Gaussian Quadratures for
State Space Approximation of Scale Mixtures
of Squared Exponential Covariance Functions

Arno Solin Simo Särkkä

Department of Biomedical Engineering and Computational Science

Aalto University, Finland

Introduction

- Gaussian processes (GPs, [2]) are a central part of both signal processing and statistical machine learning.
- In signal processing, often represented as state space models.
- In machine learning, the kernel (covariance function) representation is favored.
- This link enables the combination of the intuitive model specification from machine learning with computationally efficient signal processing methods.
- Reduces the problematic $\mathcal{O}(N^3)$ computational complexity to $\mathcal{O}(N)$ in the number of observations N by Kalman filtering methods.
- For infinitely differentiable covariance functions the representation is an approximation.
- 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.
- We focus on the rational quadratic (RQ) covariance function approximation, and we demonstrate the results in GP regression and a log-Gaussian Cox process.

Gaussian processes in machine learning

- Kernel representation: The model functions f are assumed to be realizations from a GP prior:

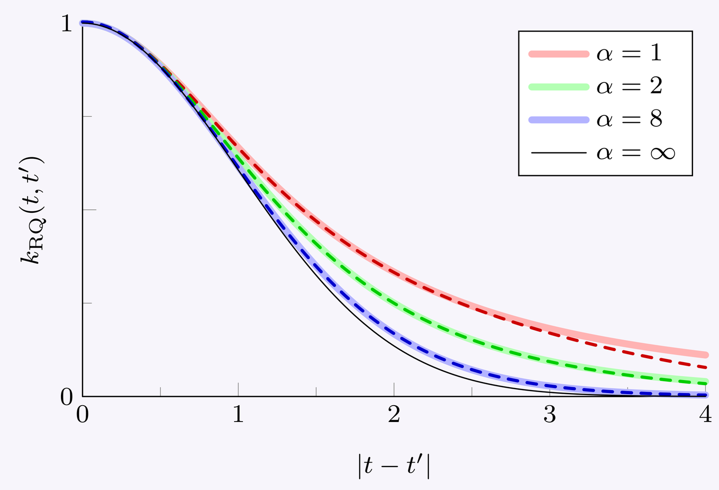
$$f(t) \sim \mathcal{GP}(0, k(t,t')).$$

- Certain classes of covariance functions allow to work with the mathematical dual, where the Gaussian process is constructed as a solution to a Mth order linear stochastic differential equation (SDE).
- State space representation: The GP regression problem can also be given as:

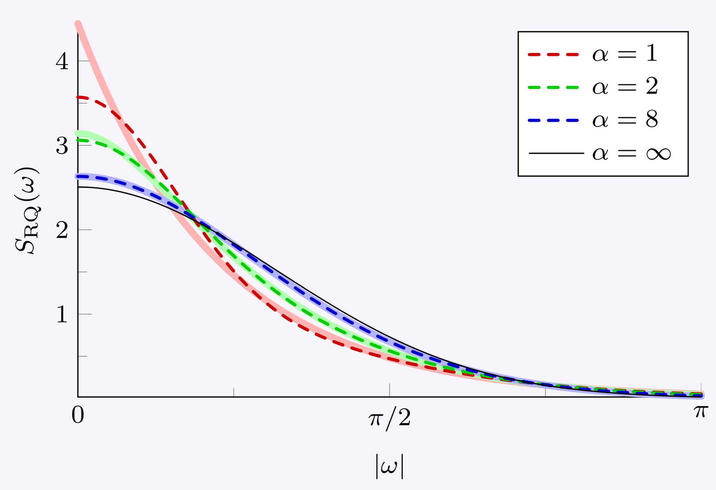
$$\frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} = \mathbf{F}\mathbf{f}(t) + \mathbf{L}\mathbf{w}(t),$$
$$f(t_k) = \mathbf{H}\mathbf{f}(t_k),$$

where $\mathbf{w}(t)$ is a multi-dimensional white noise process with spectral density $\mathbf{Q}_c$.

- The model is defined by the feedback matrix $\mathbf{F}$, the noise effect matrix $\mathbf{L}$, the spectral density $\mathbf{Q}_c$, the stationary covariance $\mathbf{P}_\infty$, and the observation model $\mathbf{H}$.
- The inference problem can now be solved using Kalman filtering [3] in $\mathcal{O}(NM^3)$ time complexity.



