Multiple Bad Data Processing using Binary PSO Algorithm Based on PC Cluster System

Hee-Myung Jeong and June Ho Park

Department of Electrical Engineering of Pusan National University, South Korea
jeonghm@pusan.ac.kr and parkjh@pusan.ac.kr

Abstract

In power systems operation, state estimation takes an important role in security control. For the state estimation problem, the weighted least squares (WLS) method and the fast decoupled method have been widely used at present. Especially when bad data are mutually interacting, the detecting of multiple bad data may be difficult to handle, since the normalized or weighted residuals may become faulty. Then the problem of detecting bad data is considered as a combinatorial decision procedure.

In this paper, the binary Particle Swarm Optimization (PSO) is used for the detecting of multiple bad data in the power system state estimation. The PSO, like other meta-heuristic algorithms, can handle constrains that would be troublesome in classical mathematical approach. However, population based algorithms require higher computing time to find optimal point. This shortcoming is overcome by a parallel processing of PSO algorithm. The parallel PSO algorithm is implemented on a PC cluster system with 8 personal computers.

The proposed approach has been tested on the IEEE-14 and 118 bus systems. The results showed that the binary PSO based procedures behave satisfactorily in the detecting multiple bad data and computing time of parallelized PSO algorithm can be reduced without losing the quality of solution.

Keywords: *Particle Swarm Optimization, multiple bad data, PC cluster system, parallel processing*

1. Introduction

In recent years, to address the challenges of the existing power grid, the new concept of smart grid has emerged. Numerous studies related to smart grid are introduced [1, 2]. Especially, state estimation is becoming more and more important in the deregulation environment of power system and there is an urgent need of faster solution technique for on-line application.

Since the beginning of state estimation studies, the problem of detecting the presence of errors within a given measurements and that of correcting or eliminating these bad data have been recognized importantly [3]. The basic identification procedures may rely on successive elimination of suspect bad data [4] or on the use of non-quadratic criteria in carrying out the state estimation [5, 6].

The use of the normalized residual vector is a very common approach of identification by elimination. In particular, the largest normalized residual (LNR) had problems in correctly identifying multiple interacting bad data especially when they are of the confirming type. In such a case, successive elimination of the measurement with the largest normalized residual may result in the suppression of correct measurements instead of the bad data [7].

In the identification of multiple interacting bad data is re-formulated as a combinatorial problem which is tackled by a branch-and-bound method. Recently, non-deterministic approaches such as artificial neural network, tabu search, and genetic algorithm have been proposed in as a viable strategy for the solution of the bad data identification problem [8-10].

A particle swarm optimization (PSO) is one of the modern heuristic algorithms [11]; it can be applied to nonlinear and discontinuous optimization problems with continuous variables. More applications of PSO are in various fields [12-14]. However, a binary PSO is introduced where the positions of particles in search space can only take on values form the set {0, 1} [15]. However, these heuristic optimization methods based on populations require a lengthy computing time to find an optimal solution.

In parallel processing, problems are divided into several sub problems, and allocated to each processor. This can reduce computing time and enhance computation efficiency [16]. To realize parallel processing, parallel computers like the transputer have been introduced. However, these computers are costly to use. Recently, PC-clustering, one of the types of parallel or distributed processing systems, which is composed of a collection of interconnected workstations or PCs working together as a single, integrated computing resource has been used for parallel computing [17].

In this paper, we used the binary PSO to identify multiple bad data in power system estimation. To overcome the shortcomings of heuristic optimization methods, we proposed parallel processing of the PSO algorithm based on the PC cluster system. The proposed approach was tested in the IEEE-14 and 118 bus systems. From the simulation results, we found that binary PSO behave satisfactorily in the identifying multiple bad data and computing time of parallelized PSO algorithm can be reduced without losing the quality of solution.

2. Formulation of the Problem

Mathematically, the information model used in power system state estimation is represented by the equation (1):

$$z = h(x) + e \tag{1}$$

where, z : $(m \times 1)$ measurement vector,
x : $(n \times 1)$ true state vector,
$h(x)$: $(m \times 1)$ state equation vector ,
e : $(m \times 1)$ measurement error vector,
m : the number of measurements ,
n : the number of state variables.

