2020 IEEE Industry Applications Society Annual Meeting | 978-1-7281-7192-0/20/$31.00 ©2020 IEEE | DOI: 10.1109/IAS44978.2020.9334900

Big Data Compression in Smart Grids via Optimal Singular Value Decomposition

Seyed Naser Hashemipour, Jamshid Aghaei,
Abdullah Kavousi-fard, Taher Niknam
Shiraz University of Technology
Shiraz, Iran
{n.hashemipour; aghaei; kavousi; niknam}@sutech.ac.ir

Ladan Salimi
Razi University
Kermanshah, Iran
salami.ladan@stu.razi.ac.ir

Pedro Crespo del Granado
Norwegian University of Science and Technology
Trondheim, Norway
pedro@ntnu.no

Miadreza Shafie-khah
School of Technology and Innovations
University of Vaasa
Vaasa 65200, Finland
mshafiek@univaasa.fi

Fei Wang
Department of Electrical Engineering
North China Electric Power University
Baoding 071003, China
feiwang@ncepu.edu.cn

João P. S. Catalão
Faculty of Engineering of the University of Porto and INESC TEC
Porto, Portugal
catalao@fe.up.pt

***Abstract*—The smart grid is a fully automatic delivery grid for electricity power with a two-way reliable flow of electricity and information among different equipment on the grid. With the rapid development of smart grids, smart meters and sensors are used to monitor the system and provide a wide reporting which produce a huge amount of data in various part of the grid. To logical manage this trouble, the presented paper proposes a new lossy data compression approach for big data compression. In the proposed method, at the first step, the optimal singular value decomposition (OSVD) is applied to a matrix that achieves the optimal number of singular values to the sending process and the other ones will be neglected. This goal is done due to the quality of retrieved data and the rate of compression ratio. In the presented scheme, to implementation of the optimization framework, various intelligent optimization methods are used to determine the number of optimal values in the elimination stage. The efficiency and capabilities of the proposed method are examined using the experimental dataset of several residential microgrid consumers and market dataset. Simulation results show the high performance and efficiency of the proposed model in smart grids with big data.**

***Keywords*—Big data, Data compression, Smart Grid, Optimization, Singular value decomposition.**

Nomenclature

A	Original data
$\overline{A}$	Reconstructed data
C_r	Compression ratio
D_r	Elements of remained data
D_d	Elements of deleted data
D_o	Elements of original data
GA_m	Genetic Algorithm with mutation
m	Rows of the original matrix
n	columns of the original matrix
N_t	the threshold of Norm 2 for comparison of original and retrieved data
$N_2(Z)$	Norm 2 of matrix Z
p	The number of deleted singular values
S_d	Number of deleted singular values
α	Weight coefficient for N_t

I. Introduction

A. Data in Smart Grids

Smart grid is an intelligent electricity grid that optimizes the generation, distribution, and consumption of electricity through the introduction of Information and Communication Technologies on the electricity grid which includes the smart meters and various sensors in different parts. The measurement and monitoring instruments to gathering the information in the transmission system and medium-voltage level distribution system are managed by supervisory control and data acquisition (SCADA) and wide area monitoring system (WAMS). Similarly, in the level of consumers, advanced metering infrastructure (AMI) and automatic meter reading (AMR) systems are employed for data gathering in the smart grid. Phasor measurement units (PMUs) are among the other units used in the smart grid to measure the required information and send it through a communications platform.

Fig. 1 shows the general structure of the WAMS system in the smart grid. Information for each PMU is transmitted through public switched telephone networks, fiber optic cables, low altitude satellites, power line carriers (PLCs) or microwave links. As a result, a huge amount of multi-source varied data is stored in the smart grids. These data if exploited properly can reveal much information about the customers and generating units and improve the power quality and smart grid efficiency. Unfortunately, a challenge in this way is the huge volume of transmitted information and the limited bandwidth available for the data transfer. To deal with this issue, the employment of data compression techniques for reducing the size of data transfer can bring great benefit to the smart grid.

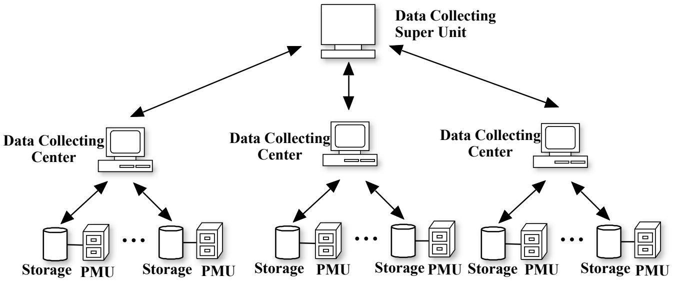

Figure 1. WAMS structure in a smart grid

In compression schemes, the initial goal is to earn data compression and simultaneously preserving the useful characteristics of original data [1]. Data compression is broadly classified into two categories, namely lossless (original data can be recovered perfectly from the compressed data) and lossy (the original data cannot be recovered from the compressed data perfectly and there are losses) [2]. Various researchers are actively engaged to derive efficient methods of data compression using the latest techniques. The following subsection reviews the important techniques that have been used commonly in data compression.

B. Literature Review

Different algorithms have been developed for smart grid data compression. These include frequency domain transformation such as wavelet decomposition (WD)[3-6] or discrete cosine transform(DCT)[6], fuzzy-based methods[12], compressed sensing theory [11], SVD-based approaches [13-14] and so on.

