Robust estimation of principal components: a literature review

Holger Cevallos-Valdiviezo, Ph.D.[1], Ariana Rodríguez-Cristiansen, M.Sc.[1], and Patricia Valdiviezo-Valenzuela, M.Sc.[1,2]

[1]Escuela Superior Politécnica del Litoral, ESPOL, (Facultad de Ciencias Naturales y Matemáticas), Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador, holgceva@espol.edu.ec, aristrod@espol.edu.ec, pvaldi@espol.edu.ec

[2]Universidad Católica de Santiago de Guayaquil, Ecuador, aristrod@espol.edu.ec, pvaldi@espol.edu.ec

***Abstract*– In this work we do a short literature review on the most relevant methods for robust estimation of Principal Component Analysis (PCA). In particular, we review methods for PCA that are resistant against rowwise outliers, cellwise outliers and against both rowwise and cellwise outliers. It is well known that classical PCA breaks down in the presence of outliers. In practical applications, we suggest to fit a robust method for PCA estimation that is resistant to rowwise and cellwise outliers. We could later compare this result with the classical fit to evaluate the influence of outliers. Robust methods for PCA can also be used to detect outliers.**

***Keywords*-- Outliers, Principal component analysis, Outlier detection, Robust estimation.**

Digital Object Identifier (DOI):
http://dx.doi.org/10.18687/LACCEI2022.1.1.574
ISBN: 978-628-95207-0-5 **ISSN:** 2414-6390

Estimación de componentes principales resistente a valores atípicos: una revisión de la literatura

Holger Cevallos-Valdiviezo, Ph.D.[1], Ariana Rodríguez-Cristiansen, M.Sc.[1], and Patricia Valdiviezo-Valenzuela, M.Sc.[1,2]

[1]Escuela Superior Politécnica del Litoral, ESPOL, (Facultad de Ciencias Naturales y Matemáticas), Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador, holgceva@espol.edu.ec, aristrod@espol.edu.ec, pvaldi@espol.edu.ec

[2]Universidad Católica de Santiago de Guayaquil, Ecuador, aristrod@espol.edu.ec, pvaldi@espol.edu.ec

***Abstract*– In this work we do a short literature review on the most relevant methods for robust estimation of Principal Component Analysis (PCA). In particular, we review methods for PCA that are resistant against rowwise outliers, cellwise outliers and against both rowwise and cellwise outliers. It is well known that classical PCA breaks down in the presence of outliers. In practical applications, we suggest to fit a robust method for PCA estimation that is resistant to rowwise and cellwise outliers. We could later compare this result with the classical fit to evaluate the influence of outliers. Robust methods for PCA can also be used to detect outliers.**

***Keywords*-- Outliers, Principal component analysis, Outlier detection, Robust estimation.**

***Resumen*– En este trabajo se realiza una revisión literaria de los métodos robustos más relevantes para Análisis de Componentes Principales (ACP). En particular, revisamos métodos robustos para ACP resistentes a valores atípicos por filas, a valores atípicos por celdas y a valores atípicos por filas y por celdas. El método clásico de ACP colapsa ante la presencia de valores atípicos. Ante la posible presencia de valores atípicos por filas y por celdas y a la proliferación de estos en datos de gran dimensión, se sugiere ajustar métodos robustos resistentes a filas y a celdas atípicas para estimar el subespacio de menor dimensión. Este ajuste robusto puede ser luego comparado con el ajuste clásico para evaluar la influencia de los valores atípicos. Los métodos robustos para ACP también pueden ser usados para detectar filas y celdas atípicas.**

***Palabras claves*-- Valores atípicos, Análisis de Componentes Principales, Detección de valores atípicos, Estimación robusta.**

