Dimension Reduction Method for Probabilistic Power Flow Computation Considering Wind Farms

Haoyan Fu, Lianghong Wu*, Hongqiang Zhang, Zhenzhu Liu, Jingzhe Li
School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China

E-mail: lhwu@hnust.edu.cn

Abstract. With the large-scale integration of renewable energy sources into the power system, a new source of uncertainty is added to the operation planning problem. In this paper, the rank correlation coefficient is introduced to characterize the dependency among random variables in power flow equations, and Nataf transformation is used to map the probabilistic power flow (PPF) problem to the independent standard normal space. Dimension reduction model is introduced to approximation the function relationship between PPF inputs and outputs. Gauss-Hermite quadrature is used to obtain the statistical moments of the univariate function, whereby the statistical moments of outputs of power flow equations are obtained. Testing on an IEEE-118 system, the dimension reduction method is compared with Hong’s point estimate method, it is found the dimension reduction method can improve the accuracy without extra computational burden.

1. Introduction

In order to alleviate the energy and environment pressure, renewable energy sources such as wind power and solar energy are connected to the power grid. The traditional deterministic power flow calculation is difficult to provide a reliable basis for the operation analysis of the power system. Probabilistic power flow regards the network structure, active and reactive power injected form nodes and other uncertainties in the power system as random variables subject to certain probability distribution, and aims to obtain bus voltage, phase angles, and statistics of active and reactive power flows (such as the mean, standard deviations, skewness, kurtosis, etc.). At present, PPF algorithms mainly include Monte Carlo Simulation (MCS) [1,2], Cumulant Method [3,4], and Point Estimation Method (PEM) [5,6].

MCS is a generally used method. As long as a distribution function of correlated non-normal random variables is established, a large number of samples can be generated for the deterministic power flow calculation, and the statistical moments of the output variables can be accurately obtained. However, the MCS requires a huge amount of samples to ensure the accuracy and convergence of the results. In order to speed up the convergence and reduce the calculation time, some researchers suggested to use Latin Hypercube Sampling (LHS) [7,8] to generate samples, but the computational time is still large. Cumulant method employs the first-order Taylor series expansion to linearize the power flow equation. When the random variables are independent of each other, cumulants of PPF outputs can be obtained by using cumulants of PPF inputs, then the Cornish-Fisher expansion is used to construct the probability density function (PDF) and cumulative distribution (CDF) of PPF outputs. But there are some problems: ① when the dependency among input random variables, it is difficult to find the
cumulants of the input variables; ② In order to describe the distribution of output variables better, it needs to obtain the 5th cumulants at least; ③ The probability distribution that can be simulated by Cornish-fisher expansion is limited, some distributions cannot be accurately described (such as partial Beta distribution).

PEM approximates the statistical information of the output variables based on the statistical information of the input variables. Usaola [9] made a systematic comparison of various point estimation schemes and found that the accuracy of the three-point method is much higher than that of the two-point method. However, PEM hold a low accuracy to provide higher statistical moments. In practice, there are different degrees of correlation between the wind speeds of wind farms. In order to obtain accurate and reliable calculation results, PPF calculations have to consider the correlation [10,11]. Researchers have proposed a method of normal transformation that is based on the principle of equal marginal transformation, transform the correlated non-normal random variables into independent standard normal variables. If the pearson linear correlation coefficient is employed to characterize the dependency among random variables, it requires to determine the equivalent correlation coefficient in the normal space for the normal transformation process, which is computationally tedious [12].

In this paper, Nataf transformation is used to handle correlated non-normal random variables in PPF problems. The Spearman rank correlation coefficient is employed to characterize the dependency among random variables, which avoids the solution of integral equations. At the same time, the univariate dimension reduction method (UDR) [13] is introduced to approximate the functional relationship between PPF input and output random variables, and to calculate the statistical moments of the output variables. Finally, the algorithm is verified on the systems of IEEE 118 bus. The results show that the univariate dimension reduction method can improve the accuracy without increasing the computational burden than Hong’s point estimation.

