# Functional Analytic Learning Unit

Ha Quang Minh

関数解析的学習ユニット ハクァンミン

RIKEN

Center for Advanced Intelligence Project


## Main Research Directions

1) Mathematical theory and structures coming from/related to **Infinite-dimensional Information Geometry** and **Optimal Transport**, in particular in the setting of **reproducing kernel Hilbert spaces** (RKHS), **infinite-dimensional Gaussian measures**, and **Gaussian processes**.
2) Theory of RKHS and related methodologies in machine learning and statistics
3) Novel algorithms and applications in functional data analysis, computer vision, signal processing, brain imaging, and brain computer interfaces.

## Current research focus: Optimal Transport in Statistics and Machine Learning

- Our current focus is on Optimal Transport in the infinite-dimensional setting, in particular in the reproducing kernel Hilbert space (RKHS) and stochastic process settings
- Our recent results show analytically that the entropic regularization formulation possess many favorable theoretical properties over the exact optimal transport formulation, such as dimension-independent convergence for empirical estimations.
- In the setting of Gaussian measures on Hilbert space and Gaussian processes, many quantities of interest admit explicit formulas.

## Optimal Transport and Entropic Regularization

**Optimal Transport distances between probability measures**

- $(X,d)$ = complete separable metric space, $c: X \times X \to \mathbb{R}_{\geq 0}$ = lower semi-continuous *cost function* (e.g. $X = \mathbb{R}^n$, $c(x,y) = ||x-y||^2$, $\mathcal{P}(X)$ = set of all probability measures on $X$.
- The *optimal transport (OT)* problem between two probability measures $\nu_0, \nu_1 \in \mathcal{P}(X)$ is

$$\mathrm{OT}_c(\nu_0,\nu_1) = \min_{\gamma \in \mathrm{Joint}(\nu_0,\nu_1)} \mathbb{E}_\gamma[c] = \min_{\gamma \in \mathrm{Joint}(\nu_0,\nu_1)} \int_{X\times X} c(x,y)d\gamma(x,y)$$

**Entropic regularization of optimal transport**

- Exact optimal transport distances are generally computationally demanding
- Exact Wasserstein distance $W_p(\nu_0,\nu_1) = \mathrm{OT}_{d^p}(\nu_0,\nu_1)^{1/p}$ can have high sample complexity, with worst case being exponential $O(n^{-1/d})$ in $\mathbb{R}^d$
- Entropic regularization problem can be solved efficiently using Sinkhorn algorithm

$$\mathrm{OT}_c^\epsilon(\mu,\nu) = \min_{\gamma\in\mathrm{Joint}(\mu,\nu)} \{\mathbb{E}_\gamma[c] + \epsilon \mathrm{KL}(\gamma || \mu\otimes\nu)\}, \quad \epsilon > 0$$

- Sinkhorn divergence $\mathrm{S}_p^\epsilon(\mu,\nu) = \mathrm{OT}^\epsilon_{d^p}(\mu,\nu) - \frac{1}{2}[\mathrm{OT}^\epsilon_{d^p}(\mu,\mu) + \mathrm{OT}^\epsilon_{d^p}(\nu,\nu)]$, with $\mathrm{S}_p^\epsilon(\mu,\mu) = 0$

## Entropic regularization of 2-Wasserstein distance - a sample of recent results

**Theorem 1 (Sinkhorn divergence between Gaussian measures on Hilbert space).** *Let $X = \mathcal{H}$ separable Hilbert space, $c(x,y) = ||x-y||^2$. Let $\mu_0 = \mathcal{N}(m_0,C_0)$,$\mu_1 = \mathcal{N}(m_1,C_1)$. For each fixed $\epsilon > 0$,*

$$\mathrm{S}_2^\epsilon(\mu_0,\mu_1) = ||m_0 - m_1||^2 + \frac{\epsilon}{4}\mathrm{tr}\,[M_{00}^\epsilon - 2M_{01}^\epsilon + M_{11}^\epsilon] + \frac{\epsilon}{4}\log\left[\frac{\det\left(I + \frac{1}{2}M_{01}^\epsilon\right)^2}{\det\left(I+\frac{1}{2}M_{00}^\epsilon\right)\det\left(I+\frac{1}{2}M_{11}^\epsilon\right)}\right]$$