I. INTRODUCCIÓN

Una definición de un valor atípico que es generalmente aceptada entre muchos estadísticos y científicos de datos, es aquella que dice que un valor atípico es una observación que se desvía del ajuste sugerido por la mayoría de las observaciones. En aplicaciones reales, los datos muy a menudo contienen valores atípicos. Los datos atípicos pueden representar errores indeseables que pueden afectar el proceso del análisis de los datos o pueden representar piezas de información valiosas que se obtienen de forma inesperada. Por ello, es importante para el investigador poder detectar estos valores atípicos para luego realizar una interpretación de ellos. Muy a menudo, antes del análisis, no se conoce si existen valores atípicos en el conjunto de datos, si existen no se conoce la proporción de valores atípicos que están presentes ni su ubicación en el conjunto de datos. Los métodos estadísticos tradicionales son sensibles ante la presencia de valores atípicos. Ante esto, métodos estadísticos robustos han sido desarrollados que tratan de ajustar el modelo impuesto por la mayoría de los datos. El objetivo principal es encontrar un ajuste robusto que sea similar al ajuste que hubiéramos encontrado sin la presencia de valores atípicos. Este ajuste robusto puede ser usado para detectar valores atípicos mediante la identificación de grandes desviaciones con respecto a este ajuste. La identificación de valores atípicos en los datos permite alertar al analista sobre posibles problemas en el proceso de recolección de datos e investigar sus causas.

El campo de la estadística robusta se ha desarrollado desde la década de los 60 en los diferentes contextos de la estadística con el desarrollo de un gran número de metodologías resistentes a la presencia de observaciones atípicas [1], [2]. Estas observaciones atípicas son también llamadas datos atípicos por fila (o rowwise outliers en inglés), ya que los datos multivariantes típicamente se representan por medio de una matriz rectangular en la cual las filas son las observaciones y las columnas son las variables. Los métodos robustos por fila requieren en general que menos de la mitad de las filas de la matriz de datos sean atípicas (ver por ej. [3]). En muchas aplicaciones prácticas, sin embargo, solo ciertas celdas en una fila de la matriz de datos contienen valores atípicos, mientras que las demás celdas de la fila representan piezas valiosas de información. Este nuevo paradigma fue presentado por primera vez por Alqallaf *et al.* [4] y se lo conoce como datos atípicos por celdas (o cellwise outliers en inglés). Alqallaf *et al.* [4] notaron además que en problemas de grandes dimensiones (por ej. cuando el número de variables p es mayor al número de observaciones n, i.e. $p > n$), es posible que un pequeño número de celdas atípicas puedan contaminar a la mayoría de las observaciones de la matriz de datos, cuando la contaminación en las celdas es producida en posiciones aleatorias. Bajo este escenario, los métodos robustos por fila "colapsan" ya que estos métodos no toleran una proporción de más del 50% de filas atípicas (ver por ej. [3]) En los últimos años se han desarrollado métodos para lidiar con el paradigma por celdas que tratan de explotar la información valiosa de las celdas regulares en cada fila y que son robustos incluso si la mayoría de las observaciones están contaminadas (ver por ej. [4]–[7]).

En este trabajo, realizaremos una revisión literaria de los métodos robustos más relevantes que se han desarrollado para

Digital Object Identifier (DOI):
http://dx.doi.org/10.18687/LACCEI2022.1.1.574
ISBN: 978-628-95207-0-5 **ISSN:** 2414-6390

Análisis de Componentes Principales (ACP). ACP es una herramienta exploratoria popular para datos multivariantes. En particular, ACP es útil para encontrar una representación en menor dimensión que produce la mejor aproximación posible a los datos originales. ACP clásico minimiza las distancias euclídeas al cuadrado entre las observaciones originales y sus proyecciones ortogonales sobre el subespacio de menor dimensión. De forma equivalente, ACP clásico busca de forma secuencial direcciones definidas por vectores unitarios que maximizan la varianza clásica (no robusta) de los datos proyectados, bajo la restricción que estas direcciones sean ortogonales una respecto de la otra. La solución de ACP clásico viene dada por los primeros vectores propios correspondientes a los valores propios de mayor valor de la matriz de covarianza clásica, los cuales pueden obtenerse al realizar la descomposición espectral de la matriz de covarianza clásica. Una alternativa para obtener la solución de ACP clásico que no requiere el cálculo de la matriz de covarianza ni su descomposición espectral es mediante el algoritmo "Nonlinear Iterative Partial Least Squares" (NIPALS) [8], [9]. NIPALS obtiene la solución de ACP clásico mediante secuencias iterativas de proyecciones ortogonales. El algoritmo NIPALS posee algunas ventajas sobre el método que utiliza la descomposición espectral. En particular, NIPALS puede trabajar con datos incompletos y puede lidiar con datos de gran dimensión y con problemas de colinealidad de las variables.

