Electrical Power and Energy Systems 88 (2017) 1–12

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Probabilistic load flow calculation with quasi-Monte Carlo and multiple linear regression

Xiaoyuan Xu *, Zheng Yan

Key Laboratory of Control of Power Transmission and Conversion, Ministry of Education, Department of Electrical Engineering, Shanghai Jiao Tong University, No. 800, Dongchuan Road, Shanghai, China

A R T I C L E I N F O

Article history:
Received 6 June 2016
Received in revised form 6 November 2016
Accepted 28 November 2016

Keywords:
Correlation
Multiple linear regression
Non-normal distribution
Probabilistic load flow
Quasi-Monte Carlo
Wind power

A B S T R A C T

In this paper, quasi-Monte Carlo combined with multiple linear regression (QMC-MLR) is proposed to solve probabilistic load flow (PLF) calculation. A distinguishing feature of the paper is that PLF is approached by a low-dimensional problem with the concept of the effective dimension, and thus QMC based on low-discrepancy sequences is used to improve the sampling efficiency of the Monte Carlo simulation (MCS). Moreover, according to the relationship between linear correlation and linear regression, the MLR-based correlation control technique is developed to arrange the orders of samples in order to introduce prescribed dependences between variables. The proposed method is tested with the IEEE 118-bus system. Simulation results indicate that the MLR-based technique is robust and efficient in handling correlated non-normal variables and the proposed method shows better performances in PLF calculation compared with other MCS techniques, including simple random sampling (SRS), Latin hypercube sampling (LHS) and Latin supercube sampling (LSS).

1. Introduction

Following the increased use of renewable energies, the uncertainties linked to their resources have had an increasing impact on the operation of power systems. Deterministic methods cannot fully depict the characteristics of systems and stochastic approaches are now attracting more attention. Probabilistic load flow (PLF) is an efficient stochastic tool to analyze the steady state of power systems considering various random variables [1–6].

Monte Carlo simulation (MCS) is an important approach to solve PLF, and simple random sampling (SRS) is usually used to check the accuracy of other methods. However, SRS requires large samples in order to obtain satisfactory results, thus various methods are proposed to improve the efficiency of MCS. Typical techniques include antithetic variables, control variables, conditional Monte Carlo, importance sampling (IS), stratified sampling, quasi-Monte Carlo (QMC) and so on [7].

IS is useful in the estimation of rare-event probabilities [8], and the cross-entry method, a classic IS technique, has been applied in reliability evaluation [9]. IS can accurately capture the means of system states but it is limited in analyzing probability distributions of output variables. Latin hypercube sampling (LHS) is a stratified sampling method and has been widely used in PLF. LHS consists of sampling and permutation [10]. Random permutation (RP), the basic permutation technique, may introduce undesired correlations between samples of independent variables. Hence Cholesky decomposition (CD) is adopted in [4] to diminish the spurious dependences and the resulting LHS-CD obtains more accurate results than LHS-RP under the same sample sizes. QMC, which is based on low-discrepancy sequences, yields a faster convergence rate than MCS using random sequences and has already been used in some stochastic problems such as probabilistic small signal stability analysis [11]. But its advantage degenerates for high dimensions and thus few researches discuss its performance in PLF. Recently, Latin supercube sampling (LSS), which combines QMC with LHS, has been used to solve PLF and shows better performances than SRS and LHS [12].

Besides the sampling strategy, introducing desired correlations between dependent variables is also important. Non-normally distributed variables are frequently used for the detailed modelling of uncertainties in power systems. However, inducing correlations between non-normal variables is more difficult and the most

Abbreviations: ANOVA, analysis of variation; CD, Cholesky decomposition; GA, genetic algorithm; IS, importance sampling; LDS, low-discrepancy sequence; LHS, Latin hypercube sampling; LSS, Latin supercube sampling; MCS, Monte Carlo simulation; MLR, multiple linear regression; NORTA, normal to anything model; PLF, probabilistic load flow; QMC, quasi-Monte Carlo; RP, random permutation; SRS, simple random sampling.

* Corresponding author.
E-mail address: xuxiaoyuansjtu@gmail.com (X. Xu).

http://dx.doi.org/10.1016/j.ijepes.2016.11.013
0142-0615/

popular method is the Normal to Anything (NORTA) model [13,14]. SRS-NORTA is used in [15] to analyze the impact of wind production on locational marginal prices. LHS-NORTA is developed in [16] to tackle PLF with dependent input variables. The main shortcoming of NORTA is that it becomes infeasible with the increase of correlation matrix dimensions [17].

Another important type of correlation control method is the intelligent optimization algorithm. In [18], a population-based algorithm, genetic algorithm (GA), is designed to control correlations by treating the arrangement of samples as a combinatorial optimization problem. In [19], a simulated annealing (SA) approach is proposed to introduce correlations between variables by exchanging the position of samples. Intelligent optimization algorithms can be robust and efficient for correlation control in small-sample MCS but suffer from a much heavier computational burden in large-scale problems. In addition, other approaches such as polynomial normal transformation [20] and unscented transformation [21] have also been developed to handle correlated variables in PLF.