The usual state variables are the voltage and angle magnitudes, while the measurements are the real and reactive power flows, node injections and voltage magnitudes. The objective function of the state estimation is the same as that of conventional state estimation as follows:

$$\min J(x) = \sum_{i=1}^{m} w_i \,(z_i - h_i(x))^2 \tag{2}$$

where, w_i : a weighting factor of measurement variable z_i .

Commonly, a criterion that is used as state estimation formulation is to minimize the sum of the differences between the estimated and true values. This approach is called weighted least squares estimation.

However, following [7], the identification of multiple interacting bad data can be handled as an optimization problem of combinatorial nature.

When multiple interacting bad data are present within the measurement set, different plausible explanations may be found for the measurement inconsistency. Any of the possible combinations of good and bad measurements can be associated with an ***m***-dimensional decision vector ***b*** in which:

$b_i = 1$ if the ***i****-th* measurement is a bad data,
$b_i = 0$ if the ***i****-th* measurement is a good data

Hence, for an observable system with m measurements there are 2^m possible decision vectors.

The removal of all the bad data can be checked, for example, by means of the χ^2 test; in this case, it is sufficient to verify that $J(x+\Delta x) < C$, where $x+\Delta x$ is the new state estimate and C is the appropriate detection threshold.

The problem of the identification of multiple bad data is then formulated as follow:

$$\begin{aligned} &Min \quad F(b) \\ &subject\ to\text{: } S(b) \text{ is observable} \\ &\qquad\qquad J[\hat{x}(b)] < C(b) \end{aligned} \tag{3}$$

For any given decision vector ***b***, $S(b)$ denotes the corresponding measurement set assuming that only the "good" data are taken into account and the suspect bad data have been eliminated. After re-estimation according to the $S(b)$ measurement set, $\hat{x}(b)$ is the new state vector, $J[\hat{x}(b)]$ is the corresponding value of the performance index, and $C(b)$ is the updated detection threshold.

The objective function is equal to the total number of suspect bad data, that is:

$$F(b) = \sum_{i=1}^{m} b_i \tag{4}$$

3. Particle Swarm Optimization

3.1. Overview of Particle Swarm Optimization

PSO has been developed through simulation of simplified social models. The features of the method are as follows:

1) The method is based on researches about swarm such as fish schooling and a flock of birds.

2) It is based on a simple concept. Therefore, the computation time is short and it requires few memories.

3) It was originally developed for nonlinear optimization problems with continuous variables. However, it is easily expanded to treat problems with discrete variables.

According to the research results for a flock of birds, birds find food by flocking (not by each individual). The observation leads the assumption that all information are shared inside flocking. Moreover, according to observation of behavior of human groups, behavior of each individual (agent) is also based on behavior patterns authorized by the groups such as customs and other behavior patterns according to the experiences by each individual. The assumption is a basic concept of PSO. PSO is basically developed through simulation of a flock of birds in two dimension space. The position of each agent is represented by XY-axis position and the velocity is expressed by v_x (the velocity of X-axis) and v_y (the velocity of Y-axis). Modification of the agent position is realized by using the position and the velocity information.

Searching procedures by PSO based on the above concept can be described as follows: a flock of agents optimizes a certain objective function. Each agent knows its best value so far (*pbest*) and its XY position. Moreover, each agent knows the best value in the group (*gbest*) among *pbest*s, namely the best value so far of the group. The modified velocity of each agent can be calculated using the current velocity and the distance from *pbest* and *gbest* as shown below:

$$v_i^{k+1} = w_i v_i^k + c_1 rand \times \left(pbest_i - s_i^k \right) + c_2 rand \times \left(gbest - s_i^k \right) \tag{5}$$

where,

v_i^k : current velocity of agent *i* at iteration *k*,

v_i^{k+1} : modified velocity of agent *i*,

$rand$: random number between 0 and 1,

s_i^k : current position of agent *i* at iteration *k*,

$pbest_i$: pbest of agent *i*,

$gbest$: gbest of the group,

w_i : weight function for velocity of agent *i*,

c_i : weight coefficients for each term.

Using the above equation, a certain velocity that gradually gets close to *pbest*s and *gbest* can be calculated. The current position (searching point in the solution space) can be modified by the following equation:

$$s_i^{k+1} = s_i^k + v_i^{k+1} \tag{6}$$

Figure 1 shows the above concept of modification of searching points.

