Synthesizing Power Flow Datasets via Constrained Diffusion Models

Milad Hoseinpour, Prof. Vladimir Dvorkin

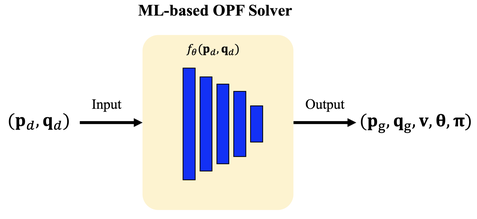
Department of Electrical and Computer Engineering, University of Michigan

Motivation

- There is an increasing demand for large-scale and high-quality datasets for machine learning and data analysis tasks in power systems.
- Challenges in Data Availability: privacy and security concerns, data sparsity, etc.

A synthetic dataset is artificially generated data that mirrors the statistical properties of real-world data without containing any actual records.

- Synthetic datasets are essential for developing and benchmarking trustworthy ML-based optimal power flow (OPF) solvers.



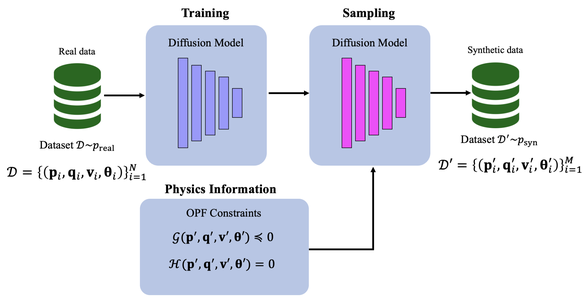
Problem setup

Goal: Given a dataset including real power flow data points, we aim to synthesize (1) statistically representative and (2) high fidelity power flow data points:

$$\text{dist}\left(p_{\text{syn}} \,||\, p_{\text{real}}\right) \le \epsilon \rightarrow \text{“statistically representative”}$$

$$\begin{cases} \mathcal{G}(\mathbf{p}_i', \mathbf{q}_i', \mathbf{v}_i', \boldsymbol{\theta}_i') \preceq 0, & \forall i = 1, \dots, M \\ \mathcal{H}(\mathbf{p}_i', \mathbf{q}_i', \mathbf{v}_i', \boldsymbol{\theta}_i') = 0, & \forall i = 1, \dots, M \end{cases} \rightarrow \text{“high fidelity”}$$

where $\mathcal{G}(\cdot)$ and $\mathcal{H}(\cdot)$ are OPF inequality and equality constraints.



A high-level view of the problem setup.

Diffusion Models

- **Forward** diffusion process that gradually adds noise to input

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, , \quad \epsilon_t \sim \mathcal{N}(0, \mathbb{I}), \quad t \in (0, T].$$

- **Reverse** denoising process that learns to generate data by denoising

$$\mathbf{x}_{t-1} = \mu_\theta(\mathbf{x}_t, t) + \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \mathbb{I}), \quad t \in (T, 0].$$

- **Training:** The loss function to train the denoiser neural network ϵ_θ is

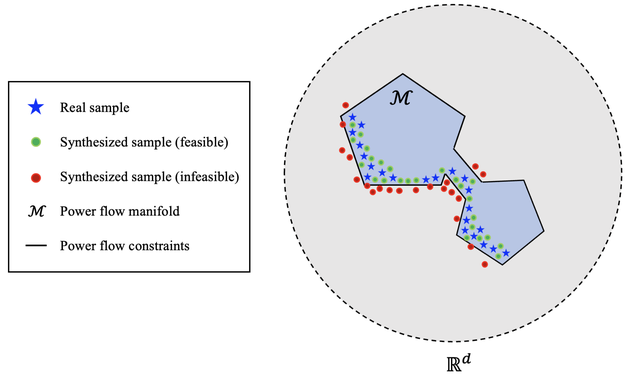
$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{\mathbf{x}_0, \epsilon, t}\left[\|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\|^2\right].$$

- **Sampling:**

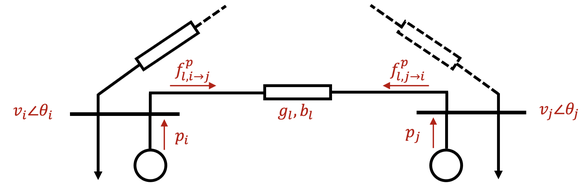
$$\mathbf{x}_{t-1} = \frac{\sqrt{\alpha_t(1-\bar{\alpha}_{t-1})}}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}\beta_t}}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \sigma_t z, \quad z \sim \mathcal{N}(0, \mathbb{I}), \quad t \in [T, 0).$$

Diffusion Guidance based on Power Flow Constraints

- **Theoretically**, a diffusion model trained on a dataset of **feasible** power flow data points should satisfy the power flow constraints.
- In practice, a diffusion model may generate power flow data points that are **infeasible** due to learning and sampling errors.



Power Flow Equality Constraints



- Active and reactive power balance constraints can be represented as follows:

$$p_b - \sum_{l\in\mathcal{L}:i=b} f^p_{l,i\to j} - \sum_{l\in\mathcal{L}:j=b} f^p_{l,j\to i} = 0, \quad \forall b \in \mathcal{B},$$

$$q_b - \sum_{l\in\mathcal{L}:i=b} f^q_{l,i\to j} - \sum_{l\in\mathcal{L}:j=b} f^q_{l,j\to i} = 0, \quad \forall b \in \mathcal{B}.$$

How can we enforce power flow constraints in generated samples?

