



**HAL**  
open science

## Compact Representation of Nodal Prices in a Distribution Grid

Guéno   Ch  rot, Roman Le Goff Latimier, H. Ben Ahmed

► **To cite this version:**

Gu  no   Ch  rot, Roman Le Goff Latimier, H. Ben Ahmed. Compact Representation of Nodal Prices in a Distribution Grid. 2024 22nd International Conference on Intelligent Systems Applications to Power Systems (ISAP), Sep 2024, Budapest, Hungary. 10.1109/ISAP63260.2024.10744350 . hal-04778059

**HAL Id: hal-04778059**

**<https://hal.science/hal-04778059v1>**

Submitted on 12 Nov 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 destin  e au d  p  t et    la diffusion de documents scientifiques de niveau recherche, publi  s ou non,   manant des   tablissements d'enseignement et de recherche fran  ais ou   trangers, des laboratoires publics ou priv  s.

# Compact Representation of Nodal Prices in a Distribution Grid

Guéno   CH  ROT, Roman LE GOFF LATIMIER, Hamid BEN AHMED  
SATIE, UMR CNRS 8029, UniR – ENS Rennes  
Rennes, France  
guenole.cherot@ens-rennes.fr

**Abstract**—As distributed generation becomes more widespread, distribution network management methods need to be adapted. To achieve an optimal network management, the resolution of the optimal power flow (OPF) problem indicates that it is necessary to use a minimum number of decision variables equal to the number of buses making up the network. However, a real deployment is likely to be limited to sending one or a few variables to the agents connected to the network, as shown by the rich literature currently devoted to this subject. A key question is then to know the optimality gap between the optimal solution and the one using a reduced number of variables. Based on the control variables derived from the OPF and using principal component analysis (PCA), this contribution proposes the construction of new variables, known as latent variables, which are decorrelated from one another. The control of the network by these variables shows, for the test case considered, that only 30% of the variables are necessary to obtain near-optimal control. This finding makes it possible to consider learning these variables for networks in which their operational calculation is impossible in real time.

**Index Terms**—Distribution network, nodal price, dimensional-ity reduction, principal component analysis, Optimal power flow

## I. INTRODUCTION

The transition to less carbon-intensive energy production will not be possible without the widespread use of renewable energies [1]. These are distributed and mainly connected to distribution networks (DN) [2]. At present, the low penetration rate of renewable energies, the size of the lines and the limited flexibilities available lead to a situation where management is unnecessary and not deployed. However, there is a trend towards networks in which 100% of players are equipped with distributed generation resources, in addition to routing consumption linked to electric mobility. Half of the DNs could then experience issues of congestion or non-compliance with the voltage limits [3]. The distribution system operator (DSO) must therefore find new levers to continue to ensure the distribution of high-quality energy.

Numerous management strategies based on the existence of distributed flexibilities – electric vehicles, heat pumps, etc. – and on the democratisation of smart meters – Linky used by Enedis in France for example – are currently being actively studied. The direct solution, formulated as an optimisation problem taking into account all the physical constraints of the network, was first proposed by Carpentier [4] under the name of optimal power flow (OPF). The solution to this problem

is usually an active and reactive power injection plan – the primary variables in the problem – to be communicated to each producer – and possibly consumers if they are flexible. The same optimal situation can also be implemented by communicating the nodal electricity prices (NP) – dual variables of the problem – to each node [5], [6]. In a perspective of operational deployment, this second communication scheme, which is non-injunctive for users, seems better suited to distributed control, for instance using broadcast of Distribution Locational Marginal Prices values [7]. Although optimal, this solution is still not perfectly adapted to the distribution grid because of the number of agents involved [8] and the uncertainties of the network [9]. Relaxing some of the constraints makes the problem convex, which greatly speeds up the optimisation. Under certain conditions, the relaxations have a zero sub-optimality gap [10], [11] but solving them requires perfect knowledge of the agents in the network. However, these agents may wish to keep their preferences private. In addition, such an approach is hampered by the computational burden and the communication challenge of gathering data from players in real time.

This problem has driven the development of peer-to-peer algorithms. Baroche [12], for example, has proposed two methods for integrating the SO into a peer-to-peer market. The operator could constrain exchanges by means of a network usage price, thus ensuring that congestion constraints were respected. Faria [13] proposes another type of penalty based on the distance between agents, and compares it with an approach based on a flexibility market in which the DSO would offer flexibilities to the agents until all the constraints are resolved. This approach, based on the presence of aggregators managing local constraints, seems to be the most promising [14]. Finally, a last approach consists in formulating the management of flexibilities as a reinforcement learning problem in which the DSO decides at each instant the powers exchanged by each agent and where the reward signal penalises non-compliance with the constraints [15], [16].

