



HAL
open science

Principal Component Analysis: A Generalized Gini Approach

Arthur Charpentier, Stéphane Mussard, Téo Ouraga

► **To cite this version:**

Arthur Charpentier, Stéphane Mussard, Téo Ouraga. Principal Component Analysis: A Generalized Gini Approach. *European Journal of Operational Research*, inPress, 10.1016/j.ejor.2021.02.010 . hal-03156702

HAL Id: hal-03156702

<https://hal.science/hal-03156702v1>

Submitted on 22 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Stochastics and Statistics

Principal component analysis: A generalized Gini approach[☆]Arthur Charpentier^a, Stéphane Mussard^{b,*}, Téa Ouraga^b^a UQAM, Québec, Canada^b Chrome Université de Nîmes, France

ARTICLE INFO

Article history:

Received 29 April 2020

Accepted 3 February 2021

Available online 19 February 2021

Keywords:

(R) Multivariate statistics

Gini

PCA

Robustness

ABSTRACT

A principal component analysis based on the generalized Gini correlation index is proposed (Gini PCA). The Gini PCA generalizes the standard PCA based on the variance. It is shown, in the Gaussian case, that the standard PCA is equivalent to the Gini PCA. It is also proven that the dimensionality reduction based on the generalized Gini correlation matrix, that relies on city-block distances, is robust to outliers. Monte Carlo simulations and an application on cars data (with outliers) show the robustness of the Gini PCA and provide different interpretations of the results compared with the variance PCA.

© 2021 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

1. Introduction

The first PCA was introduced by Pearson (1901), projecting a real matrix X onto the eigenvectors of its covariance matrix, and observing that the variances of those projections are the corresponding eigenvalues. One of the key property is that $X^T X$ is a positive matrix. Most statistical properties of PCAs (see Anderson, 1963; Flury & Riedwyl, 1988) are obtained under Gaussian assumptions. Furthermore, geometric properties can be obtained using the fact that the covariance defines an inner product on the subspace of random variables with finite second moment, which is a strong hypothesis that can be dropped with the aid of the Gini index.

This late decade, a line of research has been developed and focused on the Gini methodology, see Yitzhaki and Schechtman (2013) for a general review of different Gini approaches applied in Statistics and in Econometrics.¹ Among the Gini tools, the Gini regression has received a large audience since the Gini regression initiated by Olkin and Yitzhaki (1992). Gini regressions have been generalized by Yitzhaki and Schechtman (2013) in different areas and particularly in time series analysis. Shelef and Schechtman (2011) and Carcea and Serfling (2015) investigated ARMA processes with an identification and an estimation procedure based on Gini autocovariance functions. This robust Gini approach has been shown to be relevant to heavy tailed distributions such as Pareto

processes. Also, Shelef (2016) proposed a unit root test based on Gini regressions to deal with outlying observations in the data.

In parallel to the above literature, a second line of research on multidimensional Gini indices arose. This literature paved the way on the valuation of inequality about multiple commodities or dimensions such as education, health, income, etc., that is, to find a real-valued function that quantifies the inequality between the households of a population over each dimension, see among others, List (1999), Gajdos and Weymark (2005), Decancq and Lugo (2013). More recently, Banerjee (2010) shows that it is possible to construct multidimensional Gini indices by exploring the projection of the data in reduced subspaces based on the Euclidean norm. Accordingly, some notions of linear algebra have been increasingly included in the axiomatization of multidimensional Gini indices.

In this paper, in the same vein as in the second line of research mentioned above, we start from the recognition that linear algebra may be closely related to the maximum level of inequality that arises in a given dimension. In data analysis, the variance maximization is mainly used to further analyze projected data in reduced subspaces. The variance criterion implies many problems since it captures a very precise notion of dispersion, which does not always match some basic properties satisfied by variability measures such as the Gini index. Such a property may be, for example, an invariance condition postulating that a dispersion measure remains constant when the data are transformed by monotonic maps.² Another property typically related to the Gini index is its robustness to outlying observations, see e.g. Olkin and Yitzhaki

[☆] Research Fellow LISER Luxembourg. The authors are indebted to their three reviewers for their very helpful remarks and suggestions. A. Charpentier acknowledges the support of the Natural Sciences and Engineering Research Council of Canada Grant NSERC-2019-07077, and AXA Research Fund Joint Research Initiative.

* Corresponding author.

E-mail address: stephane.mussard@unimes.fr (S. Mussard).

¹ See Giorgi (2013) for an overview of the "Gini methodology".

² See Furman and Zitikis (2017) for the link between variability (risk) measures and the Gini correlation index. See also Laurini and Ohashi (2015) for problem of noisy PCA for pricing interest rate derivatives.

(1992) in the case of linear regressions. Accordingly, it seems natural to analyze multidimensional dispersion with the Gini index, instead of the variance, in order to provide a Principal Component Analysis (PCA) in a Gini sense (Gini PCA).³

In the field of ℓ_2 norm PCA, many authors have proposed robust estimators in order to deal with outlying observations. For instance Laurini and Ohashi (2015) point out that the classical PCA is not suitable for the estimation of principal components on some times series such as forward rate curves because of the presence of measurement errors (inherent to market microstructure effects and numerical interpolations). Then, they propose to employ long run covariance matrices that prevent temporal dependencies due to either serial correlation or to some contamination process.

Another branch of PCA, the ℓ_1 norm PCA, became famous these last two decades. The algorithms underlying the ℓ_1 norm PCA are mainly built on the minimization of the absolute difference between the coordinates of the projected data and the original data. Kwak (2008), show that the ℓ_1 norm PCA is rather less sensitive to outliers compared with the ℓ_2 norm PCA. An algorithm based on either linear or quadratic programming is proposed to obtain a ℓ_1 norm PCA. Despite the robustness of the technique, it remains quite time-consuming and not rotational invariant.

Ding, Zhou, He, and Zha (2006) propose the R1-PCA which is a rotational invariant ℓ_1 norm PCA. It deals properly with outliers and enables rotations, however it strongly depends on the dimension of the subspace in which the data are projected onto. Indeed, the projector of dimension $K - 1$ cannot be deduced from the projector of dimension K .

Since the ℓ_1 norm PCA relies on optimization problems, the literature offers some new techniques of optimization related to the ℓ_1 PCA such as ℓ_1 norm discriminant analysis for image and pattern recognitions, see for instance Li, Shaoa, and Deng (2015) who propose to replace the Euclidean norm by the ℓ_1 norm in order to maximize the between-group variability in a given sample with the aid of an iterative algorithm. Brooks, Dulá, and Boone (2013) suggest the ℓ_1 norm best-fit hyperplane problem leading to a ‘pure’ ℓ_1 PCA and Visentin, Prestwich, and Armagan (2016) a ℓ_1 norm inverse iterative process based on linear programming that minimizes the variance.

Instead of looking for optimization procedures related to the ℓ_1 norm, we propose a closed-form ℓ_1 norm PCA.⁴ Indeed, following the recent works on the Gini index done by Yitzhaki and Schechtman (2013), it is well-known that the Gini covariance function, being a ℓ_1 norm covariance, enables robustness statistical estimations. Baccini, Besse, and de Falguerolles (1996) and Korhonen and Siljamäki (1998) are among the first authors dealing with a ℓ_1 -norm PCA Gini framework. Their idea was to robustify the standard PCA by means of the Gini Mean Difference metric introduced by Gini (1912), which is a city-block distance measure of variability. The authors employ the Gini Mean Difference as an estimator of the standard deviation of each variable before running the singular value decomposition leading to a robust PCA. In what follows, we investigate the employ of the generalized Gini covariance operator in order to obtain a closed-form ℓ_1 PCA without taking recourse to algorithms of optimization.

In particular, it is shown that the variance may be seen as an inappropriate criterion for dimensionality reduction in the case of data contamination or outlying observations. A generalized Gini PCA is investigated by means of Gini correlations matrices. These matrices contain generalized Gini correlation coefficients

³ Recent PCAs derive latent variables thanks to regressions based on elastic net (a ℓ_1 regularization) that improves the quality of the regression curve estimation, see Zou, Hastie, and Tibshirani (2006).

⁴ For ℓ_1 norm PCA with new geometric perspectives, see Schölkopf, Smola, and Müller (1998) for the kernel PCA which is a non-linear PCA.

(see Yitzhaki, 2003) based on the Gini covariance operator introduced by Schechtman and Yitzhaki (1987) and Schechtman and Yitzhaki (2003). The generalized Gini correlation coefficients are: (i) bounded, (ii) invariant to monotonic transformations, (iii) and symmetric whenever the variables are exchangeable. It is shown that the standard PCA is equivalent to the Gini PCA when the variables are Gaussians. Also, it is shown that the generalized Gini PCA may be realized either in the space of the variables or in the space of the observations. In each case, some statistics are proposed to perform some interpretations of the variables and of the observations (absolute and relative contributions). To be precise, a U -statistics test is introduced to test for the significance of the correlations between the axes of the new subspace and the variables in order to assess their significance. Monte Carlo simulations are performed in order to show the superiority of the Gini PCA compared with the usual PCA when outlying observations contaminate the data. Finally, with the aid of the well-known cars data, which contain outliers, it is shown that the generalized Gini PCA leads to different results compared with the usual PCA.

The outline of the paper is as follows. Section 2 sets the notations and presents the Gini-covariance operator. Section 3 is devoted to the generalized Gini PCA. Section 4 focuses on the interpretation of the Gini PCA. Sections 5 and 6 present some Monte Carlo simulations and applications, respectively.

2. Notations, definitions and Gini-covariance operators

In this section, the properties of the Gini Covariance operator is studied with the special case of Gaussian random variables. Then, the properties of the Gini correlation matrix are reviewed.

2.1. Notations

Let \mathbb{N}^* be the set of integers and $\mathbb{R} [\mathbb{R}_{++}]$ the set of [positive] real numbers. Let \mathcal{M} be the set of all $N \times K$ matrix $\mathbf{X} = [x_{ik}]$ that describes N observations on K dimensions such that $N \gg K$, with elements $x_{ik} \in \mathbb{R}$, and \mathbb{I}_n the $n \times n$ identity matrix. The $N \times 1$ vectors representing each variable are expressed as \mathbf{x}_k , for all $k \in \{1, \dots, K\}$ and we assume that $\mathbf{x}_k \neq c\mathbf{1}_N$, with c a real constant and $\mathbf{1}_N$ a N -dimensional column vector of ones. The $K \times 1$ vectors representing each observation i (the transposed i th line of \mathbf{X}) are expressed as $\mathbf{x}_{i\cdot}$, for all $i \in \{1, \dots, N\}$. It is assumed that \mathbf{x}_k is the realization of the random variable X_k , with cumulative distribution function F_k . The arithmetic mean of each column (line) of the matrix \mathbf{X} is given by $\bar{\mathbf{x}}_k$ ($\bar{\mathbf{x}}_{i\cdot}$). The cardinal of set A is denoted $\#A$. The ℓ_1 norm, for any given real vector \mathbf{x} , is $\|\mathbf{x}\|_1 = \sum_{k=1}^K |x_k|$, whereas the ℓ_2 norm is $\|\mathbf{x}\|_2 = (\sum_{k=1}^K x_k^2)^{1/2}$. The trace and the determinant of \mathbf{X} are denoted by $\text{Tr}[\mathbf{X}]$ and $|\mathbf{X}|$, respectively.

Assumption 2.1. The random variables X_k are such that $\mathbb{E}[|X_k|] < \infty$ for all $k \in \{1, \dots, K\}$, but no assumption is made on the second moments (that may not exist).

This assumption imposes less structure compared with the classical PCA in which the existence of the second moments are necessary, as can be seen in Appendix A for a review of ℓ_2 PCA.

2.2. The Gini-covariance operator

In this section, $\mathbf{X} = (X_1, \dots, X_K)$ denotes a random vector. The covariance matrix between \mathbf{X} and \mathbf{Y} , two random vectors, is defined as the inner product between centered versions of the vectors,

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \text{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^\top]. \tag{1}$$

Hence, it is the matrix where elements are regular covariances between components of the vectors, $\text{Cov}(\mathbf{X}, \mathbf{Y}) = [\text{Cov}(X_i, Y_j)]$. It is

the upper-right block of the covariance matrix of (\mathbf{X}, \mathbf{Y}) . Note that $\text{Cov}(\mathbf{X}, \mathbf{X})$ is the standard variance-covariance matrix of vector \mathbf{X} .

Definition 2.1. Let $\mathbf{X} = (X_1, \dots, X_K)$ be collections of K identically distributed random variables. Let $h: \mathbb{R} \rightarrow \mathbb{R}$ denote a non-decreasing function. Let $h(\mathbf{X})$ denote the random vector $(h(X_1), \dots, h(X_K))$, and assume that each component has a finite variance. Then, operator $\Gamma C_h(\mathbf{X}) = \text{Cov}(\mathbf{X}, h(\mathbf{X}))$ is called h -Gini covariance matrix.

Let us now introduce the Gini-covariance. Gini (1912) introduced the Gini mean difference operator Δ , defined as:

$$\Delta(X) = \mathbb{E}(|X_1 - X_2|) \text{ where } X_1, X_2 \sim X, \text{ and } X_1 \perp\!\!\!\perp X_2, \tag{2}$$

for some random variable X (or more specifically for some distribution F with $X \sim F$, because this operator is law invariant). For $F(X)$ being uniformly distributed on $[0,1]$, we have $\text{Cov}(F(X), F(X)) = 1/12$, then:

$$\Delta(X) = 4\text{Cov}[X, F(X)] = \frac{1}{3} \frac{\text{Cov}[X, F(X)]}{\text{Cov}[F(X), F(X)]} \tag{3}$$

where the ratio on the right-hand side of (3) is interpreted as the slope of the regression curve of the observed variable X and its ‘ranks’ (up to a scaling coefficient). Thus, the Gini-covariance is obtained when the function h is equal to the cumulative distribution function of the second term, see Schechtman and Yitzhaki (1987).

