Volume 3 Issue 2 October 2012
# The MINDANAWAN
**Journal of Mathematics**
ISSN 2094-7380

# General Adaptive Sparse-PCA for High Dimensional Data with Low Sample Size

**Mark Gil T. Torres**[4]
*Department of Mathematics, College of Science and Mathematics, MSU-Iligan Institute of Technology, 9200 Iligan City*
markgil.torres@g.msuiit.edu.ph

**Abstract:** In this paper, we propose a novel solution to the general adaptive sparse-PCA (GAS-PCA), which was developed by [3], to work with high dimensional data with low sample size (HDDLSS).

## 1 Introduction

Principal Component Analysis (PCA) is a popular methodology for dimension reduction, but one of its drawbacks is the interpretability of the principal components (PCs). Let $X \in \mathbb{R}^{n\times p}$ and $x_i \in \mathbb{R}^p$ be its realization on the $i$th observation. Without loss of generality, suppose that E(X)=0 and cov(X)=$\Sigma$, where $\Sigma$ is some symmetric positive definite matrix. Then $\Sigma$ can be represented as an eigen decomposition,

$$\Sigma = Q\Lambda Q^T$$

where $\Lambda$ is a diagonal matrix whose entries are the eigenvalues of $\Sigma$ and the columns of $Q$ are its associated eigenvector. For the purpose of identifiability, we assume that $\Lambda_{11} > \Lambda_{22} > \cdots > \Lambda_{pp} > 0$ and that the first nonzero component of our eigenvectors is positive. Since the $j$th

[4]Research is partially supported by the DOST-PCIEERD, Philippines.

principal component of $x_i$ is given by $x_i^T Q_{.j}$, then the PCs are linear combinations of all the original $p$ variables.

To have a better interpretability of the PCs, [5] introduced a method called sparse-PCA (SPCA) using the lasso to produce modified PCs with sparse loadings. By focusing on the first $k$ leading PCs, they transformed the original problem into an equivalent least-squares problem. Then, they apply the lasso to minimize the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. Thus, the SPCA formulation is given by

$$(\hat{A}, \hat{B}) = \text{argmin}_{A,B} \left\{ \frac{1}{n} \sum_{i=1}^{n} ||x_i - AB^T x_i||^2 + \sum_{j=1}^{p_0} (\lambda_j \sum_{k=1}^{p} |\beta_{jk}|) \right\}, \text{ subject to } A^T A = I,$$

where $\bar{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_{p_0})^T$, and $\hat{A}_{\bar{\lambda}} = (\hat{\alpha}_{\bar{\lambda},1}, \hat{\alpha}_{\bar{\lambda},2}, \cdots, \hat{\alpha}_{\bar{\lambda},p_0})$, $\hat{B}_{\bar{\lambda}} = (\hat{\beta}_{\bar{\lambda},1}, \hat{\beta}_{\bar{\lambda},2}, \cdots, \hat{\beta}_{\bar{\lambda},p_0})$ are the resulting lasso estimator. Here, the resulting estimator $\hat{\beta}_{\bar{\lambda}}$ is the SPCA estimator.

To improve the performance of SPCA, [3] proposed a new methodology called general adaptive sparse-PCA (GAS-PCA), where they replaced the least-squares objective function and the lasso penalty by a general least-squares objective function and an adaptive lasso penalty, respectively. Under this new methodology, it allows the study of many related sparse-PCA under a unified theoretical framework. Reformulating the lasso penalty of SPCA, wherein,

$$\sum_{j=1}^{p_0} (\lambda_j \sum_{k=1}^{p} |\beta_{jk}|) \longrightarrow \sum_{j=1}^{p_0} \sum_{k=1}^{p} \lambda_{jk} |\beta_{jk}|,$$

resulted to a simple adaptive sparse-PCA (SAS-PCA). This new reformulation of changing the lasso penalty to an adaptive lasso penalty allows different tuning parameters for different loading coefficients. As stated in [3], for a fixed A and up to a constant irrelevant to B, the SAS-PCA objective function is equivalent to