The diversity of these methods highlights the dynamism of the research currently devoted to the open question of managing active distribution networks. Focuses are either on voltage control [17] or on power management [18]. Nevertheless, they all share one common feature: the use of a limited number of signals to control or influence the numerous players in the distribution network. This choice necessarily implies a sub-

optimality of control that needs to be quantified.

This contribution therefore seeks to answer the following questions: i) what is the sub-optimality induced by the control of a network using an insufficient number of variables? ii) for a fixed number of variables, how can the optimal set of variables be constructed?

To answer these questions, we will draw on the literature on nodal prices [5], [6]. These variables, derived from the OPF, allow optimal control of the network when communicated to each and every bus making the network. In addition the connected agents are assumed to behave rationally. By studying the structure of these variables using feature extraction methods [19] such as Principal Component Analysis (PCA) [20], we will investigate to what extent these variables can be gathered into a smaller latent space.

The rest of the article will be organised as follows. The section II will introduce PCA and its use as an information compression tool. Section III will describe the test case and the simulation tools used. Finally, section IV will discuss the results.

## II. METHOD FOR NODAL PRICES REDUCTION

### A. Principal Component Analysis

Principal Component Analysis (PCA) is a data analysis method that transforms correlated variables into new variables that are decorrelated from one another. The transformation induced by PCA is linear and can be seen as a change of basis [20], [21]. The unit vectors forming the new basis – called the latent space – are orthogonal to each other. They are ordered from most to least significant in terms of the variance explained in the input data.

Let be a vector of  $N$  random variables denoted  $\mathbf{x} = (x_1, \dots, x_N)$  with mean value being zero. Let  $x_i$  be the components of  $\mathbf{x}$  in the initial basis ( $\mathbf{e}_1, \dots, \mathbf{e}_N$ ). We denote  $\mathbf{x}^{[m]}$  the  $m^{\text{th}}$  observation of  $\mathbf{x}$ . The set of  $M$  observations forms the matrix  $\mathbf{X} \in \mathcal{M}_{M,N}$ :

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^{[1]} \\ \mathbf{x}^{[2]} \\ \dots \\ \mathbf{x}^{[M]} \end{pmatrix} = \begin{pmatrix} x_1^{[1]} & x_2^{[1]} & \dots & x_N^{[1]} \\ x_1^{[2]} & x_2^{[2]} & \dots & x_N^{[2]} \\ \dots & \dots & \dots & \dots \\ x_1^{[M]} & x_2^{[M]} & \dots & x_N^{[M]} \end{pmatrix} \quad (1)$$

The matrix  $\mathbf{T}$  allows us to pass from the initial space to the latent space ( $\mathbf{e}'_1, \dots, \mathbf{e}'_N$ ) in the following way:  $\mathbf{x}' = \mathbf{x}\mathbf{T}$ . This matrix consists in the eigenvectors of the covariance matrix  $\mathbf{A} = \mathbf{X}\mathbf{X}^T/(M-1)$ . The eigenvalues of  $\mathbf{A}$  indicate an order of importance of the eigenvectors. The percentage of information contained in the latent variable  $x'_i$  is:

$$\%_i = \frac{\lambda_i}{N} \quad (2)$$

where  $\lambda_i$  is the eigenvalue associated with the eigenvector  $\mathbf{e}'_i$ .

It is common for most of the information to be contained in just a few latent variables. It is then possible to define a new observation containing only part of the information:

$$\tilde{\mathbf{x}}^{(k)} = (x'_1, \dots, x'_k, 0, \dots, 0) \quad (3a)$$

$$= \mathbf{x}'\mathbf{K} \quad (3b)$$

where  $\tilde{\mathbf{x}}^{(k)}$  is the new compressed observation and  $\mathbf{K} \in \mathcal{M}_{N,N}$  the matrix preserving only the  $k$  first components of  $\mathbf{x}'$ .

### B. Application to nodal prices

In the context of this study, the impact of dimension reduction on the management performance of a distribution network is investigated. To do this, a dimension reduction method must be selected. PCA was chosen for its simplicity of implementation, its low computational cost and the explicability of the obtained results. However, this is not the only method that could have been applied here. In the light of the results obtained, the relevance of this choice and the differences that might be anticipated if it were to be modified will be discussed in section IV-D.

By using PCA for the present problem, correlations between nodal prices will be highlighted. We therefore need to collect a large number of observations representative of the operation of a distribution network. To do this, production and consumption time series must be collected in a database (DB) to describe the power demanded by each agent on the network. Combined with an OPF, it enables us to generate the data required for the PCA. For the remainder of this contribution, this step will be referred to as the *training phase*.

Once the PCA has been performed, the performance of the dimension reduction is tested on a test DB. The aim of this step is to quantify the difference between the optimal solution and the solution based on compressed latent prices - differences in voltages, line flows and total cost. This step is referred to as the *test phase*. The management method based on reduced nodal prices is described in Fig. 1. It is described in detail below.

