Computational Statistics and Data Analysis 80 (2014) 117–128

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Stable estimation of a covariance matrix guided by nuclear norm penalties

Eric C. Chi[a,*], Kenneth Lange[b]

[a] Department of Electrical and Computer Engineering, Rice University, TX, USA

[b] Departments of Human Genetics, Biomathematics, and Statistics, University of California, Los Angeles, CA, USA

ARTICLE INFO

Article history:
Received 18 May 2013
Received in revised form 19 May 2014
Accepted 24 June 2014
Available online 1 July 2014

Keywords:
Covariance estimation
Regularization
Condition number
Discriminant analysis
EM clustering

ABSTRACT

Estimation of a covariance matrix or its inverse plays a central role in many statistical methods. For these methods to work reliably, estimated matrices must not only be invertible but also well-conditioned. The current paper introduces a novel prior to ensure a well-conditioned maximum a posteriori (MAP) covariance estimate. The prior shrinks the sample covariance estimator towards a stable target and leads to a MAP estimator that is consistent and asymptotically efficient. Thus, the MAP estimator gracefully transitions towards the sample covariance matrix as the number of samples grows relative to the number of covariates. The utility of the MAP estimator is demonstrated in two standard applications – discriminant analysis and EM clustering – in challenging sampling regimes.

1. Introduction

Estimation of a covariance matrix or its inverse plays a central role in many statistical methods, ranging from least squares regression to EM clustering. In these applications it is crucial to obtain estimates that are not only non-singular but also stable under small perturbations in sample values. It is well known that the sample covariance matrix

$$\boldsymbol{S} = \frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{y}_j - \bar{\boldsymbol{y}})(\boldsymbol{y}_j - \bar{\boldsymbol{y}})^t$$

is the maximum likelihood estimate of the population covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ of a random sample $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \in \mathbb{R}^p$ from a multivariate normal distribution. When the number of components p of each sample point exceeds the sample size n, the sample covariance $\boldsymbol{S}$ is no longer invertible. Even when n slightly exceeds p, the estimate $\boldsymbol{S}$ can be unstable. Introducing a penalty in the maximum likelihood framework offers a reliable means of stabilizing covariance estimation.

To motivate our choice of penalization, consider the eigenvalues of the sample covariance matrix in a simple simulation experiment. We drew n independent samples from a 10-dimensional multivariate normal distribution $\boldsymbol{y}_i \sim N(\boldsymbol{0}, \boldsymbol{I}_{10})$. Fig. 1 presents boxplots of the sorted eigenvalues of the sample covariance matrix $\boldsymbol{S}$ over 100 trials for sample sizes n drawn from the set $\{5, 10, 20, 50, 100, 500\}$. The boxplots descend from the largest eigenvalue on the left to the smallest eigenvalue on the right. The figure vividly illustrates the previous observation that the highest eigenvalues tend to be inflated upwards above 1, while the lowest eigenvalues are deflated downwards below 1 (Ledoit and Wolf, 2004, 2012). In general, if the sample size n and the number of components p approach ∞ in such a way that the ratio p/n approaches $\zeta \in (0, 1)$, then

* Corresponding author.
E-mail address: echi@rice.edu (E.C. Chi).

