

SAMPLING VIA GRADIENT FLOWS IN THE SPACE OF PROBABILITY MEASURES

Andrew Stuart

Caltech, United States

astuart@caltech.edu

Sampling a probability distribution with an unknown normalization constant is a fundamental problem in computational science and engineering. This task may be cast as an optimization problem over all probability measures, and an initial distribution can be evolved to the desired minimizer (the target distribution) dynamically via gradient flows. Mean-field models, whose law is governed by the gradient flow in the space of probability measures, may also be identified; particle approximations of these mean-field models form the basis of algorithms. The gradient flow approach is also the basis of algorithms for variational inference, in which the optimization is performed over a parameterized family of probability distributions such as Gaussians, and the underlying gradient flow is restricted to the parameterized family. By choosing different energy functionals and metrics for the gradient flow, different algorithms with different convergence properties arise. In this work, we concentrate on the Kullback–Leibler divergence as the energy functional after showing that, up to scaling, it has the unique property (among all f -divergences) that the gradient flows resulting from this choice of energy do not depend on the normalization constant of the target distribution. For the metrics, we focus on variants of the Fisher–Rao, Wasserstein, and Stein metrics; we introduce the affine invariance property for gradient flows, and their corresponding mean-field models, determine whether a given metric leads to affine invariance, and modify it to make it affine invariant if it does not.

We study the resulting gradient flows in both the space of all probability density functions and in the subset of all Gaussian densities. The flow in the Gaussian space may be understood as a Gaussian approximation of the flow in the density space. We demonstrate that, under mild assumptions, the Gaussian approximation based on the metric and through moment closure coincide; the moment closure approach is more convenient for calculations. We establish connections between these approximate gradient flows, discuss their relation to natural gradient methods in parametric variational inference, and study their long-time convergence properties showing, for some classes of problems and metrics, the advantages of affine invariance. Furthermore, numerical experiments are included which demonstrate that affine invariant gradient flows have desirable convergence properties for a wide range of highly anisotropic target distributions.

Joint work with Y. CHEN and D.Z. HUANG (Caltech), J. HUANG (U Penn) and S. REICH (Potsdam).