Medium term electric load forecasting using Lanczos Bidiagonalization with singular value decomposition

Ghufran Ullah¹, Muhammad Aamir Aman², Kamran Khan³, Izhar ulHaq⁴, Mehr E Munir⁵

¹, ², ³, ⁴, ⁵ Department of Electrical Engineering, Iqra National University, Pakistan

³ Department of Electrical Engineering, Sarhad University of information & Technology Pakistan

*Corresponding author: Mehr E Munir, E-mail: mehre.munir@INU.edu.pk

https://doi.org/10.26782/jmcms.2019.08.00029

Abstract

The term forecast stands for predictions of future events and conditions. The process of making such predictions is called forecasting. The main purpose of forecasting is to meet future requirements, reduce unexpected cost and provide a potential input to decision making regarding electrical power production and dispatch. In operating a power system, the mission of the utility/company, from the forecasting point of view, is to match demand for electric energy with available supply. This leads to the fact that a major objective of any power company is accurately predicting future loads. In this research, medium term electrical load forecasting for Peshawar region is studied using Lanczos Bidiagonalization with Singular Value Decomposition. Here, electrical hourly loads are processed in three steps. A polynomial fit is performed to access the non-linear trend of the hourly loads of each year. This is followed by applying the SVD method to the difference between the hourly loads and their trend. SVD serves to extract both the cyclic and the random components of the numerical data. Finally, prediction is done using matrix completion via Lanczos Bidiagonalization.

Keywords: electrical load forecasting, Lanczos Bidiagonalization, SVD method,

I. Introduction

The word forecast mean information and knowledge about upcoming conditions. The way of such type of information and knowledge is named forecasting [I,II]. The purpose of forecasting is how to catch future events, how much expenses can be reduce to take a step for decision-making. In term of forecasting, the utility companies, which need to operate their power system, have a need of forecasting in order to match between the demanded value of electrical energy and supply of
electrical energy. Therefore, utility company, which has more precise power management system, must be accurate and precise forecasting.

So forecasting is a basic tool of energy management system for power related companies. Precise predictions of load and the promise of electrical utility unit help to reduce the ability of spinning reserve decision schedule care plan equipment properly. Play an important role in reducing the costs of generation, in addition to the power system reliability is important. The result is a load system operator offline network analysis as a basis for the prediction; the system uses may be weak. It is considered a time series because the electric load is a continuous variation. In order to predict upcoming load data from existing and available historical data different type of techniques and methodologies are enabled to be applied. Time series methods are based on belief that the data have an internal structure, such as seasonal variation autocorrelation or trend. Such structure has been explore and detected by time series forecasting [III]. The goal of this research and analysis is define here and estimate value in the upcoming time to provide a prediction model is to determine. In addition, we can say that the goal is to search out such a filtered component that can state and explain the composition of load behavior and having ability of such estimation.

Generally time series method show three added elements, the trend component (T), the seasonal component (S), and the irregular or random component (R). To operate power system main goal of utility companies in sense of forecasting, is to match between two values, one is required value of electrical energy and second one is supply of electrical energy. This emphasize to the reality main goal of any company is to predict accurately the upcoming load. Forecasting in case of period is categorized as: a) long-period forecasting, b) medium period forecasting, c) short-period forecasting and d) very short period forecasting.

Mainly Medium period load forecasting depend upon growth factor ;i.e., Growth factor is a factor that impact demand like main events, new load addition, variation occur due to seasons, large facility pattern demand, requirement of services of large consumers. Therefore, such types of forecast, which use historically hourly data in order to predict peak load forecasting of a day or forecasting of week ahead. Both this forecasting and short time forecasting have same way of analysis but there need less accuracy. In other words, we can also say that short term forecasting is more sensitive on power system than medium term load forecasting.

In this research, three steps taken in order to proceed electrical hourly load data. To get nonlinear trend of hourly load data of each year a polynomial fit is performed. This is followed by applying the SVD method to the difference between the hourly loads and their trend[4]. Both random and cyclic components are extracting with the help of SVD. In linear algebra, real matrix or complex matrix factorization is named SVD. It has many useful applications in signal processing and statistics. SVD is
significantly important for solving nonlinear inverse problems where usually a truncated SVD (TSVD) is used as representation of model.