Sin embargo, ACP clásico es muy sensible a la presencia de datos atípicos ya que usa la función de pérdida cuadrática. Por tanto, algunos enfoques para robustificar ACP se han propuesto en la literatura científica. El objetivo de este trabajo es presentar una perspectiva de las diferentes técnicas para ACP robusto disponibles en la literatura científica y presentar algunas de sus características. El lector puede profundizar sobre algunas definiciones presentadas en este trabajo en las referencias que se enlistan a lo largo del texto. La siguiente sección de este trabajo presenta una revisión literaria de los principales métodos robustos desarrollados para ACP, los cuales hemos clasificado en tres categorías: aquellos resistentes a datos atípicos por fila, aquellos resistentes a datos atípicos por celdas y aquellos resistentes tanto a celdas como a filas atípicas. La sección 3 describe nuestras principales conclusiones.

II. ACP Robusto: Revisión de la Literatura

A. ACP robusto contra filas atípicas

El enfoque más antiguo y sencillo para robustificar ACP contra filas atípicas consiste en tomar los vectores y valores propios de un estimador robusto de la matriz de covarianza en lugar de la matriz clásica de covarianza muestral (ver por ej. [10]–[14]). Sin embargo, este enfoque no puede ser usado en aplicaciones con datos de gran dimensión pues calcular una matriz de covarianza robusta de gran dimensión es computacionalmente complejo o inclusive no factible si el tamaño muestral es pequeño en relación a la dimensión de los datos. Además, aunque la eficiencia de los estimadores robustos de dispersión se incrementa con la dimensión, esto sucede a expensas de pérdida de robustez. Una alternativa robusta y que puede ser fácil de calcular son las ACP esféricas propuestas por Locantore *et al.* [15]. Esta técnica usa la matriz de covarianza de los datos proyectados sobre la esfera unitaria y, por tanto, su cómputo puede ser efectuado rápidamente.

Otro enfoque para robustificar ACP contra filas atípicas busca de manera secuencial direcciones univariantes ortogonales, una respecto de la otra, que maximizan un estimador robusto de escala de las proyecciones. Este enfoque robusto de búsqueda por proyecciones (también conocido como "projection pursuit" en inglés) ha sido estudiado por ejemplo por [16]–[18].

