Jannik Prüßmann*, Jan Graßhoff, Philipp Rostalski

Exploitation of Kronecker Structure in Gaussian Process Regression for Efficient Biomedical Signal Processing

Abstract: Gaussian processes are a versatile tool for data processing. Unfortunately, due to storage and runtime requirements, standard Gaussian process (GP) methods are limited to a few thousand data points. Thus, they are infeasible in most biomedical, spatio-temporal problems. The methods treated in this work cover GP inference and hyperparameter optimization, exploiting the Kronecker structure of covariance matrices. To solve regression and source separation problems, two different approaches are presented. The first approach uses efficient matrix-vector-products, whilst the second approach is based on efficient solutions to the eigendecomposition. The latter also enables efficient hyperparameter optimization. In comparison to standard GP methods, the proposed methods can be applied to very large biomedical datasets without any further performance loss and perform substantially faster. The performance is demonstrated on esophageal manometry data, where the cardiac and respiratory signal components are to be inferred by source separation.

Keywords: Gaussian Processes, Regression, GP Inference, Kronecker Product, Kronecker Structure, Covariance Matrix

https://doi.org/10.1515/cdbme-2021-2073

1 Introduction

In modern medicine, an ever-growing variety of signals can be measured from the patient. In many relevant settings these signals have spatio-temporal structure, i.e., measurements are taken at different positions and recorded over time. Important examples are electrocardiography, electroencephalography and electrical impedance tomography. These examples motivate the need for a concise framework for modeling and processing such signals. Gaussian processes provide a powerful tool in signal processing, but, due to requirements in storage and runtime, standard Gaussian process (GP) methods are limited to small data sets. As a result, standard GP methods are not feasible on large spatio-temporal data originating from biomedical, physiological processes. This work provides two methods for efficient spatio-temporal GP inference and learning to overcome these limitations. The provided methods are based on exploiting the Kronecker structure in covariance matrices. Both methods are compared with the standard GP method concerning the quality of the results and the implementation's runtime. Moreover, one of the introduced methods is applied to esophageal manometry data as an example of a biomedical source separation problem.

2 Methods

A GP, being a generalization of the Gaussian probability distribution, is a stochastic process that governs distributions over functions [1]. GPs are a widely used tool to handle spatio-temporal settings since many classical approaches model either the temporal or the spatial structure but not both [2]. In general, a GP assigns a prior likelihood to every possible function [1]. A GP $f(x) \sim \mathcal{GP}\big(m(x), k(x, x')\big)$ is fully specified by its mean function $m(x)$, which is often assumed to be zero, and its covariance function $k(x, x')$. Given a vector of measurement observations $y = (y_1, \dots, y_n)^\top$ and assuming a noisy measurement model $y_i = f(x_i) + \epsilon$, where ϵ is Gaussian noise with the variance ${\sigma_n}^2$, the posterior mean of the GP can be calculated in closed form:

$$\overline{f^*} = K_{XX}[K_{XX} + {\sigma_n}^2 I]^{-1} y, \tag{1}$$

where the entries of the kernel matrix K_{XX} are formed by evaluating the covariance function for all pairs of points in $X = \{x_1, \dots, x_N\}$ and I denotes the identity matrix.

In biomedical data, the measurements are often constituted by a mixture of different physiological processes, which can be modelled by the sum of two independent GPs

***Corresponding author: Jannik Prüßmann:** Institute for Electrical Engineering in Medicine, University of Lübeck, Lübeck, Germany, e-mail: jannik.pruessmann@student.uni-luebeck.de
Jan Graßhoff, Philipp Rostalski: Institute for Electrical Engineering in Medicine, University of Lübeck, Moislinger Allee 53-55, 23558 Lübeck, Germany

$f(x)$ and $g(x)$: $y_i = f(x_i) + g(x_i) + \epsilon$. In many biomedical problems the separation of different physiological sources is of great interest. In the additive GP model and according to Liutkus [3] the posterior mean $\overline{f^*}$ and $\overline{g^*}$ of sources f and g, respectively, can also be calculated in closed form via

$$\overline{f^*} = K_{f,XX}\left[K_{f,XX} + K_{g,XX} + \sigma_n{}^2 I\right]^{-1} y \qquad (2)$$
$$\overline{g^*} = K_{g,XX}\left[K_{f,XX} + K_{g,XX} + \sigma_n{}^2 I\right]^{-1} y,$$

Where K_f and K_g are formed using only the covariance functions of f and g, respectively. Note that the linear system must only be solved once for the extraction of all sources.

