# A General Null Space Property for Sparse Principal Component Analysis

Xuanli Han[1,2] · Jigen Peng[3] · Angang Cui[4] · Fujun Zhao[1] · Kexue Li[1]

## Abstract

Sparse principal component analysis (SPCA) has achieved great success in improving interpretable ability of the derived results and has become a powerful technique for modern data analysis. It presents that principal component can be modified to produce sparse loadings by imposing sparsity-induced penalty, which is often $l_1$-regularized constraint. In order to analyze the $l_1$-regularized sparsity-induced model, in this paper, we propose a general null space property of a matrix $\mathbf{A}$ relative to a index set $S$ and give a necessary and sufficient condition for the exact or approximate sparse principal components. Meanwhile, the conclusions with respect to the stable and robust situations are given in the case of exact or approximate sparse principal components, respectively.

**Keywords** Principal component analysis (PCA) · Sparse principal component analysis (SPCA) · General null space property (GNSP)

✉ Jigen Peng
  jgpengxjtu@126.com

  Xuanli Han
  hanxuanli@xust.edu.cn

  Angang Cui
  cuiangang@163.com

  Fujun Zhao
  fjzhao1991@163.com

  Kexue Li
  kxli@mail.xjtu.edu.cn

1   School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China

2   College of Sciences, Xi'an University of Science and Technology, Xi'an 710054, China

3   School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China

4   School of Mathematics and Statistics, Yulin University, Yulin 719000, China

# 1 Introduction

Principal component analysis (PCA) [1] has been widely used in data analysis and dimension reduction, especially in the contemporary era of big data. Standard PCA extracts principal components (PCs), which are orthogonal to each other, of sample data matrix by computing leading eigenvectors of sample covariance matrix. PCs capture most sample information in the first few principal components and guarantee least information loss in the sample variance sense. However, principal components (PCs) are linear combination of all original sample variables, which are often numerous in most modern applications, and this makes PCs difficult to interpret which variables influence the sample information further. In order to solve this problem, in the past dozen years, several alternative sparse approaches have been proposed, that is, modified PCs, which are the linear combination of a small subset of sample variables. The modified PCs can still explain most information of all original sample variables. That is the sparse principal component analysis (SPCA). SPCA is one of the most widely used techniques for machine learning, modern data analysis, finance, statistics, process fault detection [2] and many other fields [3, 4].

One main method to get SPCA, which was first proposed by Jolliffe et al. [5], is $l_1$-regularized PCA. The technique is successful in improving the sparsity of loadings of principal component analysis (PCA) by imposing $l_1$-regularization penalty on the loadings of PCA. This earlier stage research on SPCA leads to rapidly growing investigations into sparsity promoting methods due to the success of SPCA's more obtained interpretable ability. D'Aspremont et al. [6] translated the SPCA problem into a semi-definite programming-based relaxation formulation, and Shen et al. [7] proposed a method to obtain SPCA via regularized singular value decomposition. Subsequently, Journée et al. [8] developed two single-unit and two block optimization formulations of the SPCA problem, and also D'Aspremont et al. [3] formulated the SPCA problem as a new semi-definite relaxation and derived a greedy algorithm to compute a full set of good solutions for all target numbers of nonzero coefficients. Other related works including the problem of finding the dominant eigenvector of the sample covariance matrix under additional constraints on the vector proposed by Sigg et al. [9] and the generation of modified PCA with sparse loadings by using of the lasso and a new regularization and variable selection method, that is, the elastic net proposed by Zou et al. [10, 11].

One of the fundamental method to obtain the SPCA is to reformulate the PCA as a regression-type optimization problem and then to impose the $l_1$ constraint on the regression coefficients [11]. It turns out that this problem can be transformed into a lasso-type optimization problem which has the form

$$\min_{\mathbf{z} \in \mathbb{R}^N} \|\mathbf{y} - \mathbf{Az}\|_2^2 + \lambda \|\mathbf{z}\|_1, \tag{1.1}$$

where $\mathbf{A} \in \mathbb{R}^{m \times N}$ is a matrix and $\lambda$ is a nonnegative parameter [10]. According to [12], the solution of problem (1.1) has close relation to the solution of problem (1.2)

$$\min_{\mathbf{z}\in\mathbb{R}^N} \|\mathbf{z}\|_1 \quad subject \quad to \quad \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta. \tag{1.2}$$

But, in practice, it is inconceivable to measure a target vector $\mathbf{z} \in \mathbb{R}^N$ exactly, a practicable method is to search $\mathbf{z}$ consistent with the measured vector $\mathbf{y}$. Meanwhile, as measurements are often perturbed by noise, this means that the measurement data vector $\mathbf{y} \in \mathbb{R}^m$ is only an approximation of the vector $\mathbf{A}\mathbf{z} \in \mathbb{R}^m$, that is, $\|\mathbf{A}\mathbf{z} - \mathbf{y}\| \leq \eta$ for some $\eta \geq 0$ and norm $\|\cdot\|$ on $\mathbb{R}^m$, usually the norm $\|\cdot\|_2$.

