INFORMS 2025

Synthesizing Power Flow Datasets via Constrained Diffusion Models

PhD Student: Milad Hoseinpour

Advisor: Prof. Vladimir Dvorkin

Department of Electrical Engineering and Computer Science

October 27, 2025

Outline

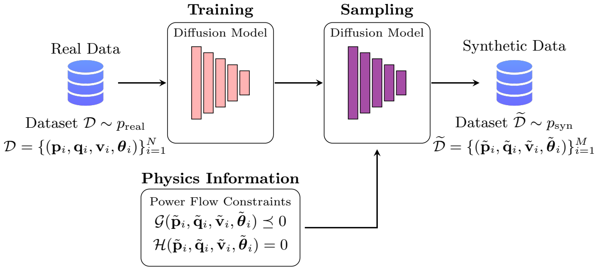
- Motivation and Goal
- Related Work
- Preliminaries
- Diffusion Guidance based on Power Flow Constraints
- Results
- Conclusion
- New Work Announcement

Motivation and Goal

- There is an increasing demand for large-scale and high-quality power flow datasets for machine learning (ML) tasks in power systems.
- Challenges in Data Availability: privacy and security concerns

Goal: generating synthetic power flow data

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



Ref.: Venzke, Andreas, Daniel K. Molzahn, and Spyros Chatzivasileiadis. "Efficient creation of datasets for data-driven power system applications." Electr. Power Syst. Res., 2021.

Related Work

(1) Generic random sampling approaches:

Iteratively perturbing system parameters (e.g., demand level) around the nominal value and solving the corresponding OPF problem.

- Drawback 1: The resulted datasets do NOT represent the true underlying distribution of real-world operating conditions.
- Drawback 2: The required number of random samples to cover the feasibility region of operating conditions grows exponentially.

Related Work

(1) Generic random sampling approaches:

Iteratively perturbing system parameters (e.g., demand level) around the nominal value and solving the corresponding OPF problem.

- Drawback 1: The resulted datasets do NOT represent the true underlying distribution of real-world operating conditions.
- Drawback 2: The required number of random samples to cover the feasibility region of operating conditions grows exponentially.

(2) Historical data-driven approaches: Generative models (e.g., VAE, GAN)

The historical data-driven approaches leverage real operational records to learn the data distribution.

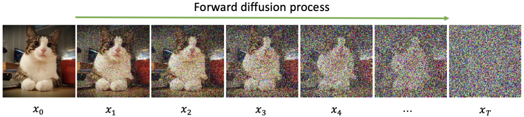
- Drawback 1: Poor generation quality (VAE), Mode Collapse (GAN)
- Drawback 2: No rigorous method to control their output (e.g., enforcing power flow constraints)

Ref.:

- S. Lovett et al., "OPFData: Large-scale datasets for AC optimal power flow with topological perturbations," arXiv preprint arXiv:2406.07234, 2024.
- A. Jabbar et al., "A survey on generative adversarial networks: Variants, applications, and training," ACM CSUR, vol. 54, no. 8, pp. 1–49, 2021.
- Z. Pan et al., "Data-driven EV load profiles generation using a variational auto-encoder," Energies, vol. 12, no. 5, p. 849, 2019.

Preliminary: Diffusion Models

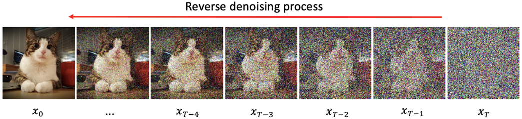
- Forward diffusion process that gradually adds noise to input



$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}\right)$$

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

- Reverse denoising process that learns to generate data by denoising



$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}\left(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \sigma_t^2 \mathbf{I}\right),$$

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

Training the Diffusion model

- The loss function to train the neural network ϵ_θ is

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

Algorithm 1 : Training the diffusion model

Inputs: initialized neural network $\boldsymbol{\epsilon}_\theta$, noise schedule $\{\alpha_t\}_{t=1}^{T}$, dataset of $\mathbf{x}_0$'s sampled from q_0
Outputs: trained neural network $\boldsymbol{\epsilon}_\theta$

1: **repeat**
2: $\quad \mathbf{x}_0 \sim q_0(\mathbf{x}_0)$
3: $\quad t \sim \text{Uniform}(\{1, \dots, T\})$
4: $\quad \epsilon \sim \mathcal{N}(0, \mathbf{I})$
5: $\quad$ Take gradient descent step on

$$\nabla_\theta \left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t\right)\right\|^2$$

6: **until** converged

Ref.: J. Ho et al., "Denoising diffusion probabilistic models," Adv. Neural Inf. Process. Syst., vol. 33, pp. 6840–6851, 2020.

