STATE ESTIMATION FOR UNOBSERVABLE DISTRIBUTION SYSTEMS VIA DEEP NEURAL NETWORKS

A Design Project Report

Presented to the School of Electrical and Computer Engineering of

Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering, Electrical and Computer Engineering

Submitted by

Jaime Luengo Rozas

M.Eng Field Advisor: Prof. Lang Tong

Degree Date: June 2018

Abstract

Master of Engineering Program

School of Electrical and Computer Engineering

Cornell University

Design Project Report

State estimation for unobservable distribution systems via deep neural networks

by Jaime Luengo Rozas

The problem of state estimation for unobservable three-phase unbalanced distribution systems is considered. A Bayesian approach is proposed that combines Bayesian inference with deep neural networks to achieve the minimum mean squared error estimation of network states for real-time applications. The proposed technique learns probability distributions of net injection from smart meter data and generates samples for training a deep neural network. Structural characteristics of the deep neural networks are investigated. Results illustrate the benefits of deep learning and robustness against distribution errors and bad data. Comparing with the pseudo measurement techniques, direct Bayesian state estimation with deep neural networks significantly outperforms existing pseudo measurement techniques, including those using neural-network based pseudo measurement approaches.

Executive Summary

The main contribution of this work is a novel application of Bayesian estimation, probability distribution learning, and deep neural network learning techniques. We demonstrate the potential of such techniques for large distribution systems and provide insights into the architectural characteristics of deep neural networks.

Bayesian inference is based on the probabilistic modeling of system states. Given the highly stochastic nature of DER, modeling voltage phasors as random variables is natural. The key challenges of developing Bayesian state estimation, are (i) the need to learn the underlying probability distributions that define the system states and measurements; (ii) the complexity of computing the conditional mean of the system states.

The main idea of the proposed technique is that historical measurements are used to learn the probability distributions of the net-injection. The learned distributions are used to generate samples to train a deep neural network to approximate the minimum mean squared error (MMSE) estimator of the system states. A stochastic gradient descent algorithm with early stopping is used in the training process. In real-time applications, the neural network directly computes the MMSE estimates with linear complexity. In contrast, WLS types of state estimators have the complexity roughly of the order of $O(N^3)$ per iteration.

Numerical results demonstrate several interesting features of the proposed approach. First, deep learning seems to be essential. Whereas existing neural network state estimation techniques typically use a flat network involving one or two layers, our results show that, for the tested 85 and 141 bus networks, a neural network of 10 layers or more can provide accurate estimates, achieving mean squared error (MSE) per bus at the level of 10^{-2} p.u. on test data sets. Such estimation errors are well within the typically required accuracy for voltage estimates. In contrast, the WLS methods with pseudo-measurements techniques have errors several orders of magnitude higher.

Second, among various training techniques, the stochastic gradient decent algorithm with early stopping is the most effective for the cases tested. Finally, simulation results show that the proposed approach exhibits a promising level of robustness against estimation errors in probability distributions and sampling errors in training.

Acknowledgements

Thank you Iberdrola Foundation for your support and making this year possible. Thank you Prof. Lang Tong for your help and advice throughout the M.Eng. and this project, in particular. Thank you Kursat Mestav for your collaboration.

Contents

List of Figures

Abbreviations

CDF	**C**umulative **D**istribution **F**unction
DER	**D**istributed **E**nergy **R**esources
DNN	**D**eep **N**eural **N**etwork
LTC	**L**oad **T**ap **C**hanger
MMSE	**M**inimum **M**ean **S**quare **E**rror
MSE	**M**ean **S**quare **E**rror
PMU	**P**hasor **M**easurement **U**nit
SCADA	**S**upervisory **C**ontrol and **D**ata **A**cquisition
WLS	**W**eighed **L**east **S**quares

Chapter 1

Design Problem

1.1 Introduction

We consider the problem of state estimation for distribution systems that have limited measurements such that they are *unobservable* [1]. The states of such a system cannot be determined uniquely from the available measurements even when there is no measurement noise, and techniques such as the weighted least squared (WLS) method [1] fail in general. A standard remedy is to use the so-called *pseudo-measurements* based on interpolated measurements or historical data. Such techniques are ad-hoc and in general suboptimal.

Transmission networks are typically observable because the supervisory control and data acquisition (SCADA) system provides redundant measurements in the form of branch and injection power flows measurements including some current magnitude measurements. Phasor measurement units (PMUs) that measure voltage phasors directly can also be incorporated. Distribution systems, in contrast, are in general unobservable for lack of measurements. Because real-time dispatch and control in distribution systems have so far not been needed, there is no compelling case to estimate and tract distribution system states at a fast time scale.

The present distribution systems are not well metered and in general unobservable. However, there have been compelling cases made for distribution system state estimation due to the rising presence of distributed energy resources (DER) in distribution systems [2]. To unlock the full potential of DER, a modernization of the distribution system is necessary to provide tighter control of power flow in real-time operations, which requires effective state estimation.

An essential barrier to state estimation for real-time control is unobservability. Although smart meters at the edge of the network have been deployed progressively, these type of measurements are typically at a much slower time scale, incompatible with the more rapid changes of DER such as solar generations. Realizing state estimation for real-time operation in distribution systems, therefore, requires a fundamentally different approach from that used in transmission systems — one that overcomes the difficulty of lack of measurements.

Realizing state estimation for real-time operation in distribution systems, therefore, requires a fundamentally different approach from that used in transmission systems, one that overcomes the observability constraint. To this end, we pursue a Bayesian inference approach where the system states (voltage phasors) are modeled as random. Given the highly stochastic nature of solar generation, such a Bayesian model are natural. Unlike the *point estimation* techniques such as the WLS method that requires rich measurements to make the system observable, a Bayesian estimator exploits the distributional dependencies of the measurements on the system states and improves the estimate based on prior distribution with available measurements, even if there are only a few such measurements.

1.2 System Requirements

Developing a state estimator for the distribution system, given the necessary hardware, requires the interface of different software modules to deal with the power flow analysis, state estimation and digital communication. In this theoretical work the focus was on the first two.