Definition 2.2. Let $\mathbf{X} = (X_1, \dots, X_K)$ be a collection of K identically distributed random variables, with cumulative distribution function F . Then, the Gini covariance is $\Gamma C_F(\mathbf{X}) = \text{Cov}(\mathbf{X}, F(\mathbf{X}))$.

On this basis, it is possible to show that the Gini covariance matrix is a positive semi-definite matrix.

Theorem 2.1. Let $Z \sim \mathcal{N}(0, 1)$. If \mathbf{X} represents identically distributed Gaussian random variables, with distribution $\mathcal{N}(\mu, \sigma^2)$, then the two following assertions hold:

- (i) $\Gamma C_F(\mathbf{X}) = \sigma^{-1} \text{Cov}(Z, \Phi(Z)) \text{Var}(\mathbf{X})$.
- (ii) $\Gamma C_F(\mathbf{X})$ is a positive semi-definite matrix.

Proof. See Appendix C. \square

Note that $\Gamma C_F(\mathbf{X}) = \text{Cov}(\mathbf{X}, -\bar{F}(\mathbf{X})) = \Gamma C_{-\bar{F}}(\mathbf{X})$, where $\bar{F} = 1 - F$ denotes the survival distribution function.

Definition 2.3. Let $\mathbf{X} = (X_1, \dots, X_K)$ be a collection of K identically distributed random variables, with survival distribution function \bar{F} . Then, the generalized Gini covariance is $G\Gamma C_\nu(\mathbf{X}) = \Gamma C_{-\bar{F}^{\nu-1}}(\mathbf{X}) = \text{Cov}(\mathbf{X}, -\bar{F}^{\nu-1}(\mathbf{X}))$, for $\nu > 1$.

This operator is related to the one introduced in Schechtman and Yitzhaki (2003), called generalized Gini mean difference GMD_ν operator. More precisely, an estimator of the generalized Gini mean difference is given by:

$$GMD_\nu(\mathbf{x}_\ell, \mathbf{x}_k) := -\frac{2}{N-1} \nu \text{Cov}(\mathbf{x}_\ell, \mathbf{r}_{\mathbf{x}_k}^{\nu-1}), \nu > 1,$$

where $\mathbf{r}_{\mathbf{x}_k} = (R(x_{1k}), \dots, R(x_{nk}))$ is the decumulative rank vector of \mathbf{x}_k , that is, the vector that assigns the smallest value (1) to the greatest observation x_{ik} , and so on.⁵ The rank of observation i with respect to variable k is:

$$R(x_{ik}) := \begin{cases} N + 1 - \#\{x \leq x_{ik}\} & \text{if no ties} \\ N + 1 - \frac{1}{p} \sum_{i=1}^p \#\{x \leq x_{ik}\} & \text{if } p \text{ ties } x_{ik}. \end{cases}$$

Hence $GMD_\nu(\mathbf{x}_\ell, \mathbf{x}_k)$ is the empirical version of $-2\nu \Gamma C_\nu(X_\ell, X_k) := -2\nu \text{Cov}(X_\ell, \bar{F}_k(X_k)^{\nu-1})$.

⁵ It is an unbiased and consistent estimator being a U -statistics, see Yitzhaki and Schechtman (2013) Chapter 9.

The index GMD_ν is a generalized version of the GMD_2 proposed earlier by Schechtman and Yitzhaki (1987), and can also be written as:

$$GMD_2(X_k, X_k) = 4\text{Cov}(X_k, F_k(X_k)) = \Delta(X_k).$$

When $k = \ell$, GMD_ν represents the variability of the variable \mathbf{x}_k itself. Focus is put on the lower tail of the distribution \mathbf{x}_k whenever $\nu \rightarrow \infty$, the approach is said to be max-min in the sense that GMD_ν inflates the minimum value of the distribution. On the contrary, whenever $\nu \rightarrow 0$, the approach is said to be max-max, in this case focus is put on the upper tail of the distribution \mathbf{x}_k . As mentioned in Yitzhaki and Schechtman (2013), the case $\nu < 1$ does not entail simple interpretations, thereby the parameter ν is used to be set as $\nu > 1$ in empirical applications.⁶

Note that even if X_k and X_ℓ have the same distribution, we might have $GMD_\nu(X_k, X_\ell) \neq GMD_\nu(X_\ell, X_k)$. In that case $\mathbb{E}[X_k h(X_\ell)] \neq \mathbb{E}[X_\ell h(X_k)]$ if $h(2) \neq 2h(1)$ (this property is nevertheless valid if h is linear). We would have $GMD_\nu(X_k, X_\ell) = GMD_\nu(X_\ell, X_k)$ when X_k and X_ℓ are exchangeable (see Appendix B for some properties on exchangeability and the positive semidefiniteness of the Gini-covariance matrix). But since generally GMD_ν is not symmetric, we have for \mathbf{x}_k being not a monotonic transformation of \mathbf{x}_ℓ and $\nu > 1$, $GMD_\nu(\mathbf{x}_k, \mathbf{x}_\ell) \neq GMD_\nu(\mathbf{x}_\ell, \mathbf{x}_k)$. In other words, the exchangeability property is not necessary for Gini PCA.

2.3. Generalized Gini correlation

In this section, \mathbf{X} is a matrix in \mathcal{M} . The Gini correlation coefficient (G -correlation from now on), is a normalized GMD_ν index such that for all $\nu > 1$, see Schechtman and Yitzhaki (2003),

$$GC_\nu(\mathbf{x}_\ell, \mathbf{x}_k) := \frac{GMD_\nu(\mathbf{x}_\ell, \mathbf{x}_k)}{GMD_\nu(\mathbf{x}_\ell, \mathbf{x}_\ell)}; GC_\nu(\mathbf{x}_k, \mathbf{x}_\ell) := \frac{GMD_\nu(\mathbf{x}_k, \mathbf{x}_\ell)}{GMD_\nu(\mathbf{x}_k, \mathbf{x}_k)},$$

with $GC_\nu(\mathbf{x}_k, \mathbf{x}_k) = 1$ and $GMD_\nu(\mathbf{x}_k, \mathbf{x}_k) \neq 0$, for all $k, \ell = 1, \dots, K$. Following Schechtman and Yitzhaki (2003), the G -correlation is well-suited for the measurement of correlations between non-normal distributions or in the presence of outlying observations in the sample.

Property 2.1. – Schechtman and Yitzhaki (2013):

- (i) $GC_\nu(\mathbf{x}_\ell, \mathbf{x}_k) \leq 1$.
- (ii) If the variables \mathbf{x}_ℓ and \mathbf{x}_k are independent, for all $k \neq \ell$, then $GC_\nu(\mathbf{x}_\ell, \mathbf{x}_k) = GC_\nu(\mathbf{x}_k, \mathbf{x}_\ell) = 0$.
- (iii) For any given monotonic increasing transformation φ , $GC_\nu(\mathbf{x}_\ell, \varphi(\mathbf{x}_k)) = GC_\nu(\mathbf{x}_\ell, \mathbf{x}_k)$.
- (iv) If $(\mathbf{x}_\ell, \mathbf{x}_k)$ have a bivariate normal distribution with Pearson correlation ρ , then $GC_\nu(\mathbf{x}_\ell, \mathbf{x}_k) = GC_\nu(\mathbf{x}_k, \mathbf{x}_\ell) = \rho$.
- (v) If \mathbf{x}_k and \mathbf{x}_ℓ are exchangeable up to a linear transformation, then $GC_\nu(\mathbf{x}_\ell, \mathbf{x}_k) = GC_\nu(\mathbf{x}_k, \mathbf{x}_\ell)$.

Whenever $\nu \rightarrow 1$, the variability of the variables is attenuated so that GMD_ν tends to zero (even if the variables exhibit a strong variance). The choice of ν is interesting to perform generalized Gini PCA with various values of ν in order to robustify the results of the PCA, since the standard PCA (based on the variance) is potentially of bad quality if outlying observations drastically affect the sample.

A G -correlation matrix is proposed to analyze the data into a new vector space. Following Property 2.1 (iv), it is possible to rescale the variables \mathbf{x}_ℓ thanks to a linear transformation, then the matrix of standardized observation is,

$$\mathbf{Z} \equiv [z_{i\ell}] := \left[\frac{X_{i\ell} - \bar{\mathbf{x}}_\ell}{GMD_\nu(\mathbf{x}_\ell, \mathbf{x}_\ell)} \right]. \tag{4}$$

⁶ In risk analysis $\nu \in (0, 1)$ denotes risk lover decision makers (max-max approach), whereas $\nu > 1$ stands for risk averse decision makers, and $\nu \rightarrow \infty$ extreme risk aversion (max-min approach).

The variable $z_{i\ell}$ is a real number without dimension. The variables \mathbf{x}_k are rescaled such that their Gini variability is equal to unity. Now, we define the $N \times K$ matrix of decumulative centered rank vectors of \mathbf{Z} , which are the same compared with those of \mathbf{X} :

$$\mathbf{R}_z^c \equiv [R^c(z_{i\ell})] := [R(z_{i\ell})^{\nu-1} - \bar{\mathbf{r}}_{z_{i\ell}}^{\nu-1}] = [R(x_{i\ell})^{\nu-1} - \bar{\mathbf{r}}_{x_{i\ell}}^{\nu-1}].$$

Note that the last equality holds since the standardization (4) is a strictly increasing affine transformation.⁷ The $K \times K$ matrix containing all G-correlation indices between all couples of variables \mathbf{z}_k and \mathbf{z}_ℓ , for all $k, \ell = 1, \dots, K$ is expressed as:

$$GC_v(\mathbf{Z}) := -\frac{2\nu}{N(N-1)} \mathbf{Z}^T \mathbf{R}_z^c.$$

If $GMD_v(\mathbf{Z}) \equiv [GMD_v(\mathbf{z}_k, \mathbf{z}_\ell)]$, then we get the following.

Proposition 2.1. For each standardized matrix \mathbf{Z} defined in (4), the following relations hold:

$$GMD_v(\mathbf{Z}) = GC_v(\mathbf{X}) = GC_v(\mathbf{Z}). \tag{5}$$

$$GMD_v(\mathbf{z}_k, \mathbf{z}_k) = 1, \forall k = 1, \dots, K. \tag{6}$$

Proof. See Appendix C. \square

Finally, under a normality assumption, the generalized Gini covariance matrix $GC_v(\mathbf{X}) \equiv [GMD_v(X_k, X_\ell)]$ is shown to be a positive semi-definite matrix.

Theorem 2.2. Let $Z \sim \mathcal{N}(0, 1)$. If \mathbf{X} represents identically distributed Gaussian random variables, with distribution $\mathcal{N}(\mu, \sigma^2)$, then the two following assertions hold:

- (i) $GC_v(\mathbf{X}) = \sigma^{-1} \text{Cov}(Z, \Phi(Z)) \text{Var}(\mathbf{X})$.
- (ii) $GC_v(\mathbf{X})$ is a positive semi-definite matrix.

Proof. The first part (i) follows from Yitzhaki and Schechtman (2013), Chapter 6. The second part follows directly from (i). \square

Theorem 2.2 shows that under the normality assumption, the variance is a special case of the Gini methodology. As a consequence, for multivariate normal distributions, it is shown in Section 4 that Gini PCA and classical PCA (based on the ℓ_2 norm and the covariance matrix) are equivalent.

3. Generalized Gini PCA

In this section, the multidimensional Gini variability of the observations $i = 1, \dots, N$, embodied by the matrix $GC_v(\mathbf{Z})$, is maximized in the \mathbb{R}^K -Euclidean space, i.e., in the set of variables $\{\mathbf{z}_1, \dots, \mathbf{z}_K\}$. This allows the observations to be projected onto the new vector space spanned by the eigenvectors of $GC_v(\mathbf{Z})$. Then, the projection of the variables is investigated in the \mathbb{R}^N -Euclidean space induced by $GC_v(\mathbf{Z})$. Both observations and variables are analyzed through the prism of absolute and relative contributions to propose relevant interpretations of the data in each subspace.

3.1. The \mathbb{R}^K -Euclidean space

It is possible to investigate the projection of the data \mathbf{Z} onto the new vector space induced by $GMD_v(\mathbf{Z})$ or alternatively by $GC_v(\mathbf{Z})$ since $GMD_v(\mathbf{Z}) = GC_v(\mathbf{Z})$. Let \mathbf{f}_k be the k th principal component, i.e. the k th axis of the new subspace, such that the $N \times K$ matrix \mathbf{F} is defined by $\mathbf{F} \equiv [\mathbf{f}_1, \dots, \mathbf{f}_K]$ with $\mathbf{R}_f^c \equiv [\mathbf{r}_{c, \mathbf{f}_1}^{\nu-1}, \dots, \mathbf{r}_{c, \mathbf{f}_K}^{\nu-1}]$ its corresponding decumulative centered rank matrix (where each decumulative rank vector is raised to an exponent of $\nu - 1$). The $K \times K$ matrix $\mathbf{B} \equiv [\mathbf{b}_1, \dots, \mathbf{b}_K]$ is the projector of the observations, with the normalization condition $\mathbf{b}_k^T \mathbf{b}_k = 1$, such that $\mathbf{F} = \mathbf{Z}\mathbf{B}$. We denote by

⁷ By definition $GMD_v(\mathbf{x}_\ell, \mathbf{x}_\ell) \geq 0$ for all $\ell = 1, \dots, K$. As we impose that $\mathbf{x}_\ell \neq \mathbf{c}1_N$, the condition becomes $GMD_v(\mathbf{x}_\ell, \mathbf{x}_\ell) > 0$.

