Application of Robust PCA with a structured outlier matrix to topology estimation in power grids

Stéphane Chrétien[a,1,*], Paul Clarkson[a], Maria Segovia Garcia[b]

[a] *National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UK.*
[b] *University of Strathclyde, Power Network Demonstration centre, Glasgow, Scotland*

Abstract

Robust PCA is a widely used technique for Principal Component Analysis when the data is corrupted by outliers. The goal of the present short note is to report on the performance results of a simple modification of the method of Netrapali et al. for estimating Low Rank + Sparse models where the sparse matrix has the structure of a tree. We demonstrate the efficiency of the approach on the problem of estimating the topology in power grid networks.

Keywords: Robust PCA, Non-Convex optimisation, Tree structured sparsity

1. Introduction

Recent changes in the electrical power industry such as Distributed Energy Resources (DER), active customer participation in emerging energy markets and deployment of measurement, communication, and control infrastructure have triggered extensive research activity. System operators are seizing a new opportunity to improve power system efficiency and stability by leveraging novel optimization techniques. Very interesting results in this direction were the recent discovery that some state estimation problems could be efficiently relaxed using Semi-Definite Programming [6][12]. One essential prerequisite for such techniques is the knowledge of network topology and the value of the admittance matrix. These data are not so often easy to know since, in many cases, only limited visibility into the state of the system is available.

The inverse power flow (IPF) problem defined in [10] concerns the estimation of the nodal admittance matrix, describing the network topology, from synchronized measurements of voltage and current magnitudes and phase angles which can be obtained from e.g. phasor measurement units (PMUs). In [10, Section III B.], the problem is transformed into the estimation of a low-rank plus sparse matrix where the sparse matrix has the structure of the adjacency matrix of a tree. Thus, the IPF problem can be recast as a Robust PCA problem with structured sparsity.

Robust PCA is a widely used technique for Principal Component Analysis when the data is corrupted by outliers. Taking into account outliers is of crucial importance in many applications such as network analytics [9], gene expression analysis [5] and video processing [1] among many others. Robustness can be ensured via different techniques inherited from Huber's theory of Robust Statistics but has also recently been addressed using Low Rank + Sparse models which can be estimated using efficient penalised least squares approaches. The paper [3] by Candès, Li, Ma and Wright is a milestone in the recent research on Robust PCA in which it is proved that the (noiseless) decomposition $X_0 = L_0 + S_0$ can be computed using the convex relaxation

$$(L^*, S^*) \in \operatorname{argmin}_{L\in\mathbb{R}^{d\times n}, S\in\mathbb{R}^{d\times n}} \ \|L\|_* + \lambda \|S\|_1 \text{ s.t. } X_0 = L + S, \tag{1}$$

*Corresponding author

Email addresses: stephane.chretien@npl.co.uk (Stéphane Chrétien), paul.clarkson@npl (Paul Clarkson), maria.segovia-garcia@strath.ac.uk (Maria Segovia Garcia)

URL: https://sites.google.com/view/stephanechretien/home (Stéphane Chrétien)

when L_0 is sufficiently sparse, the singular vectors of L_0 are "spread", i.e. far from sparse in the terminology of [8] and the matrix S_0 is sufficiently sparse. The case where the observation of X_0 is corrupted by random noise can be addressed using the following estimator

$$(L^*, S^*) \in \text{argmin}_{L \in \mathbb{R}^{d \times n}, S \in \mathbb{R}^{d \times n}} \|L\|_* + \lambda \|S\|_1 \text{ s.t. } \|X - L + S\|_{S_2} \leqslant \epsilon. \tag{2}$$

One of the main issues in using a convex relaxation scheme for (2) is the computational complexity of solving semi-definite programs. One way to get around this issue is to use a Bürer Monteiro type approach where L is reparametrized as $L = UV^t$ [2]. Another recent approach is proposed in [7]. The method in [7] consists of two steps: one projection step onto low rank matrices for updating L and one thresholding step for updating S.