Given the challenges explained above the main requirements of our system were:

1. Robustness against scarce historical data
2. Robustness against bad data
3. Running time under $O(N^3)$

4. Independent of the distribution grid model

5. Estimation Error under $10^{-}2$ p.u.

The motivation for the first two points is the existing measurement infrastructure available. This constraints the amount of existing data for each node and poses the challenge how to deal with the situation certain node measurement not being received in the control center for several time intervals. The running time benchmark is stated by the WLS method. The last two are fruit of the nature of the distribution system. There is a need of an estimator that does not depend on the meshed, radial, single-phase or three-phase topology of a system. Finally, in order to perform reactive power control, voltage regulation, or any other desired control, a level of $10^{-}2$ pu would be the minimal to at least operate a LTC transformer, or an inverter under a low resolution digital control.

1.3 Previous Work

State estimation based on deterministic models of states has been extensively studied. See [1] and references therein. Existing Bayesian techniques that model states as random are less common even though the idea was already proposed in the seminal work of Schweppe [3] where (extended) Kalman filtering techniques were proposed.

In the context of state estimation for unobservable distribution systems, Bayesian methods for state estimation can be classified into two categories: Bayesian pseudo-measurements [4–8] and Bayesian state estimation [9–23]; the former uses probability distributions to generate pseudo-measurements so that conventional (point) estimation techniques such as WLS can be applied. Such hybrid techniques are suboptimal but can be easily incorporated in the conventional state estimation methods. The latter type uses distribution information explicitly and aims to minimize the mean squared error. These techniques vary in how the conditional mean of the system states are computed. The method proposed in this project falls into this latter category.

Direct Bayesian state estimation requires the computation of the conditional mean of the state variables. One approach is based on a graphical model of the distribution system from which belief propagation techniques are used to generate state estimates [9–12]. These techniques require a dependency graph of the system states and explicit forms of probability distributions. Another approach is based on a linear approximation of the AC power flow [13]. The proposed approach belongs to the class of Monte Carlo techniques in which samples are generated and empirical conditional means are computed. The main difference between our approach and existing techniques [14–18] is the way conditional means are computed in real-time. Instead of using Monte carlo sampling to compute the conditional mean directly as in [15, 17, 18], Monte carlo sampling is used to train a neural network that, in real-time, computes the MMSE estimate directly from the measurements.

Neural networks have been proposed for state estimation as early as [19]. Different architectures of neural networks have been considered: parallel distributed processing in [20], auto-encoder in [21]. In [24] a two staged approach with artificial neural networks is used to eliminate bad data and estimate states. A radial basis function network is used with in [22]. Although not casted as a Bayesian state estimation, the approach in [23] appears to be quite close to ours. In [23], a neural network with two hidden layers is used to estimate states directly using load bus measurements as the input of the neural network. The distributions of power injections are assumed. In our approach, the distribution of net injections is learned from smart meter measurements and an architecture employing sophisticated deep learning techniques is used.

Chapter 2

Design and Implementation

2.1 Range of Solutions

The main idea of the proposed technique, as illustrated in a schematic diagram in Fig. 2.1, is that historial measurements are used to learn the probability distributions of the net-injection. The learned distributions are used to generate samples to train a deep neural network to approximate the minimum mean squared error (MMSE) estimator of the system states. A stochastic gradient descent algorithm with early stopping is used in the training process. In real-time applications, the neural network directly computes the MMSE estimates with linear complexity of $O(N)$ where N is the size of the network. In contrast, WLS types of state estimators have the complexity roughly of the order of $O(N^3)$ per iteration .Considering a network of N buses, for the weighted least squares state estimator, a direct implementation of Newton's iteration requires the computation of the inverse of Hessian matrix, which in general requires $O(N^3)$ computations per-iteration. The total number of iterations depends on the accuracy level ϵ, typically in the order of $O(\epsilon^-2)$[25]. The neural network, on the other hand, computes the N-dimensional state estimate directly by passing measurements through a fixed neural network without iteration. The computation cost is roughly $O(KN)$ where K corresponds roughly the number of neurons.

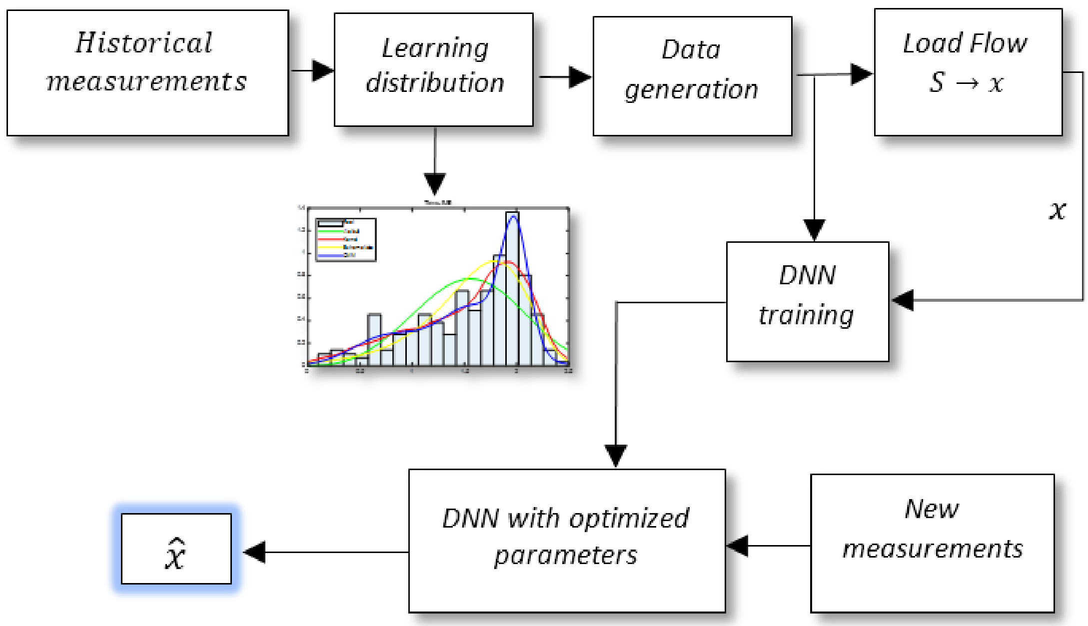

FIGURE 2.1: Block diagram of the method.

2.2 Network and Measurement Models

The distribution system is described by a graph $\mathcal{N} = \{\mathcal{V}, \mathcal{E}\}$ where $\mathcal{V} = \{1, \cdots, N\}$ is the set of buses and $\mathcal{E}$ the set of network branches. We assume that node 1 is the slack bus representing the point of common coupling (PCC), where the distribution network is connected to the main grid.

