Management Innovation, and Practice Journal
ISSN 3065-1417
Volume 11 Issue 3
July-September 2023
Impact Factor: 6.85
Enquiry: contact@continentalpub.online
Published by Continental Publication | https://cspjournals.org/index.php/Management

A WASSERSTEIN-GRADIENT FLOW APPROACH TO ENHANCING POWER FLOW DATASET QUALITY

Wang Hui Fang
China Electric Power Research Institute, Haidian District, Beijing, China

Abstract: *The application of artificial intelligence (AI) methods in power grid analysis necessitates the utilization of power flow datasets for model training. Presently, power flow data sources predominantly stem from offline simulations and real-time data collection. However, the accumulated online and offline power flow datasets have limitations that impede their direct suitability for AI model training. Online power flow data, collected during actual grid operations, offers a substantial volume of sample data. Nevertheless, this data distribution lacks uniformity and contains numerous redundant samples, falling short of the comprehensive coverage and clear boundaries required for effective analysis. On the other hand, offline power flow data, characterized by extreme operational scenarios, is manually curated and often situated at the stable boundaries of grid operations. While it possesses strong sample typicality and clear boundaries, the dataset's volume is limited and fails to represent the full spectrum of typical working conditions in grid operations. Addressing this challenge involves supplementing datasets to align with the distribution characteristics of offline analysis data. By doing so, the resulting dataset can fulfill both comprehensive coverage and clear boundary requirements. However, the methodologies for dataset adjustment that consider distribution characteristics remain underexplored, hindering the full exploitation of offline analysis data's distribution traits. This study delves into the development of advanced dataset adjustment methods that consider distribution characteristics. It aims to bridge the gap between online and offline power flow data, enabling the creation of comprehensive and boundary-clear datasets suitable for AI-driven power grid analysis. The proposed approach not only enhances the efficacy of AI methods in grid analysis but also offers a unique perspective on utilizing distribution characteristics in dataset adjustment. By addressing this gap in research, we contribute to the improved applicability of AI techniques in power grid analysis, optimizing grid performance and reliability.*

Keywords: *Artificial Intelligence (AI), Power Grid Analysis, Power Flow Datasets, Dataset Adjustment Methods, Distribution Characteristics*

1. Introduction

AI methods applied to power grid analysis require training of power flow datasets. The existing sources of power flow data are mainly generated by offline simulation and online data collection, but both the online and offline power flow datasets accumulated in the past cannot directly meet the requirements. The power flow data for online analysis is the actual operation mode collected, which constitutes a large amount of sample data, but the distribution is not uniform and there are many similar samples, which cannot meet the requirements of covering comprehensively and clear boundary; the power flow data for offline analysis is the extreme operation mode manually adjusted, which constitutes a strong sample typicality and is distributed at the stable boundary of the grid operation, which helps to achieve the requirement of clear boundary, but the data volume is small and it is difficult to cover all the typical working conditions of the grid operation, which cannot meet the requirement of covering comprehensively. If the dataset is supplemented by targeting the distribution characteristics of the data for offline analysis, the obtained dataset will satisfy the two requirements mentioned above. Since the research on data set adjustment methods considering distribution characteristics is relatively weak, it is difficult to take full advantage of the distribution characteristics of the data for offline analysis.

Optimal transport theory is the study of the relationship between distributions and distributions. Gradient flow based on optimal transport theory is an important tool in applied mathematics for constructing dynamic models in feature spaces [1], gradient flow has been extensively studied in the context of metric spaces [2] and has been found to be deeply related to partial differential equations (PDEs) [3].

In view of this, we study a power flow dataset supplementation method considering the distribution characteristics, which transforms the power flow dataset data into Wasserstein space in the form of distribution, then transforms the power flow dataset adjustment problem into the problem of solving the extreme value of the energy functional by constructing the functional, then solves the curve evolution equation by using the variational method, and finally solves the evolution equation to obtain a set of power flow dataset series labeled by process time. This paper is organized as follows: Section 2 presents the relevant technical background, including optimal transmission theory and gradient flow; Section 3 introduces the Wasserstein-gradient flow based power flow dataset supplementation method. Section 4 verifies the effectiveness of the method by testing it in the power flow dataset of the CEPRI 36 node power grid model.

