2014 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 21–24, 2014, REIMS, FRANCE

GAUSSIAN QUADRATURES FOR STATE SPACE APPROXIMATION OF SCALE MIXTURES OF SQUARED EXPONENTIAL COVARIANCE FUNCTIONS

Arno Solin and Simo Särkkä

Aalto University
Department of Biomedical Engineering and Computational Science
Rakentajanaukio 2 C, FI-02150 Espoo, Finland

ABSTRACT

Stationary one-dimensional Gaussian process models in machine learning can be reformulated as state space equations. This reduces the cubic computational complexity of the naive full GP solution to linear with respect to the number of training data points. For infinitely differentiable covariance functions the representation is an approximation. In this paper, we study a class of covariance functions that can be represented as a scale mixture of squared exponentials. We show how the generalized Gauss–Laguerre quadrature rule can be employed in a state space approximation in this class. The explicit form of the rational quadratic covariance function approximation is written out, and we demonstrate the results in a regression and log-Gaussian Cox process study.

Index Terms— Gaussian process, state space model, rational quadratic covariance function, Gaussian quadrature

1. INTRODUCTION

One-dimensional (typically temporal) Gaussian processes (GPs, [1]) are a central part of both signal processing and statistical machine learning. In signal processing they are typically represented as state space models [2–4], whereas the kernel (covariance function) formalism is favored in machine learning. The link between these two representations is interesting, because it enables the combination of the intuitive model specification from machine learning with computationally efficient signal processing methods. Most notably, this reduces the computational cost of a naive GP regression solution from $\mathcal{O}(N^3)$ to $\mathcal{O}(N)$ in the number of data points N by solving the state space inference problem by Kalman filtering methods (see, *e.g.*, [5]).

Certain classes of covariance functions lend themselves well to the state space formulation. Most notably, the Matérn class of covariance functions [1, 6] can be directly converted into state space form for half-integer values of the smoothness parameter. This link is discussed in [2, 3], where also a Taylor approximation scheme for the squared exponential (also known as exponentiated quadratic, Gaussian, or RBF) covariance function is given. Other rational function approximations to the squared exponential are discussed in [7]. Furthermore, a complementary approach for converting periodic covariance functions to state space models was recently proposed in [4]. The approach also extends to spatio-temporal modeling (see, *e.g.*, [3, 8, 9]) and non-Gaussian likelihoods (see [10] and Sec. 4.2).

In this paper, we focus on a class of covariance functions that can be represented as a scale mixture of squared exponential covariance functions (see [1, 11]). This class of covariance functions is infinitely differentiable and is therefore not well suited for the state space representation (an exact representation would require an infinite number of state variables). As will be shown, the scale mixture integral can, however, be approximated by the Gauss–Laguerre quadrature rule [12], and thus a state space approximation can be constructed as a superposition of scaled and weighted state space models for the squared exponential.

This paper is organized as follows. In Section 2 we briefly cover the connection between the kernel and state space formalisms of Gaussian processes. In Section 3 we consider a class of covariance functions that can be written as a scale mixture of squared exponentials, and write down the explicit approximation to the rational quadratic (RQ) covariance. We also analyze the convergence and error of this approximation. We present two example studies in Section 4, a simulated regression example and a log-Gaussian Cox process model with empirical data. Finally the results are discussed.

2. GAUSSIAN PROCESSES IN MACHINE LEARNING

Gaussian processes can be used as flexible priors in Bayesian statistical machine learning. The model function f is assumed to be a realization of a Gaussian random process prior

$$f(t) \sim \mathcal{GP}(0, k(t, t')), \tag{1}$$

This work was supported by grants from the Academy of Finland (266940, 273475) and the Finnish Funding Agency for Technology and Innovation (40304/09).

978-1-4799-3694-6/14/$31.00

where the covariance function $k(t, t')$ encodes the prior assumptions of the process into the model.

For stationary models, we introduce the following one-argument notation: $k(t, t,) \triangleq k(t - t')$. In this case, the covariance function is the inverse Fourier transform of the corresponding spectral density $S(\omega)$:

$$k(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \exp(-\mathrm{i}\,\omega\tau)\, \mathrm{d}\omega. \tag{2}$$

Instead of directly working with the kernel formalism of the Gaussian process $f(t)$, certain classes of one-dimensional covariance functions allow to work with the mathematical dual [3], where the Gaussian process is constructed as a solution to an Mth order linear stochastic differential equation (SDE). The corresponding inference problem can be solved with Kalman filtering type of methods [13], where the computational complexity is $\mathcal{O}(M^3 N)$. If the number of observations $N \gg M$, as typically is the case in temporal modeling, this formulation is very beneficial.