En lugar de buscar direcciones de forma secuencial, podríamos optar por estimar directamente y de manera robusta el subespacio de menor dimensión (ver por ej. [19], [20]). ROBPCA [21] busca el subespacio de menor dimensión a través de un procedimiento de pasos múltiples. Brevemente, el primer paso intenta identificar un subconjunto de $n/2 \leq h < n$ observaciones con base en la medida de atipicidad de Stahel-Donoho. Luego, se aplica ACP clásico en este subconjunto de tamaño h y se determina la dimensión óptima del subespacio. En el siguiente paso, los puntos de los datos son proyectados sobre el subespacio de menor dimensión. Finalmente, se calcula el estimador de covarianza de determinante mínimo reponderado en los datos proyectados para obtener las estimaciones de las direcciones principales del subespacio. Los autores en [21], Hubert *et al.*, proponen además un gráfico de diagnóstico que permite detectar observaciones atípicas al graficar las distancias ortogonales entre las observaciones y el subespacio estimado robusto, versus, las distancias robustas de las proyecciones de las observaciones sobre el subespacio (o scores) con respecto a su centro. Hubert *et al.* [21] proponen puntos críticos para las distancias ortogonales y para las distancias de los scores afín de clasificar cada observación en una de las siguientes categorías: observaciones regulares, observaciones buenas de apalancamiento, observaciones malas de apalancamiento y observaciones atípicas ortogonales. Las observaciones buenas de apalancamiento son aquellas que se encuentran cerca del subespacio de componentes principales pero cuya proyección sobre el subespacio está alejada de las proyecciones típicas. Aunque las observaciones buenas de apalancamiento no influyen en la estimación del subespacio de componentes principales, estas aún pueden afectar la estimación de los vectores y valores propios [22]. Las observaciones malas de apalancamiento, por otro lado, son aquellas que se encuentran alejadas del subespacio (i.e. están a una distancia ortogonal grande con respecto al subespacio) y además su proyección en el subespacio es remota con respecto a las proyecciones típicas. Por su parte, las observaciones atípicas ortogonales son aquellas que están alejadas ortogonalmente del subespacio pero que sus proyecciones se encuentran cercanas a las proyecciones típicas. Tanto las observaciones malas de apalancamiento como las observaciones atípicas ortogonales son influyentes en la estimación del subespacio. En la referencia [23] se muestra el uso de este gráfico de diagnóstico para detectar observaciones atípicas mediante ejemplos.

Siguiendo la misma línea que [21], Maronna [24] propuso estimar de forma robusta el subespacio minimizando sea un M-estimador de escala o un estimador de escala "least trimmed squares" (LTS) de las distancias euclídeas entre las observaciones y sus proyecciones ortogonales sobre el subespacio. La estimación propuesta por Maronna [24] es robusta en el sentido de que emplea estimadores robustos de escala en el procedimiento de optimización en lugar de emplear la desviación estándar clásica y no robusta. La robustez de estos estimadores del subespacio ha sido ampliamente investigada de manera empírica. Ver por ejemplo [24]–[26] para el caso del estimador basado en el M-estimador de escala y [24], [27], [28] para el caso del estimador basado en la escala LTS. El trabajo en [27] contribuye con un estudio exhaustivo de las propiedades teóricas del estimador basado en la escala LTS. Maronna [24] caracterizó las soluciones de estos estimadores por medio de las direcciones del complemento ortogonal y mostró que estas direcciones corresponden con los vectores propios asociados a los valores propios más pequeños de una matriz de covarianza ponderada. En base a esta caracterización, Maronna [24] propuso un algoritmo iterativo para calcular los estimadores del subespacio robusto. Sin embargo, en caso de una aproximación de baja dimensión para datos de gran dimensión, el algoritmo de Maronna requiere la descomposición de una matriz de covarianza de gran dimensión y necesita de un gran número de sus vectores propios para caracterizar la solución. Esto hace que el cálculo de los estimadores del subespacio de Maronna [24] requiera mucho tiempo o sea incluso imposible de calcular en grandes dimensiones. Cevallos-Valdiviezo y Van Aelst [29] proponen un algoritmo relativamente rápido para los estimadores desarrollados en [24] que hace posible la estimación del subespacio, inclusive para casos de grandes dimensiones, basado en el cálculo directo de las direcciones principales del subespacio, el uso de las condiciones de primer orden correspondientes a los estimadores para actualizar las soluciones de forma iterativa con operaciones entre vectores y matrices de baja dimensión (sin tener que manipular matrices de gran dimensión) y el empleo de cinco valores iniciales determinísticos y robustos que permiten la convergencia rápida del algoritmo. El algoritmo rápido de Cevallos-Valdiviezo y Van Aelst [29] es similar al algoritmo NIPALS. Tanto el algoritmo de Cevallos-Valdiviezo y Van Aelst [29] como el algoritmo NIPALS no requieren el cálculo y la descomposición de una matriz de covarianza para estimar el subespacio de menor dimensión, lo cual es una ventaja sobre todo en datos con grandes dimensiones. Sin embargo, el algoritmo de Cevallos-Valdiviezo y Van Aelst [29] es robusto y no necesariamente obtiene direcciones ortogonales, por lo que es posible que requiera un paso de ortogonalización al finalizar el procedimiento. Por su parte, la versión estándar de NIPALS no es robusta ante valores atípicos ya que obtiene direcciones ortogonales que corresponden con la solución del ACP clásico, aunque puede ser adaptado para obtener una solución robusta (e.g. podría ser adapatado para obtener la solución de los estimadores en [24]).

