Sparse Principal Component Regression for Generalized Linear Models

Ferney Henao-Ceballos[1], Samuella Boadi [1]

[1]The University of Iowa, Iowa City, IA.

ARTICLE HISTORY
Compiled January 12, 2026

ABSTRACT
High-dimensional data present significant challenges for modeling and interpretation, especially in the framework of generalized linear models (GLMs). Traditional Principal Component Regression (PCR) addresses multicollinearity by using principal components derived from explanatory variables. However, its two-stage approach does not take into account the response variable during the dimension reduction step, which can lead to suboptimal predictive performance. This paper explores the Sparse Principal Component Regression for Generalized Linear Models (SPCR-GLM), a one-stage method that simultaneously performs dimension reduction and model fitting. SPCR-GLM minimizes a combined loss function consisting of the negative log-likelihood from the GLM and a PCA reconstruction loss, while incorporating sparsity via elastic net regularization. We review the theoretical formulation of SPCR-GLM, compare it to related methods such as sparse PCA and sparse PCR, and demonstrate its performance on real and simulated datasets.

KEYWORDS
PCA, SPCR-GLM, GLM, dimension reduction, sparsity, elastic net regularization.

1. Introduction

High-dimensional data analysis is a common challenge in modern statistics, particularly when the number of predictors exceeds or is comparable to the number of observations. In such cases, issues like multicollinearity, overfitting, and poor interpretability often arise. A widely adopted technique to address these challenges is Principal Component Analysis (PCA), a classical method for dimension reduction that transforms a dataset with many correlated variables into a smaller set of uncorrelated variables while preserving most of the original variance (Pearson, 1901; Hotelling, 1933; Jolliffe, 2002).

PCA works by linearly projecting the original p-dimensional features into a k-dimensional subspace, where $k < p$, resulting in new variables called principal components (PCs). These components are orthogonal and ranked according to the amount of variance they capture from the original data. Transformation is typically derived through the eigendecomposition of the sample covariance matrix or the singular value decomposition of the data matrix (Bishop, 2006). Each principal component is defined as a linear combination of the original variables, with the coefficients of these combinations represented in the **loading matrix** $\mathbf{B} \in \mathbb{R}^{p \times k}$. Each column $\boldsymbol{\beta}_j$ of $\mathbf{B}$ corresponds to a **loading vector** for the jth principal component.

Because the principal components are orthogonal, the columns of $\mathbf{B}$ are constrained to be orthonormal, satisfying $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_k$, where $\mathbf{I}_k$ is the identity matrix of size k. Mathematically, PCA is formulated as the following least squares problem:

$$\min_{\mathbf{B}} \sum_{i=1}^{n} \left\| \mathbf{x}_i - \mathbf{B}\mathbf{B}^\top \mathbf{x}_i \right\|^2, \text{subject to } \mathbf{B}^\top\mathbf{B} = \mathbf{I}_k$$

where $\mathbf{x}_i$ for $i \in \{1, \cdots n\}$ is the ith row of the design matrix, $\mathbf{B} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k]$ is the loading matrix, and $\mathbf{I}_k$ is the identity matrix of size k. Each principal component is a linear combination of the original variables, and the loadings (coefficients in the matrix $\mathbf{B}$) indicate the contribution of each variable to that component. Loadings with large absolute values suggest strong influence, while those close to zero imply minimal contribution. Interpretability improves when many loadings are zero, which motivates sparse PCA methods that introduce sparsity through regularization (Zou et al., 2006; Journee et al., 2010).

Principal Component Regression (PCR) extends PCA by using principal components as predictors in a regression model (Jolliffe, 1982, 2002). This method is particularly useful for mitigating multicollinearity and reducing model complexity. However, traditional PCR selects components solely based on the variance of the covariates, disregarding their relationship with the response variable. Consequently, components that are highly predictive but explain less variance may be excluded, leading to suboptimal performance (Bair et al., 2006; Chandrasekaran et al., 2012). To address this limitation, Sparse Principal Component Regression (SPCR) was introduced by Zou et al. (2006) and later extended by Kawano et al. (2015), integrating PCA and regression into a single optimization framework with sparse regularization. SPCR aims to discover components that are both relevant for prediction and interpretable through sparsity. However, the initial formulations were tailored to continuous outcomes under linear regression.

