Bayesian global optimization for neuroimaging

Ekaterina I Lomakina[1,2,3] · Christoph Mathys[2,3] · Vera M Schäfer[4] · Kay H Brodersen[1,2,3] · Alexander Vezhnevets[1] · Klaas E Stephan[2,3,5] · Joachim M Buhmann[1]

[1] Department of Computer Science, ETH Zurich, Switzerland [2] Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich & ETH Zurich, Switzerland [3] Social and Neural Systems Research Laboratory (SNS lab), University of Zurich, Switzerland [4] Department of Physics, ETH Zurich, Switzerland [5] Wellcome Trust Centre for Neuroimaging, University College London, United Kingdom

1 Summary

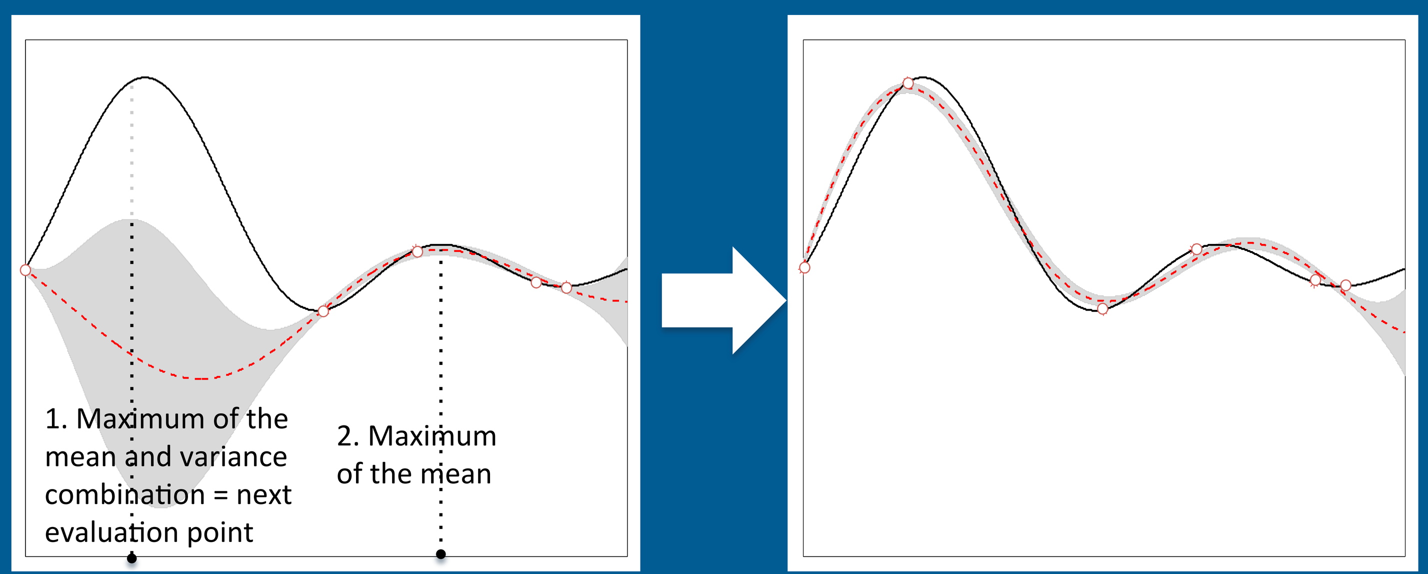
- Model-based approaches to the analysis of neuroimaging data and human behaviour have become increasingly powerful in recent years. Such models can provide mechanistic insights into latent processes that may, in particular, prove powerful for distinguishing between different groups of subjects [1].
- However, the estimation of the parameters of these models poses a challenging optimization problem. Existing inference methods are typically computationally expensive (e.g., Markov chain Monte Carlo, MCMC, [2]) or susceptible to local minima (e.g., variational Bayes, VB, [3]) .
- Bayesian global optimization (BGO, [4]) is a novel optimization approach that balances the merits of sampling methods and variational inference. In this study, we demonstrate the utility of BGO in neuroimaging by comparing it to MCMC and VB on the basis of synthetic data in application to a computational model and a dynamic causal model (DCM, [5]).

2 Gaussian processes for global optimization

Gaussian processes (GP, [6]) represent a non-parametric Bayesian method for approximating an unknown function under the assumption of smoothness.

BGO is an optimization routine based on GP and consists of the following steps:

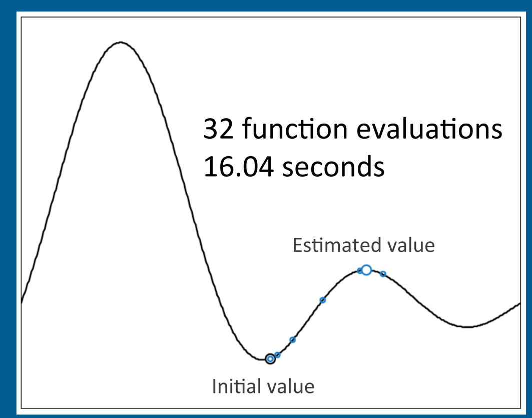
1. We evaluate the function in a set of initial points (red circles) and approximate the resulting values with a GP (whose mean is the red dashed line with the grey area indicating standard deviation).
2. Next, we evaluate the function at the point that has the best combination of estimated mean and variance (point 1) instead of approaching the maximum of the estimated mean (point 2), as in a gradient ascent. This leads to exploration along with exploitation, exploring the whole space even when local maxima have already been found.
3. We adjust the GP approximation including the new point and find the next potential maximum. This procedure is repeated until a certain stopping criterion is satisfied.



