Robust Data-Driven Linear Power Flow Model With Probability Constrained Worst-Case Errors

Yitong Liu, Student Member, IEEE, Zhengshuo Li, Senior Member, IEEE, and Junbo Zhao, Senior Member, IEEE

Abstract—To limit the probability of unacceptable worst-case linearization errors that might yield risks for power system operations, this letter proposes a robust data-driven linear power flow (RD-LPF) model. It applies to both transmission and distribution systems and can achieve better robustness than the recent data-driven models. The key idea is to probabilistically constrain the worst-case errors through distributionally robust chance-constrained programming. It also allows achieving the linearization accuracy for a chosen operating point. Comparison results with three recent LPF models demonstrate that the worst-case error of the RD-LPF model is significantly reduced over 2- to 70-fold while reducing the average error. A compromise between computational efficiency and accuracy can be achieved through different ambiguity sets and conversion methods.

Index Terms—Data-driven, distributionally robust, linear power flow, worst-case errors.

I. INTRODUCTION

Constructing a linear power flow (LPF) model for a future system operating state is of great interest to system operators and useful for several system operation functions. LPF models can be generally categorized into two types whose main difference relies on whether the historical data are utilized. In the first type, AC power flow equations are linearized completely depending on mathematical tools [1] such as Taylor expansion [2]. In the second type, with the utilization of historical measurements, data-driven techniques such as partial least squares regression [1], [3], least squares regression [4], [5], support vector regression [6], Gaussian process regression [7], etc., are employed on top of or in place of the aforementioned linearization. These data-driven LPF (DD-LPF) models [1]–[7] have demonstrated improved average linearization accuracy as compared to the first-type model.

Nevertheless, an LPF model with moderate average linearization errors but notable worst-case errors might be improper for certain practical applications. The worst-case errors mean the maximum errors under some system operating states that significantly deviate from the chosen one. For example, it may cause high risks when certain security constraints are on the verge of the limits, where the operators’ tolerance of the worst-case error is low. Thus, researchers should consider reducing both the average error and the worst-case error [8] when constructing an LPF model. Indeed, for the system with increased penetration of variable renewable energy, the future system operating state may notably deviate from the chosen one in which case unacceptable worst-case errors are likely to occur, see Fig. 1 for illustrations. However, most LPF models, such as those in [1]–[7], fail to explicitly bound the worst-case error. Reference [9] derives the error-bound of the branch LPF model with multiple times of inequality relaxations which may lead to a relatively loose bound. Also, [9] does not utilize historical operating data to improve the linearization accuracy. Reference [8] reduces the error-bound in a data-driven way, but it does not consider the uncertainty of probability distributions of the system operating state.

To bridge this gap, this letter proposes to construct a robust data-driven LPF model (RD-LPF) whose probability of yielding unacceptable worst-case linearization errors is explicitly constrained through the data-driven distributionally robust (DR) chance-constrained programming (DR-CCP). Hence, this RD-LPF model is more robust than the extant DD-LPF models. To be more specific, since the future system operating state is unknown, historical operating data are leveraged to build an ambiguity set about the probability distribution of this stochastic future state, over which the unacceptable worst-case errors are probabilistically bounded through chance constraints. The ambiguity set can be moment-based [10] or φ-divergence based [11], etc., depending on operators’ preference. In this way, no accurate prior assumption on the distribution of the stochastic future state is required. Also, this RD-LPF model can ensure a guaranteed linearization accuracy for a chosen operating point by considering it as the objective function of the DR-CCP. Case
studies will show that the proposed RD-LPF achieves improved average and worst-case accuracy for linearization as compared to [4], [5], and [9]. Also, this proposed model is generic and applicable to both transmission and distribution systems.

II. RD-LPF MODEL

We will first present a generic formulation of the DD-LPF model and its robustified version through the DR-CCP, yielding the proposed RD-LPF model. In this letter, \(\mathbb{I}\) denotes the modulus of a (complex) number.

A. Generic Formulation of DD-LPF Models

Most of the DD-LPF models describe a linear relationship between the dependent variable \(y\) and independent variable \(x\) in the way of \(y = A(w)x + b(w)\) where \(w\) is the model parameter vector to be determined based on historical data and constructs the matrix \(A(w)\) and the vector \(b(w)\) that form the linear relationship between \(y\) and \(x\). The specific meanings of \(y\) and \(x\) vary for different DD-LPF models and they will be elaborated by two examples later.