$\lambda_{\cdot k}$ (or $2\mu_{\cdot k}$) the eigenvalues of the matrix $[GC_v(\mathbf{Z}) + GC_v(\mathbf{Z})^T]$. Let the basis $\mathcal{B} := \{\mathbf{b}_1, \dots, \mathbf{b}_h\}$ with $h \leq K$ issued from the maximization of the overall Gini variability:

$$\begin{aligned} \max \mathbf{b}_k^T GC_v(\mathbf{Z}) \mathbf{b}_k &\implies [GC_v(\mathbf{Z}) + GC_v(\mathbf{Z})^T] \mathbf{b}_k = 2\mu_{\cdot k} \mathbf{b}_k, \\ \forall k = 1, \dots, K. \end{aligned}$$

Indeed, from the Lagrangian,

$$L = \mathbf{b}_k^T GC_v(\mathbf{Z}) \mathbf{b}_k - \mu_{\cdot k} [1 - \mathbf{b}_k^T \mathbf{b}_k],$$

because of the non-symmetry of $GC_v(\mathbf{Z})$, the eigenvalue equation is,

$$[GC_v(\mathbf{Z}) + GC_v(\mathbf{Z})^T] \mathbf{b}_k = 2\mu_{\cdot k} \mathbf{b}_k,$$

that is,

$$[GC_v(\mathbf{Z}) + GC_v(\mathbf{Z})^T] \mathbf{b}_k = \lambda_{\cdot k} \mathbf{b}_k. \tag{7}$$

The new subspace $\{\mathbf{f}_1, \dots, \mathbf{f}_h\}$ such that $h \leq K$ is derived from the maximization of the Gini variability of the observations over each axis \mathbf{f}_k . To be precise, this Gini variability is the Generalized Gini Mean Difference of \mathbf{f}_k given by:

$$GGMD_v(\mathbf{f}_k, \mathbf{f}_k) := -\frac{2\nu}{N(N-1)} \mathbf{f}_k^T \mathbf{R}_z^c \mathbf{b}_k. \tag{8}$$

The Generalized Gini Mean Difference measures the variability between the observations projected onto the new axis \mathbf{f}_k and the projected rank of these observations. In the literature on inequality indices, this kind of index is rather known as a generalized Gini index, because of the product between a variable \mathbf{f}_k and a function Ψ of its ranks, $\Psi(\mathbf{r}_k) := \mathbf{R}_z^c \mathbf{b}_k$. Yaari (1987) and subsequently Yaari (1988) proposes generalized Gini indices with a rank distortion function Ψ that describes the behavior of the decision maker (being either max-min or max-max).⁸

As shown in the next result, when the variability of the observations projected onto component \mathbf{f}_k is null, then the eigenvalue associated with axis \mathbf{f}_k is null. In the same time, there is neither co-variability in the Gini sense between \mathbf{f}_k and another axis \mathbf{f}_ℓ , that is $GGMD_v(\mathbf{f}_k, \mathbf{f}_\ell) = 0$.

Proposition 3.1. Let the eigenvalues of $GC_v(\mathbf{Z}) + GC_v(\mathbf{Z})^T$ be such that $\lambda_1 = 2\mu_{\cdot 1} \geq \dots \geq \lambda_K = 2\mu_{\cdot K}$. Then,

- (i) $GGMD_v(\mathbf{f}_k, \mathbf{f}_k) = GGMD_v(\mathbf{f}_k, \mathbf{f}_\ell) = GGMD_v(\mathbf{f}_\ell, \mathbf{f}_k) = 0$, for all $\ell = 1, \dots, K$, if and only if, $\lambda_{\cdot k} = 0$ and $k = 1, \dots, K$.
- (ii) $\max_{k=1, \dots, K} GGMD_v(\mathbf{f}_k, \mathbf{f}_k) = \mu_{\cdot 1}$.
- (iii) $\min_{k=1, \dots, K} GGMD_v(\mathbf{f}_k, \mathbf{f}_k) = \mu_{\cdot K}$.

Proof. See the Appendix. \square

In the Gaussian case, because the Gini correlation matrix is positive semi-definite, the eigenvalues are non-negative, then the GGMD is null whenever it reaches its minimum. Furthermore, the Gini PCA and the standard PCA are equivalent.

Proposition 3.2. Let $Z \sim \mathcal{N}(0, 1)$ and let \mathbf{X} represent identically distributed Gaussian random variables, with distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\rho})$ such that $\text{Var}(X_k) = 1$ for all $k = 1, \dots, K$ and let $\gamma_1, \dots, \gamma_K$ be the eigenvalues of $\text{Var}(\mathbf{X})$. Then the following assertions holds:

- (i) $\text{Tr}[GC_v(\mathbf{X})] = \text{Cov}(Z, \Phi(Z)) \text{Tr}[\text{Var}(\mathbf{X})]$.
- (ii) $\mu_{\cdot k} = GGMD_v(\mathbf{f}_k, \mathbf{f}_k) = \text{Cov}(Z, \Phi(Z)) \gamma_k \geq 0$ for all $k = 1, \dots, K$.
- (iii) $|GC_v(\mathbf{X})| = \text{Cov}^K(Z, \Phi(Z)) |\text{Var}(\mathbf{X})|$.
- (iv) For $\nu > 1$:

⁸ Strictly speaking Yaari (1987) and Yaari (1988) suggest probability distortion functions $\Psi : [0, 1] \rightarrow [0, 1]$, which does not necessarily coincide to our case.

$$\frac{\mu_k}{Tr[GC_v(\mathbf{X})]} = \frac{\gamma_k}{Tr[Var(\mathbf{X})]}, \forall k = 1, \dots, K.$$

Proof. See Appendix C. \square

Point (ii) shows that the eigenvalues are always non-negative in the case of Gaussian variables. On the contrary, it may appear that $GGMD(\mathbf{f}_k, \mathbf{f}_k) = \mu_k < 0$. Since the index $GGMD$ measures the association between the (projected) variable \mathbf{f}_k and its (projected) ranks, this would indicate that the variable is not correlated with its rank vector. Consequently, the information provided by component \mathbf{f}_k would be irrelevant since by definition as far as the variability over \mathbf{f}_k increases, $GGMD(\mathbf{f}_k, \mathbf{f}_k)$ increases. We can interpret the negativity of the eigenvalue as an irrelevant principal component in which the interpretation of the PCA cannot be done. This occurs when the quantity of information $\frac{\mu_k}{Tr[GC_v(\mathbf{X})]}$ captured by each axis \mathbf{f}_k is very low. In some particular cases, such as exchangeability of the variables, the matrix of Gini covariance is positive semi-definite, and so the eigenvalues are non-negative (see Appendix B).

Point (iv) shows that the eigenvalues of the standard PCA are proportional to those issued from the generalized Gini PCA. Because each eigenvalue (in proportion of the trace) represents the variability (or the quantity of information) inherent to each principal component, then both PCA techniques are equivalent when \mathbf{X} is Gaussian. In cases of non-Gaussian distributions, and in particular when outlying observations affect the sample, the Gini PCA can be used with many parameters ν (from max-min to max-max decision makers) in order to select the optimal parameter that minimizes the influence of the outliers (see Section 5.4).

As a remark, we can show that the maximization of the multidimensional variability in the Gini sense $\mathbf{b}_k^T GC_v(\mathbf{Z}) \mathbf{b}_k$ does not necessarily coincide with the maximization of the traditional Gini variability embodied by $GMD_\nu(\mathbf{f}_k, \mathbf{f}_k)$.

Remark 3.1. Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_h\}$ with $h \leq K$ be the basis issued from the maximization of $\mathbf{b}_k^T GC_v(\mathbf{Z}) \mathbf{b}_k$ for all $k = 1, \dots, K$, then: $\max GMD_\nu(\mathbf{f}_k, \mathbf{f}_k) = \mu_k$ for all $k = 1, \dots, K$, if and only if $\mathbf{r}_{c, \mathbf{f}_k}^{\nu-1} = \mathbf{R}_c^T \mathbf{b}_k$.

Proof. See Appendix C. \square

The difference between $GGMD_\nu(\mathbf{f}_k, \mathbf{f}_k)$ and $GMD_\nu(\mathbf{f}_k, \mathbf{f}_k)$ comes from the fact that $GGMD_\nu(\mathbf{f}_k, \mathbf{f}_k)$ deals with projected ranks of the observations ($\mathbf{R}_c^T \mathbf{b}_k$). It measures the co-variability (or degree of association) between the projected observations and their projected ranks.⁹

Finally, the literature has already dealt with multidimensional Gini index, see for instance Banerjee (2010). The author proposes to extract the first eigenvector \mathbf{e}_1 of $\mathbf{X}^T \mathbf{X}$ and to project the data \mathbf{X} such that $\mathbf{s} := \mathbf{X} \mathbf{e}_1$ so that the multidimensional Gini index is $G(\mathbf{s}) = \mathbf{s}^T \tilde{\Psi}(\mathbf{r}_s)$, with \mathbf{r}_s the rank vector of \mathbf{s} and with $\tilde{\Psi}$ a function that distorts the ranks. Banerjee (2010)'s index is derived from the matrix $\mathbf{X}^T \mathbf{X}$. To be precise, the maximization of the variance-covariance matrix $\mathbf{X}^T \mathbf{X}$ (based on the ℓ_2 metric) yields the projection of the data on the first component \mathbf{f}_1 , which is then employed in the multidimensional Gini index (based on the ℓ_1 metric). This approach is legitimated by the fact that $G(\mathbf{s})$ has some desirable properties linked with the Gini index. However, this Gini index deals with an information issued from the variance-covariance matrix, and as such the projection of the data onto the new subspace does not prevent the presence of outlying observations that could capture an important part of the information (variance) on the first principal component. This case occurs in the classical PCA

⁹ On the contrary, $GMD_\nu(\mathbf{f}_k, \mathbf{f}_k)$ deals with the ranks of the projected observations.

(see Section 5 for Monte Carlo simulations). Let us before investigate the \mathbb{R}^N -Euclidean space.

3.2. The \mathbb{R}^N -Euclidean space

In classical PCA, the duality between \mathbb{R}^N and \mathbb{R}^K enables the eigenvectors and eigenvalues of \mathbb{R}^N to be deduced from those of \mathbb{R}^K and conversely. This duality is not so obvious in the Gini PCA case. Indeed, in \mathbb{R}^N the Gini variability between the observations would be measured by $GC_v(\tilde{\mathbf{Z}}) := \frac{-2\nu}{N(N-1)} (\mathbf{R}_c^T)^T \mathbf{Z}$, and subsequently the idea would be to derive the eigenvalue equation related to \mathbb{R}^N ,

$$[GC_v(\tilde{\mathbf{Z}}) + GC_v(\tilde{\mathbf{Z}})^T] \tilde{\mathbf{b}}_k = \tilde{\lambda}_k \tilde{\mathbf{b}}_k.$$

The other option is to define a basis of \mathbb{R}^N from a basis already available in \mathbb{R}^K . In particular, the set of principal components $\{\tilde{\mathbf{f}}_1, \dots, \tilde{\mathbf{f}}_k\}$ provides by construction a set of normalized and orthogonal vectors. Let us rescale the vectors $\tilde{\mathbf{f}}_k$ such that:

$$\tilde{\mathbf{f}}_k = \frac{\mathbf{f}_k}{GMD_\nu(\mathbf{f}_k, \mathbf{f}_k)}.$$

Then, $\{\tilde{\mathbf{f}}_1, \dots, \tilde{\mathbf{f}}_k\}$ constitutes an orthonormal basis of \mathbb{R}^K in the Gini sense since $GMD_\nu(\tilde{\mathbf{f}}_k, \tilde{\mathbf{f}}_k) = 1$. This basis may be used as a projector of the variables \mathbf{z}_k onto \mathbb{R}^N . Let $\tilde{\mathbf{F}}$ be the $N \times K$ matrix with $\tilde{\mathbf{f}}_k$ in columns. The projection of the variables \mathbf{z}_k in \mathbb{R}^N is given by the following Gini correlation matrix:

$$\mathbf{V} := \frac{-2\nu}{N(N-1)} \tilde{\mathbf{F}}^T \mathbf{R}_c^T,$$

whereas it is given by $\frac{1}{N} \tilde{\mathbf{F}}^T \mathbf{Z}$ in the standard PCA, that is, the matrix of Pearson correlation coefficients between all $\tilde{\mathbf{f}}_k$ and \mathbf{z}_k . The same interpretation is available in the Gini case. The matrix \mathbf{V} is normalized in such a way that $\mathbf{V} \equiv [\nu_{k\ell}]$ are the G -correlations indices between $\tilde{\mathbf{f}}_k$ and \mathbf{z}_k . This yields the ability to make easier the interpretation of the variables projected onto the new subspace.

4. Interpretations of the Gini PCA

The analysis of the projections of the observations and of the variables are necessary to provide accurate interpretations. Some criteria have to be designed in order to bring out, in the new subspace, the most significant observations and variables.