We assume an unbalanced three phase distribution system. The three-phase voltage phasors at bus i is a complex column vector $x_i = [x_i^1, x_i^2, x_i^3]^\intercal$, where the superscripts are phase indices and $x_i^k = V_i^k \angle \theta_i^k$, where V_i^k is the voltage magnitude and θ_i^k is the phase angle for the state variable at phase k of bus i. The overall system state $x = [x_1, \cdots, x_N]^\intercal$ is the column vector consisting of voltage phasors at all buses.

The vector of measurements z is a function of the state x, and measurement error e modeled by

$$z = h(x) + e, \quad S = g(x), \tag{2.1}$$

where $h(x)$ is the measurement equation, S the vector of power injections, and $g(x)$ the power flow equations. When the system is unobservable, each measurement (even when $e = 0$) is associated with a manifold of states.

Different configurations of measurements can be assumed. Possible measurements include, for each phase $k \in \{1, 2, 3\}$,

P_i^k Active power injection at node i

Q_i^k Reactive power injection at node i

P_{ij}^k Active power flow from node i to j

Q_{ij}^k Reactive power flow from node i to j

I_{ij}^k Current from node i to j

Power flow equations are used to relate the state variables with the measurements:

$$P_i^k = V_i^k \sum_{l=1}^{3} \sum_{j=1}^{n} V_i^l [G_{ij}^{kl} \cos(\theta_i^k - \theta_j^l) + B_{ij}^{kl} \sin(\theta_i^k - \theta_j^l)] \tag{2.2}$$

$$Q_i^k = V_i^k \sum_{l=1}^{3} \sum_{j=1}^{n} V_i^l [G_{ij}^{kl} \sin(\theta_i^k - \theta_j^l) + B_{ij}^{kl} \cos(\theta_i^k - \theta_j^l)] \tag{2.3}$$

where G_{ij}^{kl} and B_{ij}^{kl} are the conductance and susceptance between node i and j from phase k to l. The LHS is used to define the three-phase net power injection vector at each node S_i whose elements are $S_i^k = P_i^k + jQ_i^k$. .In absence of measurement noise, given the set of active and reactive power injections, the above equation can be solved to obtain the system states, which in turn give the branch power flows:

$$\begin{aligned} P_{ij}^k = & V_i^k \sum_{l=1}^{3} V_i^l [G_{ij}^{kl} \cos(\theta_i^k - \theta_j^l) + B_{ij}^{kl} \sin(\theta_i^k - \theta_j^l)] \\ & -V_i^k \sum_{l=1}^{3} V_j^l [G_{ij}^{kl} \cos(\theta_i^k - \theta_j^l) + B_{ij}^{pq} \sin(\theta_i^k - \theta_j^l)] \end{aligned} \tag{2.4}$$

$$\begin{aligned} Q_{ij}^k = & -V_i^k \sum_{l=1}^{3} V_i^l [G_{ij}^{kl} \sin(\theta_i^k - \theta_j^l) - B_{ij}^{kl} \cos(\theta_i^k - \theta_j^l)] \\ & - V_i^k \sum_{l=1}^{3} V_j^l [G_{ij}^{kl} \sin(\theta_i^k - \theta_j^l) - B_{ij}^{kl} \cos(\theta_i^k - \theta_j^l)] \end{aligned} \tag{2.5}$$

The currents are formulated as $I_{ij}^k = \mathrm{Re}(I_{ij}^k) + j\,\mathrm{Im}(I_{ij}^k)$, where Re and Im denote the real and imaginary components respectively. These can be also calculated from the system states:

$$\begin{aligned}\mathrm{Re}(I_{ij}^k) = &\sum_{l=1}^{3} V_i^l [G_{ij}^{kl}\sin(\theta_i^l) - B_{ij}^{kl}\cos(\theta_i^l)]\\ &- \sum_{l=1}^{3} V_j^l [G_{ij}^{kl}\sin(\theta_j^l) - B_{ij}^{kl}\cos(\theta_j^l)]\end{aligned} \tag{2.6}$$

$$\begin{aligned}\mathrm{Im}(I_{ij}^k) = &-\sum_{l=1}^{3} V_i^l [G_{ij}^{kl}\sin(\theta_i^l) + B_{ij}^{kl} cos(\theta_i^l)]\\ &- \sum_{l=1}^{3} V_j^l [G_{ij}^{kl}\sin(\theta_j^l) + B_{ij}^{kl}\cos(\theta_j^l)]\end{aligned} \tag{2.7}$$

The energy consumption measurements are modeled as the accumulated values of the power injections P_i^k.

2.3 Bayesian Solution via Deep neural Networks

Bayesian state estimation starts with defining the probability space that specifies the joint distribution of the measurement z and state x. For a distribution system with stochastic injections, the probability space is defined by the independent random vector S of net-injection and measurement error e. From the power flow equations (2)-(3), S determines the system state x (in the forms of (V, θ)) , which in turn determines measurement z. Thus the joint distribution of $F_{S,e}$ specifies the joint distribution $F_{x,z}$.

A Bayesian estimator $\hat{x}(z)$ of x is a measurable function of z. The MMSE estimator is given by

$$\hat{x}^*(z) = \arg\min_{\hat{x}}(||x - \hat{x}(z)||^2) = (x|z). \tag{2.8}$$

Unfortunately, the functional form of the MMSE estimator is highly complex. One approach is to use a Monte Carlo technique that, for a given measurement z, computes samples from generated conditional distribution $F_{x|z}$. For real-time applications, such an approach is difficult to implement. We propose next an alternative that approximates $\hat{x}^*$ using a deep neural network.

2.3.1 Neural Network Approximation

Fig 2.2 shows a K layer network, where the (first) input layer receives the input z, and the output layer produces estimates of x.

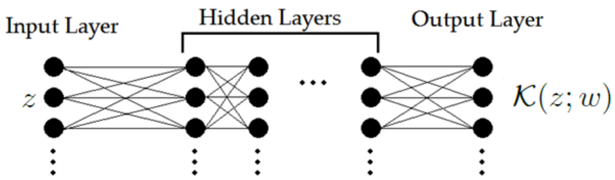

FIGURE 2.2: Multi-layer Perceptron Model.