In many interesting situations, some further prior knowledge on the sparsity is available. One especially interesting example is the joint topology and admittance matrix estimation problem in power grids [10]. In the first example, the sparse matrix S_0 is the weighted adjacency matrix of a tree. Network topology estimation based on machine learning is currently a topic of extensive research do to its importance for practical applications. Modern methodologies such as the one presented in [10] are expected to have a great impact on practical applications. For a very interesting alternative approach based on neural networks, we refer the reader to the recent preprint [11].

In this paper, we show through simulation experiments that the a priori knowledge on the structure of the sparse matrix S_0 can be leveraged in order to perform exact reconstruction of the support of the matrix S_0 and accurate estimation of the low rank matrix L_0 in a range where the S_0 may be less sparse than required for the procedure in [7] to succeed. For this purpose, we introduce appropriate modifications of the original method of [7] in order to take into account the specific properties of the prior structures. Note that previous works on Robust PCA show that certain constraints are amenable to standard convex relaxation techniques [4]. On the other hand one important remark about using such alternating projection/thresholding procedures as the one in this paper is that prior structure, such as being the adjacency of a graph, seems very hard to specify as an additional convex constraint in (2) whereas such types of constraints are straightforward to incorporate into the procedure of [7].

The plan of the paper is as follows. In Section 2 we recall the method proposed in [7] and describe natural modifications which allow the practitioner to incorporate available prior knowledge about the structure of the problem. Simulation experiments are proposed in Section 3.2 together with results on real datasets. We end the paper with a conclusion section.

2. The algorithm

In the sequel, $P_r(A)$ denotes the best approximation of rank r of the matrix A. $HT_\beta(A)$ denotes image of A by the Hard Thresholding operator which sets to zero all components of A with absolute value less than of equal to β.

2.1. The AltProj procedure

The alternating projection/thresholding procedure described in Algorithm 1 below, was proposed and thoroughly studied in [7] for Robust PCA. In the original presentation of the method, the algorithm consisted of two loops, one for estimating the rank of L_0 and the other one for estimating the sparsity of S_0. In the presentation below, we restrict our attention to the case of a fixed rank r. The reason for this is that the presentation becomes significantly simpler with this restriction, and that estimation of r can be performed using a standard model selection procedure, e.g. BIC.

2.2. A basic modified procedure

In this section, we propose a modification of the algorithm in the previous section which takes into account any prior knowledge about a special structure of S_0. For this purpose, we assume that the structure can be mathematically described as a set $\mathcal{S}_s$ possibly parameterised by an integer s. In the case where no prior information about the sparsity is available, one can always set $\mathcal{S}_s$ to be the set of s-sparse matrices.

The projection onto $\mathcal{S}_s$ is denoted by $P_{\mathcal{S}_s}$. The distance used in the definition of this projection can be chosen so as to achieve best or at least reasonable computational efficiency.

Algorithm 1 Non-convex alternating projection based robust PCA

Require: Matrix $X \in \mathbb{R}^{d \times n}$, a convergence criterion $\epsilon > 0$ and a thresholding parameter $\beta > 0$.
Set initial threshold $\zeta = \beta\ \sigma_1(X)$, $L^{(0)} = 0$, $S^{(0)} = HT_{\zeta_0}(X - L^{(0)})$ and $l = 0$.
while $l \leqslant L = 10 \log \left(n\ \beta \|X - S^{(0)}\|_{S_2} / \epsilon\right)$ **do**
Set threshold $\zeta = \beta\ \sigma_{r+1}\left(X - S^{(l)}\right) + \frac{1}{2}\ \sigma_r\left(X - S^{(l)}\right)$.
$L^{(l+1)} = P_r(X - S^{(l)})$.
$S^{(l+1)} = HT_{\zeta}\left(X - L^{(l+1)}\right)$.
$l \leftarrow l + 1$
end while
Output $L^{(L)}$ and $S^{(L)}$.

Algorithm 2 Modified AltProj for Robust PCA with structured sparsity prior.

