

Sparsity-Aware Distributed Learning for Gaussian Processes with Linear Multiple Kernel

Richard Cornelius Suwandi

Supervisor: Prof. Feng Yin & Prof. Tsung-Hui Chang

July 19, 2025

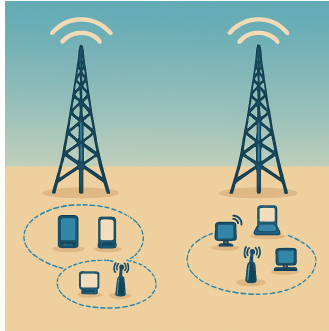


GPs are Everywhere!

- Gaussian processes (GPs) are widely used in science and engineering.



(a) Climate modeling



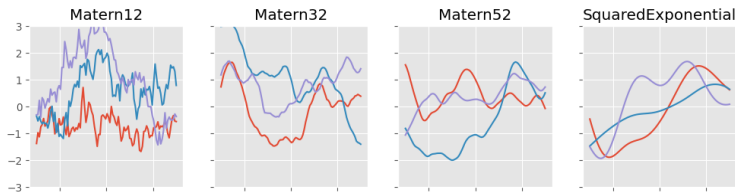
(b) Wireless traffic



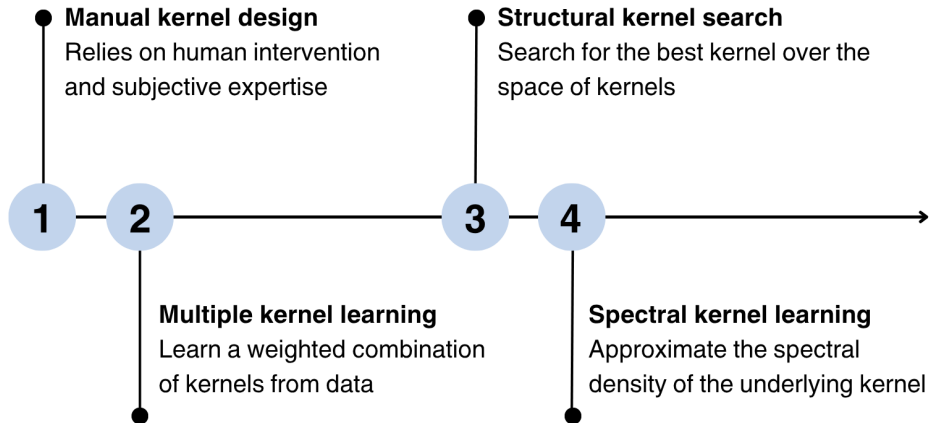
(c) Motion planning

The Kernel Design Problem

- ▶ The **representation power** of a GP relies on the kernel function.
- ▶ Manual kernel design is subjective and not scalable.

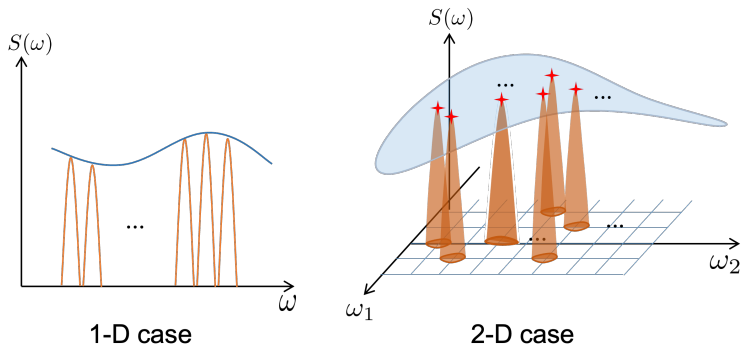


The Journey of Kernel Design



The Curse of Dimensionality

- Existing GSM kernels don't scale to high dimensions: number of parameters grows exponentially.



Grid Spectral Mixture Product (GSMP) Kernel

- **Idea:** Product of 1D sub-kernels, scalable, linear in parameters.

