Efficient Sliced Wasserstein Distance Computation Via Adaptive Bayesian Optimization

Manish Acharya [1,2] David Hyde [1,2]

[1]Department of Computer Science, Vanderbilt University [2]Institute for Software Integrated Systems, Vanderbilt University

Motivation

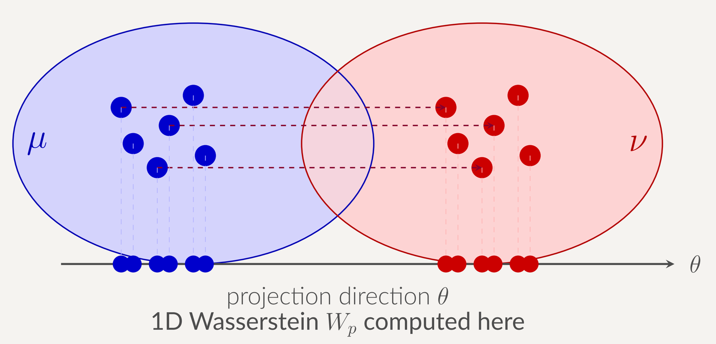
Optimal Transport provides a principled way to compare probability distributions, but computing the Wasserstein distance in high dimensions is computationally expensive.

The Sliced Wasserstein (SW) distance improves scalability by projecting distributions onto 1D directions and averaging the resulting Wasserstein distances.

However, the accuracy of SW estimates depends critically on the choice of projection directions.

This limitation directly affects applications in generative modeling, computer vision, and large scale data analysis, where accurate and efficient distance computation is essential.

Sliced Wasserstein Distance



The Sliced Wasserstein distance computes optimal transport by projecting high-dimensional distributions onto 1D directions. For a projection direction $\theta \in S^{d-1}$, the projection is

$$T_\theta(x) = \theta^\top x$$

The SW distance averages the Wasserstein distance between projected distributions:

$$SW_p^p(\mu,\nu) = \mathbb{E}_{\theta\sim U(S^{d-1})}\left[W_p^p(\theta_\#\mu, \theta_\#\nu)\right]$$

In practice, the expectation is approximated using L projection directions:

$$SW_p^p(\mu,\nu) \approx \frac{1}{L}\sum_{\ell=1}^{L} W_p^p((\theta_\ell)_\#\mu, (\theta_\ell)_\#\nu)$$

