Spectral Methods in Gaussian Modelling

Topic 2: Kernel Design

James Requiema and Wessel Bruinsma

University of Cambridge and Invenia Labs

21 January 2019

+ RFFs alleviate the $O(N^3)$ scaling.

Kernel Design

+ RFFs alleviate the $O(N^3)$ scaling.

− RFFs do not help with choice of kernel.

Kernel Design

+ RFFs alleviate the $O(N^3)$ scaling.

– RFFs do not help with choice of kernel.

How to parametrise a flexible kernel?

Kernel Design (2)

- Bochner's Theorem:

$$k(\tau) \overset{\mathcal{F}}{\longleftrightarrow} s(\omega) = \text{PSD}.$$

Kernel Design (2)

- Bochner's Theorem:

$$k(\tau) \xleftrightarrow{\mathcal{F}} s(\omega) = \text{PSD}.$$

- PSD:
  - distribution of power contained in frequencies,
  - must be **nonnegative** and **symmetric**.

Kernel Design (2)

- Bochner's Theorem:

$$k(\tau) \xleftrightarrow{\mathcal{F}} s(\omega) = \text{PSD}.$$

- PSD:
  - distribution of power contained in frequencies,
  - must be **nonnegative** and **symmetric**.
- Easier to flexibly parametrise PSD!

Spectral Mixture Kernel

- SSA (Lázaro-Gredilla et al., 2010) models PSD with **symmetric average** of **lines**:

$$s(\omega) = \frac{1}{2Q}\sum_{q=1}^{Q}(\delta(\omega - \mu^{(q)}) + \delta(\omega + \mu^{(q)})).$$

Spectral Mixture Kernel

- SSA (Lázaro-Gredilla et al., 2010) models PSD with **symmetric average** of **lines**:

$$s(\omega) = \frac{1}{2Q}\sum_{q=1}^{Q}(\delta(\omega - \mu^{(q)}) + \delta(\omega + \mu^{(q)})).$$

- Inverse Fourier transform gives kernel:

$$k(\tau) = \frac{1}{Q}\sum_{q=1}^{Q}\cos(\mu^{(q)\mathsf{T}}\tau).$$

Spectral Mixture Kernel

- SSA (Lázaro-Gredilla et al., 2010) models PSD with **symmetric average** of **lines**:

$$s(\omega) = \frac{1}{2Q} \sum_{q=1}^{Q} (\delta(\omega - \mu^{(q)}) + \delta(\omega + \mu^{(q)})).$$

- Inverse Fourier transform gives kernel:

$$k(\tau) = \frac{1}{Q} \sum_{q=1}^{Q} \cos(\mu^{(q)\mathsf{T}} \tau).$$

- Strong parametric assumption: $f(t) =$ sum of sines.

Spectral Mixture Kernel (2)

- SMK (Wilson and Adams, 2013) models PSD with **symmetric mixture** of **Gaussians**:

$$s(\omega) = \frac{1}{2} \sum_{q=1}^{Q} w^{(q)} \Big(\mathcal{N}\Big(\omega; \mu^{(q)}, \Sigma^{(q)}\Big) + \mathcal{N}\Big(\omega; -\mu^{(q)}, \Sigma^{(q)}\Big) \Big).$$

Spectral Mixture Kernel (2)

- SMK (Wilson and Adams, 2013) models PSD with **symmetric mixture** of **Gaussians**:

$$s(\omega) = \frac{1}{2}\sum_{q=1}^{Q} w^{(q)}\Big(\mathcal{N}\Big(\omega; \mu^{(q)}, \Sigma^{(q)}\Big) + \mathcal{N}\Big(\omega; -\mu^{(q)}, \Sigma^{(q)}\Big)\Big).$$

- $w^{(q)} = 1/Q$ and $\Sigma^{(q)} \to 0$ recovers SSA.