*Training:* ① For each operating point from the training DB, we calculate an OPF to obtain the nodal prices. We denote  $\mathbf{np} = (np_1, \dots, np_N)$  the vector of  $N$  nodal prices, corresponding to each network node. The nodal price database is made up of all the  $\mathbf{np}$  observed during the period considered for this training phase. Thus:  $\mathbf{NP} = (\mathbf{np}^{[1]}, \dots, \mathbf{np}^{[M]})$  with  $M$  observations – from the  $M$  operating points of the training DB. ② The nodal prices  $\mathbf{NP}$  are normalized using the method described (4). PCA is then performed to calculate the  $\mathbf{T}$  transformation matrix.

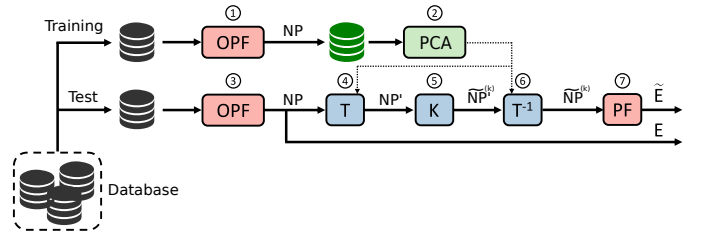


Fig. 1. Nodal price dimension reduction method. In the first phase – training data – the  $\mathbf{T}$  transformation matrix is identified using principal component analysis. In the second phase – test data – a distribution network is controlled using  $\tilde{\mathbf{NP}}$  nodal prices whose information has been compressed.

$$\mathbf{x}_{\text{norm}}^{[i]} = \frac{\mathbf{x}^{[i]} - \bar{\mathbf{X}}}{\boldsymbol{\sigma}(\mathbf{X})} = \left( \frac{x_1^{[i]} - \bar{x}_1}{\sigma(x_1)}, \dots, \frac{x_N^{[i]} - \bar{x}_N}{\sigma(x_N)} \right) \quad (4)$$

*Test:* ③ Once training is complete, original configurations – i.e. those not included in the training DB – are examined. As in ①, for each of these configurations, optimal nodal prices are calculated using an OPF. ④ They are then projected into latent space using the  $\mathbf{T}$  transformation matrix obtained in step ②. ⑤ The  $\mathbf{K}$  matrix is used to retain only the  $k$  first components, the others being replaced by zeros. As the components of a PCA are ordered, only the least important information is removed. ⑥ The compressed latent vectors are projected into the original space using  $\mathbf{T}^{-1}$ . The whole operation can be summarized as follows:  $\tilde{\mathbf{n}}\mathbf{p}^{(k)} = \mathbf{n}\mathbf{p}\mathbf{T}\mathbf{K}\mathbf{T}^{-1}$ . ⑦ Finally, the  $\tilde{\mathbf{n}}\mathbf{p}$  are transmitted to all network players so that they can deduce their power injection. A power flow is performed to compare the state of the network under optimal  $E$  management and degraded  $\tilde{E}$  management respectively.

### III. TEST CASE AND SIMULATION FRAMEWORK

The *Low Voltage European Test Feeder* [22] was selected as the test case, as it is representative of European DRs. A graphical representation of this network is given figure 5.

The players in the network are selected at random from a database. The database contains time series with minute time steps, of power exchanged by households, electric vehicles (EVs) and photovoltaic (PV) panels. The number of each type of actor is specified when the test case is created. Each actor is then assigned a price. It will be assumed fixed for the whole simulation, and drawn randomly according to a normal distribution  $\mathcal{N}(\mu, \sigma)$ . The power of each actor type, and the values of  $\mu$  and  $\sigma$  are given in table I. The power exchanged is deliberately higher than in the original test case, in order to activate the constraints – mainly voltage limits – without which nodal prices would be uniform over the entire network.

The OPF is a non-convex optimization problem and therefore has no exact solution in the general case. We will use the formulation proposed by [23] which combines a relaxation of the constraint on the angles of the tensions and a relaxation of the equality constraint on the flow in the lines [24]. They allow us to calculate the exact solution of the OPF in the case of non-meshed networks and when the reverse power flow is only active, only reactive, or zero. The optimality of the solution is systematically verified a posteriori. The reader is referred to [11] for a full analysis.

As part of an open and reproducible science approach, all the data used are publicly accessible: [25] for household

TABLE I: Selected parameters for test case creation.