4.1. Observations

The absolute contribution of an observation i to the variability of a principal component \mathbf{f}_k is:

$$ACT_{ik} = \frac{f_{ik} \Psi(R(f_{ik}))}{GGMD_\nu(\mathbf{f}_k, \mathbf{f}_k)}.$$

The absolute contribution of each observation i to the generalized Gini Mean Difference of \mathbf{f}_k (ACT_{ik}) is interpreted as a percentage of variability of $GGMD_\nu(\mathbf{f}_k, \mathbf{f}_k)$, such that $\sum_{i=1}^N ACT_{ik} = 1$. This provides the most important observations i related to component \mathbf{f}_k with respect to the information $GGMD_\nu(\mathbf{f}_k, \mathbf{f}_k)$. On the other hand, instead of employing the Euclidean distance between one observation i and the component \mathbf{f}_k , the Manhattan distance is used. The relative contribution of an observation i to component \mathbf{f}_k is then:

$$RCT_{ik} = \frac{|f_{ik}|}{\|\mathbf{f}_i\|_1}.$$

Remark that the gravity center of $\{\mathbf{f}_1, \dots, \mathbf{f}_K\}$ is $\mathbf{g} := (\bar{\mathbf{f}}_1, \dots, \bar{\mathbf{f}}_K) = \mathbf{0}$. The Manhattan distance between observation i and \mathbf{g} is then $\sum_{k=1}^K |f_{ik} - 0|$, and so

$$RCT_{ik} = \frac{|f_{ik}|}{\|\mathbf{f}_i - \mathbf{g}\|_1}.$$

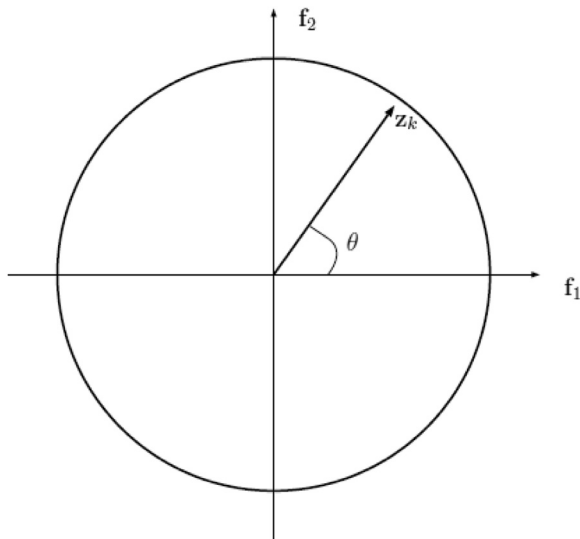


Fig. 1. Circle of correlation.

The relative contribution RCT_{ik} may be interpreted rather as the contribution of dimension k to the overall distance between observation i and \mathbf{g} .

4.2. Variables

The usual PCA enables the variables to be analyzed in the circle of correlation, which outlines the correlations between the variables \mathbf{z}_k and the principal components \mathbf{f}_ℓ measured by $\cos^2 \theta$ (see Fig. 1).

The most significant variables must be retained in order to make the interpretation of the data in the new subspace. It would be possible to test for the significance of the terms of the Gini correlation matrix $\mathbf{V} \equiv [v_{k\ell}]$:

$$U_{\ell k} := \frac{\text{Cov}(\mathbf{f}_\ell, \mathbf{R}_{\mathbf{z}_k}^c)}{\text{Cov}(\mathbf{f}_\ell, \mathbf{R}_{\mathbf{f}_\ell}^c)} = v_{k\ell}.$$

However, this option is not investigated because the U -statistics $U_{\ell k}$ is not rotational invariant.¹⁰ Consequently, the correlation between the principal components and the variables is captured in the following way:

$$\tilde{U}_{\ell k} := \frac{\text{Cov}(\mathbf{f}_\ell, \mathbf{R}_{\mathbf{z}_k}^c)}{\text{Cov}(\mathbf{z}_k, \mathbf{R}_{\mathbf{z}_k}^c)},$$

with $\mathbf{R}_{\mathbf{z}_k}^c$ the (decumulative) centered rank vector of \mathbf{z}_k raised to an exponent of $v - 1$. Since $\tilde{U}_{\ell k}$ is a ratio of two Gini covariances, then it is a U -statistics, i.e. an unbiased and consistent estimator of $\tilde{U}_{\ell k}^0 := \mathbb{E}(\tilde{U}_{\ell k})$. From Theorem 10.4 in Yitzhaki and Schechtman (2013), Chapter 10, we asymptotically get that $\sqrt{N}(\tilde{U}_{\ell k} - \tilde{U}_{\ell k}^0) \overset{a}{\approx} \mathcal{N}$. Then, the proposed test is:

$$\begin{cases} H_0 : \tilde{U}_{\ell k}^0 = 0 \\ H_1 : \tilde{U}_{\ell k}^0 \neq 0. \end{cases}$$

Let $\hat{\sigma}_{\ell k}^2$ be the Jackknife variance of $\tilde{U}_{\ell k}$, then it is possible to test for the null under the assumption $N \rightarrow \infty$ as follows:¹¹

$$\frac{\tilde{U}_{\ell k} - \hat{\sigma}_{\ell k}^2}{\hat{\sigma}_{\ell k}} \overset{a}{\approx} \mathcal{N}(0, 1). \tag{9}$$

¹⁰ When the sign of the eigenvectors is reversed, the rank vector of the projected observations $\mathbf{R}_{\mathbf{f}_\ell}^c$ does not remain invariant.

¹¹ As indicated by Yitzhaki (1991), the efficient Jackknife method may be used to find the variance of any U -statistics.

It is worth mentioning that the usual circle of correlations does not provide the significance of the variables, and moreover, the Pearson correlation $\cos^2 \theta$ may be sensitive to outliers since it depends on the ℓ_2 metric.

5. Monte Carlo simulations

In this Section, it is shown with the aid of Monte Carlo simulations that the usual PCA yields irrelevant results when outlying observations contaminate the data. To be precise, the absolute contributions computed in the standard PCA based on the variance may lead to select outlying observations on the first component in which there is the most important variability (a direct implication of the maximization of the variance). We propose three simulations: a first one with a strong correlation between the variables, a second one with lower correlations, and a third one with Clayton copula. Then, we propose an algorithm to select the optimal parameter ν .

5.1. Case 1: strong correlation

The Monte Carlo simulations are based on the following process of data contamination.

Algorithm 1: Monte Carlo Simulation

```

Result: Robust Gini PCA with data contamination
1  $\theta = 1$  [  $\theta$  is the value of the outlier ] ;
2 repeat
3   Generate a 4-variate normal distribution  $\mathbf{X} \sim \mathcal{N}$ ,  $N = 500$  ;
4   Introduce outliers in 1 row of  $\mathbf{X}$ :  $\mathbf{X}_{ji}^o := \theta \mathbf{X}_{ji}$  with
    $j = 1, \dots, 4$  [for a random row localization];
5   Compute the eigenvalues  $\lambda_k^{oi}$  of  $\text{Var}(\mathbf{X}^o)$  and of
    $GC_\nu(\mathbf{X}^o) + GC_\nu(\mathbf{X}^o)^\tau$ 
6 until  $\theta = 1000$  [increment of 1];
7 ;
8 For Variance and Gini PCA, the ACT and RCT are computed
   for the axes 1 and 2 on the contaminated matrix  $\mathbf{X}^o$ ;
9 return Root-Mean-Square Errors of eigenvalues, ACT and RCT
;
    
```

The Root-mean-square error of the eigenvalues are computed as follows:

$$RMSE_{\lambda_k} = \sqrt{\frac{\sum_{i=1}^{1,000} (\lambda_k^{oi} - \lambda_{\cdot,k})^2}{1,000}},$$

where λ_k^{oi} is the eigenvalue computed with outlying observations in the sample. The RMSE of ACT et RCT are computed in the same manner.

We first investigate the case where the variables are highly correlated in order to gauge the robustness of each technique (Gini for $\nu = 2, 4, 6$ and variance). The correlation matrix between the variables is given by:

$$\rho = \begin{pmatrix} 1 & 0.8 & 0.9 & 0.7 \\ 0.8 & 1 & 0.8 & 0.75 \\ 0.9 & 0.8 & 1 & 0.6 \\ 0.7 & 0.75 & 0.6 & 1 \end{pmatrix}$$

As can be seen in the matrix above, we can expect that all the information be gathered on the first axis because each pair of variables records an important linear correlation. The repartition of the information on each component, that is, each eigenvalue in percentage of the sum of the eigenvalues is the following.

The first axis captures around 82% of the variability of the overall sample (before contamination). Although each PCA method

Table 1
Eigenvalues % and their RMSE (level).

Eigenvalues	Gini $\nu = 2$	Gini $\nu = 4$	Gini $\nu = 6$	Variance
Axis 1 eigenvalues (%)	83.14	82.68	82.31	82.90
Axis 1 RMSE (level)	14.60	13.53	12.79	15.81
Axis 2 eigenvalues (%)	10.55	11.31	11.87	10.82
Axis 2 RMSE (level)	13.86	11.86	10.10	10.09
Axis 3 eigenvalues (%)	4.24	4.01	3.87	4.27
Axis 3 RMSE (level)	2.83	2.69	2.69	4.21

Table 2
Average of the RMSE of the ACT of the observations over each principal component (standard deviation in parenthesis).

RMSE of ACT (%)	Gini $\nu = 2$	Gini $\nu = 4$	Gini $\nu = 6$	Variance
Axis 1	2.92 (1.59)	3.02 (1.59)	3.12 (1.65)	4.01 (1.89)
Axis 2	0.56 (0.81)	0.83 (1.57)	1.32 (2.86)	0.24 (0.23)

Table 3
Average of the RMSE of the RCT of the observations over each principal component (standard deviation in parenthesis).

RMSE of RCT (%)	Gini $\nu = 2$	Gini $\nu = 4$	Gini $\nu = 6$	Variance
Axis 1	33.67 (9.40)	33.86 (9.37)	33.89 (9.41)	50.26 (12.35)
Axis 2	24.19 (7.65)	24.71 (8.33)	24.63 (8.14)	44.75 (11.62)

yields the same repartition of the information over the different components before the contamination of the data, it is possible to show that the classical PCA is less robust.

On the one hand, Table 1 depicts the RMSE of the eigenvalues, which are slightly lower for the Gini PCA, especially for the first component. On the other hand, Table 2 shows that the RMSE of the ACT (in percentage) measured during to the contamination process provide lower values for the Gini index compared with the variance. Over the first component, the mean of the RMSE is the lowest for $\nu = 2$. The same remarks holds true for the standard deviation over the first axis, the second one may be considered as non-significant.

Finally, Table 3 presents the RMSE of the RCT. Again, the RMSE of the RCT of the standard PCA are much more important compared with the Gini PCA. Since the RCT point out the distance between each observation and each component, their reliability enables to use the projection of the observations onto the two first components in order to make a graphical interpretation.

5.2. Case 2: intermediate correlation

Let us now take an example with less correlations between the variables in order to get a more equal repartition of the information on the first two principal components.

$$\rho = \begin{pmatrix} 1 & -0.5 & 0.25 & 0.5 \\ -0.5 & 1 & -0.9 & 0 \\ 0.25 & -0.9 & 1 & -0.25 \\ 0.5 & 0 & -0.25 & 1 \end{pmatrix}$$

The repartition of the information over the principal components (percentage of each eigenvalue) is given in Table 4. When the information is less concentrated on the first axis (50% for axis 1 and around 35% for axis 2), the RMSE of the eigenvalues after contamination are much more important for the standard PCA compared with the Gini approach (approximately 1.7 times more important for the two first axes).

Table 4
Eigenvalues % and their RMSE.

Eigenvalues	Gini $\nu = 2$	Gini $\nu = 4$	Gini $\nu = 6$	Variance
Axis 1 eigenvalues (%)	49.98	49.91	49.75	49.74
Axis 1 RMSE (level)	28.04	23.66	19.25	48.02
Axis 2 eigenvalues (%)	35.29	35.62	35.98	35.66
Axis 2 RMSE (level)	19.07	21.76	24.68	33.70
Axis 3 eigenvalues (%)	9.11	8.98	8.84	8.98
Axis 3 RMSE (level)	5.74	6.17	6.62	8.85

Table 5
Average of the RMSE of the ACT of the observations over each principal component (standard deviation in parenthesis).

RMSE of ACT (%)	Gini $\nu = 2$	Gini $\nu = 4$	Gini $\nu = 6$	Variance
Axis 1	3.40 (1.65)	3.46 (1.67)	3.55 (1.71)	4.00 (1.92)
Axis 2	0.27 (0.20)	0.32 (0.22)	0.31 (0.25)	0.23 (0.21)

Table 6
Average of the RMSE of the RCT of the observations over each principal component (standard deviation in parenthesis).

RMSE of RCT (%)	Gini $\nu = 2$	Gini $\nu = 4$	Gini $\nu = 6$	Variance
Axis 1	32.49 (10.89)	28.05 (10.55)	26.37 (10.48)	48.16 (12.85)
Axis 2	23.35 (8.76)	23.78 (9.41)	24.08 (9.60)	42.73 (9.45)

Let us now have a look on the RMSE of the absolute contributions of each observation ($N = 500$) for each PCA technique (Table 5). We obtain the same kind of results, with less variability for the second axis. In Table 5, it is apparent that the classical PCA based on the ℓ_2 norm exhibits more ACT variability. This means that the contamination of the data can lead to the interpretation of some observations as significant (important contribution to the variance of the axis) while they are not (and vice versa). On the other hand, the RMSE of the RCT after contamination of the data (Table 6) are less spread out for the Gini technique for $\nu = 4$ and $\nu = 6$ over the first axis, and for $\nu = 2$ over the second axis. This means that the distance from one observation to an axis is more reliable (although the interpretation of the data rather depends on the ACT).

As in the previous simulation, these results show that the PCA based on the variance is less stable about eigenvalues, ACT and RCT that outline the most important observations of the sample. This may lead to irrelevant interpretations when the contamination process is important. The same interpretations are obtained with copulas.