In WD-based data compression, Khan *et al* in [3] proposed a new method for data compression and de-noising in smart grids based on wavelet transform. This scheme uses the wavelet packet decomposition (WPD) that is more accurate than wavelet decomposition. In this method WD tree has been converted to fully binary tree using a cost function and the best tree has been selected from a number of WPD bases. This work provides an acceptable compression ratio and a good denoising tool. Also, it uses a threshold for reconstruction of the signal.

Also, Khan *et al.* [4] presented an embedded zerotree wavelet transform (EZWT)-based approach which has been used to depress the signal noise in smart grid. Since EZWT does not require tables, codebooks and etc. for extraction of original data, it is known as a simple method. In this paper, the proposed method has been evaluated using PMU and power system data and has been used to compress and elimination of noise in data.

Besides, In [5], Ning et al suggested a lossy method for data compression in a smart grid that is based on wavelet transform (WT) and multi-resolution analysis (MRA). Due to WT-MRA features, this packet is very suitable for compression and de-noising of power system signals. In decomposition process, Daubechies filter is considered as the mother signal. Experimental results illustrated that the WT-MRI con not only compresses power system signals and it must be combined with another algorithm. So, the presented method can only solve the white noise of considered signals.

Moreover, on another research [6] the wavelet transformation has been deployed for signal compression. In this method, after performing wavelet transformation, dynamic bit allocation is defined for compression. To the definition of bit allocation, the spectral shape has been estimated by several methods in which one is the neural shape estimator (NSE). Spectral Shape estimation is necessary to eliminate data redundancy and implementation of the entropy coding. The results demonstrated that NSE is successful than the other estimators which provide an acceptable ratio for compression of waveforms. In another study [7], to compression of PMU data, a general compression algorithm is presented that uses intrinsic correlation using the extraction of temporal and spatial redundancies. The mentioned approach is designed in two stages which in first stage, the principal component analysis has been defined and in the second stage, a discrete cosine transform has been used. Also, the compression parameters have been adjusted using efficient statistical techniques. The results show this approach is general, and so can be employed on a phasor data concentrators (PDC) fed from any number of PMUs.

As another application of compression in the electricity dataset, [8] provided a novel method based on Deep Stacked Auto-Encoders, which compresses the load data on the user side. The fuzzy-based methods have been insured by some researchers as another way to data compression. For instance, in [9] a fuzzy-based approach has been introduced to save the required memory and bandwidth, which reduces the computational burden for smart grids' data analysis. In this regard, data redundancy has been detected by fuzzy-based domain transformation. So, the significant amount of data has been eliminated and this scheme provides an acceptable ratio for data compression. some works using compressed sensing theory for smart grid applications emerged lately, such as those in [10] which provided the compression technique for electricity datasets. The suggested algorithm makes the sparsity condition of original data. At the decoder side, an iterative threshold algorithm has been employed to reconstruct the compressed bitstream. The experimental results illustrated that the proposed method has good performance for compression and decompression of the original data.

The use of artificial intelligence such as neural networks [11] and intelligent measurement techniques [12] have also been investigated in search of compression methods. Barrosa *et al.* [11] suggested a method for the compression of electrical power signals using genetic algorithm (GA) and an artificial neural network (ANN). The GA is used to select the best samples of the signal and the ANN is used to compression of remained samples and the reconstruction of the signal. This method proposed the compression ratio from 2.5% to 10% for a recorder installed in a 230-kV electrical power system. In [12], an approach has been suggested for compression of electricity load data based on intelligent measurement. In this method, considering the size of the input, a foundation has been provided which is suitable for low-complexity coding and decoding of smart grid transmissions. Linear algebra-based techniques are considered as another tool for data compression, which has been investigated in the search to propose lossy compression method.

For example, In [13], a new lossy method based on SVD decomposition has been presented for distribution systems, which is useful for the compression of large data. This technique reduces a large number of data by rank reduction of the matrix which is done by SVD decomposition.

Actually, some singular values of the diagonal matrix are neglected by the try-and-error approach. Hence, an approximation of the original matrix is achieved and on the decoder side, data is accurately recovered. As another research, [14] introduced a novel lossy compression method based on the combination of SVD decomposition and wavelet difference reduction (WDR). At first, considering the visual quality of original images, the rank of the original matrix has been reduced using the try-and-error technique and then WDR has been applied to reduced data. In this method, the SVD decomposition has a good Peak Signal-to-Noise Ratio (PSNR) with a low compression ratio, accordingly, the WDR has been added to SVD decomposition. The use of compression methods in general, and lossy compression in particular, with electricity dataset, has been previously investigated. As a study in loss-less compression methods, [15] introduced various lossless compression methods such as Lempel–Ziv–Markov chain algorithm (LZMA) in smart grids. the presented scheme has been employed on a dataset that is achieved by continuous monitoring. In this technique, for the increment of the compression ratio, the differential encoding has been used as analytical modules compressors.

To the best of authors' knowledge, in the previous SVD-based researches, the rank reduction is based on the try-and-error approach. To solve this issue, an optimization framework has been presented in the presented paper. In this regard, considering the accuracy of reconstructed data and compression ratio, the optimal SVD decomposition to rank reduction is suggested. In the optimization framework, The inverse of compression ratio is considered as the objective function in a minimization problem and the proximity between the original and the reduced matrix is considered as the problems' constraint. Accordingly, this paper focuses on optimal SVD with a proper compression ratio, which reduces the computational burden. The proposed algorithm is simple and efficient for the implementation of the compressor.

The remainder of the paper is organized as follows: In section II, A summary of data compression in smart grids is presented. In section III, the overview of the used methodologies in the proposed compression technique is discussed in which details of the SVD and optimization framework are addressed. Then in section IV, the proposed lossy compression is introduced. The experimental results are expressed in V and the discussions are given in Section VI followed by the conclusions in the last section.