Spectral Mixture Kernel (2)

- SMK (Wilson and Adams, 2013) models PSD with **symmetric mixture** of **Gaussians**:

$$s(\omega) = \frac{1}{2} \sum_{q=1}^{Q} w^{(q)} \Big(\mathcal{N}\Big(\omega; \mu^{(q)}, \Sigma^{(q)}\Big) + \mathcal{N}\Big(\omega; -\mu^{(q)}, \Sigma^{(q)}\Big)\Big).$$

- $w^{(q)} = 1/Q$ and $\Sigma^{(q)} \to 0$ recovers SSA.
- Inverse Fourier transform gives kernel:

$$k^{(\mathrm{SMK})}(\tau) = \sum_{q=1}^{Q} w^{(q)} \exp\Big(-\tfrac{1}{2}\tau^{\mathsf{T}}\Sigma^{(q)}\tau\Big) \cos\Big(\mu^{(q)\mathsf{T}}\tau\Big),$$

Spectral Mixture Kernel (3)

- Equivalent generative model as a truncated Fourier series:

$$f^{(\mathrm{SMK})}(t) = \sum_{q=1}^{Q} \sqrt{w^{(q)}}\left(c_1^{(q)}(t)\cos(\mu^{(q)\mathsf{T}} t) + c_2^{(q)}(t)\sin(\mu^{(q)\mathsf{T}} t)\right),$$

$$c_1^{(q)}, c_2^{(q)} \sim \mathcal{GP}(0, \exp(-\tfrac{1}{2}\tau^{\mathsf{T}}\Sigma^{(q)}\tau)).$$

Spectral Mixture Kernel (3)

- Equivalent generative model as a truncated Fourier series:

$$f^{(\mathrm{SMK})}(t) = \sum_{q=1}^{Q} \sqrt{w^{(q)}}\left(c_1^{(q)}(t)\cos(\mu^{(q)\mathsf{T}}t) + c_2^{(q)}(t)\sin(\mu^{(q)\mathsf{T}}t)\right),$$

$$c_1^{(q)}, c_2^{(q)} \sim \mathcal{GP}(0, \exp(-\tfrac{1}{2}\tau^{\mathsf{T}}\Sigma^{(q)}\tau)).$$

- In SSA, $(c_1^{(q)}, c_2^{(q)})_{q=1}^{Q}$ are constant.

Spectral Mixture Kernel (3)

- Equivalent generative model as a truncated Fourier series:

$$f^{(\mathrm{SMK})}(t) = \sum_{q=1}^{Q} \sqrt{w^{(q)}}(c_1^{(q)}(t)\cos(\mu^{(q)\mathsf{T}}t) + c_2^{(q)}(t)\sin(\mu^{(q)\mathsf{T}}t)),$$

$$c_1^{(q)}, c_2^{(q)} \sim \mathcal{GP}(0, \exp(-\tfrac{1}{2}\tau^{\mathsf{T}}\Sigma^{(q)}\tau)).$$

- In SSA, $(c_1^{(q)}, c_2^{(q)})_{q=1}^{Q}$ are constant.
- SMK fattens spectral lines by allowing $c_1^{(q)}$ and $c_2^{(q)}$ to vary with time.

Spectral Mixture Kernel (4)

+ Flexible, drop-in replacement

Spectral Mixture Kernel (4)

+ Flexible, drop-in replacement
+ Can recover many standard kernels

Spectral Mixture Kernel (4)

+ Flexible, drop-in replacement
+ Can recover many standard kernels
+ Models negative covariances

Spectral Mixture Kernel (4)

+ Flexible, drop-in replacement
+ Can recover many standard kernels
+ Models negative covariances

− Unclear how many components needed

Spectral Mixture Kernel (4)

+ Flexible, drop-in replacement
+ Can recover many standard kernels
+ Models negative covariances

– Unclear how many components needed
– Hyperparameters difficult to optimise

- MOSMK (Parra and Tobar, 2017) generalises SMK to multiple outputs.

Multi-Output Spectral Mixture Kernel

- MOSMK (Parra and Tobar, 2017) generalises SMK to multiple outputs.
- Uses multivariate extension of Bochner's Theorem: Cramér's Theorem.