Require: Matrix $X \in \mathbb{R}^{d \times n}$, a maximum number of steps $L > 0$ and (possibly) a parameter structure $s > 0$.
Set $L^{(0)} = 0$, $S^{(0)} = P_{\mathcal{S}_s}(X - L^{(0)})$ and $l = 0$.
while $l \leqslant L$ **do**
$L^{(l+1)} = P_r(X - S^{(l)})$
$S^{(l+1)} = P_{\mathcal{S}_s}\left(X - L^{(l+1)}\right)$
$l \leftarrow l + 1$
end while
Output $L^{(L)}$ and $S^{(L)}$.

3. Applications to the inverse power flow problem

The admittance matrix Y can be estimated from voltage and current measurements based on additional structure. In particular, the matrix Y can be decomposed as

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}, \tag{3}$$

where Y_{11} is indexed by the observed nodes and has a tree structure and Y_{22} is indexed by the unobserved nodes. As discussed in [10], the matrix $Y_{12}Y_{11}^{-1}Y_{12}^t$ is low rank. Then, Theorem 1 in [10] states that the matrix $\bar{Y}$ defined by

$$\bar{Y} = Y_{11} - Y_{12}Y_{11}^{-1}Y_{12}^t \tag{4}$$

can be inferred from measured time series data at the observed nodes. As a result, we obtain that $\bar{Y}$ is a sum of a sparse matrix with tree structure and a low rank matrix, which is exactly the setting we study in the present paper. Note that the tree structure in not taken into account in [10]. The next sections show that incorporating this constraint in the estimation drastically improves the estimation.

3.1. Implementation details

3.1.1. Projection on trees

At each step, we have to find the closest tree to $X - L^{(l+1)}$, i.e. solve the projection problem

$$S^{(l+1)} = P_{\mathcal{S}_s}\left(X - L^{(l+1)}\right). \tag{5}$$

For the sake of simplicity, the norm with respect to which this projection will be computed will be the ℓ_1-norm. The reason for this is that the projection can therefore be computed using the *maximum spanning tree*. This in turn can be computed using the *minimum spanning tree* using the edge-wise transformation of the weights given by $w \mapsto \exp(-w)$, which ensures non-negativity of the weights.

3.1.2. Choosing the sparsity s and rank r

In the topology estimation problem, the sparsity is naturally set to $N-1$ because this is the number of edges in a spanning tree in a connected graph with N nodes.

An appropriate value of the rank r has to be chosen in order to run the Algorithm 2. One simple way to select the value of r is to use a statistical model selection based criterion such as Akaike's AIC or Schwarz's BIC.

3.2. Numerical experiments

In these numerical experiments, the matrix $\bar{Y}$ is unknown but an estimator can easily be obtained based on current and voltage measurements; see [10].

3.2.1. Random examples

We first ran a series of random examples. The adjacency matrix S_0 of a tree was drawn uniformly at random and the low rank matrix was drawn as UV^t where U (resp. V) was drawn independently with an i.i.d. standard Gaussian distribution on the space of matrices with N rows and r columns (resp. r rows and N columns). The noise was drawn from an i.i.d. Gaussian distribution $\mathcal{N}(0, 0.01)$. In all the subsequent experiments, r was chosen in the range $\{1, \ldots, 20\}$.

- In a first series of experiments, r was assumed to be known ahead of time. The results are shown in Figures 1, 2 and 3 below.
- In a second series of experiments, r was assumed to be unknown and was selected using the BIC criterion [1].

The results in Figure 1a, 2a and 3a below show that the number of edges recovered by our modified alternating minimisation procedure is much larger than the number of recovered edges recovered by a similar procedure that does not take the sparsity structure into account. Figure 1b, 2b and 3b show that the relative estimation error of the low rank matrix is smaller with our approach than when the sparsity structure is not known. Figures 1c, 2c and 3c show that the computational time is extremely small for small networks (N=100) and that the algorithm scales fairly well with the network's dimension.

