Learning with Nonparametric Kernels

Wessel Bruinsma, Felipe Tobar, Richard Turner

UNIVERSITY OF CAMBRIDGE

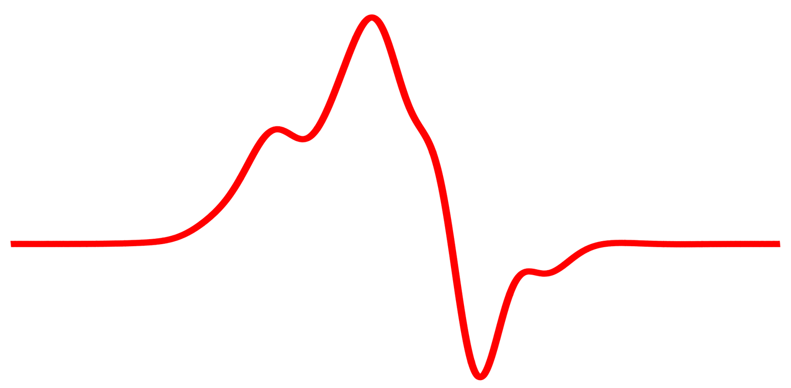
MPHIL IN MACHINE LEARNING, SPEECH AND LANGUAGE TECHNOLOGY

- Gaussian process (GP) regression models widely used for their **expressiveness**, **robustness** and **tractability**
- Choice of kernel determines nature of GP — **expressive kernels** are needed
- Recent work [1] models kernel **nonparametrically** in a single-output time series setting
- Aim is to extend this work to **multidimensional input and output spaces** — numerous important applications

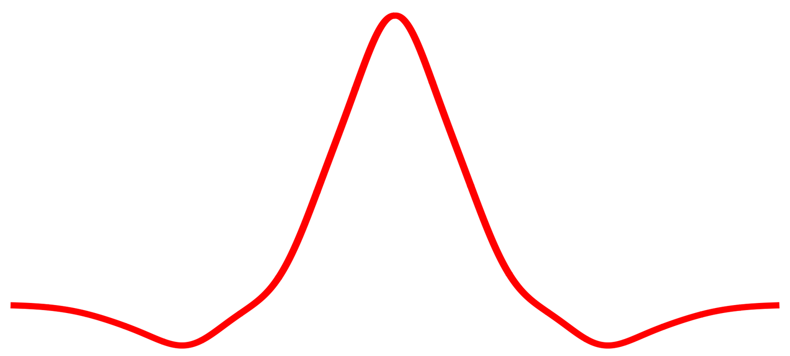
Gaussian Process Convolution Model (GPCM) [1]

GENERATIVE MODEL

❶ $h \sim \mathcal{GP}(0, \mathcal{K}_h)$

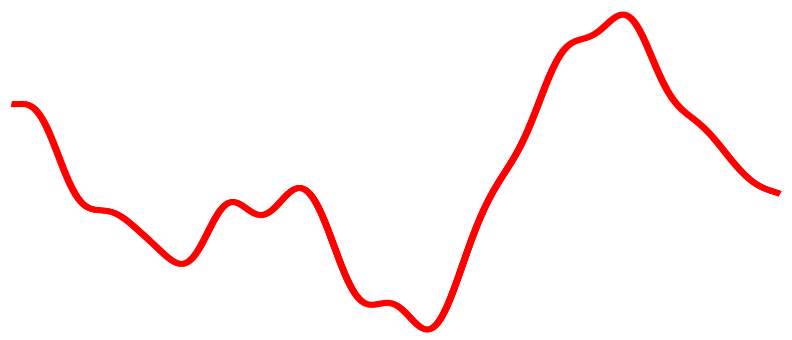


❷ $\mathcal{K}_{f\,|\,h} =^1 h(t) * h(-t)$



$^1 h(t) * h(-t) = \int_{\mathbb{R}} h(\tau + t) h(\tau) \mathrm{d}\tau$

❸ $f\,|\,h \sim \mathcal{GP}(0, \mathcal{K}_{f\,|\,h})$



- Parameterising $\mathcal{K}_{f\,|\,h}$ as the convolution between $h(t)$ and $h(-t)$ ensures that $\mathcal{K}_{f\,|\,h}$ is positive definite
- Equivalently, $f = x * h$ for $x \sim \mathcal{GP}(0, \delta(t - t'))$
- Learning and inference are performed using state-of-the-art variational free-energy approximations
- Particular choice of $\mathcal{K}_h$ reveals $\mathcal{K}_{f\,|\,h}$ as a spectral mixture kernel [2] with an infinite number of components

