SAVER: Safe Learning-Based Controller for Real-Time Voltage Regulation

Yize Chen, Yuanyuan Shi, Daniel Arnold and Sean Peisert

2022 IEEE Power & Energy Society General Meeting (PESGM) | 978-1-6654-0823-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/PESGM48719.2022.9917098

***Abstract*—Fast and safe voltage regulation algorithms can serve as fundamental schemes for achieving a high level of renewable penetration in the modern distribution power grids. Faced with uncertain or even unknown distribution grid models and fast-changing power injections, model-free deep reinforcement learning (DRL) algorithms have been proposed to find the reactive power injections for inverters while optimizing the voltage profiles. However, such data-driven controllers can not guarantee satisfaction of the hard operational constraints, such as maintaining voltage profiles within a certain range of the nominal value. To this end, we propose SAVER: SAfe VoltagE Regulator, which is composed of an RL learner and a specifically designed, computational efficient safety projection layer. SAVER provides a plug-and-play interface for a set of DRL algorithms that guarantees the system voltages to be within safe bounds most of the time. Numerical simulations on real-world data validate the performance of the proposed algorithm.**

***Index Terms*—Machine Learning, Power Systems Operations, Reinforcement Learning, Safety**

I. Introduction

The voltage regulation problem has been investigated for decades, yet the increasing penetration from distributed renewable resources keeps adding new challenges to such foundational control tasks. With the greater fluctuations coming from active power injections (e.g., rooftop solar panels and electric vehicles), along with the high r/x ratios of distribution lines/cables, unacceptable voltage swings may appear in the current distribution grids [1]. While on the other hand, smart inverters of fast-acting distributed energy resources (DERs) can provide reactive power injections in real-time, which can be systematically designed to optimize the voltage profiles [2].

Previous efforts on the voltage regulation problem focus on designing either centralized or decentralized controllers with optimality guarantees with exact grid models [3], [4]. This requires distribution grid system operators to either know the exact topology and line parameters or take extra steps to identify such modeling knowledge [5]. With the increasing availability of grid measurements and sensing data, there is a growing interest in designing model-free, data-driven voltage regulation algorithms such as reinforcement learning (RL) to achieve real-time decision making [6], [7].

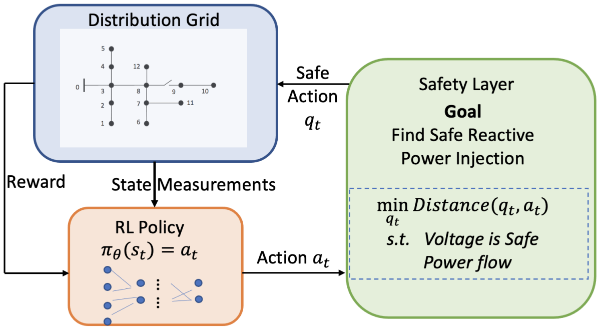

Fig. 1: The proposed learning-based control strategy to safely regulate distribution grid voltage.

The RL training process holds the promise of finding control policies with good performance in terms of minimizing the voltage deviations and regulating costs, while it does not need explicit knowledge of the grid parameters. By aggregating nodal voltage and active power injections as the states, the RL agent is trained by interacting with the power grid environment by using learned reactive power injections as actions. However, compared to the model-based counterparts [4], RL is trained to maximize the accumulated reward, while hard physical constraints such as feasible voltage magnitudes are mostly not guaranteed to be satisfied. Unsafe reactive power injections can cause severe impacts over the grids such as voltage collapse and load shedding [8], [9].

Recent works seek to promote RL safety by reward function shaping [10], whereby combining constraint violation information, the RL agent learns to avoid unsafe actions and states as such can lead to a high penalty. However, in the training process agents still have to violate the safety constraint and receive the penalty multiple times before it learns to avoid it. In addition, safety is more like an implicit regularization in such methods, as violations of the safety constraint can lead to high costs, while it does not always guarantee safety during implementation. In [11], [12], the voltage regulation problem is modeled as constrained Markov Decision Process (CMDP) to constrain the state in safe space probabilistically, yet no formal guarantees can be made regarding the voltage profiles in the training or implementation stage.