The results show that significantly more edges can be recovered using the spanning tree prior. Figure 2b shows that the

In the case where the rank has to be estimated, the results were very similar for $r_0 = 1, \ldots, 8$. The reason for this is that the rank is correctly estimated as long as r_0 is less than or equal to 8 as shown in Figure 4 below. The boxplot in Figure 4 shows that for $N = 1000$, the true rank is exactly recovered for small values of r_0 and starts being recovered less accurately for larger values of r_0.

3.2.2. Results with a real grid

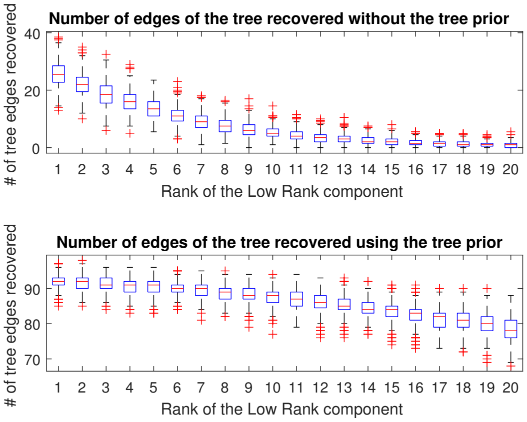
The method was tested with a distribution grid supplying a city in Spain [2], which is represented in Figure 5 below. This is an 11 kV urban network fed from a 33 kV supply point. The grid we used consists of 100 buses and 50 branches, with 2 generators supplying the medium voltage network which also feeds a low voltage network, which has 9 load buses with power demands of around a few hundred kilowatts. Since the entire grid is a tree it is therefore amenable to our method which uses tree structured sparsity. Real measurement data were acquired at 20 locations sampled at random in the grid. The true grid structure is known in full detail and can therefore serve as a baseline for assessing the methods used in the reconstruction experiments.

The algorithms were run with several prior values on the rank of the underlying matrix L_0, ranging from 1 to 20. The best fit based on BIC was $r = 5$.

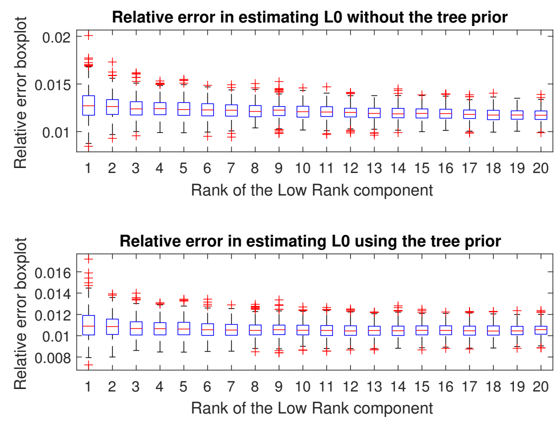
Figures 6 show the recovery results for the technique described in Algorithm 1 and the results obtained using our approach described in Algorithm 2 with BIC-based selection of the most appropriate rank (here 5 is selected).

[1] we applied the BIC as if the support of the sparse matrix was known instead of being actually estimated

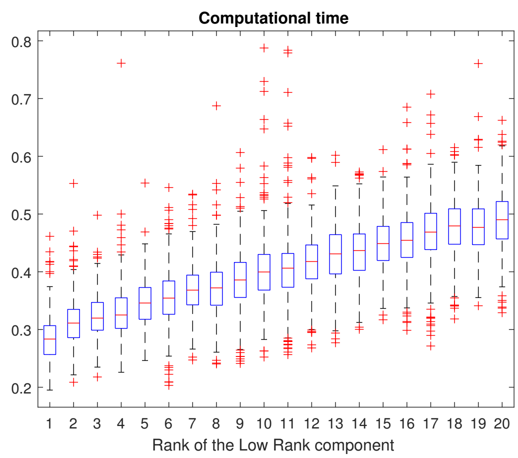
[2] the full network details are confidential



(a) Comparing recovery with $N = 100$ nodes (target 99 edges)