2. Technical Background

2.1. Optimal Transport and the Optimal Transport Dataset Distance

Optimal transport theory is the study of the problem of interconversion between distributions, where the optimal transport distance (also known as the Wasserstein distance) is a quantitative tool to describe the degree of variation between distributions. For two subsets of measures $\alpha,\beta \in \mathcal{P}(\mathcal{X})$ and the transport cost function $c: \mathcal{X}\times\mathcal{X} \to \mathbb{R}_+$, the optimal transport problem is

$$\mathrm{OT}_c(\alpha,\beta) \triangleq \min_{\pi\in\Pi(\alpha,\beta)} \int c(\boldsymbol{x}_1,\boldsymbol{x}_2)\mathrm{d}\pi(\boldsymbol{x}_1,\boldsymbol{x}_2), \quad (1)$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are features from the samples in the two measures, and $\Pi(\alpha,\beta)$ is the set of transport schemes between α and β, i.e., the coupling with these two measures as marginal measures:

$$\Pi(\alpha,\beta) \triangleq \{\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{X}) \mid P_{1\#}\pi = \alpha, P_{2\#}\pi = \beta\}. \quad (2)$$

Where for p ≥ 1, $c(x,y) = d(x,y)^p$ $\mathrm{W}_p = \mathrm{OT}_c(\alpha,\beta)^{1/p}$ is called the p-Wasserstein distance. As the name suggests, W_p defines a true distance on $\mathcal{P}(\mathcal{X})$ [4]. Thus, with the former as the distance configuration $\mathbb{W}_p(\mathcal{X}) = (\mathcal{P}(\mathcal{X}), \mathrm{W}_p)$ is the metric space, called the (p-) Wasserstein space. In practice, the solution method is often solved by a regularized version of Eq. (1) with an additional entropy term $\lambda \mathrm{H}(\pi)$ [5].

The dual formula of the Kantorovich problem is

$$\mathrm{OT}_c(\alpha,\beta) = \sup_{\varphi\in\mathcal{C}(\mathcal{X})} \int_{\mathcal{X}} \varphi \mathrm{d}\alpha + \int_{\mathcal{X}} \varphi^c \mathrm{d}\beta, \quad (3)$$

where $\varphi: \mathcal{X} \to \mathbb{R}$ is called the Kantorovich potential function and φ^c is its c-conjugate: $\varphi^c(x) = \inf_{x'\in\mathcal{X}} c(x,x') - \varphi(x)$. For $c(x,x') = \| x - x' \|^2$, ϕ^c is the Fenchel conjugate.

In the literature [6] it was demonstrated that there is also a dynamic formula for OT:

$$\mathrm{W}_p^p(\alpha,\beta)=\min_{\mu_t,V_t}\int_0^1\int_{\mathcal{X}}\|V_t(x)\|^p\mathrm{d}\mu_t(x)\mathrm{d}t, \tag{4}$$

where the minimum is taken from the measure-domain pair satisfying $\mu_0=\alpha,\mu_1=\beta$ and the continuity equation:

$$\partial_t\mu_t=-\nabla\cdot(\mu_t V_t). \tag{5}$$

This formulation corresponds to finding the shortest path satisfying the conservation of the mass constraint in the metric path μ_t from α to β and the velocity field V_t, even if the path length is the smallest (formally the integral of the metric derivative). Thus, in contrast to the global correspondence (Via γ) in the static formulation (Eq. (1)), the dynamic formulation focuses on the local transport (via μ_t).

It is appealing to use OT to define a distance between datasets, but this is non-trivial for labeled datasets. The main issue is that problem (1) would require an elementwise metric d, which for labeled datasets means defining a distance between pairs of feature-label pairs. For the general case where $\mathcal{Y}$ might be a discrete set (i.e., classification), this seems daunting. In recent work, researchers [7] propose a hybrid metric on this joint space that relies on representing the labels as distributions over features α_y. E.g., for a digit classification dataset, α_1 would be a distribution over images with label $y=1$.

