and Y. Applications of vsp include, among others, Independent Component analysis, Stochastic Block Model (SBM), Degree-Corrected Stochastic Block Model (DCBM), Overlapping, Mixed Membership and Degree-Corrected Mixed Membership Stochastic Block Models, and sparse dictionary learning.

Since each of the above research areas developed its own techniques, it would be interesting to see how vsp performs for specific types of problems. The authors do not provide any numerical examination of the precision of the vsp in various scenarios (due to their sheer multitude and the fact that the complete paper is already over 100 pages). Therefore, we carry out a limited simulation study that complements the paper.

Specifically, we study three simulations scenarios. Scenario 1 considers SBM with $k = 2$ communities, where Z is a clustering matrix with exactly one 1 per row and $Y = Z$. Scenario 2 examines DCBM, where again $Z = Y$ and matrix $Z = \Theta W$, Θ is a diagonal and W is a clustering matrix. Scenario 3 deals with matrices $Z \in \mathbb{R}^{n\times k}$ and $Y \in \mathbb{R}^{n\times d}$ comprised of independent T random variables with ν degrees of freedom. In the first two scenarios, we generated clusters using multinomial distribution with equal probabilities. Elements of Θ are generated as Uniform on [0, 1]. For SBM and DCBM, the diagonal and nondiagonal elements of matrix B are, respectively, equal to a and wa. For Scenario 3, elements of B are Uniform on [0, 1].

In order to make matrices Z and Y identifiable, we renormalise Z to have column norms $\sqrt{n}$ with the respective readjustment of matrix B. We choose $k = d = 2$, vary n, w, a, and set $X = ZBY^T$. Since $\mathbb{E}(\Theta_{i,i}) = 0.5$, values $a = 0.5$ and $a = 0.25$ for SBM corresponds to $a = 1.0$ and $a = 0.5$ for

Table 2. Δ_Z for the spectral clustering in Gao et al. (2018), vsp and adjusted vsp, averaged over 1,000 runs (standard deviations in parentheses)

Estimation in the degree corrected stochastic block model					
n	a	w	Clustering	vsp	Adjusted vsp
100	1.0	0.6	0.4349 (0.1014)	0.4527 (0.0627)	0.4374 (0.0806)
200	1.0	0.6	0.2277 (0.0462)	0.3138 (0.0251)	0.2683 (0.0327)
300	1.0	0.6	0.1543 (0.0280)	0.2555 (0.0161)	0.2055 (0.0188)
400	1.0	0.6	0.1185 (0.0176)	0.2209 (0.0117)	0.1718 (0.0122)
500	1.0	0.6	0.0973 (0.0131)	0.1975 (0.0093)	0.1505 (0.0091)
100	1.0	0.8	0.8816 (0.0941)	0.8611 (0.0963)	0.8857 (0.0981)
200	1.0	0.8	0.7075 (0.1113)	0.6810 (0.1061)	0.7026 (0.1125)
300	1.0	0.8	0.5224 (0.0687)	0.5253 (0.0446)	0.5267 (0.0561)
400	1.0	0.8	0.4108 (0.0510)	0.4488 (0.0286)	0.4321 (0.0385)
500	1.0	0.8	0.3392 (0.0404)	0.3991 (0.0215)	0.3706 (0.0290)
100	0.5	0.6	0.8321 (0.1152)	0.8077 (0.1215)	0.8318 (0.1232)
200	0.5	0.6	0.5725 (0.0862)	0.5545 (0.0622)	0.5691 (0.0744)
300	0.5	0.6	0.4073 (0.0546)	0.4382 (0.0301)	0.4255 (0.0416)
400	0.5	0.6	0.3116 (0.0406)	0.3767 (0.0206)	0.3468 (0.0291)
500	0.5	0.6	0.2489 (0.0327)	0.3366 (0.0161)	0.2967 (0.0221)
100	0.5	0.8	0.9418 (0.0566)	0.9393 (0.0580)	0.9579 (0.0577)
200	0.5	0.8	0.9407 (0.0530)	0.9291 (0.0585)	0.9539 (0.0570)
300	0.5	0.8	0.9148 (0.0638)	0.8978 (0.0743)	0.9263 (0.0710)
400	0.5	0.8	0.8718 (0.0779)	0.8467 (0.0873)	0.8780 (0.0845)
500	0.5	0.8	0.7981 (0.0861)	0.7660 (0.0920)	0.7991 (0.0913)