(b) Comparing relative estimation error for L_0 with $N = 100$ nodes



(c) Computation time in seconds with 100 nodes

Figure 1: Monte Carlo experiments with $N = 100$ and r known

4. Conclusion

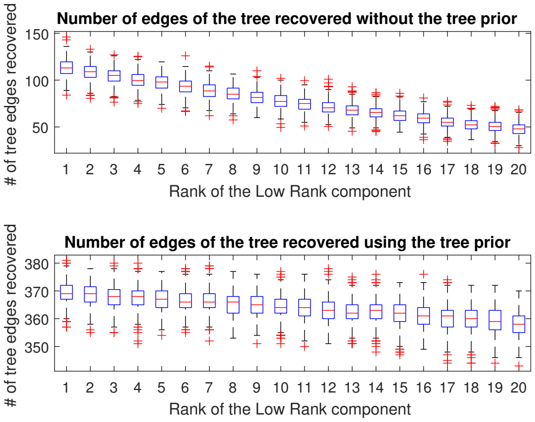
In this paper, we have presented a new efficient method for the inverse power flow problem [10] based on structured Robust PCA. We showed that great improvements can be obtained from taking into account the specific structure of the sparse matrix in the Robust PCA decomposition.

Future investigations will focus on devising a more accurate penalisation than the heuristic version of the BIC proposed in the present work.

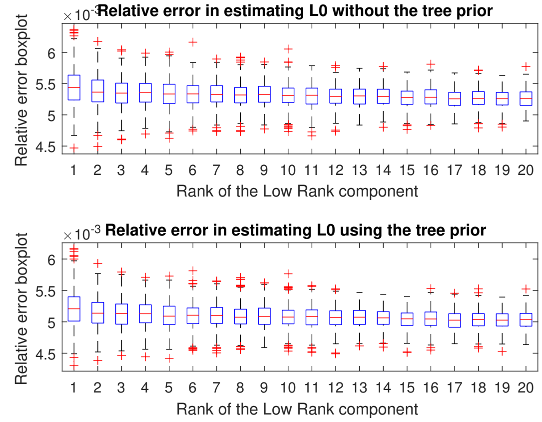
Acknowledgements: The authors would like to thank Jorge Bruna and Noemi Galan Hernandez of CIRCE, University of Zaragoza for providing network data and helping to convert it to a suitable format. The research leading to this paper has received funding from the European Union on the basis of Decision No 912/2009/EC. This work is part of the EMRP Joint Research Project ENG63. The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union.

[1] Thierry Bouwmans and El Hadi Zahzah. Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34, 2014.

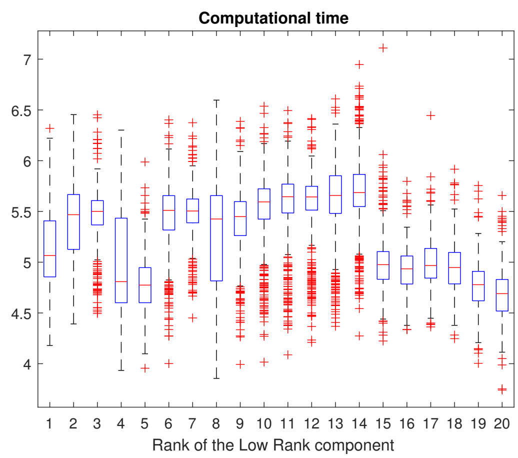
[2] Samuel Burer and Renato DC Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.