More recently, SPCR for Generalized Linear Models (SPCR-glm) was developed to accommodate binary, count, and other non-Gaussian responses within the GLM framework (Kawano and Konishi, 2018). This method jointly optimizes the negative log-likelihood of the GLM and the PCA reconstruction error, incorporating elastic net regularization to encourage sparsity and prevent overfitting. SPCR-glm achieves automatic component selection, parameter identifiability, and competitive predictive performance, as demonstrated in applications ranging from health surveys to genomic studies.

This paper investigates the theoretical foundations and practical performance of PCR, with a particular focus on its sparse and GLM-integrated variants. We study the framework proposed by Kawano and Konishi (2018) and others to evaluate how these methods balance prediction accuracy, dimension reduction, and model interpretability in modern high-dimensional settings.

2. Sparse Principal Component Regression for GLMs (SPCR-GLM)

PCA uses an orthogonal transformation to convert a set of possibly correlated variables into a set of linearly uncorrelated variables called principal components. These components are ordered such that the first few retain most of the variance present in the original variables. However, the principal components obtained by standard PCA are typically linear combinations involving all original variables, which makes interpretation difficult. Sparse PCA (SPCA) addresses this limitation by enforcing sparsity on the principal components, i.e., by encouraging many of the component loadings to be zero. This results in principal components that are easier to interpret. A more detailed description of SPCA is presented below.

Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ denote the centered matrix of explanatory variables, where each row $\mathbf{x}_i \in \mathbb{R}^p$ represents an observation, and let $\mathbf{y} \in \mathbb{R}^n$ be the corresponding response variable. Zou et al. (2006) proposed a formulation of PCA including sparsity through elastic net regularization. In their approach, the first sparse principal component is obtained by solving the following optimization problem:

$$\min_{\mathbf{A},\mathbf{B}} \left\{ \sum_{i=1}^{n} \|\mathbf{x_i} - \mathbf{A}\mathbf{B}^\top \mathbf{x_i}\|_2^2 + \sum_{j=1}^{k} \left\{ \lambda \|\boldsymbol{\beta}_j\|_2^2 + \lambda_{1,j} \|\boldsymbol{\beta}_j\|_1 \right\} \right\}, \quad \text{subject to } \mathbf{A}^\top \mathbf{A} = \mathbf{I}_k,$$

where $\mathbf{B} \in \mathbb{R}^{p \times k}$ is the loading matrix whose columns $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k \in \mathbb{R}^p$ represent the sparse principal component directions, $\mathbf{A} \in \mathbb{R}^{p \times k}$ is a orthonormal matrix whose columns $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k \in \mathbb{R}^p$ are the direction vector scores, and $\lambda, \lambda_{1,j}$ are the regularization parameters controlling sparsity and stability, respectively. The optimization is typically solved using iterative algorithms, alternating between updating $\mathbf{A}$ and $\mathbf{B}$. The solution yields sparse principal components that retain most of the data variance and improve interpretability by involving only a subset of variables in each component. Finally, Sparse Principal Component Regression (SPCR) is just linear regression applied to the linear predictor $\mathbf{B}^\top \mathbf{X}$.

In SPCR, the response variable is restricted to being continuous variable. Kawano and Konishi (2018) proposed Sparse Principal Component Regression for Generalized Linear Models (SPCR-GLM), this approach integrates dimension reduction and regression into a single-stage procedure. It incorporates the response variable directly into the optimization process, allowing it to identify components that are sparse, interpretable, and predictive. This is particularly valuable in scenarios where **predictive information is hidden in lower-variance directions**, which standard PCA and SPCA would discard. SPCR-GLM uses generalized linear models to deal with various types of data, including binary, count, and multiclass data.