Table 3. Δ_Z and Δ_Y, averaged over 1,000 runs (standard deviations in parentheses), for the vsp in the case of T distribution with ν degrees of freedom and no random errors

Estimation for T-random matrices, no noise					
n	d	ν	κ	Δ_Z	Δ_Y
100	200	5	9	0.1712 (0.1415)	0.1197 (0.1058)
200	400	5	9	0.1229 (0.1091)	0.0869 (0.0753)
300	600	5	9	0.0984 (0.0830)	0.0688 (0.0541)
400	800	5	9	0.0839 (0.0726)	0.0589 (0.0516)
500	1,000	5	9	0.0772 (0.0626)	0.0540 (0.0456)
100	200	10	4	0.2586 (0.1867)	0.2102 (0.1699)
200	400	10	4	0.2068 (0.1663)	0.1508 (0.1399)
300	600	10	4	0.1743 (0.1575)	0.1332 (0.1296)
400	800	10	4	0.1519 (0.1451)	0.1228 (0.1322)
500	1,000	10	4	0.1338 (0.1245)	0.1182 (0.1419)
100	200	16	3.5	0.2886 (0.2033)	0.2616 (0.1924)
200	400	16	3.5	0.2718 (0.2025)	0.2253 (0.1889)

(*continued*)

Table 3. Continued

Estimation for T-random matrices, no noise					
n	d	ν	κ	Δ_Z	Δ_Y
300	600	16	3.5	0.2439 (0.1925)	0.2279 (0.1968)
400	800	16	3.5	0.2253 (0.1878)	0.2375 (0.2175)
500	1,000	16	3.5	0.2255 (0.1944)	0.2762 (0.2412)
100	200	28	3.25	0.3378 (0.2094)	0.3168 (0.2130)
200	400	28	3.25	0.3287 (0.2125)	0.2902 (0.2104)
300	600	28	3.25	0.3045 (0.2157)	0.3055 (0.2193)
400	800	28	3.25	0.2932 (0.2089)	0.3616 (0.2345)
500	1,000	28	3.25	0.2966 (0.2153)	0.4184 (0.2430)

Table 4. Δ_Z and Δ_Y, averaged over 1,000 runs (standard deviations in parentheses), for the vsp in the case of T distribution with $\nu = 5$ degrees of freedom and iid Gaussian random errors with zero mean and standard deviation σ

Estimation for T-random matrices, $\nu = 5$, noise level σ				
n	d	σ	Δ_Z	Δ_Y
100	200	0.1	0.2209 (0.2064)	0.2022 (0.2090)
200	400	0.1	0.1721 (0.1799)	0.1399 (0.1753)
300	600	0.1	0.1355 (0.1634)	0.1198 (0.1614)
400	800	0.1	0.1246 (0.1626)	0.1104 (0.1657)
500	1,000	0.1	0.1027 (0.1241)	0.0926 (0.1317)
100	200	0.2	0.2850 (0.2540)	0.2665 (0.2553)
200	400	0.2	0.1870 (0.1934)	0.1757 (0.1918)
300	600	0.2	0.1641 (0.1968)	0.1576 (0.2020)
400	800	0.2	0.1486 (0.1912)	0.1446 (0.1959)
500	1,000	0.2	0.1376 (0.1811)	0.1333 (0.1873)
100	200	0.3	0.3178 (0.2716)	0.3077 (0.2718)
200	400	0.3	0.2342 (0.2454)	0.2306 (0.2486)
300	600	0.3	0.1923 (0.2262)	0.1960 (0.2337)
400	800	0.3	0.1701 (0.2061)	0.1736 (0.2129)
500	1,000	0.3	0.1503 (0.1868)	0.1563 (0.2009)
100	200	0.4	0.3604 (0.2886)	0.3537 (0.2940)
200	400	0.4	0.2624 (0.2659)	0.2633 (0.2684)
300	600	0.4	0.2292 (0.2521)	0.2331 (0.2617)
400	800	0.4	0.1959 (0.2368)	0.1991 (0.2424)
500	1,000	0.4	0.1725 (0.2184)	0.1822 (0.2262)

DCBM. We generate A as symmetric matrix with independent Bernoulli entries for SBM and DCBM, while $A = X + \sigma\Xi$ where Ξ has iid standard Gaussian entries for Scenario 3.