The state space model corresponding to the GP formulation (1) can be given as

$$\begin{aligned} \frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} &= \mathbf{F}\mathbf{f}(t) + \mathbf{L}\mathbf{w}(t), \\ f(t_k) &= \mathbf{H}\mathbf{f}(t_k), \end{aligned} \tag{3}$$

where $\mathbf{f}(t) = \big(f_1(t), f_2(t), \ldots, f_M(t)\big)^\mathsf{T}$ holds the M stochastic processes, and $\mathbf{w}(t)$ is a multi-dimensional white noise process with spectral density $\mathbf{Q}_\mathrm{c}$. The model is defined by the feedback matrix $\mathbf{F}$ and the noise effect matrix $\mathbf{L}$.

The continuous-time linear time-invariant model (3) can be solved for discrete points. The initial state $\mathbf{f}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_\infty)$ is defined by the stationary covariance $\mathbf{P}_\infty$ that is the solution to the corresponding Lyapunov equation:

$$\frac{\mathrm{d}\mathbf{P}_\infty}{\mathrm{d}t} = \mathbf{F}\mathbf{P}_\infty + \mathbf{P}_\infty\mathbf{F}^\mathsf{T} + \mathbf{L}\mathbf{Q}_\mathrm{c}\mathbf{L}^\mathsf{T} = \mathbf{0}. \tag{4}$$

The solution to (3) can be written out in closed-form at the specified time points, and it is given as

$$\mathbf{f}_{k+1} = \mathbf{A}_k\mathbf{f}_k + \mathbf{q}_k, \quad \mathbf{q}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k), \tag{5}$$

where $\mathbf{f}(t_k) = \mathbf{f}_k$, and the state transition and process noise covariance matrices can be solved analytically (see, *e.g.*, [3]):

$$\mathbf{A}_k = \mathbf{\Phi}(\Delta t_k), \tag{6}$$

$$\begin{aligned} \mathbf{Q}_k &= \int_0^{\Delta t_k} \mathbf{\Phi}(\Delta t_k - \tau)\mathbf{L}\mathbf{Q}_\mathrm{c}\mathbf{L}^\mathsf{T}\mathbf{\Phi}(\Delta t_k - \tau)^\mathsf{T}\, \mathrm{d}\tau \\ &= \mathbf{P}_\infty - \mathbf{A}_k\mathbf{P}_\infty\mathbf{A}_k^\mathsf{T}, \end{aligned} \tag{7}$$

where $\Delta t_k = t_{k+1} - t_k$ and $\mathbf{\Phi}(\tau) = \exp(\mathbf{F}\tau)$ is the matrix exponential of the feedback matrix. In (7) we have used the method by [14], which is not generally well known and provides a very efficient way of solving the stationary SDE in practice.

In the state space form, the spectral density $S(\omega)$ of $f(t)$ can be written using the SDE matrices as

$$S(\omega) = \mathbf{H}(\mathbf{F} - \mathrm{i}\,\omega\mathbf{I})^{-1}\mathbf{L}\mathbf{Q}_\mathrm{c}\mathbf{L}^\mathsf{T}\left[(\mathbf{F} + \mathrm{i}\,\omega\mathbf{I})^{-1}\right]^\mathsf{T}\mathbf{H}^\mathsf{T}. \tag{8}$$

The covariance function can also be recovered by considering the following forward and backward time models:

$$k(\tau) = \begin{cases} \mathbf{H}\mathbf{P}_\infty\mathbf{\Phi}(\tau)^\mathsf{T}\mathbf{H}^\mathsf{T}, & \text{if } \tau \geq 0, \\ \mathbf{H}\mathbf{\Phi}(-\tau)\mathbf{P}_\infty\mathbf{H}^\mathsf{T}, & \text{if } \tau < 0. \end{cases} \tag{9}$$

The problem in exploiting the link between the two representations is to come up with the actual form of the model matrices $\mathbf{F}, \mathbf{L}, \mathbf{Q}_\mathrm{c}, \mathbf{H}$, and $\mathbf{P}_\infty$ corresponding to the particular covariance function.

