High-Dimensional Gaussian Process Inference with Derivatives

Filip de Roos, Alexandra Gessner & Philipp Hennig

ICML 2021

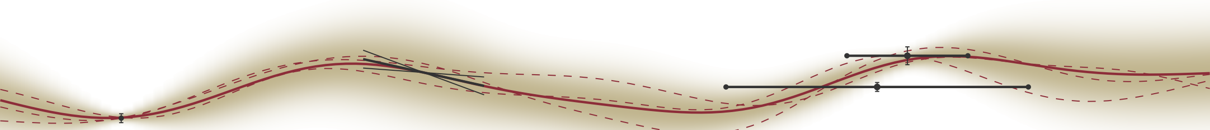
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS

imprs-is International Max Planck Research School on Intelligent Systems

some of the presented work is supported by the European Research Council.

erc



Problem: Gaussian process inference with derivatives

Model $f : \mathbb{R}^D \to \mathbb{R}$ with a GP and N observations has cost

	compute	**memory**
GP inference with functions	$\mathcal{O}(N^3)$	$\mathcal{O}(N^2)$
GP inference **with gradients**	$\mathcal{O}((DN)^3)$	$\mathcal{O}((DN)^2)$

$\rightarrow$ 1 gradient observation $\widehat{=}$ D function evaluations

This work shows that

Gradient inference requires $\mathcal{O}(DN^2 + N^6)$ **compute** and $\mathcal{O}(DN + N^2)$ **memory**

Translation: 1 gradient can be **cheaper** than D function evaluations

Solution: Structured kernels admit efficient matrix inversion

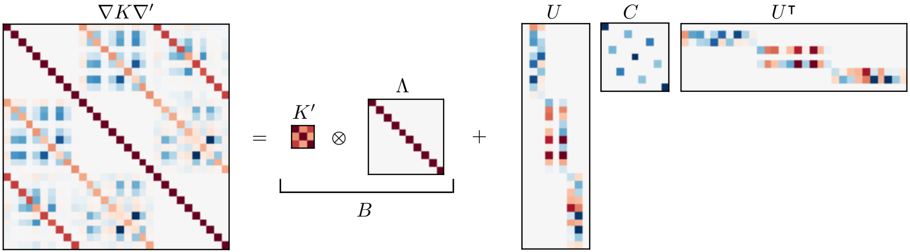


Figure: Kernel Gram matrix for RBF kernel with $N = 3$ gradient observations in $D = 10$ dimensions.

Woodbury's matrix inversion lemma

$$(B + UCU^\intercal)^{-1} = B^{-1} - B^{-1}U\left(C^{-1} + U^\intercal B^{-1}U\right)^{-1} U^\intercal B^{-1}$$

Implications: High-dimensional GP inference with gradients

Highlights (for $N < D$):

- Reduced compute and memory
- Efficient implicit matrix-vector multiplication
- Algorithms for optimization and sampling

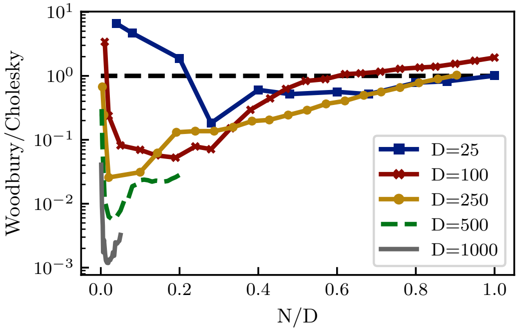


Figure: cpu(Woodbury) divided by cpu(Cholesky) for different dimensions and Gram matrices up to size 50 000.

Key takeaway:

Gradient inference is efficient in high-dimensional Gaussian processes

Paper `arxiv:2102.07542`

Code `https://github.com/fidero/gp-derivative`

Thank you!