	$\mu$	$\sigma$	Peak total power
Household	0.25	0.10	1000 kW
EV	0.15	0.05	200 kW
PV	0.08	0.02	800 kW

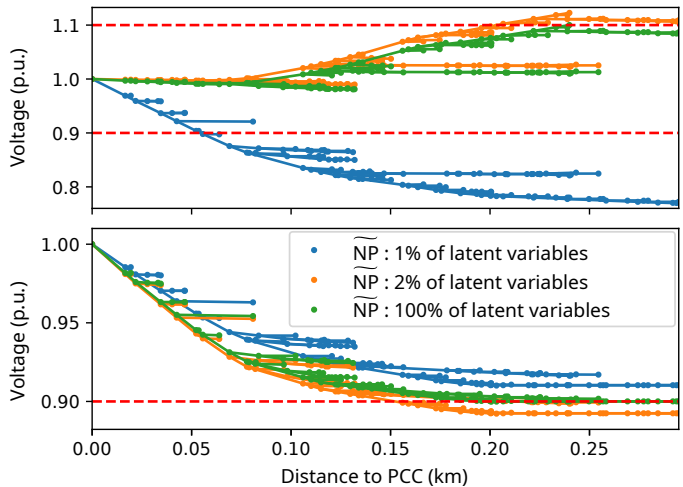


Fig. 2. Voltage evolution at each bus depending on its distance from the PCC and according to the management method used: OPF in green, using 1% latent variables in blue and using 2% in orange. Red dotted lines represent the  $\pm 10\%$  boundaries around the nominal voltage.

consumption and wind generation, [26] for electric vehicle consumption and EPEX<sup>1</sup> for energy market prices. Power flow and OPF have been realized using the PandaPower [27] and PowerModels [28] libraries. We use the PCA implementation proposed by Scikit-Learn [29]. The source codes for this study are open and accessible on GitLab<sup>2</sup>.

The simulations were conducted by separating the 365 days in the database into two groups: 60% of the days were used for training, 40% for testing.

### IV. RESULTS AND DISCUSSION

All the results presented in this section are based on the case presented in section III. The performance of an OPF-based management of this example distribution network is compared with the performance obtained when nodal prices are obtained after compression by a PCA.

#### A. Illustration on a specific configuration

Figure 2 shows two examples of voltage plans, obtained at two different times on the same case study. The first configuration – upper panel – illustrates a situation where reduction leads to inefficient or even dangerous management. The second – lower panel – is a situation where a very high reduction can be applied without difficulty. Voltages are shown as a function of their distance from the Point of Common Coupling (PCC), in order to maintain the network structure and make it easier to read the results. The PCC acts as a reference, with its voltage set at 1 p.u. The voltage values obtained using the optimal NPs from an OPF is shown in green. When the NPs are compressed using only 1% – respectively 2% – of the latent space dimensions, the resulting voltages is shown in

<sup>1</sup><https://ewoken.github.io/epex-spot-data/>

<sup>2</sup>[https://gitlab.com/satie.sete/dimred\\_nodalprices\\_distribgrid](https://gitlab.com/satie.sete/dimred_nodalprices_distribgrid)

blue – respectively orange. Permissible stress limits – set at 1.1 p.u. and 0.9 p.u. – are shown in dotted red.

Configuration n°1 – upper panel – illustrates a case where reduction works very well using 2% of latent variables, but very poorly below. Indeed, for 1% of latent variables, 90% of buses does not respect the constraints. The maximum deviation from the optimal solution, 0.22 p.u., is obtained at the end of the line. In this configuration, the optimally managed network should have been a power exporter, as evidenced by voltages in excess of 1 p.u. However, compressing nodal prices to 1% creates a gap that makes the network an importer, resulting in significant voltage differences.

Configuration n°2 – lower panel – corresponds to a time when the network is a power importer. Voltages therefore decrease when moving farther away from the PCC. They fall by 0.05 p.u. on average. It can be seen that as the number of latent variables increases, the gaps between the optimal and compressed voltage plans become smaller. In particular, deviations can be positive or negative. Note that poorly estimated NPs do not necessarily lead to a violation of the voltage plan. It is remarkable that the strategy retaining 2% of latent variables leads to more violations than that using only 1%. Indeed, a better approximation of the NPs leads the voltages to be closer to the optimal values derived from the OPF. However, the OPF has often placed them on their admissible limits, as can be seen on the green curve in Fig. 2. Thus, even the slightest error in the NPs can lead to boundaries violations. The phenomena illustrated here highlights the difficulties and risks associated with dimension reduction. Although an overall improvement in management is expected when compression is light, many specific phenomena are likely to occur on constrained configurations. The evolution of performance may therefore not be monotonous when studying a particular network and call for a risk specific attitude.

### B. Suboptimality induced by dimension reduction

After illustrating the impact of dimension reduction on the particular configurations of figure 2, this section systematizes the study of the performance of management by compressed NPs. Figure 3 describes the differences between the optimal strategy (OPF) and the strategy with compressed nodal prices, in terms of information captured, difference in costs, differences in observed voltages and frequency of voltage violations. It is essential, as it enables us – for a fixed network – to estimate the maximum performance of a management strategy using a fixed number of variables. This figure aggregates the results obtained by simulating the optimal and compressed strategies over the entire duration of the case study.