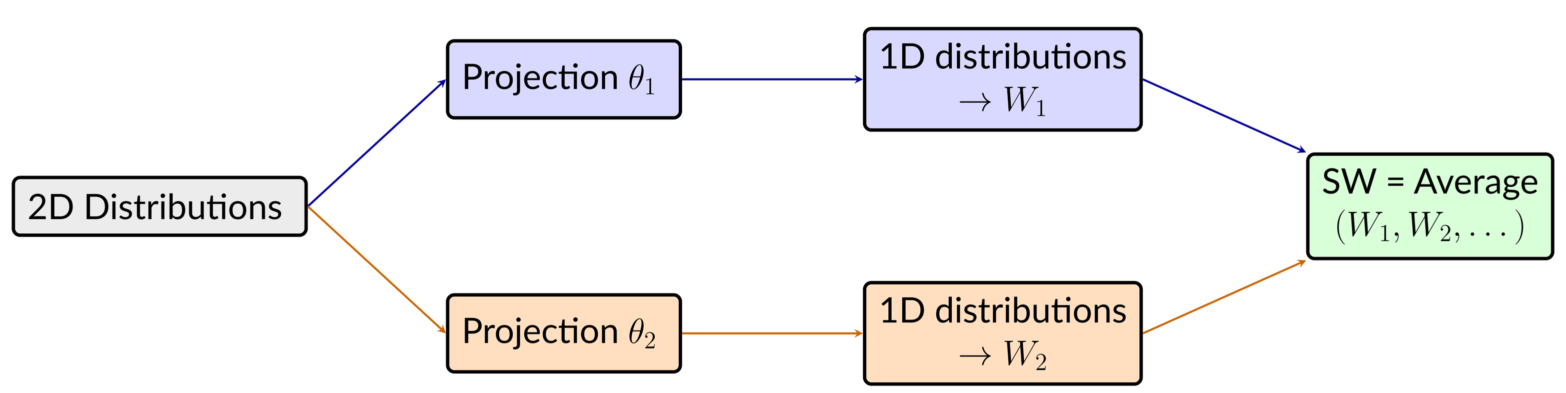


Figure 1. Sliced Wasserstein projects high-dimensional distributions onto multiple 1D directions and averages the resulting Wasserstein distances.

Key Idea: Learning Informative Directions

Existing SW methods select projection directions using data-independent strategies such as Monte Carlo or quasi–Monte Carlo, which aim to uniformly cover the sphere.

However, projection directions contribute unequally to the sliced Wasserstein distance.

Our key insight is that slice evaluations reveal information about the projection landscape. We therefore use Bayesian Optimization (BO) to adaptively select informative directions.

Method

For a projection direction $\theta \in S^{d-1}$, we define the slice cost

$$f(\theta;\mu,\nu) = W_p^p(\theta_\#\mu, \theta_\#\nu).$$

Instead of selecting directions uniformly, we treat $f(\theta)$ as a black-box function on the sphere and use Bayesian Optimization (BO) to identify informative directions. A Gaussian Process (GP) surrogate models the projection landscape, and an acquisition function selects the next directions to evaluate.

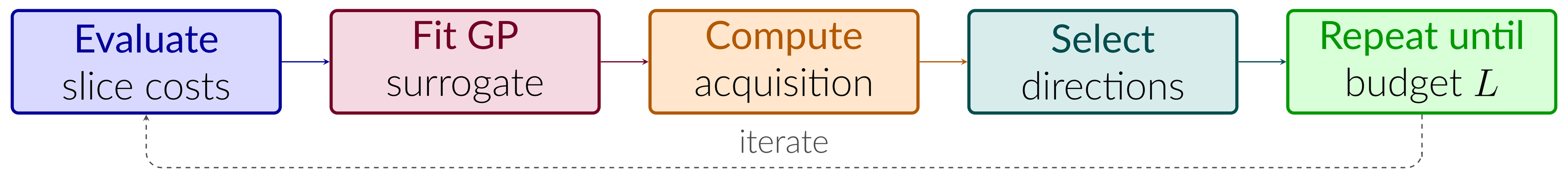


Figure 2. Bayesian Optimization pipeline for selecting informative projection directions in Sliced Wasserstein estimation.

- BOSW: one-shot BO learns a fixed set of projection directions.
- RBOSW: periodically refreshes directions during optimization.
- ABOSW: starts from a strong QSW set and applies lightweight BO refinement.
- ARBOSW: periodically restarts from QSW seeds and refines with BO.

We model the projection landscape using a GP with an angular RBF kernel:

$$k(\theta,\theta') = \exp\left(-\frac{d_S(\theta,\theta')^2}{2\ell^2}\right), \qquad d_S(\theta,\theta') = \arccos\langle\theta,\theta'\rangle.$$

