DYNAMIC CUR, AN ALTERNATIVE TO VARIABLE SELECTION IN CUR DECOMPOSITION

Greibin Villegas Barahona*[1], Carlos Manuel Martín Barreiro**,Nerea González García***, Sergio Hernández González****, Mercedes Sánchez Barba***, María Purificación Galindo Villardón***
*Universidad Estatal a Distancia, Costa Rica
** ESPOL Polytechnic University, ESPOL, FCNM, Guayaquil, Ecuador
*** Universidad de Salamanca, España
**** Universidad Veracruzana, México

ABSTRACT
CUR decomposition is one of the matrix decomposition techniques proposed in the literature for the selection of rows and/or columns of a data matrix. Dynamic CUR is proposed as an alternative to the selection criteria of the CUR decomposition based on probabilistic criteria. This alternative tries to fit the most adequate theoretical probability distribution to the empirical distribution of the leverages obtained from the start and based on it, automatically determines not only the individuals and/or variables that need to be selected, but also their numbers. In this way, Dynamic CUR sets itself apart from CUR in the information selection criteria, dynamizing the calculation of the approximation error starting from an optimal initial selection of parameters based on the most adequate probability distribution. Lastly, with the purpose of facilitating the use of this new method in any practical context, the Dynamic CUR algorithm has been developed in C#.NET and R languages.

KEYWORDS: Multivariate analysis, Principal component analysis, CUR decomposition, Correlation, Singular Value Decomposition.

MSC: 62E17, 62G30, 49M27

RESUMEN
La descomposición CUR es una de las técnicas de descomposición matricial propuesta en la literatura para la selección de filas y/o columnas de una matriz de datos. Se propone Dinamic CUR, una alternativa al criterio de selección de la descomposición CUR basada en criterios probabilísticos. Esta propuesta trata de ajustar la distribución de probabilidad teórica más adecuada a la distribución empírica de los puntajes altamente influyentes obtenidos de partida y, a partir de ella, determina de manera automática no sólo los individuos y/o variables a seleccionar sino el número de ellas. Así, DinamicCUR se diferencia de CUR en el criterio de selección de la información, dinamizando el cálculo del error de aproximación a partir de una óptima elección inicial de parámetros en base a la distribución de probabilidad más adecuada. Por último, con el fin de facilitar el uso de este nuevo método en cualquier contexto práctico, se ha desarrollado el algoritmo DinamicCUR en lenguaje C#.NET y R.

PALABRAS CLAVE: Análisis Multivariado, Análisis de Componentes Principales, CUR descomposición, Correlación, Descomposición Singular Del Valor.

1. INTRODUCTION

Data analysis has evolved in a considerable manner in recent years, going from basic descriptive analysis with few variables to the use of multivariate statistical techniques to work with "Big Data". Analysis tools must take advantage of all the information the data provides to be able to accurately reproduce reality because the power of individual studies and the understanding of all the phenomena lies in the multivariate vision of the world):

Principal Component Analysis (PCA) (Jolliffe, 2002) came to be from the classical methods to reduce data matrix dimensionality, started by Pearson (1901) who searched for a better fitted subspace and Hotelling (1933) for variance maximization. PCA is the most used multivariate dimension reducing technique used to date. It allows, starting from a p set of related variables, to extract q non-correlated latent variables (known as principal components), with q<<p, and using them to get to know the sample's behavior, absorbing the most amount of variability possible.

[1] gvillegas@uned.ac.cr

Like many other multivariate techniques, PCA is a method that can be derived straight from Singular Value Decomposition (SVD)): Given that each PC is calculated from the singular vectors obtained in the SVD of the data matrix, these are expressed considering all the original variables, each PC is calculated as a combination of all of the original variables too. The coefficients of these linear combinations, called loadings, indicate the contribution of each of the original variables to each PC.

Unfortunately, understanding linear combinations as the mathematical abstractions that they are, and due to the fact that, in practice, loadings are usually different from zero, there is no guarantee of providing meaning to these mathematical concepts, and consequently, meaning to the PCs. This hinders the informative capability of the data and raises the main inconvenient of PCA, it's interpretation.