As many literatures and research work is going on in field of load forecasting since 1960 but still due to complexity and technology development remain a challenge to electrical engineers, scholars. Really, it was a difficult job how to predict the future load from the historical load, especially for load forecasting of vacation days, days with extreme weather and other anomalous days. Improvement in forecasting result is potentially possible due to recent mathematical development, mining of data and tolls.

Recently in market de regularity of electricity exist due to which load forecasting gained greater and important challenge. The basis of electrical energy trade is precise forecasting in market base and spot price establishment for the system to reduce electricity purchasing cost. In the real time dispatch operation forecasting error causes more electricity purchasing cost or breaking contract penalty cost to keep the electricity supply and consumption balance. Medium Term Load Forecasting (MTLF) has also some sort of modification due to implementation of electricity Market. For Example, the demand side management and volatility of spot market causes the consumer active response to the electricity price. This should be considered in the forecasting model in the market environment.

Load forecasting is way of estimating what future electric load will be for a given forecast horizon based on the available information about the state of the system.

Planning efforts for all utility load predictions long has been recognized as the initial building blocks. While changing market structures have altered the types of forecasts that are most useful, the link between sound evaluation and design of infrastructure improvements is irreducible. Physical system planners condition their alternatives on future views regarding load levels and locations. Financial planners tie both revenue and expense forecasts to expected future energy sales and peak demands.

Electric power generation and distribution companies, predictions that load forecasting is a key of performance and can increase revenues. It leads them to plan on their capacity and operations so that they can reliably provide the demanded values of energy to consumers.

Load forecasting are different categories as:

- Short term forecasting (1 hour to a week)
- Medium term forecasting (month up to a year)
- Long term forecasting (over one year)
Some factors that effects load forecasting are:

- Weather influence
- Time factors
- Customer classes

**Weather’s Influence**
Electric load has an obvious correlation to weather. The most important variables responsible in load changes are:

- Dry and wet bulb temperature
- Dew point
- Humidity
- Wind speed/direction
- Sky cover
- Sunshine

**Time-factor**
In the forecasting model, we should also consider time factor such as:

- The day of the week
- The hour of the day
- Holidays

**Customer’s classes**
Electric utilities always observe different types of consumers and accordingly various load consumption pattern. For example, an industrial consumer consumes more power during morning hour whereas residential consumer consumes more power in evening.

### II. Methodology

In any matrix completion problem formulation, the need to find out its singular value decomposition is apparently unavoidable. This problem can be designed in such a way as one of an eigenvalue problem where the algorithm used for decreasing the rank of the matrix is Lanczos Bidiagonalization. In the method used in this research, we have avoided loss of orthogonality among Lanczos vectors by using Partial reorthogonalization according to research [I,II].

**Singular Value Decomposition**
The Singular Value Decomposition of a square matrix $A$ is just a decomposition or factorization of the matrix into three matrices such that the product of the inner product of two of them with the transpose of the third matrix equals the original matrix again. Mathematically this is represented as:

$$A = U \Sigma V^T$$
In the above equation, the U and V are called as the singular vectors where U is a left singular matrix and V is a right singular matrix whereas both matrices have orthonormal columns.

Visually the same mathematical equation can be represented as shown below:

\[
\begin{bmatrix}
A \\
\end{bmatrix}_{n \times d} = \begin{bmatrix}
\begin{bmatrix}
D \\
\end{bmatrix}_{r \times r} & \begin{bmatrix}
V^T \\
\end{bmatrix}_{r \times d}
\end{bmatrix}
\]

Figure -1 SVD of a matrix A

This also implies that the inverse of the matrix A can be calculated as:

\[A^{-1} = V \Sigma^{-1} U^T\]

SVD is true for all invertible matrices A, whether they are square or rectangular.