2. Nataf transformation

It is very difficult to construct the joint probability density function of multi-dimensional random variables directly when the non-normal variables are correlated. Compared with other probability distributions, the joint probability distribution function of the multivariate standard normal distribution is easy to construct. Then, the probability distribution of the multivariate non-normal random variables can be described based on the marginal transformation, which is also called Nataf transformation. Suppose \( X=(X_1, \cdots, X_m) \) is a random vector, \( F(X) \) is the cumulative distribution function (CDF) of \( X \). The marginal transformation:

\[
F(X) = \Phi(Z)
\]

\[
X = F^{-1}[\Phi(Z)]
\]

where \( Z=(Z_1, \cdots, Z_m) \) is a standard normal vector, \( \Phi(\cdot) \) is the CDF of \( Z \), \( F^{-1}(\cdot) \) is the inverse CDF of \( X \). Let \( X_1 \) and \( X_2 \) be two arbitrary random variables in \( X \). Based on the transformation in equation 2, \( X_1 \) and \( X_2 \) can be generated by \( Z_1 \) and \( Z_2 \):

\[
X_1 = F_1^{-1}[\Phi(Z_1)], \quad X_2 = F_2^{-1}[\Phi(Z_2)]
\]

Let \( \rho_s \) denotes the correlation coefficient between \( X_1 \) and \( X_2 \). Let \( \rho_t \) denotes the correlation coefficient between \( Z_1 \) and \( Z_2 \):

\[
\rho_s \sigma_1 \sigma_2 + \mu_1 \mu_2 = E[X_1 X_2] = \iiint F_1^{-1}[\Phi(Z_1)]F_2^{-1}[\Phi(Z_2)]\phi(Z_1, Z_2, \rho_t) dZ_1 dZ_2
\]

where \( \mu_1, \mu_2 \) denote the means of \( X_1, X_2 \) respectively, \( \sigma_1, \sigma_2 \) denote the standard deviations of \( Z_1, Z_2 \) respectively, \( \phi(Z_1, Z_2, \rho_t) \) is the joint PDF of two correlated standard normal variables. For a given \( \rho_s \), \( \rho_t \) can be determined by solving the integral equation of equation 4. In general, the integral is not tractable. This paper use Spearman’s rank correlation coefficient to characterize the dependency.
between \(X_1\) and \(X_2\). For the sake of simplicity: the Spearman rank correlation coefficient is recorded as \(\rho_s\).

Suppose \(S_1\) and \(S_2\) are two uniformly distributed random variables. The Spearman rank correlation coefficient of \(X_1\) and \(X_2\) are defined as the linear correlation coefficient of \(S_1\) and \(S_2\):

\[
S_1 = F_1(X_1), \quad S_2 = F_2(X_2)
\]

\[
\rho_s = 12\text{cov}(S_1, S_2) = 12\text{cov}[F_1(X_1), F_2(X_2)]
\]

Via the procedures in equation 1, the Spearman rank correlation coefficient of \(Z_1\) and \(Z_2\) can be calculated as:

\[
\rho_z = 2\sin\left(\frac{\pi}{6}\rho_s\right)
\]

It can be seen that the Spearman rank correlation coefficient is unchanged based on the marginal transformation in equation 1. In particular, for two correlated standard normal random variables, if their linear correlation coefficient is \(\rho_z\) and their rank correlation coefficient is \(\rho_s\), the two variables have the following functional relationship:

\[
\rho_z = 2\sin\left(\frac{\pi}{6}\rho_s\right)
\]

Based on the Spearman rank correlation coefficient, the steps of Nataf transformation are as follows:

- Suppose \(X=(X_1, \ldots, X_m)\) is a random vector with the Spearman rank correlation matrix \(R_s\).
- Via the procedures in equation 8, the linear correlation coefficient matrix of standard normal random vector is obtained:

\[
\mathbf{R}_z = 2\sin\left(\frac{\pi}{6}\mathbf{R}_s\right)
\]

perform Cholesky decomposition on \(\mathbf{R}_z\) to obtain the lower triangular matrix \(L: \mathbf{R}_z = LL^\top\).
- Let \(U=(U_1, \ldots, U_m)\) is an independent standard normal vector, transform \(U\) to the correlated standard normal vector \(Z\). The correlation matrix of \(Z\) would be \(\mathbf{R}_z\).

\[
Z = LU
\]

- Via the procedures in equation 2, transform \(Z\) to the correlated random vector \(X\):

\[
X_i = F_i^{-1}\Phi(Z_i) \quad (i = 1, \ldots, m)
\]

Following the above procedures, an independent standard normal vector \(U\) can be transformed to a correlated random vector \(X\), whose marginal distribution is \(F_i(x_i)\) and Spearman rank correlation matrix is \(R_s\). The Nataf transformation defines an implicit function relationship between \(X\) and \(U\):

\[
X = T(U)
\]

### 3. UDR method

Denote all input random variables in power flow equation as \(X=(X_1, \ldots, X_m)\), denote an arbitrary solution of power flow equation as \(Y\). All input random variables and an arbitrary solution of power flow equation can be expressed by an implicit function:

\[
Y = H(X) = H(X_1, \ldots, X_m)
\]

