Combining Newton-Raphson and Stochastic Gradient Descent for Power Flow Analysis

Napoleon Costilla-Enriquez, *Student Member, IEEE,* Yang Weng, *Member, IEEE,* and Baosen Zhang, *Member, IEEE*

***Abstract*—The power flow problem is an indispensable tool to solve many of the operation and planning problems in the electric grid and has been studied for the last half-century. Currently, popular algorithms require second-order methods, which may lead to poor performance when the initialization points are poor or when the system is stressed. These conditions are becoming more common as both the generation and load profiles changes in the grid. In this paper, we present a hybrid first-order and second-order method that effectively escapes local minimums that may trap existing algorithms. We demonstrate the performance of our algorithm on standard IEEE benchmarks.**

I. Introduction

Power flow (PF) analysis is one of the most important and well-studied problems in the power system community. It is commonly formulated as finding the solution to a system of nonlinear algebraic equations, and a host of algorithms have been proposed to solve this system of equations. The most common among these is the Newton-Raphson (NR) method, where the inverse of the Jacobian is used to update the solutions iteratively [1], [2]. The popularity of the NR method (and its variants) is partially due to the fact that it has a fast speed of convergence. However, convergence is not guaranteed, especially if the initial guess is not close enough to the final solution, or the Jacobian matrix becomes ill-conditioned in the iteration process [3]. A number of approaches have been proposed to address these challenges, including augmenting the system states [4], comparing polar and rectangular formulations [5], improving the starting points [6], [7], and using different approximations of the Jacobian [8–10]. The authors from [11] take a different approach. They reformulate the PF problem as an optimization one and handle PV buses with complementarity constraints.

In this paper, we present an algorithm by combining gradient descent (GD) methods and the NR methods to overcome some of the standard computational challenges in PF problems. By formulating the PF problem as an optimization one, gradient descent steps can be taken without inverting the Jacobian matrix. In addition, we use stochastic gradient descent (SGD) to escape from local optima and saddle-points that would have trapped deterministic algorithms. Once the iterations are close enough to the global optimal solution, we can then utilize NR-type methods to accelerate the convergence.

The rest of the paper is organized as follows: Section II provides the problem formulation, Section III presents the algorithm, Section IV shows numerical simulations to validate our theory and makes a comparison between existing methods and our proposed algorithm. Section V concludes the paper.

N. Costilla-Enriquez and Y. Weng are with the School of Electrical, Computer and Energy Engineering at Arizona State University, {yang.weng,ncenriqu}@asu.edu. B. Zhang is with the Department of Electrical and Computer Engineering at the University of Washington, zhangbao@uw.edu

The authors are partially supported by NSF grant ECCS-1810537.

II. Problem Formulation

Consider a power system with n buses. For bus k, we denote its complex voltage, active power and reactive power as v_k, P_k and Q_k, respectively. We use bold fonts $\mathbf{v} = (v_1, \ldots, v_n)$, $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ to denote the vector version of the quantities. Let $\mathbf{Y}$ denote the admittance bus matrix. Then the power flow equation can be written in a compact form as $f(\mathbf{v}) = \mathbf{p} + j\mathbf{q} = \text{diag}(\mathbf{v}\mathbf{v}^{\mathbf{H}}\mathbf{Y}^{\mathbf{H}})$, where $(\cdot)^H$ denotes the Hermitian transpose [12].

Given a complex load vector $\mathbf{s}$, PF solves for the complex voltage vector $\mathbf{v}$ such that $f(\mathbf{v}) = \mathbf{s}$. Instead of directly solving this nonlinear equation, we consider the following optimization problem:

$$\min_{\mathbf{v}} \frac{1}{2}||f(\mathbf{v}) - \mathbf{s}||_2^2 = \min_{\mathbf{v}} \frac{1}{2}\sum_{i=1}^{n}\left(f_i\left(\mathbf{v}\right) - s_i\right)^2, \quad (1)$$

where if the PF problem is feasible, then the optimal value of the objective is 0, and there is a $\mathbf{v}^*$ such that $f(\mathbf{v}^*) = \mathbf{s}$. Given that the optimization problem is unconstrained with a smooth objective function, it is natural to use gradient descent to solve it.