Especially, when $\eta = 0$, the problem (1.2) has the form

$$\min_{\mathbf{z}\in\mathbb{R}^N} \|\mathbf{z}\|_1 \quad subject \quad to \quad \mathbf{A}\mathbf{z} = \mathbf{y}. \tag{1.3}$$

In fact, problem (1.3) is the case considering linear measurements of a sparse data vector $\mathbf{z} \in \mathbb{R}^N$ from its measurement vector $\mathbf{y} = \mathbf{A}\mathbf{z} \in \mathbb{R}^m$, where $\mathbf{A} \in \mathbb{R}^{m\times N}$ is a known matrix. Problem (1.3), known as basis pursuit (BP) [13], is usually regarded as the convex relaxation of the original problem

$$\min_{\mathbf{z}\in\mathbb{R}^N} \|\mathbf{z}\|_0 \quad subject \quad to \quad \mathbf{A}\mathbf{z} = \mathbf{y}, \tag{1.4}$$

which is to reconstruct the sparse data vector $\mathbf{z}$ by solving the $l_0$-minimization problem, where $\|\mathbf{z}\|_0$ denotes the number of nonzero entries of the vector $\mathbf{z}$. Problem (1.3) is a tractable strategy convex relaxation of problem (1.4), which is a non-convex problem and is generally NP-hard indeed. This convex relaxation is meaningful because under certain conditions with respect to the matrix $\mathbf{A}$, the model (1.3) can exactly or approximately reconstruct the original sparse data vector $\mathbf{z}$ of problem (1.4). One of these conditions is the theorem with respect to the concept of null space property, which involves the known data matrix $\mathbf{A}$ that ensures exact or approximate reconstruction of the data vector $\mathbf{z}$. The theorem declares that given a matrix $\mathbf{A} \in \mathbb{R}^{m\times N}$, every vector $\mathbf{x} \in \mathbb{R}^N$ supported on a index set $S \subset [N]$ is the unique solution of (1.3) with $\mathbf{y} = \mathbf{A}\mathbf{x}$ if and only if matrix $\mathbf{A}$ satisfies the null space property relative to $S$ [12].

Due to the key role of $l_1$-regularization of (1.1) as a sparse reconstruction technique in compressive sensing literatures, this paper extends the results with respect to the (stable, robust) null space property, which can be used to recover (exactly or approximately) data vector, to reconstruction (exact or approximate) of sparse information matrix, and gives the corresponding results regarding the (stable, robust) general null space property.

This work develops theoretical condition for reconstruction (exact or approximate) of sparse principal components based on the definition of (stable, robust) general null space property and gives a theoretical analysis approach for sparse principal component analysis (SPCA).

The rest of this paper is organized as follows. The definitions of (stable, robust) general null space property are given, and the results respect to these concepts are proved, respectively, in Sects. 2, 3 and 4. In Sect. 5, the conclusion of this paper is presented.

*Notation* In the paper, card($S$) denotes the cardinality of a set $S$, and symbol $\|\mathbf{X}\|_{[0]}$, where $\mathbf{X}$ is a matrix, denotes the number of nonzero entries of the matrix $\mathbf{X}$. The sum of absolute value of entries of matrix $\mathbf{X}$ is denoted as $\|\mathbf{X}\|_{[1]}$, and $\sigma_s(\mathbf{X})_{[1]} = \inf_{\|\mathbf{Z}\|_{[0]} \le s} \|\mathbf{X} - \mathbf{Z}\|_{[1]}$ denotes the error of best $s$-term approximation to the matrix $\mathbf{X}$. For a $m \times n$ matrix $\mathbf{M}$, the set $\Omega = \{(1, 1), (1, 2), \ldots, (1, n), \ldots, (m, n)\}$ denotes all the subscripts of the elements of matrix $\mathbf{M}$, and for index set $S \subset \Omega$, $\mathbf{M}_S$ denotes a $m \times n$ matrix that coincides with $\mathbf{M}$ on the indices in $S$ and is extended to zero outside $S$. Moreover, the set $\overline{S}$ denotes the complement of a set $S$ with respect to set $\Omega$, that is, $\overline{S} = \Omega \setminus S$. $\|\cdot\|_F$ presents the Frobenius norm of a matrix $\cdot$, and $\mathbf{O}$ denotes a null matrix.

