We introduced the squared exponential (SE) Gaussian process prior on $S^1$ last time, where $S^1$ is the unit circle: 
\[ \{W_t : t \in S^1\} \text{ with the squared exponential kernel function} \]
\[ K(t, t') = \exp(-a^2 \|t - t'\|^2). \]

Through the map $Q : [0, 1] \rightarrow S^1, \theta \rightarrow (\cos 2\pi \theta, \sin 2\pi \theta)$ (MacKay, 1998), we obtain an equivalent GP on the unit interval: 
\[ \{W_\theta : \theta \in [0, 1]\} \text{ with the kernel } G_a(\theta, \theta') \]

\[
G_a(\theta, \theta') = \exp(-a^2 \{ (\cos 2\pi \theta - \cos 2\pi \theta')^2 + (\sin 2\pi \theta - \sin 2\pi \theta')^2 \}) \\
= \exp\{-4a^2 \sin^2(\pi \theta - \pi \theta')\},
\]

where $\theta, \theta' \in [0, 1]$.

We call $G_a(\cdot, \cdot)$ on the unit interval as *squared exponential periodic* (SEP) kernel.
Karhunen–Loève decomposition of $G_a$

The Karhunen–Loève expansion of the covariance kernel is

$$G_a(t, t') = \sum_{k=1}^{\infty} v_k(a) \psi_k(t) \psi_k(t'),$$

where the eigenvalues are given by

$$v_1(a) = e^{-2a^2 I_0(2a^2)}, \quad v_{2j}(a) = v_{2j+1}(a) = e^{-2a^2 I_j(2a^2)}, \quad j \geq 1,$$

with eigenfunctions $\psi_j(t)$, $j = 1, 2, \ldots$ given by the Fourier basis functions $\{1, \cos 2\pi t, \sin 2\pi t, \ldots\}$ in that order.

- The function $I_n(x)$ is the modified Bessel function of the first kind with order $n$ and argument $x$ where $n \in \mathbb{Z}$ and $x \in \mathbb{R}$.
- $I_n(x)$ is a library function in most software, such as `besselI` in R language.
- The analytical decomposition $G_a$ allows for efficient covariance matrix inversion.
If a curve $\gamma(\theta) \sim \text{GP}(\mu(\theta), G_a(\cdot, \cdot)/\tau)$, we then have the equivalent representation

$$
\gamma(\theta) = \mu(\theta) + \sum_{k=1}^{\infty} z_k \psi_k(\theta),
$$

where $z_k \sim N(0, v_k(a)/\tau)$ independently.

In practice, we truncate the infinite sum by some finite number. Why?

- General result: $I_n(2x) \leq I_0(2x)(2x)^n/n!$. Therefore, the eigenvalues $v_k(a)$’s decay very fast for fixed $a$.
- When $a$ increases, the smoothness level of the kernel decreases. We typically do not use values as large as 100 since then the kernel becomes very close to the identity matrix and thus the resulting prior paths become very rough.
- Consequently, some suitable finite order truncation of the Karhunen-Loève expansion is able to approximate the kernel function well.
Figure 1(a) shows that eigenvalues decay very fast at fixed $a$.

- We shall use $\text{PVE}_a = \sum_{k=1}^{L} v_k(a) / \sum_{k=1}^{\infty} v_k(a)$ the percentage of variance explained by the first $L$ basis functions to select $L$. 
The denominator in PVE is 1, i.e.,
\[ \sum_{k=1}^{\infty} v_k(a) = e^{-2a^2} \sum_{k=-\infty}^{\infty} I_k(2a^2) = 1 \]
according to a property of modified Bessel functions of the first kind.

Consider \( L = 2J + 1 \) in view of that eigenfunctions are paired up except for the first one.

Figure 1(b) shows that with \( J = 10 \), we are able to explain at least 98% of all the variability for a reasonable range of \( a \)'s from 0 to 10.

In practice, one can specify a level of PVE and a range of \( a \) (particularly an upper bound) to select \( L \).
Let $\Psi$ be the $n$ by $L$ matrix with the $k$th column comprising of the evaluations of $\psi_k(\cdot)$ at the components of $\theta$, and $\mu$ comprising of the evaluations of $\mu(\cdot)$ at the components of $\theta$.

Then the SEP Gaussian process prior can be expressed as

$$\gamma = \Psi z + \mu; \quad z \sim N(0, \Sigma_a / \tau)$$

where $\Sigma_a = \text{diag}(v_1(a), \ldots, v_L(a))$.

The inverse of $\Sigma_a$ takes $O(L)$ time as it is a diagonal matrix.

A fully Bayes approach uses priors for the hyper-parameters involved in the covariance kernel, for example, Gamma priors for $\tau$ and $a$.

SEP GPs do not require explicit matrix inverse, thus they are numerically much more appealing than SE GPs.
• Any concerns?
Uniform credible bands

- Pointwise credible bands are based on quantiles of MCMC samples at each point.
- One advantage of the Bayesian framework is that it leads to various summaries for inference, all based on the joint posterior distributions (either analytically tractable or through MCMC samples).
- One may also construct a variable-width *uniform* credible band. Specifically:
  - Let \( \{ \gamma_i(\theta) \}_{i=1}^M \) be the posterior samples and \((\hat{\gamma}(\theta), \hat{s}(\theta))\) be the posterior mean and standard deviation functions derived from \( \{ \gamma_i(\theta) \} \).
  - For each MCMC run, we calculate the distance
    \[
    u_i = \| (\gamma_i - \hat{\gamma}) / \hat{s} \|_{\infty} = \sup_{\theta} \{ |\gamma_i(\theta) - \hat{\gamma}(\theta)| / \hat{s}(\theta) \}.
    \]
  - Obtain the 95th percentile of all the \( u_i \)'s, denoted as \( L_0 \).
  - Then a 95% uniform credible band is given by
    \[
    [\hat{\gamma}(\theta) - L_0 \hat{s}(\theta), \hat{\gamma}(\theta) + L_0 \hat{s}(\theta)].
    \]