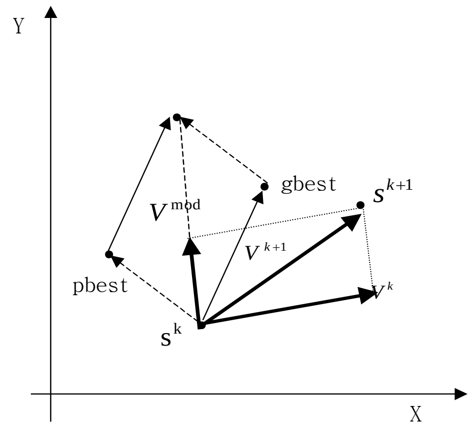

Figure 1. Concept of Modification of a Searching Point

Where, : S^k current searching point,

S^{k+1} : modified searching point,

V : current velocity,

V^{k+1} : modified velocity,

gbest : velocity based gbest,

pbest : velocity based pbest.

3.2. Binary PSO

Most PSO implementations are for continuous search spaces. However, Kenndy and Eberhart introduced a binary PSO where the positions of particles in search space can only take on values from the set {0, 1}. For the binary PSO, particle positions are updated using

$$x_{ij}(t+1) = \begin{cases} 0 & if \quad r_i(t) \geq f(v_{ij}(t)) \\ 1 & if \quad r_i(t) < f(v_{ij}(t)) \end{cases} \tag{7}$$

where

$$f(v_{ij}(t)) = \frac{1}{1+e^{-v_{ij}(t)}}$$

and $x_{ij}(t)$ is the value of the *j-th* parameter of particle P_i at time step t, $v_{ij}(t)$ is the corresponding velocity and $r_i(t) \sim U(0,1)$.

3.3. Handling Problem Constraints

The formulation of the bad data identification problem is well suited to the characteristics of binary PSO, because the unknown vector ***b*** does not need any particular coding since it is already a binary valued vector. The constraints of problem may be handled by penalty techniques, whereby problem can be reformulated as the unconstrained minimization of the following objective function:

$$\tilde{F}(b) = \sum_{i=1}^{m} b_i + K_1 J\left[x(b)\right] + K_2(b) \tag{8}$$

where K_1 is a penalization coefficient, while K_2 introduces a large penalization term when the measurement layout, corresponding to the decision vector $\boldsymbol{b}$, makes the system unobservable. In the proposed procedures K_1 has been taken equal to 1 and K_2 equal to 1000.

4. Implementation Issues

The computation burden is proportional to the number of state estimation computations. For this reason, two methods aiming at reducing the number of state estimations have been designed.

4.1. Parallel PSO Algorithm

The population size is one of the key factors that will affect the search performance of the PSO algorithm for seeking the optimal solution. A larger population size can guarantee a higher chance of obtaining the optimal solution. However, it is obvious that more computing time is needed. To reduce the computing time with the same quality of solutions, a parallel PSO algorithm is proposed and paralleled by the PC cluster system.

The most important issue of parallelizing the PSO algorithm is the exchange model of evolution information. Different methods will result in different performances. The proposed configuration is a kind of parallel algorithm based on the coarse grain model, in which the population is divided into sub-populations that evolve independently. Each sub-population exchanges the required information only between two neighboring sub-populations connected by arrowed lines, as shown in Figure 2.

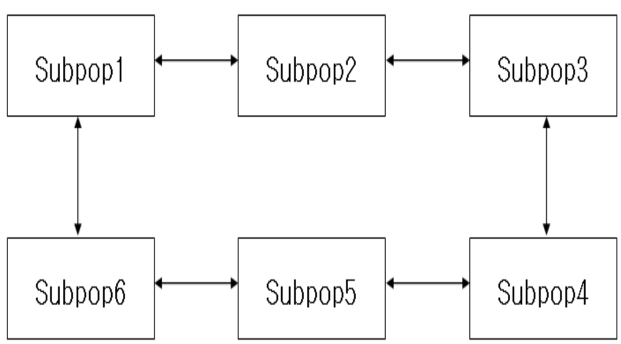

Figure 2. Configuration of Parallel Processing

Each sub-population is allocated in each processor involved in parallel computing. With each processor that can communicate with the neighboring sub-populations, the best solution of each processor is transferred to the neighboring processors by a migration operation with every generation. The major procedures of the parallel PSO algorithm are as follows:

1) Initialize population: Randomly initialize the total population. This initial population is divided into ***N*** sub-populations and allocated to each processor;

2) Evaluate fitness function: Evaluate the fitness value of each particle in the sub-population. Record the ***pbest*** and ***gbest*** of the sub-population;

3) Information exchange: The global best of each sub-population is exchanged between neighboring sub-populations by the migration operation;

4) Update position of each particle: After exchanging the evolution information, update every particle in the sub-population.