In this paper, we propose SAVER, which is composed of a novel safety layer to overcome such safety concerns regarding RL voltage regulators. By making use of the underlying grid information, we design a projection layer that projects the reactive power injection outputted by the trained RL policy into a safety set of nodal voltage magnitudes. The scheme of the resulting procedure is illustrated in Fig. 1. The proposed method can fast compute *safe reactive power injections in terms of voltage constraints with guarantees*. Moreover, the safety learning framework can be embedded as a lightweight plug-and-play module for *most if not all* standard reinforcement learning algorithms. This work is partly inspired by [13],

Y. Chen, D. Arnold and S. Peisert are with the Lawrence Berkeley National Laboratory, emails: {yizechen, dbarnold, sppeisert}@lbl.gov. Y. Shi is with the University of California San Diego, email: yyshi@eng.ucsd.edu.

978-1-6654-0823-3/22/$31.00 ©2022 IEEE

where a safe exploration strategy is proposed for reinforcement learning agents in the area of robotic control. We also note that [14] proposed a linear projection for safe voltage control policy, yet it only works for the special case where only one single node's voltage hits the safety boundary, while this paper's safety layer works for the general voltage safety set.

The advantages of applying the proposed data-driven voltage regulator are multi-fold. First, it utilizes the empirically proven advantages of RL algorithms to search the space of neural network parameterized voltage controllers, that can optimize the voltage derivations and control costs without the exact grid information. Second, it limits the search of the safe policy only to local control actions within the neighborhood of a trained RL policy, adding little computation burden for the resulting controller. Third, the use of such a safety layer encodes the physical constraints, and opens the door for designing a practical controller for inverters.

II. Preliminaries: Power Flow Models and Problem Formulation

In this work, we focus on the voltage regulation model for a radial distribution network. We use the graph $\mathcal{G} = (\mathcal{N} \cup \{0\}, E)$ to represent a radial distribution feeder with $N+1$ buses with bus 0 denoting the reference bus, and E is the set of connected transmission lines. We denote $\mathbf{p} \in \mathcal{R}^N$ and $\mathbf{q} \in \mathcal{R}^N$ as the nodal active and reactive power injections respectively.

The relationship between the voltage and power injections can be stated as power flow equations. For each line, denote $s_{ik} = p_{ik} + jq_{ik}$ as the complex power flow from bus i to bus k, and denote the line impedance as $z_{ik} = r_{ik} + jx_{ik}$. We adopt the DistFlow formulation [15] as follows

$$-p_j = p_{ij} - r_{ij}l_{ij} - \sum_{k:(j,k)\in E} p_{jk}, j = 1, ..., n \tag{1a}$$

$$-q_j = q_{ij} - x_{ij}l_{ij} - \sum_{k:(j,k)\in E} q_{jk}, j = 1, ..., n \tag{1b}$$

$$v_j = v_i - 2\left(r_{ij}p_{ij} + x_{ij}q_{ij}\right) + (r_{ij}^2 + x_{ij}^2)l_{ij}\,, (i,j) \in E \tag{1c}$$

$$l_{ij} = \frac{p_{ij}^2 + q_{ij}^2}{v_i} \tag{1d}$$

where $l_{ij} = |I_{ij}|^2$, $v_i = |V_i|^2$. Equation (1) defines a nonlinear relationship between the active power injection $\mathbf{p}$, reactive power injection $\mathbf{q}$, and the nodal voltage magnitude $\mathbf{v}$.