The performance of the GP models is highly dependent on the choice of some free hyperparameters θ that define the shape of the covariance functions. These parameters can be learned by optimizing the log marginal likelihood

$$\log \mathcal{L}(\theta|y) \propto y^\top \widehat{K}^{-1} y + \log\left(\left|\widehat{K}\right|\right), \qquad (3)$$

with $|\cdot|$ being the determinant and $\widehat{K} = K_{f,XX} + K_{g,XX} + \sigma_n{}^2 I$.

2.1 Spatio-Temporal GPs

Spatio-temporal GPs are characterized by their multi-dimensional inputs, defined by a point in time and an input vector from the spatial dimension. We consider a special case of spatio-temporal GPs where the input points are on a grid, so that for all time instances the same spatial points are considered. Then we can define a collection of spatial input vectors $S = \{s_m : m = 1, \dots, M\}$ and a collection of time points $T = \{t_n : n = 1, \dots, N\}$ such that NM observations $y_{i,j} = f(s_i, t_j) + \epsilon$ are located at the $S \times T$ grid of the Cartesian product of the inputs [2]. We further choose a separable covariance function

$$k\big((s,t),(s',t');\theta\big) = k_s(s,s';\theta_s)k_t(t,t';\theta_t), \qquad (4)$$

where θ_s and θ_t are the hyperparameter sets used in the dimension-related covariance functions [2]. Due to the separability and the definition on a grid, $k\big((s,t),(s',t');\theta\big)$ is a tensor product kernel and, as mentioned by Saatçi [4], the covariance matrices can be written as a Kronecker product of smaller covariance matrices

$$K = K_T \otimes K_S \qquad (5)$$

where the covariance matrix K is formed by computing the covariance matrices K_S and K_T and evaluating inputs S with the kernel k_s and inputs T with the kernel k_t.

2.2 Efficient GP Inference and Learning

2.2.1 Conjugate Gradient Method

To overcome the restrictions of standard GPs, concerning the data size and exceeding runtime, one critical issue is the matrix inversion in eq. 1. Considering the shown Kronecker structure of the covariance matrix K, these constraints can be eased. To solve the linear system, we denote the solution to the linear system involving the kernel matrix and y as

$$\alpha = [(K_t \otimes K_s) + \sigma_n^2 I]^{-1} y. \qquad (6)$$

This equation can be solved iteratively using the conjugate gradient method [2]. Thereby, the expensive matrix inversion is avoided, but still, there remains a matrix-vector-product with the large matrix $(K_t \otimes K_s) + \sigma_n^2 I$ to be computed, which is also not expected to be efficient. To increase efficiency, the matrix-vector-product can be solved via

$$[(K_t \otimes K_s) + \sigma_n^2 I][A]_: = [K_s A K_t]_: + \sigma_n^2 [A]_:, \qquad (7)$$

where $[A]_:$ is the vector obtained by stacking the columns of the $M \times N$ matrix A [2]. $[A]_:$ converges to α with each iteration of the conjugate gradient method. Thus, the large matrix-vector-product can be avoided and the *conjugate gradient (CG)* method can be applied to obtain the approximate solution $\hat{\alpha}$ to eq. 6. The posterior mean is then computed as

$$\overline{f^*} \approx (K_t \otimes K_s)\hat{\alpha} \qquad (8)$$

and the expensive matrix inversion is circumvented. Furthermore, this efficient approach can be applied directly to source-separation problems, using $K = K_f + K_g$ with $K_f = K_{f_t} \otimes K_{f_s}, K_g = K_{g_t} \otimes K_{g_s}$ and

$$\overline{g^*} \approx \left(K_{g_t} \otimes K_{g_s}\right)\hat{\alpha} \qquad (9)$$

for the computation of the posterior mean's component that is produced by the source corresponding to K_g [3]. To calculate the posterior mean's component $\overline{f^*}$ that is related to the K_f-source, the covariance matrices in eq. 9 are exchanged.

2.2.2 Efficient Eigendecomposition

Whilst the previously mentioned implementation of the posterior mean calculation relies on iterative methods, this section provides another approach of exploiting the Kronecker product's properties for efficient GPs using the eigendecompostion for a direct solution.

Since K, K_t and K_s are real, symmetric matrices they can be decomposed as $K = Q\Lambda Q^\top$ [4]. The eigendecomposition of a Kronecker matrix can be calculated efficiently using the eigendecompositions of the much smaller Kronecker factors, see [4] for details. Following Saatçi [4] these properties can be utilized in calculating the posterior mean using

$$[Q\Lambda Q^\top + \sigma_n^2 I]^{-1} y = Q[\Lambda + \sigma_n^2 I]^{-1} Q^\top y, \qquad (10)$$

where the inversion $[\Lambda + \sigma_n^2 I]^{-1}$ is done efficiently due to its diagonal structure.