II. Compression in Smart Grids

The generation, transmission, and distribution of the power in smart power systems are deeply impressed by data analyzing. Therefore, a considerable increase in data exchange and in the required memory is likely to occur [16] and the required data storage and bandwidth of the communication links in the smart grids have a growing trend. Besides, to obtain accurate and real-time running status information of the smart grid, the frequency of sampling should be increased. Accordingly, the importance of data compression in the smart grid will be more highlighted. The proposed compression method is presented in the following. This method can be employed effectively in different points of the grid where the volume of the sent and received data is high.

A. The SVD Decomposition

The SVD is a computational tool for approximating a matrix by three other matrices. Indeed, it decomposes matrix A into U, V, and Σ. Let m and n be arbitrary and A be a $m\times n$ matrix. A singular value decomposition of A is a factorization as can be seen in (1).

$$A = U\Sigma V^T \qquad (1)$$

where U is a $m\times m$ real or complex unitary matrix, Σ is a $m\times n$ rectangular diagonal matrix with non-negative real numbers on the diagonal, and V is a $n\times n$ real or complex unitary matrix. The diagonal entries σ_i of Σ are known as the singular values of A. briefly [17]:

- U: is $m \times m$ unitary (the left singular vectors of A)
- V: is $n \times n$ unitary (the right singular vectors of A)
- Σ: is $m \times n$ diagonal (the singular values of A)

where,

$$\Sigma_{(m\times n)} = diag(\sigma_1, \,...\,, \sigma_n) \;\; with\, \sigma_1 \geq \sigma_2 \geq ... \geq \sigma_n \;\geq\; 0$$

In Fig. 2, the SVD decomposition on a matrix has been shown. As already mentioned, Σ is a diagonal matrix whose elements on its original diameter are singular values that are placed in descending order. Each singular value is involved in the retrieving process of the original matrix. In other words, equation (1) can be rewritten in the form of equation (2) [16].

$$A = \sum_{i=1}^{m} u_i \sigma_i v_i^T \qquad (2)$$

where u_i and v_i are the left and right singular vectors of the matrix A, respectively and σ_i is the i^{th} singular value. As can be seen, the smaller singular values play a smaller role in the building of the original data. Thus, the low-rank matrix approximation can be used by the elimination of smaller values and the original information can be retrieved with appropriate approximation as can be seen in (3a). According to (3b), Σ is decomposed to a submatrix including the important singular values ($\bar{\Sigma}$) and three non-important submatrices which are approximated by zero matrices with the same dimension.

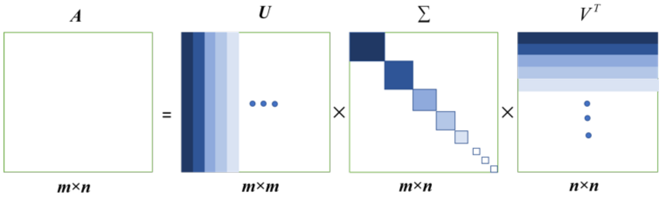

Figure 2. The SVD form of a matrix

$$\overline{A} = \overline{U}\overline{\Sigma}\overline{V}^{T} \tag{3a}$$

$$\Sigma = \begin{bmatrix} \overline{\Sigma}_{(n-p)*(n-p)} & 0_{(n-p)*p} \\ 0_{(m-n+p)*(n-p)} & 0_{(m-n+p)*p} \end{bmatrix}$$
$$U = [\overline{U}_{m*(n-p)} \quad \overline{\overline{U}}_{m*m-(n+p)}] \tag{3b}$$
$$V = [\overline{V}_{n*(n-p)} \quad \overline{\overline{V}}_{n*p}]$$

As can be seen in (4), $\overline{A}$ is a low ranked approximation $m \times n$ matrix, the matrix $\bar{U}$ is $m\times$ *(n-p)*, $\bar{V}$ is $n\times$ *(n-p)*, and $\bar{\bar{\Sigma}}$ is *(n-p)* × *(n-p)*.

$$A_{m*n} = [\overline{U}_{m*(n-p)} \quad \overline{\overline{U}}_{m*m-(n+p)}] \times \begin{bmatrix} \overline{\Sigma}_{(n-p)*(n-p)} & 0_{(n-p)*p} \\ 0_{(m-n+p)*(n-p)} & 0_{(m-n+p)*p} \end{bmatrix} \times [\overline{V}_{n*(n-p)} \quad \overline{\overline{V}}_{n*p}]^{T} \tag{4}$$

Due to the fact that $\bar{U}$, $\bar{V}$ and $\bar{\bar{\Sigma}}$ can provide an acceptable approximation of the original data (*A*) and can be sent instead of *A*, their dimension specifies the compression ratio.

Therefore, the compression ratio is defined as (5).

$$C_r = \frac{m \times n}{(n-p)(m+n+1)} \tag{5}$$

The compression is done when the ratio is more than 1. Equations (6)-(8) provide a lower bound on the number of neglected singular values.

$$\frac{m \times n}{(n-p)(m+n+1)} > 1 \tag{6}$$

$$(n-p) < \frac{m \times n}{(m+n+1)} \tag{7}$$

$$p > n - \frac{m \times n}{(m+n+1)} \tag{8}$$

In the compression schemes, along with the compression ratio, the redundancy of data can also be determined by equation (9).

$$\%R = 1 - \frac{1}{C_r} \tag{9}$$

In this process, according to (4) we have:

$$C_r = \frac{D_o}{D_r} = \frac{D_r + D_d}{D_r} \Rightarrow C_r = 1 + \frac{D_d}{D_r} \Rightarrow D_d = D_r(C_r - 1) \tag{10}$$

To find the redundant data, both sides of (10) is divided by the original data, so:

$$R = \frac{D_d}{D_o} = \frac{D_r}{D_o} \times (C_r - 1) \overset{\frac{D_r}{D_o} = \frac{1}{C_r}}{\Rightarrow} R\% = \left(1 - \frac{1}{C_r}\right) \times 100 \tag{11}$$

Therefore, by (11), the percentage of duplicate and redundant data is calculated.