(a) Comparing recovery with $N = 400$ nodes (target 399 edges)



(b) Comparing relative estimation error for L_0 with $N = 400$ nodes



(c) Computation time in seconds with 400 nodes

Figure 2: Monte Carlo experiments with $N = 400$ and r known

[3] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

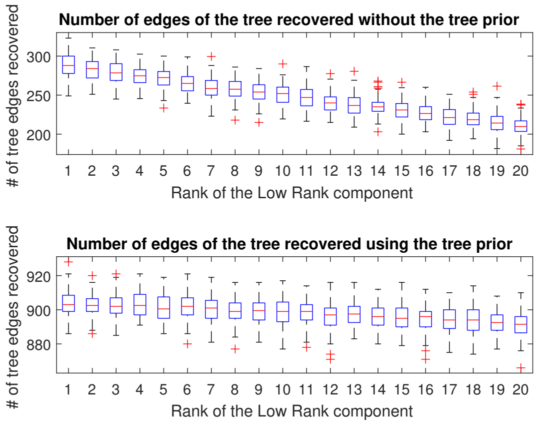
[4] Kai-Yang Chiang, Cho-Jui Hsieh, and EDU Inderjit S Dhillon. Robust principal component analysis with side information. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 2291–2299, 2016.

[5] Mia Hubert and Sanne Engelen. Robust pca and classification in biosciences. *Bioinformatics*, 20(11):1728–1736, 2004.

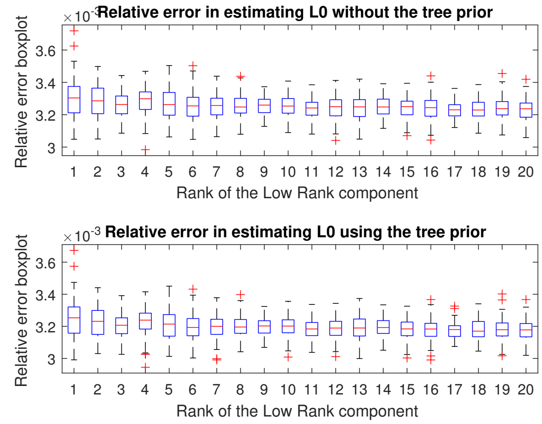
[6] Javad Lavaei and Steven H Low. Zero duality gap in optimal power flow problem. *IEEE Transactions on Power Systems*, 27(1):92–107, 2012.

[7] Praneeth Netrapalli, UN Niranjan, Sujay Sanghavi, Animashree Anandkumar, and Prateek Jain. Non-convex robust pca. In *Advances in Neural Information Processing Systems*, pages 1107–1115, 2014.

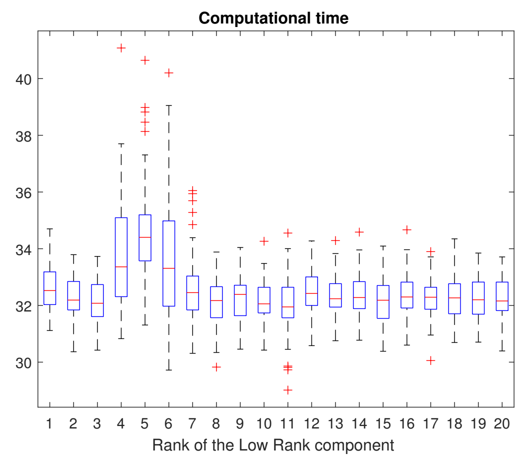
[8] M Rudelson. Probabilistic and combinatorial methods in analysis. In *Lecture notes from an NSF-CBMS Regional Research Conference at Kent State University*, 2006.



(a) Comparing recovery with $N = 1000$ nodes (target 999 edges)



(b) Comparing relative estimation error for L_0 with $N = 1000$ nodes



(c) Computation time in seconds with 1000 nodes

Figure 3: Monte Carlo experiments with $N = 1000$ and r known

[9] Konstantinos Slavakis, Georgios B Giannakis, and Gonzalo Mateos. Modeling and optimization for big data analytics:(statistical) learning tools for our era of data deluge. *IEEE Signal Processing Magazine*, 31(5):18–31, 2014.

[10] Ye Yuan, Omid Ardakanian, Steven Low, and Claire Tomlin. On the inverse power flow problem. *arXiv preprint arXiv:1610.06631*, 2016.

[11] Yue Zhao, Jianshu Chen, and H Vincent Poor. A learning-to-infer method for real-time power grid topology identification. *arXiv preprint arXiv:1710.07818*, 2017.

