Hilbert Space Methods for Reduced-Rank Gaussian Process Regression

Arno Solin and Simo Särkkä

Aalto University, Finland

Workshop on Gaussian Process Approximation
Copenhagen, Denmark, May 2015

Contents

Problem formulation

- Gaussian process (GP) regression problem:

$$
\begin{aligned}
f &\sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}')), \\
y_i &= f(\mathbf{x}_i) + \varepsilon_i.
\end{aligned}
$$

- The GP-regression has cubic computational complexity $O(n^3)$ in the number of measurements.
- This results from the inversion of an $n \times n$ matrix:

$$
\begin{aligned}
\mu(\mathbf{x}_*) &= k(\mathbf{x}_*, \mathbf{x}_{1:n}) \left(k(\mathbf{x}_{1:n}, \mathbf{x}_{1:n}) + \sigma_\mathrm{n}^2 \mathbf{I}\right)^{-1} \mathbf{y} \\
V(\mathbf{x}_*) &= k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathbf{x}_*, \mathbf{x}_{1:n}) \left(k(\mathbf{x}_{1:n}, \mathbf{x}_{1:n}) + \sigma_\mathrm{n}^2 \mathbf{I}\right)^{-1} k(\mathbf{x}_{1:n}, \mathbf{x}_*).
\end{aligned}
$$

- In practice, we use Cholesky factorization and do not invert explicitly – but still the $O(n^3)$ problem remains.
- Various sparse, reduced-rank, and related approximations have been developed for mitigating this problem.

Covariance operator

- For covariance function $k(\mathbf{x}, \mathbf{x}')$ we can define covariance operator:

$$\mathcal{K}\,\phi = \int k(\cdot, \mathbf{x}')\,\phi(\mathbf{x}')\,\mathrm{d}\mathbf{x}'.$$

- For stationary covariance function $k(\mathbf{x}, \mathbf{x}') \triangleq k(\|\mathbf{r}\|)$; $\mathbf{r} = \mathbf{x} - \mathbf{x}'$ we get

$$S(\boldsymbol{\omega}) = \int k(\mathbf{r})\, e^{-\mathrm{i}\,\boldsymbol{\omega}^\mathsf{T}\mathbf{r}}\,\mathrm{d}\mathbf{r}.$$

- $\Rightarrow$ The transfer function corresponding to the operator $\mathcal{K}$ is

$$S(\boldsymbol{\omega}) = \mathscr{F}[\mathcal{K}].$$

- The spectral density $S(\boldsymbol{\omega})$ also gives the approximate eigenvalues of the operator $\mathcal{K}$.
- We can now represent the operator $\mathcal{K}$ as a series of Laplace operators.

Laplacian operator series

- In isotropic case $S(\boldsymbol{\omega}) \triangleq S(\|\boldsymbol{\omega}\|)$, and we can expand

$$S(\|\boldsymbol{\omega}\|) = a_0 + a_1\|\boldsymbol{\omega}\|^2 + a_2(\|\boldsymbol{\omega}\|^2)^2 + a_3(\|\boldsymbol{\omega}\|^2)^3 + \cdots .$$

- The Fourier transform of the Laplace operator ∇^2 is $-\|\boldsymbol{\omega}\|^2$, *i.e.*,

$$\mathcal{K} = a_0 + a_1(-\nabla^2) + a_2(-\nabla^2)^2 + a_3(-\nabla^2)^3 + \cdots .$$

- Defines a pseudo-differential operator as a series of differential operators.
- Let us now approximate the Laplacian operators with a Hilbert method...

Background: series expansions of GPs

- Assume a covariance function $k(\mathbf{x}, \mathbf{x}')$ and an inner product, say,

$$\langle f, g \rangle = \int_\Omega f(\mathbf{x})\, g(\mathbf{x})\, w(\mathbf{x}) \, \mathrm{d}\mathbf{x}.$$

- The inner product induces a Hilbert-space of (random) functions.
- If we fix a basis $\{\phi_j(\mathbf{x})\}$, a Gaussian process $f(\mathbf{x})$ can be expanded into a series

$$f(\mathbf{x}) = \sum_{j=1}^{\infty} f_j \, \phi_j(\mathbf{x}),$$

  where f_j are jointly Gaussian.
- If we select ϕ_j to be the eigenfunctions of $k(\mathbf{x}, \mathbf{x}')$ w.r.t. $\langle \cdot, \cdot \rangle$, then this becomes the Karhunen–Loève series.
- In the Karhunen–Loève case the coefficients f_j are independent Gaussian.