Throughout the years, different alternatives have emerged in literature to try and improve the informative capability of these types of results. Among the rotation methods, Kaiser (1958) proposes the Varimax rotation with the purpose of rotating the new axes in search of the variables that present the highest loadings on them. In the late 60's, with the goal of making components interpretation easier, Jeffer suggested the use of a "subjective" threshold regarding factorial loadings, so that those variables which loadings were below the given threshold's value would not be considered. These ideas were supported by Hausman (1982) and Cadima & Jolliffe):

The first ideas in the determination of principal variables in a data matrix were introduced by McCabe (1982) who gave the first steps to get to CUR Decomposition. In the late 90's and early 2000's Tibshiriani (1996) and Vines (2000) laid out their ideas. Tibshiriani first defined the Lasso "Least Absolute Shrinkage and Selection Operator" norm in the ambit of regression. With Lasso, the sum of the absolute values of the coefficients of each coefficient vector is penalized, giving place to the automatic selection methods of relevant variables.

On the other hand, Vines (2000) continued working on the ideas previously raised by Hausman (1982) and proposed the restriction of loadings values to a small subset of whole numbers ({-1,0,1}): Later, several authors like Jolliffe, Trendafilov & Uddin (2003) and Zou, Hastie & Tibshirani (2006) reformulated PCA, giving place to Sparse PCA. This methodology is raised with the purpose of generating projection vectors in the new subspace that have some of their loadings as null, through the addition of a penalization coefficient on the optimization problem of PCA.

Jolliffe, Trendafilov & Uddin (2003) proposed SCoTLASS (Simplified Component Technique for Least Absolute Shrinkage and Selection), in which the Lasso penalization is added to the variance maximization problem on PCA developed by Hotelling in 1933. On the other hand, (Zou, Hastie, & Tibshirani, 2006) proposed SPCA, a reformulation of PCA as a problem of optimization in minimizing error with the Elastic net (Zou & Hastie, 2005) penalty. As a result, Sparse PCA became one of the first methods which simultaneously reduces dimensionality and automatically selects variables. It is important to mention that, even though a part of the variability explained by latent variables is sacrificed, an important improvement in the interpretation of the PCs is achieved, in regards to the ideas that existed up until that moment.

Conversely, Mahoney & Drineas (2009) identified the singular vectors of the SVD as the PCs' interpretation problem and proposed another type of matrix factorization known as CUR Decomposition (Mahoney & Drineas, 2009; Mahoney, Maggioni, & Drineas, 2008; Bodor, Csabai, Mahoney, & Solymosi, 2012): The goal of CUR Decomposition is to give a better interpretation of the matrix decomposition by means of relevant variable selection in the data matrix, in a way that yields a simplified structure. Its origins come from analysis in genetics. One example is the one showed in Mahoney & Drineas (2009), in which cancer microarrays are highlighted with the purpose of recognizing, based on 5000 variables, genetic patterns in patients with soft tissue tumors analyzed with cDNA microarrays.

The CUR theoretical basis comes from the SVD of the matrix of interest, in order to make a new factorization by selecting columns and rows from the original matrix. This is basically a low-rank approximation to the original matrix expressed in a small number of rows and/or columns, which are easier to interpret than the singular vectors of the SVD. The main advantage of CUR Decomposition over SVD is that the original data matrix can be expressed with a reduced number of rows and columns instead of obtaining factorial axes that result of a linear combination of all the original variables.

The objective of this article is to show an alternative to variable selection (columns) or individuals (rows) to the ones developed by Mahoney & Drineas (2009): The idea proposed consists of adjusting the probability distributions to the leverage scores and selecting the best columns and rows that minimize the reconstruction error of the matrix approximation $\|A - CUR\|$. Additionally, another objective of the article is to present the Dynamic CUR algorithm, developed in C#.NET and R languages.