In Scenarios 1 and 2, we compare vsp with the spectral clustering algorithms in Lei and Rinaldo (2015) and Gao et al. (2018), respectively, where matrix $\hat{Z}$ is based on clustering assignment and estimator $\hat{\Theta}$ of Θ. We add the third estimator, adjusted vsp, which leaves only the largest (in absolute value) element of the vsp estimator $\hat{U}$ in each row and renormalise $\hat{Z}$ accordingly. For DCBM, we adjust $\hat{Z}$ to the column norms $\sqrt{n}$. Note that, for Scenarios 1 and 2, all three algorithms recover Z perfectly if matrix X is available. Scenario 3 is remarkably different from 1 and 2 since vsp does not recover Z and Y exactly from matrix X, and there is no 'yardstick' algorithm for comparison. Hence, for Scenario 3, we study performance of vsp only, for both X and A, which corresponds to $\sigma = 0$ and $\sigma > 0$.

Results of simulations are presented in Tables 1–4. The errors are measured as Frobenius norms $\Delta_Z = \|\hat{Z} - Z\|_F/\sqrt{n\,k}$ and $\Delta_Y = \|\hat{Y} - Y\|_F/\sqrt{n\,d}$, averaged over 1,000 runs. The standard deviations of the means are reported in parentheses.

Tables 1 and 2 confirm that the algorithms designed specifically for SBM and DCBM have better precision that vsp since they 'know' that matrix Z has only one nonzero element per row. However, adjusted vsp, which makes use of this information, performs very similarly to algorithms specifically designed for SBM and DCBM, with clustering algorithm of Gao et al. (2018) being slightly more precise in the case of DCBM. Hence, adjusted vsp can be used for clustering in the SBM (with average miss-classification proportion Δ_Z^2). The errors grow as a decreases and w increases due, respectively to sparsity increase and decline of the signal-to-noise ratio.

Table 3 shows that, as ν grows and kurtosis $\kappa = 3 + 6/(\nu - 4)$ decreases, precision of the vsp declines, even when exact matrix X is available. Therefore, for $\sigma > 0$, we carry out simulations only with $\nu = 5$ ($\kappa = 9$). We set $d = 2n$ for various choices of n. Tables 3 and 4 demonstrate that small kurtosis can be as much of a problem for recovering Z and X as noise. Indeed, errors for small κ do not decline as n and d grow as they do for larger κ and σ.

Conflict of interest: None declared.

References

Gao C., Ma Z., Zhang A. Y., & Zhou H. H. (2018). Community detection in degree-corrected block models. *Annals of Statistics*, *46*(5), 2153–2185. https://www.jstor.org/stable/26542860

Lei J., & Rinaldo A. (2015). Consistency of spectral clustering in stochastic block models. *Annals of Statistics*, *43*(1), 215–237. https://doi.org/10.1214/14-AOS1274

The vote of thanks was passed by acclamation.

https://doi.org/10.1093/jrsssb/qkad031
Advance access publication 6 April 2023

Joshua Cape's contribution to the Discussion of 'Vintage Factor Analysis with Varimax Performs Statistical Inference' by Rohe & Zeng

Joshua Cape

Department of Statistics, University of Wisconsin–Madison

Address for correspondence: Joshua Cape, 1250A Medical Sciences Center, 1300 University Avenue, Madison, WI 53706, USA. Email: jrcape@wisc.edu

I congratulate Professor Rohe and Dr Zeng on their illuminating paper. Their broad contributions will no doubt redouble contemporary research activity in multivariate analysis for years to come. Even the paper's appendices are full of valuable gems, not to be overlooked by readers.