Each middle layer has a set of neurons connecting the outputs of the neurons from the previous layer and producing outputs for the next layer using a nonlinear function, in our case the hyperbolic tangent function. For the ith neuron in the interior layer j, its output z_i^{j+1} is given by

$$z_{i,j} = \frac{\exp(u_{i,j}) - \exp(-u_{i,j})}{\exp(u_{i,j}) + \exp(-u_{i,j})}, \quad u_{i,j} = w_{i,j}^{(0)} + \sum_k w_{i,j}^{(k)} z_{k,j-1},$$

where $\{w_{ij}^k\}$ is the set of weights associated with neuron (i,j). For the Kth (output) layer, the output of neuron j is an affine function of the output of the previous layer. Specifically, an estimate of a state variable given by

$$z_{i,K} = w_{i,K}^{(0)} + \sum_k w_{i,K}^{(k)} z_{k,K-1}.$$

The outputs collectively produce an estimate of the state x.

In approximating the MMSE estimator, the neural network weight parameter w is set to minimize the MSE of its estimate

$$w^* = \arg\min_w(||x - (z;w))||^2). \tag{2.9}$$

The above optimization is only conceptual, however, because the expectation operator requires explicit joint distribution of z and x. We show next how w^* can be obtained though a neural network training process.

2.3.2 Bad Data Detection and Mitigation

Bad data are anomalies in data collection that are common in transmission systems and potentially more significant in distribution systems. Bad data detection can be implemented using a generalized likelihood ratio test that has α as the probability of type I (false positive) error and β the probability of type II (false negative) error. When the bad data test is positive, the measurement used in the input of the neural network should be replaced by the mean of the prior distribution. This implies that α percentage good measurements are replaced by the mean of the prior distribution whereas β percentage of the bad data are missed, which introduces statistical deviation of the nominal measurements. Our numerical results indicate that Bayesian state estimates via neural networks is quite robust to bad data.

2.4 Training Neural Network

2.4.1 Learning Net Injection Distributions

In our proposed method, the first step is to estimate net injection distributions using the historical load data. These distributions strongly depend on the days of the week, seasons and geographical locations. We cluster the historical data with common attributes.

We consider parametric models for net injection distributions. Among commonly used models, the Gaussian mixture model appears to be the most flexible and accurate. Indeed, Gaussian mixtures have been used to model load distributions in [7]. From historical data, the maximum likelihood method is used to estimate parameters of the Gaussian mixture model.

2.4.2 Training the Neural Network

To train the neural network, we need to generate state and measurement training samples $= \{(x[k], z[k])\}$. To this end, we draw net-injection samples from the learned net-injection distributions. In particular, given an injection sample $S[k]$, the power flow equations (2)-(3) are used to solve for system state $x[k]$ and measurement $z[k]$. Two additional sets of samples are generated independently, ˘ for testing performance, ˜ for validation.

To approximate the MMSE state estimator, the weight of the neural network is chosen to minimize the empirical risk

$$\begin{aligned} w^* &= \arg\min_w L(w;\ ) \\ &= \arg\min_w \frac{1}{||} \sum_{k:(x[k],z[k])\in} ||x[k] - (z[k];w)||^2 \end{aligned} \tag{2.10}$$

The empirical risk minimization problem above is well studied for deep learning problems. For the state estimation at hand, the stochastic gradient descent algorithm [26] appears to offer the training-generalization tradeoff. In particular, the adaptive moment (ADAM) technique [27] designed for nonstationary objectives and noisy measurements, appears to be most appropriate for the considered application.

A characteristic of deep learning is overfitting, which means that the number of neurons (weight parameters) tends to be comparable or smaller than the available training samples. A key component in deep learning, therefore, is a way to regularize the optimization of (5). Standard techniques including L_1 regularization, dropout regularization [28], and early stopping [29] are tested in our numerical study.

The early stopping technique uses the validation data set ˜ to determine the stopping time for the gradient descent. These techniques are tested in our numerical study.

To decide on the optimal number of neurons in each layer of the network, a pruning algorithm based on hierarchical clustering is used. Pruning is to decrease number of neurons in a neural network which is studied in [30–32]. Hierarchical clustering is used on input of neurons for the training samples to find the highly correlated neurons. These correlated neurons are deducted to propose a simpler network which is faster and less likely to overfit.

Chapter 3

Results

3.1 Network, data and benchmarks

The simulations were performed in a 85 and 141 bus systems defined in the MATPOWER toolbox [33][1]. The results of the two systems were similar; only the results for the 85 bus system are reported here. To model the relatively high penetration of DER, two-thirds of the buses were chosen to have solar PV attached to the load.3.1 represents the previous system.

Smart meters were used to measure energy consumption of every bus with 3 types of measurement accuracy: 0.015, 0.0175 and 0.02 pu. Current magnitude meters were placed in 20% of distribution system branches to measure the current magnitude and a SCADA meter was placed at the slack bus to measure complex power injection, both with a measurement accuracy of 0.01 pu. A uniform distribution was used to model the error of the meters, where their accuracy is used as positive and negative bound of the distribution. Smart meters had a sampling rate 30 times slower than the SCADA and current magnitude measurements.

We used the data sets from the Pecan Street collection[2] for distribution learning and testing. The data set was split into 21st May to 21st September 2015 for training and same dates of 2016 for test, which coincided with a summer patterned net consumption.

[1] The 85 and 141 bus systems are single phase systems in MATPOWER.

