Adaptive Power Flow Approximations with Second-Order Sensitivity Insights

Paprapee Buason,[§] Sidhant Misra,[†] Jean-Paul Watson,[*] and Daniel K. Molzahn[§]

Abstract—**The power flow equations are fundamental to power system planning, analysis, and control. However, their inherent non-linearity and non-convexity can pose significant challenges when solving problems that involve these equations. To address these challenges, linear approximations are frequently used to simplify computations at the potential cost of compromising accuracy and feasibility. Typical power flow approximations rely on general assumptions that are not always valid. Additionally, many existing linear approximations, such as the first-order Taylor series, overlook important information contained within higher-order terms. To gain a deeper understanding of the relationships among variables within a specific operational range and to achieve more accurate approximations, this paper considers a second-order sensitivity analysis of the power flow equations. By modeling the curvature of the power flow manifold, this paper uses the second-order sensitivities in an importance sampling method to improve the accuracy of adaptive power flow linearizations. Furthermore, this paper explores approximations formulated as rational functions with linear numerators and denominators that can be rewritten as linear constraints in power system optimization problems. These approximations, inspired by the Padé approximant, utilize second-order power flow sensitivities. This approach is extended to enhance accuracy beyond what linear functions can achieve across various operational scenarios, and can also be constructed to be conservative (over- or under-estimate a quantity of interest).**

I. Introduction

The physics of power flow in a power network is described by the AC power flow (AC-PF) equations that relate bus voltages and line power flows to active and reactive power injections. These equations therefore must be included in a variety of optimization problems used for planning and decision making purposes. The implicit nonlinear nature of the power flow equations make the target optimization problems challenging to solve. Problems that involve additional complex features, such as uncertainty quantification problems for managing stochastic load and mixed-integer and bilevel problems for unit commitment and expansion planning, can become intractable due to the non-convex AC-PF equations.

To obtain tractable formulations, a variety of power flow approximations are used in place of the AC-PF equations within optimization problems. The DC approximation for high voltage transmission systems [1], LinDistFlow for distribution systems [2], and the first-order Taylor expansion are among the most commonly used approximations. There is a rich literature on alternative approximation approaches and methods to enhance existing approximations [3]. A large fraction of these approximations are unsurprisingly linear, since they result in linear constraints that significantly improve tractability. For more complex and large-scale mixed integer and multi-level problems, preserving linearity of constraints may be the only viable strategy to obtain tractable solution methods.

Several of the well-known approximations are derived from first principles and ignore the specifics of the operating ranges encountered in practice. For example, the DC approximation relies on negligible voltage variations and losses. The generality of these approximations, however, comes at the price of large errors that can lead to constraint violations and sub-optimal cost. To address this issue, several recent works have considered *adaptive* power flow approximations that are tailored to be accurate within a given operating range. These include the optimization-based approaches [4], [5] and data-driven approaches [6]–[8]; see [9]–[11] for recent surveys.

To further improve the utility of adaptive linear approximations for optimization problems, [6] introduced the concept of *conservative* linear approximations (CLAs) that are designed to either under- or over-estimate a quantity of interest such as bus voltage magnitude. Since the goal of these approximations is to formulate inequality constraints in an optimization problem (e.g., satisfying limits on voltage magnitudes), introducing conservativeness during their construction yields improved performance in terms of constraint enforcement.

Data-driven formulations, which rely on randomly sampled points from a given operating range, are the most widely used methods for constructing adaptive power flow approximations. Tractability, scalability, and suitability for parallel implementation are the primary reasons for their popularity. Naturally, the accuracy of the approximations are heavily dependent on the quantity and quality of samples used in their construction.

This paper develops an *importance sampling* method to construct linear and conservative linear approximations. The method aims to include the most informative samples to quickly improve approximation quality while keeping the training process efficient. Our approach is based on second-order sensitivity information. By preferentially including more samples from a relatively low-dimensional subspace with high curvature, we obtain highly accurate linear approximations using far fewer samples compared to random sampling.

We also introduce a class of approximations based on rational functions with linear numerator and denominator. The main motivation behind using these rational functions

§: School of Electrical and Computer Engineering, Georgia Institute of Technology. {pbuason6, molzahn}@gatech.edu
†: Los Alamos National Laboratory. sidhant@lanl.gov
*: Lawrence Livermore National Laboratory. watson61@llnl.gov

is that they result in linear inequality constraints when used in an optimization formulation. At the same time, expanding the approximating class of functions from linear to rational provides better approximation error. We first introduce a multivariate generalization of the Padé approximant which can be constructed at a given operating point using the second-order sensitivities and serves as a rational generalization of the second-order Taylor expansion [12]. We then provide a data-driven training method similar to [6] to construct rational and conservative rational approximations over an operating range.

The paper is organized as follows: Section II covers background material on the power flow equations as well as adaptive linear and conservative linear approximations. Section III describes the importance sampling approach for data-efficient construction of linear approximations. Section IV introduces rational approximations, including the multivariate generalization of the Padé approximant for the power flow equations. Section V provides numerical results demonstrating the improvements obtained from our approach. Section VI concludes the paper along with directions for our future work.

II. Background

In this section, we first present the power flow equations and then describe a so-called "conservative linear approximation" (CLA) approach for both voltage magnitudes, as we previously proposed, and a new extension to current flows.

A. The power flow equations

Consider an N-bus power system. Let the length-N vectors P, Q, and $V\angle\theta$ denote the active and reactive power injections and voltage phasors, respectively, at each bus. Designate a reference bus where the voltage phasor is set to $1\angle 0^\circ$ per unit. We use the subscript $(\cdot)_i$ to represent a quantity at bus i and the subscript $(\cdot)_{ik}$ to represent a quantity from or connecting bus i to k. The AC power flow equations for bus i are:

$$P_i = V_i^2 G_{ii} + \sum_{k\in\mathcal{B}_i} V_i V_k (G_{ik}\cos\theta_{ik} + B_{ik}\sin\theta_{ik}), \quad \text{(1a)}$$

$$Q_i = -V_i^2 B_{ii} + \sum_{k\in\mathcal{B}_i} V_i V_k (G_{ik}\sin\theta_{ik} - B_{ik}\cos\theta_{ik}), \quad \text{(1b)}$$

where $G + jB$ (with $j = \sqrt{-1}$) is the admittance matrix.

B. Conservative linear approximations

To model voltage and current limits, many optimization problems include the power flow equations (1) as constraints. The nonlinearity of these equations contributes to the complexity of the problems, often making them difficult to solve. To address this challenge, we previously introduced a sample-based CLA approach in [6]. This linear approximation seeks to over- or under-estimate a specified quantity of interest (see Fig. 1). CLAs approximate the nonlinear power flow equations using a sample-based approach that enables parallel computation, i.e., the CLA of each quantity of interest can be computed concurrently. Constructing a CLA involves sampling power injections over an operating range of interest, computing the power flow equations for each sample, and then solving

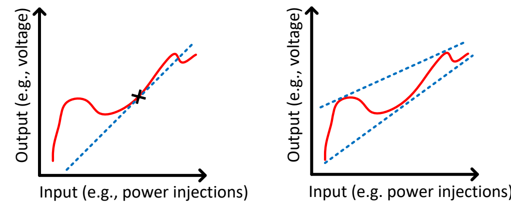


Fig. 1. A conceptual example of a traditional linear approximation (left) compared to a CLA (right). The solid line represents the nonlinear function of interest. The dotted line in the left figure is a traditional first-order Taylor approximation around point $\times$ while the dotted top (bottom) line in the right figure is an over- (under-) estimating approximation.

a constrained-regression problem. This approach tailors the approximation to a particular operating range and system of interest. Moreover, considering more complex components, such as tap-changing transformers and smart inverters, is straightforward since their behavior can be integrated into the sampled power flow solutions [7]. CLAs thus yield simplified optimization problems that are suitable for commercial optimization solvers. Finally, when applied in an optimization context, CLAs have a key advantage: one may ensure satisfaction of nonlinear constraints *while only enforcing linear inequalities* (assuming that the CLAs are indeed conservative). In this section, we next revisit the CLA approach.