## 2 General Null Space Property

In this paper, we focus on the problem

$$\min_{\mathbf{Z} \in \mathbb{R}^{N \times l}} \|\mathbf{Z}\|_{[1]} \quad subject\ to \quad \mathbf{AZ} = \mathbf{Y}, \tag{$\mathbf{P}_1$}$$

where $\mathbf{A}$ is matrix in $\mathbb{R}^{m \times N}$.

First, we give three definitions, which are needed in the following presentation, with respect to the general null space property.

**Definition 1** Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$, the sum of all absolute values of the entries of matrix $\mathbf{A}$ is denoted as $\|\mathbf{A}\|_{[1]}$. At the same time, $\|\mathbf{A}\|_{[0]}$ denotes the number of nonzero entries in the matrix $\mathbf{A}$.

**Definition 2** Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$, the set $\{\mathbf{V} \in \mathbb{R}^{N \times l} | \mathbf{AV} = \mathbf{O}, \mathbf{A} \in \mathbb{R}^{m \times N}\}$ is called the general null space of matrix $\mathbf{A}$ of order $N \times l$, denoted as $Gker(\mathbf{A})$.

As stated by Definition 2, we have $Gker(\mathbf{A}) = \{\mathbf{V} \in \mathbb{R}^{N \times l} | \mathbf{AV} = \mathbf{O}, \mathbf{A} \in \mathbb{R}^{m \times N}\}$.

**Definition 3** A matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ is said to satisfy the general null space property relative to a index set $S \subset \Omega$, if

$$\|\mathbf{V}_S\|_{[1]} < \|\mathbf{V}_{\overline{S}}\|_{[1]} \quad for\ all \quad \mathbf{V} \in Gker(\mathbf{A}) \setminus \{\mathbf{O}\}. \tag{2.1}$$

A matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ is said to satisfy the general null space property of order $s$ if it satisfies the general null space property relative to any index set $S \subset \Omega$ with card($S$)$\le s$.

**Remark 1** In the definition 3, $\Omega = \{(1, 1), (1, 2), \ldots, (1, l), \ldots, (N, l)\}$ and $\mathbf{O}$ is a null matrix in $\mathbb{R}^{N \times l}$.

**Remark 2** According to Definition 3, it can be observed that, for a given $\mathbf{V} \in Gker(\mathbf{A}) \setminus \{\mathbf{O}\}$, the condition $\|\mathbf{V}_S\|_{[1]} < \|\mathbf{V}_{\overline{S}}\|_{[1]}$ holds for any set $S \subset \Omega$ with card($S$) $\le s$ as soon as it holds for an index set of $s$ largest entries of $\mathbf{V}$ in absolute values.

**Remark 3** There are other reformulations of the Definition 3. One of them is obtained by adding $\|\mathbf{V}_S\|_{[1]}$ to both sides of the inequality $\|\mathbf{V}_S\|_{[1]} < \|\mathbf{V}_{\overline{S}}\|_{[1]}$. Thus, the general null space property relative to index set $S \subset \Omega$ is

$$2\|\mathbf{V}_S\|_{[1]} < \|\mathbf{V}\|_{[1]} \quad for \ all \quad \mathbf{V} \in Gker(\mathbf{A})\backslash\{\mathbf{O}\}.$$

The second one is obtained by choosing index set $S \subset \Omega$ as an set of $s$ largest in absolute value entries of $\mathbf{V}_S$ and by adding $\|\mathbf{V}_{\overline{S}}\|_{[1]}$ to both sides of the inequality $\|\mathbf{V}_S\|_{[1]} < \|\mathbf{V}_{\overline{S}}\|_{[1]}$. Then, the general null space property of order $s$ is

$$\|\mathbf{V}\|_{[1]} < 2\sigma_s(\mathbf{V})_{[1]} \quad for \ all \quad \mathbf{V} \in Gker(\mathbf{A})\backslash\{\mathbf{O}\},$$

where

$$\sigma_s(\mathbf{X})_{[1]} = \inf_{\|\mathbf{Z}\|_{[0]} \leq s} \|\mathbf{X} - \mathbf{Z}\|_{[1]}.$$

Based on the definitions above, we have the following Theorem 1 which indicates the link between the general null space property and the recovery of sparse matrix via ($\mathbf{P}_1$).

