A Search Method for Obtaining Initial Guesses for Smart Grid State Estimation

Yang Weng, Student Member, IEEE, Rohit Negi, Member, IEEE, and Marija D. Ilić, Fellow, IEEE

Abstract—AC power system state estimation process aims to produce a real-time “snapshot” model for the network. Therefore, a grand challenge to the newly built smart grid is how to “optimally” estimate the state with increasing uncertainties, such as intermittent wind power generation or inconsecutive vehicle charging. Mathematically, such estimation problems are usually formulated as Weighted Least Square (WLS) problems in literature. As the problems are nonconvex, current solvers, for instance the ones implementing the Newton’s method, for these problems often achieve local optimum, rather than the much desired global optimum. Due to this local optimum issue, current estimators may lead to incorrect user power cut-offs or even costly blackouts in the volatile smart grid. To initialize the iterative solver, in this paper, we propose utilizing historical data as well as fast-growing computational power of Energy Management System, to efficiently obtain a good initial state. Specifically, kernel ridge regression is proposed in a Bayesian framework based on Nearest Neighbors search. Simulation results of the proposed method show that the new method produces an initial guess excelling current industrial approach.

Index Terms—Smart grid, state estimation, iterative algorithm, historical data, kernel ridge regression.

I. INTRODUCTION

Initiated by the U.S. government, the rapid-expanding smart grid aims to evolve into an efficient, reliable and sustainable modern grid by adopting, integrating, and advancing the communication and computing technologies already exist. To achieve such an ambitious goal, namely the “smartness” of the power grid, a highly accurate State Estimation (SE) process [1], [2] is necessary in providing bases for many key functionalities in the operation and control of smart grid.

However, the nonlinear measurement model of AC power system renders the SE problem a highly nonconvex character, which can not be optimally solved without great computational expense. To deal with nonlinearity, one can approximate AC power system with DC model [3]–[5], based on which robust state estimation can be further conducted to increase robustness to bad data [6]–[8]. One can also tries to convexify the nonconvex problem by convex relaxation with Semidefinite Programming [9], [10]. However, those techniques usually come with an approximation cost, resulting in a relative poor estimate.

As a result, many successful and widely used state estimation algorithms are to directly work with nonlinear measurement model, which is formulated in the Weighted Least Square (WLS) [11]–[15] form, with Newton’s method (i.e. [3]) to be the solver. By successively finding a better approximation, Newton’s method can reach a local optimum of the nonconvex problem. However, obtaining global optimum is not guaranteed, due to the non-convexity of the problem. If the initial guess used in Newton’s method is (by pure chance) close to global optimum, then it is likely to find global optimum. Otherwise, it may merely reach a local optimum and stop. Therefore, how to efficiently obtain a good initial guess becomes a critical issue, which greatly affects state estimate, and consequently operations in smart grid.

In traditional transmission network, it is possible to use previous state estimate as a heuristic initial guess for SE, based on the belief that no significant change appears in a short time. However, such belief will no longer hold in smart grid, where intermittent generation (such as wind generation) or topological change leads to significant state shift in power system operations. In such case, a previous state estimate computed around 2 minutes ago [16] may not truly reflect the operating point of the current power system and generates suboptimal results accordingly.

Hence, instead of using the result produced by the last SE, an efficient initial guess generating algorithm must be adopted to acquire a higher probability in achieving the global optimum. In this paper, we propose a Bayesian approach based on historical data search, where a group of measurement sets and corresponding state estimates are used in combination with the current measurement in kernel ridge regression [17] to pursue a good initial estimate of current states. The proposed method is based on the idea that two similar system measurement sets usually indicate two similar operation conditions (system states). In particular, we first compute the distance between the current measurement set and the historical measurement sets. After identifying a group of measurement sets with the smallest distances to the current measurement set, the group of sets is used to compute an initial guess for the iterative algorithm to calculate the current state. A key property enabling such an efficient method is the increasing computing ability of the smart grid, which allows an algorithm to achieve good performance.