We denote vectors and matrices in bold. Consider some quantity of interest (e.g., the voltage magnitude at a particular bus or current flow on a certain line) that we generically denote as β. An *overestimating* CLA is given by the linear expression:

$$a_0 + \boldsymbol{a}_1^T \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{bmatrix}, \quad (2)$$

constructed such that the following relationship is satisfied for power injections $\boldsymbol{P}$ and $\boldsymbol{Q}$ within a specified range:

$$\beta \le a_0 + \boldsymbol{a}_1^T \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{bmatrix}, \quad (3)$$

where the superscript T denotes the transpose. Assuming that (3) is indeed satisfied, one may ensure satisfying the constraint $\beta \le \beta^{\max}$, where β is the output of a nonlinear function such as the implicit system of nonlinear AC power flow equations (1) and $\beta^{\max}$ is a specified upper bound, by instead enforcing the linear constraint $a_0 + \boldsymbol{a}_1^T \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{bmatrix} \le \beta^{\max}$.

To determine the coefficients of the affine function of power injections in (2), we solve the following regression problem:

$$\min_{a_0,\ \boldsymbol{a}_1} \quad \frac{1}{M}\sum_{m=1}^{M} \mathcal{L}\left(\beta_m - \left(a_0 + \boldsymbol{a}_1^T \begin{bmatrix} \boldsymbol{P}_m \\ \boldsymbol{Q}_m \end{bmatrix}\right)\right) \quad \text{(4a)}$$

$$\text{s.t.} \quad \beta_m - \left(a_0 + \boldsymbol{a}_1^T \begin{bmatrix} \boldsymbol{P}_m \\ \boldsymbol{Q}_m \end{bmatrix}\right) \le 0, \quad m = 1, \ldots, M. \quad \text{(4b)}$$

The subscript m denotes the m^{th} sample, M is the number of samples, and $\mathcal{L}(\,\cdot\,)$ represents a loss function, e.g., absolute value for the ℓ_1 loss and the square for the squared-ℓ_2 loss.

In this paper, our quantities of interest (β) are the magnitudes of voltages (V) and current flows (I).

Underestimating CLAs are constructed in the same manner as (4) except that the direction of the inequality in (4b) is flipped. One could also compute a linear approximation (LA) that is not conservative (minimizes approximation errors without consistently under- or over-estimating) by solving (4) without enforcing (4b).

III. Adaptive Sampling Method for Power Flow Approximations

The accuracy and conservativeness of data-driven power flow approximations like CLA depend on the set of samples, with more samples leading to higher accuracy and increased confidence in conservativeness. However, using many samples can be computationally challenging, which is problematic when power flow approximations need to be recomputed frequently during rapidly changing operating conditions.

Thus, approximation methods such as CLA can significantly benefit from the concept of importance sampling. The samples that provide the most valuable information for improving the quality of a linearization are often associated with regions of the power flow manifold that exhibit the most curvature. We propose an importance sampling method based on the second-order sensitivities of the power flow equations. This sampling method can achieve high accuracy with fewer samples compared to naive random sampling approaches.

For brevity, we rewrite the power flow equations in (1) as:

$$\boldsymbol{x} = g(\boldsymbol{y}), \tag{5}$$

where g is a vector-valued function and

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{V} \end{bmatrix}. \tag{6}$$

The second-order sensitivity matrix for a specific quantity of interest y_k, denoted as $\boldsymbol{\Lambda}_{y_k}$, takes the following form:

$$\boldsymbol{\Lambda}_{y_k} = \begin{bmatrix} \dfrac{\partial^2 y_k}{\partial x_1 \partial x_1} & \dfrac{\partial^2 y_k}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 y_k}{\partial x_1 \partial x_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 y_k}{\partial x_{2N} \partial x_1} & \dfrac{\partial^2 y_k}{\partial x_{2N} \partial x_2} & \cdots & \dfrac{\partial^2 y_k}{\partial x_{2N} \partial x_{2N}} \end{bmatrix}. \tag{7}$$

The appendix derives an explicit form for the second-order sensitivity matrix $\boldsymbol{\Lambda}_{y_k}$ (7).

After obtaining the second-order sensitivity matrix, we analyze the power flow manifold's curvature by computing the singular value decomposition (SVD) of $\boldsymbol{\Lambda}_{y_k}$. A singular vector associated with a largest singular value indicates the direction of highest curvature. As we will demonstrate empirically in Section V, the sensitivity matrices $\boldsymbol{\Lambda}_{y_k}$ typically have only a few significant singular values (i.e., these matrices are approximately low rank). Thus, the subspace spanned by only a few singular vectors gives a good characterization for the directions of highest curvature and hence suggests promising directions for prioritized sampling. By selecting a larger fraction of samples in the span of the dominant singular vectors, we seek to better inform the linear approximation to fit the most nonlinear components of the function. The second-order sensitivity analyses can also identify the convexity or concavity of the power flow equations at a specified operating point that can be used to further improve the sample selection.

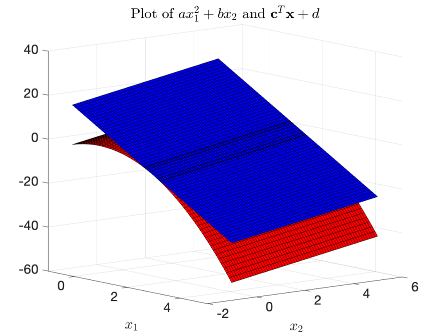


Fig. 2. An example of computing an overestimating linear function $\mathbf{c}^T\mathbf{x} + d$ (blue plane) of a quadratic function $ax_1^2 + bx_2$ (red manifold) where $\mathbf{x}^T = [x_1,\ x_2], a = -2, b = 2, \mathbf{c}^T = [-8,\ 2]$, and $d = 8$.

The choice to draw samples along the span of singular vectors associated with the first few largest singular values of the second-order sensitivity matrix is conceptually illustrated via the simple three-dimensional plot in Fig. 2. This figure shows an example quadratic function that is nonlinear in the x_1 direction and linear in the x_2 direction along with an overestimating CLA. The CLA here minimizes the volume between the linear approximation and the nonlinear function while ensuring that the overestimating linear function consistently lies above the quadratic function. In a sample-based approach to computing the CLA, samples in the linear x_2 direction would provide little benefit while samples in the nonlinear x_1 direction would be informative.

We present detailed results on power system test cases in Section V, empirically showing that the second-order sensitivity matrices associated with voltage magnitudes are approximately low-rank and have all non-positive eigenvalues. The results demonstrate that the overestimating CLA benefits most from samples near the middle of the operating range to ensure conservativeness, with a small number of more extreme samples capturing the function's curvature to maintain accuracy. Conversely, in the case of an underestimating linear function, we need more samples in the directions of maximum curvature to maintain conservativeness and fewer samples in the middle of the operating range to improve accuracy.

IV. Rational Approximations of the Power Flow Equations

The CLAs in Section II-B were developed as linear functions of power injections. The primary goal of these linear functions is to maintain linearity in an optimization problem's constraints. However, there are more general classes of functions that are nonlinear but can be represented with linear constraints, specifically, rational functions with a linear numerator and a strictly positive linear denominator. With more degrees of freedom than linear functions, rational approximations can more accurately capture power flow nonlinearities. This section introduces the Padé approximant and conservative rational approximations (CRAs) of the power flow equations.

A. Padé approximant

The Padé approximant is a rational approximation that matches the first terms of a function's Taylor expansion [13]. The Padé approximant of a univariate function $f(x)$ is:

$$R_{x_0}(x) = \frac{a_0 + a_1(x-x_0) + \cdots + a_m(x-x_0)^m}{1 + b_1(x-x_0) + \cdots + b_n(x-x_0)^n}, \tag{8}$$

where $m \geq 0$, $n \geq 1$, and both m and n are integers. Equation (8) is called the $[m/n]$ Padé approximant, which matches the $(m+n)^{\text{th}}$-degree Taylor series. To compute coefficients $a_0, \ldots, a_m$ and $b_1, \ldots, b_n$, we set $R_{x_0}(x)$ equal to the Taylor series and multiply both sides by the denominator (see [12]). By matching the coefficients of terms with degrees less than or equal to $m+n$, we can determine all the coefficients.

