Inference for Dynamic and Latent Variable Models via Plug-and-Play Automatically Differentiable Particle Filtering

Kevin Tan[a], **Giles Hooker**[a], **and Edward L. Ionides**[b,1]

This manuscript was compiled on November 2, 2023

Automatic differentiation has driven recent advances in machine learing, including deep neural networks and Hamiltonian Markov Chain Monte Carlo methods. This progress has required simultaneous advances in algorithms, software and hardware. Partially observed nonlinear stochastic dynamic systems have proved resistant to the benefits of automatic differentiation techniques; despite various attempts, widely applicable methods have not yet emerged. We present a new approach which is applicable to a general class of models, possesses a theoretical foundation, and is demonstrated to beat current state-of-the-art methods on a challenging scientific benchmark problem. Our algorithm is compatible with parallel computation on a graphical processing unit, and enjoys the plug-and-play property that its software implementation requires only a simulator for the scientific dynamic model as input (in addition to data, and a measurement model).

Sequential Monte Carlo | Automatic Differentiation | Particle Filter | Markov Process | Maximum Likelihood

The particle filter is a widely used Monte Carlo algorithm which provides an unbiased estimate of the likelihood function for partially observed Markov process (POMP) models . Maximization of the likelihood permits computation of the maximum likelihood parameter estimate, as well as profile likelihood confidence intervals, and model selection by likelihood ratio tests or Akaike's information criterion. The particle filter involves a random resampling step in which particles inconsistent with the data are likely to be pruned out and the most consistent particles are replicated. Applying automatic differentiation (AD) to supply the optimization with gradient information is complicated by the disctete nature of this resampling step. Simulations of discrete random variables are necessarily discontinuous, specifically piecewise constant, as a function of the unknown parameters for a fixed value of the seed for the random number generator. As a result of this, the Monte Carlo expectation of the derivative is not equal to the derivative of the Monte Carlo expectation. Various strategies have been proposed to address this phenomenon , but none of these algorithms have found widespread practical applicability.

A significant practical strength of the particle filter is its applicability to arbitrarily nonlinear stochastic systems, requiring only a simulator for the dynamic model. This is called the plug-and-play property , and this property is critical in some applications, such as disease modeling, where the models are complex enough that obtaining the density is intractable. Some extensions of the particle filter to parameter estimation require availability of transition probabilities for the dynamic model which violates the plug-and-play property. We have developed an AD algorithm for the particle filter which preserve the plug-and-play property while providing useful derivative estimates which demonstrably improve our ability to search for the maximum of the likelihood.

Our key insights are as follows:

1. We consider a novel one-parameter family of plug-and-play AD gradient estimators allowing flexibility in the tradeoff between Monte Carlo bias and variance. For our target problem, the results are insensitive to the choice of this parameter.

2. ADPF stochastic gradient optimization has complementary strengths to iterated filtering algorithms that are widely used for plug-and-play likelihood maximization

Significance Statement

Many scientific models involve highly nonlinear stochastic dynamical systems which can be observed only via noisy and incomplete measurements. Under the Markov assumption on system dynamics, previous work has provided methods of performing inference for these models. In particular, prior to this work, iterated filtering algorithms were the only class of algorithms for maximum likelihood estimation that did not require access to the system's transition probabilities, instead needing only a simulator of the system dynamics. We leverage recent advances in automatic differentiation to propose a hybrid algorithm that requires only a differentiable simulator for maximum likelihood estimation. Our new method outperforms previous approaches on a challenging problem in epidemiology.

Author affiliations: [a]University of Pennsylvania; [b]University of Michigan

K.T. and E.L.I. planned the study; K.T. developed the numerical results; all authors wrote the manuscript.

The authors declare no competing interests.