We consider our control devices are inverter-based DERs (such as renewable generators, battery energy storage), that can change their reactive power output $\mathbf{q}$ in a fast timescale to provide voltage regulation. Denote the set of controllable $q_i, i \in C$ using the set of nodes C. At each time step, the optimal voltage regulation problem can be then formulated as follows:

$$\min_{\boldsymbol{q}} \quad \sum_{i\in C} c_i^q\left(q_i\right) + \eta \sum_{i\in\mathcal{N}} c_i^v\left(v_i\right) \tag{2a}$$

$$\text{s.t.} \quad \underline{v}_j \leq v_j(\boldsymbol{p}, \boldsymbol{q}) \leq \bar{v}_j, \qquad j \in \mathcal{N}. \tag{2b}$$

The control objective (2a) is to reduce the total control cost for controllable inverters plus the penalty on voltage deviation for all buses, with η as a hyperparameter that balances the weights of the two costs. System operators can choose different function forms of $c_i^q(\cdot)$ and $c_i^v(\cdot)$ to achieve operational goals. Equation (2b) is the voltage safety constraints that should be ensured at each time step.

The challenge of directly solving (2) lies in the fact that even though we can design a convex cost function (e.g., a quadratic cost over reactive power injection and squared loss of voltage deviation), the underlying power flow (1) is nonlinear and non-convex, making directly solving the optimization problem involving (2b) a hard problem. For the convex relaxation methods based on SOCP or SDP formulations, even though the resulting optimization problem is tractable, it needs the exact information on grid topology and line parameters. In addition, it can take a significant amount of time to solve the optimization problem for large grids, which may not fulfill the goal of achieving fast timescale voltage regulation. These challenges lead to the design need for a computationally efficient model-free controller.

III. Learning a Voltage Regulator

In this section, we describe how we formulate the voltage regulation problem as a RL problem, and illustrate the need for explicitly incorporating safety as a constraint for standard RL algorithm.

RL provides a powerful paradigm for solving (2), in the sense that during the training process, we can train a policy network that maps the states to reactive power injections, to minimize the control objectives defined by (2a). First, we define a Markov Decision Process (MDP) of 4-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, r)$ to represent the voltage control model. The states and actions for timestep t can be defined as,

$$\mathbf{s}(t) := ((v_i)_{i\in\mathcal{N}}(t), (p_i)_{i\in\mathcal{N}}(t)); \tag{3a}$$

$$\mathbf{a}(t) := ((q_i)_{i\in C}(t)). \tag{3b}$$

Without loss of generality, in this paper, we use $\mathbf{s}(t), \mathbf{a}(t)$ and $\mathbf{s}, \mathbf{a}$ interchangeably. The state transition model $\mathcal{P} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$, is determined by the power flow equations (1) and external dynamics such as renewable generation and nodal demand; and $r : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is a scalar function that is defined as follows

$$r(\mathbf{s}(t), \mathbf{a}(t)) = ||\mathbf{v}(t) - \mathbf{v}_0||_2^2 + \eta||\mathbf{q}(t)||_2^2; \tag{4}$$

where $\mathbf{v}_0 = v_0 \cdot \mathbf{1}$ and v_0 is the feeder head voltage.

To summarize, given MDP model and an initial policy $\boldsymbol{\pi}_\theta$, the RL for optimal voltage control problem is formulated as

$$\max_{\theta} \quad J(\theta) := E_{\mathcal{P},\pi}\left[\sum_{t=1}^{\infty} \gamma^t r(\mathbf{s}(t), \mathbf{a}(t))\right] \tag{5a}$$

$$\text{s.t.} \quad \mathbf{a}(t) = \boldsymbol{\pi}_\theta(\mathbf{s}(t))\,, \tag{5b}$$

$$g(\mathbf{s}(t), \mathbf{a}(t)) \in \mathcal{G}, \tag{5c}$$

$$\boldsymbol{s}(t+1) = f\left(\boldsymbol{s}(t), \boldsymbol{a}(t), \boldsymbol{y}_{ext}(t)\right); \tag{5d}$$

where $\boldsymbol{\pi}_\theta(\mathbf{s}_t)$ are the parameters of the voltage control policy, and $\mathcal{G}$ represents the set of safety constraints. γ is a discount factor. To be more specific, $g(\mathbf{s}(t), \mathbf{a}(t))$ refers to the state

