A Comparison of Penalized and Penalty-Free Procedures for Sparse Factor Analysis

Kohei Adachi, Osaka University, Japan

1. Latent Variable Factor Analysis (LVFA) and Matrix Decomposition FA (MDFA)

Let $\mathbf{x}$ ($p \times 1$) denote the random vector containing observed variables with $E[\mathbf{x}]$ being the zero vector. In the traditional formulation of latent variable factor analysis (LVFA), the common and unique factor score vectors, which we express as $\mathbf{f}$ ($m \times 1$) and $\mathbf{u}$ ($p \times 1$), respectively, are treated as latent random vectors: FA is modeled as $\mathbf{x} = \mathbf{\Lambda f} + \mathbf{\Psi u}$ with $m < p$. Here, $\mathbf{\Lambda}$ ($p \times m$) contains factor loadings and $\mathbf{\Psi}$ ($p \times p$) is diagonal (Adachi, 2019).

On the other hand, matrix decomposition FA (MDFA) is formulated with the common and unique factor scores treated as the fixed parameters contained in $\mathbf{F}$ ($n \times m$) and $\mathbf{U}$ ($n \times p$), respectively: FA is modeled as $\mathbf{X} = \mathbf{F\Lambda}' + \mathbf{U\Psi} + \mathbf{E}$ for $n \times p$ column centered data matrix $\mathbf{X}$, with $\mathbf{E}$ containing errors (Adachi, 2019).

Recently, sparse FA (SFA) procedures have been developed for estimating sparse $\mathbf{\Lambda}$. They can be classified into two types. One of them is formulated by incorporating a penalty function in LVFA. We call this approach penalized sparse LVFA (PS-LVFA). The other is based on MDFA and penalty-free (i.e., does not use a penalty function). It is referred to as cardinality-constrained MDFA (CC-MDFA), as the un-sparsity or cardinality of $\mathbf{\Lambda}$ is directly constrained. In this paper, PS-LVFA and CC-MDFA are reviewed in the next sections, and then compared empirically and in usability.

2. Penalized Sparse LVFA (PS-LVFA)

Hirose and Yamamoto (2014) have proposed a PS-LVFA procedure, in which the penalized log likelihood

$$l(\mathbf{\Lambda},\mathbf{\Psi},\mathbf{\Phi}) = \tfrac{n}{2}\log\left|(\mathbf{\Lambda\Phi\Lambda}'+\mathbf{\Psi}^2)^{-1}\mathbf{V}\right| - \tfrac{n}{2}\operatorname{tr}(\mathbf{\Lambda\Phi\Lambda}'+\mathbf{\Psi}^2)^{-1}\mathbf{V} - nMC(\mathbf{\Lambda}\,|\,w, u) \tag{1}$$

following from the normality for $\mathbf{f}$ and $\mathbf{u}$, is maximized over $\mathbf{\Lambda}$, $\mathbf{\Psi}$ and a factor correlation matrix $\mathbf{\Phi}$ ($m \times m$) with the EM algorithm. Here, $MC(\mathbf{\Lambda}\,|\,w, u)$ is the penalty function based on MC+ with w and u the tuning parameters specifying $MC(\mathbf{\Lambda}\,|\,w, u)$. As PS-LVFA, we consider the above procedure with only the algorithm for $\mathbf{\Phi}$ in the M-step differing from the original one: $\mathbf{\Phi}$ is reparameterized as $\operatorname{diag}(\mathbf{T}'\mathbf{T})^{-1/2}\mathbf{T}'\mathbf{T}\operatorname{diag}(\mathbf{T}'\mathbf{T})^{-1/2}$ and a simple gradient method is used for updating $\mathbf{\Phi}$, with the differential of the objective function w.r.t. $\mathbf{T}$ given numerically.

3. Cardinality-Constrained MDFA (CC-MDFA)

In Adachi and Trendafilov's (2015) CC-MDFA, least squared function

$$f(\mathbf{F}, \mathbf{U}, \mathbf{\Lambda}, \mathbf{\Psi}) = \|\mathbf{X} - (\mathbf{F\Lambda}' + \mathbf{U\Psi})\|^2 = \|\mathbf{X} - (\mathbf{FA}' + \mathbf{U\Psi})\|^2 + n\|\mathbf{\Lambda} - \mathbf{A}\|^2 \tag{2}$$

is minimized over $\mathbf{F}$, $\mathbf{U}$, $\mathbf{\Lambda}$, and $\mathbf{\Psi}$ subject to $n^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_m$, $n^{-1}\mathbf{U}'\mathbf{U} = \mathbf{I}_p$, $\mathbf{F}'\mathbf{U}$ being the zero matrix, and the cardinality of Λ equaling c (a specified integer), with $\mathbf{A} = n^{-1}\mathbf{XF}$. The constraint $n^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_m$ allows the last identity in (2) to hold. It shows that the update of Λ can be attained by the constrained minimization of a simple function $\|\mathbf{\Lambda} - \mathbf{A}\|^2$. This is a trick in CC-MDFA, but follows from $n^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_m$ implying that the common factors are comnstrained to be mutually uncorrelated: $\mathbf{\Phi}$ cannot be estimated.

4. Comparison of PS-LVFA and CC-MDFA

Suitable values of w and u in PS-LVFA can be selected using BIC. Though CC-MDFA is not based on maximum likelihood (ML), the BIC-based selection of c is feasible, using the ML-FA solution with the loadings estimated as zeros in CC-MDFA being constrained to be zeros. Through such tuning parameter selection, PS-LVFA and CC-MDFA are found to provide similar solutions for a number of data sets.

PS-LVFA is superior in that $\mathbf{\Phi}$ can be estimated. On the other hand, CC-MDFA is superior in that its tuning parameter c is restricted to an integer within a certain interval, thus the suitability of all candidates for c can be assessed. Such an assessment is impossible in PS-LVFA, since its tuning parameters w and u take real values. Thus, only some representative w and u values can be evaluated, and further the combination of the two parameters must also be considered which is time-consuming.

References

Adachi, K. (2019). Factor analysis: latent variable, matrix decomposition, and constrained uniqueness formulations. *WIREs Computational Statistics,* https://doi.org/10.1002/wics.1458.

Adachi, K., & Trendafilov, N.T. (2015). Sparse orthogonal factor analysis. In E. Carpita, E. Brentari, & E.M. Qannari, (Eds.), *Advances in latent variable: Methods, models, and Applications*, pp. 227–239, Springer.

Hirose, K., & Yamamoto, M. (2014). Estimation of an oblique structure via penalized likelihood factor analysis. *Computational Statistics and Data Analysis*, **79**, 120–132.