In this context, the linear predictor for the i-th observation is defined as:

$$\eta_i = \gamma_0 + \boldsymbol{\gamma}^\top (\mathbf{B}^\top \mathbf{x}_i),$$

where $\gamma_0 \in \mathbb{R}$ is the intercept, and $\boldsymbol{\gamma} \in \mathbb{R}^k$ is the regression coefficient vector associated with the components. This formulation projects the high-dimensional predictors $\mathbf{x}_i$ onto a lower-dimensional subspace defined by $\mathbf{B}$, and models the response in this

reduced space. SPCR-GLM solves the following regularized optimization problem:

$$\min_{\mathbf{A},\mathbf{B},\gamma_0,\boldsymbol{\gamma}} \left\{ Ł_{\text{reg}}(\gamma_0, \boldsymbol{\gamma}, \mathbf{B}) + w \sum_{i=1}^{n} \|\mathbf{x_i} - \mathbf{A}\mathbf{B}^\top \mathbf{x_i}\|_2^2 + \lambda_{\boldsymbol{\beta}} \left[\xi \sum_{j=1}^{k} \|\boldsymbol{\beta}_j\|_2^2 + (1-\xi) \sum_{j=1}^{k} \|\boldsymbol{\beta}_j\|_1 \right] + \lambda_{\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_1 \right\}$$

subject to the constraint: $A^\top A = \mathbf{I}_k$.

The components of this objective function are as follows:

- $Ł_{\text{reg}}(\gamma_0, \boldsymbol{\gamma}, \mathbf{B})$ is the negative log-likelihood from the GLM, based on the chosen distribution and link function.
- The penalty on $\mathbf{B}$ is an elastic net-type regularization controlled by $\lambda_{\boldsymbol{\beta}} > 0$ and $\xi \in [0, 1)$, which balances sparsity (ℓ_1 norm) and stability/grouping effects (ℓ_2 norm).
- The ℓ_1 penalty on $\boldsymbol{\gamma}$, scaled by $\lambda_{\boldsymbol{\gamma}} > 0$, encourages sparse regression coefficients, potentially removing irrelevant components.
- The tuning parameter $w > 0$ controls the relative weight between the regression and PCA reconstruction objectives.

The orthonormality constraint $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_k$ maintains identifiability and the interpretability of the component directions. The combined use of PCA loss and GLM likelihood enables flexible modeling across a variety of outcome types, including binary and count responses, making SPCR-GLM broadly applicable in high-dimensional data analysis (Kawano and Konishi, 2018).

3. Computational Algorithm

Due to the non-convex nature of the joint optimization, the authors propose an iterative algorithm based on coordinate descent and local quadratic approximation of the log-likelihood:

(1) Initialize $\mathbf{A}$, $\mathbf{B}$, γ_0, and $\boldsymbol{\gamma}$.
(2) Approximate the GLM loss in quadratic form using the Taylor expansion.
(3) Alternate updates:
    - Update $\mathbf{B}$ through soft holding (Lasso).
    - Update $\boldsymbol{\gamma}$ through coordinate descent.
    - Update $\mathbf{A}$ through the rotation of the Procrustes.
    - Update γ_0 analytically.
(4) Repeat until convergence.

4. Simulation

A simulation experiment is designed where the true signal for predicting a binary response lies in directions of **low variance** in the explanatory variables. We generate $n = 1000$ observations with explanatory variables $\mathbf{x}_i \in \mathbb{R}^{50}$ and binary response $\mathbf{y}_i \in \{0, 1\}$. Each $\mathbf{x}_i$ is constructed from a low-dimensional signal $\mathbf{u}_i \in \mathbb{R}^2$ and a noise component $\mathbf{v}_i \in \mathbb{R}^{48}$ from a multivariate normal distribution with an exponentially decaying covariance structure.