With this, they define a metric on $\mathcal{Z}$ as $d_{\mathcal{Z}}(z,z')=d_{\mathcal{X}}(x,x')^p+\mathrm{W}_p^p(\alpha_y,\alpha_{y'})$. Using $d_{\mathcal{Z}}$ as the ground cost in eq. (1) yields a distance between measures on $\mathcal{P}(\mathcal{Z})$, and therefore between datasets, which they refer to as the Optimal Transport Dataset Distance (OTDD):

$$\mathrm{OTDD}(\mathrm{D}_\alpha,\mathrm{D}_\beta)\triangleq\left(\min_{\pi\in\Pi(\alpha,\beta)}\int_{\mathcal{Z}\times\mathcal{Z}}d_{\mathcal{Z}}(z,z')\mathrm{d}\pi(z,z')\right)^{\frac{1}{2}}. \tag{6}$$

The main appeal of this distance is that it is defined even if the label sets of the two data sets are nonoverlapping, or if there is no explicit known correspondence between them (e.g., digits to letters). It achieves this through a purely geometric treatment of features and labels. Another advantage is its computational scalability, which relies on using a Gaussian approximation on the per-label distributions, i.e., modeling each α_y as $\mathcal{N}(\mu_y,\Sigma_y)$, whose mean and covariance are estimated from samples. In that case, the $\mathrm{W}_2^2(\alpha_y,\alpha_{y'})$ distances can be computed in closed form, so no optimization is needed to evaluate $d_{\mathcal{Z}}(z,z')$ inside problem (6).

2.2. *Grandient Flows*

Consider a functional $F:\mathcal{X}\to\mathbb{R}$ and a point $x_0\in\mathcal{X}$. A gradient flow is an absolutely continuous curve $x(t)$ that evolves from x_0 in the direction of steepest descent of F. When $\mathcal{X}$ is Hilbertian and F is sufficiently smooth, its gradient flow can be succinctly expressed as the solution of a differential equation $x'(t)=-\nabla F(x(t))$ with initial condition $x(0)=x_0$. Different discretizations of this equation yield popular gradient descent schemes, such as momentum and acceleration [8]

3. Wasserstein-Gradient Flow Based Sample Replenishment Method for Power Flow Datasets

The power flow dataset data are transformed into Wasserstein space, and then the power flow dataset adjustment problem is transformed into the problem of solving the extreme value of the energy generalization function by constructing the energy generalization function, and then the curve evolution equation is obtained by using the variational method, and finally the evolution equation is solved to obtain a set of power flow dataset series labeled by process time. The distribution difference between this serial dataset and the target distribution dataset gradually decreases with the increment of the time principal scale, and finally an adjusted dataset with controllable distribution difference is obtained.

The main problem that needs to be solved for a specific implementation is how to choose the objective functional.

3.1. *Functional Minimization via Gradient Flows*

Given a dataset objective expressed as a functional $F : \mathcal{P}(\mathcal{Z}) \to \mathbb{R}$, we seek a joint measure $\rho \in \mathcal{P}(\mathcal{Z})$ realizing:

$$\min_{\rho \in \mathcal{P}(\mathcal{Z})} F(\rho) \quad (7)$$

We propose to approach this problem via gradient flows, i.e., by moving along a curve of steepest descent starting at ρ_0 until reaching a solution ρ^*. Unlike Euclidean settings, here the underlying space $\mathbb{W}_p(\mathcal{Z})$ is infinite-dimensional and non-Hilbertian, thus requiring stronger tools.

First, the notion of derivative can be extended to functionals on measures through the first variation, denoted by $\frac{\delta F}{\delta \rho}$. With this, we characterize the gradient flow $(\rho_t)_{t \geq 0}$ of F as the solution of:

$$\partial_t \rho_t = \nabla_{\mathrm{W}} F(\rho_t) \triangleq \nabla \cdot \left(\rho_t \nabla \frac{\delta F}{\delta \rho}(\rho_t)\right), \quad (8)$$

which can also be seen as a continuity equation (4) for the measure ρ_t and the velocity field $-\nabla \frac{\delta F}{\delta \rho}(\rho_t)$.

Our main functional of interest will be the Wasserstein distance to a target distribution: $\mathcal{T}_\beta(\rho) \triangleq \mathrm{W}_2(\rho, \beta)$, which we realize using the OTDD (Section 2.1).

Hence, we assume the objective of interest can be cast as:

$F(\rho) = \mathcal{T}_\beta(\rho)$

The numerical solution of the functional can be found in the literature [9].

4. Experimental Validation

4.1. *Example Introduction*

The samples in the power flow dataset of this paper describe various modes of operation of the grid model CEPRI36, and the grid structure is shown in Figure 1, where some nodes are connected to capacitors or reactors that are not involved in regulation, and there are 18 nodes of generating units or loads involved in regulation, with the nodes injecting power as the input feature values, for a total of 36 variables, i.e., the sample contains a feature dimension of 36 dimensions.