Multi-Output Spectral Mixture Kernel

- MOSMK (Parra and Tobar, 2017) generalises SMK to multiple outputs.
- Uses multivariate extension of Bochner's Theorem: Cramér's Theorem.
- Multivariate PSD $S\colon \mathbb{R}^D \to \mathbb{C}^{P\times P}$.

Multi-Output Spectral Mixture Kernel

- MOSMK (Parra and Tobar, 2017) generalises SMK to multiple outputs.
- Uses multivariate extension of Bochner's Theorem: Cramér's Theorem.
- Multivariate PSD $S\colon \mathbb{R}^D \to \mathbb{C}^{P\times P}$.
  - Must be **symmetric**: $S(\omega) = S^\dagger(-\omega)$, $S_{ii}(\omega) = S_{ii}(-\omega)$.

Multi-Output Spectral Mixture Kernel

- MOSMK (Parra and Tobar, 2017) generalises SMK to multiple outputs.
- Uses multivariate extension of Bochner's Theorem: Cramér's Theorem.
- Multivariate PSD $S\colon \mathbb{R}^D \to \mathbb{C}^{P\times P}$.
  - Must be **symmetric**: $S(\omega) = S^{\dagger}(-\omega)$, $S_{ii}(\omega) = S_{ii}(-\omega)$.
  - Must be **nonnegative**: $S(\omega) \geq 0$.

Multi-Output Spectral Mixture Kernel (2)

- MOSMK models PSD with **symmetric mixture** of **outer products** of **vectors of Gaussians**:

$$S(\omega) = \frac{1}{2}\sum_{q=1}^{Q}\Big(R^{(q)}(\omega)R^{(q)\dagger}(\omega) + R^{(q)}(-\omega)R^{(q)\dagger}(-\omega)\Big),$$

$$\begin{aligned}R_i^{(q)}(\omega) = w^{(q)} \exp\Big(&-\tfrac{1}{4}(\omega - \mu_i^{(q)})\Sigma_i^{(q)-1}(\omega - \mu_i^{(q)}) \\ &- \iota(\theta_i^{(q)\mathsf{T}}\omega + \phi_i^{(q)})\Big).\end{aligned}$$

Multi-Output Spectral Mixture Kernel (3)

- Inverse Fourier transform gives kernel:

$$K_{ij}^{(\mathrm{MOSMK})}(\tau) = \sum_{q=1}^{Q} \alpha_{ij}^{(q)} \exp\Big(-\tfrac{1}{2}(\tau + \theta_{ij}^{(q)})^{\mathsf{T}} \Sigma_{ij}^{(q)} (\tau + \theta_{ij}^{(q)})\Big) \\ \times \cos\Big((\tau + \theta_{ij}^{(q)})^{\mathsf{T}} \mu_{ij}^{(q)} + \phi_{ij}^{(q)}\Big).$$

- Equivalent generative model as truncated Fourier series:

$$
\begin{aligned}
&f_i^{(\mathrm{MOSMK})}(t) \\
&\quad = \sum_{q=1}^{Q} w_i^{(q)} \Big(c_{i1}^{(q)}(t-\theta_i^{(q)}) \cos\Big(\mu_i^{(q)\mathsf{T}}(t-\theta_i^{(q)}) + \phi_i^{(q)}\Big) \\
&\qquad\qquad + c_{i2}^{(q)}(t-\theta_i^{(q)}) \sin\Big(\mu_i^{(q)\mathsf{T}}(t-\theta_i^{(q)}) + \phi_i^{(q)}\Big)\Big),
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}[c_{ik}^{(p)}(t) c_{jl}^{(q)}(t')] \\
&\quad = \begin{cases} \frac{\alpha_{ij}^{(q)}}{w_i^{(q)} w_j^{(q)}} \exp\Big(-\frac{1}{2}(t-t')^{\mathsf{T}} \Sigma_{ij}^{(q)} (t-t')\Big) & \text{if } k = l,\, p = q, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

Generalised Spectral Mixture Kernel

- GSMK (Chen et al., 2018) generalises SMK to nonstationary signals.