In addition to increasing the probability to achieve the global optimum, the aforementioned search algorithm has another attractive feature: in the case of equal accuracy (both the previous estimate start and the proposed Nearest Neighbors (NN) start return the global optimum states), the NN algorithm...
becomes helpful in finding a better initial guess, creating the potential to reduce the iteration time vital for large system SE.

Whereafter, the performance of the NN method is verified by simulations on the standard IEEE 300-bus test case [18], [19]. Provided with enough historical data, the NN start can furnish performance equal to or better over the traditional start. Further, in such context, when a computer can not generate a prompt initial guess, a proper adjustment method is proposed to furnish a satisfactory performance with time reduction.

The rest of the paper is organized as follows: Section II reviews the WLS state estimation; Section III describes the Nearest Neighbors approach; Section IV illustrates the simulation results and section V concludes the paper.

II. REVIEW OF POWER SYSTEM STATE ESTIMATION

In this section we briefly review the state estimation (SE) problem in power systems. In general, the state estimation problem is a nonlinear problem that needs to be solved by implementing iterative algorithms.

A. Model

The following AC power system model is assumed in this paper:

\[ z = h(x) + w \]  

(1)

where the vector \( x \):

\[ x = (|v_1|e^{j\delta_1}, |v_2|e^{j\delta_2}, \ldots, |v_n|e^{j\delta_n})^T \]  

(2)

represents the power system states, \( w \) is an \( m \times 1 \) vector denoting the additive measurement noises, presumably independent Gaussian random variables with zero means, i.e., \( u \sim \mathcal{N}(0, \Sigma) \), where \( \Sigma \) is a diagonal matrix, with the \( i^{th} \) diagonal element \( \sigma_i^2 \). \( z \) is an \( m \times 1 \) vector denoting the set of telemetered measurements, such as power flows and voltage magnitudes. \( h(\cdot) \) is a vector of nonlinear functions relating the states \( x \) to the measurements \( z \). In practice, the measurement set \( z \) is usually made redundant to guarantee the observability of the whole system, i.e., \( m \geq n \).

The goal of power system SE is to find an estimate (\( \hat{x} \)) of the true states (\( x \)) that best fits the measurement set \( z \) according to the measurement model in (1). This is usually achieved by minimizing the following criterion:

\[ \hat{x} = \arg \min_x J_p(x) = \sum_{i=1}^{m} \left( \frac{z_i - h_i(x)}{\sigma_i} \right)^p \]  

(3)

where the parameter \( p \) (\( p \geq 0 \)) is used to achieve desired performance. For example, for \( p = 2 \), the above problem corresponds to the conventional Weighted Least Square (WLS) SE. For \( p = 1 \), the above problem reduces to the Weighted Least Absolute Value (WLAV) SE [3], which is well-known as robust to bad data.

B. WLS State Estimation

As we already discussed, the optimization problem in (3) is highly nonconvex, due to the non-convexity of the cost function in (3). Thus, it is very difficult to solve the problem optimally. In practice, the state estimation problem is usually solved by using Newton’s method, which is essentially a local search algorithm. As indicated in the following, if we set \( p = 2 \), then the problem becomes equivalent to:

\[ \hat{x} = \arg \min_x J_2(x) = (z - h(x))^T \Sigma^{-1} (z - h(x)) \]  

(4)

After obtaining an initial guess \( x^{(0)} \), Newton’s method updates the estimate according to the following rule:

\[ x^{(k+1)} = x^{(k)} - \frac{J'_2(x^{(k)})}{J''_2(x^{(k)})}, \quad \forall k \in \mathbb{Z} \]  

(5)

Notice that, such local search method is highly sensitive to the initial guess. As a result, for smart grid SE, the simple industrial initial guess approach (use the last SE result) may not be able to provide a sufficient initial guess for the Newton’s method to converge to the global optimum. In the next section, we are going to discuss a new systematic approach to enhance the initial guess process.