Experiments & Results

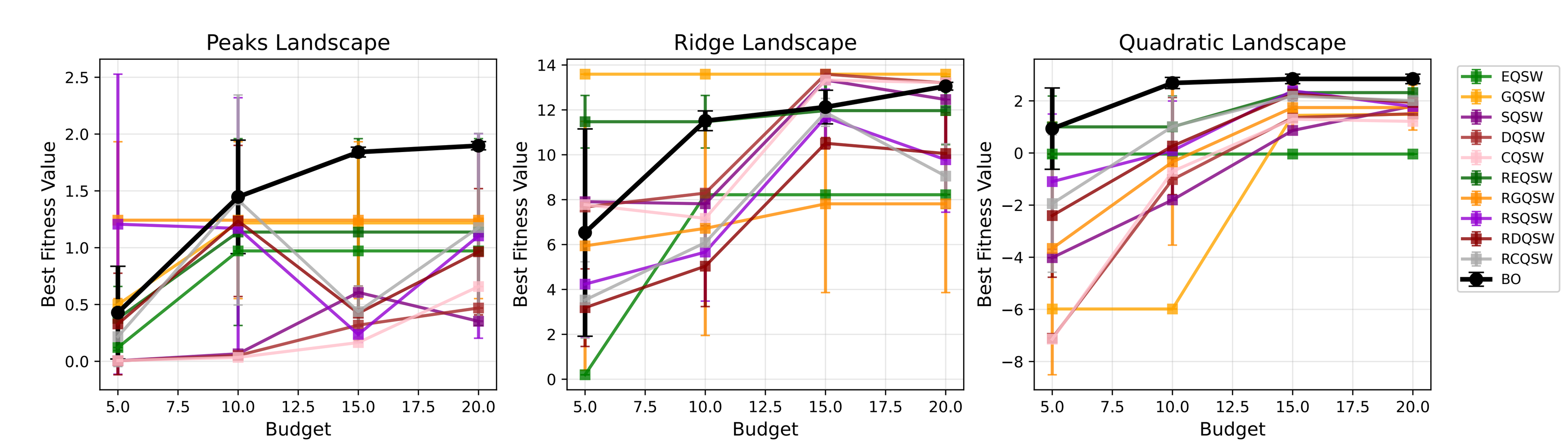


Figure 3. Projection landscape experiment. On synthetic projection landscapes, BO rapidly identifies high-value directions, confirming that adaptive slice selection can exploit feedback in ways fixed QSW designs cannot.

Approximation Error:

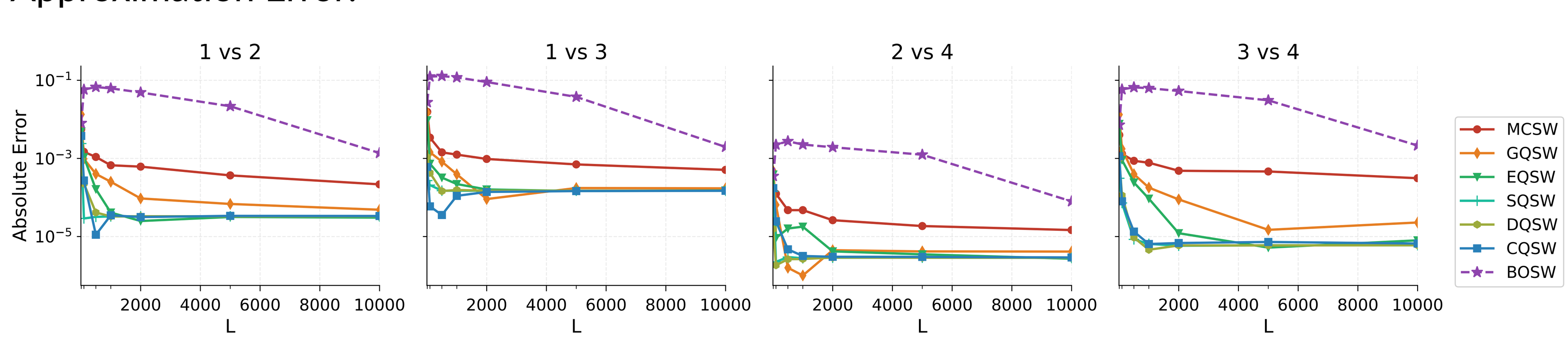


Figure 4. QSW variants outperform BOSW. This is expected because the benchmark rewards uniform sphere coverage, whereas BOSW learns task-adapted directions rather than an unbiased quadrature rule.