**Theorem 1** *Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$, every matrix $\mathbf{X} \in \mathbb{R}^{N \times l}$ supported on a index set $S$ is the unique solution of $(P_1)$ with $\mathbf{Y} = \mathbf{AX}$ if and only if $\mathbf{A}$ satisfies the general null space property relative to $S$.*

**Proof** Given a fixed index set $S$, let us first assume that every matrix $\mathbf{X} \in \mathbb{R}^{N \times l}$ supported on $S$ is the unique minimizer of $\|\mathbf{Z}\|_{[1]}$ subject to $\mathbf{AZ} = \mathbf{AX}$. Thus, for any $\mathbf{V} \in Gker(\mathbf{A})\backslash\{\mathbf{O}\}$, the matrix $\mathbf{V}_S$ is the unique minimizer of $\|\mathbf{Z}\|_{[1]}$ subject to $\mathbf{AZ} = \mathbf{AV}_S$. At the same time, we have $\mathbf{A}(-\mathbf{V}_{\overline{S}}) = \mathbf{AV}_S$ and $-\mathbf{V}_{\overline{S}} \neq \mathbf{V}_S$, because $\mathbf{A}(\mathbf{V}_{\overline{S}} + \mathbf{V}_S) = \mathbf{AV} = \mathbf{O}$ and $\mathbf{V} \neq \mathbf{O}$. We conclude that $\|\mathbf{V}_S\|_{[1]} < \|\mathbf{V}_{\overline{S}}\|_{[1]}$. This established the general null space property relative to $S$.

On the other hand, suppose the general null space property relative to index set $S$ holds. Then, given a matrix $\mathbf{X} \in \mathbb{R}^{N \times l}$ supported on $S$ and a matrix $\mathbf{Z} \in \mathbb{R}^{N \times l}$, $\mathbf{Z} \neq \mathbf{X}$, satisfying $\mathbf{AZ} = \mathbf{AX}$. We consider the matrix $\mathbf{V} := \mathbf{X} - \mathbf{Z} \in Gker(\mathbf{A})\backslash\{\mathbf{O}\}$. According to the general null space property, we get

$$\begin{aligned}
\|\mathbf{X}\|_{[1]} &\leq \|\mathbf{X} - \mathbf{Z}_S\|_{[1]} + \|\mathbf{Z}_S\|_{[1]} \\
&= \|\mathbf{V}_S\|_{[1]} + \|\mathbf{Z}_S\|_{[1]} \\
&< \|\mathbf{V}_{\overline{S}}\|_{[1]} + \|\mathbf{Z}_S\|_{[1]} \\
&= \|-\mathbf{Z}_{\overline{S}}\|_{[1]} + \|\mathbf{Z}_S\|_{[1]} \\
&= \|\mathbf{Z}\|_{[1]}.
\end{aligned}$$

The required minimality of $\|\mathbf{X}\|_{[1]}$ is obtained.                          □

We immediately get the following Theorem 2 as a consequence of Theorem 1 by varying the index set $S$.

**Theorem 2** *Given a matrix* $\mathbf{A} \in \mathbb{R}^{m \times N}$, *every s-sparse matrix* $\mathbf{X} \in \mathbb{R}^{N \times l}$ *is the unique solution of* $(P_1)$ *with* $\mathbf{Y} = \mathbf{A}\mathbf{X}$ *if and only if* $\mathbf{A}$ *satisfies the general null space property of order s.*

## 3 Stable General Null Space Property

The matrices we intend to recover via $\mathbf{P}_1$ are sparse only in idealized situations. In more realistic circumstances, we can only claim that they approximate to the sparse matrices. In such cases, we would like to recover a matrix $\mathbf{X} \in \mathbb{R}^{N \times l}$ with an error controlled by its distance to the sparse matrices. This property is often referred to as the stability of the reconstruction scheme with respect to sparsity defect. First of all, we give the following definition.

**Definition 4** A matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ is said to satisfy the stable general null space property with constant $0 < \rho < 1$ relative to a index set $S \subset \Omega$, if

$$\|\mathbf{V}_S\|_{[1]} \leq \rho \|\mathbf{V}_{\overline{S}}\|_{[1]} \quad for \ all \quad \mathbf{V} \in Gker(\mathbf{A}).$$

It is said to satisfy the stable general null space property of order $s$ with constant $0 < \rho < 1$ if it satisfies the stable general null space property with constant $0 < \rho < 1$ relative to any index set $S \subset \Omega$ with $card(S) \leq s$, where $\Omega$ is the same as in the Definition 3.