(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.

A general class of scale mixture covariance functions

- The squared exponential (RBF, Gaussian, exponentiated quadratic) covariance function (corresponds to $\alpha = \infty$ in Fig. 1):

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

where $\ell, \sigma^2 > 0$ are the characteristic length-scale and magnitude parameters.

- A general class of stationary kernels can be constructed as superpositions of squared exponential covariance functions (a scale mixture):

$$k_{\mathrm{SM}}(t) = \int_0^\infty p(\ell)\, k_{\mathrm{SE}}(t \mid \ell)\,\mathrm{d}\ell, \qquad (1)$$

where $k_{\mathrm{SE}}(t \mid \ell)$ denotes the squared exponential kernel with length-scale ℓ.

- This class includes, for example, the rational quadratic and Cauchy covariance functions.
- The corresponding stochastic process is infinitely many times mean square differentiable, and thus has no finite state space representation.
- However, the squared exponential can be approximated by a state space model [4].

The rational quadratic covariance function

- The rational quadratic (RQ) covariance function (see Fig. 1) is of the form:

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

- The scale mixture form (1) now gives the integral representation:

$$k_{\mathrm{RQ}}(t) = \int_0^\infty p(\xi \mid \alpha, \ell_{\mathrm{RQ}})\, k_{\mathrm{SE}}(t \mid \xi)\,\mathrm{d}\xi.$$

- The Gauss–Laguerre quadrature rule gives us an approximation:

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

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

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

- The state space formulation is thus a sum of state space models for the squared exponential (see [4]).

Conclusions

- We have shown how certain types of covariance functions (*e.g.*, the rational quadratic covariance) can be approximated by a Gaussian quadrature rule.
- This formulation can be used for constructing a state space approximation for this model.
- This link enables the use of efficient sequential inference methods to solve GP inference problems in $\mathcal{O}(N)$ time complexity.
- The approximation converges and a rough upper bound for the error can be given.
- Our experiments showed that this state space approximation is useful in practice both in GP regression (Fig. 2) and GP modeling in a more general setting (Fig. 3).

Codes available

- This method is now a part of the GPstuff toolbox for Matlab/Octave.
- Details and related codes are available on the author web page:

`http://becs.aalto.fi/~asolin/`

References

[1] Arno Solin and Simo Särkkä, "Gaussian quadratures for state space approximation of scale mixtures of squared exponential covariance functions," in *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2014.

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

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

[4] 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.

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

Examples

Consistency with naive GP regression

- In the first example (Fig. 2), we compare the results given by the state space approximation model against the naive full GP regression solution.
- We use the rational quadratic covariance function and simulated data.
- The results are practically equal, but the state space solution can be obtained in $\mathcal{O}(N)$ time complexity.

A temporal log-Gaussian Cox process

- In the second example (Fig. 3), we model the intensity of coal mining accidents.
- We consider a log-Gaussian Cox process, which is an inhomogeneous Poisson point process [5].
- The model is thus a GP model with a Poisson likelihood:

$$f(t) \sim \mathcal{GP}(0, k(t,t'))$$
$$p(\mathcal{D} \mid f) = \prod_{k=1}^{N} \mathrm{Poisson}(y_k \mid \exp(f(t_k))),$$

- The state space approximation can be beneficial, as the interval can be discretized into a very dense grid without running into computational limitations.