III. THE NEAREST NEIGHBORS APPROACH

Now we consider the initial guess generating method to lessen the state estimation error. We assume the availability of the following two key properties of the smart grid:

- A database recording the historical measurements and estimates (Bad data was filtered out in each estimate.)
- High performance computers

Intuitively, close-by states usually produce similar measurements. Therefore, a smaller distance between the current measurement set \( z_{current} \) and a historical measurement set \( z_k \) implies that the associated historical state vector \( x_k \) stays closer to the current true state vector \( x_{true} \). Consequently, when \( x_k \) is used as a starting point, Newton’s method achieves better accuracy. Provided with the aforesaid, we proposed a historical initial guess search method, known as the Nearest Neighbors (NN) start. To make the estimation unbiased to a single data point, a group of nearest neighbors is obtained in stead of one historical data point. A refined initial guess is solved by using Newton’s method, which is essentially a local search algorithm. As indicated in the following, if we set \( p = 2 \), then the problem becomes equivalent to:

Mathematically, the NN algorithm can be decomposed into two parts:

- A minimization problem to obtain a group of likely historical data
- A kernel ridge regression problem to obtain an ‘optimal’ initial state estimate from the group

Specific analysis will be presented in the next two subsections.
the algorithm simply looks for an index set $s(z)$ which minimizes a particular distance function $d(s)$. Here $N$ is the number of total points in the database. $s$ represents the set of natural number. $k$ represents a particular index for a data point. As a result, $z_k$ indicates the measurement set within the time slot of index $k$. Finally, $p$ indicates the cardinality of the set $s$. Essentially, during the searching step, the algorithm simply looks for an index set $s$ with $p$ elements which represents a group of measurement sets that have nearer distance to the current measurement $z_{\text{current}}$.

B. Kernel ridge regression

After obtaining the minimum index set $s$, this subsection aims to use the associated measurements and their corresponding states, as well as the current measurement set, in obtaining a good current state estimate to initialize the regular Newton’s method in the SE problem.

1) Ridge regression: In order to explain the process, we consider the Normal model below, which is a popular discriminative model with unknown hyper-parameters $q$ and $\Sigma_q$:

$$ x|z : N(q^Tz, \Sigma_q) $$

(7)

To identify such discriminative model for our inference, a regularized (ridge regression) estimator in (8) is commonly used:

$$ \hat{q} = \arg \min_q \sum_{i=1}^{n} (x_i - q^T z_i)^2 + 2\gamma ||q||^2 $$

(8)

For the Normal model, with the historical data stored in $Z_{\text{mat}}$ and $X_{\text{mat}}$ as follows,

$$ Z_{\text{mat}} = (z_1, z_2, \cdots, z_n) $$

$$ X_{\text{mat}} = (x_1, x_2, \cdots, x_n) $$

(9)

where the subscript $h$ is the total number of chosen neighbor measurement sets, we can obtain a closed-form solution:

$$ \hat{q} = (Z_{\text{mat}}Z_{\text{mat}}^T + 2\gamma I)^{-1} Z_{\text{mat}} x_{\text{current}} $$

(10)

(The unknown hyper-parameter $\Sigma_q$ has been absorbed into the penalty constant $\gamma$.) Notice that due to the ridge regularization (since $\gamma > 0$), $Z_{\text{mat}}Z_{\text{mat}}^T + 2\gamma I \geq 2\gamma I > 0$, so that this matrix is always invertible. Thus, the regularized estimator always exists. Once the hyper-parameter $\hat{q}$ is estimated, it can be used for the Bayesian inference to initialize the current state estimate $\hat{x}^{\text{B}}$, as follows:

$$ \hat{x}^{\text{B}} = \hat{q}^T z_{\text{current}} $$

(11)

$$ = x_{\text{current}}^T Z_{\text{mat}}^T (Z_{\text{mat}}Z_{\text{mat}}^T + 2\gamma I)^{-1} z_{\text{current}} $$

