Multiplicative Weights Update as a Distributed Constrained Optimization Algorithm: Convergence to Second-order Stationary Points Almost Always

Ioannis Panageas   Georgios Piliouras   Xiao Wang

Singapore University of Technology and Design

June, 2019

Motivation

- Non-concave maximization has been the subject of much recent study in the optimization and machine learning communities. Constrained maximization is of importance in many applications such as hidden Markov model and game theory.
- Results in [Lee et al., 2017] suggest the avoidance of saddle points for deterministic first-order methods with random initialization. How ever, the approach has limitations in certain constrained cases.
- Question: Is there a provable convergence for the problems of the form

$$\max_{\mathbf{x} \in D} P(\mathbf{x}),$$

where P is a non-concave, twice continuously differentiable function and D is the product of simplices, i.e., $D = \Delta_1 \times ... \times \Delta_n$?

Key Techniques

- A classic algorithm for simplicial constrained maximization is Baum-Eagon algorithm [Baum and Eagon, 1967]:

$$x_{ij}^{t+1} = x_{ij}^t \frac{\frac{\partial P}{\partial x_{ij}}\Big|_{\mathbf{x}^t}}{\sum_s x_{is}^t \frac{\partial P}{\partial x_{is}}\Big|_{\mathbf{x}^t}}. \tag{1}$$

  (1) is not a diffeomorphism in general.
- We use MWU as an instance of Baum-Eagon algorithm with learning rates:

$$x_{ij}^{t+1} = x_{ij}^t \frac{1 + \epsilon_i \frac{\partial P}{\partial x_{ij}}\Big|_{\mathbf{x}^t}}{1 + \epsilon_i \sum_s x_{is}^t \frac{\partial P}{\partial x_{is}}\Big|_{\mathbf{x}^t}}. \tag{2}$$

  (2) is a diffeomorphism with small ϵ_i.
- Use of Center-stable Manifold Theorem [Lee et al., 2017].

Our Results

Combining the classification of stationary points with constraints and the Center-stable Manifold Theorem to MWU, we prove the following:

- Assume that P is twice continuously differentiable in a set containing D. There exists small enough fixed stepsizes ϵ_i such that the set of initial conditions $\mathbf{x}^0$ of which the MWU dynamics converges to fixed points that violate second order KKT conditions is of measure zero.
- Assume μ is a measure that is absolutely continuous with respect to the Lebesgue measure and P is a rational function (fraction of polynomials) that is twice continuously differentiable in a set containing D, with isolated stationary points. It follows that with probability one (randomness induced by μ), MWU dynamics converges to second order stationary points.

Thank You!