← Go to **ICML 2023 Conference** homepage (/group?id=ICML.cc/2023/Conference)

Bayesian Spatio-temporal Model Inversion With METHO

(/pdf?id=MaqhJC9IiJ)

Dan MacKinlay (/profile?id=~Dan_MacKinlay1), Daniel Edward Pagendam (/profile?id=~Daniel_Edward_Pagendam1), Russell Tsuchida (/profile?id=~Russell_Tsuchida1), Petra Kuhnert (/profile?email=petra.kuhnert%40data61.csiro.au), Sreekanth Janardhanan (/profile?email=sreekanth.janardhanan%40csiro.au)

26 Jan 2023, 17:27 (modified: 11 Feb 2023, 06:45) ICML 2023 Conference Submission Conference, Area Chairs, Authors, Reviewers, Senior Area Chairs Revisions (/revisions?id=MaqhJC9IiJ)

Keywords: Gaussian process, Model inversion, State Space Models, Inverse problems, Monte Carlo, Physics, Bayesian inference, Spatiotemporal

TLDR: We derive a sample-wise update for Ensemble Filters based on the Matheron trick, which enables efficient inference in very high dimensional state

Abstract:
We study the problem of stochastic model inversion for physical dynamical systems. This paper introduces MaTHerOn Ensemble inversion (METHO), a simple and efficient means of using a forward prediction operator to sample from a posterior distribution over unobserved functionvalued system parameters. The cost of the resulting algorithm scales linearly with the dimension of the estimands and does not require the calculation or storage of large covariance matrices. We do not require exact likelihood evaluations of the prior or posterior samples. However, we do require the ability to simulate from the prior, and sufficient regularity of the forward operator. Our method achieves orders-of-magnitude speedup over a classical physical model inversion approach for high-dimensional models, while maintaining acceptable accuracy. This increases the dimensionality of the estimand at which inference is feasible, without resorting to dimension reduction methods.

Supplementary Material: zip (/attachment?id=MaqhJC9IiJ&name=supplementary_material)
Paper Checklist Guidelines: I certify that all co-authors of this work have read and commit to adhering to the Paper Checklist Guidelines, Call for Papers and Publication Ethics.
Submission Number: 3833

Official Review by Reviewer fxbi

Official Review Reviewer fxbi 06 Mar 2023, 06:37 (modified: 14 Mar 2023, 21:40)
Program Chairs, Area Chairs, Authors, Reviewer fxbi, Submitted, Senior Area Chairs

Summary:
The paper proposes a method for model inversion of dynamical model with support for stochastic processes. The approach uses an ensemble method using the ideas from path-wise conditioning of GPs to efficiently perform updates of the underlying Gaussian representation. Experiments show that the method converges better than a simple linearizing approach.

Strengths And Weaknesses:
As model inversion is not my research area I can not assess how this work fits within the general area with regards to sensible comparisons and novelty in the field. As such the review is more focused on the idea, understandability, and evaluation.

The overall presentation of the paper is good and it introduces the used concepts and required background logically and in a concise yet understandable manner. While the way in which the method is introduced is somewhat unorthodox, i.e. the same section describes other typical approaches to model inversion before describing the proposed one. The reasoning for it is visible in that doing so gives the required context that would be hard or inconvenient to convey if done differently.

Regarding the method some questions that arise. The linearization is mentioned as an issue, would similar tricks as used in unscented Kalman filters, to improve over a standard EKF, be applicable here as well? While moving the algorithm to the appendix is understandable for space reasons it likely would improve how easy it is to grasp how everything is tied together, so if possible in some way moving the algorithm to the main body would be great. Additionally, currently the algorithm uses function names for the individual computations, ideally this would be the actual equations or point to the equations in the paper.

The experiments appear to be on a single problem set and do not provide much in terms of comparisons. The results show that the method outperforms the used baseline and that it can handle the high-dimensional stochastic case. As mentioned in the beginning, this is not my field of research so I cannot say what good comparisons or problem setups are. Though if there are other sensible baselines, even if significantly slower it would seem sensible to include them here.