Generalised Spectral Mixture Kernel

- GSMK (Chen et al., 2018) generalises SMK to nonstationary signals.
- Uses the *Gibbs kernel* (Gibbs, 1997):

$$k^{(\mathrm{Gibbs})}(t,t') = \prod_{d=1}^{D} \sqrt{\frac{2\ell_d(t)\ell_d(t')}{\ell_d^2(t) + \ell_d^2(t')}} \exp\left(- \sum_{d=1}^{D} \frac{(t_d - t'_d)^2}{\ell_d^2(t) + \ell_d^2(t')} \right).$$

Nonstationary EQ Kernel

- Cannot simply make length scale input dependent.

Nonstationary EQ Kernel

- Cannot simply make length scale input dependent.
- Construction of EQ from basis functions:

$$\phi(t; c) = \left(\sqrt{\frac{2}{\pi}} \frac{1}{\ell} \right)^{\frac{1}{2}} \exp\left(-\frac{1}{\ell^2} (t-c)^2 \right),$$

Nonstationary EQ Kernel

- Cannot simply make length scale input dependent.
- Construction of EQ from basis functions:

$$\phi(t;c) = \left(\sqrt{\frac{2}{\pi}}\frac{1}{\ell}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{\ell^2}(t-c)^2\right),$$

$$f(t)\,|\,n = \int_{-\infty}^{\infty} \phi(t;c)n(c)\,\mathrm{d}c, \quad n(t) \sim \mathcal{GP}(0, \delta(t-t')),$$

Generalised Spectral Mixture Kernel (2)

Nonstationary EQ Kernel

- Cannot simply make length scale input dependent.
- Construction of EQ from basis functions:

$$\phi(t;c) = \left(\sqrt{\frac{2}{\pi}}\frac{1}{\ell}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{\ell^2}(t-c)^2\right),$$

$$f(t)\,|\,n = \int_{-\infty}^{\infty} \phi(t;c)n(c)\,\mathrm{d}c, \quad n(t) \sim \mathcal{GP}(0, \delta(t-t')),$$

$$\mathbb{E}[f(t)f(t')] = \int_{-\infty}^{\infty} \phi(t;c)\phi(t';c)\,\mathrm{d}c$$

Generalised Spectral Mixture Kernel (2)

Nonstationary EQ Kernel

- Cannot simply make length scale input dependent.
- Construction of EQ from basis functions:

$$
\begin{aligned}
\phi(t; c) &= \left(\sqrt{\frac{2}{\pi}}\frac{1}{\ell}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{\ell^2}(t-c)^2\right), \\
f(t) \,|\, n &= \int_{-\infty}^{\infty} \phi(t; c) n(c) \,\mathrm{d}c, \quad n(t) \sim \mathcal{GP}(0, \delta(t - t')), \\
\mathbb{E}[f(t) f(t')] &= \int_{-\infty}^{\infty} \phi(t; c)\phi(t'; c) \,\mathrm{d}c \\
&= \exp\left(-\frac{1}{2\ell^2}(t - t')^2\right).
\end{aligned}
$$

Generalised Spectral Mixture Kernel (3)

Nonstationary EQ Kernel

- Make length scale of ϕ dependent on t:

$$\phi(t; c) = \left(\sqrt{\frac{2}{\pi}} \frac{1}{\ell(t)} \right)^{\frac{1}{2}} \exp\left(-\frac{1}{\ell^2(t)} (t - c)^2 \right),$$

Generalised Spectral Mixture Kernel (3)

Nonstationary EQ Kernel

- Make length scale of ϕ dependent on t:

$$\phi(t;c) = \left(\sqrt{\frac{2}{\pi}}\frac{1}{\ell(t)}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{\ell^2(t)}(t-c)^2\right),$$

$$f(t)\,|\,n = \int_{-\infty}^{\infty} \phi(t;c)n(c)\,\mathrm{d}c, \quad n(t) \sim \mathcal{GP}(0, \delta(t-t')),$$