The stability result with respect to the stable general null space property is the following theorem.

**Theorem 3** *Suppose that a matrix* $\mathbf{A} \in \mathbb{R}^{m \times N}$ *satisfies the stable general null space property of order s with constant* $0 < \rho < 1$. *Then, for any* $\mathbf{X} \in \mathbb{R}^{N \times l}$, *a solution* $\mathbf{X}^\sharp$ *of* $(\mathbf{P}_1)$ *with* $\mathbf{Y} = \mathbf{A}\mathbf{X}$ *approximates the matrix* $\mathbf{X}$ *with* $\|\cdot\|_{[1]}$ *error*

$$\|\mathbf{X} - \mathbf{X}^\sharp\|_{[1]} \leq 2\frac{1+\rho}{1-\rho}\sigma_s(\mathbf{X})_{[1]}. \tag{3.1}$$

We would like to prove a stronger theorem in the following. The result is a statement valid for any index set $S$ in which the matrix $\mathbf{X}^\sharp \in \mathbb{R}^{N \times l}$ is replaced by any matrix $\mathbf{Z} \in \mathbb{R}^{N \times l}$ satisfying $\mathbf{A}\mathbf{Z} = \mathbf{A}\mathbf{X}$. Apart from improving Theorem 3 above, the result declares that, under the stable general null space property relative to index set $S$, the distance between a matrix $\mathbf{X} \in \mathbb{R}^{N \times l}$ supported on $S$ and a matrix $\mathbf{Z} \in \mathbb{R}^{N \times l}$ satisfying $\mathbf{A}\mathbf{Z} = \mathbf{A}\mathbf{X}$ is controlled by the difference between their $\|\cdot\|_{[1]}$ norms.

**Theorem 4** *The matrix* $\mathbf{A} \in \mathbb{R}^{m \times N}$ *satisfies the stable general null space property with constant* $0 < \rho < 1$ *relative to index set S if and only if*

$$\|\mathbf{Z} - \mathbf{X}\|_{[1]} \leq \frac{1+\rho}{1-\rho}(\|\mathbf{Z}\|_{[1]} - \|\mathbf{X}\|_{[1]} + 2\|\mathbf{X}_{\overline{S}}\|_{[1]}) \tag{3.2}$$

*for all matrices* $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{N \times l}$ *with* $\mathbf{A}\mathbf{Z} = \mathbf{A}\mathbf{X}$.

The error bound in (3.1) can be derived from Theorem 4 as follows: Take $S$ to be a index set of $s$ largest absolute elements of matrix $\mathbf{X}$, so that $\|\mathbf{X}_{\overline{S}}\|_{[1]} = \sigma_s(\mathbf{X})_{[1]}$. If $\mathbf{X}^{\sharp}$ is a minimizer of ($\mathbf{P}_1$), then $\|\mathbf{X}^{\sharp}\|_{[1]} \le \|\mathbf{X}\|_{[1]}$ and $\mathbf{AX}^{\sharp} = \mathbf{AX}$. The right hand side of inequality (3.2) with $\mathbf{Z} = \mathbf{X}^{\sharp}$ can therefore be estimated by the right hand of (3.1).

Before proving Theorem 4, we first identify the following Lemma, which will be needed in the following proof later.

**Lemma 1** *Given a index set $S \subset \Omega$ and matrices $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{N \times l}$, it follows that*

$$\|(\mathbf{X} - \mathbf{Z})_{\overline{S}}\|_{[1]} \le \|\mathbf{Z}\|_{[1]} - \|\mathbf{X}\|_{[1]} + \|(\mathbf{X} - \mathbf{Z})_S\|_{[1]} + 2\|\mathbf{X}_{\overline{S}}\|_{[1]}.$$

*Proof* The result can be obtained from

$$\begin{aligned}
\|\mathbf{X}\|_{[1]} &= \|\mathbf{X}_{\overline{S}}\|_{[1]} + \|\mathbf{X}_S\|_{[1]} \\
&\le \|\mathbf{X}_{\overline{S}}\|_{[1]} + \|(\mathbf{X} - \mathbf{Z})_S\|_{[1]} + \|\mathbf{Z}_S\|_{[1]}, \\
\|(\mathbf{X} - \mathbf{Z})_{\overline{S}}\|_{[1]} &\le \|\mathbf{X}_{\overline{S}}\|_{[1]} + \|\mathbf{Z}_{\overline{S}}\|_{[1]}.
\end{aligned}$$

Summing these two inequalities, we get

$$\|\mathbf{X}\|_{[1]} + \|(\mathbf{X} - \mathbf{Z})_{\overline{S}}\|_{[1]} \le 2\|\mathbf{X}_{\overline{S}}\|_{[1]} + \|(\mathbf{X} - \mathbf{Z})_S\|_{[1]} + \|\mathbf{Z}\|_{[1]}.$$

The proof is completed. $\square$

Having the above lemma 1, we can prove Theorem 4.