Multidimensional Signal Estimation

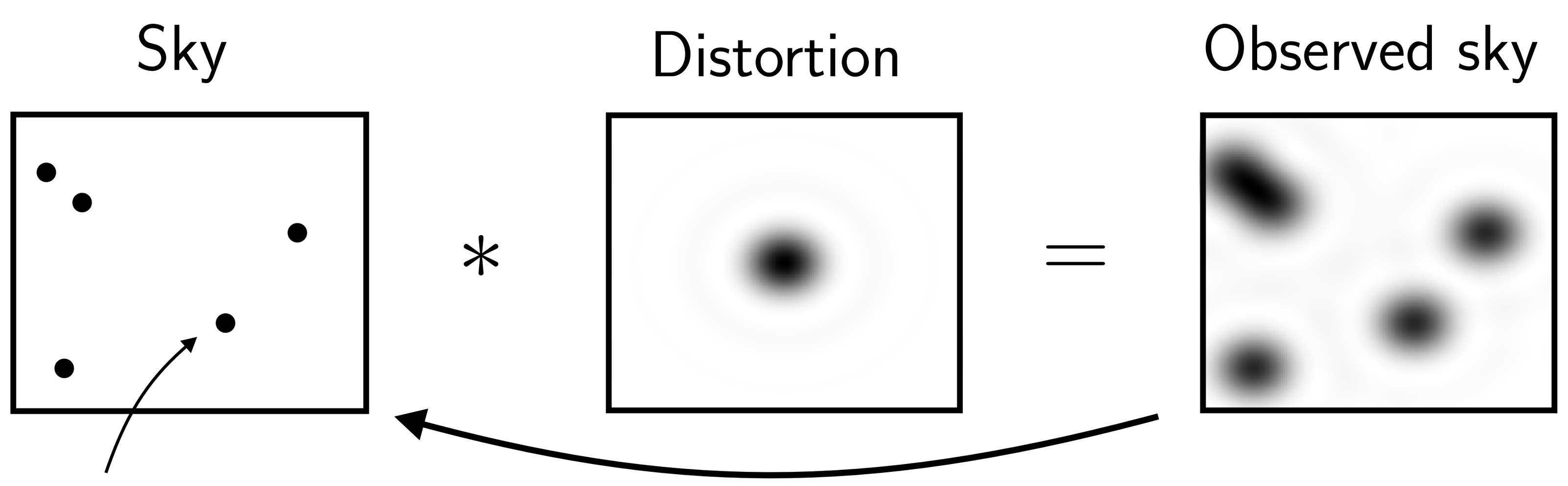
Extension to **multidimensional input space** $\mathcal{T}$ involves extending h's domain to $\mathcal{T}$; for high-dimensional $\mathcal{T}$ let

$$h(t_1, \ldots, t_n) = h_1(t_1) \cdots h_n(t_n)$$

- Separability of h avoids exponential growth in computational complexity

APPLICATION IN ASTROINFORMATICS

Unsupervised compensation of distortion due to the lens and atmosphere in astrophotography