This paper focuses is on the [1/1] Padé approximant, which represents a ratio of linear functions. Despite being a nonlinear function, the Padé approximant can be integrated into an optimization problem as a linear constraint. To be applicable to a power systems setting, we employ a multivariate generalization of the Padé approximant. Consider the constraint:

$$\frac{a_0 + \boldsymbol{a_1}^T \boldsymbol{x}}{1 + \boldsymbol{b_1}^T \boldsymbol{x}} - U \leq 0, \tag{9}$$

where $\boldsymbol{x}$ is a vector of decision variables and U is an upper bound on the constrained quantity. In the multivariate generalization of the Padé approximant shown in (9), $\boldsymbol{a_1}$ and $\boldsymbol{b_1}$ are vectors. For instance, we may seek to compute parameters $a_0, \boldsymbol{a_1}$, and $\boldsymbol{b_1}$ such that $(a_0 + \boldsymbol{a_1}^T\boldsymbol{x})/(1 + \boldsymbol{b_1}^T\boldsymbol{x})$ approximates a voltage magnitude V_i based on the power injections $\boldsymbol{x} = [\boldsymbol{P}^T \ \boldsymbol{Q}^T]^T$, with $U = V_i^{\max}$ denoting the upper bound on the voltage at bus i. As long as $1 + \boldsymbol{b_1}^T\boldsymbol{x} > 0$, the constraint (9) can be reformulated into the equivalent linear constraint:

$$(a_0 - U) + (\boldsymbol{a_1} - U\boldsymbol{b_1})^T \boldsymbol{x} \leq 0. \tag{10}$$

As we will show next, the additional degrees of freedom compared to a linear approximation enable the [1/1] multivariate Padé approximant to better capture nonlinear behavior while still yielding a linear constraint.

The $[1/1]$ multivariate generalization of the Padé approximant that we consider seeks to match the first- and second-order curvature of the second-order Taylor series approximation of the power flow equations. Consider the following equation around $\boldsymbol{x} = \boldsymbol{x_0}$:

$$\frac{a_0 + \boldsymbol{a_1}^T(\boldsymbol{x}-\boldsymbol{x_0})}{1 + \boldsymbol{b_1}^T(\boldsymbol{x}-\boldsymbol{x_0})} = f(\boldsymbol{x_0}) + \nabla_{\boldsymbol{x}} f(\boldsymbol{x_0})^T(\boldsymbol{x}-\boldsymbol{x_0}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x_0})^T \boldsymbol{\Lambda}_f(\boldsymbol{x_0})(\boldsymbol{x}-\boldsymbol{x_0}), \tag{11}$$

where $\nabla_{\boldsymbol{x}} f(\boldsymbol{x_0})$ is the gradient of f at $\boldsymbol{x} = \boldsymbol{x_0}$ and $\boldsymbol{\Lambda}_f(\boldsymbol{x_0})$ is the second-order sensitivity matrix of f at $\boldsymbol{x} = \boldsymbol{x_0}$. Multiplying by the denominator in (11) and comparing coefficients up to the quadratic term, we obtain:

$$a_0 = f(\boldsymbol{x_0}), \tag{12a}$$
$$\boldsymbol{a_1} = \nabla_{\boldsymbol{x}} f(\boldsymbol{x_0}) + f(\boldsymbol{x_0})\boldsymbol{b_1}, \tag{12b}$$
$$\boldsymbol{A} = \boldsymbol{b_1} \nabla_{\boldsymbol{x}} f(\boldsymbol{x_0})^T + \frac{1}{2}\boldsymbol{\Lambda}_f(\boldsymbol{x_0}), \tag{12c}$$

for some skew-symmetric matrix $\boldsymbol{A}$[1]. Since the matrix given by the outer product $\boldsymbol{b_1}\nabla_{\boldsymbol{x}} f(\boldsymbol{x_0})^T$ has rank 1, the matrix equation (12c) is unsolvable, in general. Accordingly, there are many possible generalizations of univariate Padé approximations to multivariate settings [14]. We choose to approximately solve (12c) in the Frobenius-norm sense. We start by adding the expression in (12c) to its transpose to obtain:

$$\underline{\boldsymbol{0}} = \boldsymbol{b_1}\nabla_{\boldsymbol{x}} f(\boldsymbol{x_0})^T + \nabla_{\boldsymbol{x}} f(\boldsymbol{x_0})\boldsymbol{b_1}^T + \boldsymbol{\Lambda}_f(\boldsymbol{x_0}), \tag{13}$$

where $\underline{\boldsymbol{0}}$ is a zeros matrix. The left-hand side equals $\underline{\boldsymbol{0}}$ since $\boldsymbol{A} + \boldsymbol{A}^T = \boldsymbol{0}$ for any skew-symmetric matrix $\boldsymbol{A}$. We then minimize the Frobenius norm (denoted as $\|\cdot\|_{\mathcal{F}}$) to select $\boldsymbol{b_1}$:

$$\boldsymbol{b_1} = \arg\min_{\boldsymbol{b_1}} \|\boldsymbol{b_1}\nabla_{\boldsymbol{x}} f(\boldsymbol{x_0})^T + \nabla_{\boldsymbol{x}} f(\boldsymbol{x_0})\boldsymbol{b_1}^T + \boldsymbol{\Lambda}_f(\boldsymbol{x_0})\|_{\mathcal{F}}. \tag{14}$$

We note that the second-order sensitivity $\boldsymbol{\Lambda}_f(\boldsymbol{x_0})$ discussed in Section III is again needed here to compute the $[1/1]$ multivariate generalization of the Padé approximant. By incorporating information from the second-order sensitivity, this approximant can better capture power flow nonlinearities while still yielding linear constraints in optimization problems.

B. Conservative rational approximations

The first-order Taylor series is the best-fitting linear function around a specified point. The CLAs in [6] extend this concept to fit linear functions in an operating region of interest. Analogously, the [1/1] Padé approximant is the best-fitting ratio of linear functions around a specified point. Motivated by the [1/1] multivariate generalization of the Padé approximant, we next extend this concept to construct conservative rational approximations (CRAs) with linear numerator and denominator, called [1/1] CRAs, within an operational region of interest.

A [1/1] CRA is defined as a rational function whose numerator and denominator are linear functions and, similar to CLAs, are conservative (under- or over-estimate the nonlinear power flow equations) in an operational region of interest. In an attempt to achieve this, we enforce conservativeness of the CRAs over a set of sampled power injections in this region.

Analogous to (4) for an overestimating CLA, the regression problem for computing an overestimating CRA is:

$$\min_{a_0, \boldsymbol{a_1}, \boldsymbol{b_1}} \quad \frac{1}{M}\sum_{m=1}^{M} \mathcal{L}\left(\beta_m - \frac{a_0 + \boldsymbol{a_1}^T\boldsymbol{x_m}}{1 + \boldsymbol{b_1}^T\boldsymbol{x_m}}\right) \tag{15a}$$
$$\text{s.t.} \quad a_0 + \boldsymbol{a_1}^T\boldsymbol{x_m} - \beta_m(1 + \boldsymbol{b_1}^T\boldsymbol{x_m}) \geq 0, \tag{15b}$$
$$1 + \boldsymbol{b_1}^T\boldsymbol{x_m} \geq \epsilon, \quad m = 1, \ldots, M, \tag{15c}$$

[1]If $\boldsymbol{A}$ is skew symmetric (i.e., $\boldsymbol{A} = -\boldsymbol{A}^T$), then $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = 0$ for any $\boldsymbol{x}$.

where M again represents the number of samples. The objective (15a) minimizes the mismatch between the rational approximation and the value of the quantity of interest, again generically denoted as β, obtained from solving the power flow equations at each sampled set of power injections. Constraint (15b) ensures an overestimating property. Constraint (15c) maintains the correct inequality sign in (15b), assuming that ϵ is a specified positive number. The analogous formulation for an underestimating CRA is given by (15) with a flipped inequality sign in (15b). One could also compute a rational approximation (RA) that was not conservative by solving (15) without enforcing (15b). Similar to those in Section IV-A, the rational approximations from solving (15) can better capture the curvature the power flow manifold compared to linear approximations, here considering a specified operational region instead of at a particular point.

