Paper

ITSPACE: Monotone Gaussian Optimal Transport Updates

Woojoo Na, Jennifer Dy

Northeastern University

Motivation

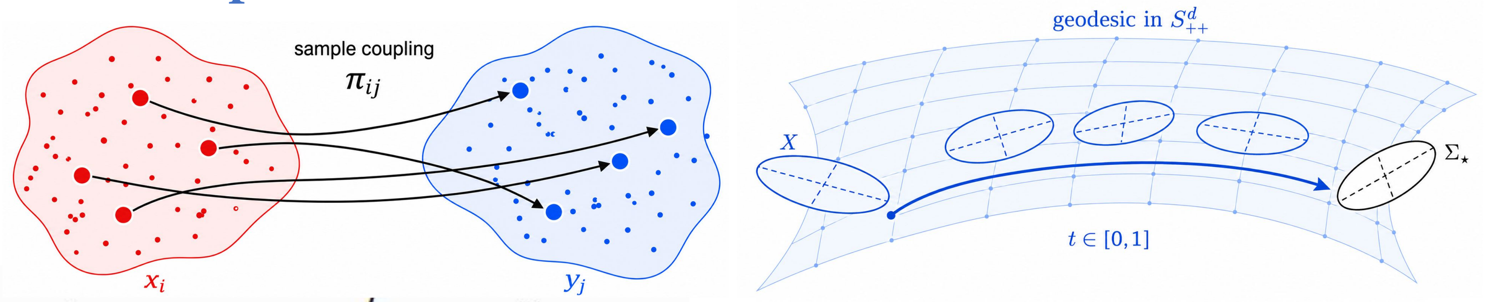
- In ML, we often compare diverse probability distributions, and optimal transport maps are used to align them, either at the empirical sample level or through higher-level statistics such as covariances.
- Covariance matrices serve as compact descriptors of feature distributions in many machine-learning pipelines, including domain adaptation and Gaussian embeddings.
- Under a centered Gaussian approximation, the unregularized Wasserstein–2 optimal-transport (OT) discrepancy admits a closed form on covariances given by the Bures–Wasserstein (BW) objective on the symmetric positive definite (SPD) cone.
- However, computing OT problems on non-Euclidean surfaces is computationally expensive.

Contributions

- We introduce ITSPACE (Iterative Transport for Stable Proximal Alignment of Covariance Embeddings), a rank budgeted, few step covariance alignment algorithm.
- Theoretical Guarantees: Exact polar certificates yield sufficient decrease of the true BW objective; inexact polar computation is controlled by an explicit certificate-gap bound.
- Empirical Support:
  - Across Camelyon17, VisDA-2017, and Terra/CCT-20, ITSPACE rapidly contracts the exact BW gap under the same rank and step budget.
  - In downstream tasks, ITSPACE updates recover downstream performance inside a frozen feature/classifier pipeline in few (1~2) steps.

Sample-level OT vs Gaussian OT at a glance

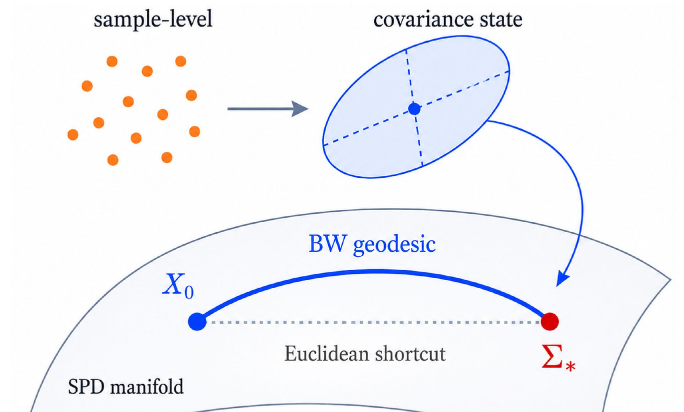
Sample-level OT / Gaussian OT



$W_2^2(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \int \|x-y\|_2^2 \, d\pi(x,y)$

