Functional Analytic Learning Team

Ha Quang Minh

関数解析的学習チーム ハクァンミン

RIKEN

Center for Advanced Intelligence Project

Main Research Directions

1) Mathematical theory and structures coming from/related **to Operator Theory**, **Infinite-dimensional Information Geometry,** and **infinite-dimensional Optimal Transport**, in particular in the settings of **Hilbert spaces and stochastic processes,** and their applications **in machine learning and statistics**
2) Focus: **infinite-dimensional Gaussian measures** and **Gaussian processes**

Focus: Optimal Transport and Information Geometry in Statistics and Machine Learning

- Our focus is on Optimal Transport and Information Geometry in infinite dimension, in particular in the settings of infinite-dimensional probability measures on Hilbert space and stochastic processes with sample paths lying in infinite-dimensional function spaces. In the setting of Gaussian measures on Hilbert space and Gaussian processes, many quantities admit explicit formulas.
- Our theoretical results provide a rigorous mathematical framework for Gaussian process methods in machine learning and statistics. Examples of recent applications utilizing these results include Functional Bayesian Neural Networks and Linear Gaussian inverse problems on Hilbert space.

Example of recent results: Optimal Transport of infinite-dimensional Gaussian mixture models (GMM)

Optimal Transport distances between probability measures. Let (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 $\nu_0, \nu_1 \in \mathcal{P}(X)$ is

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

For $c(x,y) = d^p(x,y)$, $W_p(\nu_0,\nu_1) = \mathrm{OT}_{d^p}(\nu_0,\nu_1)^{1/p}$ is the p-Wasserstein distance between ν_0 and ν_1.
Let $\mathcal{H}$ be an infinite-dimensional separable Hilbert space and $\mathrm{Gauss}(\mathcal{H})$ be the set of all Gaussian measures on $\mathcal{H}$. For $\mu_i = \mathcal{N}(m_i, C_i)$, $i = 1, 2$,

$$\mathrm{W}_2^2(\mu_0,\mu_1) = \inf_{\gamma \in \mathrm{Joint}(\mu_0,\mu_1)} \int_{\mathcal{H}\times\mathcal{H}} ||x-y||^2 d\gamma(x,y) = ||m_0 - m_1||^2 + \mathrm{tr}(C_0) + \mathrm{tr}(C_1) - 2\mathrm{tr}(C_0^{1/2}C_1C_0^{1/2})^{1/2}$$

with the optimal transport plan γ being a joint Gaussian measure of μ_0 and μ_1.
Gaussian mixture models (GMM) on $\mathbb{R}^n$: Chen, Georgiou, and Tannenbaum (2019), Delon and Desolneux (2020)
Gaussian mixture models (GMM) on $\mathcal{H}$: Let $\mathrm{GMM}_{\mathcal{H}}(M)$ be the set of probability measures on $\mathcal{H}$ that can be expressed as sum of M or fewer Gaussian measures. Let $\mathrm{GMM}_{\mathcal{H}}(\infty) = \cup_{M>0}\mathrm{GMM}_{\mathcal{H}}(M)$ be the set of all Gaussian mixtures on $\mathcal{H}$.
Optimal transport between two GMMs: For μ_0, μ_1 both being GMMs the optimal transport plan γ may not necessarily be a GMM. Consider the OT problem restricted to the set of joint measures which are GMMs, which gives rise to a Wasserstein-type distance $\mathrm{MW}_2(\mu_0,\mu_1)$ between μ_0 and μ_1

$$\mathrm{W}_2^2(\mu_0,\mu_1) \leq \mathrm{MW}_2^2(\mu_0,\mu_1) = \inf_{\gamma \in \mathrm{Joint}(\mu_0,\mu_1)\cap \mathrm{GMM}_{\mathcal{H}\times\mathcal{H}}(\infty)} \int_{\mathcal{H}\times\mathcal{H}} ||x-y||^2 d\gamma(x,y)$$

Equivalent discrete formulation: For two discrete probability measures $\pi_i \in \mathbb{R}_+^{M_i}, \sum_{j=1}^{M_i}\pi_i^j = 1$, define $\mathrm{Joint}(\pi_0,\pi_1) = \{w \in \mathbb{R}_+^{M_0\times M_1} : w\mathbf{1}_{M_1} = \pi_0, w^T\mathbf{1}_{M_0} = \pi_1\}$