B. *ACP robusto contra celdas atípicas*

En los últimos años ha recibido atención el problema de datos atípicos por celdas. En el contexto de ACP, el método de búsqueda de componentes principales llamado en inglés "Principal Component Pursuit" (PCP) [30] es una de las primeras contribuciones en este sentido. PCP robustifica ACP contra celdas atípicas de forma natural en su formulación al descomponer la matriz de datos en un componente de rango bajo y un componente disperso que refleja los valores atípicos en las celdas de la matriz de datos. Las referencia [31], por ejemplo, contiene un trabajo relacionado. PCP, sin embargo, puede fracasar en la detección de datos atípicos presentes en el complemento ortogonal del subespacio, tal como se muestra en [32] y [33]. She *et al.* [33] investigan este problema al proponer el método ROCPCA. El método ROCPCA modela los datos proyectados sobre el complemento ortogonal del subespacio a través de un término para la media, una matriz dispersa que identifica los valores atípicos ortogonales y un término asociado al ruido inherente del proceso. She *et al.* [33] proponen dos formulaciones para ROCPCA. Una es robusta contra datos atípicos por fila en el complemento ortogonal (llamada r-Type o rowwise Type). La otra formulación es robusta contra datos atípicos por celdas en el complemento ortogonal (llamada e-Type o elementwise Type). ROCPCA introduce penalizaciones sobre la matriz dispersa para forzar la dispersión por fila o por celda, según sea la necesidad. Debido a que ROCPCA modela los datos atípicos en el complemento ortogonal, la estimación del subespacio por este método puede ser influenciada por datos atípicos en el espacio original de las observaciones. Por ello, Brahma *et al.* [34] combinaron las ideas de los métodos PCP y ROCPCA para construir un método que ellos llamaron PCP Robusto Reforzado (RRPCA por sus siglas en inglés). RRPCA considera datos atípicos en el espacio de las observaciones originales (como lo hace PCP) y datos atípicos en el complemento ortogonal del subespacio (como lo hace ROCPCA). RRPCA descompone la matriz de datos en un componente de bajo rango, una matriz dispersa para representar los datos atípicos en el espacio de las observaciones originales, un término de ruido, un término para la media y otra matriz dispersa para representar datos atípicos en el complemento ortogonal del subespacio. RRPCA representa estos dos últimos componentes de la descomposición en el complemento ortogonal del subespacio por lo que luego son transformados al espacio de las observaciones originales. RRPCA puede ser formulado para ser robusto contra valores atípicos por fila o por celda en el espacio de las observaciones, y contra valores atípicos por fila o por celda en el espacio del complemento ortogonal. Dependiendo de la formulación que el usuario desee aplicar, RRPCA emplea penalizaciones sobre la respectiva matriz dispersa para forzar la dispersión, ya sea por fila o por celda. Una desventaja de los métodos ROCPCA y RRPCA es que, similar a los estimadores propuestos por Maronna [24], caracterizan el subespacio en términos de las direcciones del complemento ortogonal, lo cual puede resultar computacionalmente intensivo en aplicaciones con datos de gran dimensión. Note que PCP, ROCPCA y RRPCA son métodos basados en la aproximación de bajo rango de la matriz de datos.