Panel 1 shows the percentage of variance explained as a function of the number of components retained in the latent space – expressed in % for greater generality. For this network, 89% of the information contained in the NPs is captured from the first component and 99% from two components. According to this metric, very few dimensions in latent space are sufficient to explain all the observed NPs. This observation was to be expected: all NPs are identical if no tension or

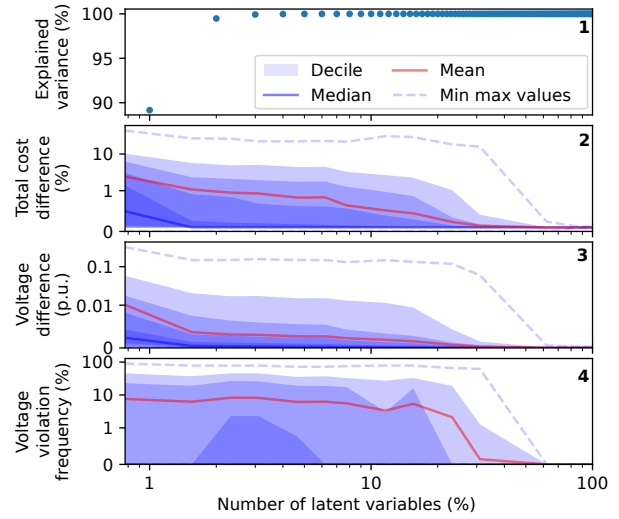


Fig. 3. Effect of dimension reduction on control sub-optimality. Panel 1 represents the information contained in each component of the latent representation. The following panels represent different  $|E - \tilde{E}|$  deviations between the OPF strategy and the proposed strategy as a function of the number of components retained in the latent space (step ⑤). Panels 2 and 3 show cost and voltage plan deviations respectively. Panel 5 describes the frequency of voltage violations.

congestion constraint is activated. In our case, and even after overloading the network (see section III), this only happens 27% of the time. However, in the light of the developments described in the following panels, this loss of information, although seemingly minimal at this stage, has very strong repercussions on the electrical quantities of the network.

Panels 2, 3 and 4 describe the differences between the two strategies. Note that for all panels, the y-axis is linear between 0 and the first scale, then logarithmic for higher values. Extreme values are represented by dotted lines, and deciles by color shading. Thus, the first solid line from the top corresponds to the decile at 90%. The median and mean values are also shown. These statistics have been constructed from the sets described equation 5:

$$\text{Tot cost diff}(k) = \left\{ 100 \cdot \left| \sum_{i \in \mathcal{A}} \frac{\theta_i P_i^{[j]} - \theta_i \tilde{P}_i^{[j]}}{\theta_i P_i^{[j]}} \right|, \forall j \in \mathcal{O}_k \right\} \quad (5a)$$

$$\text{Voltage diff}(k) = \{ |V_i^{[j]} - \tilde{V}_i^{[j]}|, \forall (j, i) \in \mathcal{O}_k \times \mathcal{N} \} \quad (5b)$$

$$\text{Voltage violation freq}(k) = \left\{ \frac{100}{\#\mathcal{N}} \sum_{i \in \mathcal{N}} \mathbb{1}_{\tilde{V}_i^{[j]}}, \forall j \in \mathcal{O}_k \right\} \quad (5c)$$

where  $k$  is the number of dimensions conserved in the latent space,  $\mathcal{O}_k$  is the set of observations conserving  $k$  dimensions,  $x^{[j]}$  denotes the observation  $j$  of a quantity  $x$  and  $\tilde{x}$  of a quantity resulting from control by reduced nodal prices,  $\theta_i$  and  $P_i$  are the price and power exchanged by agent  $i$ ,  $\mathcal{A}$  is the set of agents in the network,  $\mathcal{N}$  is the set of nodes in the network and  $\#\mathcal{N}$  their number,  $\mathbb{1}_{\tilde{V}_i^{[j]}}$  is the indicator function for violation of a voltage constraint.



Panel 2 illustrates the differences in total cost, i.e. the sum of cost paid by each agent in the network. As indicated in the previous section, the general trend is indeed that performance approaches optimal management when more dimensions are taken into account. However, this trend is not entirely monotonic or uniformly paced. In particular, extreme values are the most difficult to estimate in a Monte-Carlo simulation. By keeping only one variable during compression – the extreme left of the figure – performance differences of the order of a few percent can be observed on average. This value is pulled upwards by a few particularly poorly managed and costly configurations – see extreme and last decile. At the maximum, there is a 60% difference. The median is significantly lower than the mean. A management showing good overall performance can therefore be achieved with very high compression ratios. However, to guarantee that all configurations will be well handled, it is necessary to retain many more variables: a compression rate of 25% is thus necessary for the last decile to fall below 1% of additional cost. A difference of the magnitude of 0.1% is still observed when all variables are used. This discrepancy, which should be zero, is due to approximations in numerical precision and convergence criteria, and could be mitigated by reducing the tolerance of the OPF.