Sampling the Diffusion Model

- First, predict the clean data $\hat{\mathbf{x}}_0$:

$$\hat{\mathbf{x}}_0(\mathbf{x}_t, t) = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(\mathbf{x}_t - \sqrt{1 - \bar{\alpha}_t}\, \epsilon_\theta(\mathbf{x}_t, t) \right),$$

$$\boxed{\mathbf{x}_{t-1} = \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} \hat{\mathbf{x}}_0 + \sigma_t z.}$$

Algorithm 2 : Sampling new data points

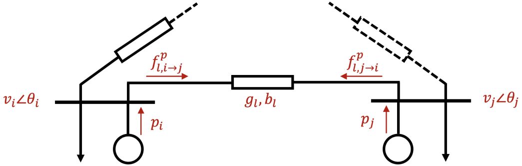
Inputs: trained neural network $\boldsymbol{\epsilon}_\theta$, noise schedule $\{\alpha_t\}_{t=1}^T$, noise scale σ_t

Outputs: new data point $\tilde{\mathbf{x}}_0$

1: $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\quad \hat{\mathbf{x}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}} \left(\mathbf{x}_t - \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right)$
4: $\quad \mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = 0$
5: $\quad \mathbf{x}_{t-1} \leftarrow \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} \mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} \hat{\mathbf{x}}_0 + \sigma_t \mathbf{z}$
6: **return** $\tilde{\mathbf{x}}_0$

Ref.: J. Ho et al., "Denoising diffusion probabilistic models," Adv. Neural Inf. Process. Syst., vol. 33, pp. 6840–6851, 2020.

Power Flow Equality Constraints



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

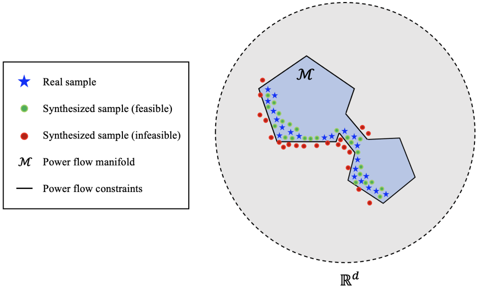
- The expressions for $f^{p}_{l,i\to j}$ and $f^{q}_{l,i\to j}$ are

$$f^{p}_{l,i\to j} = v_i v_j \left[g_l \cos(\theta_i - \theta_j) + b_l \sin(\theta_i - \theta_j)\right], \quad \forall l \in \mathcal{L},$$
$$f^{q}_{l,i\to j} = v_i v_j \left[g_l \sin(\theta_i - \theta_j) - b_l \cos(\theta_i - \theta_j)\right], \quad \forall l \in \mathcal{L},$$

where g_l and b_l are the real and imaginary parts of $Y = G + jB$.

Ref.: D. K. Molzahn et al., "A survey of relaxations and approximations of the power flow equations," Found. Trends Electr. Energy Syst., vol. 4, no. 1-2, pp. 1–221, 2019.

- **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.



How can we enforce power flow constraints in generated samples?

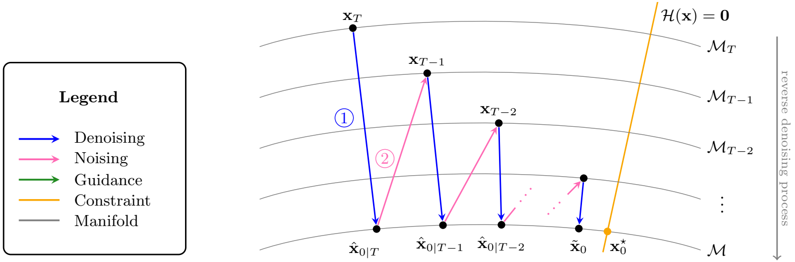
Ref.:

- Feng, Berthy T., Ricardo Baptista, and Katherine L. Bouman. "Neural approximate mirror maps for constrained diffusion models." arXiv preprint arXiv:2406.12816, 2024.

- G. Daras et al., "Consistent diffusion models: Mitigating sampling drift by learning to be consistent," Adv. Neural Inf. Process. Syst., vol. 36, pp. 42 038–42 063, 2023.