The goal of probabilistic power flow is to find the statistical moments of the output variable \(y\). Based on the Nataf transformation in Section 2, \(Y\) can be expressed as a function of independent standard normal variables:

\[
Y = H(X) = H[T(U)] = G(U) = G(U_1, \ldots, U_m)
\]
The rth-order raw moment of $Y$ is:

$$E[Y^r] = \int \cdots \int G'(U_1, \cdots, U_m)\phi(U_i) \cdots \phi(U_m) dU_1 \cdots dU_m$$

where $\phi(\cdot)$ is the probability distribution function of an independent standard normal variable. If the tensor product is used to calculate the above m-dimensional integral:

$$E[Y^r] = \sum_{k=1}^{m} P_k G'(t_{1,k}, \cdots, t_{n,k})$$

where $(t_{1,k}, \cdots, t_{n,k})$ and $P_k (k=1, \cdots, n)$ are the kth quadrature nodes and weights, respectively. When the number of inputs $m$ is large, the model in equation 2 based on tensor product would suffer the curse of dimensionality. In reference [14], a method of univariate dimension reduction (UDR) is proposed, which can greatly reduce the amount of computation. UDR uses the sum of univariate functions to approximate the function $Y = G(U)$:

$$G(U) \approx G_0 + \sum_{i=1}^{m} [G(U) - G_0]$$

$$G_0 = G(0, \cdots, 0, 0)$$

$$U = (0, \cdots, U_i, \cdots, 0)$$

The rth-order raw moment of $Y$ is:

$$E[Y^r] = E[G_0 + \sum_{i=1}^{m} [G(U) - G_0]]'$$

Via the procedures in equation 18, calculate the mean and standard deviation of $Y$:

$$\mu = G_0 + \sum_{i=1}^{m} [\mu_i - G_0]$$

$$\sigma^2 = \sum_{i=1}^{m} \sigma_i^2$$

where $\mu_i = E[G(0, \cdots, U_i, \cdots, 0)]$ and $\sigma_i^2 = E[G^2(0, \cdots, U_i, \cdots, 0) - \mu_i^2]$ can be seen as a univariate function on the standard normal random variable $U_i$:

$$\mu_i = \int G(0, \cdots, U_i, \cdots, 0)\phi(U_i) dU_i \approx \sum_{k=1}^{n} P_k G(0, \cdots, t_{i,k}, \cdots, 0)$$

$$\sigma_i^2 = \sum_{k=1}^{n} P_k G^2(0, \cdots, t_{i,k}, \cdots, 0) - \mu_i^2$$

Because $U_i$ is a standard normal random variable, its corresponding numerical integral is Gauss-Hermite quadrature. Table 1 presents the quadrature weights and points of 3(5)-points Gauss-Hermite quadrature.

**Table 1. Weights and nodes of Gauss-Hermite quadrature.**

| $n$ | Weights ($P_k$)            | Points ($t_{i,k}$) |
|-----|---------------------------|-------------------|
| 3   | 0.66666666666667          | 0                 |
|     | 0.16666666666667          | ±1.732050807569   |
|     | 0.94530872048294          | 0                 |
| 5   | 0.3936193215224           | ±0.958572464613   |
It can be known from Table 1 that when the number of Gauss-Hermite quadrature points is odd, one of them is zero. For an m-dimensional input random vector, the number of calculations is: \((n-1)m + 1(n=3,5)\).

4. Computational procedure
The algorithm to solve the PPF problem with correlated input variables via the Nataf transformation and UDR method is as follows.

- For an m-dimensional input random vector \(X=(X_1,\cdots,X_m)\), based on Gauss-Hermite quadrature on 3 and 5 nodes, construct \(2^m(4^m)\) sample vectors \(U_{i,k}=(0,\cdots,t_{i,k},\cdots,0)\) and 1 zero vector \(U_0=(0,\cdots,0,\cdots,0)\).
- For each random variable in the power flow equation, determine its probability distribution \(F_i(X_i)(i=1,\cdots,m)\), and construct the Spearman rank correlation matrix \(R\), obtain the lower triangular matrix \(L\) by Cholesky decomposition of \(R\), and transform the vector \(U\) into correlated vector \(Z\) by equation 10.
- Transform the \(Z\) form the standard normal space to the original space by the marginal transformation, obtain \(X_{i,k}\) and \(X_0\).
- Substitute \(X_{i,k}\) and \(X_0\) into power flow equations, obtain \(m\) samples \(Y_i (i = 1,\cdots,m)\) for each PPF outputs.
- Calculate the mean and standard deviation of \(Y\).