4.2. PC Cluster Systems

PC cluster system PC cluster system provides higher availability as well as greater performance by lower cost with interconnecting several PCs or workstations. PC cluster systems are very competitive with parallel machines in terms of the ratio of cost to performance because clustering is one of the types of parallel or distributed processing systems, and is composed of a collection of interconnected low cost PCs working together as single and integrated computing resource. Also, it is easy to add processors in constructing the PC cluster. A basic construction diagram for the PC cluster is shown in Figure 3.

The PC cluster system implemented in this paper is composed of 8 processors. Figure 4 and Table 1 show the picture and the specifications of the 8-processor PC cluster system developed in this paper.

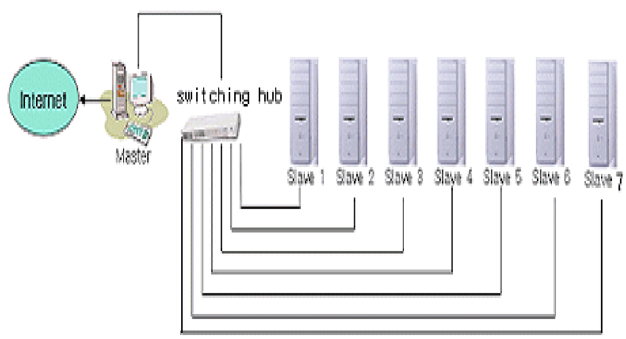

Figure 3. Structure of PC Cluster System

Figure 4. 8-processor PC Cluster System

Table 1. Specification of 8-processor PC Cluster System

CPU	Intel Core2 Duo 2.33GHz
Main Board	GA-P35-DS3
Chipset	Intel P35 chipset
RAM	DDR RAM 1 GB
NIC	Realtek RTL 8168/8111 PCI-E Gigabit Ethernet
Network Switch	3Com 3C16478 Switch
Operating System	Window 2000 Server, Window 2000 Pro
MPI Library	MPICH 1.2.5
Compiler	Visual C++ 6.0

4.3. Reducing the Search Space

The adopted bad data identification approach is formulated as a search problem, in which all the measurement data could be considered for elimination, at least in principle. The search space would contain as many as 2^m points and working with artificial particles of length $\boldsymbol{m}$ would be the simplest choice in view of a PSO implementation. A better efficiency can be achieved by reducing the search space, taking into account that, in actual operation, the number of bad data is not generally large.

After carrying out the first state estimation computation, with the original data set, the vector of normalized residuals is evaluated and re-arranged in decreasing order of absolute values. The bad data are searched within the subset S of measurements with the p largest normalized residuals and the artificial particle length is similarly reduced to $\boldsymbol{p}$. parameter **p** is taken in the range between 10 and 20. The objective functions (4) and (7) have to be modified to reflect the change in vector $\boldsymbol{b}$.

5. Case studies

5.1. Single Processor

The program has been tested in the identification of multiple bad data with reference to IEEE-14 bus systems. All tests have been carried out on an Intel Core 2 2.33GHz and Visual C++. Figure 5 is IEEE-14 bus system and the set of multiple bad data shown in Table 2 has tested.

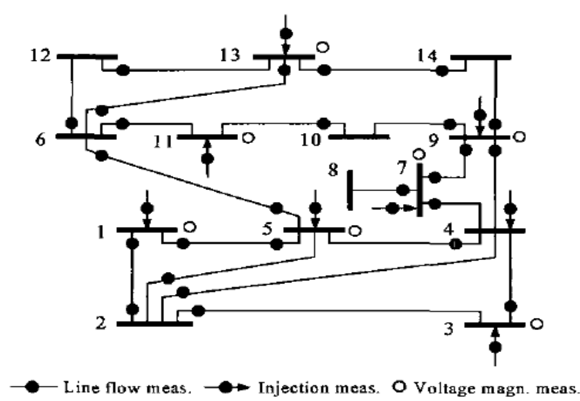

Figure 5. IEEE-14 Bus Systems with Measurements

Table 2. Conforming Bad Data

bad data type	good measurements (p.u.)	bad data values (p.u.)
P_{12}	1.55	0.85
Q_{12}	-0.2	0
P_1	2.32	1.62
Q_1	-.3	-0.1

The repeated LNR method fails in this case because of the conforming nature of the four bad data.