[2] `http://www.pecanstreet.org/`

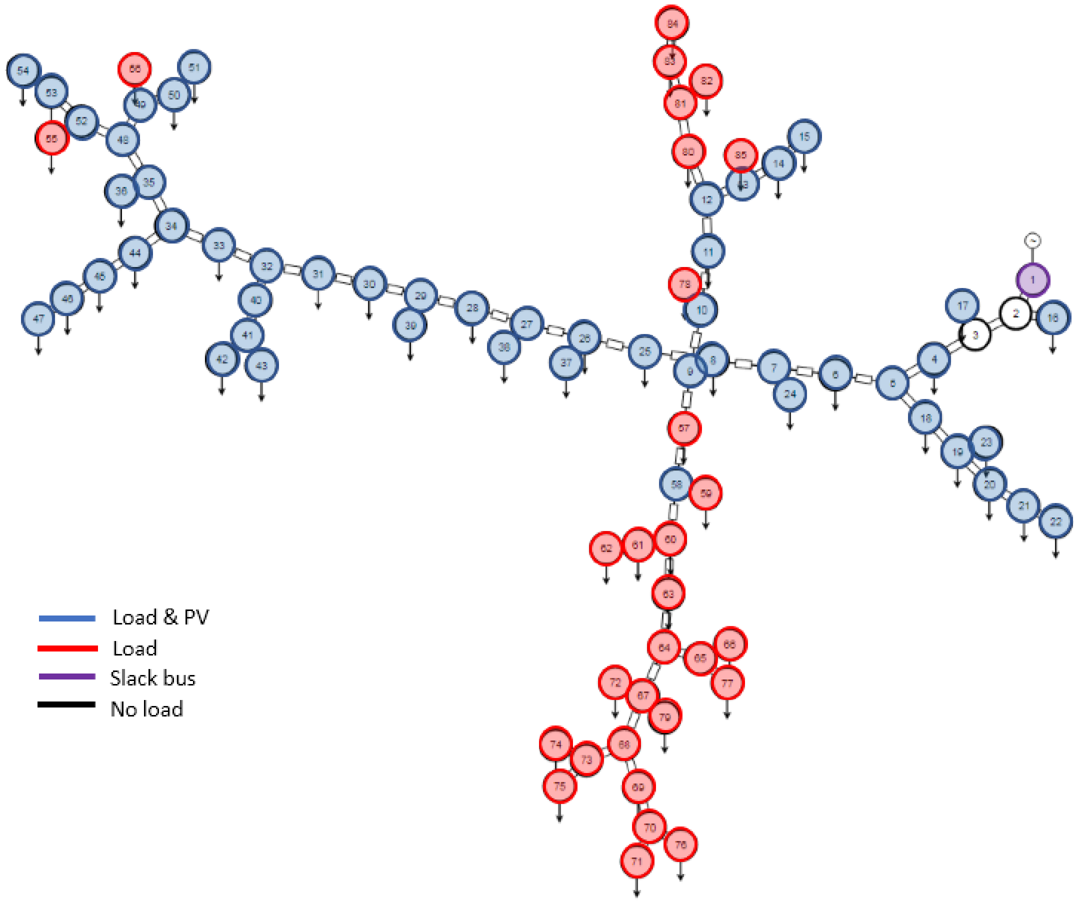

FIGURE 3.1: 85 bus system.

The performance was evaluated based on the per-node average squared error (ASE) of state estimate using test data set S defined as:

$$\text{ASE} = \frac{1}{MN} \sum_{k} ||\hat{x}[k] - x[k]||^2 \tag{3.1}$$

where M is number of cases, N is number of nodes.

The method in [23] was implemented. Given the relatively scarce amount of historical data, the results were not comparable with the proposed approach since its ASE was several orders of magnitude worse. Our method was then compared with WLS methods with two types of pseudo-measurements of net power injections in the literature [5, 6] : i) averaging the last energy consumption measurement over the number of samples; ii) using a neural network whose inputs are the last energy consumption vector and output the net power injection measurements for each sample.

3.2 Learning Distributions

The first attempt to learn the distributions was using bootstrap methods to find the best estimates of the parametric models. However, the small sample size available hindered the convergence of the maximum likelihood methods. This size constraint resulted in a empirical cdf too discretized, as shown in 3.2. Hence we turned to a more ad-hoc method for our system. Different models were proposed and for each node and hour they were tested what percentage of all these tests they fell inside the confidence intervals of the cdf. The discrete computation of the cdf added a new varaible to the test, that is the number of steps in the x-axis that the distributions were allowed to be out the intervals to be considered errors. As shown in 3.2, the distributions that were more accurate in these tests were the Kernel and Gaussian Mixture models.

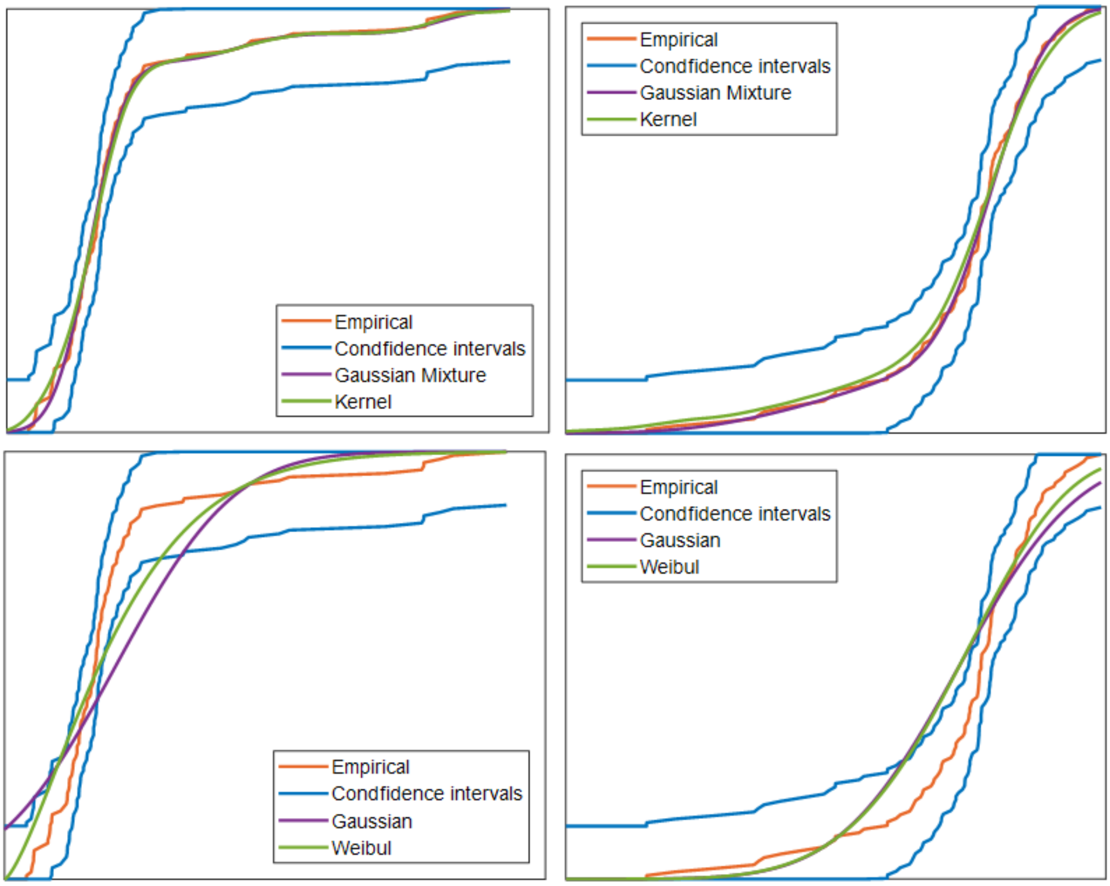

FIGURE 3.2: Cumulative distribution function (cdf) of consumption (left column), solar generation (right column) for models that lie inside (upper row) and outside (lower row) the confidence intervals of the empirical cdf.