Geometry of Sampling without Guidance

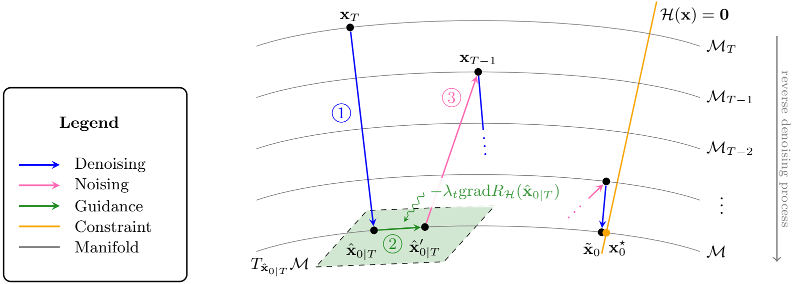
- Sampling steps can be characterized as transitions from $\mathcal{M}_i$ to $\mathcal{M}_{i-1}$:
  - (1) we do a denoising step based on $\mathbf{x}_t$ and estimate the clean data $\hat{\mathbf{x}}_0$,
  - (2) add noise w.r.t. the corresponding noise schedule and obtain $\mathbf{x}_{t-1}$.



The sampling trajectory is oblivious to the power flow constraints.

Geometry of Sampling with Guidance

- 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 $\mathbf{x}_{t-1}$.



The gradient guidance steers the sampling trajectory toward feasible power flow data points.

Gradient Guidance

Main Idea: The guidance term corresponds to a single iteration of Riemannian gradient descent on the clean data manifold.

- We minimize the data consistency loss $R_{\mathcal{H}}(\mathbf{x})$ on the learned data manifold $\mathcal{M}$:

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

  where

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

- We take one step of Riemannian gradient descent:

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

  where

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

Ref.:

- H. Chung et al., "Improving diffusion models for inverse problems using manifold constraints," Adv. Neural Inf. Process. Syst., vol. 35, pp. 25 683–25 696, 2022.

- N. Boumal, An introduction to optimization on smooth manifolds. Cambridge University Press, 2023.

Diffusion Guidance based on Power Flow Constraints

Gradient Guidance

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

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

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

Algorithm 4 : Sampling with gradient guidance

Inputs: trained neural network $\boldsymbol{\epsilon}_\theta$, noise schedule $\{\alpha_t\}_{t=1}^T$, noise scale σ_t, guidance scale λ_t

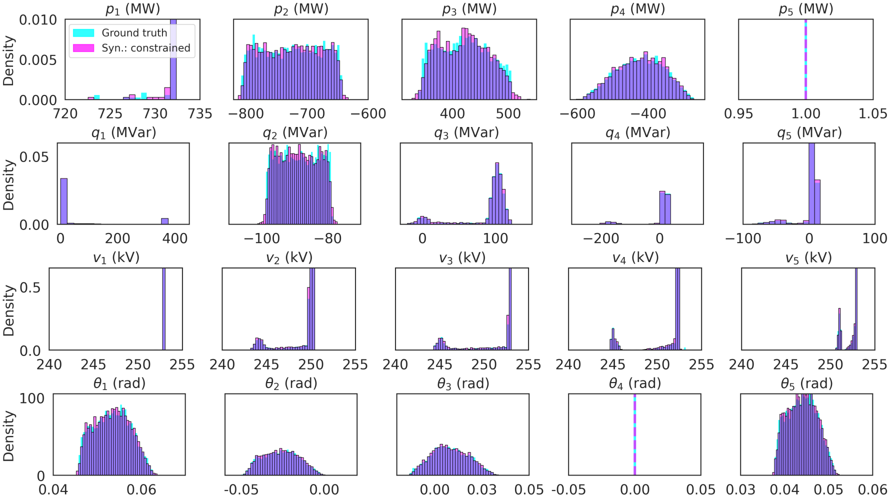
Outputs: new data point $\tilde{\mathbf{x}}_0$

1: $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I}_{4B})$
2: **for** $t = T-1$ to 0 **do**
3: $\quad\hat{\mathbf{x}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right)$
4: $\quad\hat{\mathbf{x}}'_0 \leftarrow \hat{\mathbf{x}}_0 - \lambda_t \nabla_{\mathbf{x}_t} R_{\mathcal{H}}(\hat{\mathbf{x}}_0)$
5: $\quad z \sim \mathcal{N}(0, \mathbf{I}_{4B})$
6: $\quad\mathbf{x}_{t-1} \leftarrow \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}'_0 + \sigma_t \mathbf{z}$
7: **return** $\tilde{\mathbf{x}}_0$