In the following section, the theoretical basis of CUR Decomposition is presented, giving emphasis to the leverage influence factors' method of calculation, to the columns and rows selection criteria, and to the available R library for its practical application. Furthermore, a short revision of some extreme value probability distributions and the theoretical basis of Dynamic CUR is presented. Lastly, the details regarding the use of the Dynamic CUR program, its application on a real practical case and some conclusions are presented.

2 . $\boldsymbol{CUR}$ DECOMPOSITION

Given a real matrix A of dimension MxN which contains information of N variables for M individuals or analysis objects. Given $A \in \mathbb{R}^{MxN}$, the CUR Decomposition of A tries to approximate this matrix by means of the product of three matrixes $C,\ U$ and R, where C is a matrix formed by a subset of c columns from the original matrix, R contains a subset of r rows of A, and the U matrix is defined as $U = C^{+}AR^{+}$ (where C^{+} and R^{+} are the pseudoinverse matrixes of C and R respectively) in a way that the CUR matrix product approximates A satisfactorily): Thus,

$$A_{MxN} \approx C_{Mxc}U_{cxr}R_{rxN}$$

The rows and columns that make up the R and C matrixes respectively are selected based on the influence level (or leverage) of the observations and variables of the global model. For each row (individuals) and/or column (variables) of the data matrix, these factors of importance are defined based on the Singular Value Decomposition of the original matrix. Given the considered matrix A of rank $r \leq \min\,(M, N)$, the Singular Value Decomposition (SVD) of A is defined as the matrix approximation

$$A_{MxN} = U_{Mxr}\Sigma_{rxr}V^{T}{}_{rxN}$$

where $U = [u_1, \dots, u_r]$ and $V = [v_1, \dots, v_r]$ are the orthonormal matrixes whose column vectors are the singular vectors to the left and right respectively, $U^TU = I\ and\ V^TV = I$, and Σ the diagonal matrix that stores the singular values of X, conveniently expressed in a way that: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$. From the point of view of the SVD, for $k < r$, the low rank approximation A_k is defined as:

$$A_k = U_{mxk}\Sigma_{kxk}V^{T}{}_{kxn}$$

Furthermore, given A_k the k rank approximation to the A matrix, the best starting matrix approximation will be that which minimizes the squared Frobenius norm between A and A_k):

$$\|A - A_k\|_F^2$$

3. THE STATISTICAL LEVERAGE AND THE BEST DECOMPOSITION OF THE $\boldsymbol{A_{mxn} \approx CUR}$ MATRIX

The selection criteria for the relevant columns and rows of the original matrix is crucial in the CUR Decomposition. In both cases, a leverage score which takes into consideration the importance of each column or row of A_{MxN} is calculated. Said influence factors' calculation is done as follows. Given v_n $(\mathrm{n} = 1, \dots, \mathrm{N})$ the singular vectors to the right, obtained in the SVD of the A matrix. The leverage score of each variable is defined as:

$$\pi_n = \frac{1}{k}\sum_{i=1}^{k}(v_{in})^2 \tag{8}$$

where k is the number of new axes in the dimensionality reduction. The leverage scores calculation for each of the matrix's individuals, when selecting the most influential, is done respectively from the u_i vectors of the U matrix obtained in the SVD.

Additionally, using this normalization, it is simple to demonstrate that $\pi_n \geq 0$ and that $\sum_{j=1}^{N}\pi_j = 1$, therefore the leverage scores create a probability distribution over the N original columns. This last point that Mahoney & Drineas (2009) make is crucial in this article, due to the fact that the leverages have special properties that allow them to be fitted to a probability distribution, being this one of the points not approached by the CUR decomposition authors.

Given that the leverage scores are a quadratic combination of the factor loadings of the k retained PCs, it is possible to see the value π_j takes as the importance that the column or row has on the original data matrix.