Ideally, we would like to directly solve (15). However, due to the objective (15a), this is a nonlinear program that can be challenging to solve. Alternatively, we consider the following regression problem for computing an overestimating CRA with an ℓ_1 loss function:

$$\min_{a_0,\boldsymbol{a_1},\boldsymbol{b_1}} \frac{1}{M}\sum_{m=1}^{M} w_m \left|\left(a_0 + \boldsymbol{a_1}^T\boldsymbol{x_m} - \beta_m(1+\boldsymbol{b_1}^T\boldsymbol{x_m})\right)\right| \tag{16a}$$

$$\text{s.t.} \quad a_0 + \boldsymbol{a_1}^T\boldsymbol{x_m} - \beta_m(1+\boldsymbol{b_1}^T\boldsymbol{x_m}) \geq 0, \tag{16b}$$

$$1 + \boldsymbol{b_1}^T\boldsymbol{x_m} \geq \epsilon, \quad m = 1,\ldots,M, \tag{16c}$$

where w_m is a specified parameter vector that weights each term in the objective. If w_m were equal to $(1+\boldsymbol{b_1}^T\boldsymbol{x_m})^{-1}$, then (16) would be equivalent to (15) since multiplying by the weights w_m would then compensate for multiplying by the denominators $1+\boldsymbol{b_1}^T\boldsymbol{x_m}$ in (16a) relative to (15a). However, this choice of w_m requires prior knowledge of $\boldsymbol{b_1}$, which is what we are trying to compute.

We therefore employ an iterative algorithm that updates the weights w_m at each iteration based on the values of $\boldsymbol{b_1}$ from the previous iteration. Let superscript $(\,\cdot\,)^k$ denote quantities at the k^{th} iteration. This algorithm first initializes $\boldsymbol{w}$ as $\boldsymbol{w}^0$ (using, for instance, the $\boldsymbol{b_1}$ parameters computed from the [1/1] multivariate generalization of the Padé approximant described in Section IV-A). For each iteration k, the algorithm then solves (16) to compute $\boldsymbol{b_1}^k$ and updates $w_m^{k+1} = (1+(\boldsymbol{b_1}^k)^T\boldsymbol{x_m})^{-1}$, repeating until $\|\boldsymbol{w}^k - \boldsymbol{w}^{k-1}\|_1$ reaches a specified tolerance.

V. Numerical results

This section validates the proposed importance sampling and rational function approximation methods for the power flow equations. We begin by comparing the [1/1] multivariate generalization of the Padé approximant to the first- and second-order Taylor approximations for voltage magnitudes. Following that, we evaluate the performance of linear approximations (LAs), rational approximations (RAs), as well as their conservative versions, CLAs and CRAs. Additionally, we utilize the second-order sensitivities in the importance sampling. Lastly, we provide a conceptual demonstration through a simplified optimal power flow problem.

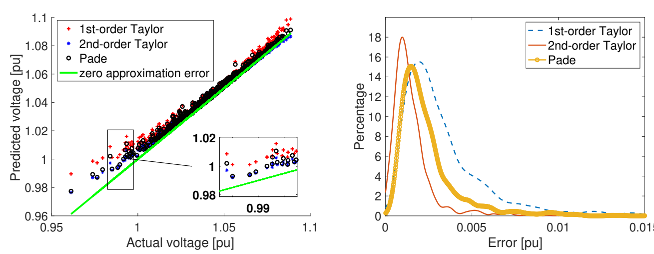
The test cases used in the simulations include the IEEE 24-bus system, the 30-bus system *case30*, the 33-bus system *case33bw*, the 85-bus system *case85*, the 141-bus system *case141*, and the 2383-bus system *case2383wp*, all of which are available in MATPOWER [15]. For the voltage and current flow approximations, we draw 1000 samples by varying the power injections from 70% to 130% of their nominal values. The voltage and current flow values are reported in per unit (pu). We use ℓ_1 for a loss function $\mathcal{L}(\,\cdot\,)$.

A. Multivariate generalization of the Padé approximant

By incorporating curvature information from the second-order Taylor approximation, the [1/1] multivariate generalization of the Padé approximant can surpass the accuracy of the first-order Taylor approximation. We empirically demonstrate these accuracy advantages by comparison to both the first- and second-order Taylor approximation across a range of operation around a nominal point. For the sake of illustration, we choose, in each test case, the voltage magnitude at a bus for which the first-order Taylor approximation has a large error.

Fig. 3 presents an illustrative example of voltage approximation at bus 1 in the IEEE 24-bus system. The results depicted in Fig. 3(a) indicate that the [1/1] multivariate generalization of the Padé approximant has approximation accuracy between the first- and second-order Taylor approximations. To provide a clearer view of the errors associated with each approximation, we present a histogram plot in Fig. 3(b).

To quantify these results, Table I presents the average errors per sample in voltage magnitudes resulting from three different approximation methods: the first-order Taylor approximation, the [1/1] multivariate generalization of the Padé approximant,



(a) Predicted vs actual voltage (b) Histogram of errors

Fig. 3. The left plot shows a comparison between the first- (red crosses) and second-order Taylor approximations (blue asterisks) and the Padé approximant (black circles) for voltage magnitudes at bus 1 in the IEEE 24-bus system. The green line at 45° represents zero approximation error. The smoothed histogram in the right plot displays the errors from the first- (blue dashed line) and second-order Taylor approximations (red solid line), along with the [1/1] multivariate generalization of the Padé approximant (thick yellow line).

TABLE I
VOLTAGE MAGNITUDE ERRORS AT A SPECIFIC BUS

Cases	Bus	Average errors per sample [10^{-4} pu]		
		1^{st}-order Taylor	[1/1] multivariate Padé (*)	2^{nd}-order Taylor (*)
IEEE 24-bus	22	47.2	27.9 (40.9%)	9.27 (80.4%)
case30	30	13.8	9.78 (29.1%)	0.76 (94.5%)
case33bw	33	0.32	0.16 (50%)	0.0036 (98.9%)
case141	80	0.065	0.030 (53.8%)	0.00038 (99.4%)

*Percentage reduction in errors compared the first-order Taylor approximation.

TABLE II
APPROXIMATION ERRORS FOR VOLTAGE MAGNITUDES AT A SPECIFIC BUS

Cases	Bus	Errors per sample [10^{-4} pu]							
		LA		RA		CLA		CRA	
		Mean	Max	Mean (*)	Max	Mean	Max	Mean (†)	Max
case30	25	4.40	27.6	3.03 (31.14%)	25.3	8.30	35.6	5.46 (34.22%)	28.5
case33bw	33	0.165	1.07	0.115 (30.30%)	1.07	0.245	1.38	0.175 (28.57%)	1.42
case85	54	0.0869	0.864	0.0816 (6.10%)	0.745	0.1265	1.09	0.1173 (7.27%)	0.89
case141	80	0.0318	0.366	0.0225 (29.24%)	0.255	0.0439	0.397	0.0326 (25.74%)	0.289

*The percentage reduction in errors compared to the mean errors from linear approximation (LA).
†The percentage reduction in errors compared to the mean errors from conservative linear approximation (CLA).