Hilbert-space approximation of Laplacian

- Consider the eigenvalue problem for the Laplacian operators:

$$\begin{cases} -\nabla^2 \phi_j(\mathbf{x}) = \lambda_j \, \phi_j(\mathbf{x}), & \mathbf{x} \in \Omega, \\ \phi_j(\mathbf{x}) = 0, & \mathbf{x} \in \partial\Omega. \end{cases}$$

- The eigenfunctions $\phi_j(\cdot)$ are orthonormal w.r.t. inner product

$$\langle f, g \rangle = \int_\Omega f(\mathbf{x}) \, g(\mathbf{x}) \, \mathrm{d}\mathbf{x},$$

$$\int_\Omega \phi_i(\mathbf{x}) \, \phi_j(\mathbf{x}) \, \mathrm{d}\mathbf{x} = \delta_{ij}.$$

- The negative Laplacian has the formal kernel

$$l(\mathbf{x}, \mathbf{x}') = \sum_j \lambda_j \, \phi_j(\mathbf{x}) \, \phi_j(\mathbf{x}')$$

in the sense that

$$-\nabla^2 f(\mathbf{x}) = \int l(\mathbf{x}, \mathbf{x}') \, f(\mathbf{x}') \, \mathrm{d}\mathbf{x}'.$$

Hilbert-space approximation of covariance function

- Recall that we have the expansion

$$\mathcal{K} = a_0 + a_1(-\nabla^2) + a_2(-\nabla^2)^2 + a_3(-\nabla^2)^3 + \cdots .$$

- Substituting the formal kernel gives

$$\begin{aligned} k(\mathbf{x}, \mathbf{x}') &\approx a_0 + a_1\, l^1(\mathbf{x}, \mathbf{x}') + a_2\, l^2(\mathbf{x}, \mathbf{x}') + a_3\, l^3(\mathbf{x}, \mathbf{x}') + \cdots \\ &= \sum_j \left[a_0 + a_1\, \lambda_j^1 + a_2\, \lambda_j^2 + a_3\, \lambda_j^3 + \cdots \right] \phi_j(\mathbf{x})\, \phi_j(\mathbf{x}'). \end{aligned}$$

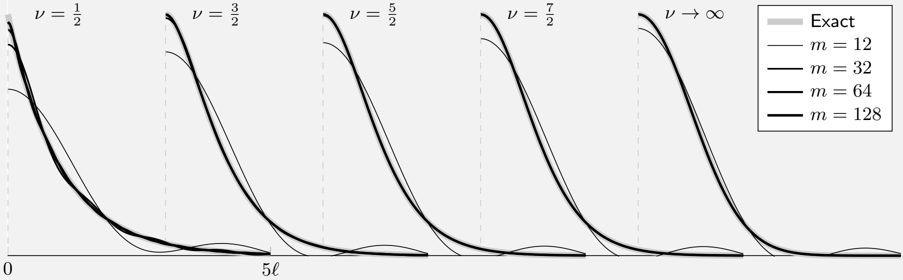
- Evaluating the spectral density series at $\|\boldsymbol{\omega}\|^2 = \lambda_j$ gives

$$S(\sqrt{\lambda_j}) = a_0 + a_1\lambda_j^1 + a_2\lambda_j^2 + a_3\lambda_j^3 + \cdots .$$

- This leads to the final approximation

$$\boxed{k(\mathbf{x}, \mathbf{x}') \approx \sum_j S(\sqrt{\lambda_j})\, \phi_j(\mathbf{x})\, \phi_j(\mathbf{x}').}$$

Accuracy of the approximation



Approximations to covariance functions of the Matérn class of various degrees of smoothness; $\nu = 1/2$ corresponds to the exponential Ornstein–Uhlenbeck covariance function, and $\nu \to \infty$ to the squared exponential (exponentiated quadratic) covariance function. Approximations are shown for 12, 32, 64, and 128 eigenfunctions.