Questions:
In the introduction the paper mentions that the method approximates observations as Gaussian variates. How limiting is this assumption for real-world problems and how does it impact solution quality when this assumption is not met?

Limitations:
The paper does not address possible negative societal impacts or limitations of their method.

Ethics Flag: No
Soundness: 3 good
Presentation: 3 good
Contribution: 2 fair
Rating: 5: Borderline accept: Technically solid paper where reasons to accept outweigh reasons to reject, e.g., limited evaluation. Please use sparingly.
Confidence: 2: You are willing to defend your assessment, but it is quite likely that you did not understand the central parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.
Code Of Conduct: Yes

Add: **Official Comment** **Rebuttal**

Official Review by Reviewer 3T1w

Official Review Reviewer 3T1w 05 Mar 2023, 07:45 (modified: 14 Mar 2023, 21:40)
Program Chairs, Area Chairs, Authors, Reviewer 3T1w, Submitted, Senior Area Chairs

Summary:
This paper focuses on the model inversion for dynamical systems (in the context of the paper, identifying hidden parameters in the dynamics function that are time-invariant but differ across each time series, from the corresponding trajectory of observations). In contrast to existing methods that compute exact posterior but are hard to scale to large-dimension hidden variables, the proposed method uses an ensemble of samples to represent

the posterior and updates the ensemble sample-wisely with Matheron update, thus greatly improving computation efficiency and scalability.

Strengths And Weaknesses:
Identifying static (time-invariant) but trajectory-specific hidden parameters is an important and relevant problem in dynamics learning. The method proposes to use an ensemble of samples to approximate the posterior and the update can be computed sample-wise with Matheron method.

Strengths:

1. The technical part of the paper is sound and the proposed method significantly reduces the computational complexity of the model inversion from $O(D_z^3)$ (D_z: hidden variable dimension) to $O(N^2 D_z)$ (N: # of samples in the ensemble), enabling it to solve problems with hundreds of dimensions.
2. The paper is well-written and easy to read. I especially appreciate the detailed introduction to existing methods and the background of the techniques applied to model inversion.

Weakness:

1. The novelty of the proposed method is unclear and seems to be minimal. IIUC, the proposed method applies existing techniques (ensemble approximation, Matheron updates) to model inversion, while from the last paragraph of the intro, it seems Ensemble Kalman filter has already applied similar techniques to dynamic systems to discover states which is a harder task.
2. The assumption of having a forward operator only holds for limited scenarios.
3. The experiment evaluations are limited, only conducted on two simulated cases, with linearised Gaussian Process as the only baseline.

Questions:
For suggestions,

1. I hope the authors could discuss more the assumption of having a forward operator: when we do have access to it, if not, how we can approximate it, etc.
2. For the experiment datasets, I wonder if the authors could test on more realistic (or real-world) datasets, or discuss how the used Navier-stoke dataset is challenging to make the results more convincing.
3. For other baselines, though authors mention conceptually in Sec 4.1 why optimization methods may have worse performance than the proposed method (e.g., point estimate doesn't take uncertainty into consideration), I wonder if it's possible to empirically show that. After all, Sec 5.1 is evaluated on a deterministic system only. Or if other methods need extra assumptions, e.g., access to likelihood, it would be more compelling to see that your method achieves similar performance without those assumptions)

Limitations:
As discussed in suggestions point 1, I would see more discussion of the assumption of having access to the forward operator and linearization of the dynamic systems.