Experiments & Results

Image Style Transfer:

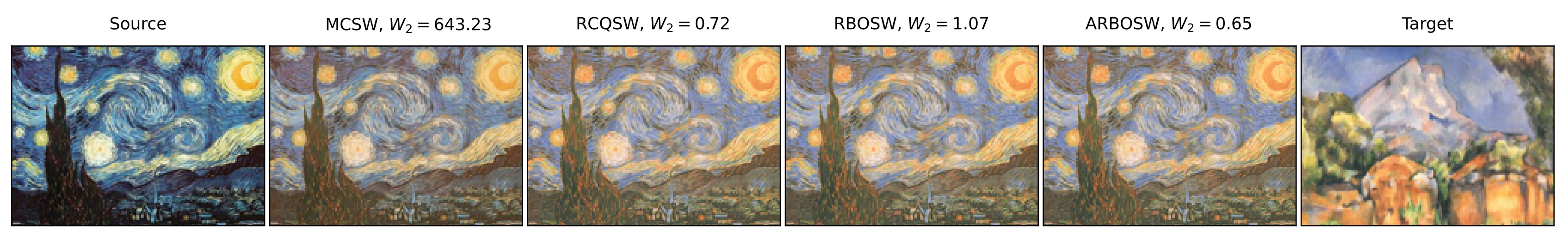
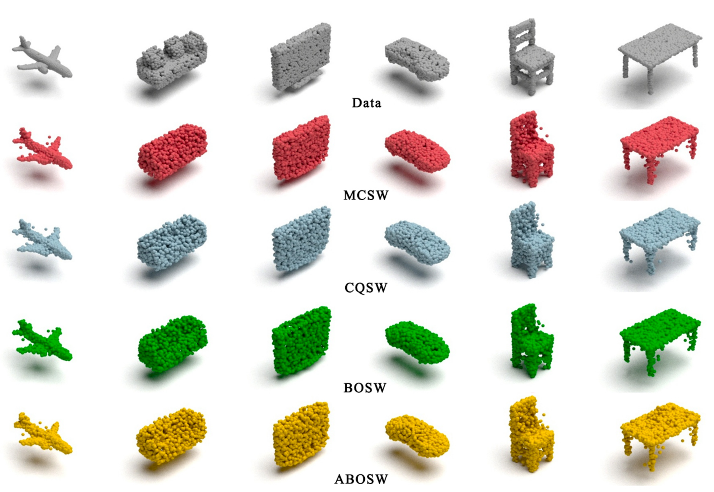


Figure 5. Color transfer results using Sliced Wasserstein variants.

Deep Point-Cloud Autoencoder:



Method	Final Loss ↓	Time ↓
SW	0.052	1.00x
CQSW	0.041	1.10x
BOSW (ours)	0.039	1.05x
ABOSW (ours)	0.036	1.07x
RBOSW (ours)	0.044	1.32x
ARBOSW (ours)	0.042	1.18x

Figure 6. Point-cloud autoencoder reconstructions using different projection strategies.

Learning projection directions via Bayesian Optimization enables faster convergence and also competitive accuracy compared to fixed sampling methods.

Conclusion & Future Work

- We introduce BO-based projection selection for Sliced Wasserstein estimation.
- Projection selection can be framed as a learning problem rather than a sampling problem.
- While BO is not ideal for unbiased quadrature, it improves performance when SW is used within optimization tasks.
- Hybrid methods such as ARBOSW and ABOSW achieve strong performance across tasks.

Future Work:

- Extending BO-based projection selection to higher-dimensional settings where recent work suggests BO can avoid the curse of dimensionality.
- Integrating adaptive methods into generative models and optimal transport pipelines.

Author Contributions & Acknowledgments

Prof. David Hyde proposed the use of Bayesian optimization as the core research direction and provided guidance on manuscript writing. Manish Acharya designed the algorithm, implemented the methodology, conducted experiments, analyzed results, and wrote manuscript. This research was supported by the SyBBURE Searle Undergraduate Research Fellowship.

References

[1] N. Bonneel, J. Rabin, G. Peyré, H. Pfister. *Sliced and Radon Wasserstein Barycenters of Measures*. JMIV, 2015.

[2] K. Nguyen, H. Janati, F. Bach, M. Cuturi. *Quasi-Monte Carlo for 3D Sliced Wasserstein*. ICML, 2023.