http://dx.doi.org/10.1016/j.csda.2014.06.018
0167-9473/

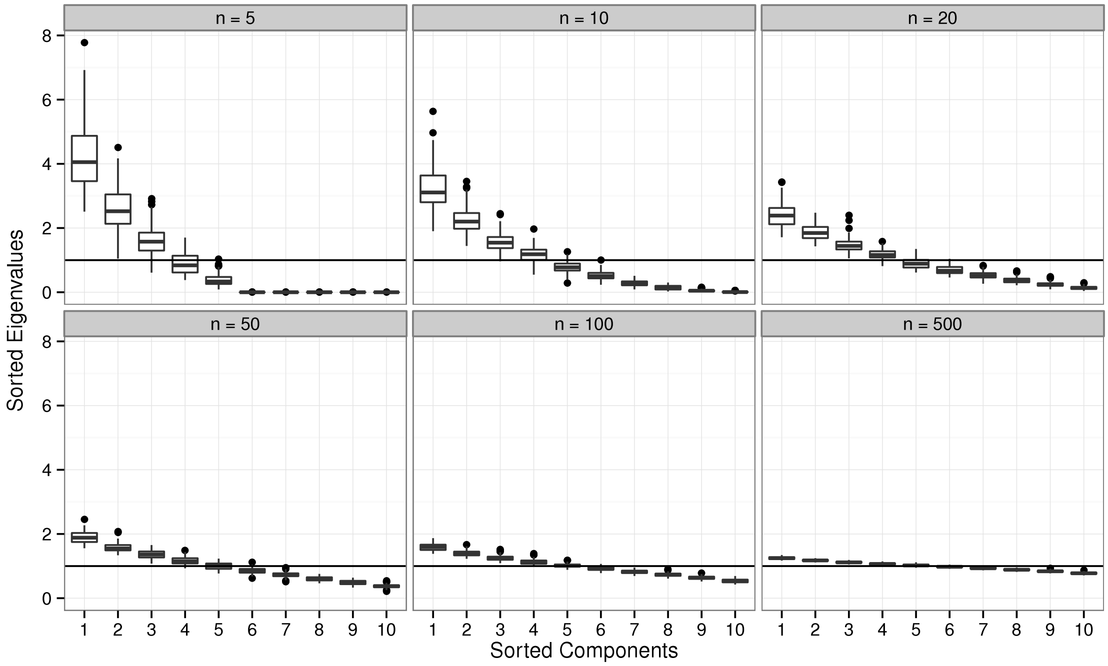

Fig. 1. Boxplots of the sorted eigenvalues of the sample covariance matrix $\boldsymbol{S}$ over 100 random trials. Here the number of components $p = 10$, and the sample size n is drawn from the set $\{5, 10, 20, 50, 100, 500\}$.

the eigenvalues of $\boldsymbol{S}$ tend to the Marĉenko–Pastur law (Marĉenko and Pastur, 1967), which is supported on the interval $([1-\sqrt{\zeta}\,]^2, [1+\sqrt{\zeta}\,]^2)$. Thus, the distortion worsens as ζ approaches 1. The obvious remedy is to pull the highest eigenvalues down and push the lowest eigenvalues up.

In this paper, we introduce a novel prior which effects the desired adjustment on the sample eigenvalues. Maximum a posteriori (MAP) estimation under the prior boils down to a simple nonlinear transformation of the sample eigenvalues. In addition to proving that our estimator has desirable theoretical properties, we also demonstrate its utility in extending two fundamental statistical methods – discriminant analysis and EM clustering – to contexts where the number of samples n is either on the order of or dominated by the number of parameters p.

The rest of our paper is organized as follows. Section 2 discusses the history of stable estimation of structured and unstructured covariance matrices. Section 3 specifies our Bayesian prior and derives the MAP estimator under the prior. Section 4 proves that the estimator is consistent and asymptotically efficient. Section 5 reports finite sample studies comparing our MAP estimator to relevant existing estimators. Section 6 illustrates the estimator for some common tasks in statistics. Finally, Section 7 discusses limitations, generalizations, and further applications of the estimator.

2. Related work

Structured estimation of covariance matrices can be attacked from two complementary perspectives: generalized linear models and regularization (Pourahmadi, 2011, 2013). In this work we consider the problem from the latter perspective. Regularized estimation of covariance matrices and their inverses has been a topic of intense scrutiny (Wu and Pourahmadi, 2003; Bickel and Levina, 2008), and the current literature reflects a wide spectrum of structural assumptions. For instance, banded covariance matrices are appropriate for time series and spatial data, where the order of the components is important. It is also helpful to impose sparsity on a covariance matrix, its inverse, or its factors in a Cholesky decomposition or other factorization (Huang et al., 2006; Rohde and Tsybakov, 2011; Cai and Zhou, 2012; Ravikumar et al., 2011; Rajaratnam et al., 2008; Khare and Rajaratnam, 2011; Fan et al., 2011; Banerjee et al., 2008; Friedman et al., 2008; Hero and Rajaratnam, 2011, 2012; Peng et al., 2009).

In this work, we do not assume any special prior structure. Our sole concern is to directly address the distortion in the eigenvalues of the sample covariance matrix. Thus, we work in the context of rotationally-invariant estimators first proposed by Stein (1975). If $\boldsymbol{S} = \boldsymbol{UDU}^t$ is the spectral decomposition of $\boldsymbol{S}$, then Stein suggests alternative estimators of the form

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{U}\mathrm{diag}(e_1, \ldots, e_p)\,\boldsymbol{U}^t$$

that modify the eigenvalues but not the eigenvectors of $\boldsymbol{S}$. In particular, Stein (1975), Haff (1991), Ledoit and Wolf (2004) and Warton (2008) study the family

$$\hat{\boldsymbol{\Sigma}} = (1-\gamma)\boldsymbol{S} + \gamma\boldsymbol{T} \tag{1}$$