$$\sum_{j=1}^{p_0}\left\{\frac{1}{n}\sum_{i=1}^{n}(\alpha_j^T x_i-\beta_j^T x_i)^2+\sum_{k=1}^{p}\lambda_{jk}|\beta_{jk}|\right\}$$
$$=\sum_{j=1}^{p_0}\left\{(\alpha_j-\beta_j)^T\tilde{\Sigma}(\alpha_j-\beta_j)+\sum_{k=1}^{p}\lambda_{jk}|\beta_{jk}|\right\},$$

which consists of $p_0$ independent lasso problems. Now, given the matrix A, GAS-PCA minimizes the general least-squares objective function

$$\sum_{j=1}^{p_0}\left\{(\alpha_j-\beta_j)^T\tilde{\Omega}(\alpha_j-\beta_j)+\sum_{k=1}^{p}\lambda_{jk}|\beta_{jk}|\right\},$$

where $\tilde{\Omega}$ is a positive definite matrix with probabilistic limit $\Omega$ and $\Omega$ is the kernel matrix. Then from the set of finite solutions, select the optimal solution using the BIC-type selection criterion

$$\mathrm{BIC}_{\lambda j}=(\hat{\alpha}_{\lambda j}-\hat{\beta}_{\lambda j})^T\tilde{\Omega}(\hat{\alpha}_{\lambda j}-\hat{\beta}_{\lambda j})+\mathrm{df}_{\lambda j}\times\frac{\log n}{n},$$

where $\mathrm{df}_{\lambda j}$ is the nummber of nonzero coefficients identified in $\hat{\beta}_{\lambda j}$.

According to [3], $\tilde{\Omega}=\tilde{\Sigma}$ is unlikely the optimal choice of $\tilde{\Omega}$ and suggest the use of $\tilde{\Omega}=\hat{\mathrm{cov}}^{-1}(\tilde{\beta}_j)$. Since there is no existing simple formula for $\mathrm{cov}(\hat{\beta}_j)$, [3] proposed the following bootstrap method to estimate $\mathrm{cov}(\hat{\beta}_j)$. For given

$\hat{\Sigma}$, apply PCA to bootstrap samples from $N(0, \hat{\Sigma})$ to produce a sufficient number of bootstrap estimates (denoted by $\hat{\beta}_j^{\text{boot}}$) for $\beta_j$. The sample covariance of $\hat{\beta}_j^{\text{boot}}$ can easily be computed (denoted as $\hat{C}_j$)as a natural estimate of $\text{cov}(\hat{\beta}_j)$. And then follow the procedure on [4] and fix $\hat{\Omega} = \hat{C}_j^{-1}$. For HDDLSS, the $\text{cov}(\hat{\beta}_j)$ tends to be singular due to the collinearity of the bootstrap samples. Thus, this presents us with the problem of solving for the $\widehat{\text{cov}}^{-1}(\tilde{\beta}_j)$.

In the next section, we propose a solution for solving GAS-PCA for HDDLSS and its methodological details. Simulation results are provided in Section 3 and the article concludes a short discussion in Section 4.

# 2 Methodology

Following the procedure in [3], we modify the ordinary method of solving the inverse of $\text{cov}(\hat{\beta}_j)$. A simple approach to solving $\widehat{\text{cov}}^{-1}(\tilde{\beta}_j)$ is to compute a best fit solution using the Moore-Penrose general inverse. And since for any covariance matrix $\Sigma$, the Moore-Penrose general inverse exists and is unique, then we always have a solution for the inverse covariance estimates.