**Lanczos Bidiagonalization**

An m x n matrix A with complex elements can be decomposed using Singular Value Decomposition and can be rewritten as:

\[A = U \Sigma V^*\]

Where \(U = [u_1, u_2, ... u_m]\) and \(V = [v_1, v_2, ... v_n]\) represent Unitary matrices such that \(U^*U = I\) and \(V^*V = I\) while \(\Sigma\) is an \(m \times n\) matrix of type diagonal containing non-negative real diagonal elements. The elements of \(\Sigma\) are represented by \(\Sigma_{ii} = \sigma_i\) where \(i = 1, ..., \min{m, n}\). The vectors \(u_i\) and \(v_i\) are called as the left singular and right singular vectors respectively. Also, \(\sigma_i, u_i, v_i\) are called as singular triplets collectively. For any matrix A, consider a cross product matrix found out by multiplying A with \(A^*\), and a cyclic matrix obtained after find the Hermitian matrix of the matrix A represented by:

\[
\text{Hermitian of matrix } A = \begin{bmatrix}
0 & A \\
A^* & 0
\end{bmatrix}
\]  

(1)

If the matrix A is a sparse matrix, then calculating the cross product \(A^*A\) is not feasible. For that purpose, a Bidiagonalization algorithm was proposed in [5] termed as Golub - Kahan - Lanczos Bidiagonalization.
Consider a decomposition of the same matrix $A$:

$$A = PBQ^*$$  \hspace{1cm} (2)

Where $B$ is an upper Bidiagonal matrix, $P$, and $B$ are unitary matrices. If we find out the SVD of $B$ matrix, we come to find that the matrices $B*B$ and $A*A$ are similar in unitarily and these are tridiagonal matrices. If the SVD of $B$ is taken i.e.:

$$B = X\Sigma Y^*$$  \hspace{1cm} (3)

Then it can be easily deduced that $U=PX$, and $V=QY$ which is a solution to the original problem $A=U\Sigma V^*$.

**Derivation of Lanczos Algorithm**

Consider a matrix SVD decomposition of $A$:

$$A = P_BQ^*$$  \hspace{1cm} (4)

Where $P_{nxn}$ contains orthonormal columns by removing the zeros, $Q_{nxn}$ is a unitary matrix and $B_{nxn}$ is a matrix representable by:

$$B_n = P_n^*AQ_n = \begin{bmatrix} 
\alpha_1 & \beta_1 \\
\alpha_2 & \beta_2 \\
& \ddots & \ddots \\
& & \ddots & \beta_{n-1} \\
& & & \alpha_n 
\end{bmatrix}$$  \hspace{1cm} (5)

The diagonal elements of this matrix can be found and represented by:

$$\alpha_i = p_i^*Aq_i \text{ and } \beta_i = p_i^*Aq_{i+1}$$

Eq. 5 can be rewritten in another way when the term $P_n$ is pre-multiplied with it as follows:

$$AQ_n = P_nB_n$$  \hspace{1cm} (6)

If the same equation is transposed on both sides and then $Q_n$ is multiplied on both sides, it would result in the following equation:

$$A^*P_n = Q_nB_n^*$$  \hspace{1cm} (7)

Finding the first $k$ columns of this matrix is represented mathematically as follows:

$$AQ_k = P_kB_k$$
And \( A^*P_k = Q_kB^*_k + \beta_k q_k + 1^e_k \) \( (8) \)

Here in Eq. 8, \( B_k \) represents principal submatrix of the original matrix with order \( k \times k \).

If a specific column is required in vector form, the equation can be transformed to:

\[
Aq_j = \beta_{j-1}p_{j-1} + \alpha_j p_j
\]

\( (9) \)

And

\[
A^*p_j = \alpha_j q_j + \beta_j q_{j+1}
\]

\( (10) \)

In addition, these equations can be used to produce the double recursion:

\[
\alpha_jp_j = Aq_j - \beta_{j-1}p_{j-1}
\]

\( (11) \)

\[
\beta_jq_{j+1} = A^*p_j - \alpha_j q_j
\]

\( (12) \)

In Eq. 11 and 12\( \alpha_j = \|Aq_j - \beta_{j-1}p_{j-1}\|_2 \) and \( \beta_j = \|A^*p_j - \alpha_j q_j\|_2 \) because both the columns of Q as well as P are normalized. From these two equations, the Bidiagonalization has been created which will be summarized in the next topic. Both of these equations can be merged to generate:

\[
A^*AQ_k = Q_kB^*_kB_k + \alpha_k \beta_k q_k + 1^e_k
\]

\( (13) \)

It can be noted from calculations that the \( B_k^*B_k \) results in a tridiagonal and positive definite symmetric matrix.