It can easily be verified that this approach cannot be applied in the previous source separation case, because a sum of Kronecker matrices does not have Kronecker structure and cannot be eigendecomposed efficiently. Nevertheless, the original source separation problem can be simplified by assuming that both sources have the same spatial covariance function k_s. Including this in the example, the covariance functions of the independent sources are

$$k_f(s, s'; \theta_f) = k_s(s, s'; \theta_s) k_{f_t}(t, t'; \theta_{f_t}), \qquad (11)$$
$$k_g(s, s'; \theta_g) = k_s(s, s'; \theta_s) k_{g_t}(t, t'; \theta_{g_t}) \qquad (12)$$

and therefore, the total covariance function $k = k_f + K_g$ can now be represented as a product with $k = k_s(k_{f_t} + k_{g_t})$, which results in the covariance matrix $K = K_s \otimes (K_{f_t} + K_{g_t})$. Thus K can be represented as a Kronecker product and the method can be applied [4] for efficient source separation.

2.2.3 Efficient GP Learning

To increase efficiency of evaluations of the log marginal likelihood, the benefits of Kronecker structure and eigendecomposition can also be exploited. Saatçi [4] showed, that the log-determinant can be derived as

$$\log(\det(K + \sigma_n^2)) = \textstyle\sum_i \log([\mathrm{diag}(\Lambda_T) \otimes \mathrm{diag}(\Lambda_S)]_i + \sigma_n^2), \qquad (13)$$

since $\Lambda = \Lambda_t \otimes \Lambda_s$ contains the eigenvalues of K on its diagonal. Thus $\log(\mathcal{L}(\theta))$ can also be calculated efficiently. Note, that this again relies on the assumption that the kernel matrix is not a sum of independent sources but shares the spatial kernel across all sources as in eq. 11 and 12.

3 Results

In order to examine how the utilization of the Kronecker structure can affect the performance of GPs, the methods introduced were implemented in *MATLAB* for comparison with standard GP methods provided by the *GPstuff*-Library [5], further referred to as *batch*-method.

3.1 Simulation Data

First, the provided *conjugate gradient* method *(CG)* and *eigendecomposition*-method *(ED)* as well as the batch-method were tested on simulated data. Therefore, two random samples were drawn from different priors, with respect to eq. 11 and 12. The samples were then added and perturbed by artificial measurement noise to simulate a measured signal resulting from two different sources.

The simulation showed that the introduced *CG* and *ED* method performed equally well as the batch-method, yielding the same results up to numerical accuracy and being suitable to solve the regression. All three methods were then set up to solve the source separation problem. As shown in Figure 1, CG and ED were substantially faster than the batch method.

Next, we tested the performance of hyperparameter learning. To this end, we report the required runtime of the ED

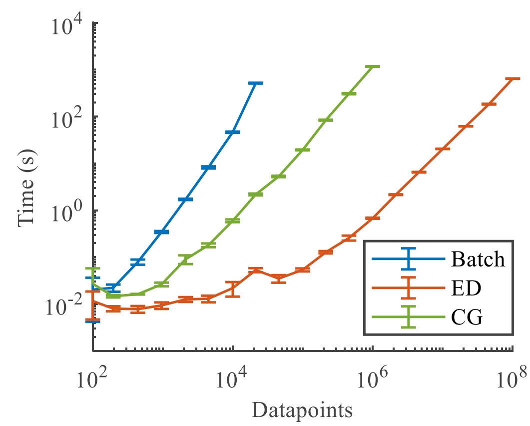

Figure 1: Runtime of the source separation in seconds. The error bars show standard deviations (n=10).

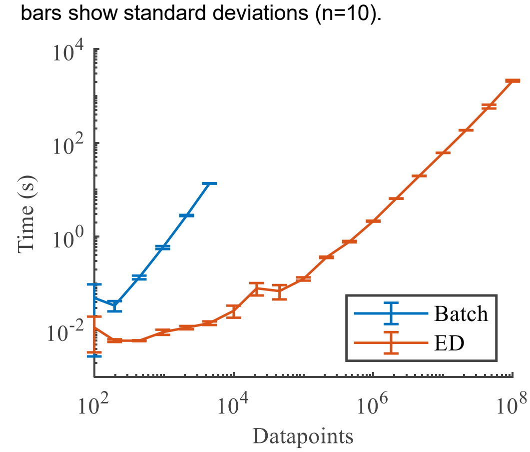

Figure 2: Runtime of computing negative log marginal likelihood and its derivatives. The error bars show standard deviations (n=10).

and batch method for evaluation of the log marginal likelihood and its derivatives for different input data sizes. As shown in Figure 2, the *ED*-method performed substantially faster.