[1]To whom correspondence should be addressed. E-mail: ionides@umich.edu

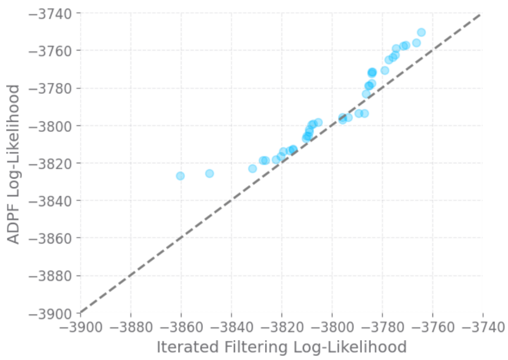

Fig. 1. Comparing the ADPF algorithm with IF2 on the Dacca model and data of (1)

for POMP models. Iterated filtering algorithms provide a computationally efficient way to approach the maximum, but they are poor at identifying the precise location of the maximum. Combining both these algorithms substantially out-performs either alone.

3. The IFAD algorithm can be implemented on a GPU to achieve massive parallelization.

Together, these innovations lead to methodology which advances capability for statistical analysis of scientific models arising in biological sciences and elsewhere . Our benchmark task is a model previously used to evaluate iterated filtering algorithms .

Historical cholera transmission in Dacca, Bangladesh. . We tested our method on the cholera transmission model and data of (1). This is a POMP model based on the fundamental Susceptible-Exposed-Infected-Recovered (SEIR) epidemiological model, but incorporating additional features that are required for practical application of this model to various real disease systems. Seasonality is included as a periodic cubic B-spline covariate process. Loss of immunity is modeled by return from the Recovered class to the Susceptible class. Asymptomatic infections can be modeled separately to allow for the possibility that they differ from symptomatic infections in their infectivity and acquired immunity. This model, and the Dacca case reports to which it was originally fitted, provide a benchmark for a challening scientific model fitting exercise. It was used oringally to test IF1 (2) and IF2 (3), and a simplified version has been used to compare with alternative methodologies (4). We will therefore benchmark on this well-established model before moving on to target the methods on models with anticipated scientific goals.

Assorted draft stuff. [naive ADPF] intruduces bias into the Monte Carlo estimate of the derivative. This bias has previously been reported to be non-negligible [REF]. An approach to avoiding this bias is the REINFORCE estimator, also known as the stop-gradient trick (5). The Monte Carlo variance of this estimator has been found to be problematic (5)

1. A new derivation of the stop-gradient estimator. We show that this estimator can be obtained as the fixed seed derivative of an extension of the basic particle filter which we call the off policy particle filter.

2. Combining the reparameterization trick with the stop-gradient estimator to obtain a plug-and-play gradient estimator.

3. Apply this gradient estimator within a stochastic gradient descent algorithm using a normalized gradient.

4. Multiple searches from diverse starting points to reduce, quantify and control maximization error due to multimodality and Monte Carlo variability.

Many approaches to inference in highly nonlinear stochastic dynamical systems assume access to the probability density of next states given the current state.

Still, maximum likelihood parameter estimation can be challenging, especially when the Monte Carlo variance of the evaluation is high and the number of parameters is not small. Existing methods like the improved iterated filtering algorithm of Ionides et. al. (3) converge quickly to a neighborhood of the MLE, but struggle to optimize the last few units of log-likelihood.

Algorithmic differentiation potentially facilitates numerical optimization, but currently its use for particle filters is limited. We investigate ways to use algorithmic differentiation of particle filters within the confines of the plug-and-play property, with the goal of enhancing current inference capabilities for general POMP models.

The issue: Particle filters provide convenient approaches to evaluating the log-likelihood function for partially observed Markov process (POMP) models. However, using this evaluator to obtain a maximum likelihood parameter estimate can be challenging – especially when the Monte Carlo variance of the evaluation is high and the number of parameters is not small. Empirically, methods such as the improved iterated filtering algorithm (IF2) from Ionides et. al. (3) rapidly converge to a neighborhood of the optimum but struggle at finding the exact optimum due to Monte Carlo variance, even with an annealing random walk standard deviation.

A potential solution to this could lie in auto-differentiation (AD). This would allow for the use of first and second-order iterative optimization techniques. However, though AD could potentially facilitate numerical optimization, its use for plug-and-play particle filters has so far been limited. This is potentially because particle filtering methods are inherently non-differentiable due to the resampling step that may take place in between iterations.

The importance of the plug-and-play property:. Performing inference in highly nonlinear stochastic dynamical systems is a challenging problem. Although many methods for inference assume access to the density of state transitions, this is often not available, especially in critical applications like epidemiology.