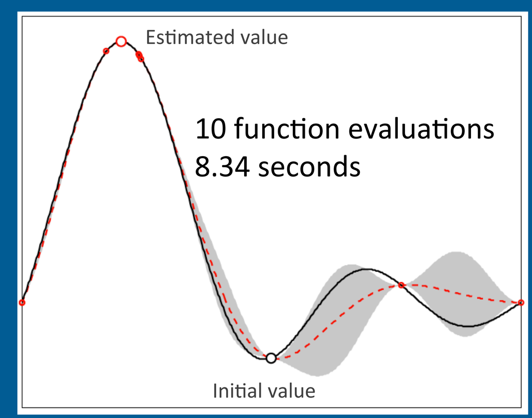
We obtain not only the maximum but also an approximation of the function, which we can use to approximate the integral of the function, i.e., the evidence.

3 Global vs local optimization

- We illustrate here the importance of global optimization by comparing gradient ascent (GA) and Bayesian global optimization (BGO) given a function with two maxima.
- Both GA and BGO were informed with the same initial value placed in the centre of the interval (black circle). However, GA (blue circle) only found a local maximum while BGO (the red dashed line shows the estimation and the red circle shows the maximum) found the global maximum.



Gradient ascent



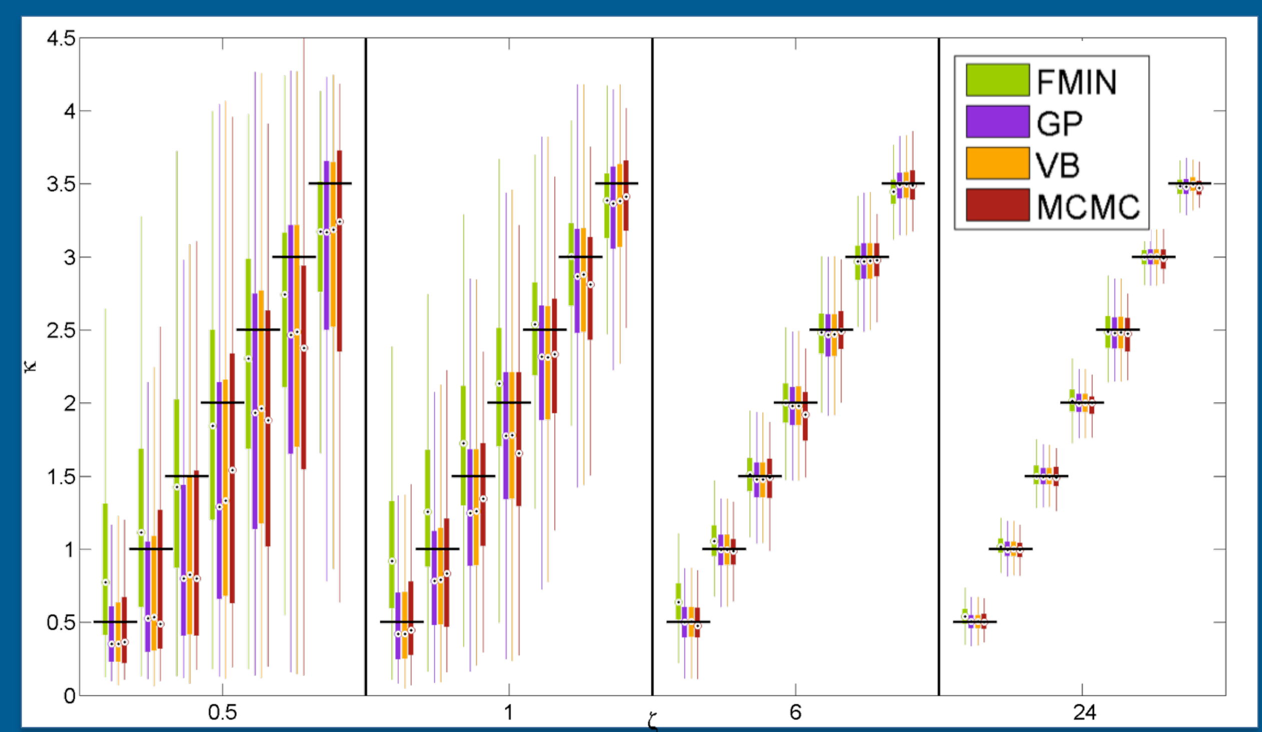
Bayesian global optimization

4 Application: belief-precision model

The belief-precision model deals with quantities whose variability exhibits time-varying volatility. It describes these quantities in terms of a hierarchy of coupled Gaussian random walks [7].