5.3. Simulations with copulas

In order to move away from the classical multivariate Gaussian distribution it is possible to use copulas with pairwise exchangeability (although exchangeability is not necessary for Gini PCA). A natural class of copulas is the class of Archimedean copulas that could also exhibit heavy tails, interpreted as outliers (see Charpentier & Segers, 2009; Nelsen, 2007 for states of the art). Given a generator $\psi : [0, \infty] \rightarrow [0, 1]$ which is some completely monotone function (that could be interpreted as the Laplace transform of a positive random variable), a bivariate Archimedean copula is $C_\psi : (u_1, u_2) \mapsto \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2))$. This can be extended to higher dimension using $(u_1, \dots, u_d) \mapsto \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d))$, which will be an exchangeable symmetric distribution in dimension d . Actually, such a copula can also be writ-

Table 7
Eigenvalues % and their RMSE (level).

Eigenvalues	Gini $\nu = 2$	Gini $\nu = 4$	Gini $\nu = 6$	Variance
Axis 1 eigenvalues (%)	67.56	72.36	74.77	67.12
RMSE (level)	16.88	15.96	13.77	31.15
Axis 2 eigenvalues (%)	21.30	20.11	19.30	21.23
RMSE (level)	11.06	12.36	13.06	19.81
Axis 3 eigenvalues (%)	9.70	6.91	5.53	9.89
RMSE (level)	6.46	5.22	4.51	9.80

Table 8
Average of the RMSE of the ACT of the observations over each principal component (standard deviation in parenthesis).

RMSE of ACT (%)	Gini $\nu = 2$	Gini $\nu = 4$	Gini $\nu = 6$	Variance
Axis 1	3.10 (1.65)	3.13 (1.72)	3.24 (1.77)	3.98 (1.98)
Axis 2	0.45 (0.51)	1.45 (2.69)	1.48 (3.69)	0.24 (0.24)

Table 9
Average of the RMSE of the RCT of the observations over each principal component (standard deviation in parenthesis).

RMSE of RCT (%)	Gini $\nu = 2$	Gini $\nu = 4$	Gini $\nu = 6$	Variance
Axis 1	33.15 (10.38)	32.27 (10.23)	31.58 (9.74)	49.31 (10.03)
Axis 2	25.22 (9.43)	24.82 (9.84)	25.00 (10.10)	44.20 (12.23)

ten (in dimension 3) as $(u_1, u_2, u_3) \mapsto C_\psi(u_1, C_\psi(u_2, u_3))$, with a simple extension in dimension d . It is possible to have more general dependence structure using nested Archimedean copulas, as in McNeil (2008). For instance, in dimension 3, one could consider $(u_1, u_2, u_3) \mapsto C_{\psi_1}(u_1, C_{\psi_2}(u_2, u_3))$ given two generators ψ_1 and ψ_2 , that should satisfy some ordering property, namely $\psi_1^{-1} \circ \psi_2$ must be completely monotone. This property implies that the dependence between components 2 and 3 should be stronger than the one between 1 and the pair (2,3). Such a property can easily be obtained when generators ψ remain in the same parametric family, such as Gumbel, Clayton or Frank.

In order to illustrate the robustness of our algorithm, we generate some multivariate random vectors $\mathbf{X} = (X_1, X_2, X_3, X_4)$ with Gaussian margins, $X_j = \Phi_{-1}(U_j)$ where vector $\mathbf{U} = (U_1, U_2, U_3, U_4)$ is generated from a 4-dimensional Nested-Clayton copula,

$$(u_1, u_2, u_3, u_4) \mapsto C_{0.2}(C_{0.5}(C_{0.8}(u_1, u_2), u_3), u_4)$$

where C_τ is the bivariate Clayton copula with Kendall's tau τ , using the algorithm described in Hofert (2011).

We generate the same contamination process explained in Algorithm 1: 1,000 Gini PCA (for $\nu = 2, 4, 6$) and 1,000 standard PCA with contamination are performed in order to compute the RMSE of the eigen values, and the RMSE of the ACT and RCT of each observation.

Table 7 shows that the first axis captures around 70% of the variability of the overall sample (before contamination). The RMSE of the eigenvalues are approximately 2 times lower for the Gini PCA over the first and second principal components compared with the standard PCA.

Table 8 shows that the RMSE of the ACT, as in the previous Monte Carlo simulations, are 1.3 times lower for the Gini PCA over axis 1 and 3 times lower over axis 2.

Finally, the same conclusion can be drawn about the robustness of the RCT for each Gini PCA (Table 9).

5.4. The choice of the ν parameter

As shown in the previous Monte Carlo simulations, the choice of the ν parameter may help to attenuate the presence of outlying observations in the sample. In our study, outlying observations are not measurement errors that should be removed from the sample. These are atypical observations for which the information on the different dimensions must be necessarily kept. Consequently, a simple algorithm for the choice of ν consists in performing two types of Gini PCA for each value of ν :

- a first Gini PCA with outliers;
- a second Gini PCA without outliers: an outlier test is performed (Grubbs test) in order to delete the outliers from the sample before running again the Gini PCA.

Then, the optimal ν parameter is selected by minimizing the influence of the outliers. The Manhattan distance of the share of each eigenvalue (in percentage of the sum of all eigenvalues) between the two Gini PCA over the significant principal components j^* are computed:

$$\nu^* = \arg \min_{\nu} \sum_{j=1}^{j^*} \left| \frac{\mu_{j^*}(\nu)}{\sum_{k=1}^K \mu_{k^*}(\nu)} - \frac{\mu'_{j^*}(\nu)}{\sum_{k=1}^K \mu'_{k^*}(\nu)} \right|,$$

where $\mu'_{j^*}(\nu)$ and $\mu_{j^*}(\nu)$ stands for eigenvalues with and without outliers of the Gini PCA related to the parameter ν . The optimal parameter ν^* is the one that minimizes the difference of information over the principal components between the actual Gini PCA and the Gini PCA if no outliers were in the sample. The idea is to keep those outliers (since they are not measurement errors) while making the interpretation of the Gini PCA without distorting the information on each component. This grid search is used in Section 6 devoted to the empirical application.

In contrast to the classical PCA, for which the projection of the data yields a unique Euclidean distance between each couple of observations with possible bias in eigenvalues and absolute contributions, the optimal parameter ν^* aims at limiting the influence of outliers. Of course, other rules could have been chosen such that the minimization of the number of outliers over each principal component (for each value of ν from $\nu = 1.1$ to $\nu = 6$).¹² An application of Algorithm 2 is performed in the next Section.

Algorithm 2: The choice of the optimal ν^* parameter.

Result: ν^*

- $\nu = 1.1$;
- repeat**
- Run Gini PCA with outliers: $\max \mathbf{b}_k^\top \mathbf{G} C_\nu(\mathbf{Z}) \mathbf{b}_k \forall k \implies \mu'_{j^*}(\nu)$;
- Grubbs test: all outliers are removed (over components 1 to j^*) ;
- Run Gini PCA without outliers: $\max \mathbf{b}_k^\top \mathbf{G} C_\nu(\mathbf{Z}) \mathbf{b}_k \forall k \implies \mu_{j^*}(\nu)$;
- until** $\nu = 6$ [$\nu = \nu + 0.1$];
- return** $\nu^* = \arg \min_{\nu} \sum_{j=1}^{j^*} \left| \frac{\mu_{j^*}(\nu)}{\sum_{k=1}^K \mu_{k^*}(\nu)} - \frac{\mu'_{j^*}(\nu)}{\sum_{k=1}^K \mu'_{k^*}(\nu)} \right|$;

¹² The maximal value $\nu = 6$ has been chosen in view of Monte Carlo simulations showing that the correlations between the projected variables and the principal components decrease significantly.

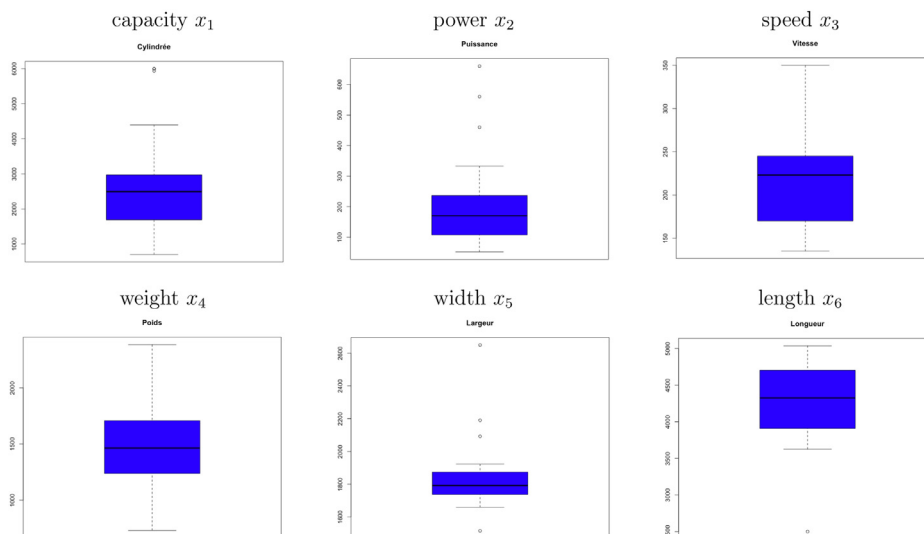


Fig. 2. Box plots.

Table 10
Pearson Correlation matrix.

	capacity x_1	power x_2	speed x_3	weight x_4	width x_5	length x_6
x_1	1.000	0.954	0.885	0.692	0.706	0.663
x_2	0.954	1.000	0.933	0.528	0.729	0.663
x_3	0.885	0.933	1.000	0.466	0.618	0.578
x_4	0.692	0.528	0.466	1.000	0.477	0.794
x_5	0.706	0.729	0.618	0.477	1.000	0.591
x_6	0.663	0.663	0.578	0.794	0.591	1.000

Table 11
Eigenvalues (%) for Gini and Variance and proportion of ℓ_1 dispersion explained by the loadings vectors for Visentin et al. (2016) and Brooks et al. (2013).

Information	Gini $\nu^* = 1.4$	Visentin et al.	Brooks et al.	Variance
Axis 1	78.45	52.12	52.12	73.52
Axis 2	12.94	7.46	7.48	14.22
Axis 3	4.85	1.53	1.53	7.26
Axis 4	3.35	1.03	1.03	3.93

6. Application on cars data

We propose a simple application with the celebrated cars data (see our github).¹³ The dataset is particularly interesting since there are highly correlated variables as can be seen in the Pearson correlation matrix given in Table 10.

Also, the dataset is composed of some outlying observations (Fig. 2): Ferrari Enzo (x_1, x_2, x_5), Bentley Continental (x_2), Aston Martin Vanquish (x_2), Land Rover Discovery (x_5), Mercedes Class S 400 CDI (x_5), Smart Fortwo Coupé (x_5, x_6).

In what follows, we compare the Gini PCA of optimal parameter $\nu^* = 1.4$ ¹⁴ with the standard PCA, with Visentin et al. (2016) (ℓ_1 PCA), and with Brooks et al. (2013) (ℓ_1 PCA). The information (variability) captured by the principal components is the following (Table 11).

Gini and Variance yield approximately 6.5 times more variability over the first axis compared with the second one, 7.5 times for the other ℓ_1 PCA. Two axes may be chosen to analyze the data. As shown in the previous Section about the Monte Carlo simula-

¹³ A python code for Gini PCA is available <https://github.com/mussard974/Gini-PCA> and also an R markdown <https://github.com/freakonometrics/GiniACP/>.

¹⁴ The grid search has been performed over the two first principal components by removing the six most important outliers.

Table 12
Number of outliers.

	Grubbs test	Gini $\nu^* = 1.4$	Visentin et al.	Brooks et al.	Variance
Axis 1	0	0	0	0	0
Axis 2	0	0	0	0	1

Table 13
Correlations Axis 1,2 / variables (U -stat: * 1%, ** 5%).

Axis 1	Capacity	Power	Speed	Weight	Width	Length
Gini $\nu^* = 1.4$	-2.12*	-2.09*	-1.98*	-1.60*	-2.00*	-1.68*
Variance	0.92*	0.88*	0.85*	0.73*	0.77**	0.76*
Visentin et al.	-0.47*	-0.03*	-0.04*	-0.27*	-0.32*	-0.78*
Brooks et al.	-0.47*	-0.03*	-0.04*	-0.26*	-0.32*	-0.78*
Axis 2	capacity	power	speed	weight	width	length
Gini $\nu^* = 1.4$	-0.05	-0.22	-0.32	0.53**	0.16	0.42*
Variance	-0.12	-0.34	-0.32	0.55**	-0.11	0.48
Visentin et al.	-0.86	-0.08	0.00	-0.07	0.23	0.75
Brooks et al.	0.85	0.1	0.01	0.07	-0.23	-0.45

tions, when the data are highly correlated such that two axes are sufficient to project the data, the Gini PCA and the standard PCA yield the same share of information on each axis. However, we can expect some differences in absolute contributions ACT and relative contributions RCT.

The projection of the data is depicted for each method in Fig. 3.

As depicted in Fig. 3, the projections are very similar for the Gini and the variance. The cars with extraordinary (or very low) efficiency are in the same relative position in the four projections: Land Rover Discovery Td5 at the top, Ferrari Enzo at the bottom right, Smart Fortwo Coupé at the bottom left (Fig. 3(a) and (b)). In order to analyze the ability of each technique to dilute the outlying observations after projection onto the new subspace, we give below the Grubbs test over the two first principal components. Before projection, six observations were considered as outliers. After projection, we depict in Table 12 the number of outliers (both maximum and minimum values) over the two first principal components.