978-1-6654-0823-3/22/$31.00 ©2022 IEEE

safety constraint $\underline{v}_j \le v_j(\boldsymbol{p},\boldsymbol{q}) \le \bar{v}_j, j \in \boldsymbol{N}$. The underlying dynamics including external variable $\boldsymbol{y}_{ext}$ is modeled as (5d). During the training phase, RL agent interacts with the environment driven by the power flow equations, and learns the reactive power dispatch $\mathbf{a} = \pi_\theta(\mathbf{s})$ to maximize the accumulated discounted reward. Recent developments of deep RL models have enabled a set of deep learning based algorithms to efficiently solve such reward maximization problem, such as Deep Q Learning for discrete actions (e.g., tap position) [16], and Deep Deterministic Policy Gradient (DDPG) for continuous actions [17].

However, one intrinsic challenge for DRL policies is to validate the policies given by deep neural networks can always guarantee the voltage staying within the safe region $[\underline{v}_i, \bar{v}_i], i \in \boldsymbol{N}$. This is due to the fact that during the DRL training stage, the RL agent only focused on maximizing the reward while there is no hard constraint on $\boldsymbol{\pi}_\theta$ to ensure safety. In addition, even if the learned policy appears "safe" on the training data (with finite training episodes), it is not guaranteed to be safe under all scenarios. The lack of formal safety guarantees is a major obstacle in the deployment of RL to real-world power systems, as a violation of safe operation constraints can lead to severe impacts such as voltage collapse and cascading failure. Motivated by the challenge above, our goal is to ensure the RL algorithm obtain provably safety guarantee *during both policy training and policy deployment (after training)*.

IV. Safety Voltage Layer

In this section, we will first discuss how to explicitly model the safety constraint set C for the RL controller with a tractable form. We will then describe our safe controller design.

To ensure safety during both training and execution, we first need to identify unsafe actions from RL agents, and then modify the reactive power injections so that the voltage is within safe bounds. We achieve this by incorporating the power flow relationship in a compact way for the learned RL agents. In the original DistFlow formulation, we can further neglect the line loss via setting $l_{ij} = 0$ for all $(i,j) \in E$. For each node, by assuming $v_i \approx 1$, while by approximating $v_j^2 - v_i^2 \approx 2(v_j - v_i)$, we can get the linearized version of DistFlow model

$$\begin{aligned} -p_j &= p_{ij} - \sum_{k:(j,k)\in E} p_{jk} \\ -q_j &= q_{ij} - \sum_{k:(j,k)\in E} q_{jk} \\ v_i - v_j &= 2(r_{ij}p_{ij} + x_{ij}q_{ij}). \end{aligned} \tag{6}$$

By collecting $\mathbf{v} = [v_1, ..., v_n]^T$, and substituting p_{ij}, q_{ij} into the last equation of (6), we can represent the voltage profile in a more compact form

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{R}\mathbf{p} + \mathbf{X}\mathbf{q}, \tag{7}$$

where v_0 is the voltage for the feeder head, and $\mathbf{R}$ and $\mathbf{X}$ are positive matrices.

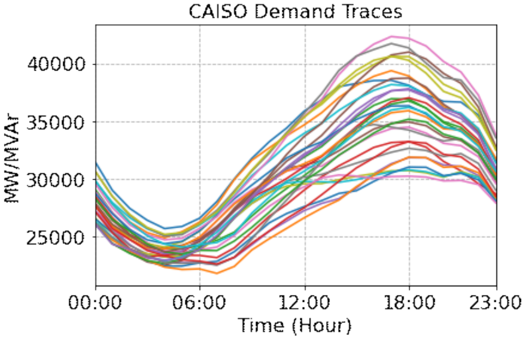

Fig. 2: CAISO daily demand data samples used for RL training.