III. Gradient Descent Based Methods

For notational simplicity, let $\mathcal{L}$ denote the objective function in (1). Its gradient with respect to $\mathbf{v}$ is given by the chain rule:

$$\nabla_{\mathbf{v}}\mathcal{L} = \mathbf{J}^T(f(\mathbf{v}) - \mathbf{s}), \quad (2)$$

where $\mathbf{J}$ is the power flow Jacobian. The standard GD algorithm is given by

$$\mathbf{v}_{t+1} = \mathbf{v}_t - \eta\nabla_{\mathbf{v}}\mathcal{L}(\mathbf{v}_t), \quad (3)$$

where t denotes the iteration number, and η is the step size or learning rate, which may be constant or adaptive. Let $\mathcal{I}_{ref}, \mathcal{I}_{PV}$, and $\mathcal{I}_{PQ}$ denote the sets of bus indices of the reference bus, PV buses, and PQ buses, respectively. Then, in (3), only the voltage angles $\delta_{\{i\}}$ for $i \notin \mathcal{I}_{ref}$ and the voltage magnitudes $v_{\{m\}}$ for $i \in \mathcal{I}_{PQ}$ will be updated. Note that by controlling which variables are updated at each iteration, we can also set specified voltage magnitudes on PV buses. From (2) and (3), the GD algorithm would stop under two conditions: 1) global optimal is reached and $f(\mathbf{v}) - \mathbf{s} = 0$, or 2) $\mathbf{J}$ loses rank and $f(\mathbf{v}) - \mathbf{s}$ is in the null space of $\mathbf{J}^T$.

The latter case means that the iterates $\mathbf{v}_t$ is trapped in a local minimum or in a saddle-point. To escape this minimum, it needs to stop following the gradient (since it is zero) and move in another direction. Of course, the direction it moves in should not be random. In this paper, we advocate for a type of stochastic gradient (SGD) approach [13]. Instead of computing the exact gradient, each SGD iteration performs a parameter update for a single term of the objective function in (1). Specifically, a randomly index i is picked at an iteration, and the gradient with respect to $\frac{1}{2}\left(f_i\left(\mathbf{v}\right) - s_i\right)^2$. We denote this gradient as $\nabla_{\mathbf{v}}\mathcal{L}_i\mathbf{v}_t$. The algorithm is shown in Fig. 1.

procedure SGD FOR PF
Select an initial vector $\mathbf{v}_0$, the number of maximum iterations $n_{iter,max}$, and $t = 0$
while $f(\mathbf{v}) - \mathbf{s} \le \epsilon$ or $n_{iter} \le n_{iter,max}$ **do** $\triangleright$ ϵ is the desired accuracy, and n_{iter} is the iteration number
Pick a random bus i from $\{1, \ldots, n\}$
Update the voltage vector:

$$\mathbf{v}_{t+1} = \mathbf{v}_t - \eta \cdot \nabla_{\mathbf{v}} \mathcal{L}_i (\mathbf{v}_t)$$

end while
return $\mathbf{v}_t$
end procedure

Fig. 1: SGD algorithm for the PF problem.

Even when the Jacobian $\mathbf{J}$ loses rank, it is typically close to full rank. Therefore, the SGD gradients $\nabla_{\mathbf{v}} \mathcal{L}_i \mathbf{v}_t$ would not all be in the null space of $\mathbf{J}$, and some of them would still provide useful directions for updating the voltage vector. Since the PF problem is non-convex, it typically has many local minima. The SGD algorithm allows the updates to "jump" out of local minima by the fluctuations induced by the randomness in the bus index selection process. Note if a global minimum is reached, then $\frac{1}{2}\left(f_i\left(\mathbf{v}\right) - s_i\right)^2$ is zero for all i, so the SGD algorithm terminates as well. However, SGD can also lead to slow convergence. Hence, we propose a hybrid method which consists in starting solving the PF problem with NR. If it stalls (when the condition number of the Jacobian degrades), we switch to the SGD method. After we escape from the local minimum, then the method switches back to the NR to reach the final solution, which is depicted in Fig. 2. The computational complexity for k iterations for the Newton-Raphson's method is $\mathcal{O}(k \times n^3)$ [14]. For the GD-based methods the computational complexity for k iterations is $\mathcal{O}(k \times n)$.

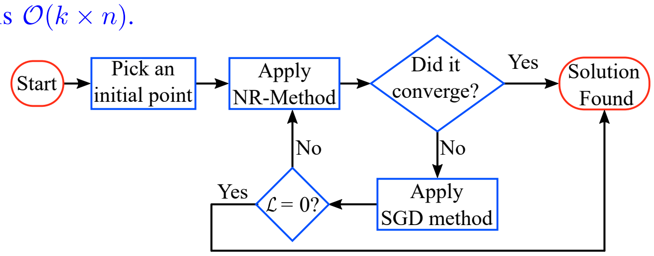
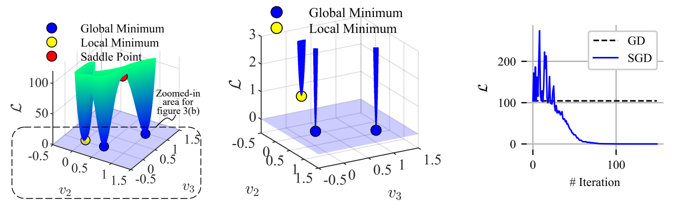


Fig. 2: Hybrid algorithm to solve PF problem.