Another important issue refers to the calculation of the retrieved data precision in the decoding process. The mean square error (MSE) criteria in (12) has been used to check the accuracy and proximity between original and retrieved data.

$$MSE = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (A - \overline{A}) \tag{12}$$

B. *Optimization Framework*

Another consideration is the amount of rank reduction without damaging the original data. Accordingly, it is not allowed to remove any large number of singular values. So far, in the related researches such as [13], the number of eliminated singular values have empirically been determined by the try and error approach. However, in this paper, an optimization framework is proposed to obtain the optimal number of singular values to be ignored. Before presenting the optimal singular value decomposition in section III, the general form of an optimization problem that can be seen in (13) is reviewed.

$$\begin{aligned} &\min F(X) \\ &s.t \quad H(X) = 0 \qquad G(X) \le 0 \end{aligned} \tag{13}$$

Where *F* is the objective function, *X* is the set of decision variables, *H* is equality constraints and *G* is in-equality constraints [18-19]. Taking the feasible region of the problem into account, the aim of the minimization problem is to find the optimal point x^* which satisfies (14).

$$\forall x \in X \Rightarrow F(x^*) \le F(x) \tag{14}$$

For this purpose, evolutionary algorithms have been used to solve the formulated optimization problem [20-21].

III. Proposed Method

As already mentioned, data types with significant volumes should be transmitted to the different points of the smart grid, for example, information of various participants on the electricity market, weather data including solar, wind, humidity and temperature data and etc., which are very important for optimal, resilient and reliable operation. Also, the selected dataset normally is in the form of a matrix for sending it to the network operator. Generally, the ways of data transmission are divided into encoding and decoding processes. In fact, in the encoding phase, data processing is performed to prepare data. After this operation, the data is transmitted through the communication channel and will be recovered to the initial form in the decoding phase. This process is shown in Fig. 3.

In the proposed method, the original data matrix is decomposed into *U, V* and Σ by the SVD decomposition. Then, the optimal rank reduction is specified and three rank-reduced matrices will be ready for sending. At first, the optimization problem will be presented. The decision variable *x* in this problem is the number of singular values that will be eliminated. In fact, the number of eliminated singular values influences the compression ratio and data quality. Since the main goal of the problem is data compression, the inverse of the compression ratio can be considered as the objective function in (15).

$$F(x) = \frac{1}{C_r(x)} = \frac{(n-x)(m+n+1)}{m \times n} \tag{15}$$

By the minimization of (15), the compression ratio will be minimized.

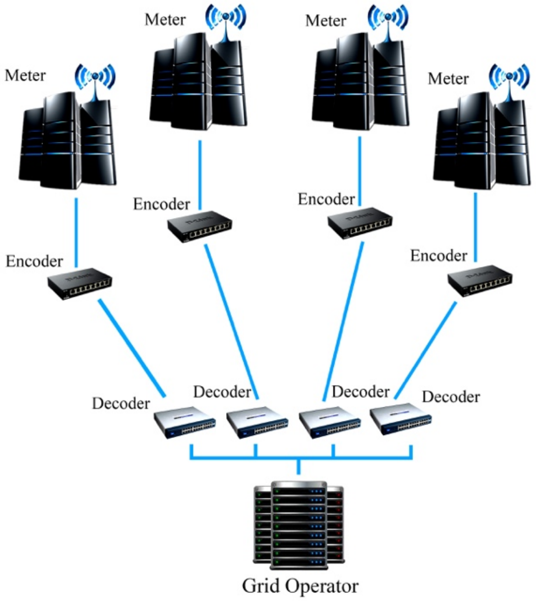

Figure 3. A data communication channel and transmission process

Alternatively, the objective function can be considered as the negative value of x as can be seen in (16). In fact, more compression can be gained by maximizing the number of neglected singular values.

$$F(x) = -x \tag{16}$$

As an important consideration, the compression is valuable if the information can be retrieved with acceptable accuracy. Therefore, a constraint is defined by (17) for the optimization problem.

$$N_2\left(A-\bar{A}\right) \le N_t \tag{17}$$

This constraint emphasizes the proximity of two matrices with norm 2 criteria. In other words, introducing (17) indicates the norm 2 of the difference between the original and recovered matrix is less than the N_t value. In this step, an important issue refers to determining the upper bound of the constraint (N_t). N_t can be determined based on the matrix A as can be seen in (18). In this way, we will have a vision at this threshold. In this regard, the upper bound of (17) is restricted to a fraction of norm$_2$ of the original matrix.

$$N_t = \alpha \times N_2(A) \tag{18}$$

To convert a constrained problem to an unconstrained one, the penalty coefficient is added to the cost function. Accordingly, the fitness function is calculated as (19).

$$fit = F(x) + K \times \max\left\{0 \;\;, N_2\left(A-\bar{A}\right) - N_t\right\} \tag{19}$$

Where K is the penalty coefficient. Thus, failing to satisfy the constraint (17), will impose a penalty on the fitness function and the solution has a lower probability for selection.