The signal vector $\mathbf{u}_i$ is drawn from a 4-component Gaussian model with equal probability, the means for the Gaussian distributions are located at $\mu_1 = (3, 3)^\top$, $\mu_2 = (-3, 3)^\top$, $\mu_3 = (-3, -3)^\top$, $\mu_4 = (3, -3)^\top$, and $\mu_5 = (3, -3)^\top$. Each component has a covariance matrix $0.25 \cdot \mathbf{I}_2$. The noise component $\mathbf{v}_i$ is sampled from a multivariate normal distribution with zero mean and covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{48\times 48}$, where $\Sigma_{jk} = 0.1^{|j-k|}$.

The binary response $\mathbf{y}$ is generated from a logistic regression model where the linear predictor is $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ with $\boldsymbol{\beta} = (1, -1, 0, \cdots, 0)^\top$, and the link function is logit. The response is sampled as:

$$\mathbf{y} \sim \text{Bernoulli}(\boldsymbol{\pi}), \quad \boldsymbol{\pi} = \frac{1}{1 + \exp(-\boldsymbol{\eta})}.$$

The goal is to assess how well different dimension reduction methods recover the signal directions and predict $\mathbf{y}$ using only a reduced set of components. Figure 1 shows the scatter plot of the true predictors and the scatter plot of the principal components using the traditional PCA and the SPCR-GLM approaches.