IV. Simulation Results

A. Escaping Local Minima

We illustrate the behavior of the SGD algorithm using a 3-bus resistive network that is arranged in a line. We choose this example since we can explicitly plot the local minima and the saddle-point of the PF problem, as shown in Fig. 3a. To compare the behavior of the SGD and the standard NR algorithms, we initialize a NR solver at some point and track the error through the iterations. Of course, because of the local minimum and the saddle-point, a NR solver can get trapped at a suboptimal solution. At this point, a vanilla gradient algorithm also gets trapped, since the Jacobian loses rank. According to the algorithm in Fig. 1, we apply the SGD algorithm. As shown by Fig 3c, the SGD escapes this point and is able to converge to one of the global optimal solutions.



(a) Saddle point and minima. (b) Zoomed-in view. (c) Objective function values with the GD and SGD algorithms.

Fig. 3: Global, local minima, and objective function values of the 3-bus network.

B. Larger Systems

In this section, we illustrate the behavior and benefit of using the hybrid method in Fig. 2 to solve the PF problem for standard IEEE test systems. In the hybrid method, we start with the NR algorithms, and if we detect divergence (when the condition number of the Jacobian deteriorates), then we switch over to the SGD algorithm. After running a few SGD steps, we then again switch over to the NR iterates and repeat the process until an optimal solution is found.

The reason we switch back and forth between the NR and SGD updates is to utilize the NR algorithm as much as possible. Because if NR is able to converge, it will converge much faster (quadratic in the iterations) than when SGD is used. For a system where the operating points do not change appreciably, the NR algorithm can usually converge in a few iterations from a good starting point. However, when the operating conditions vary considerably, for example, when the penetration of renewable resources is significant, then finding a good start point becomes challenging [15], [16]. Therefore, the role of SGD is to "correct" the actions of the NR algorithm by escaping from suboptimal solutions and saddle-points when a bad starting point is used.

In order to show the usefulness of this hybrid approach, we performed a set of simulations. We compare the NR against our hybrid approach. For both PF methods, we model transformers and phase shifters with specified tap ratios and phase shift angles that are kept constant throughout the simulation. For all the simulations, we set specified voltage magnitudes and generator reactive power limits on PV buses. We enforce reactive power limits by using the conventional procedure [17]. That is, if any generator has a violated reactive power limit, its reactive injection is fixed at the limit, and the corresponding bus is converted to a PQ bus. For the SGD simulations, we used an adaptive learning rate. Specifically, we used the Adam adaptive learning algorithm [18] with stepsize $\eta = 0.01$, and exponential decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We set $n_{iter,max} = 100$. First, as a baseline, we perform simulations under normal load conditions with a flat start with the NR method and our proposed one. We can see the results in Table I, which shows that both methods are successful. However, when the conditions are changed, the NR method will struggle to find a solution to the PF problem, as we will show in the next simulations.

Now, we will change the starting points for the simulations. This means that the initial guesses will be further away from a solution. We first randomly pick a voltage

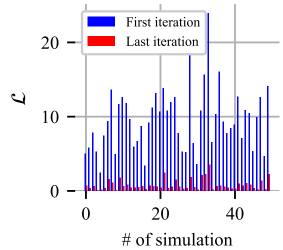


Fig. 4: Values of the first and last iteration.

TABLE I: Convergence results with a flat start.

Case	Convergence result with flat start NR	Hybrid
14-bus	Yes	Yes
39-bus	Yes	Yes
57-bus	Yes	Yes
118-bus	Yes	Yes
300-bus	Yes	Yes

solution vector $\mathbf{v}^*$ and compute the associated active and reactive power. Subsequently, we randomly pick initial starting points from a uniform distribution as follows: $|\mathbf{v}| \sim \mathcal{U}(\min(|\mathbf{v}|^*), \max(|\mathbf{v}|^*))$ and $\boldsymbol{\theta} \sim \mathcal{U}(\min(\boldsymbol{\theta}^*), \max(\boldsymbol{\theta}^*))$ and test whether the algorithms can reach $\mathbf{v}^*$ from the starting points. We use this random initialization to make more challenging the convergence when a PF solution exits. We use the 14-bus system as an illustration to explore the convergence of our hybrid method. The experiment design is as follows. We randomly initialize voltage angles and magnitudes (as we described before). Then, we use the SGD method for 50 iterations. Fig. 4 shows the result of 50 simulations, where the value of the objective function at the first iteration (in blue) is quite large, but the value of the last iteration (in red) is minimal. Then, we initialize the NR method with the starting points associated with the first value iteration and the points associated with the last iteration. The result of doing this experiment is shown in Table II (under normal load), where the convergence rate is 10% with the NR method. On the other hand, we have a 100% convergence rate with our hybrid approach. We carry out the same experiments for the 39-, 57-, 118-, and 300-bus cases, in which we obtained better results than the conventional NR approach, as shown in Table II.