(12)

$$ = x_{\text{current}}^T T $$

(13)

Further by employing the Matrix Inversion Lemma $(A + BCD)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)CA^{-1}$ to expand the inversion above, the alternative form of $T$ can be obtained, which may further simplify the computational need:

$$ T = Z_{\text{mat}}^T (Z_{\text{mat}}Z_{\text{mat}}^T + 2\gamma I)^{-1} z_{\text{current}} $$

(14)

$$ = (Z_{\text{mat}}Z_{\text{mat}}^T + 2\gamma I)^{-1} Z_{\text{mat}}^T z_{\text{current}} $$

(15)

where

$$ Z_{\text{mat}}^T z_{\text{current}} = (z_1, z_2, \cdots, z_n)^T (z_1, z_2, \cdots, z_n) $$

$$ = (z_1^T z_1 \cdots z_1^T z_n) $$

$$ \vdots $$

$$ = (z_n^T z_1 \cdots z_n^T z_n) $$

(16)

$$ Z_{\text{mat}}^T z_{\text{current}} = (z_1, z_2, \cdots, z_n)^T z_{\text{current}} $$

$$ = (z_1^T z_{\text{current}} \cdots z_n^T z_{\text{current}}) $$

(17)

Notice that the matrix $Z_{\text{mat}}^T Z_{\text{mat}}$ appears in the calculation (15), as opposed to the original calculation (14) involving $Z_{\text{mat}}Z_{\text{mat}}^T$, which creates the potential to reduce computation cost.

2) The Kernel trick for Normal Discriminative model: for the problem above, there exists high-dimensional mapping $z = w(u)$, from which the inner product $z_i^T z_j = (w(u_i))^T w(u_j)$ can be calculated by a kernel $K(\cdot, \cdot)$, as below,

$$ z_i^T z_j = K(u_i, u_j) $$

(18)

The point to note is that kernel calculation uses only the (low-dimensional) $u$’s, rather than the high-dimensional $z$’s. Therefore, the computational complexity of calculating the inner products in (16) and (17) is low, even though $z_{\text{current}}$ itself may be very large. This idea of using a cost-effective kernel calculation to implement a high-dimensional Normal model is called ‘the kernel trick’. In this paper, we employ the following kernel forms as candidates.

- Homogeneous polynomial: $K(u_i, u_j) = (u_i^T u_j)^d$.
- Inhomogeneous polynomial: $K(u_i, u_j) = (1 + u_i^T u_j)^d$.
- Gaussian (Radial Basis function): $K(u_i, u_j) = \exp(-\mu ||u_i^T u_j||^2)$, $\mu > 0$.

To choose a proper $\gamma$ and kernel pair, we use the following training and validation phases.
• In the training phase, one applies part of the historical data on different kernel function and $\gamma$ to calculate different $T$s.
• In the validating phase, another part of historical data are used to validate the kernel function and $\gamma$.

Finally, the chosen $\hat{x}_T^B$, computed from the validated $T$, is used in (13) for the testing phase as an NN initial guess.

C. Computation cost and improvements
With the introduction of the search criteria above, we discuss next the computation cost in (6). In general, the computation time of the proposed method is consuming when the historical data size is large, resulting in excess of permitted time $^1$, in solving the optimization problems optimally. Therefore, the application of proposed method can be computational infeasible for large networks with a time span of long historical data. As a consequence, a suboptimal search algorithm is considered to reduce computation time by randomly choosing a subset of measurement set. As the dimensionality of the data is reduced, the computational time can accordingly be greatly reduced.

Notice that a way to overcome the defect of overly reducing the dimension is by running another Newton’s method in parallel with the previous state estimate, then choose the initial guess with a decreased error after comparing the sum square errors in between.