*Here $M_{ij}^\epsilon = -I + \left(I + \frac{16}{\epsilon^2}C_i^{1/2}C_jC_i^{1/2}\right)^{1/2}$ is a trace class operator, $\det$ is the Fredholm determinant, and $\lim_{\epsilon\to 0} S_2^\epsilon(\mu_0,\mu_1) = W_2^2(\mu_0,\mu_1) = ||m_0-m_1||^2 + \mathrm{tr}[C_0 + C_1 - 2(C_0^{1/2}C_1C_0^{1/2})^{1/2}]$, $\lim_{\epsilon\to\infty} S_2^\epsilon(\mu_0,\mu_1) = ||m_0-m_1||^2$ ($\mathrm{S}_2^\epsilon$ provides an explicit interpolation between 2-Wasserstein distance and MMD)*

**Theorem 2 (Dimension-independent Estimation of Sinkhorn divergence between Gaussian processes from finite covariance matrices - bounded kernels).** *Assume $\sup_{x\in T} K^i(x,x) \leq \kappa_i^2$ Let $\mathbf{X} = (x_i)_{i=1}^m$ be independently sampled from $(T,\nu)$. For any $0<\delta<1$, with probability at least $1-\delta$,*

$$\left|\mathrm{S}_2^\epsilon\left[\mathcal{N}\left(0,\frac{1}{m}K^1[\mathbf{X}]\right), \mathcal{N}\left(0,\frac{1}{m}K^2[\mathbf{X}]\right) - \mathrm{S}_2^\epsilon[\mathcal{N}(0,C_{K^1}),\mathcal{N}(0,C_{K^2})]\right]\right| \leq \frac{6}{\epsilon}(\kappa_1^2+\kappa_2^2)^2\left[\frac{2\log\frac{6}{\delta}}{m} + \sqrt{\frac{2\log\frac{6}{\delta}}{m}}\right]$$

**Theorem 3 (Dimension-dependent Estimation of 2-Wasserstein distance from finite covariance matrices).** *Let $\mathbf{X} = (x_i)_{i=1}^m$ be independently sampled from $(T,\nu)$. Assume further that $\dim(\mathcal{H}_{K^2}) < \infty$, where $\mathcal{H}_{K^2}$ is the RKHS induced by $K^2$. For any $0<\delta<1$, with probability at least $1-\delta$,*

$$\left|W_2^2\left[\mathcal{N}\left(0,\frac{1}{m}K^1[\mathbf{X}]\right), \mathcal{N}\left(0,\frac{1}{m}K^2[\mathbf{X}]\right)\right] - W_2^2[\mathcal{N}(0,C_{K^1}),\mathcal{N}(0,C_{K^2})]\right| \leq (\kappa_1^2+\kappa_2^2)\left[\frac{2\log\frac{6}{\delta}}{m} + \sqrt{\frac{2\log\frac{6}{\delta}}{m}}\right] + 2\sqrt{2}\kappa_1\kappa_2\sqrt{\dim(\mathcal{H}_{K^2})}\sqrt{\frac{2\log\frac{6}{\delta}}{m} + \sqrt{\frac{2\log\frac{6}{\delta}}{m}}}$$