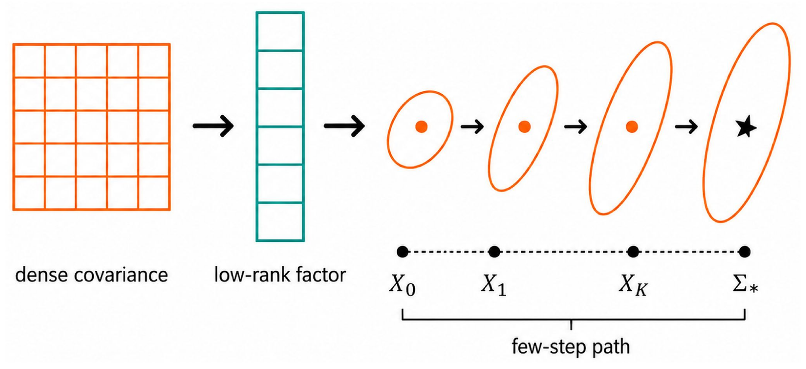
$W_2^2(X,\Sigma_\star) = \mathrm{tr}(X) + \mathrm{tr}(\Sigma_\star) - 2\,\mathrm{tr}\left[(\Sigma_\star^{1/2} X \Sigma_\star^{1/2})^{1/2}\right]$

Figure: A sample-level OT (left) moves pointwise, while a Gaussian OT (right) can move higher statistics, such as covariances



Why Bures-Wasserstein (BW) Geometry?

- For covariance matrices, BW gives the exact Gaussian Wasserstein-2 objective; sample-level transport solves a different empirical problem with popular methods like Sinkhorn.
- Covariances live on the PSD/SPD cone, so transport should respect the underlying BW geometry.



Scalability: Low Rank, Few Steps

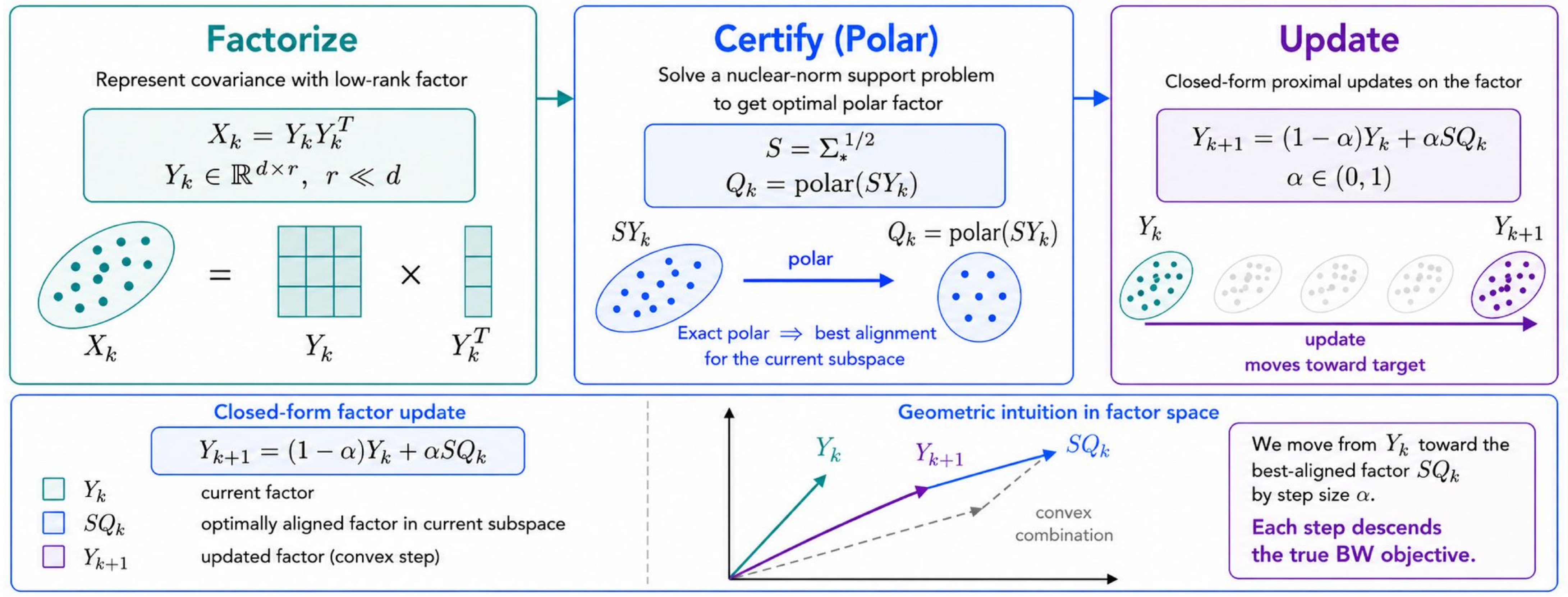
- Full-rank covariance transport is expensive in high-dimensional feature spaces; practical ML modules often maintain only a compact low-rank covariance state.
- ITSPACE updates the factor directly, giving valid PSD, rank-constrained intermediate covariances with certified descent under a small step budget.