IV. NUMERICAL RESULT
In this section, we simulate and verify the performance of the nearest neighbors (NN) start, and compare it to the previous state estimate start in the standard IEEE 300 test case.

A. Historical (training / validating) Data and Testing Data Preparation
Such simulations are completed in MATLAB environment in accordance with MATLAB Power System Simulation Package (MATPOWER) [18], [19]. Further, to simulate the power system behavior in a more practical pattern, online load profile from New York ISO [20] is also adopted in the subsequent simulation. Specifically, the load data used is between February 2005 and April 2012 with a consistent data format. It has 11 online load profiles in New York ISO area, namely ‘CAPITL’, ‘CENTRL’, ‘DUNWOD’, ‘GENESE’, ‘HUD VL’, ‘LONGIL’, ‘MHK VL’, ‘MILLWD’, ‘N.Y.C.’, ‘NORTH’, and ‘WEST’. Finally, the data are recorded every five minutes.

Firstly, in order to obtain the 199 nonzero active load power consumptions in the IEEE 300 test case file, we employ a downsampling method to extend the 11 load profiles to 209 (11 x 19) profiles. Basically, instead of the 5-minute interval from the online resource, we use a new interval of 95 (5 x 19) minutes. As a result of such downsampling, we obtained 38476 valid historical load data for each load bus from February 2005 to December 2011. The testing load profiles between February and April 2012 are generated using the same approach.

$^1$SE is usually done every two minutes [16]. Secondly, besides fitting the normalized loads data into the case file, one topology connection is changed with probability 10% to imitate certain feature of the smart grid. Subsequently, an AC power flow is run to generate the true states of the power system, followed by creating true measurement sets with Gaussian noises (standard deviations in Table I). Hereby, we assume that the measurement set includes (1) The power injection on each bus; (2) The transmission line power flow ‘from’ or ‘to’ each bus that it connects; (3) The direct voltage magnitude of each bus.

B. Applying the proposed algorithm
1) Training phase: by randomly selecting one measurement from between February and April 2012 as a test case, a group of nearest neighbors measurements in (6) between 2005 and 2009 is selected. Then the matrix $T$ in (15) is computed for different choices of $\gamma$ and the kernel function.

2) Validating phase: the T matrices are validated on the data in 2011 to obtain the best choice for the kernel and $\gamma$.

3) Testing phase: use the $T$ matrix chosen in the validating phase to calculate Bayesian initial guess $\hat{x}_T^B$ via (13), which is then tested with Newton’s method. Besides, for comparison purpose, the industrial method of choosing the last state estimate is also applied to the same data.

C. Comparison Results
1) Accuracy Improvement: the performance of the proposed Nearest Neighbors (NN) method is first demonstrated in Fig.2 with the metric Weighted Residual Sum of Squares error (WRSS) defined as follows:

$$WRSS = \sum_{i=1}^{m} \left( \frac{z_i - h_i(x)}{\sigma_i} \right)^2$$

(19)

For fairness, and to outline the difference between the two method, comparison is conducted for 400 times, with the condition that the previous state estimate start and the NN initial start each returns different results. In particular, the histogram of the relative errors

$$\frac{WRSS_{NNStart}}{WRSS_{Pre.Est.Start}}$$

(20)

drawn in Fig.2, showing that the NN approach has greatly reduced estimation errors, from which we can reasonably conclude that the proposed NN initial guess method is approximately globally optimal. This leads to a rational interpretation for the proposed SE procedure: since the NN method finds a closer initial state than the conventional method, the possibility is greatly increased in tending to the global optimum using the Newton’s method.
Further to the comparison in the error domain, Fig. 3a and Fig. 3b provide state domain plots. The $x$ coordinate represents the bus number. The $y$ coordinate represents the voltage magnitude ratio ($\frac{|V_{NN}|}{|V_{True}|}$ and $\frac{|V_{Pre:Est:}|}{|V_{True}|}$) in Fig. 3a and the voltage phase angle ratio ($\frac{\angle V_{NN}}{\angle V_{True}}$ and $\frac{\angle V_{Pre:Est:}}{\angle V_{True}}$) in Fig. 3b, respectively. It can be observed that the NN method has a ratio consistent with 1, and its less variance indicates its ability to track the true system states while the previous estimate start result is deviating from 1 with large variance. Such deviation in the previous estimate start result is caused by the local-search behavior of Newton’s method, which is suboptimal with an inferior initial guess.