TABLE III
APPROXIMATION ERRORS FOR THE CURRENT FLOWS ON A SPECIFIC LINE

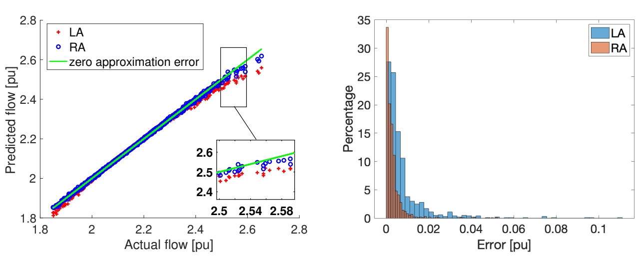
Cases	Line	Errors per sample [10^{-2} pu]							
		LA		RA		CLA		CRA	
		Mean	Max	Mean (*)	Max	Mean	Max	Mean (†)	Max
case30	1-2	1.25	8.82	1.20 (4%)	8.56	4.28	14.8	3.63 (15.19%)	13.4
case33bw	29-30	0.0261	0.215	0.0249 (4.60%)	0.169	0.0932	0.325	0.0568 (39.06%)	0.195
case85	3-17	0.0923	0.741	0.0833 (9.75%)	0.651	0.281	0.91	0.192 (31.67%)	0.817
case141	92-93	0.0303	0.376	0.0173 (42.90%)	0.201	0.0698	0.309	0.0257 (63.18%)	0.342

*The percentage reduction in errors compared to the mean errors from linear approximation (LA).
†The percentage reduction in errors compared to the mean errors from conservative linear approximation (CLA).

and the second-order Taylor approximation. The findings reveal substantial error reductions, ranging from 29.1% to 53.8%, when employing the [1/1] multivariate Padé approximant in comparison to the first-order Taylor approximation. The second-order Taylor approximation yields larger error reductions, ranging from 80.4% to 99.4%, compared to the first-order Taylor approximation, but unlike the [1/1] multivariate Padé approximant, it does not maintain linearity (or even convexity) of constraints when used for optimization.

B. Linear and rational approximations

Next, we compare the performance of four approximation methods: 1) linear approximation (LA), 2) rational approximation (RA), 3) conservative linear approximation (CLA), and 4) conservative rational approximation (CRA). RA and CRA are rational functions with linear numerators and denominators. The LAs and RAs are similar to the CLAs and CRAs, respectively, except that they do not enforce conservativeness.



(a) Predicted vs actual flow (b) Histogram of errors

Fig. 4. The left plot compares linear approximation (LA) in red crosses and rational approximation (RA) in blue circles for current flow on a branch connecting buses 15 and bus 21 in IEEE 24-bus system. The green line indicates the zero approximation error. The right plot shows the error histogram from the linear approximation (LA) and the rational approximation (RA).

In Fig. 4, we provide an illustrative example comparing the performance of LA and RA for current flow from bus 15 to bus 21 in the IEEE 24-bus system. Fig. 4(a) clearly demonstrates that RA predicts the current flow much more accurately in comparison to LA. For a more detailed inspection of the errors, Fig. 4(b) presents a histogram plot. The RA yields significantly lower errors as evident from the left-shit in the histogram, with very few samples exhibiting errors in excess of 0.01 pu.

Table III provides a comparison of RA and CRA to LA and CLA, respectively, for approximating voltage magnitudes. The results show notable error reductions for RAs, ranging from 6.10% to 31.14% improvements, when compared to LAs. Similarly, CRAs demonstrate substantial error reductions over CLAs, ranging from 7.27% to 34.22% improvements.

Table III presents a similar comparison for current flow approximations. RAs provide error reductions ranging from 4% to 42.90% compared to LAs. CRAs achieve even more significant error reductions, ranging from 15.19% to 63.18% improvements compared to CLAs. These results demonstrate the advantages inherent to rational approximations.

C. Second-order sensitivity: Computation time

This section presents results on the computation time required for calculating the second-order sensitivity information. The computational process is divided into three distinct steps: $a)$ power flow solution, $b)$ second-order sensitivity matrix calculation, and $c)$ singular value decomposition (SVD).

The computation times presented in Table IV are specific to the analysis of a single bus for each test case. However, note that Step $a)$ and most parts of Step $b)$ (e.g., computing Jacobian and Hessian matrices) only need to be computed once for all output quantities, which can result in time savings and increased efficiency during the overall process if multiple

TABLE IV
COMPUTATION TIME FOR THE SECOND-ORDER SENSITIVITY MATRIX

	Computation time [s]		
	Step a)	Step b)	Step c)*
IEEE 24-bus	0.0054	0.021	0.0002
case30	0.0040	0.022	0.002
case33bw	0.0042	0.026	0.001
case141	0.0029	0.1185	0.0036
case2383wp	0.048	13.07	3.58

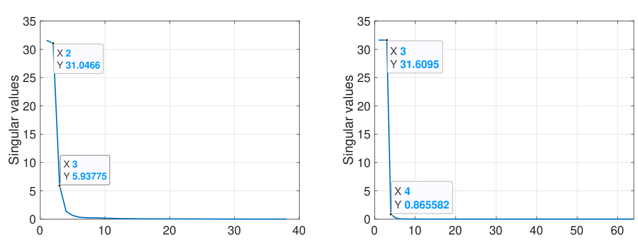
*The computation time for finding all singular values.

output quantities are of interest (e.g., voltage magnitudes at several buses). Additionally, each case shows only a few significant singular values (at least 10% of the maximum singular value), e.g., two for IEEE 24-bus at bus 7 and three for *case33* at bus 18. Consequently, we can speed up the computation in Step c) by evaluating only a limited number of the most significant singular values.

D. Second-order sensitivity: Span of singular vectors

The total time for computing sample-based power flow approximations depends on the number of sampled sets of power injections. Using second-order sensitivities, the importance sampling approach described in Section III aims to reduce computing time by prioritizing power injection samples in regions of high curvature. If there were substantial variability in the second-order sensitivities across typical operational ranges, this importance sampling approach would likely perform poorly as the second-order sensitivity information at a subset of samples would not help characterize the entire operational range. We next provide empirical evidence showing that the space spanned by the dominant singular values change very little over wide operational ranges.

We consider 1000 randomly sampled power injections that range from 70% to 130% of their nominal values and compute the first k significant singular values ($k = 2$ for IEEE 24-bus system and $k = 3$ for *case33bw*) for the second-order sensitivity matrix for each of these points. We stack the corresponding singular vectors to form a matrix, denoted as $\mathbf{M}$, consisting of $1000k$ columns and compute the singular values of this matrix. The number of significant singular values of $\mathbf{M}$, i.e., its approximate rank, indicates the extent to which the space spanned by the dominant singular vectors vary across the power flow samples. If M has approximate rank k, this indicates that the singular vectors have no variation, whereas if $\mathbf{M}$ has full rank then the singular vectors vary significantly over the operating range.



(a) IEEE 24-bus system at bus 7 (b) *case33bw* at bus 18

Fig. 5. The singular values of $\mathbf{M}$ for 1000 spans of the singular vectors.

Fig. 5 plots the sorted singular values of M for IEEE 24-bus system and *case33bw*. The plot reveals that for the IEEE 24-bus system ($k = 2$), the stacked matrix $\mathbf{M}$ has two large singular values, followed by two non-negligible but considerably smaller ones. The singular values beyond the eighth position for both test cases are negligible. The same behaviour is observed for *case33bw* with $k = 3$. From the results on these and other test cases, we conclude that the second-order sensitivities at a single nominal point is sufficient to characterize the second-order behaviour of the power flow manifold for large ranges of power injections.

E. Second-order sensitivity: Concavity

To further analyze the curvature of the power flow manifold, we examine the eigenvalues of the second-order sensitivity matrix. Since the sensitivity matrix is symmetric, all eigenvalues are real-valued. If all eigenvalues are positive, the second-order sensitivity matrix is positive definite, indicating local convexity of the second-order approximation. Conversely, all eigenvalues being negative would indicate local concavity of the second-order approximation. Otherwise, the second-order approximation is indefinite.