Generalised Spectral Mixture Kernel (3)

Nonstationary EQ Kernel

- Make length scale of ϕ dependent on t:

$$\phi(t;c) = \left(\sqrt{\frac{2}{\pi}}\frac{1}{\ell(t)}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{\ell^2(t)}(t-c)^2\right),$$

$$f(t)\,|\,n = \int_{-\infty}^{\infty} \phi(t;c)n(c)\,\mathrm{d}c, \quad n(t) \sim \mathcal{GP}(0, \delta(t-t')),$$

$$\mathbb{E}[f(t)f(t')] = \int_{-\infty}^{\infty} \phi(t;c)\phi(t';c)\,\mathrm{d}c$$

Nonstationary EQ Kernel

- Make length scale of ϕ dependent on t:

$$\phi(t;c) = \left(\sqrt{\frac{2}{\pi}}\frac{1}{\ell(t)}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{\ell^2(t)}(t-c)^2\right),$$

$$f(t)\,|\,n = \int_{-\infty}^{\infty} \phi(t;c)n(c)\,\mathrm{d}c, \quad n(t) \sim \mathcal{GP}(0, \delta(t-t')),$$

$$\begin{aligned}\mathbb{E}[f(t)f(t')] &= \int_{-\infty}^{\infty} \phi(t;c)\phi(t';c)\,\mathrm{d}c \\ &= \sqrt{\frac{2\ell(t)\ell(t')}{\ell^2(t)+\ell^2(t')}}\exp\left(-\frac{(t-t')^2}{\ell^2(t)+\ell^2(t')}\right).\end{aligned}$$

Generalised Spectral Mixture Kernel (4)

- GSMK replaces the EQs with Gibbs kernels:

$$k^{(\mathrm{GSMK})}(t,t') = \sum_{q=1}^{Q} w^{(q)}(t) w^{(q)}(t') k_q^{(\mathrm{Gibbs})}(t,t') \\ \times \cos\Big(\mu^{(q)\mathsf{T}}(t)t - \mu^{(q)\mathsf{T}}(t')t'\Big).$$

Generalised Spectral Mixture Kernel (4)

- GSMK replaces the EQs with Gibbs kernels:

$$k^{(\mathrm{GSMK})}(t,t') = \sum_{q=1}^{Q} w^{(q)}(t) w^{(q)}(t') k_q^{(\mathrm{Gibbs})}(t,t') \times \cos\Big(\mu^{(q)\mathsf{T}}(t)t - \mu^{(q)\mathsf{T}}(t')t'\Big).$$

- $(w^{(q)}, \ell^{(q)} \mu^{(q)})_{q=1}^{Q}$ given log-GP priors.

Generalised Spectral Mixture Kernel (4)

- GSMK replaces the EQs with Gibbs kernels:

$$k^{(\mathrm{GSMK})}(t,t') = \sum_{q=1}^{Q} w^{(q)}(t) w^{(q)}(t') k_q^{(\mathrm{Gibbs})}(t,t') \\ \times \cos\Big(\mu^{(q)\mathsf{T}}(t)t - \mu^{(q)\mathsf{T}}(t')t'\Big).$$

- $(w^{(q)}, \ell^{(q)} \mu^{(q)})_{q=1}^{Q}$ given log-GP priors.
- Estimated using MAP.

- Equivalent generative model as truncated Fourier series:

$$
\begin{aligned}
f^{(\mathrm{GSMK})}(t) = \sum_{q=1}^{Q} w^{(q)}(t)(c_1^{(q)}(t)\cos(\mu^{(q)\mathsf{T}}(t)t) \\
+ c_2^{(q)}(t)\sin(\mu^{(q)\mathsf{T}}(t)t)),
\end{aligned}
$$

$$
c_1^{(q)}, c_2^{(q)} \sim \mathcal{GP}(0, k^{(\mathrm{Gibbs})}(t,t')).
$$

Gaussian Process Convolution Model

- SMK and extensions assume **parametric** model.