Repeated LNR starts by removing the correct measurement P_{21} which looks bad because the bad data on P_{12} and P_1 are consistent. Correct measurements on real power flows P_{15}, P_{51} are similarly eliminated by the LNR algorithm together with the bad data on P_{12} and P_1. Reactive power flows Q_{21}, Q_{15} and Q_{51} are removed together with the bad data on Q_{12} while the bad data on Q_1 is not recognized by this algorithm.

The proposed procedures successfully identify the four bad data within 200 iterations. In Figure 6, the fitness of the best individual is plotted against the iteration number.

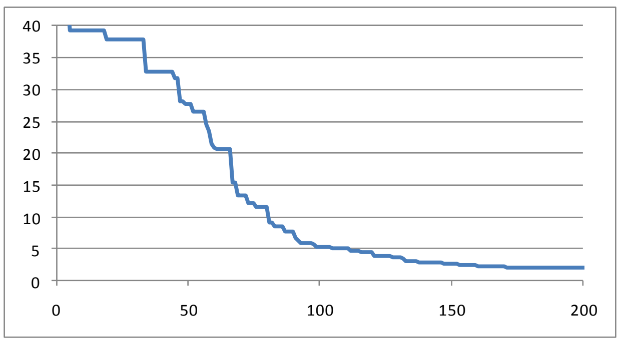

Figure 6. Convergence Histories for Test Case

A nice feature of the proposed binary PSO is that a feasible solution of problem is generally found after very few generations. This means that the system operator can be given a clean state estimation solution, with all the bad data removed very soon during the identification process, possibly leaving time to the binary PSO based procedure to find improved solutions. Such a bad data free solution is generally found within 4s.

5.2. 8-Processor PC Cluster System

In order to test and validate the proposed algorithm, IEEE-118 bus system was used as test case system. In this test system, a set of 716 measurement data were collected, including 358 pairs of active and reactive line flows and 358 pairs of active and reactive line flows in the opposite direction.

The best objective function of PSO using a single processor was 487.9 and that of the proposed parallel PSO with 8 processors had a similar result of 486.5. Both of methods show almost the same quality solution. These results indicate that the parallel PSO can guarantee the same performance with shorter computation time. The computing times for single and parallel PSO were 82.7sec and 17.2sec respectively, with a computational speedup of 4.8 for the parallel PSO.

The results of the IEEE-118 bus test system are summarized in Table 3.

Table 3. Summary of Searching Performance

No. of Processor	Computing time (sec)	Objective function
1	Min.: 72.6 Max.: 96.3 Ave.: 82.7	Min.: 487.9 Max.: 522.4 Ave.: 500.2
2	Min.: 34.6 Max.: 57.2 Ave.: 43.6	Min.: 487 Max.: 529 Ave.: 502.5
4	Min.: 23.1 Max.: 33.3 Ave.: 27.3	Min.: 486.8 Max.: 527.5 Ave.: 501.6
8	Min.: 13.8 Max.: 20.1 Ave.: 17.2	Min.: 485.5 Max.: 525.2 Ave.: 501.7

6. Conclusion

In this paper, binary PSO for identifying multiple bad data are considered in the framework of the least squares state estimation. The identification of bad data is formulated as a combinatorial optimization problem. Its solution gives a binary valued decision vector denoting the erroneous measurements.

To overcome the shortcomings of heuristic optimization methods, we proposed parallel processing of the PSO algorithm based on the PC cluster system.

The proposed approach was tested in the IEEE-14 and 118 bus systems. From the simulation results, we found that binary PSO behave satisfactorily in the identifying multiple bad data and the parallel PSO method can estimate the appropriate state variables within a relatively short calculation time. Therefore, the proposed parallel PSO algorithm can be an appropriate method for the state estimation, which is one of the online functions in power systems.

Acknowledgements

This work was supported by the International Cooperation and Human Resources Development of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government Ministry of Knowledge Economy (No. 20101030100010, No. 20104010100670).

References

[1] P. Sharma, “Transient Stability Investigations of the Wind-Diesel Hybrid Power Systems”, International Journal of Energy, Information, and Communications, vol. 1, no. 1, **(2010)**, pp. 49-63.