3.3 shows how the Gaussian Mixture of 3 components was in the end the least prone to error against the Kernel. Hence, a Gaussian Mixture Model of 3 components was selected as the best fit for both random variables, active power consumption U_i and solar power generation V_i, where i represents each node.

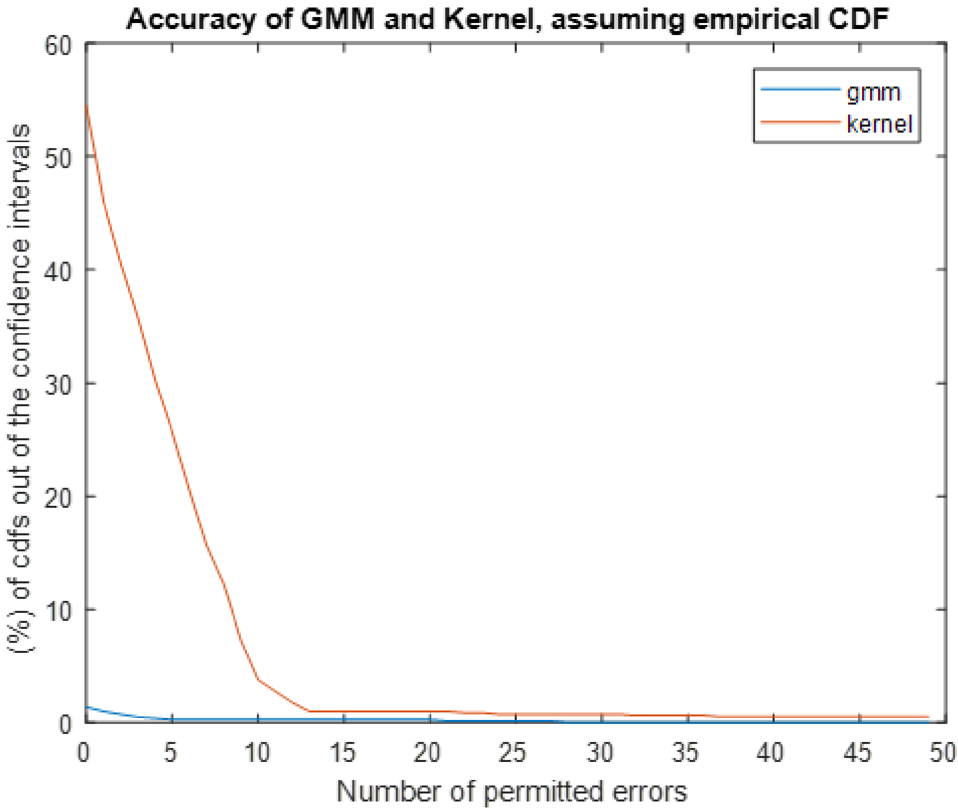

Figure 3.3: Error comparison of Kernel and Gaussian Mixture Model.

From these distributions we generated the training and validation set comprised of 600 and 300 cases respectively. The validation set is used for early stopping. We assumed a power factor of 0.86 giving a ratio of 0.6 between active and reactive power injections. Therefore, each element of the net complex power injections vector $S[k]$ was generated as $S_i[k] = (u_i[k] - v_i[k]) + j(0.6u_i[k])$where $u_i[k]$ and $v_i[k]$ are realizations of U_i and V_i.

3.3 Neural Network Architecture

The first experiment consisted of exploring which network architectures would be more suitable for the Distribution System State Estimation problem. In Fig 3.5 the total number of neurons was fixed and the ASE was calculated at different number of hidden layers. Three total number of neurons were specified.