The Algorithm of the Lanczos Bidiagonalization can be summarized in the following pseudo code.

Step a) Choose a unitarily normalized vector, call it \( q_1 \)
Step b) Set \( \beta_0 = 0 \)
Step c) For every \( j = 1, 2, ... k \)
  a. Find \( p_j = Aq_j - \beta_{j-1}p_{j-1} \)
  b. \( \alpha_j = \|p_j\|_2 \)
  c. Set new value of \( p_j = p_j/\alpha_j \)
  d. Find next element of \( q, q_{j+1} = A^*p_j - \alpha_j q_j \)
  e. Find coefficient \( \beta_j = \|q_{j+1}\|_2 \)
  f. Then, \( q_{j+1} = q_{j+1}/\beta_j \)

Step d) End
DATA Detail

The data for this research work has been taken from NTDC Pakistan’s reports of 2015 as published and seen on their website. This document is a very detailed collection of surveyed and actual data as in the national grid for the past 10 years. But some discrepancies can be seen in the document i.e. that there are tens of hundreds of missing and incomplete values of some of the most important parameters of the grid such as transmission line lengths, demands per area per hour and on various dates, a full day’s worth of data was missing.

This discrepancy made this data an ideal candidate for matrix completion. The Standard deviation of a set of data is taken, which is used to find out how much the given matrix data is deviating from its norm. In addition, the average value of the set with its minimum and maximum values also showed us in which range the missing data is to be calculated.

The data from the document is then checked for errors, the errors are corrected and the corrected data is arranged in a matrix where each column is a separate hour of the day and each row is a day of the month. This way, all the data is fed to Matlab.

Total data fed to Matlab contains 14 months’ worth of values. The total quantity is 10224 values. In order to convert these to a perfect square matrix, the data is truncated into 99 x 99 matrix. Matrix Completion utilizing Partial Reorthogonalization is run on the data and then results are tabulated, plotted and analyzed. Criteria for checking the forecasted value is the MAE (Mean absolute error) which resides below ranges of 10% in our research.

III. Results/Discussion

The data profile of the 14 months’ data is shown below

![Figure -2  14 months NTDC data from July 2015 to August 2016](image)

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ghufran Ullah et al
Figure -2 shows the power demand in kilowatts in NTDC from July 2015 to August 2016. The data is arranged w.r.t. hour number and it shows us that there are many variations in the electric power demand throughout the year. This type of data is very hard to analyze and sort into a set of equations and therefore there is a dire need to pass this data through a matrix completion algorithm so that the pattern can be understood. There are, however, superior techniques like the Artificial Neural Network.

Comparisons

After the running of algorithm for matrix completion, the data where discrepancies occurred originally, have been calculated and inserted in the new calculated profile. This data is termed as a recovery matrix, and the plot of the data w.r.t the original data is shown below.

![Full Comparison](image)

Figure -3 Overall comparison of the data profile and the recovery matrix over 14 months

The comparison shows us a Mean Square Percent Error of the whole recovery matrix to be 14.9227% to be exact. Also the data calculation inside Matlab shows us a statistical 7.43% error as a Mean Absolute Error; both of these values are in a challengeable range of less than 15% and less than 10% respectively.

Below is a table summarizing the results of the case study
Table 1 summarizing result of the case study

| Month       | Quantity of hours | MSE     | RMS Error  |
|-------------|-------------------|---------|------------|
| July 2015   | 744               | 17.55%  | 5.4222%    |
| August 2015 | 720               | 22.54%  | 4.3886%    |
| September 2015 | 720          | 14.43%  | 3.6623%    |
| October 2015 | 744              | 12.68%  | 6.0265%    |
| November 2015 | 720             | 7.70%   | 7.3546%    |
| December 2015 | 744            | 8.45%   | 4.6773%    |
| January 2016 | 744              | 6.24%   | 3.4043%    |
| February 2016 | 672             | 7.15%   | 4.3407%    |
| March 2016   | 744              | 15.82%  | 5.9280%    |
| April 2016   | 720              | 12.69%  | 2.5985%    |
| May 2016     | 744              | 17.55%  | 6.1634%    |
| June 2016    | 720              | 20.13%  | 5.4030%    |
| July 2016    | 744              | 19.63%  | 4.5105%    |
| **Total**    | **9480**         | **Mean = 14.04%** | **Mean = 4.9130%** |

Relevant Research Comparisons

The following table illustrates the comparison of this study along the various relevant researches done in this field of using SVD and matrix completion applied to electrical load forecast.