Gaussian Process Convolution Model

- SMK and extensions assume **parametric** model.
- More flexible to use **nonparametric** model:

$$s(\omega) = |\hat{h}(\omega)|^2.$$

Gaussian Process Convolution Model

- SMK and extensions assume **parametric** model.
- More flexible to use **nonparametric** model:

$$s(\omega) = |\hat{h}(\omega)|^2.$$

- Inverse Fourier transform gives kernel:

$$k(t, t') = \int_{-\infty}^{\infty} h(t - z)h(t' - z)\,\mathrm{d}z = h * R(h)(t - t').$$

- SMK and extensions assume **parametric** model.
- More flexible to use **nonparametric** model:

$$s(\omega) = |\hat{h}(\omega)|^2.$$

- Inverse Fourier transform gives kernel:

$$k(t, t') = \int_{-\infty}^{\infty} h(t - z)h(t' - z)\,\mathrm{d}z = h * R(h)(t - t').$$

- GPCM (Tobar et al., 2015) models $h \sim \mathcal{GP}(0, k_h)$.

Gaussian Process Convolution Model

- SMK and extensions assume **parametric** model.
- More flexible to use **nonparametric** model:

$$s(\omega) = |\hat{h}(\omega)|^2.$$

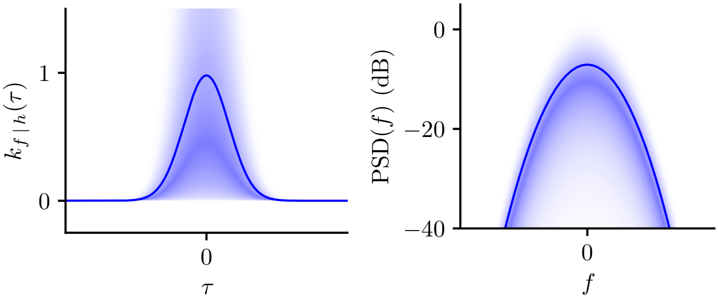
- Inverse Fourier transform gives kernel:

$$k(t, t') = \int_{-\infty}^{\infty} h(t - z)h(t' - z)\,\mathrm{d}z = h * R(h)(t - t').$$

- GPCM (Tobar et al., 2015) models $h \sim \mathcal{GP}(0, k_h)$.
  - $\int_{-\infty}^{\infty} k_h(t, t)\,\mathrm{d}t < \infty$ (finite trace).

Gaussian Process Convolution Model (2)

- Nonparametric prior over kernels and PSDs.



Gaussian Process Convolution Model (3)

- Interpretation as linear system:

Gaussian Process Convolution Model (3)

- Interpretation as linear system:

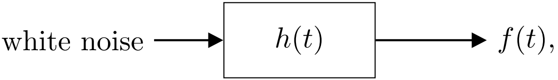
$$\text{white noise} \longrightarrow \boxed{h(t)} \longrightarrow f(t),$$

$$\text{white noise} \sim \mathcal{GP}(0, \delta(t - t')),$$

$$h \sim \mathcal{GP}(0, k_h).$$

Gaussian Process Convolution Model (3)

- Interpretation as linear system:



$$\text{white noise} \sim \mathcal{GP}(0, \delta(t - t')),$$

$$h \sim \mathcal{GP}(0, k_h).$$

- Inference complicated.

Conclusion

- Instead of designing kernel, design PSD.

Conclusion

- Instead of designing kernel, design PSD.
- Parametric approaches:
    - line spectrum (SSA),
    - mixture of Gaussians (SMK, MOSMK, GSMK).

Conclusion

- Instead of designing kernel, design PSD.
- Parametric approaches:
    - line spectrum (SSA),
    - mixture of Gaussians (SMK, MOSMK, GSMK).
- Nonparametric approach also possible (GPCM).