- Our goal is to minimize the data consistency loss $R_{\mathcal{H}}(\mathbf{x})$ on the clean data manifold $\mathcal{M}$:

$$\min_{\mathbf{x}\in\mathcal{M}} R_{\mathcal{H}}(\mathbf{x}),$$

where $\mathcal{H}(\cdot)$ encodes the equality constraints and

$$R_{\mathcal{H}}(\mathbf{x}) = \|\mathcal{H}(\mathbf{x})\|_2^2.$$

- We take one step of Riemannian gradient descent on $\mathcal{M}$:

$$\hat{\mathbf{x}}'_{0|t} = \hat{\mathbf{x}}_{0|t} - \tau_t \text{ grad } R_{\mathcal{H}}(\hat{\mathbf{x}}_{0|t}),$$

where

$$\text{grad } R_{\mathcal{H}}(\hat{\mathbf{x}}_{0|t}) = \mathcal{P}_{T_{\hat{\mathbf{x}}_{0|t}}\mathcal{M}}\left(\nabla_{\hat{\mathbf{x}}_{0|t}} R_{\mathcal{H}}(\hat{\mathbf{x}}_{0|t})\right).$$

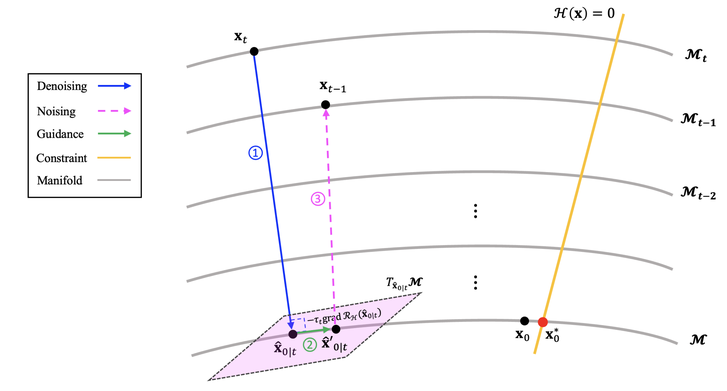
- Affine subspace assumption of clean data manifold $\mathcal{M}$:

$$\mathcal{P}_{T_{\hat{\mathbf{x}}_{0|t}}\mathcal{M}}\left(\nabla_{\hat{\mathbf{x}}_{0|t}} R_{\mathcal{H}}(\hat{\mathbf{x}}_{0|t})\right) \approx \nabla_{\hat{\mathbf{x}}_{0|t}} R_{\mathcal{H}}(\hat{\mathbf{x}}_{0|t}).$$

$$\hat{\mathbf{x}}'_{0|t} = \hat{\mathbf{x}}_{0|t} - \lambda_t \nabla_{\hat{\mathbf{x}}_{0|t}} R_{\mathcal{H}}(\hat{\mathbf{x}}_{0|t}).$$

Sampling steps can be characterized as transitions from $\mathcal{M}_i$ to $\mathcal{M}_{i-1}$:

- (1) we do a denoising step based on $\mathbf{x}_t$ and estimate the clean data $\hat{\mathbf{x}}_0$,
- (2) add the gradient guidance term,
- (3) add noise w.r.t. the corresponding noise schedule and obtain $\mathbf{x}_{t-1}$.

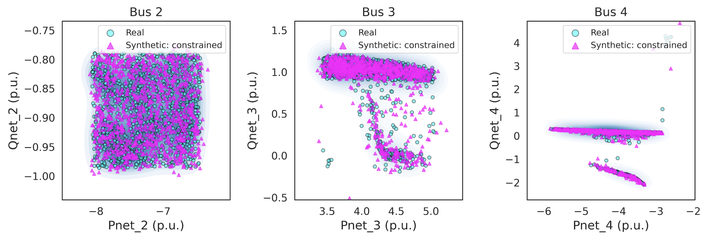


Geometry of sampling with guidance.

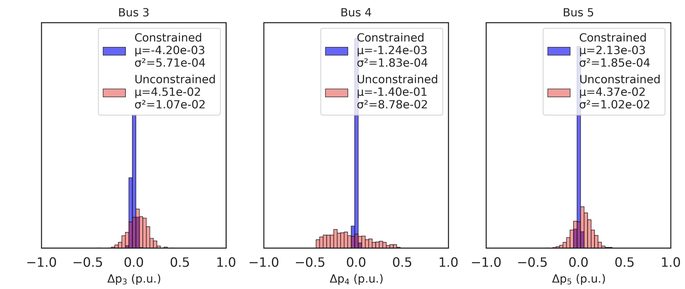
Results

Test System: PJM 5-BUS System

- Distribution Matching: joint distribution



- Histograms of violation magnitude for active power balance constraints



Conclusion

- Synthesized power flow data points effectively capture the pattern, domain, and modes of underlying distributions of the real power flow data.
- The proposed gradient guidance approach successfully enforces power flow constraints during sampling, ensuring the feasibility of the generated data.

miladh@umich.edu