3.2 Esophageal Manometry Data

The performance of the *ED*-method on real spatio-temporal biomedical data is shown, using esophageal manometry data from Persson et. al. [6]. The objective for applying the *eigendecomposition*-method on the esophageal manometry data was to separate the data signals corresponding to heart activity and ventilation, respectively, as esophageal pressure artefact removal is a relevant problem [7]. To model ventilation and heart activities, the time-domain covariance functions were chosen to be quasiperiodic, which allows minor changes of waveforms over multiple periods. To model changes in heart and breath frequency, a warping function [8] was used. The remaining parameters were set regarding the expected physiological behaviour of the two sources.

The data did not show a simple distance-based covariance structure in its spatial direction, therefore, the spatial covariance was modelled by a multi-task covariance function, see [9] for details. The multi-task covariance matrix was initialized with independent tasks for each measurement height, relying on the *ED*-method to learn the positive and negative inter-task correlations. Using this optimized covariance matrix, the source separation was performed. The result for the source separation of the data is shown in Figure 3.

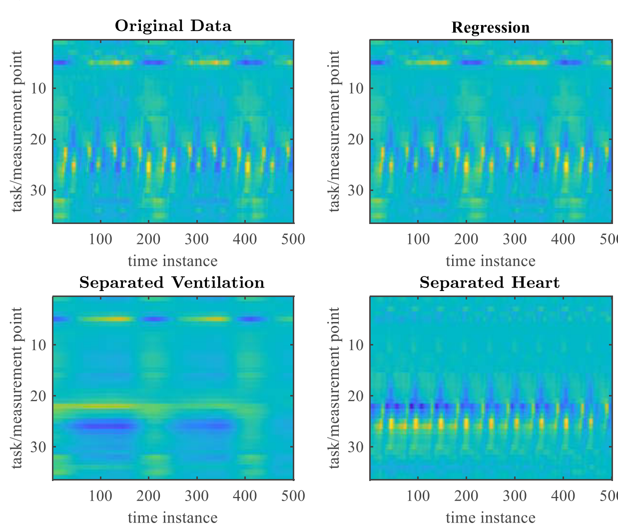

Figure 3: Separation results for esophageal manometry measurement. The figure depicts the original data [6] in the upper left, and the regression result of the *ED*-method in the upper right subfigure. The separation results of the ventilation and cardiac signals are shown in the lower subfigures.

4 Conclusion

In conclusion, this work covered two methods to efficiently solve GP inference and hyperparameter learning by exploiting the Kronecker structure of the covariance matrix, that is often present in biomedical data. Both methods were able to perform substantially faster than the commonly used standard method on simulated data, thus, relaxing restriction on the size of data and allowing to solve large regression and source separation problems on spatio-temporal biomedical data sets. The methods of learning and regression were demonstrated on a source separation problem in esophageal manometry data, where the separation of cardiac and respiratory effects was treated.

Author Statement

Conflict of interest: Authors state no conflict of interest. We would like to thank Per Persson for providing study data.

References

[1] C. E. Rasmussen and C. K. I. Williams, “Gaussian processes for machine learning”, *Adaptive computation and machine learning*. Cambridge, Mass.: MIT Press, 2008.

[2] J. Luttinen and A. Ilin, “Efficient Gaussian Process Inference for Short-Scale Spatio-Temporal Modeling”, *Artificial Intelligence and Statistics*, pp. 741–750,2012.

[3] A. Liutkus, R. Badeau, and G. Richard, “Gaussian Processes for Underdetermined Source Separation”, *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3155–3167, 2011.

[4] Y. Saatçi, “Scalable inference for structured Gaussian process models: Ph.D”, Ph.D. dissertation, University of Cambridge and Department of Engineering, 2011.

[5] J. Vanhatalo et al., “Bayesian Modeling with Gaussian Processes using the GPstuff Toolbox”, 2012.

[6] P. Persson, R. Ahlstrand, M. Gudmundsson, A. de Leon, and S. Lundin, “Detailed measurements of oesophageal pressure during mechanical ventilation with an advanced high-resolution manometry catheter”, *Critical care* (London, England), vol. 23, no. 1, p. 217, 2019.

[7] J. Graßhoff, E. Petersen, M. Eger, G. Bellani and P. Rostalski, "A template subtraction method for the removal of cardiogenic oscillations on esophageal pressure signals," *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* pp. 2235-2238, 2017.

[8] J. Graßhoff, A. Jankowski, and P. Rostalski, “Scalable Gaussian Process Separation for Kernels with a Non-Stationary Phase”, *International Conference on Machine Learning*, pp. 3722–3731, 2020.

[9] E. V. Bonilla, K. Chai, and C. Williams, “Multi-task Gaussian Process Prediction”, *Advances in Neural Information Processing Systems*, vol. 20, Curran Associates, Inc, 2008, pp. 153–160.