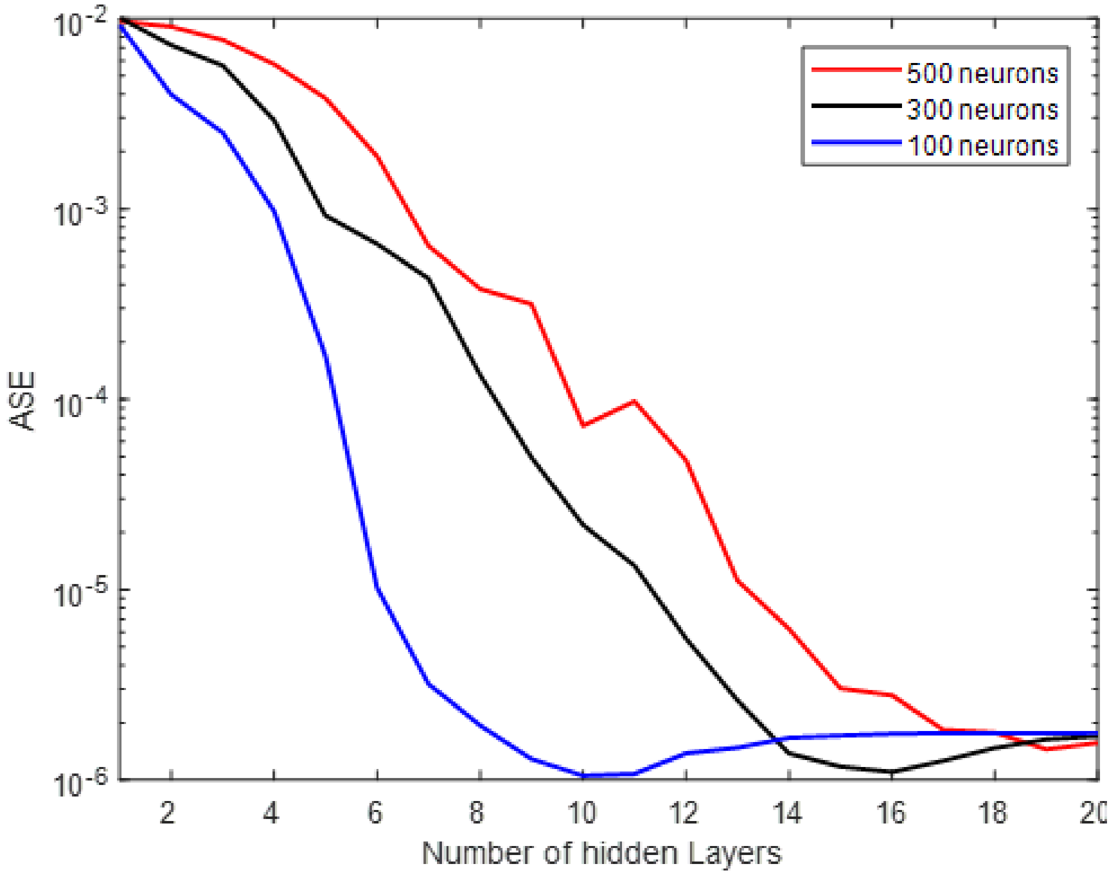

Figure 3.4: NN Architecture - Total number of neurons fixed.

The results suggest a 'deeper' neural network architecture performs better than another with a fewer number of hidden layers, as long as the different methods presented previously are followed to prevent overfitting.

3.3.1 Performance

We conducted simulations on the performance of the proposed state estimation technique for each hour of the day. The mean net consumption is presented as a reference, although it is shown unitless to focus on the ASE magnitude. For every hour a different neural network was trained. Fig **??** shows that the ASE of deep neural networks was clearly better than the first pseudo-measurements method. Comparison between the second pseudo-measurements method shows that training deep neural networks to estimate the states directly is more efficient than estimating pseudo-measurements.

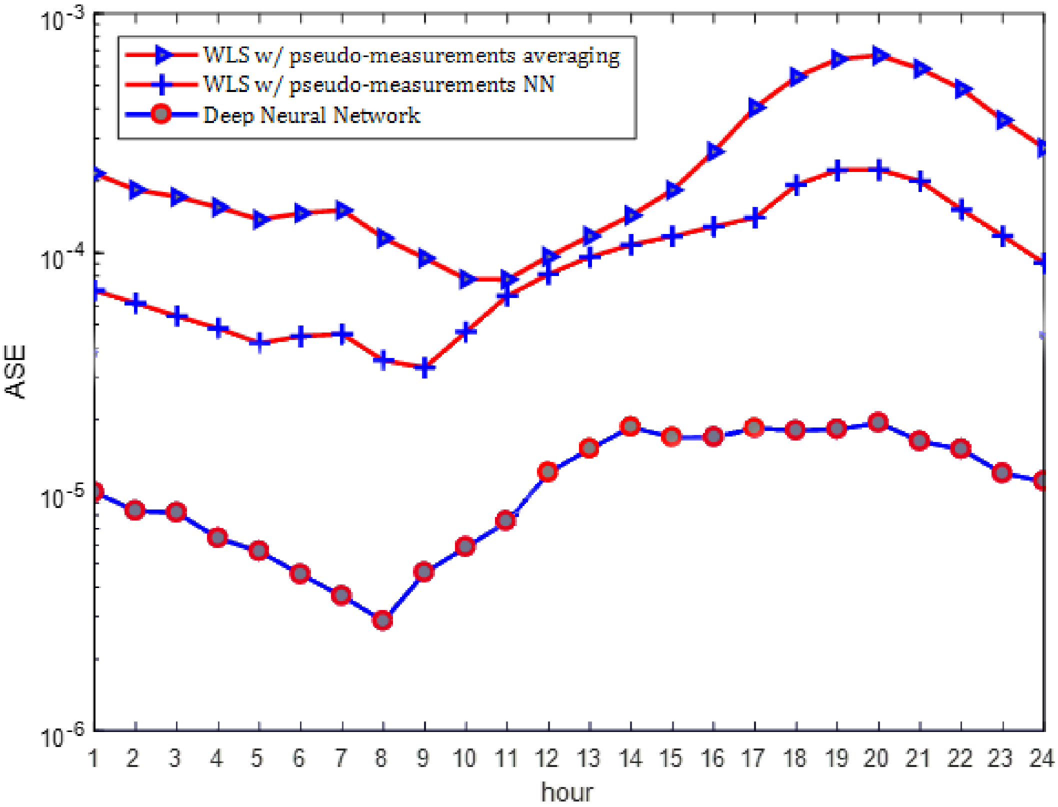

FIGURE 3.5: Hourly simulation

For testing the robustness of our method against bad data, a generalized likelihood ratio test with $\alpha = \beta = 0.05$ was implemented.Bad data measurements are assumed to have five times higher error variance than that of the uncorrupted. In the simulation the percentage of bad data is increased in steps of 10% from 0 to 50%. The outcome of the experiment shows that the DNN ASE stayed within 5% when the number of bad data measurements increased, while PM methods' ASE increased 50% from the initial uncorrupted case. We have also conducted comparison studies between the proposed technique and WLS methods with increasing SCADA complex power measurements. With additional complex power measurements, the network became observable. As expected, the WLS technique gradually outperform the Bayesian methods. Details of these results will be reported in separate work.

Chapter 4

Conclusion

This project presents a Bayesian approach using deep neural networks for state estimation in unobservable distribution systems. To benefit from deep neural network architectures, this approach learns net power injection distributions from historical measurements and generates more training samples. Furthermore, stochastic optimization methods and regularization methods are used to avoid overfitting. This method is computationally efficient and robust against bad data, variation of net consumption distributions and high penetration of DERs; which makes it suitable for real-time operation in the distribution system. In a further work, a 3-phase system should be included in the simulations as a proof concept for distribution model independence.

Bibliography

[1] Ali Abur and Antonio G. Expósito. *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

[2] A. Primadianto and C. N. Lu. A review on distribution system state estimation. *IEEE Transactions on Power Systems*, 32(5):3875–3883, Sept 2017. ISSN 0885-8950. doi: 10.1109/TPWRS.2016.2632156.

[3] F. C. Schweppe, J. Wildes, and D. P. Rom. Power system static state estimation, Parts I, II, III. *IEEE Tran. on Power Appar. & Syst.*, PAS-89:120–135, 1970.