Once the influence factors for each variable and/or individual of the original matrix is defined, the columns and/or rows of the matrix are selected according to different criteria that take into consideration the distribution established by the leverages. It becomes clear that, unlike what happened in the SVD, the CUR approximation of a matrix is not unique.
Multiple algorithms that differ from each other in the obtained error bounds and in the criteria for selecting rows and columns exist. For the selection of the C and R matrixes, Mahoney & Drineas (2009) proposed five different selection criteria, based on the obtained leverages, always taking into consideration that the error in the $\|A - CUR\|$ approximation has to be minimal, according to the c columns and r rows selected.

rCUR algorithm 1.3 R library for the selection of columns and rows

The algorithm for the CUR decomposition calculation is implemented in the rCUR library (Bodor, Csabai, Mahoney & Solymosi, 2012) of the R software):

Column and row selection criteria according to rCUR

Once the leverage scores have been calculated, the rCUR algorithm selects the columns and rows with one of the following selection criteria: a) "random", b) "exact.num.random", c) "top.scores", d) "highest. ranks" y e) "ortho.top.scores".

Column and row selecton criteria	**Criteria description**
random	With this method, the probability of selection of each column is defined as the minimum between 1 and $c\pi$. This p_j probability is then compared to the probability obtained from a uniform distribution with parameter m. This probability distribution is defined in R as runif(m), which provides m probabilities with values between 0 and 1. The selection criterion is if $p_j \geq uniform\ probability$ then that column is selected, for as many columns as indicated by parameter c. When the number of columns is reached, the algorithm stops, even if there are more $p_j \geq uniform\ probability$. The number of columns, even if set in parameter c, could differ from the one actually obtained, which means that the final size of selected columns and rows is also random.
exact.num.random	With this method a leverage scores vector which considers the first k components of the V matrix is created. The difference between the leverage score multiplied by c and the probability obtained under the uniform distribution with parameter n is calculated. This difference is then ordered from highest to lowest and the exact number of variables corresponding to the top c variables are selected.
top.scores	With this method, the highest leverages scoring c columns are selected.
highest.ranks	With this method, ranks are assigned to the leverage scores and then the ones with the highest rank selected.
ortho.top.scores	With this method, columns and rows are selected, considering the combination of the leverage scores with the orthogonality of the subspace created by the columns and/or rows previously selected.

Dynamic CUR Proposal

In order to select columns and rows and to create their respective C and R matrixes, while also guaranteeing a minimization in the $\|A - CUR\|$ difference, the following proposal has been developed. The proposal consists of the selection of the best columns and rows, according to the leverage scores probability distribution on the k components used for their calculation. Like Mahoney & Drineas (2009) point out, the leverage scores have a probability distribution over the n columns, given that $\pi_j \geq 0$ and that $\sum_{j=1}^{N} \pi_j = 1$ Our proposal corresponds to a new column and row selection criterion (keeping the leverage scores calculation method defined by Mahoney & Drineas (2009), which consists of selecting the best leverage scores according to a dynamic criterion that combines two aspects:

a) the definition of a probability or area under the curve (0.20, 0.10, 0.05, 0.01 or 0.005) in order to estimate the number of columns (variables) and/or rows (individuals) with statistical criteria and
b) fitting the empirical distribution of the leverage scores to an existing theoretical distribution for each of the k components with which the leverage scores $(k < N)$, can be calculated, choosing those columns and rows where that probability is accumulated.

This dynamizes the calculation of $\|A - CUR\|$ and allows the establishment of the better combination for a given probability. In most cases, it has been observed that the leverage score distribution follows a positive-skew distribution with a long tail to the right, which indicates that very high scores exist and are associated with high leverage scores. Even though the empirical distributions of the leverage scores can take a particular distribution, it is also possible to find a known theoretical distribution to which they can be fitted. (Figure 1) In those situations, it is possible to utilize known distributions as extreme value distributions; also known as generalized extreme value distributions, among which the Fréchet, Gumbel, and Weibull distributions can be found. Our proposal is based on the extreme value theory, as evidence that the highest leverage scores are associated to the columns (variables) and rows (individuals) that take in more information from the data matrix.