[12] Hao Zhu and Georgios B Giannakis. Power system nonlinear state estimation using distributed semidefinite programming. *IEEE Journal of Selected Topics in Signal Processing*, 8(6):1039–1050, 2014.

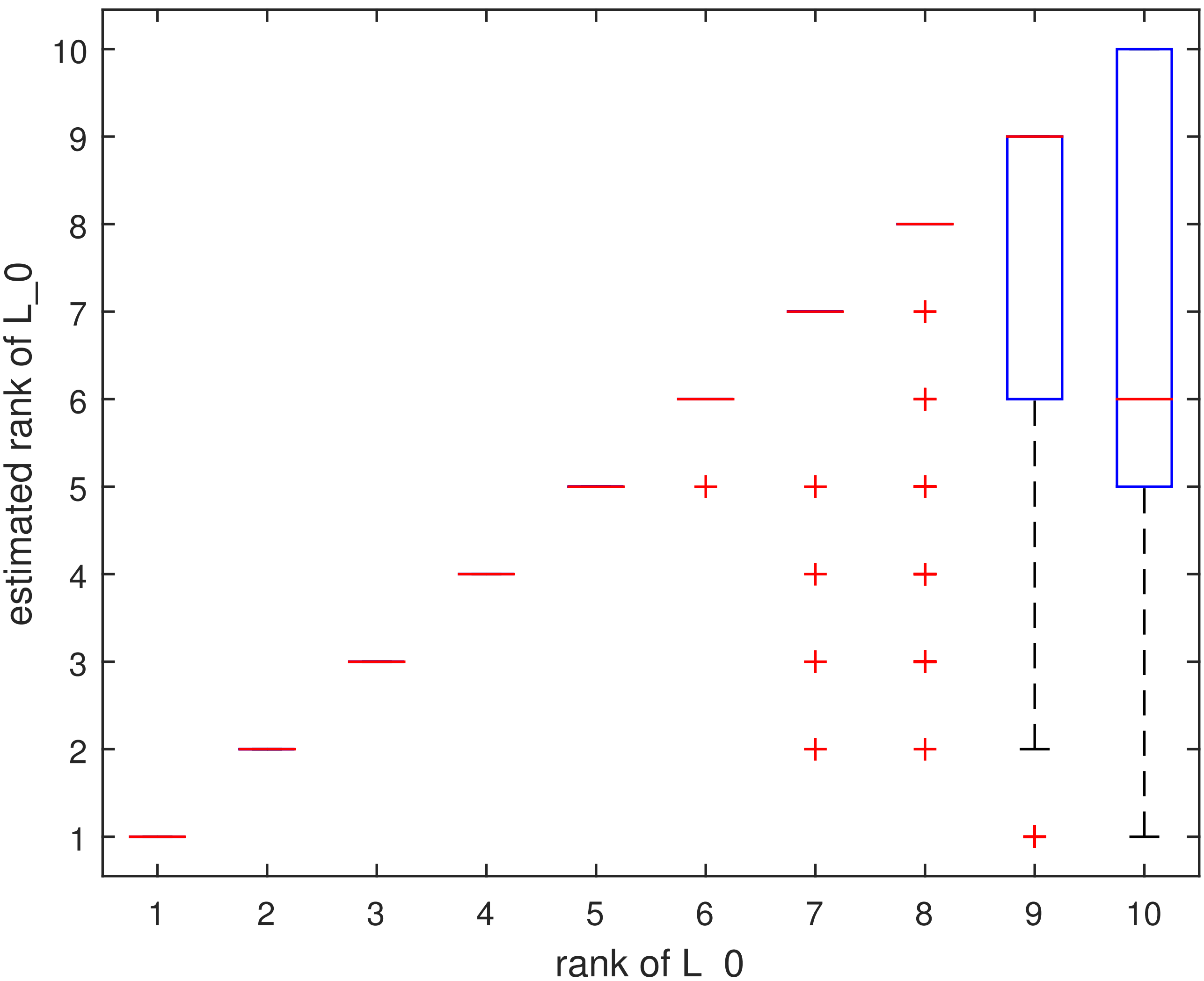


Figure 4: Estimation of the rank r_0.