Given the historical dataset \(\{(x^1, y^1), \ldots, (x^K, y^K)\}\), \(w\) can be obtained by solving the following optimization problem:

\[
w = \arg \min_w \sum_{k=1}^{K} \sum_{i=1}^{N_y} r_i(y^k, x^k, w).
\]

where \(y^k\) and \(x^k\) denote the \(k\)-th historical data, respectively; \(r_i(\cdot)\) denotes an error indicator of measuring the linearization error rated to the \(i\)-th element of \(y^k\); \(K\) and \(N_y\) are the numbers of historical data and \(r_i(\cdot)\), respectively.

To facilitate understanding of this generic formulation, two DD-LPF models are taken as examples below to explain the specific meaning of the variables mentioned above.

The first example is the least-squares distribution factors (LSDF) model in [4], suitable for transmission systems. To relate it with the above generic model, let \(y\) be the vector consisting of branch active and reactive power flows; let \(x\) be the vector consisting of nodal active and reactive power injections; \(r_i(y^k, x^k, w) = |y^k_i - a_i(w)x^k_i|^2\), where \(y^k_i\) is the \(i\)-th element of \(y^k\), and \(a_i(w)\) is the \(i\)-th row of \(A(w)\); each row of \(A(w)\) consists of the elements in \(w\) related to the corresponding row of \(y\), and \(b(w) = 0\).

Another example is the DD-LPF model in [5], which is applied to distribution systems. Similarly, let \(y\) and \(x\) be the vectors consisting of nodal voltages and power injections, respectively; and let \(w\) be the parameters to be solved by (1) in which \(r_i(y^k, x^k, w) = |\varepsilon^k_i(y^k, x^k, w)|^2\), where \(\varepsilon^k_i(\cdot)\) is the \(i\)-th row of the error-related matrix \(\varepsilon(\cdot)\) consisting of \(w\) and voltages whose detailed expressions can be found in [5]. Moreover, \(A(w)\) and \(b(w)\) can be obtained through simple mathematical operation of nodal admittance matrix and \(w\) [5].

B. General Idea of Constructing the RD-LPF Model

The solution of (1) is a sample-average approximation [12] and does not explicitly constrain the magnitude and occurrence probability of worst-case linearization errors. Hence, the resultant DD-LPF models are likely to generate the worst-case errors that may cause risk for system operations. This issue is tackled by constructing an RD-LPF model. The key idea is that the probability of yielding unacceptable worst-case errors should be explicitly constrained. To this end, when determining the model parameter \(w\), one need to modify (1) by integrating an explicit probability constraint to ensure that the value of the error indicator not exceeding an acceptable value \(\delta\) is credible.

Meanwhile, because the future system state may not emerge equally among all the possible values and is often likely to be close to a certain chosen (e.g., contemporary) operating point, operators may well be interested in minimizing the error indicator for this chosen point. This allows satisfying linearization accuracy for this chosen point. It will be achieved by formulating it as the objective function in constructing the RD-LPF model.

C. Method

Forcing the solution to be robust against any possible uncertain scenario, including the rare ones, leading to the conservativeness of the solution [13]–[14]. By contrast, we tend to probabilistically constrain the unacceptable worst-case error for the unknown future system state \(x\), as shown in (3). If the DD-LPF model \(y = A(w)x + b(w)\) is constructed by the parameters solved by the problem in (2)-(3) rather than (1), the DD-LPF then becomes the RD-LPF model. Note that the notation \(\xi\) rather than \(x\) is used to represent to the system state because it is deemed to be stochastic variables in the RD-LPF model.

\[
\min_w \sum_{i=1}^{N_y} r_i(\xi^c, w)
\]

\[
s.t. \quad \mathbb{P}\{r_i(\xi, w) \leq \delta_i\} \geq 1 - \epsilon, \quad i = 1, \ldots, N_y.
\]

where \(\xi^c\) denotes the \(c\)-th historical data pertinent to the contemporary or a typical operating point; \(\mathbb{P}\{\cdot\}\) denotes a probability distribution; \(r_i(\cdot)\) is a linear equivalent or approximation of \(r_i(\cdot)\), for example, for [4]'s \(r_i(\cdot)\), \(r_i(\cdot) = y^k_i - a_i(w)x^k_i\), then both \(\hat{r}_i(\cdot) \leq \delta_i\) and \(r_i(\cdot) \leq -\delta_i\) should be satisfied, and for [5], \(\varepsilon(\cdot)\) is a complex number and \(r_i(\cdot)\) denotes its real part or imaginary part; \(\delta_i\) is the upper bound of the acceptable worst-case error; \(\epsilon\) is the risk level, i.e., the maximum allowed probability of constraint violation.