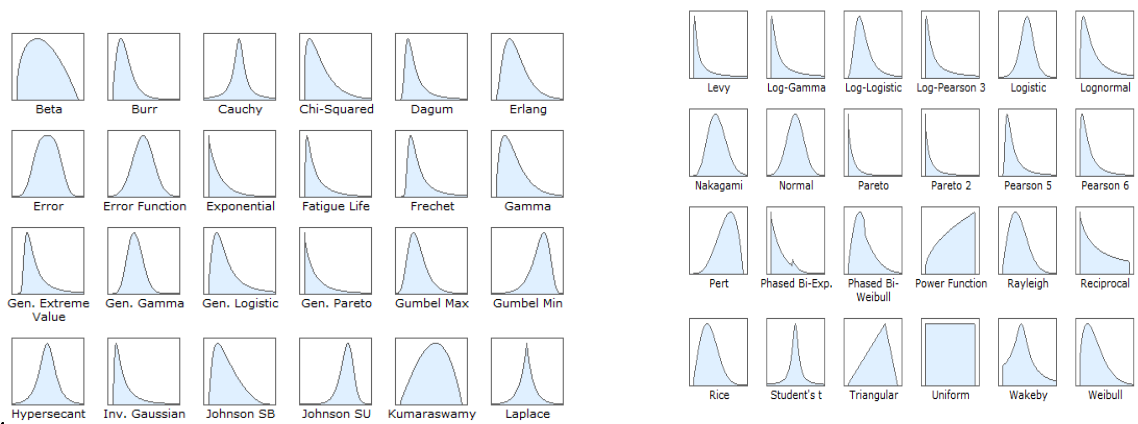

Figure 1: Different probability distributions that the leverage scores can take

Image taken from: EasyFit Professional V5.6 MathWare Technologies. http://www.mathwave.com

rCUR *versus DYNAMIC CUR*

The advantages presented by the Dynamic CUR proposal compared to the rCUR algorithm are:

1. Dynamically calculates $\|A - CUR\|$ and $\frac{\|A-CUR\|}{\|A\|}$, which correspond to the absolute error and relative error, going through all the probability distributions of the leverage scores generated by each of the k components, finding the adequate combination between k, the number of columns (variables) and the number of rows (individuals) that minimizes error. This selection can be made through the use of the observed distribution or by fitting to a theoretical distribution.
2. Allows the finding of the best columns and rows according to the calculation of the leverage scores, taking into consideration from 1 to k principal components. The investigator can even go through all the possible principal components up until $k = n - 1$ to determine which k which minimizes this error.
3. Eases the visualization of variables that arise as relevant and their relative importance within the selected set and the way in which they alternate according to how the different distributions are calculated until the results are stabilized.
4. Allows for the visualization of the relative error's progression and the observed histograms of the leverage scores as the search for the parameters which minimize error goes on.

CUR Dynamic Software

The CUR Dynamic software was implemented using C#.NET as the programming language and R. It is a two-tier client/server application. This architecture is key for the software's proper functioning, as it allows for dynamism in the processing of CUR decompositions. The software (client) is programmed in C#.NET, which is tasked with sending the requests to R for the calculation of the SVD and CUR decompositions. The use of the statistical software R, which stands at the server side, is crucial for successfully completing said decompositions. For these calculations, the libraries "Matrix", "rCUR" and "MASS" and the "svd" R function where used.
Communication between C#.NET and R was possible with the use of the middleware R.NET, which was installed in the corresponding code project as a NuGet package. The data matrix is stored in files with the "txt" extension and is read from C#.NET code in execution time using R as an intermediary. This allowed for the reduction of processing time in relation to the data size, especially for large-sized matrixes. When the Dynamic CUR software is opened, a menu pops up. This menu includes: a) Load data, b) dCUR parameters, c) Execute Dynamic CUR, d) Error Estimation (Figure 2): The procedure is initialized with the following steps:

1. The data is loaded, indicating the number of rows/individuals (objects) and variables/columns.
2. The variables of interest are selected and moved to selected variables.
3. The SVD (Singular Value Decomposition) calculation is requested in order to obtain the accumulated variance percentage from the n possible PCA. The investigator has to choose a starting k of interest. They then have the option of selecting the variables for a unique k, from 1 to k or from 1 to $k = N - 1$.
4. The area under the curve for the estimation of the number of variables (columns) and objects (rows) is determined. This value is set by the investigator.
5. Finally, the type of fitness has to be indicated. The options are through the observed distribution (histogram) or through the theoretical distribution fitted to the data.