Panel 3 describes the voltage deviation in p.u. Similarly to the interpretation in panel 2, these deviations decrease monotonically until they reach zero error. Using only 1% of the latent variables, only 10% – 9<sup>th</sup> decile – of the voltages have a deviation greater than 0.09 p.u.. We can also note that 90% of the deviations are zero with the use of 30% of latent variables.

Panel 4 shows the frequency of voltage violations. On the ordinate, 100% means that all buses - except the PCC - violate the voltage plan. Such a situation can occur in the worst configuration in the dataset, even when up to 20% of latent variables are used. This means that some configurations are highly sensitive to the values taken by nodal prices. These are the configurations for which the optimal operating voltages are very close to the constraints. This phenomenon also explains the non-monotonicity of the 7<sup>th</sup> decile, as we saw in section IV-A.

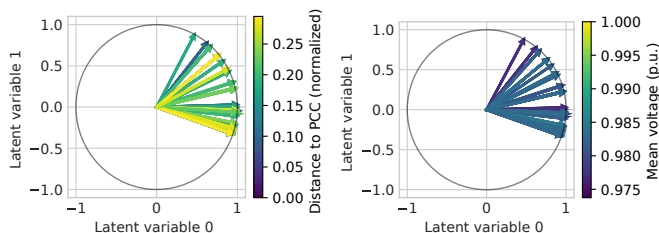


Fig. 4. PCA correlation circles represent the correlations between the old variables (nodal prices) and the new, decorrelated variables (latent space vector). All nodal prices have an equivalent contribution. The distance to the PCC and the average voltage do not allow us to dissociate the nodal prices.

### C. Analysis of the PCA decomposition and discussion

The correlation circle is a tool for analyzing the transformation constructed by PCA. To obtain it, each vector of the initial space  $e_i$  corresponding to the price of node  $i$  is projected onto the first two components of the latent space ( $e'_1, e'_2$ ) forming a vector whose distance from the origin describes the importance of this price with regard to the PCA. Positively correlated prices are grouped together. Each vector can then be colored according to an exogenous criterion to check if it can explain the groups thus formed. To the best of our knowledge, six exogenous criteria could explain the structure of the data: distance to the PCC and the minimum, maximum and average values of voltage and power at each bus.

Two correlation circles are shown in figure 4, for which the vectors representing each nodal price are colored as a function of distance from the point of common coupling (PCC) and average bus voltage respectively. All nodal prices are very close to the circle with radius 1. They are therefore of similar importance. What's more, the two exogenous criteria represented do not allow us to identify a trend to explain the distribution of nodal prices. The unrepresented criteria lead to the same conclusions.

Figure 5 represents the relative importance of each nodal price in the PCA of the first four latent vectors. Average voltage and average power at each bus are also shown. Once again, these variables do not intuitively explain the importance of nodal prices in the PCA.

### D. Discussion of the dimension reduction algorithm

As mentioned in section II, PCA is one dimension reduction method among many others. This choice seemed appropriate for this first study, due to its simplicity of implementation and explainability. However, its limitations need to be discussed in the light of the results obtained.

PCA is a linear transformation. It is therefore unable to explain the data structure of a problem with any certainty, particularly if the underlying pattern is non-linear, non-orthogonal or if there are hidden groups [30].

A more general approach might be to use an auto-encoder [31], [32], a particular type of neural network for which one of the deep layers deliberately has a reduced number of neurons. By minimizing the reconstruction error, the algorithm is forced to learn a compact representation of the data. A natural extension of the here proposed approach would then be to transpose the presented study to a comparison of the performances obtained using several dimension reduction tools.

## V. CONCLUSION AND FUTURE WORKS

As distribution networks are an essential part of the energy transition, optimal management of the agents connected to them is a key issue in the literature devoted to active distribution networks. The present contribution focuses on the number of signals required to achieve satisfactory management, moving upstream of many of the contributions in the literature that propose specific management methods. The nodal prices