**Proof** of Theorem 4. First, we assume the matrix $\mathbf{A}$ satisfies (3.2) for all matrices $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{N \times l}$ with $\mathbf{AZ} = \mathbf{AX}$. Given a matrix $\mathbf{V} \in Gker(\mathbf{A})$, since $\mathbf{AV}_{\overline{S}} = \mathbf{A}(-\mathbf{V}_S)$, we can apply (3.2) with $\mathbf{X} = -\mathbf{V}_S$ and $\mathbf{Z} = \mathbf{V}_{\overline{S}}$. Then, we have

$$\|\mathbf{V}\|_{[1]} \le \frac{1 + \rho}{1 - \rho}(\|\mathbf{V}_{\overline{S}}\|_{[1]} - \|\mathbf{V}_S\|_{[1]}).$$

That is

$$(1 - \rho)(\|\mathbf{V}_S\|_{[1]} + \|\mathbf{V}_{\overline{S}}\|_{[1]}) \le (1 + \rho)(\|\mathbf{V}_{\overline{S}}\|_{[1]} - \|\mathbf{V}_S\|_{[1]}).$$

By rearranging the above inequality's terms, we get

$$\|\mathbf{V}_S\|_{[1]} \le \rho\|\mathbf{V}_{\overline{S}}\|_{[1]},$$

and the stable general null space property with constant $0 < \rho < 1$ with respect to index set $S$ is obtained.

On the other hand, suppose the matrix $\mathbf{A}$ satisfies the stable general null space property with constant $0 < \rho < 1$ with respect to index set $S$. For $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{N \times l}$ with $\mathbf{AZ} = \mathbf{AX}$, let $\mathbf{V} := \mathbf{Z} - \mathbf{X}$, then $\mathbf{V} \in Gker(A)$, the stable general null space property declares

$$\|\mathbf{V}_S\|_{[1]} \le \rho\|\mathbf{V}_{\overline{S}}\|_{[1]}. \tag{3.3}$$

At the same time, Lemma 1 gives

$$\|\mathbf{V}_{\overline{S}}\|_{[1]} \le \|\mathbf{Z}\|_{[1]} - \|\mathbf{X}\|_{[1]} + \|\mathbf{V}_S\|_{[1]} + 2\|\mathbf{X}_{\overline{S}}\|_{[1]}. \tag{3.4}$$

Substituting (3.3) into (3.4), we have

$$\|\mathbf{V}_{\overline{S}}\|_{[1]} \le \|\mathbf{Z}\|_{[1]} - \|\mathbf{X}\|_{[1]} + \rho\|\mathbf{V}_{\overline{S}}\|_{[1]} + 2\|\mathbf{X}_{\overline{S}}\|_{[1]}.$$

Owing to $\rho < 1$, the above inequality can be transformed into

$$\|\mathbf{V}_{\overline{S}}\|_{[1]} \le \frac{1}{1-\rho}(\|\mathbf{Z}\|_{[1]} - \|\mathbf{X}\|_{[1]} + 2\|\mathbf{X}_{\overline{S}}\|_{[1]}).$$

By reusing (3.3), we get

$$\begin{aligned}
\|\mathbf{V}\|_{[1]} &= \|\mathbf{V}_{\overline{S}}\|_{[1]} + \|\mathbf{V}_S\|_{[1]} \\
&\le (1+\rho)\|\mathbf{V}_{\overline{S}}\|_{[1]} \\
&\le \frac{1+\rho}{1-\rho}(\|\mathbf{Z}\|_{[1]} - \|\mathbf{X}\|_{[1]} + 2\|\mathbf{X}_{\overline{S}}\|_{[1]}),
\end{aligned}$$