3. GAUSSIAN QUADRATURE APPROXIMATION FOR SCALE MIXTURES OF COVARIANCES

We study a class of covariance functions (*i.e.* kernels) which can be written as superpositions of squared exponential kernels. We start of by defining how the squared exponential can be written out, and then consider the scale mixture expansion defining the class of smooth kernels. The scale mixture can be approximated by the Gauss–Laguerre quadrature. Finally, we use this technique to come up with a state space approximation for the rational quadratic covariance function.

3.1. A general class of scale mixture covariance functions

As is discussed in [1, 11], we can construct a general class of stationary kernels by writing them as superpositions of squared exponential covariance functions with a distribution $p(\ell)$ over the length-scales ℓ. This gives rise to the following *scale mixture* formulation

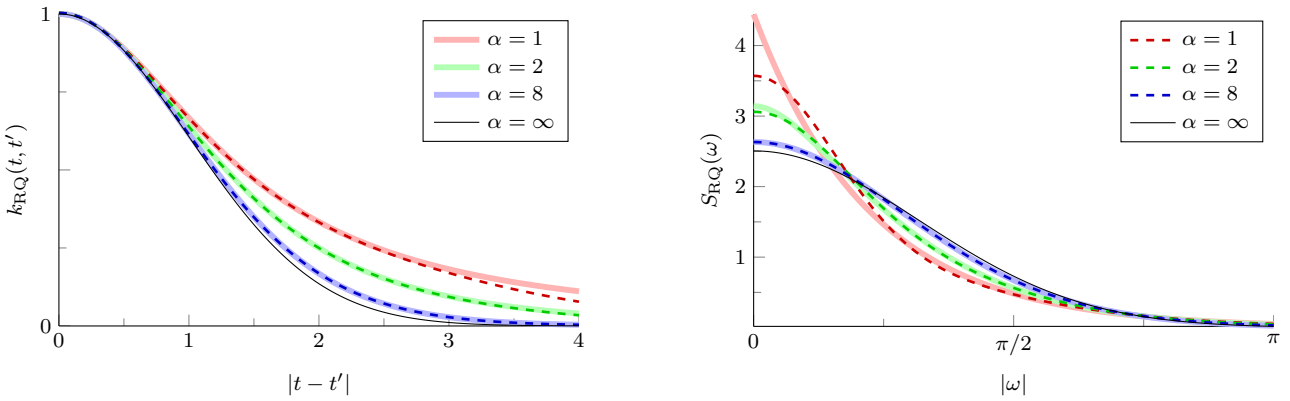
$$k_\mathrm{SM}(t) = \int_0^\infty p(\ell)\, k_\mathrm{SE}(t \mid \ell)\, \mathrm{d}\ell, \tag{10}$$

where $k_\mathrm{SE}(t \mid \ell)$ denotes the squared exponential kernel (11) with length-scale ℓ. Covariance functions of this class include, for example, the *rational quadratic* and the *Cauchy covariance functions*. The squared exponential can be recovered by imposing a delta function on $p(\ell)$.

We parametrize the squared exponential as in [1] such that

$$k_\mathrm{SE}(t) = \sigma^2 \exp\!\left(-\frac{t^2}{2\,\ell^2}\right), \tag{11}$$

where $\ell, \sigma^2 > 0$ are the characteristic length-scale and magnitude parameters. The corresponding stochastic process is infinitely many times mean square differentiable, and thus also (10) corresponds to infinitely smooth processes.



(a) Covariance function

(b) Spectral density

Fig. 1: Approximations to the rational quadratic covariance function with different shape parameters α. The degree of approximation was $n = 6$ (quadrature degree) and $m = 6$ (state space). The thick solid lines show the exact values and the dashed lines denote the approximations for each α. The thin solid line shows the values for the squared exponential covariance function.

Hartikainen and Särkkä [2] have presented a state space approximation for this covariance function by approximating the corresponding rational power spectrum

$$S_{\mathrm{SE}}(\omega) = \sigma^2 \sqrt{2\pi}\, \ell \, \exp\left(-\frac{\ell^2 \omega^2}{2}\right), \tag{12}$$

by a truncated Taylor expansion of $1/S(\omega)$ around origin. For details on what the actual state space model matrices $\mathbf{F}, \mathbf{L}, \mathbf{Q}_{\mathrm{c}}$, and $\mathbf{H}$ look like, see [2]. The state space approximation is not infinitely differentiable, but already a truncated approximation by a six-dimensional state gives good empirical results (see, *e.g.*, [7]).