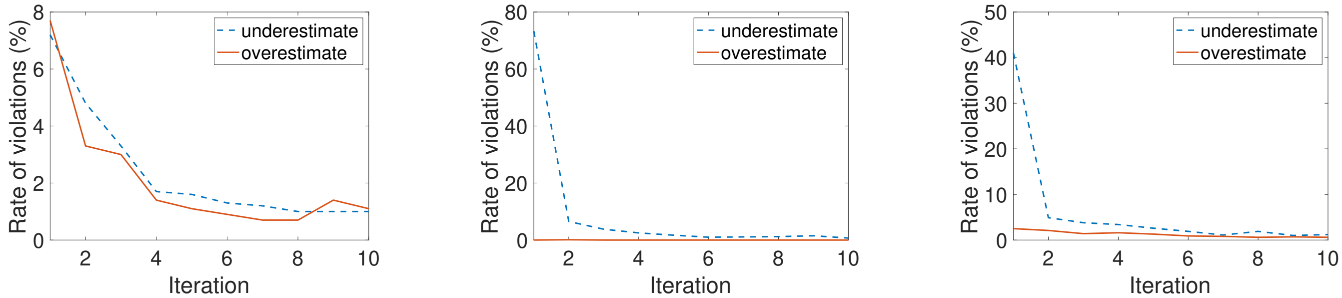
Table V presents the maximum and minimum eigenvalues of the second-order sensitivity matrices. In each case, we select the bus characterized by the most extreme curvature, which corresponds to the smallest eigenvalue. As suggested by the results in Section V-D, the second-order sensitivity matrices at a bus exhibit similarity across a range of operational conditions. Therefore, the eigenvalues for each test case are computed at a nominal value. These results empirically demonstrate that all eigenvalues of the second-order sensitivity matrix are non-positive, suggesting local concavity of the second-order approximation. Based on this concavity, we bias the importance sampling procedure to draw more samples at extreme points (i.e., away from the nominal point) for underestimating approximations and around the nominal points for overestimating approximations. The implications of these results will be further discussed in the next section.

F. Adaptive sampling

Our prior work in [6] introduced a sample selection method aimed at enhancing the conservativeness of the conservative linear approximations with limited impacts on the computation time for solving the regression problem (4). The sample selection approach in [6] is based on new samples drawn uniformly at random within the predefined range of power injections, which may not adequately capture the nonlinearity of the

TABLE V
MAXIMUM AND MINIMUM EIGENVALUES OF THE SECOND-ORDER SENSITIVITY MATRICES FOR VOLTAGE MAGNITUDES

Cases	Bus	Eigenvalues	
		Max	Min
IEEE 24-bus	7	-0.002	-0.72
case30	30	-0.009	-3.02
case33bw	18	0	-10.45
case85	50	0	-0.34
case141	52	0	-0.65

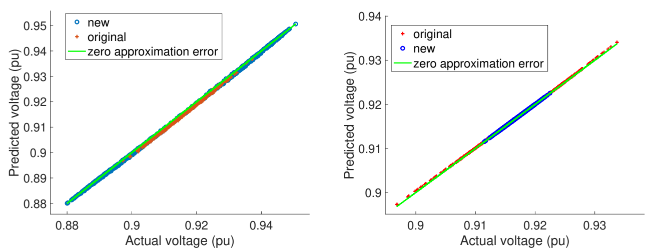


(a) All uniform random samples

(b) All samples in the dominant singular vectors' span

(c) Half uniform random samples and half samples in the dominant singular vectors' span

Fig. 6. Plots depicting the rate of violations versus iterations obtained by employing three different sampling strategies: (a) drawing all additional samples randomly, (b) selecting all additional samples in alignment with the subspace spanned by the first three singular vectors, and (c) one half of the additional samples randomly and the other half in accordance with the subspace spanned by the first three singular vectors. These simulations are the approximations of the voltage magnitude at bus 33 for *case33bw*. The dashed blue and solid red lines denote the violation rates for underestimates and overestimates, respectively.



(a) Underestimate

(b) Overestimate

Fig. 7. Plots of the results from (a) underestimating and (b) overestimating linear approximations of the voltage magnitude at bus 33 for *case33bw*. The red and blue points represent the original and new samples, respectively. The new samples are those that violate the original linear approximations. The green line at 45° represents zero approximation error.

power flow equations or effectively address high curvature regions where more violations can occur.

To address these limitations, our proposed importance sampling approach selects new samples along the span of the dominant singular vectors of the second-order sensitivity matrix. As in [6], we iterate between selecting new samples of power injections and updating the linearization based on the samples that are not conservative (are above the overestimating function or below the underestimating function). However, as discussed in Section III, we also select samples along the span of the dominant singular vectors to prioritize directions associated with the largest curvature of the power flow equations.

We also examine the specific neighborhoods from which we draw samples. This later approach takes into account the insights gained from the convexity/concavity analysis in Section V-E. By considering the influence of these directions, we aim to exploit the power flow equations' nonlinear behavior as well as their local convexity or local concavity.

To assess this importance sampling approach, Fig. 6 shows the rates of violations where the newly drawn samples' predicted values are below the approximation's underestimates or exceed the overestimates. When new samples are drawn uniformly at random (Fig. 6(a)), the first iteration's violation rates for both underestimates and overestimates range from 7% to 8%, meaning that relatively little new information is provided by additional samples with respect to the conservativeness of the approximation. However, when new samples are drawn in the direction of the subspace spanned by the three singular vectors corresponding to the three largest singular values (Fig. 6(b)), there is a 72% violation rate for underestimates, but the violation rate for overestimates remains close to 0% in the first iteration. Violation rates significantly reduce after the first iteration and continue to decrease over the next iterations. These results suggest that, as expected, samples in the span of the dominant singular vectors effectively identify violations for underestimating approximations but have little value for the conservativeness of overestimating approximations.

To balance the quality of the over- and under-estimating approximations, we assess performance when one half of the additional samples are drawn uniformly at random and the the other half is selected in the direction of the dominant singular vectors. As shown in Fig. 6(c), the violation rate is 41% for underestimates and 3% for overestimates, which suggests that a mix of sampling approaches using the second-order sensitivity matrix can substantially improve the performance of the underestimating approximations without overly detrimental impacts to overestimating approximations.

Finally, to provide insights into the characteristics of the violated samples, Fig. 7 shows both the under- and over-estimating linear approximations of voltages after the importance sampling method (half uniform random samples and the rest in the dominant singular vectors' span). Consistent with the eigenvalues, the voltage magnitude, when considered as a function of power injections, has local concavity in its second-order curvature. As shown in Figure 7(a), which presents the underestimating linear approximation, the newly drawn samples that violate the original CLA are mainly positioned away from the middle of the voltage range. Conversely, the violated samples are concentrated in middle of the voltage range for the overestimating case (see Fig. 7(b)). These results emphasize the concavity characteristics discussed in Section V-E.

G. Application: Simplified optimal power flow

While the primary intended applications of our RA and CRA methods are bilevel problems [7], [16] and capacity

TABLE VI
RESULTS COMPARING SOLUTIONS FROM
AC-, DC-, LA-, CLA-, RA-, AND CRA-OPF

	Case		
	case6ww	*case9*	*case14*
AC-OPF	2986.04	1456.83	5368.30
DC-OPF	2995.15 (0.31%)	1502.82 (3.16%)	5368.52 (0.004%)
Violation	V (0.029 pu)	-	-
LA-OPF	2989.95 (0.13%)	1473.67 (1.16%)	5368.52 (0.004%)
Violation	-	-	V (0.004 pu)
CLA-OPF	2991.14 (0.17%)	1475.92 (1.31%)	5368.52 (0.004%)
Violation	-	-	-
RA-OPF	2988.93 (0.10%)	1471.25 (0.99%)	5368.51 (0.004%)
Violation	-	-	V (0.002 pu)
CRA-OPF	2990.25 (0.14%)	1473.91 (1.17%)	5368.51 (0.004%)
Violation	-	-	-

The values in parentheses $(\cdot)$ indicate the percentage difference in cost when compared to the cost obtained from AC-OPF.

expansion planning problems [17], demonstration on these problems is beyond the scope of this paper. Instead, we focus on presenting results in a simplified optimal power flow (OPF) setting, which serves as a conceptual demonstration and a basis for comparison between various linear approximations (DC power flow, LA, CLA, RA, and CRA). The simplified versions of the OPF problem enforce constraints on voltages at load buses (where P and Q are specified) and power generations within specified limits, without considering line flow limits.