Ethics Flag: No
Soundness: 3 good
Presentation: 3 good
Contribution: 3 good
Rating: 4: Borderline reject: Technically solid paper where reasons to reject, e.g., limited evaluation, outweigh reasons to accept, e.g., good evaluation. Please use sparingly.
Confidence: 2: You are willing to defend your assessment, but it is quite likely that you did not understand the central parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.
Code Of Conduct: Yes

Add: **Official Comment** **Rebuttal**

Official Review by Reviewer KT5D

Official Review Reviewer KT5D 01 Mar 2023, 18:37 (modified: 14 Mar 2023, 21:40)

Program Chairs, Area Chairs, Authors, Reviewer KT5D, Submitted, Senior Area Chairs

Summary:

The paper introduces a Bayesian ensemble method for model inversion, that is, for the ifnerence of unknown latent variables (u) from observations (z_t) that follow some (potentially stochastic) forward operator (z_ = P(z_t, u, noise)). The paper provides helpful background, and compares the proposed method to an extended Kalman filter-type alternative.

Strengths And Weaknesses:

Strengths

The paper is very well structured and written. The presentation really makes it easy for the reader to understand the proposed method well, and it leaves no open questions regarding how the proposed method works.

The proposed method itself also appears sound, its motivation is clear, its linear scaling is clear, and the results demonstrate that it is indeed able to perform model inversion.

Weaknesses

My two main concerns are with respect to the empirical evaluation and to the originality of the approach.

On originality

I am not familiar enough with the data assimilation literature, but from my understanding the ensemble Kalman filter (EnKF) is a very commonly used method there, and the proposed method seems *very* closely related to this method. Vaguely speaking, it almost seems like the method *is* actually an EnKF. It would therefore be very helpful to discuss similarities and differences between "METHO" and commonly used EnKF methods in more detail, to clarify such questions and help readers situate the paper better.

On the evaluation

METHO is compared only to the "linearised Gaussian method", which to me seems to be (at the very least closely related to) an extended Kalman filter (EKF). In an ideal comparison, both methods aim to compute the same quantity, and we could then compare how well they approximate this quantity, and how much computational effort was necessary. But since they differ in their use of σ, their probabilistic model differs, and they compute different posteriors. I personally see two improvements that would, from my perspective, significantly strengthen the experiments:

1. Incude results of METHO for $\sigma^2 = 0.1$, which could then be compared to the EKF.
2. Implement the EKF in square-root form for improved numerical stability. Then, much lower σ^2 should be possible without running into numerical issues. Note that the square-root EKF typically requires some QR decompositions, which hurt its the runtime (and thus in comparison, METHO might have an advantage)

Generally speaking, I also believe that having more quantitative evaluations would be very beneficial; right now only Figure 4 compares the proposed method to an alternative. This includes both having more different experimental setups, but maybe more importantly also comparing to other methods (such as those EnKF approaches commonly used in data assimilation; if applicable).

Questions:
As mentioned above, my main concern is with the relation of METHO to the ensemble Kalman filtering methods used in data assimilation; a clarification by the authors on their similarities and differences would be very helpful.

On *4.1 Inversion by optimization*: The end of the second paragraph discusses the lack of uncertainties in MAP estimates. I believe that "Laplace approximations" should be at least mentioned in this context, as they target exactly this setup: First computing a MAP with optimization, and then forming a Gaussian approximate posterior by essentially just computing Hessians.

Other minor comments:

- Equation 5: I believe w and w should be flipped; see Wilson 2021, Eq 4.
- L. 191: "(IS))"
- Eq. 16 introduces $\bar{Z}_t$ and $\breve{Z}_t$, which I think were not introduced.
- "This cost [$O(N^2 D_z + N^3)$] is favourable in comparison with $O(D_z)$": The last part is probably a typo and should probably be "$O(D_z^3)$".

Limitations:
There does not seem to be a dedicated "limitations" section in the paper, and I could not really find much discussion on drawbacks of the proposed method; the limited empirical evaluation also does not make it easy for readers to

assess these. More information on this would be helpful.

Ethics Flag: No
Soundness: 2 fair
Presentation: 4 excellent
Contribution: 2 fair
Rating: 5: Borderline accept: Technically solid paper where reasons to accept outweigh reasons to reject, e.g., limited evaluation. Please use sparingly.
Confidence: 3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.
Code Of Conduct: Yes

Add: **Official Comment** **Rebuttal**