In (2), we minimize the error of the chosen typical operating point. Meanwhile, in (3), the worst-case errors are constrained considering the uncertainty of the system operating state and its distribution, which contributes to the robustness of the solution. Specifically, constraint (3) indicates that the probability of the equivalent error indicator \(\hat{r}_i\) exceeding the acceptable value \(\delta_i\) should be smaller than the risk level. It is often difficult, if not impossible, to accurately presume \textit{a priori} probability distribution function \(f(\xi)\) of \(\xi\), which can be addressed by building an ambiguity set \(D\) that is composed of a set of probability distributions equipped with some common characteristics [11]. Then, (3) can be reformulated to a DR chance constraint as:

\[
\inf_{f(\xi) \in D} \mathbb{P}\{r_i(\xi, w) \leq \delta_i\} \geq 1 - \epsilon, \quad i = 1, \ldots, N_y.
\]

Based on the mean vector \(\mu_0\) and covariance matrix \(\Sigma_0\) of the historical data, \(D\) is constructed as a moment-based ambiguity set. Then, we can apply equivalence to (4) for being tractable. Specifically, we can recast (4) into an integral format and convert it into its conic dual counterpart, whose semi-infinite constraints are further replaced by finite semi-definite programming (SDP)

\[\text{(5)}\]
constraints [10], [12]; thus (4) is recast into (5). Detailed derivations can be referred to [10], [12]. The RD-LPF in this case is remarked as Model-1 hereafter.

\[
\begin{align*}
\gamma_2 \Sigma_0 \cdot G_i + 1 - l_i + \sum_i \beta_i \leq \epsilon_i \lambda_i, \\
\lambda_i \geq 0, \quad \eta \geq 0, \quad \lambda_i \geq 0, \\
\end{align*}
\]

(5)

where the operation \( \mathbf{X} \cdot \mathbf{Y} \) denotes the trace of \( \mathbf{XY} \); \( G_i, l_i, H_i, \), \( \alpha_i, \), and \( \beta_i \) denote dual variables of the dual problem; \( \lambda_i \) denotes the nonnegative decision variable replacing the semi-infinite constraints with finite SDP constraints to construct (5); \( \gamma_1 \) and \( \gamma_2 \) are the variables that represent the decision maker’s risk preference; \( \lambda_i \) denotes the variables that represent the decision maker’s risk preference; \( A^w \) and \( b^w \) are vectors consisting of \( w \) [10], [12]. In this way, (4) becomes tractable. But the computational efficiency of the SDP model in (5) will be low if the scale of the optimal variables is relatively large.

In summary, the model in (2)-(3) can restrict the worst-case errors in an operational region under uncertainty and keep a satisfying accuracy for a chosen point in the meantime. Moreover, the proposed RD-LPF model is suitable for both transmission and distribution systems, verified in case studies.

### D. Discussions

There are effective alternatives for building the ambiguity set, e.g., the \( \phi \)-divergence based ambiguity set [11], or the \( \phi \)-Wasserstein distance ambiguity set [15]. For example, in the case of \( \phi \)-divergence ambiguity set, the DR chance constraint in (4) is equivalent to a traditional chance constraint [11] with a chosen reference distribution \( P_0 \), as shown in (6), and the RD-LPF in this case is remarked as Model-2 hereafter.

\[
\mathbb{P}_0 \{ \hat{r}_i (\hat{z}, w) \leq \delta_i \} \geq 1 - \epsilon_{i+}, \quad i = 1, \ldots, N_y. \tag{6}
\]

where \( \epsilon_{i+} = \max \{ \epsilon_i, 0 \} \). Taking the KL divergence as an example, \( \epsilon_i = \inf_{\epsilon \in (0,1)} (e^{-d \epsilon} - 1)/(z - 1) \), where \( z \) is the decision variable, and \( d \) denotes the tolerance of the distance between the particular density function and the reference one. Notice that (6) is tractable if we take some special distributions, such as the Gaussian distribution as the reference distribution. For general cases, (6) is intractable but sometimes can be approximately recast into a solvable form, as shown in [16]. In all, the formulation of ambiguity sets can be determined following the principles shown in [11], and we will analyze the accuracy and computational efficiency of Model-1 and Model-2 in case studies.