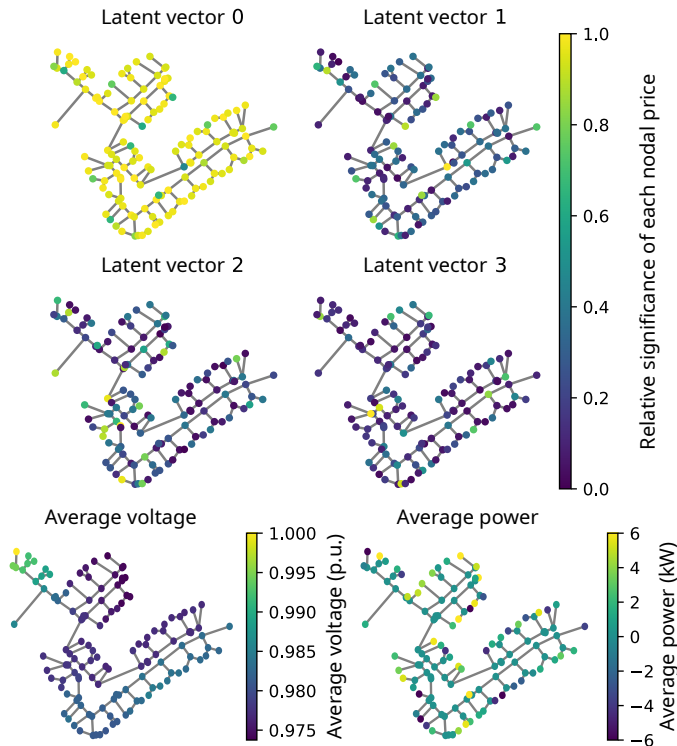


Fig. 5. Eigenvector analysis of the PCA matrix via the location of each nodal price on the network. The first four latent vectors are shown. The contributions of each nodal price are normalized so that the maximum importance for each vector is equal to 1. The figure also shows the average voltage and average power observed on each bus during the simulation.

obtained by an OPF have been used here as a reference for optimal performance. This is compared with prices obtained via PCA dimension reduction. It has been shown that 99% of the variance can be explained by 2% of the components of the latent space. However, this metric is not sufficient for DN management, as nodal prices are structurally highly correlated. Strong violations of the voltage limits can therefore be observed up to the use of 20% of latent components. To ensure that these rare configurations are eliminated, only compressions that retain at least half of the original dimensions can be considered. The contributions of each PCA component were examined. No interpretable structure in its construction matrix could be identified at this stage.

At the end of this work, three main perspectives are open. First, this article constitutes a preliminary contribution that needs to be systematized in order to gain in genericity. In particular, dimension reduction by PCA is a justified choice, but requires comparison with other dimension reduction methods. More specifically, repeating this study by replacing PCA with algorithms capable of capturing non-linear behaviors – such as an auto-encoding neural network – is necessary to confirm the results obtained. Similarly, repeating the study on other network configurations would allow the conclusions to be generalized. Dimension reduction performance is potentially not the same when the network is importing or exporting, as

illustrated in figure 2. Discussing performance according to network characteristics would therefore be promising.

On the other hand, this study makes it possible to predict an achievable performance as a function of the number of variables retained. However, it does not provide an operational management algorithm for achieving this performance. Comparing active distribution network management methods against the performance predicted here would therefore be of particular interest. This would enable us to assess the potential remaining gaps in the methods proposed in the literature. The remaining potential for progress would then be quantified.

Finally, this contribution opens up the possibility of new methods for operational management of the distribution network based solely on nodal prices. Their learning or prediction, initially too complex due to the number of variables involved, is made possible by dimension reduction. The objective of these management algorithms would then be to predict latent variables.

## REFERENCES

- [1] L. Li, J. Lin, N. Wu, S. Xie, C. Meng, Y. Zheng, X. Wang, and Y. Zhao, "Review and outlook on the international renewable energy development," *Energy and Built Environment*, vol. 3, no. 2, pp. 139–157, 4 2022.
- [2] S. Kerscher and P. Arbolea, "The key role of aggregators in the energy transition under the latest European regulatory framework," *International Journal of Electrical Power and Energy Systems*, vol. 134, p. 107361, 1 2022.
- [3] A. Navarro-Espinosa and L. F. Ochoa, "Probabilistic Impact Assessment of Low Carbon Technologies in LV Distribution Systems," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2192–2203, 5 2016.
- [4] J. Carpentier, "Optimal power flows," *International Journal of Electrical Power & Energy Systems*, vol. 1, no. 1, pp. 3–15, 4 1979.
- [5] F. L. Alvarado, "Is system control entirely by price feasible?" *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS 2003*, 2003.
- [6] A. Papavasiliou, "Analysis of distribution locational marginal prices," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4872–4882, 9 2018.
- [7] L. Bai, J. Wang, C. Wang, C. Chen, and F. Li, "Distribution locational marginal pricing (dlmp) for congestion management and voltage support," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4061–4073, 2017.
- [8] Z. Yang, H. Zhong, Q. Xia, and C. Kang, "Fundamental Review of the OPF Problem: Challenges, Solutions, and State-of-the-Art Algorithms," *Journal of Energy Engineering*, vol. 144, no. 1, p. 04017075, 11 2017. [Online]. Available: <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29EY.1943-7897.0000510>
- [9] G. Chérot, R. Le Goff Latimier, F. Sanchez, and H. Ben Ahmed, "Misestimation of Impedance Values Within a Distribution Network Optimal Power Flow," in *2023 IEEE Belgrad PowerTech, PowerTech 2023*. IEEE, 2023.
- [10] M. Farivar and S. H. Low, "Branch flow model: Relaxations and convexification-part II," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2565–2572, 2013.
- [11] S. Huang, Q. Wu, J. Wang, and H. Zhao, "A Sufficient Condition on Convex Relaxation of AC Optimal Power Flow in Distribution Networks," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1359–1368, 2017.
- [12] T. Baroche, P. Pinson, R. Le Goff Latimier, and H. Ben Ahmed, "Exogenous Cost Allocation in Peer-to-Peer Electricity Markets," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2553–2564, 2019.
- [13] A. S. Faria, T. Soares, T. Orlandini, C. Oliveira, T. Sousa, P. Pinson, and M. Matos, "P2P market coordination methodologies with distribution grid management," *Sustainable Energy, Grids and Networks*, p. 101075, 5 2023. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2352467723000838>