The theoretical Section 4 indicates that the Gini methodology for $\nu = 2$ is equivalent to the variance when the variables are Gaussian. About cars data for $\nu^* = 1.4$, we observe this similarity except over axis 2. Table 13 outlines: Gini correlations ($\nu^* = 1.4$) between the variables and axis 1 (and axis 2) with their significance (U -statistics tests (9)), Pearson correlation (Variance PCA),

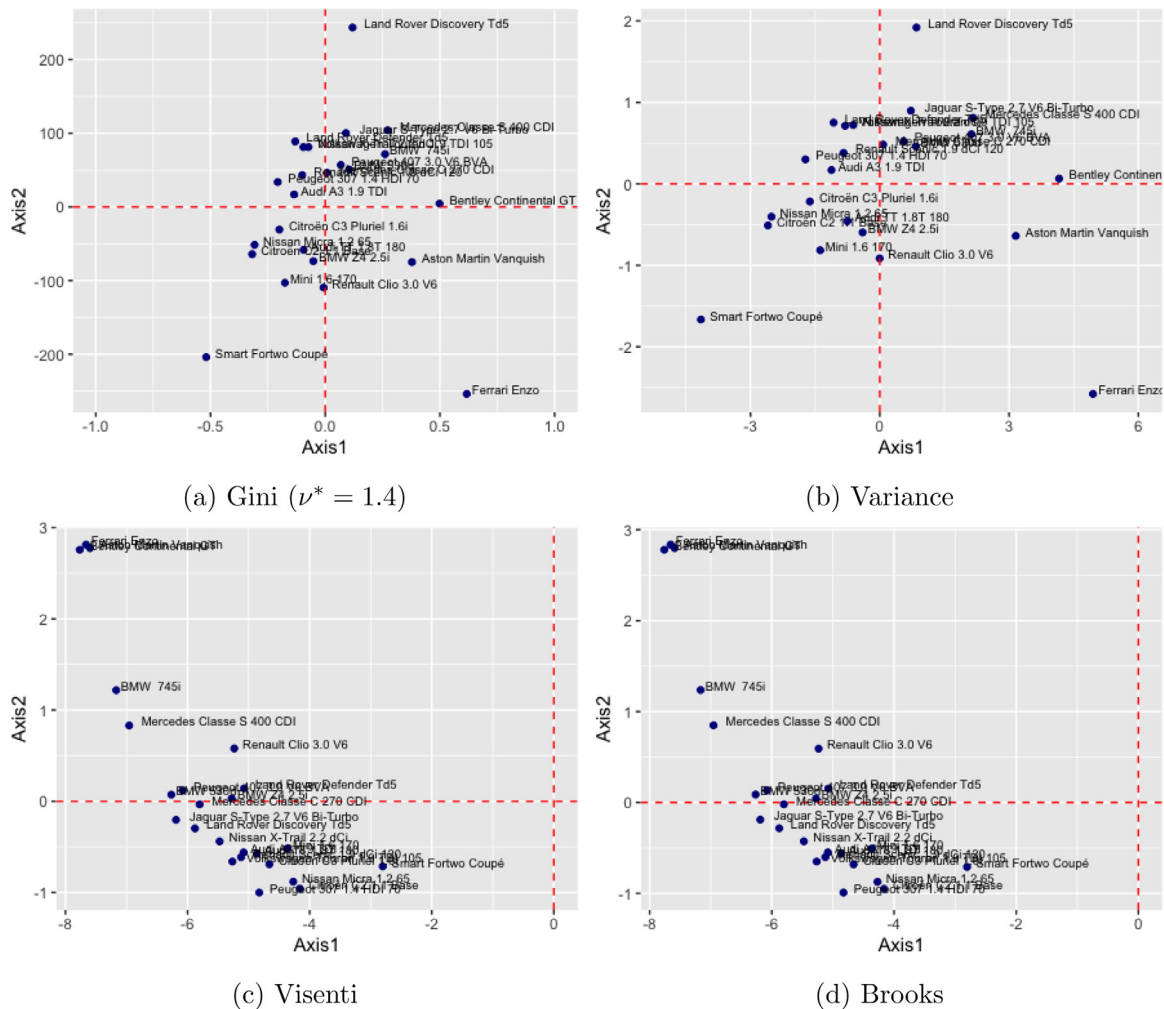


Fig. 3. Projections of the cars.

and factors loadings for the other ℓ_1 PCA. All correlations with the first principal component are significant.

Over the second principal component some differences appear: weight and length are correlated with axis 2 for the Gini PCA, whereas weight is the sole variable correlated with axis 2 in the standard PCA. The other ℓ_1 PCA, in which the variability is lower over axis 2, outline no significant correlation between the variables and axis 2.¹⁵

Let us now investigate absolute contributions in order to select the observations that explain an important part of variability (Figs. 4 and 5). Since there is no absolute contribution for the ℓ_1 PCA, we focus on the Gini PCA and the standard PCA. As expected, the difference between both PCA lies in the second principal component, in which the significance of the variables is different. Over the first principal component, each technique yields the same observations with a contribution higher than $100\%/24 \text{ cars} = 4.17\%$ per car (equal contribution): Citron C2, Smart Fortwo Coupé, Nissan Micra, BMW 745i, Mercedes Class S 400 CDI, Aston Martin Vanquish, Bentley Continental and Ferrari Enzo.

¹⁵ For Gini PCA and standard PCA, the Jackknife variance has been built by withdrawing a couple (f_{ik}, z_{ik}) and by re-estimating the Gini correlation, for each $i = 1, \dots, N$. For the other ℓ_1 PCA, we have removed one observation x_i from the sample and re-estimated the entire PCA. This may cause some slight differences in the U -statistics and then in the significance of the variables.

Table 14

ACT (%) Axis 2.

Gini ($\nu^* = 1.4$)	Variance
–	Smart Fortwo Coupé (13.55%)
Land Rover Discovery (27.95%)	Land Rover Discovery (18%)
Ferrari Enzo (27.14%)	Ferrari Enzo (32.5%)
Renault Clio 3.0 V6 (7.54%)	
Mercedes Class S 400 CDI (4.89%)	
Jaguar S-Type 2.7 V6 Bi-Turbo (4.36%)	
Mini 1.6 170 (4.28%)	

However, over the second principal component, the results are very different. The standard PCA yields Smart Fortwo Coupé, Ferrari Enzo and Land Rover Discovery with an ACT higher than 4.17%. The problem is that Smart Fortwo Coupé is an outlier in the distributions of width and length. As a consequence, the outlier absorbs a large part of the variability of axis 2 in the standard PCA (13.55%), so that the other observations are judged insignificant. Table 14 depicts the most important ACT of axis 2.

The Gini PCA yields length and weight associated with the second axis, and since the method dilute the importance of outliers (such as Smart Fortwo Coupé), more observations are related to the second principal component compared with the standard PCA. Indeed, the Gini PCA outlines cars with important weight and length

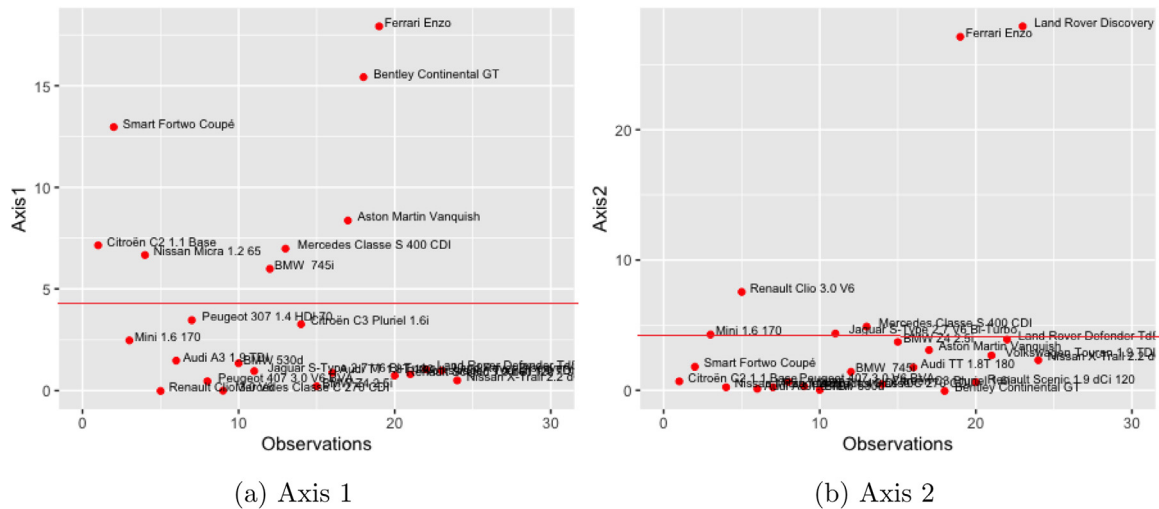


Fig. 4. ACT (Gini $\nu^* = 1.4$).

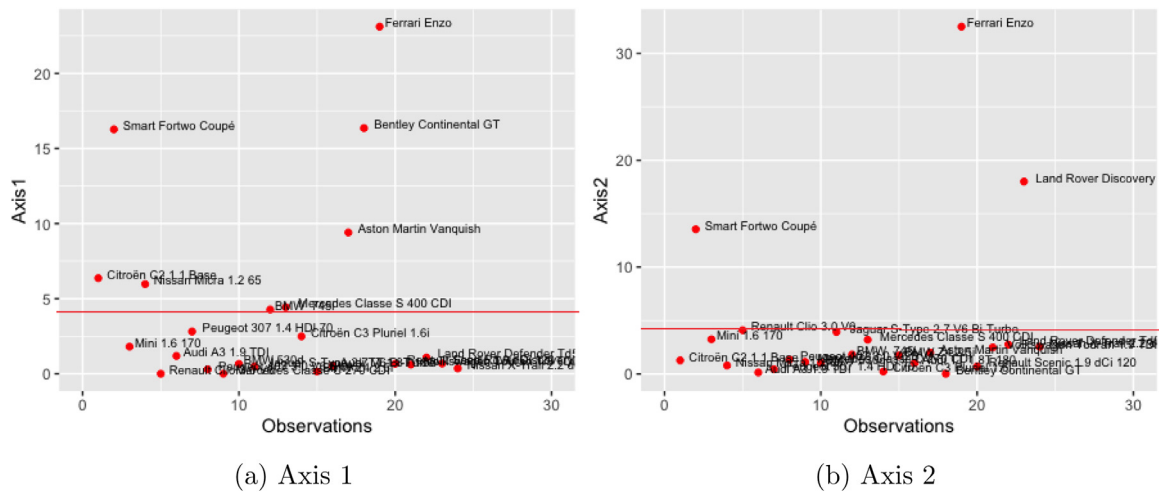


Fig. 5. ACT (Variance).

such as Mercedes Class S 400 CDI and Jaguar S-Type 2.7 V6 Bi-Turbo.

Finally, some cars are not correlated with axis 2 in the standard PCA, whereas this is the case in the Gini PCA. This example of application shows that the use of the Gini metric robust to outliers may involve some interesting changes in the interpretation of the results.

7. Conclusion

In this paper, it has been shown that the geometry of the Gini covariance operator allows one to perform Gini PCA, that is, a robust principal component analysis based on the ℓ_1 norm.

To be precise, the variance may be replaced by the Gini Mean Difference, which captures the variability of couples of variables based on the rank of the observations in order to attenuate the influence of the outliers. The Gini Mean Difference may be rather interpreted with the aid of the generalized Gini index $GGMD_\nu$ in the new subspace for a better understanding of the variability of the components, that is, $GGMD_\nu$ is both a rank-dependent measure of variability in Yaari (1987) sense and also an eigenvalue of the Gini correlation matrix.

Contrary to many approaches in multidimensional statistics in which the standard variance-covariance matrix is used to project

the data onto a new subspace before deriving multidimensional Gini indices (see e.g. Banerjee (2010)), we propose to employ the Gini correlation indices (see Yitzhaki & Schechtman, 2013). This provides the ability to interpret the results with the ℓ_1 norm and the use of U -statistics to measure the significance of the correlation between the new axes and the variables.

This research may open the way on data analysis based on Gini metrics in order to study multivariate correlations with categorical variables or discriminant analyses when outlying observations drastically affect the sample.

