Statistica Sinica **34** (2024), 1-21
doi:https://doi.org/10.5705/ss.202022.0172

A LANGEVINIZED ENSEMBLE KALMAN FILTER FOR LARGE-SCALE DYNAMIC LEARNING

Peiyi Zhang, Qifan Song and Faming Liang*

Purdue University

Abstract: The ensemble Kalman filter (EnKF) performs well in terms of data assimilation in atmospheric and oceanic sciences. However, it fails to converge to the correct filtering distribution, which precludes its use for uncertainty quantification in dynamic systems. Thus, we reformulate the EnKF under the framework of Langevin dynamics, yielding a new particle filtering algorithm, which we call the Langevinized EnKF (LEnKF). The LEnKF inherits the forecast-analysis procedure from the EnKF, and uses mini-batch data from stochastic gradient Langevin dynamics (SGLD). We prove that the LEnKF is a sequential preconditioned SGLD sampler, like the EnKF, but with its execution accelerated by the forecast-analysis procedure. Furthermore, the LEnKF converges to the correct filtering distribution in terms of the 2-Wasserstein distance as the number of iterations per i stage increases. We demonstrate the performance of the LEnKF using a variety of examples. The LEnKF is not only scalable with respect to the state dimension and the sample size, but also tends to be immune to sample degeneracy for long-series dynamic data.

Key words and phrases: Data assimilation, inverse problem, state space model, stochastic gradient Markov chain Monte Carlo, uncertainty quantification.

1. Introduction

The integration of computer technology into science and daily life has enabled scientists to collect massive volumes of data, such as climate data, high-throughput biological assay data, and website transaction logs. To address the computational difficulties that arise in Bayesian analyses of big data, several scalable MCMC algorithms have been developed, including the stochastic gradient MCMC algorithms (Welling and Teh (2011); Ding et al. (2014); Chen, Fox and Guestrin (2014); Li et al. (2016); Ma, Chen and Fox (2015); Nemeth and Fearnhead (2021)), split-and-merge algorithms (Scott et al. (2016); Li, Srivastava and Dunson (2017); Srivastava, Li and Dunson (2018)), mini-batch Metropolis–Hastings algorithms (Chen et al. (2018); Maclaurin and Adams (2014); Bardenet, Doucet and Holmes (2017)), and nonreversible Markov process-based algorithms (Bierkens, Fearnhead and Roberts (2019); Bouchard-Côté, Vollmer and Doucet (2018)).

*Corresponding author