The obtained results can be visualized as the program executes, making it a dynamic analysis.

6. Relative error graph according to each k
7. Observed distribution (histogram) and fitted distribution graph
8. List of relevant variables from the data matrix
9. Leverage scores according to selected variables
10. Absolute and relative error

Dynamic CUR as a Learning Analytics and Educational Data Mining (EDM) tool

Learning analytics is the measurement, analysis, and reporting of data about learners and their contexts. Its focus is to understand and optimize the learning process and the environment which surrounds it. On the other hand, educational data mining (EDM) seeks to develop and improve methods for exploring this data, in order to discover new insights about how people learn in the context of such settings.
An example is presented in which the relevant variables have to be selected from a set of 200 variables that have been investigated to explain academic performance of distance university students in Costa Rica. These variables are related with psychological factors, pedagogical factors, learning strategies, institution's academic and administrative management, the way in which the student manages their studies, student's health, tutor performance, family and class climate, student's relationship with their peers, employment and economic situation**.** Table 1 shows the number of variables/dimensions that each factor includes.
With the parameters set for k=50, (same number of components that accumulate 99%), superior area under the curve of 0.10 for rows (r=64) and columns (c=30), the best solution is reached with $k = 26$ with an error of 1.15%. The selected variables according to their relative importance considering the leverage scores can be observed in Figure 2.
According to the information provided by each of the variables to the global matrix, the four most important variables are: total physical activity done by the student, weekly worked hours, socioeconomic status and physical self-concept, followed by weekly study hours and graphical study techniques used during the learning process. Subsequently, the Neuroticism dimension from the "Big Five personality traits", group techniques, academic self-concept and academic conducts self-efficacy regarding communication. The active learning style from the CHAEA scale (Alonso, Gallego, & P Honey, 2018; Villegas, Sánchez, Sánchez-

García, Galindo-Villardón, 2018), the number of students per classroom, the creativity in design and art dimension and tutor performance are next. Following these, with lesser importance, the empowerment climate dimension, the reflexive CHAEA and the deep approach from the ASSIST scale (Entwistle, McCune, & Tait, 2013; Brown, White, Wakeling, & Naiker, 2015) can be found. Lastly, the agreeableness dimension from the personality traits, the surface approach from the ASSIST scale and the theoretical learning style from the CHAEA scale are found.

Table 1. Number of variables included in the analysis according to factor

Factor	**N. of variables**
Student's psychological factor	67
Student's pedagogical factor	33
Learning strategies factor	24
Institution's academic and administrative management factor	21
Student's academic management	15
Learnings evaluation factor	10
Health factor	9
Teacher's performance factor	5
Family social climate factor	3
Student's relationship	3
Scholarship factor	3
Student's academic achievement factor	3
Administrative management factor	2
Total	**200**
Work factor and student's economic factor	2