Appendix A. Variants of PCA based on the ℓ_2 norm

The classical formulation of the PCA, to obtain the first component, can be obtained by solving

$$\omega_1^* \in \operatorname{argmax}\{\operatorname{Var}\{X\omega\}\} \text{ subject to } \|\omega\|_2^2 = \omega^\top \omega = 1, \tag{A.1}$$

or equivalently

$$\omega_1^* \in \operatorname{argmax}\{\omega^\top \Sigma \omega\} \text{ subject to } \|\omega\|_2^2 = \omega^\top \omega = 1, \tag{A.2}$$

where $\omega \in \mathbb{R}^K$, and Σ is the (symmetric positive semi-definite) $K \times K$ sample covariance matrix. Mardia, Kent, and Bibby (1979) sug-

gest to write

$$\omega_1^* \in \operatorname{argmax} \left\{ \sum_{j=1}^K \operatorname{Var}[\mathbf{x}_{\cdot,j}] \cdot \operatorname{Cor}[\mathbf{x}_{\cdot,j}, \mathbf{X}\omega] \right\} \text{ subject to } \|\omega\|_2^2 = \omega^\top \omega = 1.$$

With scaled variables¹⁶ (i.e. $\operatorname{Var}[\mathbf{x}_{\cdot,j}] = 1, \forall j$)

$$\omega_1^* \in \operatorname{argmax} \left\{ \sum_{j=1}^K \operatorname{Cor}[\mathbf{x}_{\cdot,j}, \mathbf{X}\omega] \right\} \text{ subject to } \|\omega\|_2^2 = \omega^\top \omega = 1. \tag{A.3}$$

Then a Principal Component Pursuit can start: we consider the ‘residuals’, $\mathbf{X}_{(1)} = \mathbf{X} - \mathbf{X}\omega_1^*\omega_1^{*\top}$, its covariance matrix $\Sigma_{(1)}$, and we solve

$$\omega_2^* \in \operatorname{argmax} \{ \omega^\top \Sigma_{(1)} \omega \} \text{ subject to } \|\omega\|_2^2 = \omega^\top \omega = 1.$$

The part $\mathbf{X}\omega_1^*\omega_1^{*\top}$ is actually a constraint that we add to ensure the orthogonality of the two first components. This problem is equivalent to finding the maxima of $\operatorname{Var}[\mathbf{X}\omega]$ subject to $\|\omega\|_2^2 = 1$ and $\omega \perp \omega_1^*$. This idea is also called Hotelling (or Wielandt) deflation technique. On the k -th iteration, we extract the leading eigenvector

$$\omega_k^* \in \operatorname{argmax} \{ \omega^\top \Sigma_{(k-1)} \omega \} \text{ subject to } \|\omega\|_2^2 = \omega^\top \omega = 1,$$

where $\Sigma_{(k-1)} = \Sigma_{(k-2)} - \omega_{k-1}^*\omega_{k-1}^{*\top} \Sigma_{(k-1)} \omega_{k-1}^*\omega_{k-1}^{*\top}$ (see e.g. Saad, 1998). Note that, following Hotelling (1933) and Eckart and Young (1936), that it is also possible to write this problem as

$$\min \{ \|\mathbf{X} - \tilde{\mathbf{X}}\|_* \} \text{ subject to } \operatorname{rank}[\tilde{\mathbf{X}}] \leq k$$

where $\|\cdot\|_*$ denotes the nuclear norm of a matrix (i.e. the sum of its singular values).¹⁷

One extension, introduced in d’Aspremont, Ghaoui, Jordan, and Lanckriet (2007), was to add a constraint based on the cardinality of ω (also called ℓ_0 norm) corresponding to the number of non-zero coefficients of ω . The penalized objective function is then

$$\max \{ \omega^\top \Sigma \omega - \lambda \operatorname{card}[\omega] \} \text{ subject to } \|\omega\|_2^2 = \omega^\top \omega = 1,$$

for some $\lambda > 0$. This is called sparse PCA, and can be related to sparse regression, introduced in Tibshirani (1996). But as pointed out in Mackey (2009), interpretation is not easy and the components obtained are not orthogonal. Gorban, Kegl, Wunsch, and Zinovjev (2007) considered an extension to nonlinear Principal Manifolds to take into account nonlinearities.

Another direction for extensions was to consider Robust Principal Component Analysis. Candès, Li, Ma, and Wright (2009) suggested an approach based on the fact that principal component pursuit can be obtained by solving

$$\min \{ \|\mathbf{X} - \tilde{\mathbf{X}}\|_* + \lambda \|\tilde{\mathbf{X}}\|_1 \}.$$

But other methods were also considered to obtain Robust PCA. A natural ‘scale-free’ version is obtained by considering a rank matrix instead of \mathbf{X} . This is also called ‘ordinal’ PCA in the literature, see Korhonen and Siljamäki (1998). The first ‘ordinal’ component is

$$\omega_1^* \in \operatorname{argmax} \left\{ \sum_{j=1}^K \mathcal{R}[\mathbf{x}_{\cdot,j}, \mathbf{X}\omega] \right\} \text{ subject to } \|\omega\|_2^2 = \omega^\top \omega = 1 \tag{A.4}$$

¹⁶ In most cases, PCA is performed on scaled (and centered) variables, otherwise variables with large scales might alter interpretations. Thus, it will make sense, later on, to assume that components of \mathbf{X} have identical distributions. At least the first two moments will be equal.

¹⁷ but other norms have also been considered in statistical literature, such as the Frobenius norm in the Eckart-Young theorem, or the maximum of singular values – also called 2-(induced)-norm.

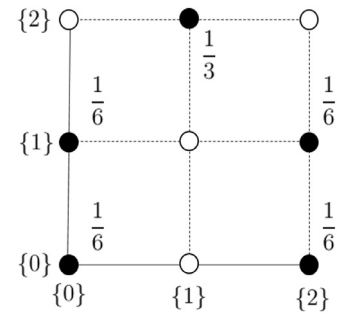


Fig. B1. Joint distribution of a random pair (X_k, X_l) such that $\mathbb{E}[X_k h(X_l)] \neq \mathbb{E}[X_l h(X_k)]$, with non-exchangeable components $X_k \not\stackrel{d}{=} X_l$.

where \mathcal{R} denotes some rank based correlation, e.g. Spearman’s rank correlation, as an extension of Eq. (A.3). So, quite naturally, one possible extension of Eq. (A.2) would be

$$\omega_1^* \in \operatorname{argmax} \{ \omega^\top \mathcal{R}[\mathbf{X}] \omega \} \text{ subject to } \|\omega\|_2^2 = \omega^\top \omega = 1$$

where $\mathcal{R}[\mathbf{X}]$ denotes Spearman’s rank correlation. In this section, instead of using Pearson’s correlation (as in Eq. (A.2) when the variables are scaled) or Spearman’s (as in this ordinal PCA), we will consider the multidimensional Gini correlation based on the h -covariance operator.

Appendix B. Pairwise exchangeability and the Gini-covariance

Pairwise exchangeability is one property respected by the Gini-covariance as well as the Gini correlation coefficient. Note that it is not required for Gini PCA.

Definition B.1. \mathbf{X} is said to be pairwise-exchangeable if for all pair $(i, j) \in \{1, \dots, K\}^2$, (X_i, X_j) is exchangeable, in the sense that $(X_i, X_j) \stackrel{d}{=} (X_j, X_i)$.

Pairwise-exchangeability is a stronger concept than having only one vector with identically distributed components, and a weaker concept than (full) exchangeability. In the Gaussian case where $h(X_k) = \Phi(X_k)$ with $\Phi(X_k)$ being the normal cdf of X_k for all $k = 1, \dots, K$, pairwise-exchangeability is equivalent to components identically distributed.

Proposition B.1. If \mathbf{X} is a Gaussian vector with identically distributed components, then \mathbf{X} is pairwise-exchangeable.

Proof. For simplicity, assume that components of \mathbf{X} are $\mathcal{N}(0, 1)$ random variables, then $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\rho})$ where $\boldsymbol{\rho}$ is a correlation matrix. In that case

$$\begin{bmatrix} X_i \\ X_j \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ji} & 1 \end{bmatrix} \right),$$

with Pearson correlation $\rho_{ij} = \rho_{ji}$, thus (X_i, X_j) is exchangeable. \square

In the very general case of a function h , results can be derived in dimension 2.

Proposition B.2. If $\mathbf{X} = (X_1, X_2)$ is a Gaussian vector with Pearson correlation r , with $\mathcal{N}(\mu, \sigma^2)$ margins, then

$$\operatorname{Cov}(X_2, h(X_1)) = r\sigma_2 \cdot \operatorname{Cov}(Z, h(\mu + \sigma Z)), \text{ with } Z \sim \mathcal{N}(0, 1).$$

Proof. If $\mathbf{X} = (X_1, X_2)$ is a Gaussian vector with Pearson correlation r , with $\mathcal{N}(\mu, \sigma^2)$ margins. Then

$$\mathbb{E}[X_2|X_1] = \mu + r\sigma \frac{X_1 - \mu}{\sigma} = \mu + r\sigma Z$$

where $Z \sim \mathcal{N}(0, 1)$. Further

$$\text{Cov}(X_2, h(X_1)) = \mathbb{E}[(\mathbb{E}[X_2|X_1] - \mu) \cdot (h(X_1) - \bar{h}_1)]$$

$$\text{where } \bar{h}_1 = \mathbb{E}[h(X_1)],$$

which can be written

$$\text{Cov}(X_2, h(X_1)) = r\sigma_2 \cdot \text{Cov}(Z, h(\mu + \sigma Z)), \text{ with } Z \sim \mathcal{N}(0, 1). \square$$

In dimension 2, even if we do not assume normality of X_1 and X_2 , the following property holds

Proposition B.3. *If \mathbf{X} is a pair of exchangeable variables, and if h is a non-decreasing function, then $C_h(\mathbf{X}) = \text{Cov}(\mathbf{X}, h(\mathbf{X}))$ is a symmetric positive semi-definite matrix.*

Proof. Observe that

$$C_h(\mathbf{X}) = \text{Cov}(\mathbf{X}, h(\mathbf{X})) = \begin{bmatrix} \text{Cov}(X_1, h(X_1)) & \text{Cov}(X_1, h(X_2)) \\ \text{Cov}(X_2, h(X_1)) & \text{Cov}(X_2, h(X_2)) \end{bmatrix}$$

is a symmetric matrix.

As mentioned in Yitzhaki (2003), page 150, terms on the diagonal $\text{Cov}(X_i, h(X_i))$ are positive since h is an increasing function. Thus $\text{trace}[C_h(\mathbf{X})] \geq 0$. Further, because diagonal elements are identical in the matrix (namely $\text{Cov}(X_1, h(X_1)) = \text{Cov}(X_2, h(X_2))$ and $\text{Cov}(X_1, h(X_2)) = \text{Cov}(X_2, h(X_1))$), the determinant can be written

$$\det[C_h(\mathbf{X})] = \text{Cov}(X_1, h(X_1))^2 - \text{Cov}(X_2, h(X_1))^2.$$

But from Hardy-Littlewood-Polya inequality, $\text{Cov}(X_2, h(X_1))$ is maximal when components are comonotonic, and

$$\max_{X_2: X_2 \stackrel{c}{\leq} X_1} \{\text{Cov}(X_2, h(X_1))\} = \text{Cov}(X_2^*, h(X_1))$$

where X_2^* is comonotonic with $h(X_1)$, meaning $X_2^* = X_1$. So $\text{Cov}(X_1, h(X_1)) \geq \text{Cov}(X_2, h(X_1))^2$, and the determinant is positive.

So, since $C_h(\mathbf{X})$ is a symmetric matrix, with $\text{trace}[C_h(\mathbf{X})] \geq 0$ and $\det[C_h(\mathbf{X})] \geq 0$, then $C_h(\mathbf{X})$ is positive semi-definite. \square

Appendix C. Proofs

Theorem 2.1:

(i) In the Gaussian case, if h is the cumulative distribution function of the X_k 's, then $\text{Cov}(X_k, h(X_\ell)) = r\sigma \cdot \text{Cov}(Z, \Phi(Z))$, where Φ is the normal cdf, see Yitzhaki and Schechtman (2013), Chapter 3. Observe that $\text{Cov}(X_k, h(X_k)) = \sigma \cdot \text{Cov}(Z, \Phi(Z))$, if h is the cdf of X_k . Thus, $\lambda := \text{Cov}(Z, \Phi(Z))$ yields:

$$\Gamma_{C_F}((X_k, X_\ell)) = \lambda \begin{bmatrix} \sigma & \rho\sigma \\ \rho\sigma & \sigma \end{bmatrix} = \lambda\sigma \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \frac{\lambda}{\sigma} \text{Var}((X_k, X_\ell)).$$

(ii) We have $\text{Cov}(Z, \Phi(Z)) \geq 0$, then it follows that $C_F((X_k, X_\ell)) \geq 0$:

$$\mathbf{x}^\top C_F(\mathbf{X}) \mathbf{x} = \mathbf{x}^\top \frac{\text{Cov}(Z, \Phi(Z))}{\sigma} \text{Var}(\mathbf{X}) \mathbf{x} \geq 0,$$

which ends the proof.

Proposition 2.1: We have $GMD_v(\mathbf{Z}) \equiv [GMD_v(\mathbf{z}_k, \mathbf{z}_\ell)]$ being a $K \times K$ matrix. The extra diagonal terms may be rewritten as,

$$\begin{aligned} GMD_v(\mathbf{z}_k, \mathbf{z}_\ell) &= -\frac{2}{N-1} \nu \text{Cov}(\mathbf{z}_k, \mathbf{r}_{\mathbf{z}_\ell}^{\nu-1}) \\ &= -\frac{2}{N-1} \nu \text{Cov}\left(\frac{\mathbf{x}_k - \bar{\mathbf{x}}_k \mathbf{1}_N}{GMD_v(\mathbf{x}_k, \mathbf{x}_k)}, \mathbf{r}_{\mathbf{z}_\ell}^{\nu-1}\right) \\ &= -\frac{2}{GMD_v(\mathbf{x}_k, \mathbf{x}_k)} \left[\frac{\nu \text{Cov}(\mathbf{x}_k, \mathbf{r}_{\mathbf{z}_\ell}^{\nu-1})}{N-1} - \frac{\nu \text{Cov}(\bar{\mathbf{x}}_k \mathbf{1}_N, \mathbf{r}_{\mathbf{z}_\ell}^{\nu-1})}{N-1} \right] \\ &= \frac{GMD_v(\mathbf{x}_k, \mathbf{x}_\ell)}{GMD_v(\mathbf{x}_k, \mathbf{x}_k)} = GC_v(\mathbf{x}_k, \mathbf{x}_\ell). \end{aligned}$$

Finally, using the same approach as before, we get:

$$\begin{aligned} GMD_v(\mathbf{z}_k, \mathbf{z}_k) &= -\frac{2}{N-1} \nu \text{Cov}(\mathbf{z}_k, \mathbf{r}_{\mathbf{z}_k}^{\nu-1}) \\ &= \frac{GMD_v(\mathbf{x}_k, \mathbf{x}_k)}{GMD_v(\mathbf{x}_k, \mathbf{x}_k)} \\ &= GC_v(\mathbf{x}_k, \mathbf{x}_k) = 1. \end{aligned}$$

By Property 2.1 (iv), since $\mathbf{r}_{\mathbf{x}_k} = \mathbf{r}_{\mathbf{z}_k}$, then $GC_v(\mathbf{x}_k, \mathbf{x}_\ell) = GC_v(\mathbf{z}_k, \mathbf{z}_\ell)$. Thus,

$$GMD_v(\mathbf{Z}) = -\frac{2\nu}{N(N-1)} \mathbf{Z}^\top \mathbf{R}_Z^c = GC_v(\mathbf{X}) = GC_v(\mathbf{Z}),$$

which ends the proof.