GSMP Kernel

$$k_{\text{GSMP}}(\boldsymbol{\tau}; \boldsymbol{\theta}) = \sum_{q=1}^Q \theta_q \prod_{p=1}^P \exp \left\{ -2\pi^2 \tau^{(p)^2} v_q^{(p)} \right\} \cos \left(2\pi \tau^{(p)} \mu_q^{(p)} \right)$$

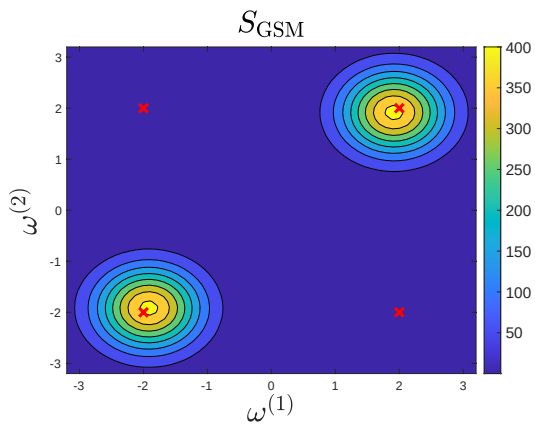
Properties of GSMP Kernel

- ▶ Fewer parameters needed for high-dimensional data.
- ▶ Sparse solutions: Only significant sub-kernels are used.
- ▶ Interpretability and generalization improved.

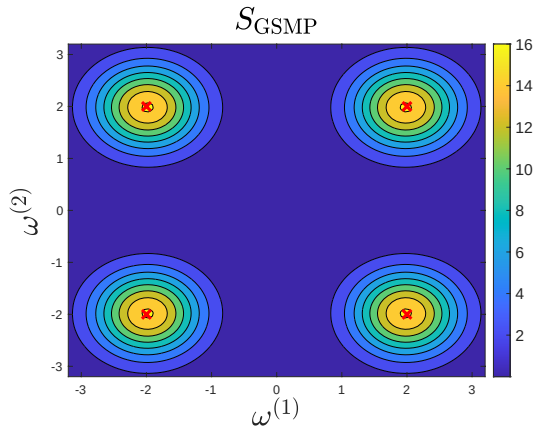
Spectral Density

$$S_{\text{GSMP}}(\omega) = \frac{1}{2^P} \sum_{q=1}^Q \theta_q \prod_{p=1}^P \left[\mathcal{N} \left(\omega^{(p)} | \mu_q^{(p)}, v_q^{(p)} \right) + \mathcal{N} \left(\omega^{(p)} | -\mu_q^{(p)}, v_q^{(p)} \right) \right]$$

GSMP vs GSM: Visual Comparison



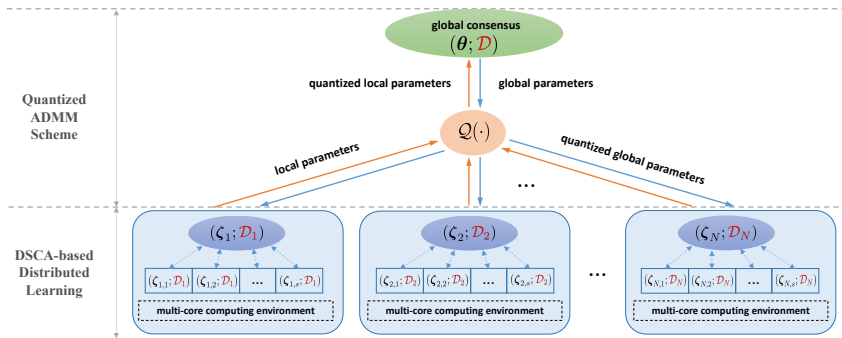
GSM Kernel



GSMP Kernel

Efficient Learning Framework: SLIM-KL

- **SLIM-KL**: Sparse Linear Multiple Kernel Learning framework.
- Distributed, quantized ADMM + DSCA for scalable, privacy-preserving learning.



Algorithmic Details Theoretical Guarantees

- ▶ Distributed optimization: Each agent solves local problem, consensus via quantized ADMM.
- ▶ Theoretical guarantees: Convergence, scalability, privacy.