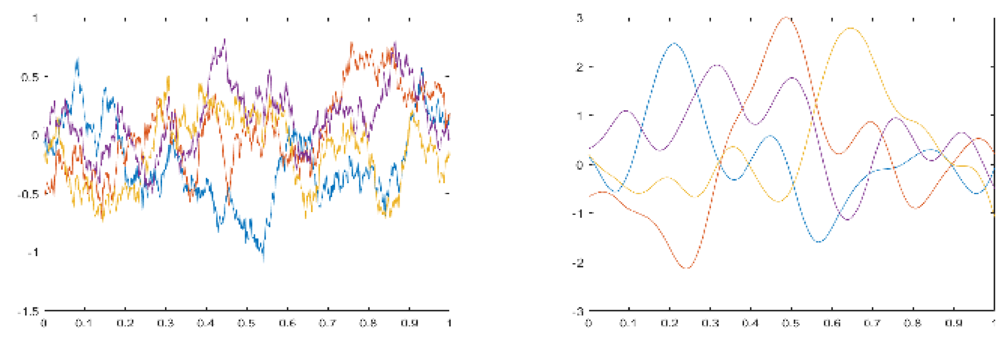
## Gaussian process setting

Let $T$ = compact metric space (in general $\sigma$-compact metric space), $\nu$ = nondegenerate Borel probability measure on $T$. Consider the Gaussian process $\xi = (\xi)_{t\in T} = (\xi(\omega,t))_{t\in T}$ on a probability space $(\Omega,\mathcal{F},P)$ with mean function $m(t) = \mathbb{E}\xi(t)$ and covariance function $K(s,t) = \mathbb{E}[(\xi(s)-m(s))(\xi(t)-m(t))]$. Assume that $\int_T m^2(t)d\nu(t) < \infty$, $\int_T K(t,t)d\nu(t) < \infty$.
There is a one-to-one correspondence between Gaussian process $\mathrm{GP}(m,K) \iff \mathcal{N}(m,C_K)$ (Gaussian measure) on $\mathcal{H} = \mathcal{L}^2(T,\nu)$, with covariance operator

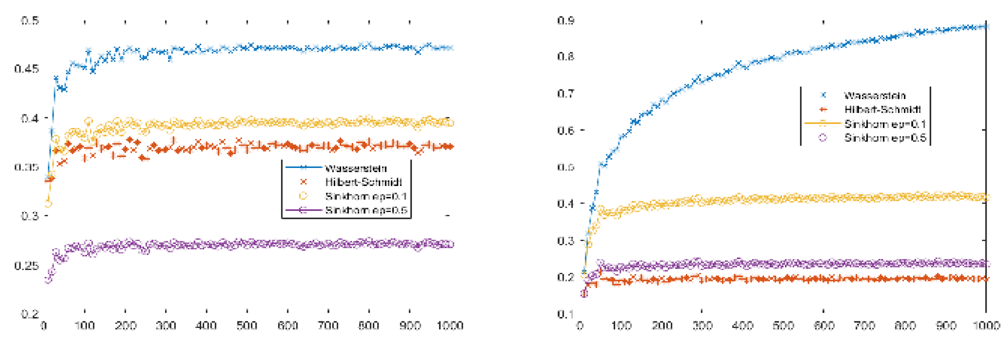
$$(C_K f)(s) = \int_T K(s,t)f(t)d\nu(t)$$

**Distances/divergences between Gaussian processes $\xi^i = \mathrm{GP}(m_i,K^i)$, $i=1,2$**

$$D_{\mathrm{GP}}(\xi^1||\xi^2) = D(\mathcal{N}(m_1,C_{K^1})||\mathcal{N}(m_2,C_{K^2}))$$



Samples of the centered Gaussian processes $\mathrm{GP}(0,K^1)$, $\mathrm{GP}(0,K^2)$ on $T = [0,1]$. Left: $K^1(x,y) = \exp(-a||x-y||)$, $a=1$. Right: $K^2(x,y) = \exp(-||x-y||^2/\sigma^2)$, $\sigma = 0.1$



Approximate squared distances/divergences between the above centered Gaussian processes on $T = [0,1]^d$ from finite covariance matrices, with $m = 10, 20, \ldots, 1000$. Left: $d=1$, right: $d=5$. The convergence of the empirical entropic regularized Sinkhorn divergence towards the theoretical value is dimension-independent, in contrast to that of the exact 2-Wasserstein distance.

## References


[1] H.Q.Minh. Entropic regularization of Wasserstein distance between infinite-dimensional Gaussian measures and Gaussian processes. **Journal of Theoretical Probability**, 2022.

[2] H.Q.Minh. Finite sample approximations of exact and entropic Wasserstein distances between covariance operators and Gaussian processes. **SIAM/ASA Journal on Uncertainty Quantification**, 2022.

[3] H.Q.Minh. Alpha Procrustes metrics between positive definite operators: a unifying formulation for the Bures-Wasserstein and Log-Euclidean/Log-Hilbert-Schmidt metrics. **Linear Algebra and Its Applications**, 2022

[4] A.Mallasto, A.Gerolin, H.Q.Minh. Entropy-regularized 2-Wasserstein distance between Gaussian measures. **Information Geometry**, 2021