Another consideration is the range of decision variables. X is an integer number whose upper limit is equal to the number of singular values. As described in the previous section, according to (8), its lower limit is determined by (20).

$$x > n - \frac{m \times n}{(m+n+1)} \Rightarrow x_{\min} = round\left(n - \frac{m \times n}{(m+n+1)}\right) + 1 \tag{20}$$

To ensure more compression, the lower limit can be calculated as follows:

$$\frac{m \times n}{(n-x)(m+n+1)} > cr_0 \tag{21}$$

$$x > n - \frac{1}{cr_0}\frac{m \times n}{(m+n+1)} \tag{22}$$

Where cr_0 is the pre-specified minimum compression ratio. But, if this value is much greater than 1, then the problem may become infeasible. Accordingly, cr_0 should be selected rationally.

IV. Numerical Results and Discussions

The presented scheme has been implemented using MATLAB 2018a on the market data. All implementations have been done on a PC Intel core i7, processor 1.8 GHz, with RAM 8GB. The data information and the simulation results are expressed in the following.

A. Data

To evaluate the efficiency of the proposed method, two types of datasets have been tested. In case study 1, a dataset from Day-Ahead Energy Market of New England's wholesale electricity marketplace on January 1, 2018, for 315 participants has been tested. It should be noted that each agent must send the 5-segment (price-power curve) for 24-hour. Hence, the information matrix has 315 rows (equal to the number of market participants) and 240 columns (5 segments of power and 5 segments of the price for 24 hours). These datasets are available in [22]. The available data for case study 2 is electrical data over a single 24-hour period from 443 unique homes on February 4, 2011. This database refers to the different features of each home in a micro-grid for a 24-period. The size of the database is 1440 × 443 and is available in [23].

B. Case study 1

To implement the proposed method and to find the number of optimal singular values to be eliminated, the evolutionary algorithms like differential evolution (DE) [24], simulated annealing (SA) [25], teaching-learning based optimization (TLBO) [26], particle swarm optimization (PSO) [27], genetic algorithm with mutation (GA-M) and GA [28] have been used on the above mentioned market data. In Table I, the optimal solution (the mean value of the optimal number of eliminated singular values over 20 runs for each algorithm) has been shown with some statistical indexes to evaluate the quality of each algorithm.

Also, the run time of these algorithms is seen in Table I. it should be noted in this table the condition is almost similar for all algorithms. As it turns out, in Table I, in order to achieve a robust answer, the number of required iterations is observed in each algorithm.

TABLE I:
THE NUMNER OF RECOMMENDED SINGULAR VALUE FOR DELETION BY EVOLUTIONARY ALGORITHMS

	DE [24]	SA [25]	TLBO [26]	PSO [27]	GA-M [28]	GA [28]
Iteration	14	150	5	6	6	100
Mean	209	209	210	210	210	≈206
Standard deviation	0	0	0	0	0	4.51
Run time (second)	1.58	2.76	1.17	0.77	1.82	4.37

In fact, the first four algorithms with a fixed number of iterations have been converged to the specific solutions with the standard deviations of 0 for 20 runs. But in GA, despite the 100 number of iterations, still it is not converged to a unique solution, but the average of the answers is close to the optimal answers of the other algorithms. Besides, DE, TLBO, PSO, GA with mutation and SA have converged to the optimal solution. According to Table I, the PSO algorithm is the fastest algorithm among all to solve the problem.

In the following, the compression ratio and the percentage of redundant data for the achieved solutions have been calculated and addressed in Table II. As can be seen from this table, the compression ratio by the presented scheme has a high performance and offers a high compression ratio and also removes a huge redundant data. Accordingly, the proposed method can be used for data reduction with reasonable accuracy.

In decoding and restoration process, the precision of the recovered matrix is satisfactory and this matter has been investigated by using the MSE and norm$_2$ value of the difference between the original and recovered matrices in Table III. The number of the original matrix elements is 75600. According to the results of Table III, it is inferred that the restoration accuracy of the proposed method with various algorithms is acceptable. The following analysis is performed on the result of PSO algorithm. Since the accuracy of the data recovery is dependent on compression ratio, the optimization problem is solved with the different thresholds (N_t). In this case, in order to cover a wider range of N_t, α has been selected from 0.000001 to 0.01. According to Table IV, by the selection of a bigger N_t, the accuracy of retrieval information will be increased. For example, with threshold 0.17, a very high precision result is achieved by the elimination of 166 singular values while the minimum MSE of the retrieved information is 1.674e-06 proofing the high accuracy. Also, this accuracy presents that the compression ratio is equal to 1.84. In general, if N_t becomes larger, the problem is relaxed to have more compression and, as it is expected, the accuracy will be decreased. For example, in N_t = 1701, the compression ratio is 15.11, but the error in this situation is high and reaches about 174 of the largest element in the original matrix.

Fig. 4(a) illustrates the relationship between MSE criteria and the percentage of the redundant data based on the number of deleted singular values in the proposed lossy compression scheme for market data.