Por otro lado, Boente y Salibián-Barrera [35] proponen estimar directamente el subespacio de menor dimensión minimizando la suma de estimadores de escala robustos de las coordenadas de las desviaciones (o residuos). Los autores emplearon M-estimadores de escala y llamaron al procedimiento S-estimadores de componentes principales. Aunque Boente y Salibián-Barrera [35] estudian el problema de ACP robusto en el contexto de datos funcionales, utilizan el resultado de la estimación de los S-estimadores en dimensiones finitas (datos multivariantes) para extender el método a la configuración funcional. Además, los autores demuestran que los S-estimadores de ACP son consistentes para vectores aleatorios elípticos y Fisher consistentes para elementos aleatorios que siguen una distribución elíptica en espacios Hilbert arbitrarios. El método de S-estimadores de ACP es robusto contra celdas atípicas, especialmente si la contaminación se produce por coordenadas en la matriz de datos.

Más recientemente, Hubert *et al.* [22], desarrollaron un método para ACP que estima directamente el subespacio de menor dimensión y que es robusto contra datos atípicos por celdas y por filas simultáneamente. Al mismo tiempo, el algoritmo puede lidiar con valores perdidos. Hubert *et al.* [22] lo llamaron MacroPCA por su nombre en inglés "PCA allowing for Missingness And Cellwise & Rowwise Outliers". MacroPCA empieza aplicando el método llamado DetectDeviatingCells [36] para detectar datos atípicos por celdas y para iniciar las imputaciones para las celdas atípicas y para los valores perdidos. DetectDeviatingCells permite también identificar posibles filas atípicas. En los pasos siguientes, MacroPCA combina ideas del algoritmo de ACP clásico iterativo (ICPCA por sus siglas en inglés) que lidia con valores perdidos [37], [38] con ROBPCA [21] con el fin de proteger la estimación del subespacio contra datos atípicos por fila y para mejorar las imputaciones de celdas atípicas y de valores perdidos.

III. Conclusión

En este trabajo se ha realizado una revisión de los métodos más relevantes para estimación robusta de componentes principales disponibles en la literatura científica. El problema de valores atípicos es un problema muy común en datos de gran dimensión y con muchas observaciones como los que se dispone hoy en día en algunas aplicaciones. Los datos atípicos pueden representar contaminación o información valiosa que ha sido obtenida de forma inesperada. El objetivo de los métodos robustos en general es el de obtener un ajuste resistente a valores atípicos que sea similar al ajuste que obtendríamos en base a los datos regulares. Ya que se espera que valores atípicos no tengan un buen ajuste en relación a la solución robusta, en general los métodos robustos pueden ser también usados para detectar valores atípicos. En una matriz de datos pueden existir filas atípicas y celdas atípicas. Investigadores en el campo de la estadística robusta han investigado desde hace ya mucho tiempo el problema de estimación resistente a observaciones (filas) atípicas. Recientemente ha habido un gran interés en la comunidad estadística por el desarrollo de métodos robustos contra celdas atípicas.

Con datos de gran dimensión, analistas a menudo utilizan ACP para obtener la mejor aproximación de menor dimensión a los datos. El método clásico para ACP colapsa ante la presencia de valores atípicos en la matriz de datos. Dada la posibilidad de que valores atípicos estén presentes por filas y por celdas en los datos a analizar, especialmente en datos de gran dimensión y con muchas observaciones, se sugiere el ajuste de métodos para ACP que sean resistentes a valores atípicos por filas y por celdas. Este ajuste puede ser luego comparado con el ajuste obtenido con el método de ACP clásico. Ante la presencia de valores atípicos potenciales se espera que los ajustes clásicos y robustos difieran considerablemente. El ajuste de ACP robusto puede luego ser usado para detectar valores atípicos. La identificación de valores atípicos proporciona una alarma al investigador sobre posibles problemas en los datos, los cuales pueden ser luego investigados e interpretados.