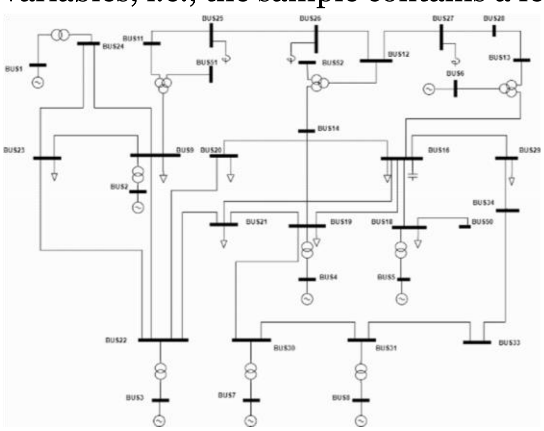

Figure 1: CEPRI36 grid model topology connection diagram

For sample supplementation of the target distribution dataset using a Wasserstein gradient flow method. Among them, the target distribution dataset uses 5000 manually generated samples with distribution characteristics similar to those of the power flow dataset for offline analysis, whose samples are mainly distributed near the stability boundary. The initial dataset for the sample adjustment generation process is chosen from the randomly generated dataset.

For this purpose, the experimental design is as follows:

1) The original random dataset is denoted as D_A, the target distribution dataset is denoted as D_B, and then four randomly generated data sets are denoted as D_{a_i}, where i=1,2,3,4.

2) Using the four data sets D_{a_i} as the initial data set and D_B as the target data set, a gradient flow operation is performed to select the appropriate four data sets according to OTDD, denoted as D_{b_i}, where i=1,2,3,4, and there is a correspondence with i in D_{a_i}.

(3) The generated new datasets are then merged into the original dataset separately to form two datasets with increasing sample capacity and maintaining the original distribution characteristics, denoted as D_{A_i} and D_{B_i}, where i=1,2,3,4 and have correspondence with i in D_{a_i}. The formation can be expressed as follows:

$$D_{A_i}(i) = D_A + \sum_{j=1}^{i} D_{a_j} \qquad (9)$$

$$D_{B_i}(i) = D_B + \sum_{j=1}^{i} D_{b_j} \qquad (10)$$

It should be noted that the "+" operator here does not indicate the operation of a set, but the direct merging of data sets. The sample sizes of D_{A_1}, D_{A_2}, D_{A_3} and D_{A_4} are 10,000, 15,000, 20,000, and 25,000, respectively. Similarly, the data set sequence D_{B_i} also has the same sample size. The D_{B_i} is the data set of the target distribution after supplementation.

The experimental hardware environment is 3.30 GHz, the CPU is AMD Ryzen9 5900HS, and the GPU is RTX-3060. in the Wasserstein gradient descent flow procedure in part 1 of the experiment, the optimal transmission distance of the power flow dataset is computed with the help of solvers for the optimal transmission distance provided by the geomloss [10] and POT [11] libraries, and the above Both libraries have the option of CUDA acceleration, which accelerates the solution of the Wasserstein distance using GPU parallel computing. One of them is the Compute Unified Device Architecture (CUDA), a computing platform introduced by NVIDIA, a graphics card manufacturer.

4.2. *Results and Discussion*

The effect of the power flow dataset supplementation method is analyzed using the optimal transport distance calculation method for power flow datasets given in Section 2.1. Comparing the distribution differences between the four randomly sampled datasets D_{a_i} used as initial values and the four target distribution datasets D_{b_i} generated by the method in this paper, the $d_{\mathrm{OT}}(\mathcal{D}_1,\mathcal{D}_2)$ between the two datasets is found, where $\mathcal{D}_1$ takes the value of the 1st column and $\mathcal{D}_2$ takes the value of the 1st row, the result is shown in Table 1 and Table 2 as follows:

Table 1: d_{OT} values between D_{a_i}

$d_{\mathrm{OT}}(\mathcal{D}_1,\mathcal{D}_2)$	D_{a_1}	D_{a_2}	D_{a_3}	D_{a_4}
D_{a_1}	0	1.59	1.57	1.93
D_{a_2}	1.59	0	1.76	1.48
D_{a_3}	1.57	1.76	0	1.61

D_{a_4}	1.93	1.48	1.61	0.