Once we get the policy $\boldsymbol{\pi}_\theta$ and the safety constraint model for the voltage profile, we want to find the reactive power injections that avoid unsafe voltage deviations. At each control time step, it leads to the following optimization problem

$$\begin{aligned} \min_{\boldsymbol{q}} \quad & \frac{1}{2}||\boldsymbol{q} - \boldsymbol{\pi}_\theta(\mathbf{s}(t))||_2^2 \\ \text{s.t.} \quad & \boldsymbol{R}\boldsymbol{p} + \boldsymbol{X}\boldsymbol{q} + \mathrm{v}_0 \le \bar{\boldsymbol{v}} \\ & \boldsymbol{R}\boldsymbol{p} + \boldsymbol{X}\boldsymbol{q} + \mathrm{v}_0 \ge \underline{\boldsymbol{v}}. \end{aligned} \tag{8}$$

By solving (8), we find $\mathbf{q}$ which are not only close to the action given by the deep RL agent, but also satisfying the hard constraints over nodal voltage.

Since (8) is a convex, quadratic optimization problem, and we can use standard QP-solvers to solve it efficiently in polynomial time [18]. Alternatively, we can first learn the active constraints, and then use the closed-form solution for the equality constrained problem. If there is at most one constraint that is tight, [13] developed a closed-form solution. However, for voltage control, there is usually more than one constraint on voltage magnitude that can be active. Therefore, we need to solve (8) rather than using the closed-form solution, unless all active constraints can be identified beforehand [19]. In the following, we also give two remarks regarding the linear power flow model and the model knowledge.

Remark 1 *Effects of Power Flow Linear Approximation* (7)*:* We note that proposed method is a combination of model-free RL and model-based optimization approach. In our safety layer design, the use of linearized power flow is restricted to validate if the policy $\boldsymbol{\pi}_\theta$ gives safe voltage. Thus we are not losing the representation capability of deep RL on the complex relationship between power injections and voltage. Moreover, the linear DistFlow model has been proved accurate verified by real-world data when estimating the voltage box constraints in (8) [20].

Remark 2 *Knowledge of Distribution Grid:* Another assumption we make for the proposed safety layer is that we can explicitly write out the inequality constraints involving grid parameters $\mathbf{R}$ and $\mathbf{X}$. Such assumption can be justified when the RL agent has access to historical grid operational data, and can therefore employ machine learning or statistical methods to estimate these parameters. There are also emerging techniques for estimating both the network topology and line parameters for distribution grid with observational data [5], [21]. We leave the discussion of incorporating model learning error to tighten voltage safety range $[\underline{v}, \overline{v}]$ to future work.

V. Case Study

In this section, we demonstrate the effectiveness of the proposed safety layer method for voltage control. Specially, we show that by incorporating the safety layer in a standard deep deterministic policy gradient (DDPG) [17] algorithm, we can always find actions satisfying voltage constraints. We compare the proposed safety layer with baseline linear policy or deep RL methods, and validate the improved performance in terms of control costs and safety measures.

Experiment Setup Throughout the simulation, we use the IEEE 13-bus test feeder to validate the algorithm performance. We use real-world load data from CAISO[1], and normalize the demand value based on the single-phase 13-bus system configurations. The nominal voltage magnitude at each bus is $12kV$, and we allow $\pm 5\%$ voltage deviation for each node. We visualize 20 sample days of training data in Fig. 2.

We employ DDPG algorithm to train the deep RL agent, and we employ 3-layer neural networks for both the policy network and value network in the DDPG agent. For each episode, we collect rollouts for 24 hours, and we also keep a replay buffer for states throughout RL training process.