Estimate "real sky" from observed sky

FURTHER APPLICATIONS

- Image denoising
- Non-stationary signal estimation

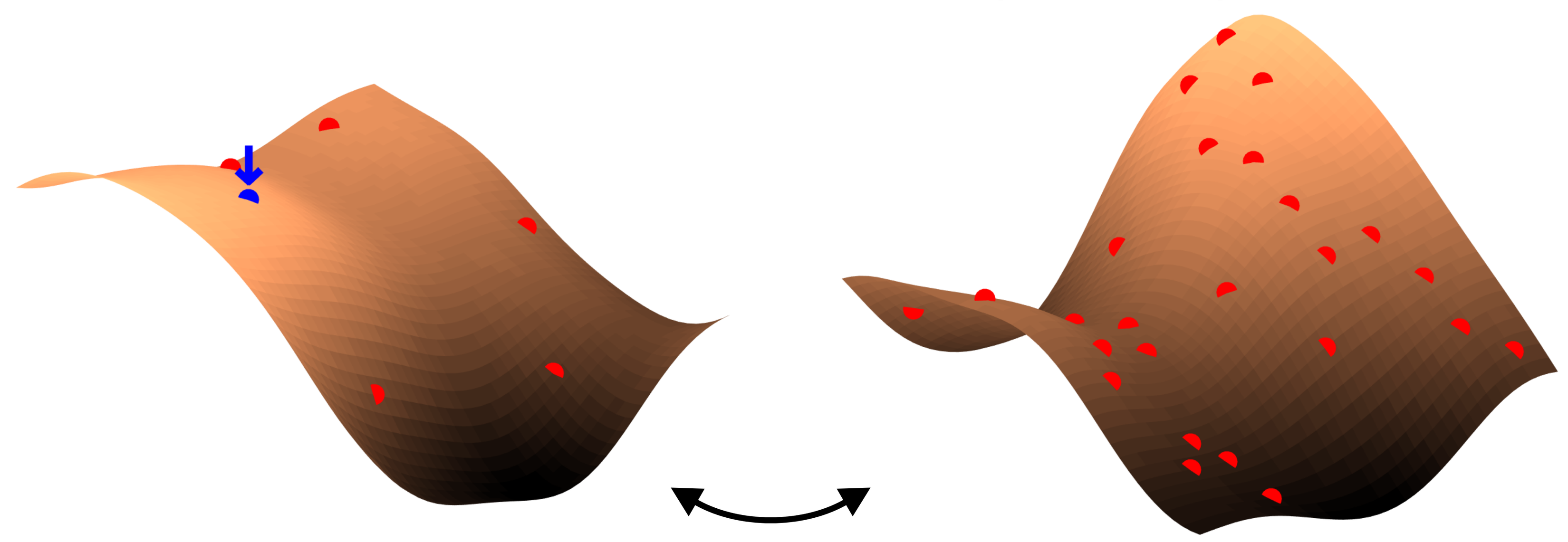
Multi-Task Learning

Extension to incorporate **multiple outputs** f_i:

$$\mathcal{K}_{f_i, f_j\,|\,h} = \sum_k h_{i,k}(t) * h_{j,k}(-t)$$

APPLICATION IN GEOSTATISTICS

Interpolation of concentration level surfaces of expensive to sample minerals



Exploit correlations between the surfaces to better estimate a new point on the expensive one

FURTHER APPLICATIONS

- Study of transcription factors in gene expression
- Prediction of exchange rates

Bayesian Power Spectrum Estimation

The GPCM defines a distribution over stationary kernels, which thus implies a distribution over power spectra

$$\mathcal{F}\left\{\mathcal{K}_{f\,|\,h}\right\}(f) = |H(f)|^2$$

APPLICATIONS

- Extension to multiple outputs to perform Bayesian cross power spectrum estimation:

$$\mathcal{F}\left\{\mathcal{K}_{f_i, f_j\,|\,h}\right\}(f) = \sum_k H^*_{i,k}(f) H_{j,k}(f)$$

- Bayes optimal signal detection in the power spectrum to improve dynamic spectrum management

References

[1] F. Tobar, T. D. Bui, and R. E. Turner, "Learning Stationary Time Series using Gaussian Processes with Nonparametric Kernels," *Advances in Neural Information Processing Systems*, vol. 29, pp. 3501–3509, 2015.

[2] A. G. Wilson and R. P. Adams, "Gaussian Process Kernels for Pattern Discovery and Extrapolation," *International Conference on Machine Learning*, vol. 3, pp. 1067–1075, 2013.