In addition, researchers may also think of two other models, i.e., the Wasserstein distance-based worst-case expectation model in (7) and the regularization model in (8) as alternatives to the above DR-CCP model, where \( D_p(\epsilon) \) denotes the \( p \)-Wasserstein distance ambiguity set with radius \( \epsilon \) and \( \alpha \) denotes the regularization coefficient. The advantage of (7) is that it is always feasible and needs fewer parameters. However, it is inconvenient to separate the errors of the typical operating point and minimize them intentionally in (7). Meanwhile, though (8) contributes to improving robustness and its solution is relatively simple, it does not explicitly constrain unacceptable worst-case linearization errors. These two models are also compared with DR-CCP model in case studies.

\[
\begin{align*}
\min_w & \quad \max_{\mathbf{P} \in D_p(\epsilon)} \sum_i \mathbb{E}_\mathbf{P} \left[ r_i (\hat{z}, w) \right] \\
\min_w & \quad \sum_i \sum_k r_i^2 (\hat{z}^k, w) + \alpha \| w \|_2
\end{align*}
\]

(7)

(8)

### III. Case Studies

The linearization error of our model is compared with the DD-LPF models in [4], [5], and the error-bound model in [9]. Since the models in [4] and [9] are mainly for transmission systems, the IEEE 118-bus and Polish 2383-bus transmission systems from the MATPOWER are used for comparisons. As the model in [5] is for distribution systems, the IEEE 123-node and 8500-node three-phase distribution systems from OpenDSS are used for comparison. The historical data are obtained by randomly changing the net load level within \( \pm 20\% \) of the base net load level, where the net load fluctuations consist of those from renewable energy and loads. The sizes of the historical data sets are 200, 500, 200, and 500 for the above four systems respectively. Moreover, we set \( \epsilon_i = 0.05 \) [12], \( \gamma_1 = 0 \) and \( \gamma_2 = 1 \) [10], and \( d = 0.1 \) [12]. The equivalent SDP model in (5) is solved by MATLAB/MOSEK and the model in (6) is solved by MATLAB/GUROBI. The codes are implemented on a PC with a 64-bit Intel Core i7 2.9 GHz CPU and 16 GB of RAM.

Linearization errors are measured in terms of relative errors, where the exact values of the power flow are obtained by the power flow engines in MATPOWER and OpenDSS. Tables I and II show the average and worst-case relative errors pertinent to the voltage of our model and the others for the net load levels.
being 60%, 80%, 100%, 120%, and 140% of the base net load level. The abbreviation WC denotes the worst-case. The results indicate that compared with the DD-LPF models in [4] and [5], both the average and the worst-case errors of our two models are notably smaller, e.g., reduced 2- to 3-fold. Table I also indicates that as compared to the error-bound model in [9], the worst-case error of our model is reduced by approximately 10- to 70-fold. The relatively large worst-case error of [9]'s model might be due to the loose error-bound, which is caused by the multiple times of inequality relaxations and relatively imprecise operating parameters without leveraging historical data. In addition, it should be noted that the training and testing data are different so the errors shown in the two tables are independent of the over-fitting.

Next, we will show the comparison between our models and the two models in (7) and (8). We conducted multiple times of tests based on cross-validation and select the parameters \( \varepsilon = 0.1 \) and \( \alpha = 0.01 \) which perform best and the results are shown in Table III. The comparison between Tables I, II, and III indicates that our model generally achieves smaller worst-case errors than the other two models at least under the above parameter settings. Moreover, since we consider minimizing the error of a chosen typical operating point as the objective function, the accuracy of our model around the chosen point (selected as the 100% net load level point) is higher. This demonstrates that compared with the two models, our model can restrict the worst-case error effectively while keeping a satisfying accuracy around the chosen operating point.