TABLE II:
CALCULATION OF COMPRESSION RATIO AND PERCENTAGE OF REDUNDANT DATA IN CASE STUDY 1

Algorithm	C_r	R%
DE [24]	4.3861	77.20
SA [25]	4.3861	77.20
TLBO [26]	4.5323	77.93
PSO [27]	4.5323	77.93
GA_m [28]	4.5323	77.93
GA [28]	3.999	74.99

TABLE III:
CALCULATING MSE AND N_2 CRITERIA BETWEEN ORIGINAL AND RECOVERED MATRIX

Algorithm	D_r	MSE	$N_2\left(A-\bar{A}\right)$
DE	17236	0.6007	113.4006
SA	17236	0.6007	113.4006
TLBO	16680	0.9152	154.1976
PSO	16680	0.9152	154.1976
GA_m	16680	0.9152	154.1976
GA	18904	0.1818	55.6600

According to Fig. 4(a), if the number of deleted singular values becomes greater, the percentage of the redundant data will also become more. Besides, If the amount of redundancy becomes higher, the data will become more compressed and a larger amount of data is deleted, so the magnitude of MSE error increases with increasing the redundancy as shown in Fig. 4(a). Fig. 4(b) is the presentation of the relationship between N_2 criteria and the compression ratio due to the number of the deleted singular values. As can be seen from Fig. 4(b), the behavior of the compression ratio and N_2 is in the same direction, since the high compression ratio implies the elimination of more values from the original data. As already mentioned, with a reduction in the amount of data and an increase in the number of deleted values, the accuracy will be reduced and will increase the error. Hence, with increasing the compression ratio, the amount of N_2 will be increased.

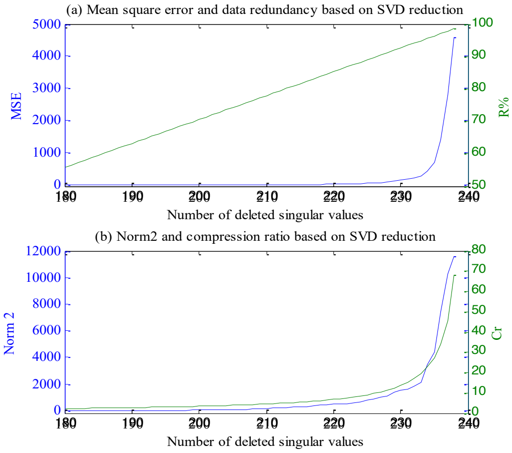

Figure. 4. The relationship of the neglected singular values with MSE (a) and Norm2 (b) for market dataset

C. Case study 2

To display the capabilities of the presented method on a different database, the information of electrical data for 24-hour has been considered as the original matrix. In this regard, evolutionary optimization methods have been used to find the number of optimal singular values.

Referring to Table V, each method gives a compression ratio and an acceptable percentage of the redundancy. The number of elements in the original matrix is 637920.

As the *Cr* and R% values show in Table VI, the robust TLBO with *Cr* = 2.015 and R% = 50.38 provide an acceptable compression rate and GA, GA-M and PSO don't reveal competence over other algorithms and they need more iterations. The mutation of GA in GA-M imposes a lot of calculations in each iteration. Since the considered database has a large number of elements, only TLBO, DE and SA algorithms can converge to the best answer and the other ones encounter computational errors.

Referring to Table VI, by increasing *Nt*, the compression ratio will be increased as well as the accuracy will be decreased. For instance, in *Nt*=17.593, MSE is better than the *Nt* =175.93. Generally, if *Nt* becomes larger, the problem is relaxed for more compression and the accuracy will be decreased. Also, due to the data type and their obtained singular values, the problem does not reach any solution for α=0.0001. Accordingly, by changing the *α* coefficient, some solution has been achieved, in which the amount of *α*= 0.001 gives an acceptable answer.

To evaluate the accuracy of the proposed method, Fig. 5 shows significant considerations. Concerning Fig. 5 (a) and (b), the increment of the number of the neglected values leads to the increment of the *Cr* and R%. Besides, with the increment of data redundancy and compression ratio, the accuracy criteria don't change over a wide range and these errors have been suddenly augmented. Hence the TLBO algorithm which has the closest solution to this point performs better than other ones.

TABLE IV.
CALCULATING OF VARIOUS ACCURACY WITH DIFFERENT N_t FOR MARKET DATA

α	N_t	*MSE*	S_d	C_r
0.000001	0.17	1.674e-06	166	1.84
0.00001	1.701	1.837e-04	176	2.12
0.0001	17.01	0.0150	196	3.09
0.001	170.1	0.9152	210	4.53
0.01	1701	174.4055	231	15.11

TABLE V.
CALCULATION OF COMPRESSION RATIO AND PERCENTAGE OF REDUNDANT DATA FOR CASE STUDY 2

	DE	**SA**	**TLBO**	**PSO**	**GA-M**	**GA**
C_r	2.003	2.003	2.015	-	-	-
R%	50.08	50.08	50.38	-	-	-
D_r	318396	318396	316512	-	-	-

TABLE VI.
CALCULATION OF VARIOUS ACCURACY WITH DIFFERENT N_t FOR ELECTRICAL DATASET OF HOMES

α	N_t	*MSE*	S_d	C_r
0.0001	1.7593	-	infeasible	-
0.001	17.593	0.0261	275	2.015
0.01	175.93	0.9803	428	22.57

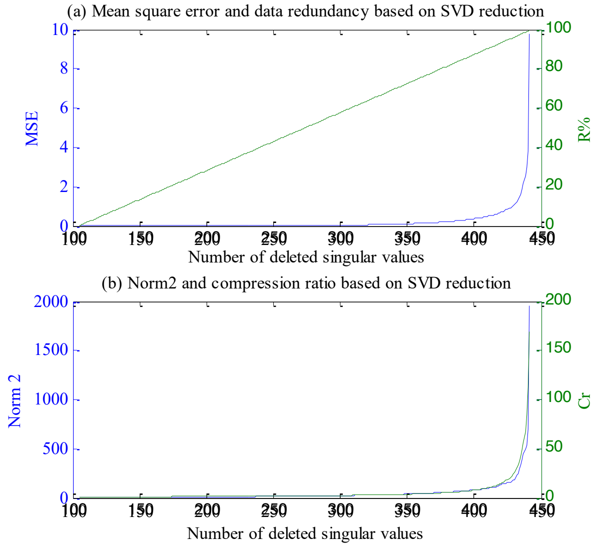

Figure. 5. The relationship of the neglected singular values with MSE (a) and Norm2 (b) for electrical dataset of homes