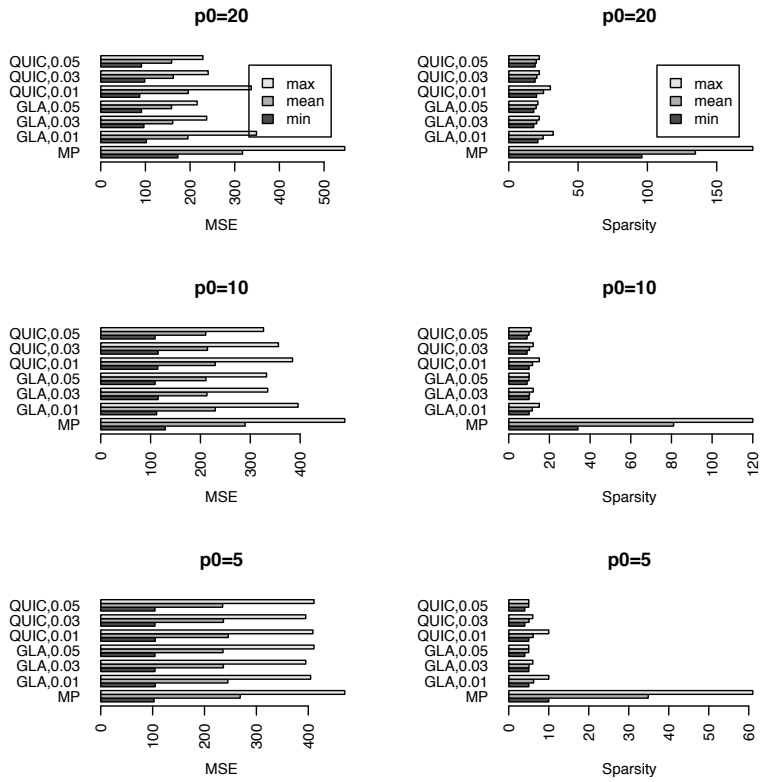
In recent years, [1] develop a method called "glasso" where they apply the lasso penalty in estimating the inverse covariance matrix which results to a sparse inverse covariance matrix. An improvement over glasso, [2] develop a method called "QUIC" which uses quadratic apprroximation in estimating the inverse covariance matrix. Here, we propose to use sparse inverse covariance estimates as an alternative to the Moore-Penrose general-inverse. We compare two methods of sparse inverse covariance estimates namely, glasso and QUIC, at different tuning parameters and use the Moore-Penrose general inverse as a basis for

comparison.

# 3 Simulation Results

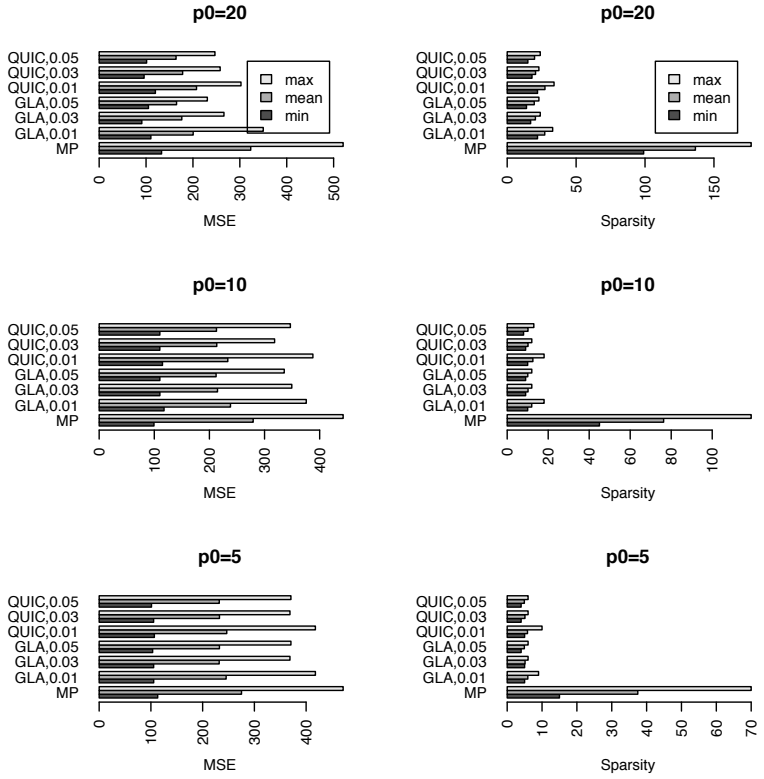
For a fixed $\beta$, let $X$ be a sample of $n$ observations from $N(0, \Sigma)$, where $\Sigma$ is some $p \times p$ positive definite matrix. Then, let $Y = X\beta + \varepsilon$, where $\varepsilon \sim N(0, 1)$, be the true model and $\hat{Y} = \hat{X}\beta$ be the estimated model, where $\hat{X} = XQ$ and $Q$ be the PCA loadings. Setting our sample size to $n < p$, we determine $Q_{\text{MP}}, Q_{\text{glasso}}$, and $Q_{\text{QUIC}}$, the PCA loadings using GAS-PCA with Moore-Penrose general inverse, with glasso, and with QUIC method respectively for some fixed number of PCs $p_0$. We then generate a new sample of 1000 observations from $N(0, \Sigma)$ and compute the MSE of the estimated model. To have a way of comparing the sparsity of the PCA loadings between different methods, we just count the number of non-zero elements of our PCA loadings $Q$. We repeat this procedure for 50 random $\Sigma$ and determine the minimum, mean, maximum value of the MSE and the number of non-zero elements of $Q$.