2) Iteration Time Facilitation: we next consider the reduction of iteration time for the Newton’s method as another benefit of the proposed algorithm, by running simulations in a personal computer with an Intel Core 2 Duo Processor and 3GB RAM. Provided that both initial guess methods return the same global optimum estimate, Fig. 4 shows the relative computation time used by NN method compared to the previous estimate start method. On average, the NN start achieves shorter iteration time on all cases with approximately 75% iteration time reduction.

3) Reducing the Computation Time: finally, we evaluate the performance of NN method with random measurement reduction essential to promptly generating an initial guess. Particularly, before computing the distance between the current measurement vector and the historical measurement vector in (6), a random measurement subset of $z$ is chosen for comparison. By simulation, in obtaining the initial guess, the range of measurement can be narrowed to 25% on average with similar performance. Finally, for the proposed method, once the current measurement set is available, distance comparisons to historical measurement sets are independent of each other. As a result, one can explore the possibility to parallelize the programming part for faster computation, if parallel computers are available. The hope is that, with these methods, the state estimation can be conducted every one or two minute, which is
the standard for the traditional power system state estimation.

V. CONCLUSIONS

In this paper, we discuss how to systematically find an initial guess for the iterative algorithm used by AC power system state estimation. Based on the intuition that similar measurements reflect similar power system states we formulate the finding of the initial guess as a minimum distance search problem. Further, a Bayesian estimate is obtained via kernel ridge regression. We develop that such method of generating ridge regression. We develop that such method of generating

ACKNOWLEDGMENT

This work was supported in part by US NSF awards 0931978, 0831973 and 0347455.

REFERENCES


Yang Weng (S’08) received his B.Engg. in Electrical Engineering in 2006 from Huazhong University of Science and Technology, Wuhan, China, and a M.S. in Statistics from University of Illinois at Chicago in 2009. Currently, he is a Ph.D candidate in the Department of Electrical and Computer Engineering at Carnegie Mellon University, Pittsburgh, PA. His research interest includes power systems, smart grid, information theory, and wireless communications.

Rohit Negi (M’00) received the B.Tech. degree in Electrical Engineering from the Indian Institute of Technology, Bombay, India in 1995. He received the M.S. and Ph.D. degrees from Stanford University, CA, USA, in 1996 and 2000 respectively, both in Electrical Engineering. He has received the President of India Gold medal in 1995. Since 2000, he has been with the Electrical and Computer Engineering department at Carnegie Mellon University, Pittsburgh, PA, USA, where he is a Professor. His research interests include network analysis, communications systems, information theory.

Marija D. Ilić (M’80-SM’86-F’99) is currently a Professor at Carnegie Mellon University, Pittsburgh, PA, with a joint appointment in the Electrical and Computer Engineering and Engineering and Public Policy Departments. She is also the Honorary Chaired Professor for Control of Future Electricity Network Operations at Delft University of Technology in Delft, The Netherlands. She was an assistant professor at Cornell University, Ithaca, NY, and tenured Associate Professor at the University of Illinois at Urbana-Champaign. She was then a Senior Research Scientist in Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, from 1987 to 2002. She has 30 years of experience in teaching and research in the area of electrical power system modeling and control. Her main interest is in the systems aspects of operations, planning, and economics of the electric power industry. She has co-authored several books in her field of interest. Prof. Ilić is an IEEE Distinguished Lecturer.