[2] D. -J. Kang and S. Park, “A Conceptual Approach to Data Visualization for User Interface Design of Smart Grid Operation Tools”, International Journal of Energy, Information, and Communications, vol. 1, no. 1, **(2010)**, pp. 64-76.

[3] F. C. Schweppe, J. Wildes and D. P. Rom, "Power System Static State Estimation", part 1, 2, and 3, IEEE Trans. Power Apparatus and Systems, vol. PAS-89, **(1970)**, pp. 120-135.
[4] E. Handschin, F. C. Schweppe, J. Kohlas and A. Fiechter, "Bad data analysis for power system state estimation", IEEE Trans. Power Apparatus and Systems, vol. PAS-94, **(1975)**, pp. 329-337.
[5] A. Abur, "A bad data identification method for linear programming state estimation", IEEE Trans. Power Systems, vol. PWRS-5, **(1990)**, pp. 894-901.
[6] R. Baldick, K. A. Clements, Z. Pinjo-Dzigal and P. W. Davis, "Implementing nonquadratic objective functions for state estimation and bad data rejection", IEEE Trans. Power Systems, vol. PWRS-12, **(1997)**, pp. 376-382.
[7] A. Monticelli, F. Wu and M. Yen, "Multiple bad data identification for state estimation by combinatorial optimization", IEEE Trans. Power Delivery, vol. PWRD-1, **(1986)**, pp. 361-369.
[8] N. H. Abbasy and W. El-Hassawy, "Power system state estimation ANN application to bad data detection and identification", Proc. 4-th IEEE AFRICON Conference, vol. 2, **(1996)**, pp. 611-615.
[9] A. Monticelli, "State estimation in electric power systems, A generalized approach", Kluver Academic Publishers, **(1999)**.
[10] S. Gastoni, G. P. Granelli and M. Montagna, "Multiple bad data processing by genetic algorithms", Proc. 2003 IEEE Bologna Power Tech Conference, vol. 1, **(2003)**.
[11] J. Kennedy and R. C. Eberhart, "Particle Swarm Optimization", Proceedings of IEEE international Conference on Neural Networks, vol. 4, **(1995)**, pp. 1942-1948.
[12] S. Singla, D. Kumar, H. M. Rai and P. Singla, "A Hybrid PSO Approach to Automate Test Data Generation for Data Flow Coverage with Dominance Concepts", International Journal of Advanced Science and Technology, vol. 37, **(2011)**, pp. 15-26.
[13] S. Mahapatra, A. K. Jagadev and B. Naik, "Performance Evaluation of PSO Based Classifier for Classification of Multidimensional Data with Variation of PSO Parameters in Knowledge Discovery Database", International Journal of Advanced Science and Technology, vol. 34, **(2011)**, pp. 27-44.
[14] H. Li, L. Li, T. Kim and S. Xie, "An Improved PSO-based of Harmony Search for Complicated Optimization Problems", International Journal of Hybrid Information Technology, vol. 1, no. 1, **(2008)**, pp. 57-64.
[15] J. Kennedy and R. C. Eberhart, "A Discrete Binary Version of the Particle Swarm Algorithm", Proceedings of the Conference of Systems, Man, and Cybernetics, vol. 5, **(1997)**, pp. 4104-4109.
[16] S. H. Chung, K. R. Ryu, S. C. O and T. W. Park, "Parallel Processing System for High Speed Information Retrieval", Parallel Processing System Newsletters. vol. 7, no. 2, **(1996)**, pp. 3-19.
[17] D. A. Lifka, "High performance computing with Microsoft windows 2000", Proc. of the 2001 IEEE Conference on Cluster Computing, **(2001)**, pp. 47-54.

Authors

Hee-Myung Jeong

He received his B.S., M.S. degrees in electrical engineering from Pusan National University. He is currently a Doctor course at Pusan National University. His research interests are intelligent systems, and power system operation and control, especially related to state estimation.

June Ho Park

He received his B.S., M.S., and Ph.D. degrees from Seoul National University of Korea in 1978, 1989, and 1987, respectively, all in Electrical Engineering. He is currently a Professor at Pusan National University. His areas of interest are intelligent systems applications to power systems. Dr. Park has been a member of the IEEE Power Engineering Society, Control System Society, and Systems, Man and Cybernetics Society.