[4] A. K. Ghosh, D. L. Lubkeman, M. J. Downey, and R. H. Jones. Distribution circuit state estimation using a probabilistic approach. *IEEE Transactions on Power Systems*, 12(1):45–51, Feb 1997. ISSN 0885-8950. doi: 10.1109/59.574922.

[5] Andrea Bernieri, Giovanni Betta, Consolatina Liguori, and Arturo Losi. Neural networks and pseudo- measurements for real-time monitoring of distribution systems. *IEEE Transactions on Instrumentation on Measurements*, 1996.

[6] Efthymios Manitsas, Ravindra Singh, Bikash Pal, and Goran Strbac. Modelling of pseudo-measurements for distribution system state estimation. *SmartGrids for Distribution*, 2008.

[7] R. Singh, B. C. Pal, and R. A. Jabr. Distribution system state estimation through gaussian mixture model of the load as pseudo-measurement. *IET Generation, Transmission Distribution*, 4(1), 2010. ISSN 1751-8687. doi: 10.1049/iet-gtd.2009.0167.

[8] Georgia Pieri, Markos Asprou, and Elias Kyriakides. Load pseudomeasurements in distribution system state estimation. *IEEE PowerTech*, 2015.

[9] Y. Hu, A. Kuh, T. Yang, and A. Kavcic. A belief propagation based power distribution system state estimator. *IEEE Computational Intelligence Magazine*, 6(3): 36–46, Aug 2011. ISSN 1556-603X. doi: 10.1109/MCI.2011.941589.

[10] P. Chavali and A. Nehorai. Distributed power system state estimation using factor graphs. *IEEE Transactions on Signal Processing*, 63(11):2864–2876, June 2015. ISSN 1053-587X. doi: 10.1109/TSP.2015.2413297.

[11] Mirsad Cosovic and Dejan Vukobratovic. Distributed gauss-newton method for ac stateestimation using belief propagation. *IEEE International Conference on Smart Grid Communications*, 2017.

[12] Chenhui Yin, Dechang Yang, and Xiaoyu Zhao. State estimation of active distribution system based on the factor graph analysis and belief propagation algorithm. *IEEE International Conference on Environment and Electrical Engineering and IEEE Industrial and Commercial Power Systems Europe*, 2017.

[13] L. Schenato, G. Barchi, D. Macii, R. Arghandeh, K. Poolla, and A. Von Meier. Bayesian linear state estimation using smart metersand pmus measurements in distribution grids. *IEEE International Conference on Smart Grid Communications*, 2014.

[14] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F - Radar and Signal Processing*, 140(2):107–113, April 1993. ISSN 0956-375X. doi: 10.1049/ip-f-2.1993.0015.

[15] Kianoush Emami, Tyrone Fernando, Herbert Ho-Ching Iu, Hieu Trinh, and Kit Po Wong. Particle filter approach to dynamic stateestimation of generators in power systems. *IEEE Transactions on Power Systems*, 2015.

[16] A. Jahic, T. Konjic, and A. Jahic. Forecast-aided distribution system state estimation. In *Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MedPower 2016)*, pages 1–5, Nov 2016. doi: 10.1049/cp.2016.1020.

[17] Paolo Attilio Pegoraro, Andrea Angioni, Marco Pau, Antonello Monti, Carlo Muscas, Ferdinanda Ponci, and Sara Sulis. Bayesian approach for distribution system stateestimation with non-gaussian uncertainty models. *IEEE Transactions on Instrumentation and Measurement*, 2017.

[18] Andreas Martin Kettner and Mario Paolone. Sequential discrete kalman filter for real-timestate estimation in power distributionsystems: Theory and implementation. *IEEE Trans. Instrum. Meas.*, 2017.

[19] T. Nakagawa, Y. Hayashi, and S. Iwamoto. Neural network application to state estimation computation. In *Proceedings of the First International Forum on Applications of Neural Networks to Power Systems*, pages 188–192, Jul 1991. doi: 10.1109/ANN.1991.213480.

[20] Jianzhong Wu, Yan He, and Nick Jenkins. A robust state estimator for medium voltage distribution networks. *IEEE Transactions on Power Systems*, 2013.

[21] P.N.Pereira Barbeiroa, H. Teixeira, J. Krstulovica, J. Pereiraa, and F. J. Soares. Exploiting autoencoders for three-phase state estimation in unbalanced distributions grids. *Electric Power Systems Research*, 2015.

[22] D. Singh, J.P.Pandey, and D.S. Chauhan. Radial basis neural network state estimation of electric power networks. *IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies. Proceedings*, 2004.

[23] Amamihe Onwuachumba, Yunhui Wu, and Mohamad Musavi. Reduced model for power system state estimation using artificial neural networks. *IEEE Green Technologies Conference*, 2013.

[24] N. H. Abbasy and W. El-Hassawy. Power system state estimation: Ann application to bad data detection and identification. In *AFRICON, 1996., IEEE AFRICON 4th*, volume 2, pages 611–615 vol.2, Sep 1996. doi: 10.1109/AFRCON.1996.562959.

[25] Coralia Cartis, Nicholas IM Gould, and Ph L Toint. On the complexity of steepest descent, newton's and regularized newton's methods for nonconvex unconstrained optimization problems. *Siam journal on optimization*, 20(6):2833–2852, 2010.

[26] L. Deng, G. Hinton, and B. Kingsbury. New types of deep neural network learning for speech recognition and related applications: an overview. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8599–8603, May 2013. doi: 10.1109/ICASSP.2013.6639344.

[27] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. *International Conference for Learning Representations*, 2015.

[28] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 2014.

[29] Lutz Prechelt. *Neural Networks: Tricks of the Trade*. Springer, 2012.

[30] Asriel U. Levin, Todd K. Leen, and John E. Moody. Fast pruning using principal components. In J. D. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems 6*, pages 35–42. Morgan-Kaufmann, 1994. URL `http://papers.nips.cc/paper/754-fast-pruning-using-principal-components.pdf`.

[31] Gang Li, Xing San Qian, Chun Ming Ye, and Lin Zhao. A clustering method for pruning fully connected neural network. 204-210:600–603, 02 2011.

[32] P Monika and D Venkatesan. Di-ann clustering algorithm for pruning in mlp neural network. 8, 08 2015.

[33] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, 26(1):12–19, Feb 2011. ISSN 0885-8950. doi: 10.1109/TPWRS.2010.2051168.