Reduced-rank method for GP regression

- Recall the GP-regression problem

$$f \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$$
$$y_i = f(\mathbf{x}_i) + \varepsilon_i.$$

- Let us now approximate

$$f(\mathbf{x}) \approx \sum_{j=1}^{m} f_j \, \phi_j(\mathbf{x}),$$

where $f_j \sim \mathcal{N}(0, S(\sqrt{\lambda_j}))$.
- The approximating GP $f_m(\mathbf{x})$ now approximately has the desired covariance function.
- Via the matrix inversion lemma we then get

$$\mu_* \approx \boldsymbol{\phi}_*^\mathsf{T}(\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi} + \sigma_\mathrm{n}^2\boldsymbol{\Lambda}^{-1})^{-1}\boldsymbol{\Phi}^\mathsf{T}\mathbf{y},$$
$$V_* \approx \sigma_\mathrm{n}^2\boldsymbol{\phi}_*^\mathsf{T}(\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi} + \sigma_\mathrm{n}^2\boldsymbol{\Lambda}^{-1})^{-1}\boldsymbol{\phi}_*.$$

Computational complexity

- The computation of $\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}$ takes $\mathcal{O}(nm^2)$ operations.
- The covariance function parameters do not enter $\mathbf{\Phi}$ and we need to evaluate $\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}$ only once (nice in parameter estimation).
- If the observations are on a grid, we can use FFT-kind of methods.
- The scaling in input dimensionality can be quite bad—but depends on the chosen domain.

Convergence analysis

Theorem

There exists a constant C_d such that

$$\left| k(\mathbf{x}, \mathbf{x}') - \widetilde{k}_m(\mathbf{x}, \mathbf{x}') \right| \leq \frac{C_d}{L} + \frac{1}{\pi^d} \int_{\|\boldsymbol{\omega}\| \geq \frac{\pi \, \hat{m}}{2L}} S(\boldsymbol{\omega}) \, \mathrm{d}\boldsymbol{\omega},$$

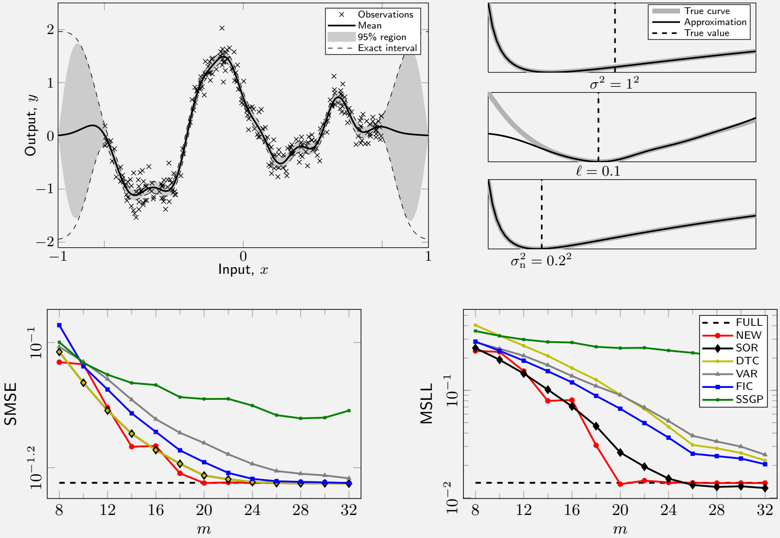
where $L = \min_k L_k$, which in turn implies that uniformly

$$\lim_{L_1, \ldots, L_d \to \infty} \left[\lim_{m \to \infty} \widetilde{k}_m(\mathbf{x}, \mathbf{x}') \right] = k(\mathbf{x}, \mathbf{x}').$$

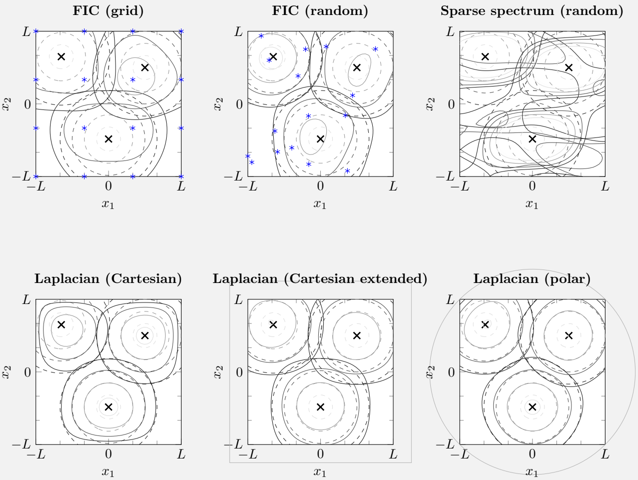
Corollary

The uniform convergence of the prior covariance function also implies uniform convergence of the posterior mean and covariance in the limit $m, L_1, \ldots, L_d \to \infty$.