Table 2 Comparative analysis of Load forecasting using Matrix Completion with SVD

| Technique                                                                 | Error Type | Value          |
|---------------------------------------------------------------------------|------------|----------------|
| RNN (Vishal Sharma) 2011 (Ontario) – 1                                    | MAPE       | 6.92%, 5.90%, 5.97%, 7.01% |
| Load Forecasting for Industry using Artificial Neural Network              | MAPE       | 2.6%, 3.58%, 5.28%, 3.93% |
| (Eric Lynn Taylor) 2013 Summer, Fall, Winter, Spring                      |            |                |
| Long Term Probabilistic Load Forecasting with Hourly Information           | MAPE       | 4.7%, 5.3%, 5.0%  |
| (Tao Hong, Jason Wilson) 2014 Extension 1, 2, 3                           |            |                |
| SVD Using Lanczos Bidiagonalization for Load Forecasting with Hourly Information (This Research) 2017 | MAPE | 4.9 % |
|                                                                            | MSE        | 14.04%         |
IV. Conclusion and Future work

From the comparison of all the above graphs and comparative results, SVD using Bidianalization technique is the best method or competitive method with the latest trends in Load Forecasting and Matrix Completion problems. The data used to verify this research work is obtained from NTDC 2015 June to 2016 July – 14 months of data. The overall MAPE error in the calculations has been shown to be less than 5% while the MSE was less than 15% at 14.9% exactly, and it could have been a lower number if the power consumption of Pakistan’s grid has been somehow regular and not peaking and odd hours. This result is closer to Mr. Yasir Khan’s (M.S. Electrical Department, Abasyn University Peshawar) Work in 2017 June, which was 3.09%. His work was based on Nuclear Norm Minimization. No recent works other than that exist on Pakistani Grid about forecasting, so we have analyzed our work with relevant Indian and Canadian researchers’ works and found to be in the competent range.

In summary new method has been presented to compute medium term load forecasting which is based on hourly load to apply SVD method. Results and performance showed that the given method is generic and give outputs with less error in percent.

Future work

In literature review section, many methods and results are discussed. Every techniques have different results depend upon facts it have, by increasing numbers of historical data mean previous years so the accuracy of predicted data will be decrease. As forecasting, whose base is made on assumption so it is not completely an accurate and perfect result? Some error will always exist so for researcher it is one of the advance fields so that they work on it and to reduce the value of error in predicted value. In the future one can work on very long term forecasting using same method, as it is theoretically presented to be best for matrix completion problems. Hence, this will be used to verify and check its working credibility and error ranges.

V. Acknowledgement

The authors would acknowledge to Iqra National University Electrical Engineering Department for providing sufficient environment and guidance.
References

I. H. M Al-Hamadi and S. A. Soliman. "Long-term/mid-term electric load forecasting based on short-term correlation and annual growth." Electric power systems research 74, no. 3 (2015): 353-361

II. N. Abu-Shikhah, F. Elkarmi. "Medium-term electric load forecasting using singular value decomposition." Energy 36, no. 7 (2011): 4259-4271.

III. Paterek, Arkadiusz. "Improving regularized singular value decomposition for collaborative filtering." In Proceedings of KDD cup and workshop, vol. 2017, pp. 5-8. 2007.

IV. S. Miloš, M. Božić, Z. Stajić, and M. Milošević. "LS-SVM model for electrical load prediction based on incremental training set update." Przegląd Elektrotechniczn 4 (2013): 195-199

V. X. Changhao, J. Wang, and K. McMenemy. "Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks." International Journal of Electrical Power & Energy Systems 32, no. 7 (2010): 743-750.