which is the desired result.                                                   □

## 4 Robust General Null Space Property

In practice, it is impossible to measure a data matrix $\mathbf{Z} \in \mathbb{R}^{N \times l}$ precisely. This implies that the measurable matrix $\mathbf{Y} \in \mathbb{R}^{m \times l}$ is only an approximation of the matrix $\mathbf{AZ} \in \mathbb{R}^{m \times l}$, with

$$\|\mathbf{AZ} - \mathbf{Y}\|_F \le \eta$$

for some $\eta \ge 0$. In this case, the recovery strategy should be required to produce a data matrix $\mathbf{Z}^* \in \mathbb{R}^{N \times l}$ whose distance to the original data matrix $\mathbf{Z} \in \mathbb{R}^{N \times l}$ is controlled by the measurement error $\eta \ge 0$. This property is called the robustness of the recovery strategy in respect of the measurement error. We will investigate the following problem in this section

$$\min_{\mathbf{Z} \in \mathbb{R}^{N \times l}} \|\mathbf{Z}\|_{[1]} \quad subject \quad to \quad \|\mathbf{AZ} - \mathbf{Y}\|_F \le \eta. \tag{4.1}$$

Before investigating the problem (4.1), we first give a more demanding definition of the general null space property, which guarantees the robustness of the reconstruction scheme.

**Definition 5** The matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ is said to satisfy the robust general null space property with constants $0 < \rho < 1$ and $\tau > 0$ relative to a index set $S$ if

$$\|\mathbf{V}_S\|_{[1]} \leq \rho \|\mathbf{V}_{\overline{S}}\|_{[1]} + \tau \|\mathbf{A}\mathbf{V}\|_F \quad for \quad all \quad \mathbf{V} \in \mathbb{R}^{N \times l}. \tag{4.2}$$

It is said to satisfy the robust general null space property of order $s$ with constants $0 < \rho < 1$ and $\tau > 0$ if it satisfies the robust general null space property with constants $\rho, \tau$ relative to any index matrix $S$ with $card(S) \leq s$.

Notice that the definition above does not require that $\mathbf{V}$ is contained in $Gker(\mathbf{A})$. Especially, when $\mathbf{V} \in Gker(\mathbf{A})$, the term $\|\mathbf{A}\mathbf{V}\|_F$ in (4.2) disappears, and that is the case of stable general null space property in Definition 4.

Based on the Definition 5, we have the result that the following Theorem 5 affirms.

**Theorem 5** *The matrix* $\mathbf{A} \in \mathbb{R}^{m \times N}$ *satisfies the robust general null space property with constant* $0 < \rho < 1$ *and* $\tau > 0$ *relative to index set* $S$ *if and only if*

$$\|\mathbf{Z} - \mathbf{X}\|_{[1]} \leq \frac{1+\rho}{1-\rho}(\|\mathbf{Z}\|_{[1]} - \|\mathbf{X}\|_{[1]} + 2\|\mathbf{X}_{\overline{S}}\|_{[1]}) + \frac{2\tau}{1-\rho}\|\mathbf{A}(\mathbf{Z} - \mathbf{X})\|_F \tag{4.3}$$

*for all matrices* $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{N \times l}$.

**Proof** Firstly, let the matrix $\mathbf{A}$ satisfy (4.3) for all matrices $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{N \times l}$. Then, for $\mathbf{V} \in \mathbb{R}^{N \times l}$, taking $\mathbf{X} = -\mathbf{V}_S$ and $\mathbf{Z} = \mathbf{V}_{\overline{S}}$, we have

$$\|\mathbf{V}\|_{[1]} \leq \frac{1+\rho}{1-\rho}(\|\mathbf{V}_{\overline{S}}\|_{[1]} - \|\mathbf{V}_S\|_{[1]}) + \frac{2\tau}{1-\rho}\|\mathbf{A}\mathbf{V}\|_F.$$

After rearranging the terms, we have

$$(1-\rho)(\|\mathbf{V}_S\|_{[1]} + \|\mathbf{V}_{\overline{S}}\|_{[1]}) \leq (1+\rho)(\|\mathbf{V}_{\overline{S}}\|_{[1]} - \|\mathbf{V}_S\|_{[1]}) + 2\tau\|\mathbf{A}\mathbf{V}\|_F,$$

that is

$$\|\mathbf{V}_S\|_{[1]} \leq \rho\|\mathbf{V}_{\overline{S}}\|_{[1]} + \tau\|\mathbf{A}\mathbf{V}\|_F.$$

This is the robust general null space property with constants $0 < \rho < 1$ and $\tau > 0$ relative to index set $S$.