D. Discussion

Generally, in the proposed scheme, GA, GA-M, DE, SA, PSO, and TLBO are used to solve the optimization framework. In the first case study with a size of 315×240, the highest compression ratio is 4.5323, i.e. 77.93% of the original data are redundant and the original matrix can be reconstructed by 22.07% of elements. In fig. 4 (a), since the neglected singular values are small and don't have a significant effect on the data, about 225 singular values can be eliminated with constant MSE. According to fig. 4 (b) the behavior of N_2 and C_r is similar to MSE. These properties are affected by the nature of the data. In other words, more dependency of rows in the matrix leads to more accuracy in compression. Also, in the second dataset with the size of 1440×443, the highest compression ratio is 2.015. Accordingly, the proposed method eliminates 50.38% of the original data considering the high accuracy on the decoder side.

Moreover, Fig. 5 (a) shows that if the number of neglected singular values is high, the MSE will be augmented. So, accuracy of the reconstructed data will be decreased and the rate of the redundant data will be augmented. It is observed that the MSE approximately has remained constant and then it has been sharply increased. Referring to Fig. 5 (a) and (b), the MSE and N_2 do not change in wide range and errors are suddenly increased at one point. Hence, the PSO and TLBO which have the closest response to this point, perform better than the other ones. Consequently, PSO and TLBO have generally good performance in optimality and runtime, while in the second case study, the optimal performance is achieved by TLBO algorithm. The results of this application on both datasets have been demonstrated in Table II and Table V, respectively, which demonstrate the examples of the optimally compressed data on the smart grid. This paper addresses the optimal compression of data in smart grids based on the SVD decomposition and evolutionary algorithms.

V. Conclusion

This paper proposed a scheme for the data compression that is based on the optimal SVD. The presented method can individually solve the issues appearing by the big volume of data such as required bandwidth and data storage by reducing the number of data elements. The proposed framework achieves the higher compression ratio as well as the satisfaction of accuracy constraint simultaneously and the redundant section is cut down. It is simple and efficient and could be utilized to market operator [29], load aggregator [30-32] and electricity retailers [33, 34]. The optimization using evolutionary methods such as DE, TLBO, PSO, GA, and GA with mutation show significant improvements in the performance of the proposed compression which is used to enhance the performance of SVD-based compression methods. Results show the effectiveness of the method on various datasets with satisfactory responses and accuracy.

VI. Acknowledgment

João P. S. Catalão acknowledges the support by FEDER funds through COMPETE 2020 and by Portuguese funds through FCT, under POCI-01-0145-FEDER-029803 (02/SAICT/2017).

References

[1] K.M. Aishwarya, R. Ramesh, P.M. Sobarad, V. Singh, “Lossy image compression using SVD coding algorithm”, *IEEE International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET),* 2016.

[2] kh. Sayood, “introduction to data compression”, 4rd ed., United States of America, 2012..

[3] J. Khan., SH. Bhuiyan, G. Murphy, and J. Williams ,”Data Denoising and Compression for Smart Grid Communication”, *IEEE Trans. Signal and Information Processing over Networks*, pp. 1-15, 2016.

[4] J. Khan, Sh. M. A. Bhuiyan, G. Murphy, M. Arline, “Embedded-Zerotree-Wavelet-Based Data Denoising and Compression for Smart Grid”, *IEEE Trans. Ind. Appl.*, vol.51, no. 5, pp. 4190-4200, 2015.

[5] P. H. Gadde, M. Biswal, S. Brahma and H. Cao,“Efficient Compression of PMU Data in WAMS”, *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2406-2413, 2016.

[6] J. Cormane, F. Nascimento, “Spectral Shape Estimation in Data Compression for Smart Grid Monitoring”, *IEEE Trans. Smart Grid*, vol.7, no. 3, pp.1214-1221, 2016.

[7] J. Ning, J. Wang, W. Gao, and C. Liu, “A Wavelet-Based Data Compression Technique for Smart Grid”, *IEEE Trans. Smart Grid.*, vol. 2, no. 1, pp. 212-219, 2011.

[8] X. Huang, T. Hu, C. Ye, G. Xu, X. Wang, L. Chen,”Electric Load Data Compression and Classification Based on Deep Stacked Auto-Encoders”. *Energies*, vol. 12, no. 4, 653. Feb. 2019.

[9] V. Loia , S. Tomasiello , A. Vaccaro, “Fuzzy Transform Based Compression of Electric Signal Waveforms for Smart Grid”, *IEEE Trans. Syst.*, Man, and Cybernetics: Systems., vol. 47, no. 1, pp. 121-132, 2017.

[10] Y. Sun, C. Cui, J. Lu, and Q. Wang, “Data compression and reconstruction of smart grid customers based on compressed sensing theory”, ELSEVIER Electrical Power and Energy Systems, vol. 83, pp. 21–25, 2016.

[11] F. Barrosa, W. Fonsecab, U. Bezerraa, M. Nunesaa, “Compression of electrical power signals from waveform records usinggenetic algorithm and artificial neural network”, ELSEVIER Electric Power Systems Research,vol.142, pp. 207–214, 2017.

[12] Unterweger, and D. Engel, “Resumable Load Data Compression in Smart Grids”, *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 919-929, 2015.