For LA, CLA, RA, and CRA, we compute approximations for voltages at load buses, reactive power outputs at generator buses (where P and V are specified), and active power generation at a reference bus. These approximations are functions of active power injections and voltages at generator buses.

Table VI presents a comparison of results obtained from various approximations with AC-OPF for *case6ww*, *case9*, and *case14*. In each cell of this table, the first row displays the actual cost associated with the AC-PF feasible solution, determined by the set points prescribed in different OPF formulations. The second row indicates the presence of voltage violations and quantifies the maximum voltage violation.

The results show that both CLA-OPF and CRA-OPF do not result in any voltage violations due to their conservativeness property. In *case6ww*, DC-OPF leads to a maximum voltage violation of 0.029 pu. For *case14*, LA-OPF and RA-OPF produce a maximum voltage violation of 0.002 pu and 0.004 pu, respectively. Moreover, RA-OPF exhibits a cost advantage over LA-OPF and DC-OPF in *case6ww* and *case9*. Similarly, CRA-OPF demonstrates cost improvements in comparison to CLA-OPF and DC-OPF in these cases.

VI. CONCLUSION AND FUTURE WORK

This paper presents an importance sampling approach designed to enhance the accuracy and conservativeness (i.e., the tendency to over- or under-estimate a quantity of interest) of power flow approximations. Our method leverages second-order sensitivity information to provide a deeper understanding of the relationships between various quantities. Additionally, inspired by the Padé approximant and second-order sensitivities, we introduce the [1/1] multivariate Padé approximant, expressed as a rational function with a linear numerator and denominator, enhancing accuracy beyond the capabilities of linear functions. This rational approximation, when used as a constraint, preserves linearity in decision variables. Our numerical results reveal the benefits of second-order sensitivities for the importance sampling method and demonstrate improved accuracy compared to other linear approximations.

In our future work, we plan to extend our research by developing second-order sensitivities for different output functions, such as V^2. We will also explore the [2/2] multivariate Padé approximant to construct convex quadratic approximations. Additionally, we will focus on the applications of our proposed method to power system planning and resilience tasks, including capacity expansion planning problems.

APPENDIX: SECOND-ORDER SENSITIVITY ANALYSIS

The appendix formulates the second-order sensitivities of the voltage magnitudes with respect to the active and reactive power injections. Recall the length-$2N$ vectors x (of active and reactive power injections) and y (of voltage magnitudes and angles) from (6). We first compute the first-order sensitivities by evaluating the gradient of each row of the vectors in (5) with respect to $\boldsymbol{x}$. The derivative expression is:

$$\boldsymbol{I} = (\boldsymbol{J})(\nabla_{\boldsymbol{x}}\boldsymbol{y}), \tag{17a}$$

$$\nabla_{\boldsymbol{x}}\boldsymbol{y} := \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{2N}}{\partial x_1} & \frac{\partial y_{2N}}{\partial x_2} & \cdots & \frac{\partial y_{2N}}{\partial x_{2N}} \end{bmatrix}, \tag{17b}$$

where $\boldsymbol{I}$ represents an identity matrix with appropriate dimensions and $\boldsymbol{J}$ is the Jacobian matrix associated with the power flow equations, expressed as:

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_{2N}}{\partial y_1} & \frac{\partial x_{2N}}{\partial y_2} & \cdots & \frac{\partial x_{2N}}{\partial y_{2N}} \end{bmatrix}. \tag{18}$$

From (17), $\frac{\partial y_k}{\partial x_i}$ is simply $[\boldsymbol{J}^{-1}]_{ki}$, i.e., the (k,i) entry of the inverse Jacobian matrix. The second-order sensitivities are computed by evaluating the partial derivative of each element in the expression for the first-order sensitivities (17b). The partial derivatives of the first-order sensitivities with respect to each variable x_i are given by:

$$\underline{\boldsymbol{0}} = \boldsymbol{J}\left(\frac{\partial}{\partial x_i}\nabla_{\boldsymbol{x}}\boldsymbol{y}\right) + \left(\sum_k \left[\frac{\partial}{\partial y_k}\boldsymbol{J}\right]\frac{\partial y_k}{\partial x_i}\right)\nabla_{\boldsymbol{x}}\boldsymbol{y}, \tag{19}$$

where $\underline{\boldsymbol{0}}$ is a zero matrix. Note that the second-order sensitivity of y_k with respect to x_i and x_l is given by

$$\left[\frac{\partial}{\partial x_i}\nabla_{\boldsymbol{x}}\boldsymbol{y}\right]_{kl} = \frac{\partial^2 y_k}{\partial x_i \partial x_l}. \tag{20}$$

Thus, we can compute

$$\left[\frac{\partial}{\partial x_i}\nabla_{\boldsymbol{x}}\boldsymbol{y}\right] = -\boldsymbol{J}^{-1}\left(\sum_k \left[\frac{\partial}{\partial y_k}\boldsymbol{J}\right]\frac{\partial y_k}{\partial x_i}\right)\boldsymbol{J}^{-1}. \tag{21}$$

We next describe how to compute the terms in (21). Let diag$(\cdot)$ represent the diagonal matrix with the vector argument on the diagonal. We denote Re$(\cdot)$ and Im$(\cdot)$ as the real and imaginary part of the quantity $(\cdot)$, respectively. The asterisk $(*)$ represents the complex conjugate. We can now express the Jacobian matrix in terms of the derivative of complex power S (i.e., $P + jQ$) with respect to the voltage magnitudes and angles:

$$\boldsymbol{J} = \begin{bmatrix} \text{Re}\left\{\nabla_{\boldsymbol{\theta}}\boldsymbol{S}\right\} & \text{Re}\left\{\nabla_{\boldsymbol{V}}\boldsymbol{S}\right\} \\ \text{Im}\left\{\nabla_{\boldsymbol{\theta}}\boldsymbol{S}\right\} & \text{Im}\left\{\nabla_{\boldsymbol{V}}\boldsymbol{S}\right\} \end{bmatrix}. \tag{22}$$

Let $\boldsymbol{Y}$ be the network admittance matrix. To compute the partial derivatives of the Jacobian matrix $\frac{\partial}{\partial y_k}\boldsymbol{J}$ in (21), we start by differentiating the entries of the Jacobian, as shown in [18]:

$$\nabla_{\boldsymbol{\theta}}\boldsymbol{S} = j\ \text{diag}\left(\boldsymbol{V}\right)\left(\text{diag}\left(\boldsymbol{I}\right) - \boldsymbol{Y}\text{diag}\left(\boldsymbol{V}\right)\right)^*, \tag{23a}$$

$$\nabla_{\boldsymbol{V}}\boldsymbol{S} = \text{diag}\left(e^{j\boldsymbol{\theta}}\right)\text{diag}\left(\boldsymbol{I}\right)^* + \text{diag}(\boldsymbol{V})\left(\boldsymbol{Y}\text{diag}\left(e^{j\boldsymbol{\theta}}\right)\right)^*, \tag{23b}$$

where $\boldsymbol{I} = \boldsymbol{Y}\boldsymbol{V}$ represents the vector of bus current injections. Differentiating the entries of the Jacobian matrix, we obtain:

$$\frac{\partial}{\partial\theta_k}\left(\nabla_{\boldsymbol{\theta}}\boldsymbol{S}\right) = \text{diag}\underline{\boldsymbol{V}}_{\boldsymbol{m}}(-\text{diag}\left(\boldsymbol{I}\right)^* + \boldsymbol{Y}\text{diag}(\boldsymbol{V}))^* + \text{diag}(\boldsymbol{V})(\text{diag}(\boldsymbol{Y}\underline{\boldsymbol{V}}_{\boldsymbol{k}}) - \boldsymbol{Y}\text{diag}\underline{\boldsymbol{V}}_{\boldsymbol{k}})^*, \tag{24a}$$