Reversely, suppose the matrix $\mathbf{A}$ satisfy the robust general null space property with constants $0 < \rho < 1$ and $\tau > 0$ relative to index set $S$. For $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{N \times l}$, letting $\mathbf{V} := \mathbf{Z} - \mathbf{X}$, according to the robust general null space property and lemma 1, we have

$$\|\mathbf{V}_S\|_{[1]} \leq \rho\|\mathbf{V}_{\overline{S}}\|_{[1]} + \tau\|\mathbf{A}\mathbf{V}\|_F,$$
$$\|\mathbf{V}_{\overline{S}}\|_{[1]} \leq \|\mathbf{Z}\|_{[1]} - \|\mathbf{X}\|_{[1]} + \|\mathbf{V}_S\|_{[1]} + 2\|\mathbf{X}_{\overline{S}}\|_{[1]}.$$

From these two inequalities, we can get

$$\|\mathbf{V}_{\overline{S}}\|_{[1]} \leq \frac{1}{1-\rho}(\|\mathbf{Z}\|_{[1]} - \|\mathbf{X}\|_{[1]} + 2\|\mathbf{X}_{\overline{S}}\|_{[1]} + \tau\|\mathbf{AV}\|_F).$$

According to the robust general null space property once again, we can obtain

$$\begin{aligned} \|\mathbf{V}\|_{[1]} &= \|\mathbf{V}_{\overline{S}}\|_{[1]} + \|\mathbf{V}_S\|_{[1]} \\ &\leq (1+\rho)\|\mathbf{V}_{\overline{S}}\|_{[1]} + \tau\|\mathbf{AV}\|_F \\ &\leq \frac{1+\rho}{1-\rho}(\|\mathbf{Z}\|_{[1]} - \|\mathbf{X}\|_{[1]} + 2\|\mathbf{X}_{\overline{S}}\|_{[1]}) + \frac{2\tau}{1-\rho}\|\mathbf{AV}\|_F, \end{aligned}$$

which is the desired result. □

## 5 Conclusions

In this paper, the null space property has been extended to the general null space property, and the corresponding results with respect to the general null space property have been proved including the stable and the robust cases. The extended concepts can be used to analysis sparse principal components theoretically. In addition to these, several possible avenues of further research work still remain to be done. Such as how to combine these theoretical analysis with practice effectively.

**Data Availability Statement** Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

## Declarations

**Conflict of interest** The authors declare that there are no known conflicts of interest associated with this publication.

## References

1. I.T. Jolliffe, *Principal Component Analysis*, 2nd edn. (Springer, New York, 2002)
2. S. Gajjar, M. Kulahci, A. Palazoglu, Use of sparse principal component analysis (SPCA) for fault detection. IFAC-PapersOnLine **49**(7), 693–698 (2016)
3. A. D'Aspremont, F. Bach, L.E. Ghaoui, Optimal solutions for sparse principal component analysis. J. Mach. Learn. Res. **9**, 1269–1294 (2008)
4. K. Fang, X. Fan, Q. Zhang, S. Ma, Integrative sparse principal component analysis. J. Multivar. Anal. **166**, 1–16 (2018)
5. I.T. Jolliffe, N.T. Trendafilov, M. Uddin, A modified principal component technique based on the lasso. J. Comput. Graph. Stat. **12**(3), 531–547 (2003)

6. A. D'Aspremont, L.E. Ghaoui, M.I. Jordan, G.R.G. Lanckriet, A direct formulation for sparse PCA using semidefinite programming. SIAM Rev. **49**(3), 434–448 (2007)
7. H. Shen, J.Z. Huang, Sparse principal component analysis via regularized low rank matrix approximation. J. Multivar. Anal. **99**(6), 1015–1034 (2008)
8. M. Journée, Y. Nesterov, P. Richtrik, R. Sepulchre, Generalized power method for sparse principal component analysis. Core Discuss. Pap. **11**(2008070), 517–553 (2010)
9. C. Sigg, J.M. Buhmann, Expectation-maximization for sparse and non-negative PCA, in Proceedings of 25th International Conference on Machine Learning (ICML), ACM, 960–967 (2008)
10. H. Zou, T. Hastie, Regularization and variable selection via the elastic net. J. R. Stat. Soc. **67**(2), 301–320 (2005)
11. H. Zou, T. Hastie, R. Tibshirani, Sparse principal component analysis. J. Comput. Graph. Stat. **15**, 265–286 (2004)
12. S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing* (Springer, New York, 2013)
13. S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit. SIAM Rev. **43**(1), 129–159 (2001)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.