[13] J. Cesar Stacchini de Souza, T. Mariano Lessa Assis. , and B. Pal, “Data Compression in Smart Distribution Systems via Singular Value Decomposition”, *IEEE Trans. Smart Grid*, pp.1-10,2016.

[14] A.M. Rufai, GH. Anbarjafari, H. Demirel, “Lossy image compression using singular value decomposition and wavelet difference reduction”, ELSEVIER Digital Signal Processing, vol.24, pp. 117–123, 2014.

[15] J. Kraus, P. Stephan, L. Kukacka, “Optimal Data Compression Ttechniques for Smart Grid and Power Quality Trend Data”, *15th IEEE International Conference on Harmonics and Quality of Power (ICHQP)*, 2012.

[16] M.P. Tcheou, L. Lovisolo, M.V. Ribeiro, E. A. B. da Silva, M. A. M. Rodrigues, J. M. T. Romano, P. S. R. Diniz, “The Compression of Electric Signal Waveforms for Smart Grids: State of the Art and Future Trends”, *IEEE Trans. Smart Grid*, Vol.5, issue.1, 2014.

[17] D.C. Lay, J. Goldstein, D. Schneider, “Linear Algebra and Its Applications”, *4rd ed., United States:The University of Texas at Dallas*, 1993.

[18] S. Zaferanlouei, M. Korpås, J. Aghaei, H. Farahmand, N. Hashemipour, “Computational efficiency assessment of multi-period AC optimal power flow including energy storage systems”, *IEEE International Conference on Smart Energy Systems and Technologies (SEST)*, pp.1-6, 2018.

[19] N. Hashemipour, T. Niknam, J. Aghaei, H. Farahmand, M. Korpås, M. Shafie-Khah, G. J. Osorio, J. P. S. Catalão, “A Linear Multi-Objective Operation Model for Smart Distribution Systems Coordinating Tap-Changers, Photovoltaics and Battery Energy Storage”, *IEEE Power Systems Computation Conference (PSCC)*, pp.1-7, 2018.

[20] Sk. Minhazul Islam, S. Das, S. Ghosh, S.Roy, and P. Nagaratnam Suganthan, “An Adaptive Differential Evolution Algorithm With Novel Mutation and Crossover Strategies for Global Numerical Optimization”, *IEEE Trans. Systems*, *Man and Cybernetics*. vol. 42, no. 2, 2012.

[21] Y.L. Li, Z.H. Zhan, Y.J. Gong, W.N. Chen, J. Zhang, and Y. Li, “Differential Evolution with an Evolution Path: A DEEP Evolutionary Algorithm”, *IEEE Trans. Cybernetics.*, vol. 45, no. 9, 2015.

[22] *Day-Ahead and Real-Time Energy Markets.* [online]., accessed https://www.iso-ne.com/markets operations/markets/da-rt-energy-markets, May. 3, 2018.

[23] *Smart* Data Set for Sustainability.* [online]., accessed http://traces.cs.umass.edu/index.php/Smart/Smart, May. 8, 2018.

[24] A. K. Qin, V. L. Huang, P. N. Suganthan, “Differential Evolution Algorithm With Strategy Adaptation for Global Numerical Optimization”*, IEEE Trans. Evolutionary Computation*., vol. 13, no. 2, 2009.

[25] S. Kirkpatrick; C. D. Gelatt; M. P. Vecchi, “Optimization by Simulated Annealing”, Science, New Series, vol. 220, no. 4598, pp. 671-680, 1983.

[26] R.V. Rao, V.J. Savsani, D.P. Vakharia, “Teaching–Learning-Based Optimization: An optimization method for continuous non-linear large scale problems”, *ELSEVIER Information Sciences*, vol.183, no.1, pp-1–15, 2012.

[27] M. Clerc and J. Kennedy,” The particle swarm-explosion, stability and convergence in a multidimensional complex space”, *IEEE Trans.Evolutionary Computation*, vol.6, no.1,pp. 58-73, 2002.

[28] M. Srinivas, L.M. Patnaik, “Genetic algorithms: a survey”*, IEEE Computer*., vol. 27, no. 6,pp.17-26, 1994.

[29] H. Xu, Y. Lin, X. Zhang, and F. Wang, “Power system parameter attack for financial profits in electricity markets,” *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3438-3446, Jul. 2020.

[30] F. Wang, B. Xiang, K. Li, X. Ge, H. Lu, J. Lai, and P. Dehghanian, “Smart households' aggregated capacity forecasting for load aggregators under incentive-based demand response programs,” *IEEE Trans. Ind. Appl.*, vol. 56, no. 2, pp. 1086-1097, Mar.-Apr. 2020.

[31] X. Lu, K. Li, H. Xu, F. Wang, Z. Zhou, Y. Zhang, “Fundamentals and business model for resource aggregator of demand response in electricity markets,” *Energy*, vol. 204, Art. no. 117885, May. 2020.

[32] K. Li, F. Wang, Z. Mi, M. Fotuhi-Firuzabad, et al, “Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation,” *Appl. Energy*, vol. 253, Art. no. 113595, Nov. 2019.

[33] F. Wang, K. Li, N. Duić, Z. Mi, B.M. Hodge, M. Shafie-khah, and J. P. S Catalão, “Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns,” *Energy Convers. Manag.*, vol. 171, pp. 839-854, Sep. 2018.

[34] S. Yan, K. Li, F. Wang, X. Ge, X. Lu, Z. Mi, H. Chen, and S. Chang, “Time-frequency features combination-based household characteristics identification approach using smart meter data,” *IEEE Trans. Ind. Appl.*, vol. 56, no. 3, , pp. 2251-2262, May-Jun. 2020.