Last Iterate Convergence of Extra-Gradient Method for Variationally Coherent Min-Max Problems

Rumen Dangovski, Ileana Rugina, Kristian Georgiev

December 14, 2020

Setup

Consider

$$\min_{\mathbf{x}\in\mathbb{R}^{n/2}} \max_{\mathbf{y}\in\mathbb{R}^{n/2}} f(\mathbf{x},\mathbf{y}).$$

Let $z = (x, y)$ and $F = [\nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{y}), -\nabla_{\mathbf{y}} f(\mathbf{x},\mathbf{y})]^T$. Now suppose F is L-Lipschitz and has a Λ-Lipschitz derivative w.r.t. spectral norm.

For f being convex in x and concave in y, we know the Extra-Gradient Method (EG) has its last iterate converge at a rate of $O\left(\frac{1}{\sqrt{T}}\right)$[1].

The goal of our project is to show a convergence bound once we relax f from convex-concave to **variationally coherent** (VC).

Definitions and a rigorous statement of the objective

We say f is VC if f satisfies the variational inequality

$$\langle F(z), z - z^* \rangle \geq 0.$$

If f is VC, $\{z | z = \arg\min f\}$ is convex and compact [2].

If f is VC, f has a global saddle point.

If f is VC, EG converges in last iterate [3] (no rate stated).

We consider the Hamiltonian $||F(z)||^2$ as a measurement of convergence.

Convergence of last iterate of EG for f VC

We show the following key result carries over to the VC case.

Lemma

Let $z^{(i)}$ be the iterates of EG. Then $\sum_{t=1}^{T} ||F(z^{(t)})||^2 \leq \frac{||z^{(0)}-z^*||^2}{\eta^2(1-\eta^2 L^2)}$.

Corollary

For some $t^* \leq T$ we have $||F(z^{(t^*)})|| \leq \frac{||z^{(0)}-z^*||}{\eta^2(1-\eta^2 L^2)} \cdot \frac{1}{\sqrt{T}}$.

It remains to further use smoothness and VC to bound variations of $z^{(T)}$ from $z^{(t*)}$ and conclude rate for last iterate convergence.

References

[1] Golowich, N., Pattathil, S., Daskalakis, C., and Ozdaglar, A. Last iterate is slower than averagediterate in smooth convex-concave saddle point problems, 2020.

[2] Zhou, Zhengyuan, et al. "Stochastic mirror descent in variationally coherent optimization problems." Advances in Neural Information Processing Systems. 2017.

[3] Mertikopoulos, Panayotis, et al. "Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile." arXiv preprint arXiv:1807.02629 (2018).