Table IV presents the solution time of different models. Since Model-1 is an SDP based model, its solution time is significantly affected by the number of decision variables determined by \( r_i(\cdot) \). For example, Table IV indicates that when \( r_i(\cdot) \) comes from [5], the average solution time of both test systems for each node is about \( 10^2 \)-to-\( 10^3 \)-s. However, the number of decision variables for [4] are related to the system scale which leads to about tens of minutes of solution time for the 2383-bus system. The computational efficiency can be enhanced with the utilization of \( \psi \)-divergence-based ambiguity set if the problem can be equivalent to a traditional chance constraint. System operators can choose which kind of ambiguity set to be used according to the actual demand as discussed fore and our proposed model is flexible enough to offer that.

### IV. CONCLUSION

This letter proposes constructing an RD-LPF model through DR-CCP that explicitly constrains the probability of unacceptable worst-case errors based on two kinds of ambiguity sets. This model is suitable for both transmission systems and distribution systems. Case studies confirm that compared with the recent LPF models, the RD-LPF model's worst-case errors can be reduced over 2- to 70-fold and its average error is also reduced. Future works will focus on testing more efficient computing algorithms, e.g., parallel alternating direction method of multipliers-based methods, to deal with constraints in the form of SDP.

### REFERENCES

[1] Y. Tan, Y. Chen, Y. Li, and Y. Cao, “Linearizing power flow model: A hybrid physical model-driven and data-driven approach,” *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 2475–2487, May 2020.

[2] Z. Yang, K. Xie, J. Yu, H. Zhong, and Q. Xia, “A general formulation of linear power flow models: Basic theory and error analysis,” *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1315–1324, Mar. 2019.

[3] Y. Liu, N. Zhang, Y. Jiang, and C. Kang, “Data-driven power flow linearization: A regression approach,” *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2569–2580, May 2019.

[4] Z. Shao, Q. Zhai, J. Wu, and X. Guan, “Data-based linear power flow model: Investigation of a least-squares based approximation,” *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4246–4258, Sep. 2021.

[5] Y. Liu, Z. Li, and Y. Zhou, “Data-driven aided linear three-phase power flow model for distribution power systems,” *IEEE Trans. Power Syst.*, vol. 37, no. 4, pp. 2783–2795, Jul. 2022.

[6] J. Chen, W. Li, W. Wu, T. Zhu, Z. Wang, and C. Zhao, “Robust data-driven linearization for distribution three-phase power flow,” in *Proc. IEEE 4th Int. Conf. Energy Internet Energy Syst. Integration*, 2020, pp. 1527–1532.

[7] P. Pareek, W. Yu, and H. D. Nguyen, “Optimal steady-state voltage control using gaussian process learning,” *IEEE Trans. Ind. Inform.*, vol. 17, no. 10, pp. 7017–7027, Oct. 2021.

[8] Y. Liu, B. Xu, A. Botterud, N. Zhang, and C. Kang, “Bounding regression errors in data-driven power grid steady-state models,” *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 1023–1033, Mar. 2021.

[9] Z. Fan, Z. Yang, and J. Yu, “Error bound restriction of linear power flow model,” *IEEE Trans. Power Syst.*, vol. 37, no. 1, pp. 808–811, Jan. 2022.

[10] Y. Zhang, Y. Shen and, J. L. Mathieu, “Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads,” *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1378–1388, Mar. 2017.

[11] W. Wei, “Data-driven robust stochastic program,” in *Tutorials on Advanced Optimization Methods*. New York, NY, USA: Springer, 2020, pp. 134–135.

[12] R. Jiang and Y. Guan, “Data-driven chance constrained stochastic program,” *Math. Prog.*, vol. 158, no. 1, pp. 291–327, Jul. 2016.

[13] M. Chamanbaz, F. Dabbene, and C. M. Lagoa, “Probabilistically robust ac optimal power flow,” *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 3, pp. 1135–1147, Sep. 2019.

[14] R. Tempo, E. W. Bai, and F. Dabbene, “Probabilistic robustness analysis: Explicit bounds for the minimum number of samples,” in *Proc. IEEE 35th Conf. Decis. Control*, vol. 3, 1996, pp. 3424–3428.

[15] D. K. Z.Chen and W. Wiesemann, “Data-driven chance constrained programs over wasserstein balls,” Sep. 2018. [Online]. Available: https://arxiv.org/abs/1809.00210

[16] R. T. Rockafellár and S. Uryasev, “Optimization of conditional value-at-risk,” *J. Risk*, vol. 2, pp. 21–24, 2002.