3.2. The generalized Gauss–Laguerre quadrature

When applicable, solving the integral equation (10) gives the corresponding covariance function (*e.g.*, the rational quadratic in Sec. 3.3). However, the scale mixture integral can also be directly approximated in terms of the Gauss–Laguerre quadrature rule (see, *e.g.*, [12, 15]).

The generalized Gauss–Laguerre quadrature rule approximates integrals of the following form:

$$\int_0^\infty x^\gamma e^{-x} f(x)\, \mathrm{d}x \approx \sum_{i=1}^n w_i\, f(x_i), \tag{13}$$

where $x^\gamma \exp(-x)$ is the weight function and $f(x)$ the integrand.

The quadrature rule is given by the abscissae x_i and corresponding weights w_i, $i = 1, 2, \ldots, n$, where the abscissae are given by the roots of the generalized Laguerre polynomial $\mathrm{L}_n^\gamma(x)$ of degree n [12], and the weights are given as:

$$w_i = \frac{\Gamma(n + \gamma + 1)\, x_i}{n!\, (n+1)^2 \left[\mathrm{L}_{n+1}^\gamma(x_i)\right]^2}, \tag{14}$$

where $\Gamma(\cdot)$ is the gamma function.

3.3. The rational quadratic covariance function

The rational quadratic (RQ) covariance function (see, *e.g.*, [1, 6]) is of the form:

$$k_{\mathrm{RQ}}(t) = \sigma^2 \left(1 + \frac{t^2}{2\alpha\ell^2}\right)^{-\alpha}, \tag{15}$$

where $\sigma^2, \ell, \alpha > 0$. The shape parameter α defines the decay of the tail, and includes the Cauchy covariance as a special case at $\alpha = 1$. As $\alpha \to \infty$, the covariance function converges to the squared exponential.

The RQ covariance function is stationary, and there exists a corresponding spectral density. The spectral density can be obtained by Fourier transforming the above expression for the covariance. The spectral density expression is given by

$$S_{\mathrm{RQ}}(\omega) = \sigma^2 \sqrt{\pi\alpha}\, \ell\, \frac{2^{2-\alpha}}{\Gamma(\alpha)} \times \left(\sqrt{2\alpha}\, \ell\, |\omega|\right)^{\alpha - \frac{1}{2}} \mathrm{K}_{\alpha - \frac{1}{2}}\left(\sqrt{2\alpha}\, \ell\, |\omega|\right), \tag{16}$$

where $\mathrm{K}_\nu(\cdot)$ denotes the modified Bessel function of the second kind.

The rational quadratic covariance bears a lot of resemblance to the Matérn family of covariance functions (see [1]). The rational quadratic family complements the Matérn family as a sort of spectral dual.

The rational quadratic covariance function can be seen as a scale mixture of squared exponentials with different characteristic length-scales. Following [1], we parametrize the squared exponential in terms of inverse squared length-scales, $\xi = \ell_{\mathrm{SE}}^{-2}$, and put a gamma distribution prior on ξ:

$$p(\xi \mid \alpha, \beta) \propto \xi^{\alpha - 1} \exp(-\alpha\, \xi / \beta). \tag{17}$$

The scale mixture form (10) now gives the integral representation:

$$k_{\mathrm{RQ}}(t) = \int_0^\infty p(\xi \mid \alpha, \beta)\, k_{\mathrm{SE}}(t \mid \xi)\, \mathrm{d}\xi \tag{18}$$
$$= \frac{\sigma_{\mathrm{SE}}^2}{\Gamma(\alpha)} \left(\frac{\beta}{\alpha}\right)^{-\alpha} \times \int_0^\infty \xi^{\alpha-1} \exp\!\left(-\frac{\alpha\,\xi}{\beta}\right) \exp\!\left(-\frac{\xi\, t^2}{2}\right) \mathrm{d}\xi,$$

where we do a change of variables in order to match the Gauss–Laguerre formulation in (13) such that

$$= \frac{\sigma_{\mathrm{SE}}^2}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} \exp(-x) \exp\!\left(-\frac{\beta\, x\, t^2}{2\,\alpha}\right) \mathrm{d}x$$
$$\approx \frac{\sigma_{\mathrm{SE}}^2}{\Gamma(\alpha)} \sum_{i=1}^{n} w_i \exp\!\left(-\frac{\beta\, x_i\, t^2}{2\,\alpha}\right). \tag{19}$$