Our Idea:

Instead of replacing the current covariance by a full dense target, update the **low-rank covariance factor directly**. This keeps every intermediate covariance positive semidefinite, rank-constrained, and descending for the exact BW objective.

Algorithm

ITSPACE: Factorize, Certify, Update



Theory supports empirical results

- Gaussian OT gives the Bures–Wasserstein geometry on covariances, but generic dense or projected updates do not inherently preserve a valid low-rank covariance trajectory.
- ITSPACE updates the square-root factor itself: PSD/rank validity is built in, and the polar certificate gives sufficient descent for the exact BW objective evaluated in our experiments.

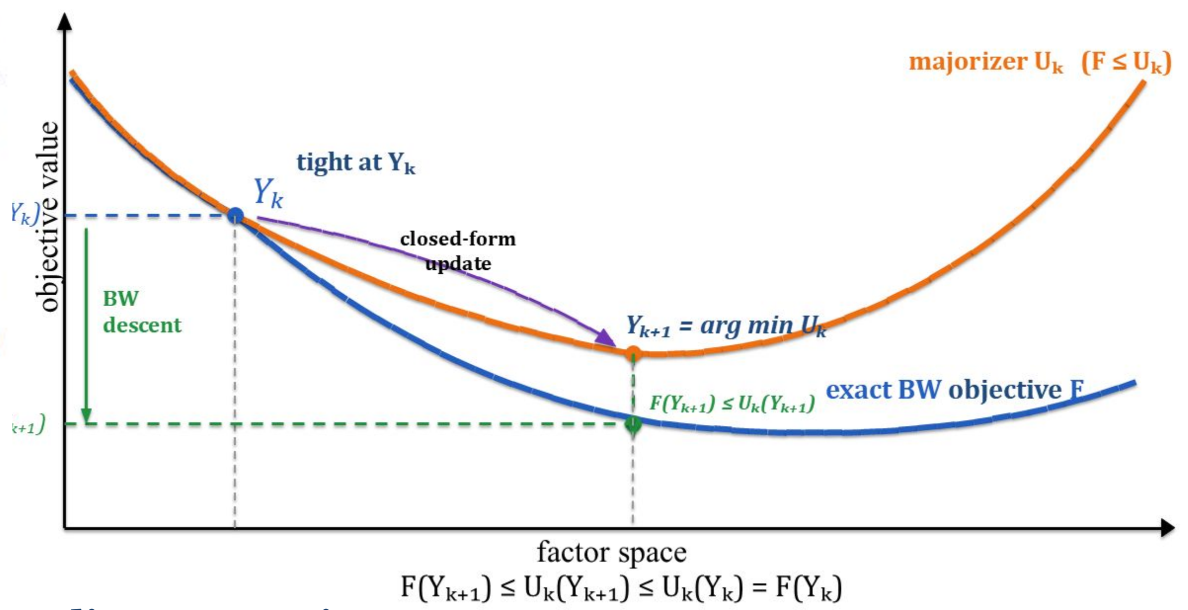
Theoretical Results

Theorem 4.2 (Sufficient BW Descent). *Assume that $Q_k = \mathrm{polar}(SY_k)$ is an exact maximizer of the nuclear-norm support problem for SY_k. Let Y_{k+1} be the* ITSPACE *update in Eq. (12). Then*

$$F(Y_k) - F(Y_{k+1}) \geq \frac{1}{\alpha}\|Y_{k+1} - Y_k\|_F^2 = \alpha\|B_k - Y_k\|_F^2. \tag{14}$$

Consequently,

$$W_2^2(X_{k+1},\Sigma_\star) \leq W_2^2(X_k,\Sigma_\star) \quad \text{for all } k.$$