Proposition 1 (Equivalent discrete formulation). *Let $\mu_i = \sum_{j=1}^{M_i}\pi_i^j\mu_i^j$, $i = 0, 1$, be two mixtures of Gaussian measures on $\mathcal{H}$, with $\mu_i^j = \mathcal{N}(m_i^j, C_i^j)$. Then*

$$\mathrm{MW}_2^2(\mu_0,\mu_1) = \inf_{\gamma \in \mathrm{Joint}(\mu_0,\mu_1)\cap \mathrm{GMM}_{\mathcal{H}\times\mathcal{H}}(\infty)} \int_{\mathcal{H}\times\mathcal{H}} ||x-y||^2 d\gamma(x,y) = \min_{w \in \mathrm{Joint}(\pi_0,\pi_1)} \sum_{j=1}^{M_0}\sum_{k=1}^{M_1} w_{jk}\mathrm{W}_2^2(\mu_0^j,\mu_1^k)$$

Stochastic process setting. Let T = compact metric space (in general σ-compact metric space), ν = 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 $\mathrm{GP}(m, K) \Longleftrightarrow \mathcal{N}(m, C_K)$ (Gaussian measure) on $\mathcal{H} = \mathcal{L}^2(T,\nu)$, with covariance operator $(C_K f)(s) = \int_T K(s,t)f(t)d\nu(t)$.
Consider GMM of the form $\mu = \sum_{j=1}^{M}\pi^j\mathcal{N}(0, C_{K^j})$. Let $\mathbf{X} = (x_i)_{i=1}^m$ be independently sampled from (T,ν). Define $\mu_{\mathbf{X}} = \sum_{j=1}^{M}\pi^j\mathcal{N}(0, \frac{1}{m}K^j[\mathbf{X}])$.

Theorem 1 (Finite-dimensional estimation of MW_2). *Let $\mu_i = \sum_{j=1}^{M_i}\pi_i^j\mathcal{N}(0, C_{K_i^j})$, $i = 0, 1$. Let $\mathbf{X} = (x_i)_{i=1}^m$ be independently sampled from (T,ν). Assume that $\sup_{x,t\in T} K_i^j(x,t) = \kappa_i^j$. For any $0 < \delta < 1$, with probability at least $1 - \delta$,*

$$\begin{aligned}|\mathrm{MW}_2^2(\mu_{0,\mathbf{X}},\mu_{1,\mathbf{X}}) - \mathrm{MW}_2^2(\mu_0,\mu_1)| &\leq \sum_{j=1}^{M_0}\sum_{k=1}^{M_1} w_{jk}^*\left|W_2^2\left[\mathcal{N}\left(0,\frac{1}{m}K_0^j[\mathbf{X}]\right),\mathcal{N}\left(0,\frac{1}{m}K_1^k[\mathbf{X}]\right)\right] - W_2^2[\mathcal{N}(0,C_{K_0^j}),\mathcal{N}(0,C_{K_1^k})]\right| \\ &\leq \sum_{j=1}^{M_0}\sum_{k=1}^{M_1}\left(w_{jk}^*((\kappa_0^j)^2 + (\kappa_1^k)^2)\left[\frac{2\log\frac{6}{\delta}}{m} + \sqrt{\frac{2\log\frac{6}{\delta}}{m}}\right] + 2\sqrt{2}\kappa_0^j\kappa_1^k\sqrt{\dim(\mathcal{H}_{K_1^k})}\sqrt{\frac{2\log\frac{6}{\delta}}{m} + \sqrt{\frac{2\log\frac{6}{\delta}}{m}}}\right)\end{aligned}$$

Recent publications

[1] H.Q.Minh. *Fisher-Rao geometry of equivalent Gaussian measures on infinite-dimensional Hilbert spaces*, **Information Geometry**, 2024.

[2] H.Q.Minh. *Infinite-dimensional distances and divergences between positive definite operators, Gaussian measures, and Gaussian processes*, **Information Geometry**, 2024.

[3] Guillaume Braun, Minh Ha Quang, Masaaki Imaizumi. *Learning a Single Index Model from Anisotropic Data with Vanilla Stochastic Gradient Descent*, **International Conference on Artificial Intelligence and Statistics (AISTATS 2025)**, accepted for publication.