Now we may re-parametrize $\beta^{-1} = \ell_{\mathrm{RQ}}^2$. The final approximation to the rational quadratic covariance function (parameterized in terms of σ_{RQ}^2, ℓ_{RQ}, and α) is of the form:

$$k_{\mathrm{RQ}}(t) \approx \sum_{i=1}^{n} k_{\mathrm{SE}}(t \mid \sigma_i^2, \ell_i), \tag{20}$$

where the squared exponentials are evaluated with magnitudes $\sigma_i^2 = \sigma_{\mathrm{RQ}}^2\, w_i/\Gamma(\alpha)$ and length-scales $\ell_i^2 = \ell_{\mathrm{RQ}}^2 \alpha / x_i$. The quadrature points and weights are given such that $x_i, i = 1, 2, \ldots, n$ are the roots of the generalized Laguerre polynomial $\mathrm{L}_n^{\alpha-1}(x)$, and the weights w_i are given as:

$$w_i = \frac{\Gamma(n+\alpha)\, x_i}{n!\,(n+1)^2 \left[\mathrm{L}_{n+1}^{\alpha-1}(x_i)\right]^2}. \tag{21}$$

Now in terms of state space models, this means that if we are able to approximate the squared exponential covariance function by a state space model, the rational quadratic can be approximated as a sum of such models by suitably coupling the magnitude and length-scale hyperparameters σ_i^2 and ℓ_i.

The corresponding state space model can now be formed by a sum of state space models approximating the squared exponential covariance function (see [7] for ways of forming different approximations). Summing covariance functions corresponds to stacking state space models such that the feedback matrix is a block-diagonal matrix of the distinct feedback matrices. Figure 1 shows the covariance function and spectral density of the rational quadratic covariance for different values of α. The thick solid lines correspond to the exact values for given α, and the dashed lines are the values corresponding to a state space approximation with $m = 6$ (state dimension of the squared exponential using a Taylor expansion) and $n = 6$ (quadrature approximation degree). The full state dimension is thus $M = nm$. For $\alpha = 1$, the approximation has trouble capturing the long tail, whereas for larger values of alpha, the lines are practically indistinguishable.

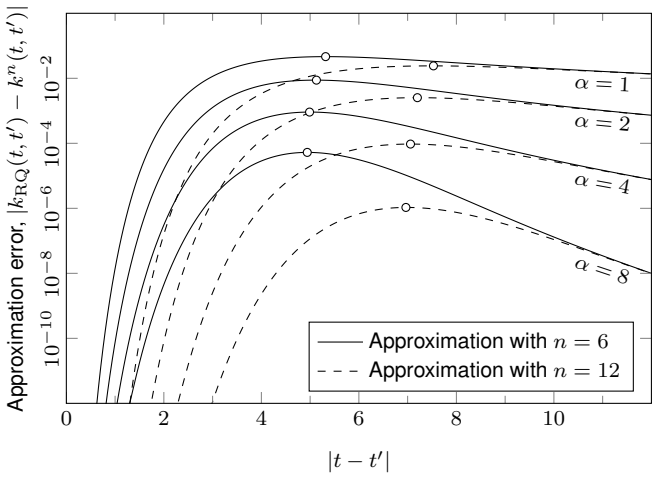


Fig. 2: The approximation error between the exact rational quadratic covariance function and the quadrature approximation of it, with $\alpha = 1, 2, 4, 8$, $n = 6, 12$, $\ell_{\mathrm{RQ}} = 1, \sigma_{\mathrm{RQ}}^2 = 1$ (*cf.* Fig. 1a). The maximum is show by the point markers (∘).

3.4. Convergence and computational aspects

The quadrature form of the RQ covariance in Equation (20) is an approximation that depends on the degree n of the quadrature and on the hyperparameter values of α, ℓ_{RQ}, and σ_{RQ}^2.