5. Example analysis
The performance of the proposed method is validated using the IEEE 118-bus system. The wind turbine data and nodes to which farms are connected can be found in Table 2. The total installed capacity of the wind turbine is 692MW.

| Wind farm | Node | Group | Rated power (MW) | Number | Wind farm | Node | Group | Rated power (MW) | Number |
|-----------|------|-------|------------------|--------|-----------|------|-------|------------------|--------|
| 1         | 52   | I     |                  | 2      | 49        | 8    | II    |                  | 2      | 41        |
| 2         | 44   | I     |                  | 2      | 25        | 9    | III   |                  | 2      | 27        |
| 3         | 53   | I     |                  | 2      | 13        | 10   | III   |                  | 2      | 18        |
| 4         | 50   | I     |                  | 2      | 7         | 11   | III   |                  | 2      | 22        |
| 5         | 84   | II    |                  | 2      | 18        | 12   | III   |                  | 2      | 31        |
| 6         | 86   | II    |                  | 2      | 14        | 13   | III   |                  | 2      | 21        |
| 7         | 83   | II    |                  | 2      | 29        | 14   | III   |                  | 2      | 41        |

The active power wind speed model of wind field is expressed as:

\[
P(v) = \begin{cases} 
0, & v \leq v_{ci} \\
\frac{P_r(v-v_{ci})}{v_{r}-v_{ci}}, & v_{ci} \leq v \leq v_{r} \\
P_r, & v_{r} \leq v \leq v_{co} \\
0, & v \geq v_{co}
\end{cases}
\]

\[(21)\]

where \(P_r\) is the installed capacity of the wind farm, \(v_{ci}, v_{r}\) and \(v_{co}\) are the cut-in wind speed, rated wind speed and cut-out wind speed respectively. The wind speed is assumed to follow Weibull distribution with scale parameter \(\alpha=10.7\) and shape parameter \(\beta=3.97\). Area-I and Area-II contains four wind farms respectively, and Area-III contains six wind farms. The wind speed of wind farms in Area-I and Area-
II is correlated, while that in Area-III is not. The correlation coefficient matrix between wind farms in Area-I and Area-II are as follows:

\[
R_1 = \begin{bmatrix}
1.00 & 0.86 & 0.82 & 0.93 \\
0.86 & 1.00 & 0.84 & 0.82 \\
0.82 & 0.84 & 1.00 & 0.84 \\
0.93 & 0.82 & 0.84 & 1.00
\end{bmatrix}
\]

\[
R_2 = \begin{bmatrix}
1.00 & 0.87 & 0.85 & 0.88 \\
0.87 & 1.00 & 0.86 & 0.92 \\
0.85 & 0.86 & 1.00 & 0.86 \\
0.88 & 0.92 & 0.86 & 1.00
\end{bmatrix}
\]

Suppose the active load follows the normal distribution, the mean equal to the active load value at the static equilibrium point and the standard deviation equal to 5% of the mean. The loads demands with the correlation coefficient is 0.5. In this section, the probabilistic power flow is calculated based on Hong’s point estimation and the UDR in equation (18). To assess the accuracy of the proposed method, following error indices are employed:

\[
\varepsilon_r = \frac{1}{N} \sum_{k=1}^{N} \left| \frac{\mu_{r,k} - \mu_{r,k}^*}{\mu_{r,k}} \right| \times 100\%
\]  

(22)

where \(\mu_{r,k}\) is the mean (\(r = 1\)) or standard deviation (\(r = 2\)) of \(k\)th output variable given by MCS, \(\mu_{r,k}^*\) is the mean or standard deviation of \(k\)th output variable given by UDR. \(N\) is the number of the output variable in the system. The results obtained by MCS method with \(10^5\) trials are assumed to be accurate, and compare with 2\(m+1\), 4\(m+1\) PEMs. The results are shown in Table 3.

**Table 3.** Errors of Hong’s point estimate method and univariate dimension reduction method (2\(m+1\) ) (4\(m+1\) )

| \(2m+1\) | Mean Voltage | Standard deviation | Voltage | Standard deviation |
|---------|--------------|--------------------|---------|--------------------|
| PEM | 0.001 | 0.001 | 0.028 | 0.028 |
| UDR | 0.001 | 0.001 | 0.006 | 0.004 |
| Angle | 0.034 | 0.034 | 1.89 | 1.81 |
| PEM | 0.023 | 0.023 | 1.25 | 1.02 |
| UDR | 0.03 | 0.03 | 0.43 | 0.35 |
| Active Power | 0.06 | 0.06 | 1.70 | 1.25 |
| PEM | 0.03 | 0.03 | 0.43 | 0.35 |
| UDR | 0.09 | 0.09 | 1.26 | 1.02 |
| Reactive Power | 0.13 | 0.13 | 2.76 | 2.03 |
| PEM | 0.09 | 0.09 | 1.26 | 1.02 |

