Seconder's discussion of "Large covariance estimation by thresholding principal orthogonal complements" by Fan, Liao and Mincheva

Piotr Fryzlewicz   Na Huang
Department of Statistics
London School of Economics
Houghton Street
London WC2A 2AE
UK

February 25, 2013

We would like to start by congratulating Professors Fan, Liao and Mincheva for the stimulating and thought-provoking article.

The POET estimator is the sum of two parts: the non-sparse, low-rank part resulting from the factor model, and the sparse part arising as a result of thresholding the "principal orthogonal complement". The estimator has been designed with a particular factor model in mind, and therefore it is natural to ask, firstly, whether and how one could verify this model assumption, and secondly, whether POET offers acceptable performance if the assumption does not hold.

We may be wrong here, but we are unaware of a reliable technique for estimating the number of factors K which works well except in the most "textbook" cases of the first few eigenvalues being "visibly" larger than others. Even if a factor structure is present, the presence of both stronger and less strong factors may well lead to the cut-off in the eigenvalues being less obvious, in which case any inference for the number of factors may not be reliable. However, it is important to get K right from the point of view of the usability of POET: the authors warn us that POET may perform poorly if K is underestimated. It is therefore tempting to ask whether POET may benefit from averaging over K as a possible guard against picking one "wrong" (e.g. underestimated) value of K. Averaging may also be beneficial in cases when the factor model assumption is not satisfied.

An appealing aspect of the construction of POET is the inclusion of the non-sparse part (which is done in case the target matrix Σ is not sparse) and the sparse part (to ensure the invertibility of the estimator). It is tempting to consider other possible estimators along these lines. Motivated by POET, we propose an estimator of Σ of the form

$$\hat{\Sigma}^N = \delta\,\hat{\Sigma}_{\text{sam}} + (1-\delta)\,t(\hat{\Sigma}_{\text{sam}}, \lambda),$$

	Σ			Σ^{-1}		
	$\delta = 0$	$\delta = \frac{1}{2}$	P	$\delta = 0$	$\delta = \frac{1}{2}$	P
L_∞	**34**	**34**	61	42	**38**	41
Fro	**30**	32	50	34	**30**	33
max	2.09	**2.03**	2.29	4.88	**3.47**	3.92
L_2	10	**9**	19	17	**15**	17

Table 1: Averaged (and rounded except max) distances to Σ (left table) and Σ^{-1} (right table) for $\hat{\Sigma}^N$ with $\delta = 0$, with $\delta = 1/2$ and for the POET estimator (P), in the L_∞, Frobenius, max and spectral norms. Distances to Σ^{-1} were multiplied by 10 before averaging. Best results boxed.

where $\hat{\Sigma}_{\text{sam}}$ is the $p \times p$ sample covariance matrix, δ is a constant in $[0, 1]$, λ is a $p \times p$ matrix with nonnegative entries, and $t(\cdot, \cdot)$ is a function that applies soft, hard, or other thresholding to each non-diagonal entry of its first argument, with the threshold value equal to the corresponding entry of its second argument. λ will typically be parameterised by one scalar parameter. Obviously, $\delta\, \hat{\Sigma}_{\text{sam}}$ and $(1 - \delta)\, t(\hat{\Sigma}_{\text{sam}}, \lambda)$ are the non-sparse and sparse components, respectively.

$\hat{\Sigma}^N$ performs "shrinkage of the sample covariance towards a sparse target". To the best of our knowledge, $\hat{\Sigma}^N$ is a new proposal, although shrinkage towards some other targets has been studied extensively before, notably by Ledoit and Wolf (2003), who propose shrinkage towards a one-factor target and Schaefer and Strimmer (2005), who review and discuss six commonly used targets. Some ideas for the "optimal" choice of δ are proposed in these articles, and can be adopted in the context of $\hat{\Sigma}^N$, thereby reducing the number of "free" parameters of the procedure to the single scalar parameter of the threshold matrix λ. These findings will be reported in more detail elsewhere. If all new covariance estimators were required to have 'literary' names (such as POET), we would name ours 'NOVELIST', for 'NOVEL Integration of the Sample and Thresholded covariance estimators'. The benefits of NOVELIST include simplicity, ease of implementation, and the fact that its application avoids eigenanalysis, which is unfamiliar to many practitioners.

We now briefly exhibit the performance of POET versus NOVELIST on a simulated covariance matrix Σ of size 100×100, available from `http://stats.lse.ac.uk/fryzlewicz/testcov.RData` (use `load("testcov.RData")` in R, the variable name is `testcov`). Σ was not generated from a factor model and is not sparse. The range of its diagonal elements is $[3.32, 7.09]$, while only 56 of the non-diagonal entries are larger than 1 in absolute value. The sample size is $n = 100$, so $\hat{\Sigma}_{\text{sam}}$ itself is not invertible. In NOVELIST, we use both $\delta = 0$ and $\delta = 1/2$, and the constant matrix $\lambda \equiv 1$. In POET, we use $K = 7$, following the authors' advice, given in the R package `POET`, to choose a large K (in order to avoid issues related to K being underestimated), but preferably smaller than 8. Both POET and NOVELIST use soft thresholding. Other POET parameters are set to default. The data are Gaussian, and there are $N = 100$ repetitions. Table 1 shows the results. POET performs poorly for Σ: it is the worst in all norms by a large margin. NOVELIST with $\delta = 0$ (which reduces to the simple soft thresholding estimator) and with $\delta = 1/2$ are hard to tell apart in terms of their performance. However, as far as Σ^{-1} is concerned, NOVELIST with $\delta = 1/2$ is the best, followed by POET and then by the simple soft thresholding. The overall clear 'winner' in this example is NOVELIST with $\delta = 1/2$.

By way of summary, POET is an elegant construction which combines parsimony of representation in the low-rank component with sparsity in the thresholded part. This brief discussion (a) attempts to list some research questions regarding POET which we believe are worth exploring further, and (b) proposes a simple competitor. We found the article a pleasure to read and thought it was written in a clear and pedagogical way. We are convinced that POET will stimulate further research in the important field of large covariance estimation. It therefore gives us great pleasure to second the vote of thanks for this paper.

References

O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10:603–621, 2003.

J. Schaefer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4, 2005. Article no. 32.