Results

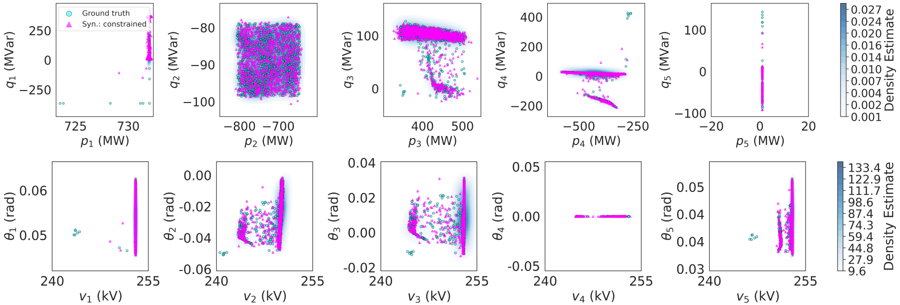
Statistical Similarity: marginal distribution

- Histograms of the ground truth versus synthetic power flow data points:



Statistical Similarity: joint distribution

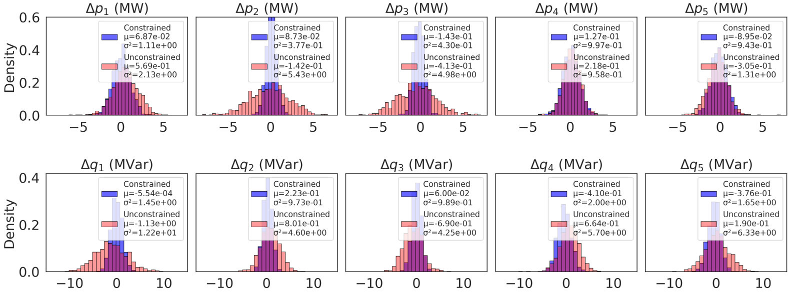
- 2D scatter plots with density estimates of $\mathbf{p}-\mathbf{q}$ and $\mathbf{v}-\boldsymbol{\theta}$:



- The synthetic data points
  - closely follow the distributional pattern of the real data.
  - closely span the entire domain of the real joint probability distributions.
  - captures the multi-modal structure of the real data distribution.

Constraint Satisfaction

- Histograms of violation magnitudes for the active and reactive power balance constraints in the **PJM 5-bus** system ($\lambda = 10^{-2}$ vs $\lambda = 0$):



- Wasserstein distances between the ground truth $\mathcal{D}$ and synthetic data $\widetilde{\mathcal{D}}$

Distance between...	5-Bus	24-Bus	118-Bus
...$\mathcal{D}$ and $\widetilde{\mathcal{D}}$ w/o guidance	0.442	0.607	0.622
...$\mathcal{D}$ and $\widetilde{\mathcal{D}}$ w/ guidance	0.382	0.585	0.597

Conclusion

- The synthesized power flow data points effectively capture both the marginal and joint 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.
- The gradient guidance mechanism maintains the sampling trajectory within the data manifold, preventing divergence from the learned data distribution.

All materials in this talk are based on our preprint:

Constrained Diffusion Models for Synthesizing Representative Power Flow Datasets
(available on arXiv)

Our latest work:

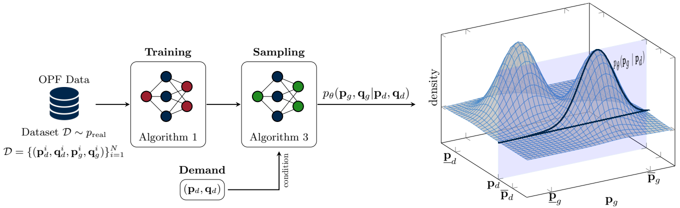
DiffOPF: Diffusion Solver for Optimal Power Flow
(available on arXiv)

- We saw how to learn the underlying distribution of OPF data and generate synthetic OPF data points. But, can we use this learned distribution as a prior and solve new OPF problems?

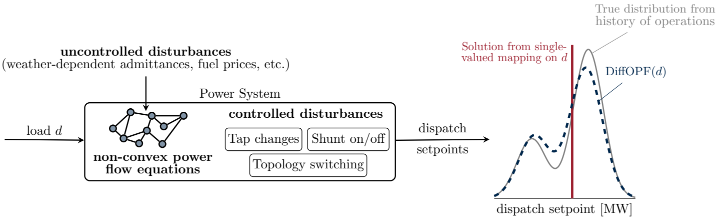
Goal: solving OPF problem via conditional sampling

Given a learned distribution of OPF data points, we aim to sample from this distribution conditioned on a new demand input.

- We introduce a diffusion-based OPF solver, termed DiffOPF, that treats OPF as a conditional sampling problem.



- The OPF problem is a multi-valued (non-unique) mapping from loads to dispatch setpoints.



Thank You!

Google Scholar Profile