Lemma 3.1. *Following the results by [16], the convergence of the quadrature is secured for any integrable function satisfying the condition*

$$|f(y)| \leq \frac{e^y}{y^{\gamma+1+\rho}}, \quad \textit{for some } \rho > 0, \tag{22}$$

for all sufficiently large values of y. For a proof, see [16].

Theorem 3.2. *Let $k^{nm}(t)$ denote the approximation of $k_{\mathrm{RQ}}(t)$ with n quadrature points and an mth order state space model, and let $k^n(t)$ denote an approximation with n quadrature points and the exact SE covariance function. The overall error,*

$$|k(t) - k^{nm}(t)| = |k(t) - k^n(t) + k^n(t) - k^{nm}(t)|$$
$$\leq |k(t) - k^n(t)| + |k^n(t) - k^{nm}(t)|, \tag{23}$$

converges to zero as $n, m \to \infty$.

Proof. The integrand $f(\cdot)$ in Equation (19) may be written as $f(x) = \exp(-\beta\, t^2\, x/(2\alpha))$, and the convergence criterion from Lemma 3.1, $|f(x)| \leq e^x\, x^{-\alpha-\rho}$, is satisfied for any positive x and ρ. Thus the first part of (23) converges pointwise. Because $\sum_{i=1}^n \sigma_i^2 = \sigma_{\mathrm{RQ}}^2$, by Theorem 3.1 of [7] the approximations of the SE covariance function converge uniformly, and thus the second part converges pointwise. □

As seen in Figure 1a, the approximation seems to diverge from the true covariance curve as $|t - t'|$ grows. This is, however, not the case. Figure 2 shows the absolute error between the approximate covariance function from (20).

4. EXPERIMENTS

As a first example we compare the results given by the state space approximation model against the naive full GP regression solution by using the rational quadratic covariance function, and we show that the results are practically equal. As a more elaborate example we consider a log-Gaussian Cox process model with a Poisson likelihood.

4.1. A simulated example study

In Gaussian process regression we want to predict an unknown scalar output $f(t_*)$ associated with a known input $t_* \in \mathbb{R}$, given a set of training data $\mathcal{D} = \{(t_k, y_k) \mid k = 1, 2, \ldots, N\}$. The model functions $f(t)$ are seen as realizations of a Gaussian process prior with observations corrupted by Gaussian noise:

$$\begin{aligned} f(t) &\sim \mathcal{GP}(0, k(t, t')), \\ y_k &= f(t_k) + \varepsilon_k, \end{aligned} \tag{24}$$

where $\varepsilon_k \sim \mathcal{N}(0, \sigma_{\text{n}}^2)$. The solution to the GP regression problem can be computed in closed-form [1] such that

$$\begin{aligned} \mathbb{E}[f(t_*)] &= \mathbf{k}_*^{\mathsf{T}} (\mathbf{K} + \sigma_{\text{n}}^2 \mathbf{I})^{-1} \mathbf{y}, \\ \mathbb{V}[f(t_*)] &= k(t_*, t_*) - \mathbf{k}_*^{\mathsf{T}} (\mathbf{K} + \sigma_{\text{n}}^2 \mathbf{I})^{-1} \mathbf{k}_*, \end{aligned} \tag{25}$$

where the predictions are given for $p(f(t_*) \mid t_*, \mathcal{D}) = \mathcal{N}(\mathbb{E}[f(t_*)], \mathbb{V}[f(t_*)])$. However, the direct solution to the GP regression problem scales as $\mathcal{O}(N^3)$ which is due to the inversion of the $N \times N$ matrix in (25).

Therefore, as the number of data points N becomes large, using the state space methodology becomes appealing. This is possible if the model in (24) can be written in the SDE form as explained in Section 2. The inference problem can then be solved by Kalman filtering and smoothing methods in $\mathcal{O}(NM^3)$ time complexity, where M is the full state dimensionality.

We consider a small simulated data set of $N = 32$ data points from a sinc function, where the measurements have been corrupted by zero-mean Gaussian noise. We train both a full GP model (24) and a state space model for this regression problem, and compare the results. We use a rational quadratic covariance function with $\alpha = 1$ and optimize the hyperparameters σ_{n}^2, σ_{RQ}^2 and ℓ_{RQ} with respect to marginal likelihood (see, *e.g.*, [1, 4] for details).