- [14] T. Morstyn, A. Teytelboym, and M. D. McCulloch, "Bilateral contract networks for peer-to-peer energy trading," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2026–2035, 3 2019.
- [15] A. Petrushev, M. A. Putratama, R. Rigo-Mariani, V. Debusschere, P. Reignier, and N. Hadjsaid, "Reinforcement learning for robust voltage control in distribution grids under uncertainties," *Sustainable Energy, Grids and Networks*, vol. 33, p. 100959, 3 2023.
- [16] G. Chérot, R. Le Goff Latimier, B. Cajna, and H. Ben Ahmed, "Learning-based optimization of the exchange price in a heterogeneous market," in *22nd International Conference on Intelligent Systems Applications to Power Systems*, 2024.
- [17] J. Wang, W. Xu, Y. Gu, W. Song, and T. C. Green, "Multi-agent reinforcement learning for active voltage control on power distribution networks," *Advances in Neural Information Processing Systems*, vol. 34, pp. 3271–3284, 2021.
- [18] S. Bahrami, Y. C. Chen, and V. W. Wong, "Deep reinforcement learning for demand response in distribution networks," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1496–1506, 2020.
- [19] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Computing Surveys*, vol. 50, no. 6, 12 2017. [Online]. Available: <https://doi.org/10.1145/3136625>
- [20] J. Shlens, "PCA Tutorial," *Measurement*, pp. 1–13, 2005. [Online]. Available: [papers3://publication/uuid/4D1DBE59-7625-4528-BAB6-E076486F0C77](https://papers3://publication/uuid/4D1DBE59-7625-4528-BAB6-E076486F0C77)
- [21] M. Rouaud, *Statistiques Et Analyses Multicritères*, 2017.
- [22] K. P. Schneider, B. A. Mather, B. C. Pal, C. W. Ten, G. J. Shirek, H. Zhu, J. C. Fuller, J. L. Pereira, L. F. Ochoa, L. R. De Araujo, R. C. Dugan, S. Matthias, S. Paudyal, T. E. McDermott, and W. Kersting, "Analytic Considerations and Design Basis for the IEEE Distribution Test Feeders," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3181–3188, 5 2018.
- [23] R. A. Jabr, "Radial distribution load flow using conic programming," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1458–1459, 8 2006.
- [24] L. Gan, N. Li, U. Topcu, and S. H. Low, "Exact Convex Relaxation of Optimal Power Flow in Radial Networks," *IEEE Transactions on Automatic Control*, vol. 60, no. 1, pp. 72–87, 1 2014.
- [25] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, J. Albrecht, and others, "Smart\*: An open data set and tools for enabling research in sustainable homes," *SustKDD, August*, vol. 111, no. 112, p. 108, 2012.
- [26] A. L. Sørensen, K. B. Lindberg, I. Sartori, and I. Andresen, "Residential electric vehicle charging datasets from apartment buildings," *Data in Brief*, vol. 36, p. 107105, 6 2021.
- [27] L. Thurner, A. Scheidler, F. Schafer, J. H. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun, "Pandapower - An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510–6521, 2018.
- [28] C. Coffrin, R. Bent, K. Sundar, Y. Ng, and M. Lubin, "PowerModels.J1: An Open-source framework for exploring power flow formulations," *20th Power Systems Computation Conference, PSCC 2018*, 2018.
- [29] C. M. Bishop, A. M. Saxe, J. L. McClelland, S. Ganguli, F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in {P}ython," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: <http://scikit-learn.sourceforge.net>.
- [30] J. Lever, M. Krzywinski, and N. Altman, "Principal Component Analysis," *Nature Methods*, vol. 14, pp. 641–642, 6 2017. [Online]. Available: <https://www.nature.com/articles/nmeth.4346http://link.springer.com/10.1007/b98835>
- [31] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework," in *International conference on learning representations.*, 2017.
- [32] Q. Meng, D. Catchpoole, D. Skillicom, and P. J. Kennedy, "Relational autoencoder for feature extraction," in *Proceedings of the International Joint Conference on Neural Networks*, vol. 2017-May. Institute of Electrical and Electronics Engineers Inc., 6 2017, pp. 364–371.