For model comparison, we also use a linear policy to output the reactive power injections. Such linear policy actively considers the voltage magnitude bound, but may take non-optimal control actions which incur larger system costs or short-term voltage violations. We refer readers to [22] for the details of such an algorithm. For the implementation of all the algorithms, we also constrain the reactive power injections within box constraints considering the regulating capabilities of the inverter of each DER. Once we get the control actions from each controller, we use the full power flow model to calculate the resulting voltage. We test the performance of the proposed safe RL approach, standard RL approach, and baseline linear policy using a high-resolution real-world load and renewable data in Fig. 3 (left). Based on the time resolution of the PV and load trajectory, controllers adjust their control output every 6s. We use Pytorch to build all RL models and run the training process. We report the computation time by using a MacBook Pro Personal Laptop with 16 GB 2400 MHz DDR4 memory and 2.2 GHz Intel Core i7 processor. Training time for our case is within 10 minutes.

Simulation Results We first compare the voltage profile using three different control schemes. In Fig. 3, we show the nodal voltage magnitude for one day's test data. It can be seen that all three algorithms take actions to try to stabilize the voltage within the safe region ($[11.4kV, 12.6kV]$). But during the middle of the day, both linear controller and standard RL agent lead to greater voltage deviation than the operational limits. The linear policy is relatively "slow" in the sense of acting to load and renewables generation change, causing the spikes in voltage profiles. The RL agent finds the control actions using least control efforts (which will be explained in detail in Table I), but more than half of nodal voltage at noon are exceeding the upper limits. This shows that even though the trained RL agent is able to find control actions to maximize the reward, such training scheme can not exclude unsafe voltage deviations during test time. On the contrary, the proposed safe layer helps maintain voltage staying within the safety bounds. We can further observe that the resulting safe policy manages to reduce the voltages for multiple buses at the same time, meaning that it is possible to refine a trained RL agent to explicitly handle the hard, safety constraints.

Method	Time (s)	Average q_i (kVAR)	$v_i > \%5$ limit
Linear	9.02×10^{-3}	1.968	9.96%
RL	8.88×10^{-3}	1.543	10.82%
Safe RL	1.92×10^{-2}	1.829	0.01%

TABLE I: Statistics for average computation time (per instance), average reactive power injection, and the frequency of infeasible voltages for linear policy, standard RL policy and safe RL policy.

In Fig. 4, we look into the nodal voltage profile, where mean and variance of voltage deviation are plotted for the three methods. We can observe that linear policy leads to voltage profile with the largest deviation across all buses, which shows that the linear model may not achieve satisfactory performance faced with renewables integration in the distribution grid. On average, the safe RL policy can reduce the voltage deviation by more than 30% compared to RL counterparts, showing the necessity of incorporating safety constraints into model-free algorithm design.

We report the statistics for solution time, average control efforts, and control results in Table I. Evaluations are based on simulation results for all 14420 test data samples. Compared to linear and standard RL policies, safe RL tends to use a bit more reactive power to realize safety. The linear policy and RL agent take nearly the same time to compute the reactive power injections. The safe RL does not add much burden to the fast RL policy inference process. The average computation time for the safe RL method is still much smaller than the control step resolution (6s), making it realistic to implement for real-time voltage regulation. This shows that the benefits of the proposed algorithm: with minimal added computational costs, we can find safe policies as a plug-in-play module for off-the-shelf RL algorithms. And all three methods take much shorter time than solving a model-based optimization problem, e.g., by taking SOCP relaxation on the voltage regulation task [6]. Note that in practice, we can also resort to the safety layer computation to mini-batches for parallel computation, which can further accelerate the solution process of the proposed algorithm. Such features allow over design can be scaled up to larger systems.