Table 2: d_{OT} values between D_{b_i}

$d_{\mathrm{OT}}(\mathcal{D}_1,\mathcal{D}_2)$	D_{b_1}	D_{b_2}	D_{b_3}	D_{b_4}
D_{b_1}	0	0.58	0.64	0.63
D_{b_2}	0.58	0	0.67	0.66
D_{b_3}	0.64	0.67	0	0.78
D_{b_4}	0.63	0.66	0.78	0.

Where D_{a_i} is also at the same level as D_{a_i}, D_{b_i} and with D_{b_i} d_{OT} . Based on the above results, it can be seen that:

1) It is logical that d_{OT} the between the initial randomly sampled distributed datasets of the motion is larger than the d_{OT} between the generated datasets, whose distribution properties dictate that the samples will appear randomly in a smaller range. This is also a side verification that the Wasserstein gradient flow method generates indeed datasets with the target distribution.

2) The values of d_{b_i} between two d_{OT} are at the same order of magnitude level, and there are no values that are significantly smaller than others and converge to zero. This phenomenon reflects the significance of initial dataset selection in Wasserstein gradient flow, setting different initial datasets, and the datasets of the final generated target distribution will not be exactly the same, still maintaining the same distribution but the data are not duplicated.

5. Conclusions

In order to take full advantage of the distribution characteristics of the power flow data for offline analysis and adjust the dataset flexibly and efficiently, this paper investigates the method of adjusting the power flow dataset considering the distribution characteristics. The Wasserstein gradient flow-based sample supplementation method for power flow datasets is proposed to convert the dataset generation process into a generalized optimization problem of finding extrema, and our goal is to obtain the complete motion trajectory of the dataset under the gradient flow. The motion trajectory can provide a sequence of datasets with progressively decreasing variance from the target dataset distribution, in which we can select the datasets with the appropriate degree of variance to add to the original data set as needed, where the initial value of the evolution equation also has an important influence on this process. This operation also enables a sample supplementation method that maintains the distribution properties, i.e., the supplemented samples still maintain the same or similar distribution properties but are not simple duplicates of the data in the original dataset. Finally, the effectiveness of the Wasserstein gradient flow method is verified by experimental examples.

Acknowledgements

This work was supported by State Grid Corporation of China Science & Technology Project - Research on rapid exploration and learning method of key information in high proportion new energy power system - under Grant 5100-202155429A-0-0-00.

References

Ambrosio L, Gigli N, Savare G (2005). Gradient flows in metric spaces and in the Wasserstein space of probability measures [Z]. Lectures in Mathematics, ETH Zurich, Birkhäuser.

Santambrogio F (2017). {Euclidean, metric, and Wasserstein} gradient flows: an overview [J]. Bulletin of Mathematical Sciences, 7(1): 87-154.

Jordan R, Kinderlehrer D, Otto F (1998). The variational formulation of the Fokker–Planck equation [J]. SIAM journal on mathematical analysis, 29(1): 1-17.

Villani C (2009). Optimal transport: old and new: Vol. 338 [M]. Springer.

Cuturi M (2013). Sinkhorn distances: Lightspeed computation of optimal transport[C]. Advances in neural information processing systems: Vol. 26. Curran Associates, Inc.

Benamou J D, Brenier Y (2000). A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem [J]. Numerische Mathematik, 84(3): 375-393.

Alvarez-Melis D, Fusi N (2020). Geometric Dataset Distances via Optimal Transport [J]. arXiv: 2002. 02923 [cs, stat], 2020.

Wilson A C, Recht B, Jordan M I (2016). A Lyapunov analysis of momentum methods in optimization [J]. arXiv preprint arXiv:1611.02635.

Alvarez-Melis D, Fusi N (2021). Dataset dynamics via gradient flows in probability space[C]. International conference on machine learning. 219-230.

Feydy J, Sejourne T, Vlalard F X, et al (2019). Interpolating between optimal transport and mmd using sinkhorn divergences [C]. The 22nd International Conference on Artificial Intelligence and Statistics. PMLR, 2681-2690.

Flamary R, Courty N, Gramfort A, et al (2021). POT: Python Optimal Transport [J]. Journal of Machine Learning Research, 22(78): 1-8.