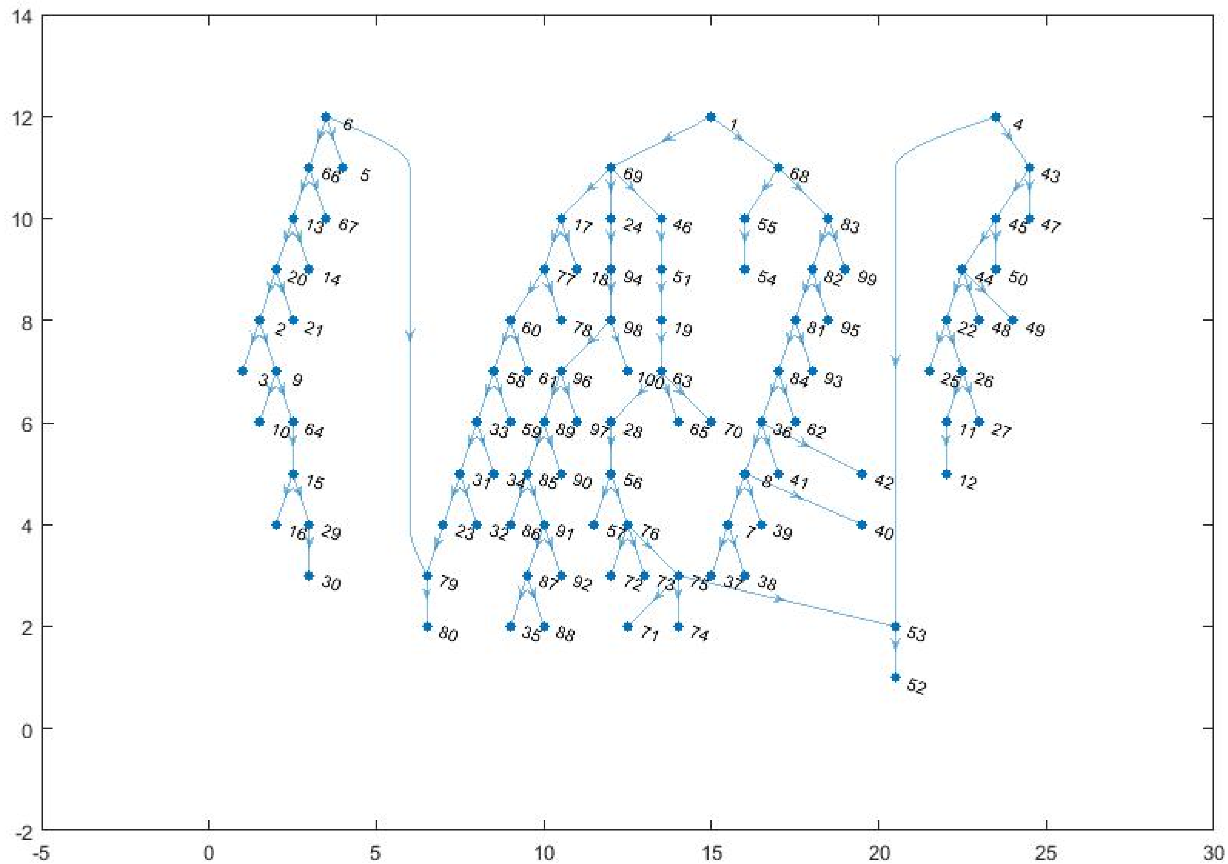


Figure 5: A grid in a spanish city

Figure 6: Comparing recovery for a real network in a Spanish city with $N = 100$ nodes (target 99 edges). The figure on the left shows the admittance matrix of the network. The figure in the middle shows the best reconstruction obtained using Algorithm 1. The figure on the right shows the reconstructed admittance matrix using Algorithm 2 and BIC-based selection of the optimal rank.