Figure 3 shows the state space inference outcome for the GP regression problem with the rational quadratic covariance function. The shaded region marks the 95% confidence interval. The naive full GP solution corresponding to (25) is shown by the dashed red lines, and they agree with the state space approximation result. The degree of the approximations were $n = 6$ and $m = 6$ leading to state dimension $M = nm$.

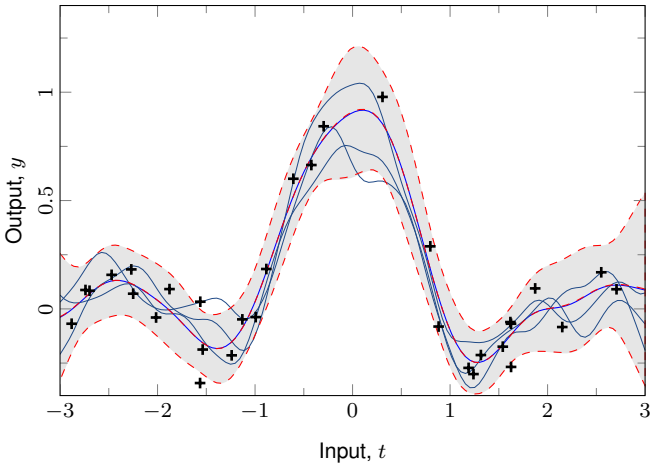


Fig. 3: GP regression results for simulated data (shown by the point markers). The state space mean and 95% confidence interval estimates are shown by the solid blue line and the grey patch. The corresponding full GP regression results is shown by dashed red lines. The thin solid lines are random draws from the state space posterior.

4.2. A temporal log-Gaussian Cox process

As a second example we consider a log-Gaussian Cox process, which is an inhomogeneous Poisson point process model with an unknown intensity function $\lambda(t)$. The log-intensity is modeled as a GP such that $f(t) = \log \lambda(t)$ (see, *e.g.*, [17]). We use the method proposed by Møller *et al.* [17], where the data range is discretized into N intervals and a locally constant intensity is assumed in each interval. The model is thus a GP model with a Poisson likelihood:

$$\begin{aligned} f(t) &\sim \mathcal{GP}(0, k(t, t')) \\ p(\mathcal{D} \mid f) &= \prod_{k=1}^{N} \text{Poisson}(y_k \mid \exp(f(t_k))), \end{aligned} \tag{26}$$

where $t_k, k = 1, 2, \ldots, N$, denotes the coordinate of the kth interval and y_k the number of incidents in the interval. As the model is now non-Gaussian we have to resort to the Laplace approximation (finding the mode by a Newton scheme and forming a Gaussian approximation in the mode, see [1, 10]) for doing the updates inside the Kalman filter.

The data[1] contain the dates of 191 coal mine explosions that killed ten or more men in Britain between years 1851 and 1962, where the RQ assumptions of smoothness and long-range correlations are justified. We use a grid discretization of $N = 1024$ points, a RQ covariance function with $\alpha = 1$, and optimize the model hyperparameters with respect to marginal likelihood. Figure 4 shows the estimated intensity curve for the log-Gaussian Cox process with a shaded 90% confidence region.

[1] The data set is available as a part of the GPSTUFF software package: `http://becs.aalto.fi/en/research/bayes/gpstuff/`

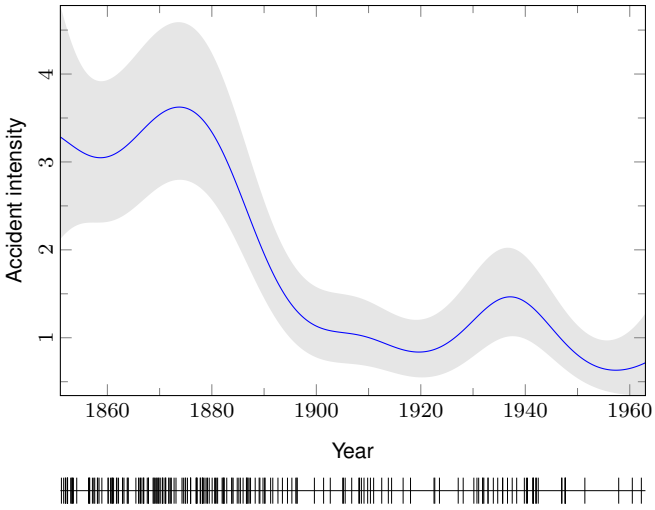


Fig. 4: Coal mining accident data, with $N = 1024$ intervals and 191 incidents. The bar shows the actual incidents, and the modeling outcome for the intensity in the log-Gaussian Cox process model with an approximate 90% confidence region is shown in the figure above.