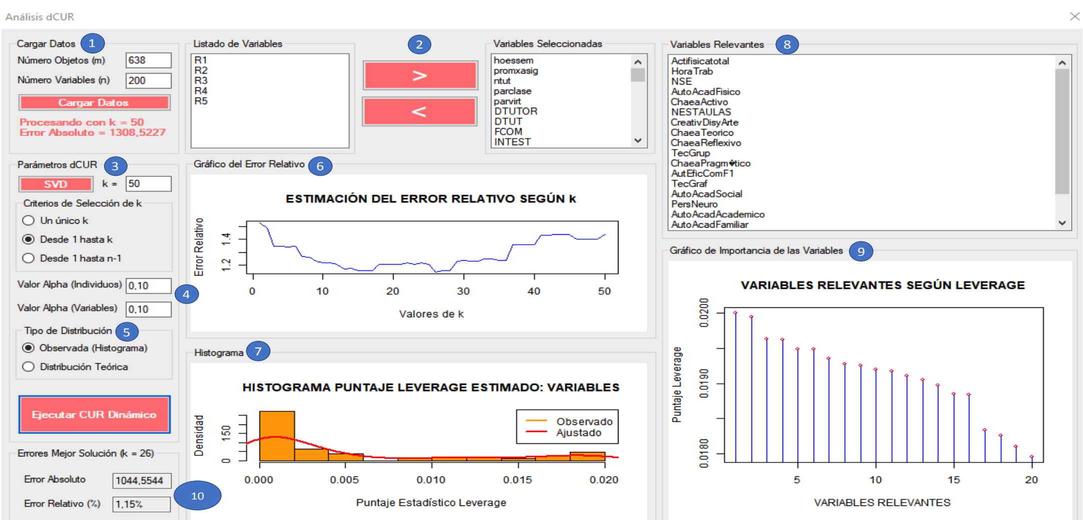

Figure 2: Dynamic CUR parameter panel

4. CONCLUSION

The options offered by the rCUR algorithm (Bodor, Csabai, Mahoney, & Solymosi, 2012) for individual and/or variable selection have some limitations. For example, the random method of the rCUR algorithm selects a different number of individuals and/or variables with each run and this selection may not include the best individuals and/or variables that minimize the approximation difference, while the Dynamic CUR sets in place the number of individuals and/or variables when setting the superior alpha probability of the established probability distribution. Once this probability is entered, the set number of individuals and/or variables of interest is estimated.

Furthermore, even though the exact.num.random method selects the exact requested number of individuals and/or variables, it doesn't necessarily select the best individuals and/or variables, given that it incorporates a random element in its configuration. Dynamic CUR selects the same number of set individuals and/or columns, albeit the best selection according to the highest leverage scores.
The highest.ranks method, apart from being the most complex from a mathematical point of view, presents the inconvenient that it degrades the measurement level of the leverage score by assigning rank as selection criteria, and ultimately ends up comparing ranks between different options in order to select individuals and/or variables, instead of using the natural measurement value and unit of the leverage scores. Dynamic CUR utilizes the highest natural values among the leverage scores.
The top.scores and ortho.top.scores are the methods that select the best individuals and/or variables because they restrict selection to the highest scores according to each case. Nevertheless, it has the limitation that they only assess leverage scores for a specific principal component, giving the perspective that it is a unique solution. Dynamic CUR has three search options for optimizing individuals and/or variable selection (for one k, for values from 1 to k and for values from 1 to $k = N - 1$):
With this situation in mind, possibilities presented with the Dynamic CUR program are important because it offers several alternatives that dynamize the configuration of selection criteria which take into consideration different numbers of individuals and/or variables, several fitness criteria and the analysis of the approximation error until the lowest error according to the combination of specified options is achieved.
There is a gap left in the rCUR algorithm since it does not take into consideration the probability distribution of the leverage scores as an alternative to selecting the best individuals and/or variables. This is addressed in our proposal due to the fact that the scores are fitted to their probability distribution in a dynamic way, in concordance with the desired k components.

