Improving Function Space Flow Matching with Kernel Optimal Transport

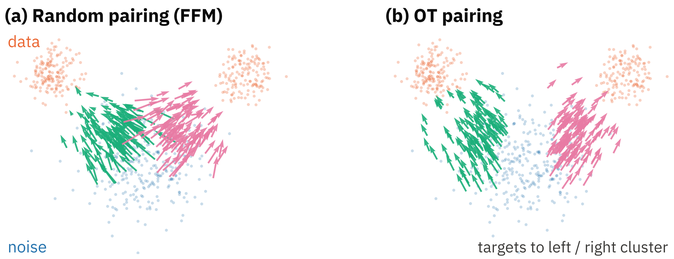
Fred Xu[1,*], Thomas Markovich[2], Barbora Barancikova[3], Yizhou Sun[1]

[1]UCLA [2]Block, Inc. [3]Imperial College London [*]Work done in part during an internship at Block, Inc.

Motivation

- Time series and PDE solutions are *functions*, with distributions on infinite-dimensional spaces. **FFM** (Kerrigan et al.) does flow matching in Hilbert space with Gaussian bridges and a neural operator, but pairs noise and data **independently**

Random Pairs Give Conflicting Targets



- The network regresses $f_1 - f_0$, the difference of two unrelated draws; OT pairing fixes this in $\mathbb{R}^d$ (Tong et al.; Pooladian et al.)
- **Hard in function space**: no densities (no Lebesgue measure), restrictive Cameron–Martin conditions, geometry-blind L^2 costs

Background

- **FFM bridge**: prior $\mu_0 = \mathcal{N}(0, C_0)$, $\mu_t^f = \mathcal{N}(tf, \sigma_t^2 C_0)$, $\sigma_t = 1 - (1 - \sigma_{\min})t$; the network regresses the bridge velocity v_t^f
- **Kernel cost**: feature map ϕ into the RKHS $\mathcal{H}_\kappa$

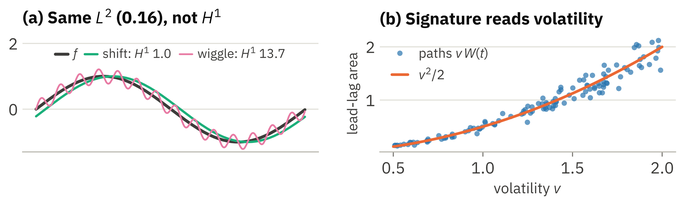
$$c_\kappa(x, y) = \|\phi(x) - \phi(y)\|^2_{\mathcal{H}_\kappa} = \kappa(x, x) + \kappa(y, y) - 2\kappa(x, y)$$

- **Hilbert Sinkhorn Divergence** (Li et al., 2021): debiased entropic OT under c_κ,

$$\mathrm{OT}_{\kappa,\epsilon}(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int c_\kappa \, d\pi + \epsilon \, \mathrm{KL}(\pi \,\|\, \mu \otimes \nu),$$

$$S_{\kappa,\epsilon} = \mathrm{OT}_{\kappa,\epsilon}(\mu, \nu) - \tfrac{1}{2}\mathrm{OT}_{\kappa,\epsilon}(\mu, \mu) - \tfrac{1}{2}\mathrm{OT}_{\kappa,\epsilon}(\nu, \nu)$$

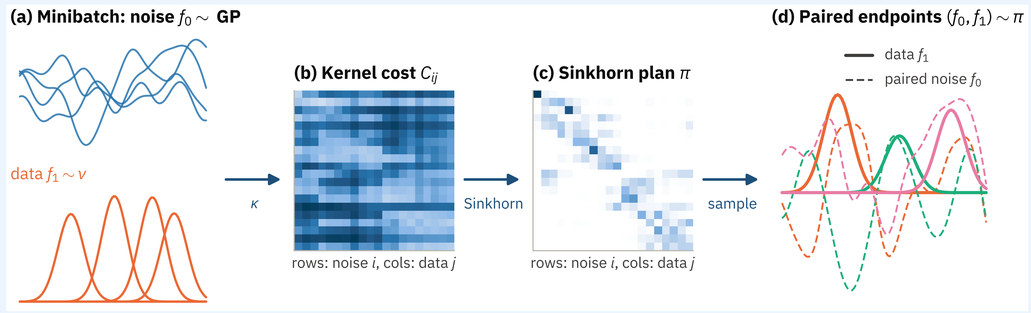
Kernels Inject Function-Space Geometry



- **RBF on L^2 or H^k**: $\kappa(f, g) = e^{-\|f-g\|^2/2\sigma^2}$, bounded ($c_\kappa \le 2$) and characteristic; the Sobolev norm sees derivatives
- **Signature**: $\kappa_{\mathrm{sig}}(x, y) = \langle S(x), S(y) \rangle$, $S^{(m)}(x) = \int_{t_1 < \cdots < t_m} dx_{t_1} \otimes \cdots \otimes dx_{t_m}$, universal on compact sets of time-augmented paths

kFFM: Pair Before You Flow

TL;DR: kFFM replaces FFM's independent endpoint pairing with entropic OT under a kernel-induced cost, the coupling of the Hilbert Sinkhorn Divergence, solved on every minibatch. The neural operator, the loss and the sampler are unchanged