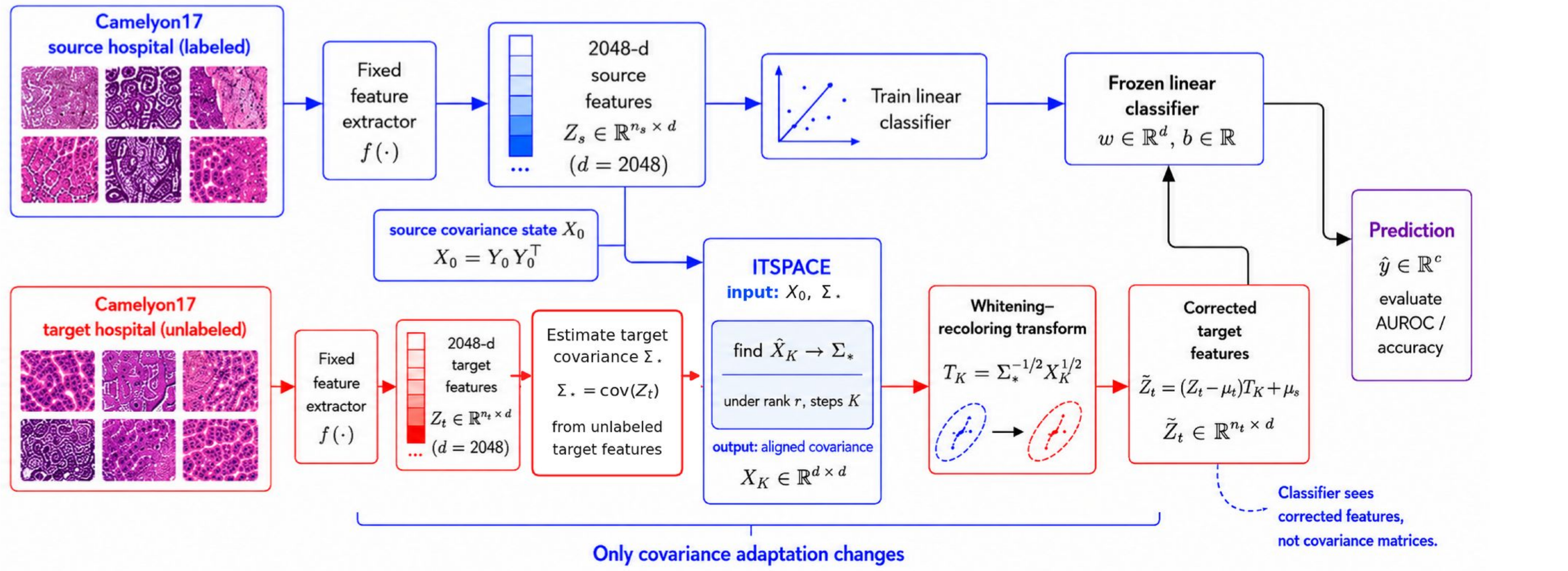
$F(Y_{k+1}) \leq U_k(Y_{k+1}) \leq U_k(Y_k) = F(Y_k)$

Baseline comparisons

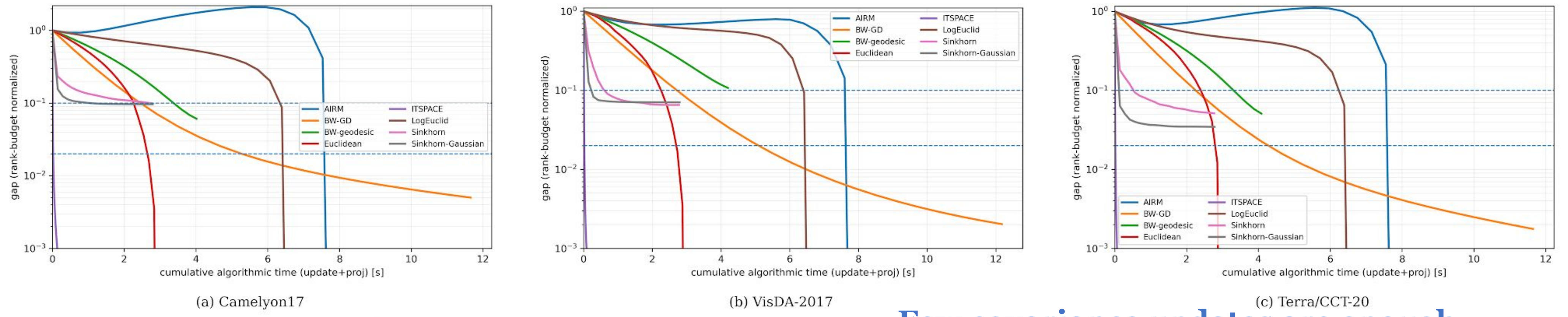
Method	Objective / update rule	Targets Gaussian W_2^2?	Monotone $W_2^2(\cdot,\Sigma_\star)$?	Cost / step
ITSPACE	Proximal MM updates for BW in a square-root factorization; polar/Procrustes certificate + damping	Yes	Yes†	$\mathcal{O}(d^2\tau + dr^2 + r^3)$
BW geodesic	Closed-form BW geodesic interpolation toward $\Sigma_\star$ (displacement interpolation), then rank-r truncation in rank-budget runs	Yes	Yes (in geodesic parameter)	$\mathcal{O}(d^3) + \mathrm{proj}(d,r)$
BW-GD (direct BW optimizer)	BW-targeting gradient/Riemannian updates with optional backtracking, then rank-r truncation in rank-budget runs	Yes	Not in general*	$\mathcal{O}(d^3) + \mathrm{proj}(d,r)$
Euclidean / Frobenius	Ambient-space interpolation/update, then rank-r truncation in rank-budget runs	No	No	$\mathcal{O}(d^2) + \mathrm{proj}(d,r)$
Log–Euclidean	Log-domain geodesic / Log–Euclidean objective (full-matrix functions), then truncation	No	No	$\mathcal{O}(d^3) + \mathrm{proj}(d,r)$
AIRM	Affine-invariant metric geodesic / AIRM objective (full-matrix functions), then truncation	No	No	$\mathcal{O}(d^3) + \mathrm{proj}(d,r)$
CORAL	Whitening/re-coloring (second-moment match); used as a one-shot endpoint reference in downstream	No†	N/A	$\mathcal{O}(d^3) + \mathrm{proj}(d,r)$
Sinkhorn (entropic OT)	Regularized OT between samples (biased objective)	No (regularized)	No	$\mathcal{O}(Tnm)$
Sinkhorn–Gaussian	Regularized OT on Gaussian samples/embeddings (biased objective)	No (regularized)	No	$\mathcal{O}(Tnm)$