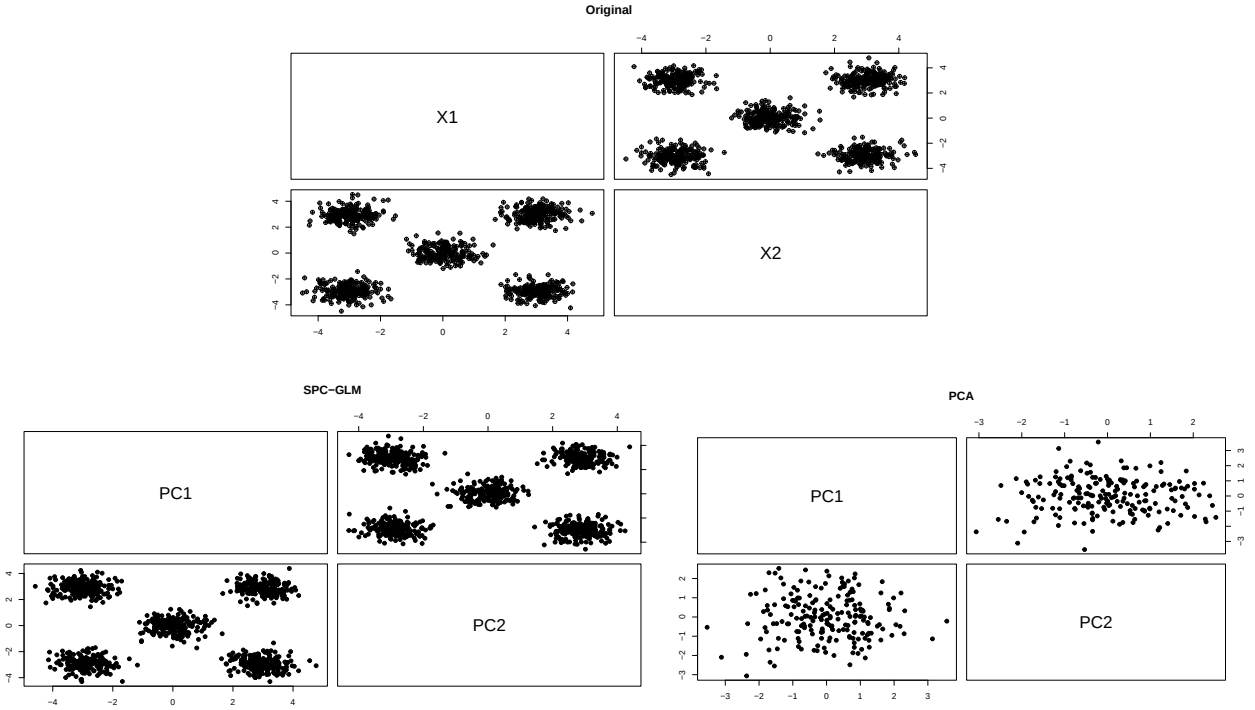

Figure 1. Scatter plots of principal components

In the simulation, traditional PCR and PCA-based approaches fail to recover the structure of the data because they prioritize variance over predictive relevance. PCA, by construction, does not see the response and thus ranks components based on how much variance they capture in $\mathbf{X}$. SPCR-GLM, on the other hand, incorporates the response into the component selection process. It successfully identifies sparse loadings that align with the directions carrying signal for $\mathbf{y}$.

This simulation highlights the importance of response-aware dimension reduction. In modern high-dimensional datasets, it is common for the most informative features for prediction to be weak in terms of variance. SPCR-GLM is specifically designed for such situations, providing a powerful framework that unifies regression, dimension reduction, and variable selection.

5. Application

We applied SPCR-GLM to the doctor visits data set from the Australian Health Survey. The dataset includes 5190 individuals and examines how 11 demographic and health-related variables influence the number of doctor visits, a count outcome modeled using Poisson regression.

The model was configured with the following parameters, $k = 5$, $w = 0.1$, $\xi = 0.001$, $\lambda_{\boldsymbol{\gamma}} = 0$, $\lambda_{\boldsymbol{\beta}} = 10$. Table presents the estimated values for the principal components:

Table 1. Principal components

	PC1	**PC2**	**PC3**	**PC4**	**PC5**
Gender	-0.535	0.011	-0.083	0.000	0.535
Age	-0.451	0.000	0.322	-0.090	-0.062
Income	0.497	0.000	0.352	0.000	0.000
Illness	-0.047	-0.531	0.000	-0.227	0.000
Reduced	0.020	-0.688	0.085	0.000	0.000
Health	0.061	-0.417	-0.212	-0.002	0.000
Private	0.085	-0.009	0.195	0.000	0.710
Freepoor	0.000	0.000	-0.780	0.000	0.000
Freerepeat	-0.460	0.000	0.153	0.000	-0.422
Nchronic	-0.035	-0.043	0.033	-0.752	0.000
Lchronic	-0.131	-0.259	0.090	0.595	0.000

The estimates $\hat{\gamma}_0$ and $\hat{\gamma}$ for SPCR-GLM were given by:

$$\hat{\gamma}_0 = -1.484, \quad \hat{\gamma} = (-0.106, 0.433, -0.124, -0.087, 0.065)^\top$$

Using this approach, we can interpret several principal components, such as:

- **PC2:** An index of health status (Illness, Reduced activity, Health perception)
- **PC3:** Accessibility to healthcare (Freepoor status, Age)
- **PC4:** Chronic disease burden (Lchronic, Nchronic)

These components provide clear insights into the factors influencing doctor visit frequency. The results demonstrate the utility of SPCR-GLM in health data analysis, particularly for complex data sets where both prediction accuracy and variable interpretability are essential.

6. Conclusion

Sparse Principal Component Regression for Generalized Linear Models (SPCR-GLM) presents a powerful framework for modeling high-dimensional data, especially when response variables do not follow a Gaussian distribution. By integrating dimension reduction and regression into a single-stage procedure, SPCR-GLM effectively identifies sparse, interpretable components that are also predictive. This approach addresses limitations of traditional PCR and PCA-based methods, particularly their neglect of the response variable during component selection. Through simulations and real-data applications, SPCR-GLM has demonstrated superior predictive accuracy and interpretability in scenarios where informative features reside in low-variance directions.

Despite its strengths, one significant drawback of SPCR-GLM is the complexity involved in tuning multiple regularization parameters, such as the sparsity-inducing penalties and the weight balancing the PCA and GLM objectives. Selecting appropriate tuning parameters often requires extensive cross-validation techniques, which can be computationally demanding and may affect model stability. Future work could explore adaptive or data-driven tuning strategies to enhance the robustness and usability of SPCR-GLM in practical applications.