The effects of the safety layer are further reflected in the average reactive power injection, and the occurrences of unsafe voltage deviation. As is shown in Table I, safe RL may take more reactive power resources to achieve the regulation task. But compared to both linear policy and standard DDPG agent, the safe RL agent can guarantee for almost all test instances, the resulting voltages are safe. Only for 0.01% of the test samples, safe RL can not find a safe policy, which may be caused by the representation limitation of the linearized power flow model, or the inability to find feasible control actions by

[1] https://www.caiso.com/Documents/HistoricalEMSHourlyLoadDataAvailable.html

978-1-6654-0823-3/22/$31.00 ©2022 IEEE

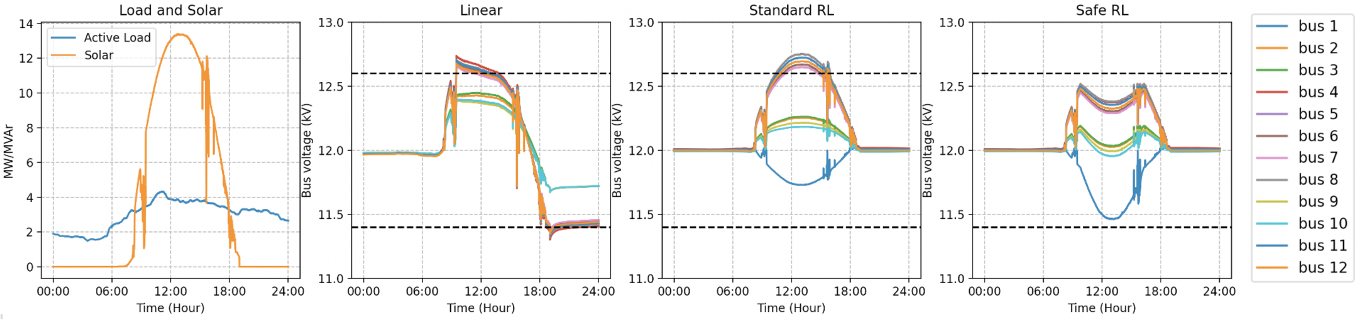

Fig. 3: Results on voltage regulation. Dark dashed lines denote the $[\underline{v}_i, \bar{v}_i]$ region. By using proposed safety layer, the resulting reactive power injections enforce the voltage magnitude to be within the safe regions.

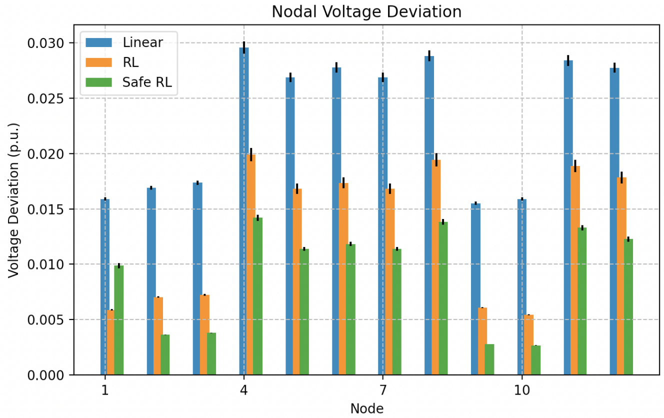

Fig. 4: The mean and variance of voltage deviation on each bus under different control schemes.

only controlling reactive power injections at DER buses.

VI. Conclusion and Discussion

In this work, we present SAVER, a computationally efficient safe layer for learning-based voltage regulators. By explicitly taking the constraints on voltage magnitude into controller design, we show the resulting controller can output reactive power injections that guarantee voltage safety almost surely. Simulation results demonstrate such safety layer involving linear constraints over the voltage magnitude can be an effective plug-in module for off-the-shelf deep RL algorithms. For future work, we will explore how to autonomously identify the nonlinear relationship between reactive power injections and nodal voltages and ensure fine-grained, safe control at the same time. More advanced safe control schemes based on both active and reactive power injections will also be investigated.

References

[1] P. M. Carvalho, P. F. Correia, and L. A. Ferreira, “Distributed reactive power generation control for voltage rise mitigation in distribution networks,” *IEEE transactions on Power Systems*, vol. 23, no. 2, pp. 766–772, 2008.