References

[1] R. A. Maronna, R. D. Martin, V. Yohai, and M. Salibián-Barrera, *Robust statistics: theory and practice.* Ed. New York, NY, USA: Wiley, 2006.

[2] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. John wiley & sons, 2005, pp. 1525–1538.

[3] H. P. Lopuhaa and P. J. Rousseeuw, "Breakdown points of affine equivariant estimators of multivariate location and covariance matrices," *Ann. Stat.*, vol. 19, no. 1, pp. 229–248, Mar. 1991.

[4] F. Alqallaf, S. Van Aelst, V. J. Yohai, and R. H. Zamar, "Propagation of outliers in multivariate data," *Ann. Stat.*, vol. 37, no. 1, pp. 311–331, Feb. 2009.

[5] C. Agostinelli, A. Leung, V. J. Yohai, and R. H. Zamar, "Robust estimation of multivariate location and scatter in the presence of cellwise and casewise contamination," *Test*, vol. 24, no. 3, pp. 441–461, Jun. 2015.

[6] V. Öllerer, A. Alfons, and C. Croux, "The shooting S-estimator for robust regression," *Comput. Stat.*, vol. 31, no. 3, pp. 829–844, Jun. 2016.

[7] S. Van Aelst, E. Vandervieren, and G. Willems, "A Stahel–Donoho estimator based on huberized outlyingness," *Comput. Stat. Data Anal.*, vol. 56, no. 3, pp. 531–542, Mar. 2012.

[8] H. Wold, "Estimation of principal components and related models by iterative least squares," *J. Multivar. Anal.*, pp. 391–420, Mar. 1966.

[9] H. Wold and E. Lyttkens, "Nonlinear iterative partial least squares (NIPALS) estimation procedures," *Bull. Int. Stat. Inst.*, vol. 43, no. 1, pp. 29–47, Apr. 1969.

[10] N. A. Campbell, "Robust procedures in multivariate analysis I: Robust covariance estimation," *J. R. Stat. Soc. Ser. C. Appl. Stat.*, vol. 29, no. 3, pp. 231–237, Mar. 1980.

[11] C. Croux and G. Haesbroeck, "Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies," *Biometrika*, vol. 87, no. 3, pp. 603–618, Sep. 2000.

[12] S. J. Devlin, R. Gnanadesikan, and J. R. Kettenring, "Robust estimation of dispersion matrices and principal components," *J. Am. Stat. Assoc.*, vol. 76, no. 374, pp. 354–362, Jun. 1981.

[13] R. A. Naga and G. Antille, "Stability of robust and non-robust principal components analysis," *Comput. Stat. Data Anal.*, vol. 10, no. 2, pp. 169–174, Oct. 1990.

[14] M. Salibián-Barrera, S. Van Aelst, and G. Willems, "Principal components analysis based on multivariate MM estimators with fast and robust bootstrap," *J. Am. Stat. Assoc.*, vol. 101, no. 475, pp. 1198–1211, Sep. 2006.

[15] N. Locantore, J. S. Marron, D. G. Simpson, N. Tripoli, J. T. Zhang, K. L. Cohen, "Robust principal component analysis for functional data," *Test*, vol. 8, no. 1, pp. 1–73, Jun. 1999.

[16] C. Croux and A. Ruiz-Gazen, "A fast algorithm for robust principal components based on projection pursuit," in *Compstat: Proc. Comput. Stat.*, Aug. 1996, pp. 211–216.

[17] C. Croux and A. Ruiz-Gazen, "High breakdown estimators for principal components: the projection-pursuit approach revisited," *J. Multivar. Anal.*, vol. 95, no. 1, pp. 206–226, Jul. 2005.