Experiments

ML Application pipeline for Domain Adaptation



Rank budgeted BW contraction in few steps



Downstream task performance

CovDrift-MR downstream on Camelyon17 under strict step budgets. Mean±std over three seeds. t_5: adaptation time at K=5 seconds (shared preprocessing excluded). Settings: r=16, drift rank 16, $s_{\max}$=1.70.

Method	K=1	K=2	K=5	K=20	t_5
No adapt	70.22 ±2.17	70.22 ±2.17	70.22 ±2.17	70.22 ±2.17	—
ITSPACE	73.94 ±1.79	74.01 ±1.78	74.01 ±1.78	74.01 ±1.78	0.035 ±0.014
BW-geodesic	71.92 ±2.09	72.86 ±1.97	73.54 ±1.92	74.01 ±1.78	0.015 ±0.001
BW-GD	73.97 ±1.78	73.93 ±1.80	73.97 ±1.77	74.01 ±1.78	0.037 ±0.004
Euclidean	72.11 ±2.10	72.11 ±2.10	72.11 ±2.10	74.01 ±1.78	0.021 ±0.006
Log-Euclidean	70.56 ±1.89	70.91 ±1.95	71.92 ±2.09	74.01 ±1.78	0.013 ±0.002
AIRM	70.56 ±1.89	70.91 ±1.95	71.92 ±2.09	74.01 ±1.78	0.013 ±0.002
Sinkhorn	72.67 ±2.45	72.67 ±2.45	72.67 ±2.45	72.67 ±2.45	0.233 ±0.014
Sinkhorn-Gaussian	72.88 ±2.92	72.88 ±2.92	72.88 ±2.92	72.88 ±2.92	0.164 ±0.006

Few covariance updates are enough

Use **ITSPACE** where the pipeline maintains a covariance state: update the covariance factor, then feed the resulting whitening–recoloring map back into the fixed representation pipeline.

In **Experiment I**, across Camelyon17, VisDA-2017, and Terra/CCT-20, the same few-step update rapidly contracts the exact BW gap under the shared rank budget.

In **Experiment II**, the downstream pipeline is deliberately frozen: fixed 2048-d features, fixed linear head, and only covariance adaptation changes.

On downstream tasks, ITSPACE reaches the endpoint performance band after 1–2 updates, suggesting that Gaussian OT can be used as a practical inner-loop covariance transport primitive.

Optimal transport beyond samples:
Align the statistical representation your model uses.

Takeaways

- Move OT from samples to the statistical representations used inside ML pipelines.
- ITSPACE makes Gaussian/covariance OT a plug-in primitive: closed-form updates, valid low-rank iterates, and certified BW descent.
- This gives a lightweight, objective-matched alternative to sample OT, dense BW optimization, or projected SPD updates under strict rank and step budgets.

Acknowledgements

This work was supported by grant NIH/NHLBI 5R01HL167072.

ML Lab: mllabneu.github.io