Toy example


Observations
Mean
95% region
Exact interval
Output, y
Input, x
True curve
Approximation
True value
$\sigma^2 = 1^2$
$\ell = 0.1$
$\sigma_n^2 = 0.2^2$
SMSE
MSLL
m
FULL
NEW
SOR
DTC
VAR
FIC
SSGP

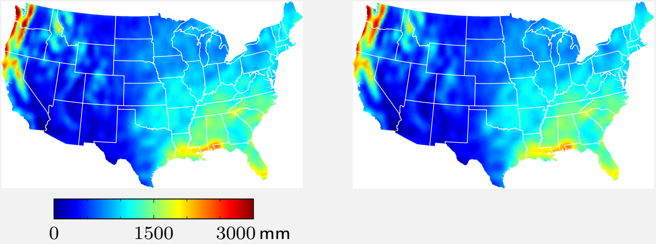
Correlation contours in two-dimensions ($m = 16$)

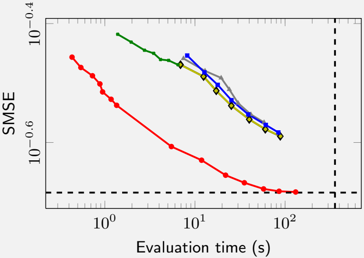


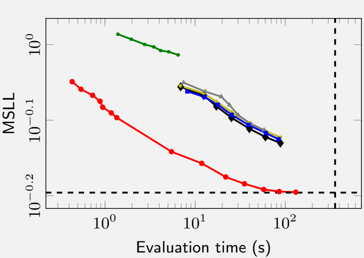
Correlation contours computed for three locations corresponding to the squared exponential covariance function (exact contours dashed).

Precipitation data

Interpolation of yearly precipitation ($n = 5776$) using a full GP and the Laplacian GP.

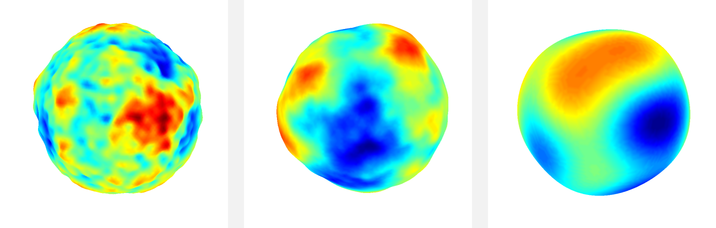


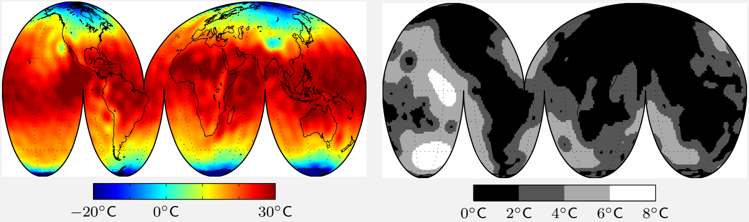




Surface temperature of the globe

Random draws (Matérn, $\nu = \{1/2, 3/2, \infty\}$) in a spherical domain.





GP regression (mean and standard deviation)
of temperature data ($n = 11028$).

Summary

- Hilbert-space method for reduced-rank Gaussian process regression.
- The covariance function is approximated by eigenfunctions of the Laplace operator.
- The eigenvalues are approximated by spectral density values.
- The approximation converges to exact limit in well-defined conditions.
- Experimental results show that the method works well in practice.

Related and future work

- Spatio-temporal Gaussian processes – combination with Kalman filtering and smoothing.
- Periodic covariance functions and other specific classes of covariances.
- Static and spatio-temporal inverse problems.
- Non-Gaussian processes.
- Optimization for high-dimensional inputs.