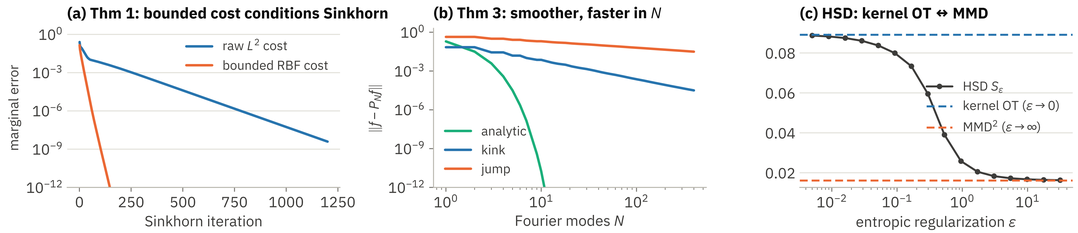
Per minibatch: median-normalized cost, log-domain Sinkhorn with $\epsilon = 0.05$, pairs sampled from the plan

$$C_{ij} = \kappa(f_0^{(i)}, f_0^{(i)}) + \kappa(f_1^{(j)}, f_1^{(j)}) - 2\kappa(f_0^{(i)}, f_1^{(j)}), \qquad (f_0, f_1) \sim \pi = \mathrm{Sinkhorn}(C, \epsilon)$$

$$g_t = t f_1 + \sigma_t f_0, \qquad \mathcal{L}(\theta) = \| f_1 - (1 - \sigma_{\min}) f_0 - v_\theta(g_t, t) \|^2$$

The kernel enters only the cost: no sample is mapped into the RKHS, and the independent plan $\pi = \mu_0 \otimes \nu$ recovers FFM exactly. The expected minibatch plan is a valid coupling for every batch size, so CFM consistency still holds

Theory: Certifying the Kernel Coupling



- **Thm 1, well-posed on Banach spaces**: if $\kappa(f, f) \le B$, then $0 \le c_\kappa \le 4B$ and $|S_{\kappa,\epsilon}(\mu, \nu)| \le 8B$ for *all* μ, ν. The Gibbs kernel $e^{-c_\kappa/\epsilon}$ stays away from 0 and 1, so Sinkhorn is well conditioned ($\approx 8\times$ faster than L^2 at $b = 512$), and no inner product is needed, which licenses the signature kernel on path space
- **Thm 2, error decomposition**: on a compact metric space $(\mathcal{X}, d)$ of diameter M,

$$\left|\mathrm{OT}_{\kappa,\epsilon}(\mu, \nu) - \mathcal{W}(\mu, \nu)\right| \le \Delta_\kappa + 2\epsilon \log \mathcal{N}(\mathcal{X}, \delta) + 8M\delta$$

  The kernel mismatch $\Delta_\kappa = \sup |c_\kappa - d^2|$ is irreducible: small for the RBF cost at large bandwidth, large by design for the signature kernel (rank correlation 0.33–0.37 with L^2), where the gains are largest
- **Thm 3, discretization invariance**: Sobolev-RBF kernel, μ, ν on the ball $B_R^{H^\alpha} \subset H^k$, N Fourier modes:

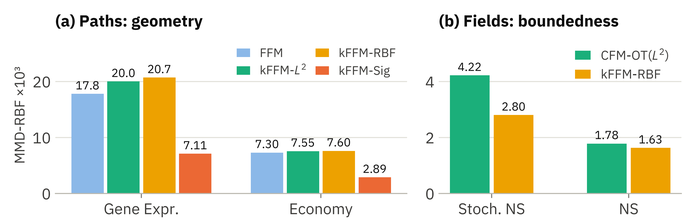
  $\left|\mathrm{KOT}_\epsilon(\mu, \nu) - \mathrm{KOT}_\epsilon(\mu_N, \nu_N)\right| \le \frac{4R^2}{\sigma^2} N^{-2(\alpha-k)/d}$
- **Scope**: the theory certifies the surrogate coupling; we make no straightness or sampling-efficiency claim, and the sample-quality gains are empirical

Distributional Quality (MMD-RBF $\times 10^3$)

	Dataset	Strongest baseline		kFFM	p
Paths	Gene Expr.	FFM	17.8	**7.11**	<0.001
	Economy	FFM	7.30	**2.89**	<0.001
	AEMET	CFM-OT(L^2)	−4.11	−5.10	0.015
	Heston	CFM-OT(L^2)	−0.08	−0.13	0.021
PDEs	KdV	CFM-OT(L^2)	−1.32	−2.56	0.039
	Stoch. KdV	DDPM	0.38	**0.10**	0.037
	NS	CFM-OT(L^2)	1.78	**1.63**	<0.001
	Stoch. NS	CFM-OT(L^2)	4.22	**2.80**	<0.001

- **Best on all 8** vs. FFM, DDPM, DDO/NCSN, GANO, CFM-OT(L^2); config picked by held-out sliced-W
- p: seed-paired t-test vs. FFM; on Heston and NS rows the base also changes
- **Non-kernel**: sliced-W and marginal W_1 improve 18–20% (Gene Expr., Economy) and 2.9–4.1× (NS, Stoch. NS)

Bounded Cost or Function-Space Geometry?



- **Paths**: no Euclidean cost recovers the signature gains (66–105× lower autocorrelation error than CFM-OT)
- **Fields**: same ranking as L^2, outliers compressed: better conditioning, equal or better MMD at the same base

Turbulent Navier–Stokes ($\nu = 10^{-5}$)

Method	MMD $\times 10^3$	Sliced-W	Vort. W_1	Enstr. W_1
kFFM	**1.48**	**0.193**	**0.038**	**0.155**
CFM-OT(L^2)	**1.47**	0.194	0.041	0.157
FFM (GP)	108	0.601	0.107	0.201
DDO/NCSN	569	1.39	1.22	1.07
DDPM	380	1.18	0.848	0.889

- OT-paired flow matching beats score and diffusion baselines 23–32× on vorticity W_1
- Gain over FFM is mostly the white-noise base; there, kernel and L^2 costs tie

Cost and Limitations

- **Cheap**: 5–7% overhead at 128^2. **Limits**: theory weakest on rough data; signature memory grows with length