In this paper, MCS is improved in two aspects, including the sampling strategy and the correlation control. Firstly, the analysis of variation (ANOVA) decomposition and the concept of the effective dimension are described. Then the method to calculate the effective dimension of PLF is designed. Based on the consideration that PLF is a low-dimensional problem, Sobol sequences, which have a good uniformity in low-dimensional projections, are employed to obtain samples of variables. Moreover, according to the relationship between correlation and regression, an MLR-based technique is designed to introduce prescribed dependences among variables. Finally, the proposed PLF calculation method is tested with the IEEE 118-bus system. Simulation results indicate that this new method is more accurate and efficient than popular MCS methods.

The rest of the paper is organized as follows. In Section 2 the theoretical analysis of QMC for solving PLF is given. Section 3 gives the relationship between linear correlation and linear regression and describes the design of the technique to control correlations between variables. The proposed PLF calculation method is given in Section 4, followed by the case studies in Section 5. Finally, the conclusion is presented in Section 6.

2. Quasi-Monte Carlo and probabilistic load flow

In this section the theoretical basis of using QMC in PLF calculation is given. Firstly, the deficiency of QMC in solving high-dimensional problems is described. Then, the ANOVA decomposition and the concept of the effective dimension are introduced. If the effective dimension is low, QMC can be efficient in handling the high-dimensional problem. Finally, the procedure of calculating the effective dimension of PLF is designed.

2.1. Introduction of quasi-Monte Carlo

MCS can be used to solve the following integration

$$Q = \int_{C^s} f(x)dx \tag{1}$$

where f is an integrable function and C^s is the unit cube in s dimensions. Then the MCS estimation of (1) is

$$Q_n = \frac{1}{n}\sum_{i=1}^{n} f(x^{\{i\}}) \tag{2}$$

where $x^{\{i\}}$ are n independent and identically distributed random points drawn from the s-dimensional problem. The estimated error of (2) is given as follows [22]:

$$|Q - Q_n| \leqslant V(f)D_n^* \tag{3}$$

where $V(f)$ is the variation of f in the sense of Hardy and Krause, and $D_n{}^*$ is the discrepancy which reflects the geometric nonuniformity of points in the set.

According to (3), the error of MCS estimation is bounded and dominated by $D_n{}^*$ since $V(f)$ is a constant as long as the function f is given. For random points uniformly distributed over C^s, it has been shown that [23]

$$D_n^* = O\left((\log\log n)^{1/2} n^{-1/2}\right) \tag{4}$$

The convergence rate, $O(n^{-1/2})$, is independent of the dimension of problems, which shows that MCS based on random points is very robust but not efficient. Instead of random points, deterministic low-discrepancy sequences (LDSs) are used in QMC and the discrepancy of LDSs is [24]

$$D_n^* = O((\log n)^s n^{-1}) \tag{5}$$

We see from the comparison of (4) and (5) that the convergence rate of QMC is much faster than MCS in low-dimensional problems. But the $O((\log n)^s\ n^{-1})$ error bound of QMC may not present any improvement over the $O(n^{-1/2})$ error bound of MCS in high dimensions and it should take very large samples before the $O(n^{-1})$ convergence rate can manifest [25].

Although the convergence rate decreases with higher dimensions, QMC has been seen to outperform MCS in some problems with large dimensions and moderate sample sizes. In [11], the probabilistic small signal stability is studied by a QMC-based technique, and the New England ten-generator 39-bus system is analyzed in the case study. There are 55 input variables and QMC obtains relatively accurate results with hundreds of samples. In [20], QMC is used to solve probabilistic optimal power flow. More than 100 input random variables are considered in the test system, and QMC gives better performances compared with SRS for 2000 samples. In [27], researchers at IBM studied the pricing of a five-year discount bond, comprising a 1439-dimensional statistical integral. They observed a QMC speedup of about 150 for an accuracy level of one basis point (i.e., a relative accuracy of 10^{-4}) compared with random Monte Carlo. Some of these successful applications have been explained by the concept of the effective dimension. In the following, this concept is reviewed since it lays the foundation of using QMC in PLF.

2.2. ANOVA decomposition and effective dimension

The function $f(x)$ defined in C^s can be represented in the following form

$$f(x) = f_0 + \sum_{k=1}^{s}\sum_{i_1<\cdots<i_k}^{s} f_{i_1\cdots i_k}(x_{i_1},\ldots,x_{i_k}) \tag{6}$$

where $1 \leqslant i_1 < \cdots < i_k \leqslant s$. Eq. (6) means that

$$f(x) = f_0 + \sum_i f_i(x_i) + \sum_{i<j} f_{ij}(x_i,x_j) + \cdots + f_{12\cdots s}(x_1,x_2,\ldots,x_s) \tag{7}$$

The total number of the summands in (6) is 2^s.

Eq. (6) is called the analysis of variation (ANOVA) decomposition of $f(x)$ if

$$\int_0^1 f_{i_1\cdots i_k}(x_{i_1},\ldots,x_{i_k})dx_m = 0 \quad for\ m = i_1,\ldots,i_k \tag{8}$$

It follows from (8) that the members in (6) are orthogonal and can be expressed as integrals of $f(x)$.

Assume that $f(x)$ is square integrable. Then all the $f_{i1\cdots ik}$ in (6) are also square integrable. Squaring (6) and integrating over C^s we get