For $n = 40$ observations and $p = 50$ predictors, we have the following results:

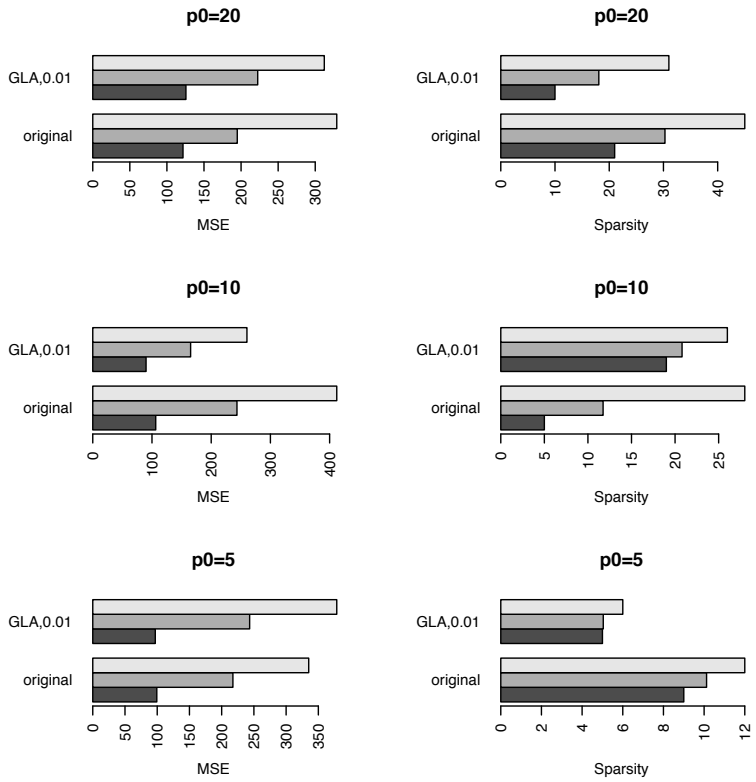



We use $0.01, 0.03$, and $0.05$ as our tuning parameter for glasso and QUIC methods. The results above shows a minimal MSE and highly sparse PCA loadings on methods using sparse inverse covariance estimates over the Moore-Penrose general inverse. As we decrease the number of PCs, it is evident that there is an increase in MSE and sparsity. And as we vary our tuning parameters on glasso and QUIC, it shows smaller MSE and higher sparsity.