Proposition 3.1:

(i) The result comes from the rank-nullity theorem. From the eigenvalue Eq. (7), we have:

$$\mathbf{b}_k^\top GC_v(\mathbf{Z}) \mathbf{b}_k = \lambda_{\cdot k} / 2 = \mu_{\cdot k}.$$

Let f be the linear application issued from the matrix $GC_v(\mathbf{Z})$. Whenever $\lambda_{\cdot k} = 0$, two columns (or rows) of $GC_v(\mathbf{Z})$ are collinear, then the dimension of the image set of f is $\dim(f) = K - 1$. Hence, $\mathbf{f}_k = \mathbf{0}$. Since $\mathbf{b}_k^\top GC_v(\mathbf{Z}) \mathbf{b}_k = GGMD_v(\mathbf{f}_k, \mathbf{f}_k)$ for all $k = 1, \dots, K$, then for $\lambda_{\cdot k}$ we get:

$$\mathbf{b}_k^\top GC_v(\mathbf{Z}) \mathbf{b}_k = GGMD_v(\mathbf{f}_k, \mathbf{f}_k) = \lambda_{\cdot k} / 2 = \mu_{\cdot k} = 0.$$

On the other hand, since $\mathbf{f}_k = \mathbf{0}$, it follows that $GGMD_v(\mathbf{f}_k, \mathbf{f}_\ell) = 0$ for all $\ell = 1, \dots, K$. Also, if $\mathbf{f}_k = \mathbf{0}$ then the centered rank vector $\mathbf{r}_{\mathbf{f}_k}^c = \mathbf{0}$, and so $GGMD_v(\mathbf{f}_\ell, \mathbf{f}_k) = 0$ for all $\ell = 1, \dots, K$.

(ii) The proof comes from the Rayleigh-Ritz identity:

$$\lambda_{\max} := \max \frac{\mathbf{b}_1^\top [GC_v(\mathbf{Z}) + GC_v(\mathbf{Z})^\top] \mathbf{b}_1}{\mathbf{b}_1^\top \mathbf{b}_1} = \lambda_1.$$

Since $\mathbf{b}_1^\top GC_v(\mathbf{Z}) \mathbf{b}_1 = \lambda_1 / 2$ and because $\mathbf{b}_1^\top GC_v(\mathbf{Z}) \mathbf{b}_1 = GGMD_v(\mathbf{f}_1, \mathbf{f}_1)$, the result follows.

(iii) Again, the Rayleigh-Ritz identity yields:

$$\lambda_{\min} := \min \frac{\mathbf{b}_{\cdot K}^\top [GC_v(\mathbf{Z}) + GC_v(\mathbf{Z})^\top] \mathbf{b}_{\cdot K}}{\mathbf{b}_{\cdot K}^\top \mathbf{b}_{\cdot K}} = \lambda_K.$$

Then, $\mathbf{b}_{\cdot K}^\top GC_v(\mathbf{Z}) \mathbf{b}_{\cdot K} = GGMD_v(\mathbf{f}_{\cdot K}, \mathbf{f}_{\cdot K}) = \lambda_{\cdot K} / 2$.

Proposition 3.2:

From Theorem 2.2:

$$GC_v(\mathbf{X}) = \sigma^{-1} \text{Cov}(Z, \Phi(Z)) \text{Var}(\mathbf{X}).$$

From Abramowitz and Stegun (1964) (Chapter 26), when $Z \sim \mathcal{N}(0, 1)$,

$$\text{Cov}(Z, \Phi(Z)) = \frac{1}{2\sqrt{\pi}} \approx 0.2821.$$

Then the results (i), (iii), and (iv) follow directly. (ii) Since by Proposition 3.1 $GGMD_v(\mathbf{f}_k, \mathbf{f}_k) = \mu_{\cdot k}$, and because $GC_v(\mathbf{Z}) + GC_v(\mathbf{Z})^\top$ is positive semi-definite, then $\mu_{\cdot k} \geq 0$ for all $k = 1, \dots, K$.

Remark 3.1: Note that $GMD_v(\mathbf{f}_{\cdot k}, \mathbf{f}_{\cdot k}) = -\frac{2\nu}{N(N-1)} (\mathbf{f}_{\cdot k} - \bar{\mathbf{f}})^\top \mathbf{r}_{c, \mathbf{f}_{\cdot k}}^{\nu-1}$, where $\mathbf{r}_{c, \mathbf{f}_{\cdot k}}^{\nu-1}$ is the k th column of the centered (decumulative) rank matrix \mathbf{R}_f^c . Since $\mathbf{f}_{\cdot k} = \mathbf{Z} \mathbf{b}_{\cdot k}$ and $\bar{\mathbf{f}} = (\bar{\mathbf{f}}_1, \dots, \bar{\mathbf{f}}_K) = \mathbf{0}$ then:¹⁸

$$\mathbf{b}_k^\top GC_v(\mathbf{Z}) \mathbf{b}_k = -\frac{2\nu}{N(N-1)} \mathbf{b}_k^\top \mathbf{Z}^\top \mathbf{R}_Z^c \mathbf{b}_k \tag{C.1}$$

¹⁸ We have:

$$\bar{\mathbf{f}}_k = 1/N \sum_{i=1}^N f_{ik} = 1/N \left[\sum_{i=1}^N \mathbf{z}_i^\top \mathbf{b}_k \right] = 1/N \left[\sum_{i=1}^N \mathbf{z}_i^\top \right] \mathbf{b}_k = \mathbf{0}.$$

$$\begin{aligned}
 &= -\frac{2\nu}{N(N-1)} \mathbf{b}_k^\top \mathbf{Z}^\top \mathbf{r}_{c,\mathbf{f}_k}^{\nu-1} (\mathbf{b}_k \mathbf{r}_{c,\mathbf{f}_k}^{\nu-1} = \mathbf{R}_z^c \mathbf{b}_k) \\
 &= -\frac{2\nu}{N(N-1)} \mathbf{f}_k^\top \mathbf{r}_{c,\mathbf{f}_k}^{\nu-1} \\
 &= \text{GMD}_\nu(\mathbf{f}_k, \mathbf{f}_k).
 \end{aligned}$$

Then, maximizing the multidimensional variability $\mathbf{b}_k^\top \text{GC}_\nu(\mathbf{Z}) \mathbf{b}_k$ yields from (7):

$$\mathbf{b}_k^\top \text{GC}_\nu(\mathbf{Z}) + \text{GC}_\nu(\mathbf{Z})^\top \mathbf{b}_k = \mathbf{b}_k^\top \lambda_{\cdot,k} \mathbf{b}_k$$

$$\iff \mathbf{b}_k^\top \text{GC}_\nu(\mathbf{Z}) \mathbf{b}_k + \mathbf{b}_k^\top \text{GC}_\nu(\mathbf{Z})^\top \mathbf{b}_k = \lambda_{\cdot,k}.$$

Since $\mathbf{b}_k^\top \text{GC}_\nu(\mathbf{Z}) \mathbf{b}_k = \mathbf{b}_k^\top \text{GC}_\nu(\mathbf{Z})^\top \mathbf{b}_k$, then

$$\mathbf{b}_k^\top \text{GC}_\nu(\mathbf{Z})^\top \mathbf{b}_k = \lambda_{\cdot,k}/2 = \mu_{\cdot,k},$$

and so $\text{GMD}_\nu(\mathbf{f}_k, \mathbf{f}_k) = \lambda_{\cdot,k}$ for all $k = 1, \dots, K$. The results (ii) and (iii) are straightforward.

References

Abramowitz, M., & Stegun, I. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. National Bureau of Standards Applied Mathematics Series No. 55.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics*, 34, 122–148.

Baccini, A., Besse, P., & de Falguerolles, A. (1996). A l_1 norm PCA and a heuristic approach. In E. Didday, Y. Lechevalier, & O. Opitz (Eds.), *Ordinal and symbolic data analysis* (pp. 359–368). Springer.

Banerjee, A. K. (2010). A multidimensional gini index. *Mathematical Social Sciences*, 60, 87–93.

Brooks, J. P., Dulá, J. H., & Boone, E. L. (2013). A pure l_1 -norm principal component analysis. *Computational Statistics & Data Analysis*, 61, 83–98.

Candes, E. J., Li, X., Ma, Y., & Wright, J. (2009). Robust principal component analysis? ArXiv:0912.3599.

Carcea, M., & Serfling, R. (2015). A gini autocovariance function for time series modeling. *Journal of Time Series Analysis*, 36, 817–838.

Charpentier, A., & Segers, J. (2009). Tails of multivariate archimedean copulas. *Journal of Multivariate Analysis*, 100, 1521–1537.

d’Aspremont, A., Ghaoui, L. E., Jordan, M. I., & Lanckriet, G. R. G. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Review*, 49(3), 434–448.

Decancq, K., & Lugo, M. A. (2013). Weights in multidimensional indices of well-being: An overview. *Econometric Reviews*, 32(1), 7–34.

Ding, C., Zhou, D., He, X., & Zha, H. (2006). r_1 -PCA: Rotational invariant l_1 -norm principal component analysis for robust subspace factorization. In *ICML ’06 Proceedings of the 23rd international conference on Machine learning* (pp. 281–288).

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211–218.

Flury, B., & Riedwyl, H. (1988). *Multivariate statistics: A practical approach*. Chapman & Hall.

Furman, E., & Zitikis, R. (2017). Beyond the pearson correlation: Heavy-tailed risks, weighted Gini correlations, and a Gini-type weighted insurance pricing model. *ASTIN Bulletin*, 47(03), 919–942.

Gajdos, T., & Weymark, J. (2005). Multidimensional generalized Gini indices. *Economic Theory*, 26(3), 471–496.

Gini, C. (1912). Variabilità e mutabilità, memori di metodologia statistica. *Variabilità e Concentrazione. Libreria Eredi Virgilio Veschi, Rome*, 1, 211–382.

Giorgi, G. M. (2013). Back to the future: Some considerations on shlomo yitzhaki and edna schechtman’s book the gini methodology: A primer on a statistical methodology. *Metron*, 71(2), 189–195.

Gorban, A. N., Kegl, B., Wunsch, D. C., & Zinovyev, A. (2007). Principal manifolds for data visualisation and dimension reduction. In *LNCSE 58*. Springer Verlag.

Hofert, M. (2011). Efficiently sampling nested archimedean copulas. *Computational Statistics & Data Analysis*, 55, 57–70.

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

Korhonen, P., & Siljamäki, A. (1998). Ordinal principal component analysis theory and an application. *Computational Statistics & Data Analysis*, 26(4), 411–424.

Kwak, N. (2008). Principal component analysis based on l_1 -norm maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9), 1672–1680.

Laurini, M. P., & Ohashi, A. (2015). A noisy principal component analysis for forward rate curves. *European Journal of Operational Research*, 246(1), 140–153.

Li, C., Shao, Y.-H., & Deng, N. Y. (2015). Robust l_1 -norm two-dimensional linear discriminant analysis. *Neural Networks*, 65, 92–104.

List, C. (1999). Multidimensional inequality measurement: A proposal. *Working paper*. Nuffield College.

Mackey, L. (2009). Deflation methods for sparse PCA. *Advances in Neural Information Processing Systems*, 21, 1017–1024.

Mardia, K., Kent, J., & Bibby, J. (1979). *Multivariate analysis*. London: Academic Press.

McNeil, A. J. (2008). Sampling nested archimedean copulas. *Journal of Statistical Computation and Simulation*, 78, 567–581.

Nelsen, R. (2007). *Introduction to copulas*. Springer Verlag.

Olkin, I., & Yitzhaki, S. (1992). Gini regression analysis. *International Statistical Review*, 60, 185–196.

Pearson, K. (1901). On lines and planes of closest fit to system of points in space. *Philosophical Magazine*, 2, 559–572.

Saad, Y. (1998). Projection and deflation methods for partial pole assignment in linear state feedback. *IEEE Transactions on Automatic Control*, 33, 290–297.

Schechtman, E., & Yitzhaki, S. (1987). A measure of association based on Gini’s mean difference. *Communications in Statistics*, 16, 207–231.

Schechtman, E., & Yitzhaki, S. (2003). A family of correlation coefficients based on the extended Gini index. *Journal of Economic Inequality*, 1(2), 129–146.

Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5), 1299–1319.

Shelef, A. (2016). A Gini-based unit root test. *Computational Statistics & Data Analysis*, 100, 763–772.

Shelef, A., & Schechtman, E. (2011). A Gini-based methodology for identifying and analyzing time series with non-normal innovations. *SSNR Electronic Journal*. July: 1–26

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal statistical society, series B*, 58(1), 267–288.

Visentini, A., Prestwich, S., & Armagan, S. T. (2016). Robust principal component analysis by reverse iterative linear programming. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 593–605.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, 55, 99–115.

Yaari, M. E. (1988). A controversial proposal concerning inequality measurement. *Journal of Economic Theory*, 44, 381–397.

Yitzhaki, S. (1991). Calculating jackknife variance estimators for parameters of the Gini method. *Journal of Business and Economic Statistics*, 9, 235–239.

Yitzhaki, S. (2003). Gini’s mean difference: A superior measure of variability for non-normal distributions. *Metron*, LXI(2), 285–316.

Yitzhaki, S., & Schechtman, E. (2013). The Gini methodology. In *A primer on a statistical methodology*. Springer.

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