REFERENCES

[1] ALONSO, C., GALLEGO, D., & P HONEY, Y. (2018): Cuestionario Honey Alonso de Estilos de Aprendizaje CHAEA. **Retrieved from https://www.researchgate.net/publication/265265933_** CUESTIONARIO_HONEY-ALONSO_DE_ESTILOS_DE_APRENDIZAJE_CHAEA
[2] BODOR, A., CSABAI, I., MAHONEY, M. W., & SOLYMOSI, N. (2012): rCUR: an R package for CUR matrix decomposition. **BMC Bioinformatics**, 13, 103.
[3] BROWN, S., WHITE, S., WAKELING, L., & NAIKER, M. (2015): Approaches and Study Skills Inventory for Students (ASSIST) in an Introductory Course in Chemistry. Approaches and Study Skills Inventory for Students (ASSIST) in an Introductory Course in Chemistry. **Journal Journal of University Teaching & Learning Practice**, 12.
[4] CADIMA, J., & JOLLIFFE, I. T. (1995): Loadings and correlations in the interpretation of principal components. **Journal of Applied Statistics**, 22, 203–214.
[5] ECKART, C., & YOUNG, G. (1936): The approximation of one matrix by another of lower rank. **Psychometrika**, 1. Retrieved from https://link.springer.com/article/10.1007/BF02288367
[6] ENTWISTLE, N., MCCUNE, V., & TAIT, H. (2013): Approaches and Study Skills Inventory for Students (ASSIST): **Report of the Development and Use of the Inventories**. Retrieved from http://www.etl.tla.ed.ac.uk/questionnaires/ASSIST.pdf
[7] HAUSMAN, R. (1982): Constrained multivariate analysis. **Optimization in Statistics**, 137–151.
[8] HOTELLING, H. (1933): Analysis of a Complex of Statistical Variables Into Principal Components. **Journal of Educational Psychology**, 24, 417–441.
[9] JOLLIFFE, I. T. (2002): Principal Component Analysis. (Springer, Ed.) **Encyclopedia of Statistics in Behavioral Scienc**e (2nd ed., Vol. 30): Springer, New York, NY.
[10] JOLLIFFE, I. T., TRENDAFILOV, N. T., & UDDIN, M. (2003): A Modified Principal Component Technique Based on the LASSO. **Journal of Computational and Graphical Statistics**, 12, 531–547.
[11] KAISER, H. F. (1958): The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, 23, 187–200.
[12] MAHONEY, M. W., & DRINEAS, P. (2009): {CUR} matrix decompositions for improved data analysis. **Proceedings of the National Academy of Sciences of the United States of America**, 106, 697–702.
[13] MAHONEY, M. W., MAGGIONI, M., & DRINEAS, P. (2008): Tensor-CUR Decompositions for Tensor-Based Data. **SIAM Journal on Matrix Analysis and Applications**, 30, 957–987.

[14] MCCABE, G. (1982): Principal Variables. Retrieved from **http://www.stat.purdue.edu/research/technical_reports/pdfs/1982/tr82-03.pdf**
[15] PEARSON, K. (1901): On lines and planes of closest fit to systems of points in space. **Philosophical Magazine** Series 6, 2, 559–572.
[16] SMILDE, A. K., BRO, R., & GELADI, P. (2004): **Multi-way analysis with applications in the chemical sciences.**
[17] TEAM, R. C. (2017): The R Project for Statistical Computing. **Retrieved from https://www.r-project.org/.**
[18] TIBSHIRANI, R. (1996): Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Statistical Society Series B**, 58, 267–288.
[19] VILLEGAS, G., SÁNCHEZ, M., SÁNCHEZ-GARCÍA, A.B., GALINDO, M. P. (2018): Correlación entre cuatro escalas de estilos de aprendizaje y rendimiento académico. **In Redine (Ed.), Innovative strategies for higher Education in Spain** (Primera ed, pp. 151–174): Einhoven, The Netherlands: Adaya press H.H. van Brabantplein.
[20] VINES, S. K. (2000): Simple principal components. **Journal of the Royal Statistical Society. Series C: Applied Statistics**, 49, 441–451.
[21] Zou, H., & Hastie, T. (2005): Regularization and Variable Selection via the Elastic Net. **Journal of the Royal Statistical Society Series B (Statistical Methodology),** 67, 301–320.
[22] ZOU, H., HASTIE, T., & TIBSHIRANI, R. (2006): Sparse principal component analysis. **Journal of Computational and Graphical Statistics**, 15, 265–286.