We argue that the state space approximation can be beneficial in this type of modeling, as the interval can be discretized into a very dense grid without running into computational limitations. Dense grids are appealing in this type of modeling, as it is known [18] that this approximation reaches posterior consistency in the limit of widths of the intervals going to zero.

5. CONCLUSION AND DISCUSSION

In this paper we have shown how certain types of covariance functions (most notably the rational quadratic covariance) that are constructed as scale mixtures of the squared exponential covariance function can be approximated by a Gaussian quadrature rule. In state space estimation, this enables us to use existing methods for conversion of the squared exponential covariance functions in approximating the rational quadratic covariance function.

In Section 3.3, we have written down approximation to the rational quadratic covariance function, and the convergence of this approximation is analyzed. Furthermore, the results in the experiments section showed that this state space approximation is useful in practice both in GP regression (Sec. 4.1) and GP modeling in a more general setting (Sec. 4.2).

6. REFERENCES

[1] Carl Edward Rasmussen and Christopher K.I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.

[2] Jouni Hartikainen and Simo Särkkä, "Kalman filtering and smoothing solutions to temporal Gaussian process regression models," in *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2010.

[3] Simo Särkkä, Arno Solin, and Jouni Hartikainen, "Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 51–61, 2013.

[4] Arno Solin and Simo Särkkä, "Explicit link between periodic covariance functions and state space models," in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, 2014, vol. 33 of *JMLR W&CP*, pp. 904–912.

[5] Simo Särkkä, *Bayesian Filtering and Smoothing*, vol. 3 of *Institute of Mathematical Statistics Textbooks*, Cambridge University Press, 2013.

[6] Bertil Matérn, "Spatial variation: Stochastic models and their applications to some problems in forest surveys and other sampling investigations," *Meddelanden från statens skogsforskningsinstitut*, vol. 49, pp. 1–144, 1960.

[7] Simo Särkkä and Robert Piché, "On convergence and accuracy of state-space approximations of squared exponential covariance functions," in *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2014.

[8] Simo Särkkä and Jouni Hartikainen, "Infinite-dimensional Kalman filtering approach to spatio-temporal Gaussian process regression," in *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, 2012, vol. 22 of *JMLR W&CP*, pp. 993–1001.

[9] Arno Solin and Simo Särkkä, "Infinite-dimensional Bayesian filtering for detection of quasiperiodic phenomena in spatiotemporal data," *Physical Review E*, vol. 88, pp. 052909, 2013.

[10] Jouni Hartikainen, Jaakko Riihimäki, and Simo Särkkä, "Sparse spatio-temporal Gaussian processes with general likelihoods," in *Artificial Neural Networks and Machine Learning (ICANN)*, 2011.

[11] Michael Leonard Stein, *Interpolation of Spatial Data*, Springer Series in Statistics. Springer, New York, 1999.

[12] Milton Abramowitz and Irene Stegun, *Handbook of Mathematical Functions*, Dover Publishing, New York, 1970.

[13] Mohinder S. Grewal and Angus P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB*, Wiley-Interscience, second edition, 2001.

[14] Edward J. Davison and Fu-Tin Man, "The numerical solution of $A'Q + QA = -C$," *IEEE Transactions on Automatic Control*, vol. 13, no. 4, pp. 448–449, 1968.

[15] Frank W.J. Olver, *NIST Handbook of Mathematical Functions*, Cambridge University Press, 2010.

[16] James Victor Uspensky, "On the convergence of quadrature formulas related to an infinite interval," *Transactions of the American Mathematical Society*, vol. 30, no. 3, pp. 542–559, 1928.

[17] Jesper Møller, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen, "Log Gaussian Cox processes," *Scandinavian Journal of Statistics*, vol. 25, pp. 451–482, 1998.

[18] Surya T. Tokdar and Jayanta K. Ghosh, "Posterior consistency of logistic Gaussian process priors in density estimation," *Journal of Statistical Planning and Inference*, vol. 137, no. 1, pp. 34–42, 2007.