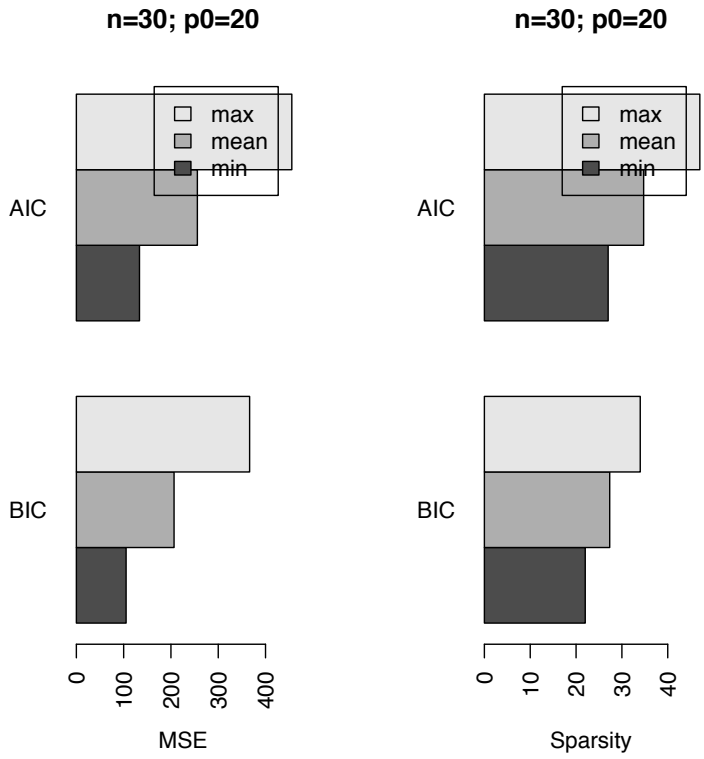
Fixing $p = 50$ while decreasing our number of observations to $n = 30$, we have the following results:




From the preceding results, the inverse covariance estimates greatly improves the GAS-PCA over the Moore-Penrose pseudoinverse estimates in terms of minimal MSE and sparsity. This motivates us to check if the same is true for high dimensional data where $n \gg p$. For $n = 1000$ and $p = 50$, we compare the original GAS-PCA procedure with our modified procedure using the glasso as our method for estimating the inverse covariance matrix.




Lastly, we want to check if using Akaike Information Criterion (AIC) as our model selection for GAS-PCA with HDDLSS will improved our procedure. Thus, we compare a GAS-PCA procedure using glasso method with tuning parameter rho=0.01 using AIC as its model selection over the original model selection which is Bayesian Information Criterion (BIC).




# 4 Discussion

In this paper, we extend GAS-PCA to HDDLSS using estimates of the inverse covariance matrix. In conclusion, using sparse inverse covariance estimates resulted to minimal MSE and greatly improved the sparsity of the PCA loadings with HDDLSS. The numerical studies show that even for data with $n \gg p$, the sparse inverse covariance estimates is comparable to the ordinary method in terms of MSE and sparsity of the PCA loadings. With respect to model selection, BIC is still the better choice even for HDDLSS.

# References

[1] Friedman, J., Hastie, T., and Tibshirani, R., *Sparse inverse covariance estimation with the graphical lasso*, Biostatitics, 9, 432-441, 2008.

[2] Hsieh, C., Sustik, M., Dhillon, I., and Ravikumar, P., *Sparse Inverse Covariance Matrix Estimation Using Quadratic Approximation*, Advances in Neural Information Processing Systems 24, 2330-2338, 2011.

[3] Leng, C. and Wang, H., *On General Adaptive Sparse Principal Component Analysis*, Journal of Computational and Graphical Statistics, 18, 201-215, 2009.

[4] Wang, H. and Leng, C., *Unified Lasso Estimation by Least Squares Approximation*, Journal of the American Statistical Association, 102, 1039-1048, 2007.

[5] Zou, H., Hastie, T., and Tibshirani, R., *Sparse Principal Component Analysis*, Journal of Computational and Graphical Statistics, 15, 265-286, 2006.