[2] D. G. Photovoltaics and E. Storage, “Ieee standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces,” *IEEE Std*, pp. 1547–2018, 2018.

[3] G. Qu and N. Li, “Optimal distributed feedback voltage control under limited reactive power,” *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 315–331, 2019.

[4] M. Farivar, C. R. Clarke, S. H. Low, and K. M. Chandy, “Inverter var control for distribution systems with renewables,” in *2011 IEEE international conference on smart grid communications (SmartGridComm)*. IEEE, 2011, pp. 457–462.

[5] D. Deka, S. Backhaus, and M. Chertkov, “Structure learning in power distribution networks,” *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1061–1074, 2017.

[6] Y. Chen, Y. Shi, and B. Zhang, “Input convex neural networks for optimal voltage regulation,” *arXiv preprint arXiv:2002.08684*, 2020.

[7] H. Xu, A. D. Domínguez-García, and P. W. Sauer, “Optimal tap setting of voltage regulation transformers using batch reinforcement learning,” *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1990–2001, 2019.

[8] T. L. Vandoorn, J. De Kooning, B. Meersman, and L. Vandevelde, “Voltage-based droop control of renewables to avoid on–off oscillations caused by overvoltages,” *IEEE Transactions on Power Delivery*, vol. 28, no. 2, pp. 845–854, 2013.

[9] Y. Chen, D. Arnold, Y. Shi, and S. Peisert, “Understanding the safety requirements for learning-based power systems operations,” *arXiv preprint arXiv:2110.04983*, 2021.

[10] X. Huang, Z. Ding, and Z. Zhang, “A guided deep reinforcement learning method for distribution voltage regulation via battery systems,” in *2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2021, pp. 1–5.

[11] H. Liu and W. Wu, “Online multi-agent reinforcement learning for decentralized inverter-based volt-var control,” *IEEE Transactions on Smart Grid*, 2021.

[12] W. Wang, N. Yu, Y. Gao, and J. Shi, “Safe off-policy deep reinforcement learning algorithm for volt-var control in power distribution systems,” *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3008–3018, 2019.

[13] G. Dalal, K. Dvijotham, M. Vecerik, T. Hester, C. Paduraru, and Y. Tassa, “Safe exploration in continuous action spaces,” *arXiv preprint arXiv:1801.08757*, 2018.

[14] P. Kou, D. Liang, C. Wang, Z. Wu, and L. Gao, “Safe deep reinforcement learning-based constrained optimal control scheme for active distribution networks,” *Applied Energy*, vol. 264, p. 114772, 2020.

[15] M. E. Baran and F. F. Wu, “Optimal capacitor placement on radial distribution systems,” *IEEE Transactions on power Delivery*, vol. 4, no. 1, pp. 725–734, 1989.

[16] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, “Playing atari with deep reinforcement learning,” *arXiv preprint arXiv:1312.5602*, 2013.

[17] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” *arXiv preprint arXiv:1509.02971*, 2015.

[18] Y. Ye and E. Tse, “An extension of karmarkar’s projective algorithm for convex quadratic programming,” *Mathematical programming*, vol. 44, no. 1, pp. 157–179, 1989.

[19] Y. Chen and B. Zhang, “Learning to solve network flow problems via neural decoding,” *arXiv preprint arXiv:2002.04091*, 2020.

[20] S. Lin and H. Zhu, “Data-driven modeling for distribution grids under partial observability,” *arXiv preprint arXiv:2108.08350*, 2021.

[21] S. Park, D. Deka, and M. Chcrtkov, “Exact topology and parameter estimation in distribution grids with minimal observability,” in *2018 Power Systems Computation Conference (PSCC)*. IEEE, 2018, pp. 1–6.

[22] N. Li, G. Qu, and M. Dahleh, “Real-time decentralized voltage control in distribution networks,” in *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2014, pp. 582–588.

978-1-6654-0823-3/22/$31.00 ©2022 IEEE