[18] G. Li and Z. Chen, "Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo," *J. Am. Stat. Assoc.*, vol. 80, no. 391, pp. 759–766, Sep. 1985.
[19] L. Liu, D. M. Hawkins, S. Ghosh, and S. S. Young, "Robust singular value decomposition analysis of microarray data," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, no. 23, pp. 13167–13172, Oct. 2003.
[20] C. Croux, P. Filzmoser, G. Pison, and P. J. Rousseeuw, "Fitting multiplicative models by robust alternating regressions," *Stat. Comput.*, vol. 13, no. 1, pp. 23–36, Feb. 2003.
[21] M. Hubert, P. J. Rousseeuw, and K. Vanden Branden, "Robpca: a new approach to robust principal component analysis," *Technometrics*, vol. 47, no. 1, pp. 64–79, Feb. 2005.
[22] M. Hubert, P. J. Rousseeuw, and W. Van den Bossche, "Macropca: An all-in-one PCA method allowing for missing values as well as cellwise and rowwise outliers," *Technometrics*, vol. 61, no. 4, pp. 459–473, Jan. 2019.
[23] K. Varmuza and P. Filzmoser, *Introduction to multivariate statistical analysis in chemometrics.* CRC press, 2016.
[24] R. Maronna, "Principal components and orthogonal regression based on robust scales," *Technometrics*, vol. 47, no. 3, pp. 264–273, Aug. 2005.
[25] S. Serneels and T. Verdonck, "Principal component analysis for data containing outliers and missing elements," *Comput. Stat. Data Anal.*, vol. 52, no. 3, pp. 1712–1727, Jan. 2008.
[26] Y. Tharrault, G. Mourot, and J. Ragot, "Fault detection and isolation with robust principal component analysis," in 16*th* *Mediterr. Conf. Control Automation.* Jun. 2008, pp. 59–64.
[27] C. Croux, L. A. García-Escudero, A. Gordaliza, C. Ruwet, and R. S. Martín, "Robust principal component analysis based on trimming around affine subspaces," *Stat. Sin.*, pp. 1437–1459, Jul. 2017.
[28] S. Engelen, M. Hubert, and K. V. Branden, "A comparison of three procedures for robust PCA in high dimensions," *Austrian J. Stat.*, vol. 34, no. 2, pp. 117–126, Jan. 2005.
[29] H. Cevallos-Valdiviezo and S. Van Aelst, "Fast computation of robust subspace estimators," *Comput. Stat. Data Anal.*, vol. 134, pp. 171–185, Jun. 2019.
[30] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM.*, vol. 58, no. 3, pp. 1–37, May. 2011.
[31] K. Y. Chiang, C. J. Hsieh, and I. Dhillon, "Robust principal component analysis with side information," in *ICML.*, Jun. 2016, pp. 2291–2299.
[32] R. A. Maronna, F. Méndez, and V. J. Yohai, "Robust nonlinear principal components," *Stat. Comput.*, vol. 25, no. 2, pp. 439–448, Dec. 2015.
[33] Y. She, S. Li, and D. Wu, "Robust orthogonal complement principal component analysis," *J. Am. Stat. Assoc.*, vol. 111, no. 514, pp. 763–771, Aug. 2016.
[34] P. P. Brahma, Y. She, S. Li, J. Li, and D. Wu, "Reinforced robust principal component pursuit," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1525–1538, May. 2017.
[35] G. Boente and M. Salibian-Barrera, "S-estimators for functional principal component analysis," *J. Am. Stat. Assoc.*, vol. 110, no. 511, pp. 1100–1111, Nov. 2015.
[36] P. J. Rousseeuw and W. V. D. Bossche, "Detecting deviating data cells," *Technometrics*, vol. 60, no. 2, pp. 135–145, Jan. 2018.
[37] H. A. L. Kiers, "Weighted least squares fitting using ordinary least squares algorithms," *Psychometrika*, vol. 62, no. 2, pp. 251–266, Jun. 1997.
[38] P. R. C. Nelson, P. A. Taylor, and J. F. MacGregor, "Missing data methods in PCA and PLS: Score calculations with incomplete observations," *Chemom. Intell. Lab. Syst.*, vol. 35, no. 1, pp. 45–65, Nov. 1996.