$$\frac{\partial}{\partial\theta_k}\left(\nabla_{\boldsymbol{V}}\boldsymbol{S}\right) = j(-\text{diag}(e^{j\boldsymbol{\theta}})(\text{diag}(\boldsymbol{Y}\underline{\boldsymbol{V}}_{\boldsymbol{k}}))^* + \text{diag}(\underline{e}^{j\theta_k})\text{diag}(\boldsymbol{I})^* - \text{diag}(\boldsymbol{V})(\boldsymbol{Y}\text{diag}(\underline{e}^{j\theta_k}))^* + \text{diag}\underline{\boldsymbol{V}}_{\boldsymbol{k}}(\boldsymbol{Y}\text{diag}(e^{j\boldsymbol{\theta}}))^*), \tag{24b}$$

$$\frac{\partial}{\partial V_k}\left(\nabla_{\boldsymbol{\theta}}\boldsymbol{S}\right) = j(\text{diag}(\underline{e}^{j\theta_k})(\text{diag}\left(\boldsymbol{I}\right) - \boldsymbol{Y}\text{diag}(\boldsymbol{V}))^* + \text{diag}(\boldsymbol{V})(\text{diag}(\boldsymbol{Y}\underline{e}^{j\theta_k})^* - (\boldsymbol{Y}\text{diag}(\underline{e}^{j\theta_k}))^*), \tag{24c}$$

$$\frac{\partial}{\partial V_k}\left(\nabla_{\boldsymbol{V}}\boldsymbol{S}\right) = \text{diag}\left(e^{j\boldsymbol{\theta}}\right)\text{diag}\left(\boldsymbol{Y}\underline{e}^{j\theta_k}\right)^* + \text{diag}\left(\underline{e}^{j\theta_k}\right)\left(\boldsymbol{Y}\text{diag}\left(e^{j\boldsymbol{\theta}}\right)\right)^*, \tag{24d}$$

where $\underline{\boldsymbol{V}}_{\boldsymbol{k}}$ represents an all zero vector except for the k^{th} entry which has the value V_k. Similarly, $\underline{e}^{j\theta_k}$ represents an all zero vector except for the k^{th} entry which has the value $e^{j\theta_k}$. Let $\boldsymbol{\Gamma}_{(\cdot)}$ denote the derivative of the Jacobian matrix with respect to the subscript $(\cdot)$, i.e., $\frac{\partial}{\partial(\cdot)}\boldsymbol{J}$ in (21). The derivative of the Jacobian matrix respect to, for example, the voltage angle θ_k is given by:

$$\boldsymbol{\Gamma}_{\theta_k} = \begin{bmatrix} \text{Re}\left\{\dfrac{\partial}{\partial\theta_k}(\nabla_{\boldsymbol{\theta}}\boldsymbol{S})\right\} & \text{Re}\left\{\dfrac{\partial}{\partial\theta_k}(\nabla_{\boldsymbol{V}}\boldsymbol{S})\right\} \\ \text{Im}\left\{\dfrac{\partial}{\partial\theta_k}(\nabla_{\boldsymbol{\theta}}\boldsymbol{S})\right\} & \text{Im}\left\{\dfrac{\partial}{\partial\theta_k}(\nabla_{\boldsymbol{V}}\boldsymbol{S})\right\} \end{bmatrix}. \tag{25}$$

We can similarly compute the derivative of the Jacobian matrix with respect to a voltage magnitude.

Let $[\boldsymbol{J}^{-1}]_k$ represent the k^{th} row of the inverse Jacobian matrix $\boldsymbol{J}^{-1}$. We can then compute the second-order sensitivity matrix $\boldsymbol{\Lambda}_{y_k}$ as:

$$\boldsymbol{\Lambda}_{y_k} = \begin{bmatrix} -[\boldsymbol{J}^{-1}]_k\left(\sum_k \boldsymbol{\Gamma}_{y_k}\frac{\partial y_k}{\partial x_1}\right)\boldsymbol{J}^{-1} \\ -[\boldsymbol{J}^{-1}]_k\left(\sum_k \boldsymbol{\Gamma}_{y_k}\frac{\partial y_k}{\partial x_2}\right)\boldsymbol{J}^{-1} \\ \vdots \\ -[\boldsymbol{J}^{-1}]_k\left(\sum_k \boldsymbol{\Gamma}_{y_k}\frac{\partial y_k}{\partial x_{2N}}\right)\boldsymbol{J}^{-1} \end{bmatrix}. \tag{26}$$

References

[1] B. Stott, J. Jardim, and O. Alsaç, "DC power flow revisited," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1290–1300, Aug. 2009.

[2] M. E. Baran and F. F. Wu, "Optimal capacitor placement on radial distribution systems," *IEEE Trans. Power Del.*, vol. 4, no. 1, pp. 725–734, Jan. 1989.

[3] D. K. Molzahn and I. A. Hiskens, "A survey of relaxations and approximations of the power flow equations," *Foundations and Trends in Electric Energy Systems*, vol. 4, no. 1-2, pp. 1–221, Feb. 2019.

[4] S. Misra, D. K. Molzahn, and K. Dvijotham, "Optimal adaptive linearizations of the AC power flow equations," in *20th Power Syst. Comput. Conf. (PSCC)*, June 2018.

[5] T. Mühlpfordt, V. Hagenmeyer, D. K. Molzahn, and S. Misra, "Optimal adaptive power flow linearizations: Expected error minimization using polynomial chaos expansion," in *IEEE Milan PowerTech*, 2019.

[6] P. Buason, S. Misra, and D. K. Molzahn, "A sample-based approach for computing conservative linear power flow approximations," *Electric Power Systems Research*, vol. 212, p. 108579, 2022, presented at the *22nd Power Systems Computation Conference (PSCC 2022)*.

[7] ——, "A data-driven method for locating sensors and selecting alarm thresholds to identify violations of voltage limits in distribution systems," *arXiv:2210.09414*, 2022.

[8] J. Chen and L. A. Roald, "A data-driven linearization approach to analyze the three-phase unbalance in active distribution systems,," *Electric Power Systems Research*, vol. 211, p. 108573, 2022, presented at the *22nd Power Systems Computation Conference (PSCC 2022)*.

[9] M. Jia and G. Hug, "Overview of data-driven power flow linearization," in *IEEE Belgrade PowerTech*, 2023.

[10] M. Jia, G. Hug, N. Zhang, Z. Wang, and Y. Wang, "Tutorial on data-driven power flow linearization–Part I: Challenges and training algorithms," preprint available at https://doi.org/10.3929/ethz-b-000606654, 2023.

[11] ——, "Tutorial on data-driven power flow linearization–Part II: Supportive techniques and experiments," preprint available at https://doi.org/10.3929/ethz-b-000606656, 2023.

[12] A. Cuyt, *Abstract Padé Approximants in Operator Theory*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1984, pp. 1–58.

[13] G. A. Baker and J. Gammel, "The Padé approximant," *Journal of Mathematical Analysis and Applications*, vol. 2, no. 1, pp. 21–30, 1961.

[14] A. Cuyt, "How well can the concept of Padé approximant be generalized to the multivariate case?" *Journal of Computational and Applied Mathematics*, vol. 105, no. 1-2, pp. 25–50, 1999.

[15] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[16] S. Wogrin, S. Pineda, and D. A. Tejada-Arango, *Applications of Bilevel Optimization in Energy and Electricity Markets*. Cham: Springer International Publishing, 2020, pp. 139–168.

[17] R. S. Go, F. D. Munoz, and J.-P. Watson, "Assessing the economic value of co-optimized grid-scale energy storage investments in supporting high renewable portfolio standards," *Applied Energy*, vol. 183, no. C, pp. 902–913, 2016.

[18] R. D. Zimmerman, "AC power flows, generalized OPF costs and their derivatives using complex matrix notation," MATPOWER *Technical Note 2*, Feb. 2010.