Basic particle filtering algorithms do not require evaluation of the transition density of the latent Markov process, in a feature known as the **plug-and-play property** (6) since it enables an arbitrary model simulator to be plugged into the algorithm. We investigate ways to use algorithmic differentiation of particle filters within the confines of the plug-and-play property, with the goal of enhancing current inference capabilities for general POMP models.

Other potential applications:. This has applications beyond the obvious one of learning model parameters via first or second-order optimization routines. For example, it could be

Table 1. Comparison of the fitted potential energy surfaces and ab initio benchmark electronic energy calculations

Species	CBS	CV	G3
1. Acetaldehyde	0.0	0.0	0.0
2. Vinyl alcohol	9.1	9.6	13.5
3. Hydroxyethylidene	50.8	51.2	54.0

nomenclature for the TSs refers to the numbered species in the table.

a step towards developing very general Hamiltonian Monte Carlo methods for particle MCMC, as Rosato et. al. (7) do by using previous work such as (5, 8) (and we conjecture that the seed-fixing derivatives of Rosato et. al. are the same as these as an immediate consequence of section **??**) to differentiate the particle filter.

Format. Many authors find it useful to organize their manuscripts with the following order of sections: title, author line and affiliations, keywords, abstract, significance statement, introduction, results, discussion, materials and methods, acknowledgments, and references. Other orders and headings are permitted.

Data Archival. PNAS must be able to archive the data essential to a published article. Where such archiving is not possible, deposition of data in public databases, such as GenBank, ArrayExpress, Protein Data Bank, Unidata, and others outlined in the Information for Authors, is acceptable.

Supporting Information Appendix (SI). Authors should submit SI as a single separate SI Appendix PDF file, combining all text, figures, tables, movie legends, and SI references. SI will be published as provided by the authors; it will not be edited or composed. Additional details can be found in the PNAS Author Center. The PNAS Overleaf SI template can be found here. Refer to the SI Appendix in the manuscript at an appropriate point in the text. Number supporting figures and tables starting with S1, S2, etc.

Authors who place detailed materials and methods in an SI Appendix must provide sufficient detail in the main text methods to enable a reader to follow the logic of the procedures and results and also must reference the SI methods. If a paper is fundamentally a study of a new method or technique, then the methods must be described completely in the main text.

SI Datasets. Supply .xlsx, .csv, .txt, .rtf, or .pdf files. This file type will be published in raw format and will not be edited or composed.

Materials and Methods

Please describe your materials and methods here. This can be more than one paragraph, and may contain subsections and equations as required.

Subsection for Method. Example text for subsection.

ACKNOWLEDGMENTS. Please include your acknowledgments here, set in a single paragraph. Please do not include any acknowledgments in the Supporting Information, or anywhere else in the manuscript.

1. AA King, EL Ionides, M Pascual, MJ Bouma, Inapparent infections and cholera dynamics. *Nature* **454**, 877–880 (2008).
2. EL Ionides, C Bretó, AA King, Inference for nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **103**, 18438–18443 (2006).
3. EL Ionides, D Nguyen, Y Atchadé, S Stoev, AA King, Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proc. Natl. Acad. Sci. USA* **112**, 719—-724 (2015).
4. M Fasiolo, N Pya, SN Wood, A comparison of inferential methods for highly nonlinear state space models in ecology and epidemiology. *Stat. Sci.* **31**, 96–118 (2016).
5. A Ścibior, F Wood, Differentiable particle filtering without modifying the forward pass. *arXiv:2106.10314* (2021).
6. C Bretó, D He, EL Ionides, AA King, Time series analysis via mechanistic models. *Annals Appl. Stat.* **3**, 319–348 (2009).
7. C Rosato, J Harris, J Panovska-Griffiths, S Maskell, Inference of stochastic disease transmission models using particle-mcmc and a gradient based proposal in *25th International Conference on Information Fusion (FUSION 2022).* (IEEE), pp. 1–8 (2022).
8. G Poyiadjis, A Doucet, SS Singh, Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika* **98**, 65–80 (2011).