Consistency with naive GP regression

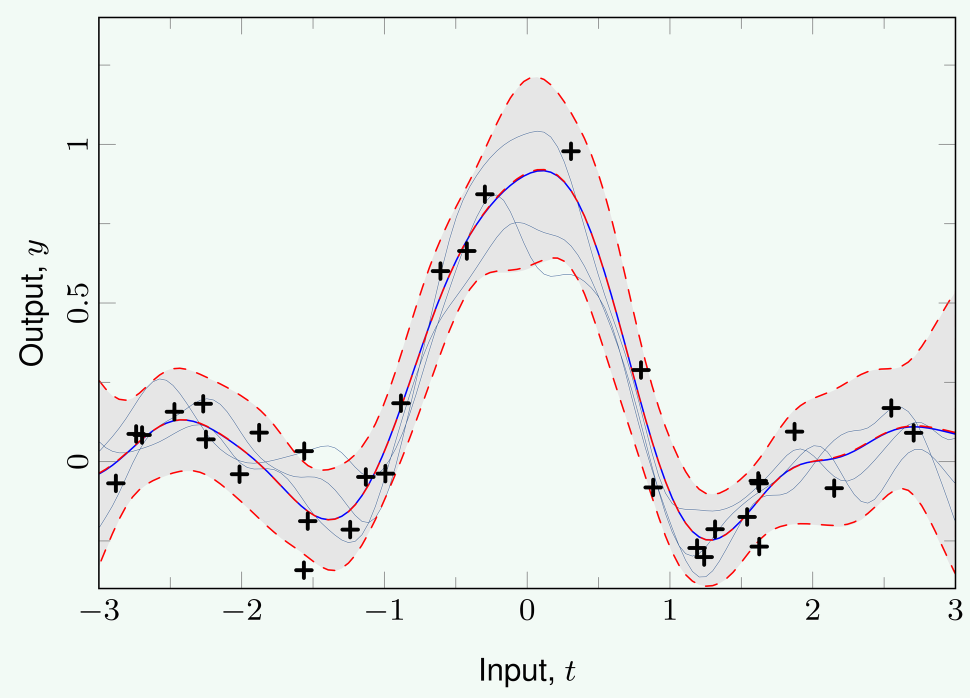


Fig. 2: GP regression results for simulated data (shown by the point markers). The state space mean and 95% confidence interval estimates are shown by the solid blue line and the grey patch. The corresponding full GP regression results is shown by dashed red lines. The thin solid lines are random draws from the state space posterior.

A temporal log-Gaussian Cox process

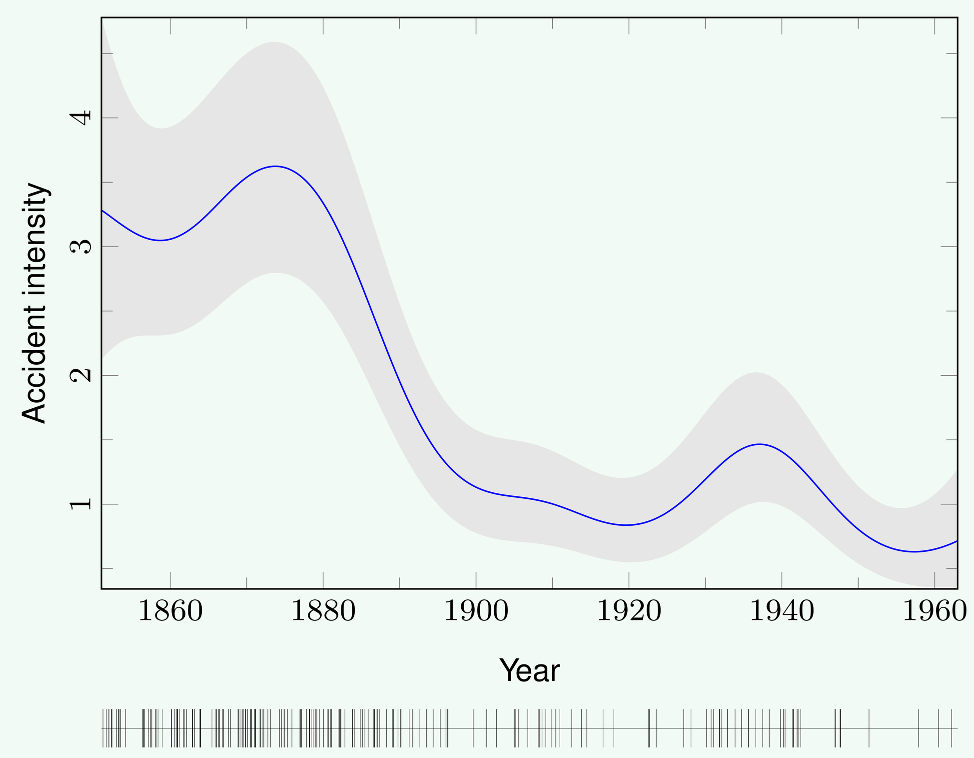


Fig. 3: Coal mining accident data, with $N = 1024$ intervals and 191 incidents. The bar shows the actual incidents, and the modeling outcome for the intensity in the log-Gaussian Cox process model with an approximate 90% confidence region is shown in the figure above.

Contact details: arno.solin@aalto.fi, web address: http://arno.solin.fi. This work was presented at MLSP 2014, Reims, France.