We illustrate the utility of our approach by applying it to a parameter estimation of the belief-precision model. Baseline methods: MCMC, VB, and a local optimization method (FMIN, Nelder-Mead simplex algorithm).

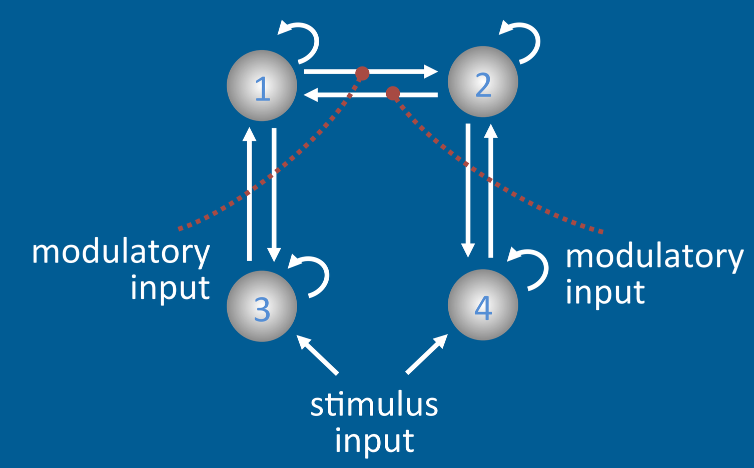
We systematically varied ground truth, generated noisy observations, and tested across 1,000 simulations how well the different methods recovered the true values. For details of the simulations see companion poster # 362 MT [8].

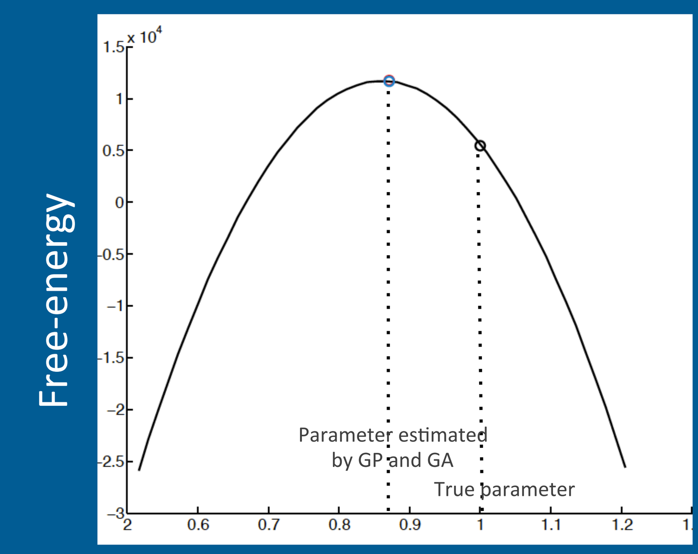


Distributions of estimates of the parameter κ by different methods. Boxplots for all four methods at different noise levels ζ. Higher ζ means less noise. FMIN stands for local optimization, GP for Bayesian Global Optimization, VB for variational Bayes and MCMC for Markov chain Monte Carlo.

5 Application: dynamic casual modelling

- BGO can be used to infer the parameters of dynamic casual models (DCMs).
- We use variational Bayes under the Laplace approximation to optimize parameters with respect to the negative free-energy using BGO as an alternative to the conventional Gauss-Newton gradient ascent.





Coefficient of the right stimulus input

- On the basis of the synthetic DCM above we found in an initial analysis with a reduced parameter space that BGO gives the same results as a conventional gradient ascent. Furthermore, BGO provides greater certainty that the result is indeed the global maximum since it is a global optimization method.

6 Conclusions

- Bayesian global optimization is an approach that can be usefully applied to the problem of inference on computational models, where it unfolds three particular strengths.
- First, being a global optimization method, BGO can deal with multimodal problems and avoids maxima that are merely local.
- Second, it replaces parametric assumptions about the objective function by structural constraints which are typically less restrictive and easier to define.
- Third, BGO is computationally highly efficient, especially when the objective function is expensive to evaluate.

References

1. Brodersen, K.H. et al., 2011. Generative embedding for model-based classification of fMRI data. *PLoS Computational Biology*, 7(6): e1002079. doi:10.1371/journal.pcbi.1002079.
2. Metropolis, N., et al., 1949. The Monte Carlo method. *Journal of American Statistical Association*, 44, 335-341.
3. Beal, M.J., 2003. Variational Algorithms for Approximate Bayesian Inference. PhD Thesis, Gatsby Computational Neuroscience Unit, University College London.
4. Osborne, M.A., et al., 2009. Gaussian processes for global optimization. *3rd International Conference on Learning and Intelligent Optimization (LION3), Trento, Italy*.
5. Friston, K.J., et al., 2003. Dynamic causal modelling. *NeuroImage*, 19(4), 1273-1302.
6. Rasmussen, C.E. et al., 2006. Gaussian Processes for Machine Learning, MIT Press.
7. Mathys, C., et al., 2011. A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, 5(39).
8. Mathys, C., et al., 2012. Parameter estimation in a Bayesian hierarchical model of learning: a comparison of four methods. *HBM 2012*, Abstract #6354.