Key Update (Quantized ADMM)

$$\begin{aligned}\boldsymbol{\theta}^{t+1} &= \frac{1}{N} \sum_{j=1}^N \left(\boldsymbol{\zeta}_{j,[Q]}^t + \frac{1}{\rho_j} \right) \\ \boldsymbol{\theta}_{[Q]}^{t+1} &= \mathcal{Q}(\boldsymbol{\theta}^{t+1})\end{aligned}$$

Experimental Setup

- ▶ Diverse real-world datasets: 1D, multi-D, large-scale.
- ▶ Metrics: Prediction MSE, sparsity, scalability.

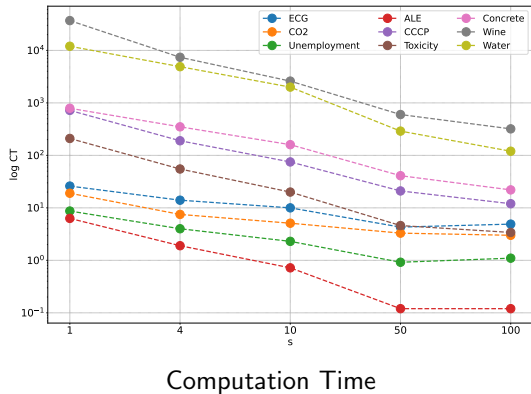
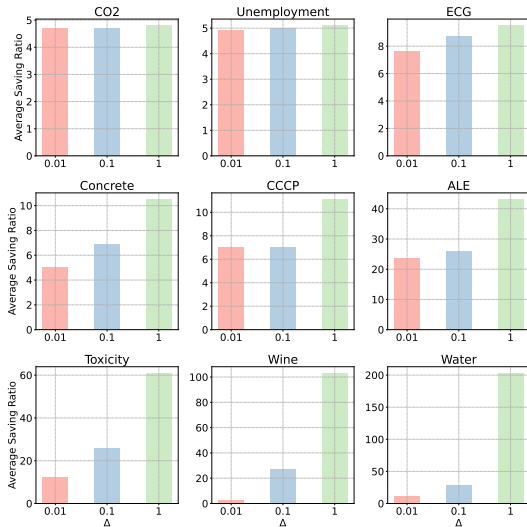
Dataset	ECG	CO2	Unemp.	ALE	CCCP	Toxicity	Concrete	Wine
Train Size	680	481	380	80	1000	436	824	1279
Test Size	20	20	20	20	250	110	206	320
Dim	1	1	1	4	4	8	8	11

Prediction Performance

- GSMPGP consistently outperforms baselines in MSE and sparsity.

Dataset	GSMPGP	GSMGP	SMGP	SEGP	LSTM	Informer
ECG	1.2E-02	1.2E-02	1.9E-02	1.6E-01	1.6E-01	5.4E-02
CO2	3.7E-01	6.2E-01	7.4E-01	1.5E+03	2.9E+02	8.4E+01
Unemp.	2.0E+03	2.2E+03	7.7E+03	5.6E+05	1.7E+05	3.8E+03
ALE	2.5E-02	1.9E+00	3.7E-01	3.7E-02	3.4E-02	1.8E-01
CCCP	1.6E+01	1.9E+02	2.1E+05	1.7E+01	2.8E+02	1.4E+05
Toxicity	1.4E+00	2.7E+02	5.3E+00	1.4E+00	1.7E+00	2.9E+00
Concrete	5.9E+01	5.6E+02	1.7E+03	1.3E+02	1.4E+02	2.0E+02
Wine	4.6E-01	7.4E+03	3.1E+01	4.4E+00	4.7E-01	1.6E+00

Scalability Communication Efficiency



Practical Impact: Long Time Series Forecasting

- ▶ GSMPGP achieves SOTA or near-SOTA results on real-world forecasting tasks.

Dataset	GSMPGP	PatchTST	SegRNN	MICN	TiDE
ETTh1	0.431	0.447	0.434	0.770	0.454
ETTh2	0.382	0.379	0.394	0.956	0.419
ETTm1	0.407	0.416	0.410	0.481	0.413
ETTm2	0.341	0.362	0.330	0.502	0.352
Weather	0.313	0.314	0.310	0.311	0.313