7. appendix

Code:

```
library(spcr)
library(ggplot2)
library(Matrix)
library(MASS)
library(elasticnet)
library(gridExtra)
library(mice)
library(nnet)
library(AER)
library(kableExtra)


# Data generation
set.seed(123)
p<-18
a1 <- c(3, 3); a2 <- c(-3, 3); a3 <- c(-3, -3); a4 <- c(3, -3); a5 <- c(0, 0)
centers <- list(a1, a2, a3, a4, a5)
n <- 1000
u_list <- lapply(1:n, function(i) {
  j <- sample(1:5, 1)
mvrnorm(1, mu = centers[[j]], Sigma = diag(0.5^2, 2))
})
u <- do.call(rbind, u_list)
Sigma_v <- outer(1:p, 1:p, function(i, j) 0.1^abs(i - j))
v <- mvrnorm(n, mu = rep(0, p), Sigma = Sigma_v)
z <- cbind(1,u, v)

beta<-c(1,1,-1,rep(0, p))
# Informative directions and response
logit_p <- as.vector(z%*%beta)
p <- 1 / (1 + exp(-logit_p))
y <- rbinom(n, size = 1, prob = p)
```

```
pairs(u,pch = 10, label=c("X1","X2"), main = "Original")

# Apply SPCR-GLM
result <- spcrglm(x = z, y = y, family = "binomial", k = 2,
                  w = 0.01, xi = 0.001, lambda.B = 2, lambda.gamma = 0.1)

# PC scores
scores <- z %*% result$loadings.B
pairs(scores[, 1:2], labels=c("PC1","PC2","PC3"), pch = 19,
      main = "SPC-GLM")


pca_result <- prcomp(x, scale. = TRUE)
pca_scores <- pca_result$x[, 1:2]
pairs(pca_scores[, 1:2], labels=c("PC1","PC2","PC3"), pch = 19,
      main = "PCA")



# Application Doctor data set

data("DoctorVisits")

x <- model.matrix(~.,data=DoctorVisits)[,-c(1,2)]
y <- DoctorVisits$visits

model <- spcrglm(x = x, y = y, k = 5, family = "poisson",
                 lambda.B = 10, lambda.gamma = 0, w = 0.1, xi = 0.001, center=TRUE,scale = TRUE)


xs<-scale(x, center = TRUE, scale = TRUE)


loadings <- model$loadings.B

kbl(round(loadings,3),format="latex")

scores <- xs%*%loadings
pairs(scores,labels=c("PC1","PC2","PC3","PC4","PC5"))

pca_result <- prcomp(x, center = TRUE, scale. = TRUE)
pca_scores <- pca_result$x[, 1:5]

pairs(pca_scores)
```

References

Bair, E., T. Hastie, D. Paul, and R. Tibshirani (2006). Prediction by supervised principal components. *Journal of the American Statistical Association 101*(473), 119–137.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Chandrasekaran, V., P. A. Parrilo, and A. S. Willsky (2012). Latent variable graphical model selection via convex optimization. *Annals of Statistics 40*(4), 1935–1967.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology 24*(6), 417–441.

Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 31*(3), 300–303.

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer.

Journee, M., Y. Nesterov, P. Richtarik, and R. Sepulchre (2010). Generalized power method for sparse principal component analysis. *Journal of Machine Learning Research 11*, 517–553.

Kawano, S. and S. Konishi (2018). Sparse principal component regression for generalized linear models. *Biometrical Journal 60*(3), 496–514.

Kawano, S., S. Konishi, and Y. Fujikoshi (2015). Sparse principal component regression with adaptive loading. *Computational Statistics & Data Analysis 86*, 42–52.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2*(11), 559–572.

Zou, H., T. Hastie, and R. Tibshirani (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics 15*(2), 265–286.