It can be seen that the accuracy of UDR and PEM is the same when calculating the mean of the output variable; when calculating the standard deviation, the accuracy of UDR is higher than that of PEM. Moreover, whether PEM or UDR, the accuracy of 4\(m+1\) method is higher than that of 2\(m+1\) method. Through mathematical analysis, the approximation form of Hong’s point estimation and UDR to multivariate functions can be restored:

\[
G'(U) = G'_0 + \sum_{i=1}^{m} (G'(U) - G'_0)
\]  

(23)

\[
G'(U) = \left( G'_0 + \sum_{i=1}^{m} (G(U) - G'_0) \right)'
\]  

(24)

It can be seen from the comparison formula (23) and formula (24): when \(r=1\), the approximation form of Hong’s point estimation and UDR method are exactly the same; while when \(r=2\), Hong’s approximation form is omitted Staggered terms, which leads to the low accuracy of the standard deviation.
6. Conclusion
Based on Nataf transformation and UDR, this paper transform the probabilistic power flow problem to an independent standard normal space, and the UDR method is used to solve the statistical moment of the output. Through numerical experiments and theoretical analysis, the following conclusions can be drawn:

- Using Spearman’s rank correlation coefficient to describe the correlation of input random variables is not only simple to calculate, but also not affected by the marginal distribution of input variables.
- When calculating the mean of output variables, the accuracy of UDR and PEM is the same; when calculating standard deviation, because UDR can consider some staggered terms, the calculation accuracy is higher than PEM.

7. Reference
[1] Fang S, Cheng H, Xu G 2016 A Modified Nataf Transformation-based Extended Quasi-Monte Carlo Simulation Method for Solving Probabilistic Load Flow (Electric Power Components and Systems). vol 44 chapter 15 pp 1-10
[2] Morales J M, Baringo L, Conejo A J, et al 2010 Probabilistic power flow with correlated wind sources (Generation Transmission & Distribution). vol 4 chapter 5 pp 641-651
[3] Shi F, Yu Y J, Feng S H, et al 2014 Probabilistic Load Flow Based on Cumulant Method Considering Multi-Slack Balance (Advanced Materials Research). 1070-1072:943-951
[4] Willy Akbar Ajami, Ardiaty Arief, Muhammad Bachtiar Nappu 2019 Optimal power flow for power system interconnection considering wind power plants intermittency (International Journal of Smart Grid and Clean Energy). vol 8 no 3 pp 372-376
[5] Usaola J 2010 Probabilistic load flow with correlated wind power injections (Electric Power Systems Research). vol 80 chapter 5 pp 528-536
[6] J Üpping, F Bollhöfer, K Forche 2020 Size of a distributed electrical storage for a rural area with a wind farm (International Journal of Smart Grid and Clean Energy). vol 9 no 1 pp 1-7
[7] Ontoseno Penangsang, Primaditya Sulistijono, Suyantoa 2014 Optimal power flow using multi-objective genetic algorithm to minimize generation emission and operational cost in micro-grid(International Journal of Smart Grid and Clean Energy). vol 3 no 4 pp 410-416
[8] Chen Y, Wen J, Cheng S 2013 Probabilistic Load Flow Method Based on Nataf Transformation and Latin Hypercube Sampling (IEEE Transactions on Sustainable Energy). vol 4 chapter 2 pp 294-301
[9] Usaola J, M 2009 Probabilistic load flow in systems with wind generation(Generation Transmission & Distribution). vol 3 chapter 12 pp 1031-1041
[10] Xu X, Yan Z 2016 Probabilistic load flow evaluation considering correlated input random variables ( International Transactions on Electrical Energy Systems). vol 26 chapter 3 pp 555-572
[11] Chen C, Wu W, Zhang B, et al 2015 Correlated probabilistic load flow using a point estimate method with Nataf transformation (International Journal of Electrical Power & Energy Systems). 65:325-333
[12] Li H S, Lu Z Z, Yuan X K 2008 Nataf transformation based point estimate method (Chinese Science Bulletin). vol 53 chapter 17 pp 2586—2592
[13] Mohamed M H Adam, Naem Hannoon, Muhamed Nabil bin Hidayat 2019 Performance analysis of grid-connected and stand alone wind farm (International Journal of Smart Grid and Clean Energy). vol 8 no 5 pp 611-619

Acknowledgments
This work was supported by National Natural Science Foundation of Hunan Province, China(2018JJ2137); The Outstanding Youth Project of Education Department of Hunan Province of China (19B200);