For the proposed method under worse conditions, we make simulations with a higher load level. We increase the power system load by multiplying the active load by a factor α that will produce an ill-conditioned test case. Table II shows the result (under heavy load), where the convergence rate of the hybrid method is better than the NR method.

Finally, we explore the convergence of both PF methods around local minimum and saddle points. For this numerical experiment, we perform two simulations by test case. In one simulation, we choose the starting point to be a local minimum one, and in the other simulation, we choose the starting point to be a saddle-point. The results are shown in Table II. We can see that the NR method does not converge, whereas we achieve convergence in both simulations with our method. This result is expected due to the NR Jacobian is singular at the first iteration.

V. Conclusion

A novel hybrid method for the power flow problem is proposed in this paper. This method combines the Newton-Raphson and stochastic gradient algorithms to achieve fast convergence speed as well as the ability to escape local minima and saddle points.

Numerical tests on power systems of various sizes and topologies demonstrated the effectiveness and efficiency of the

TABLE II: Convergence of different test cases for the NR and Hybrid methods with original and increased loads.

	Convergence rate (%)				Convergence Result	
	Normal Load		Heavy Load		Local minimums and saddle points	
Case	**NR**	**Hybrid**	**NR**	**Hybrid**	**NR**	**Hybrid**
14-bus	10	100	10	96	No	Yes
39-bus	10	100	10	84	No	Yes
57-bus	10	90	0	82	No	Yes
118-bus	5	80	5	66	No	Yes
300-bus	5	75	0	61	No	Yes

proposed approach in solving fast and reliable the PF problem under different load conditions and initial starting points.

In our future work, we would like to explore the voltage stability problem with our proposed framework. Also, we will investigate the possibility of finding and ranking problematic buses that need power compensation to converge to a solution with our proposed method.

References

[1] B. Stott, "Review of load-flow calculation methods," *Proceedings of the IEEE*, vol. 62, no. 7, pp. 916–929, 1974.
[2] W. F. Tinney and C. E. Hart, "Power flow solution by newton's method," *IEEE Transactions on Power Apparatus and systems*, no. 11, pp. 1449–1460, 1967.
[3] F. Milano, "Continuous newton's method for power flow analysis," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 50–57, 2009.
[4] A. G. Expósito and E. R. Ramos, "Augmented rectangular load flow model," *IEEE Transactions on Power Systems*, vol. 17, no. 2, pp. 271–276, 2002.
[5] V. Da Costa and A. Rosa, "A comparative analysis of different power flow methodologies," *IEEE PES Transmission and Distribution Conference and Exposition: Latin America*, pp. 1–7, 2008.
[6] A. M. Sasson, C. Trevino, and F. Aboytes, "Improved newton's load flow through a minimization technique," *IEEE Transactions on Power Apparatus and Systems*, no. 5, pp. 1974–1981, 1971.
[7] L. M. Braz, C. A. Castro, and C. Murati, "A critical evaluation of step size optimization based load flow methods," *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 202–207, 2000.
[8] B. Stott and O. Alsac, "Fast decoupled load flow," *IEEE transactions on power apparatus and systems*, no. 3, pp. 859–869, 1974.
[9] Y. Chen and C. Shen, "A Jacobian-free newton method with adaptive preconditioner and its application for power flow calculations," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1096–1103, 2006.
[10] S. Abhyankar, Q. Cui, and A. J. Flueck, "Fast power flow analysis using a hybrid current-power balance formulation in rectangular coordinates," in *IEEE PES T&D Conference and Exposition*, 2014, pp. 1–5.
[11] M. Pirnia, C. A. Cañizares, and K. Bhattacharya, "Revisiting the power flow problem based on a mixed complementarity formulation approach," *IET Generation, Transmission & Distribution*, vol. 7, no. 11, pp. 1194–1201, 2013.
[12] B. Zhang and D. Tse, "Geometry of injection regions of power networks," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 788–797, 2013.
[13] D. P. Bertsekas, "Nonlinear programming," *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 334–334, 1997.
[14] R. Battiti, "First-and second-order methods for learning: between steepest descent and newton's method," *Neural computation*, vol. 4, no. 2, pp. 141–166, 1992.
[15] Y. Weng, R. Rajagopal, and B. Zhang, "A geometric analysis of power system loadability regions," *IEEE Transactions on Smart Grid*, 2019.
[16] D. K. Molzahn, V. Dawar, B. C. Lesieutre, and C. L. DeMarco, "Sufficient conditions for power flow insolvability considering reactive power limited generators with applications to voltage stability margins," in *IREP Symposium Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid*, 2013, pp. 1–11.
[17] D. P. Kothari and I. Nagrath, *Modern power system analysis*. Tata McGraw-Hill Education, 1989.
[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.