



Chair: Optimization & Learning

Jérôme MALICK (jerome.malick@univ-grenoble-alpes.fr) with a group of 4 researchers and 4 phd students

Last-iterate convergence for min-max problems: explore aggressively and update conservatively

Work of Yu-Guan He (MIAI PhD) with JM & Raphaël Hertzberg (LIG)
 spot light at NeurIPS 20

Context: GAN training adversarial learning } need to solve { min-max problems

Challenge: convergence issues no guarantees } stochasticity in the 2-player system

Remedy: double stepsize strategy } already found in simple examples

stochastic extra-gradient in the case $V(x^*)=0$ } known by a stochastic oracle $E[V(x)] = V(x^*)$

$$\begin{cases} X_{t+1}^1 = X_t^1 - \eta_t \nabla_x f_t(X_t^1, X_t^2) \\ X_{t+1}^2 = X_t^2 - \frac{\eta_t}{2} \nabla_x g_t(X_t^1, X_t^2) \end{cases}$$

← explore aggressively
 ← update conservatively
 with $\frac{\eta_t}{2} \rightarrow 0$

⇒ global convergence rate + local results } beyond monotonicity/convexity

Illustration: simple example: $\min_x \max_y xy$
 only solution (0,0)
 stochastic extra-gradient never converges
 it does converge with our double stepsize strategy

Federated Learning with Heterogeneous Users: A Superquantile/CVaR Approach

work of Yassine Laguel (UGA PhD) in revision in Machine Learning Journal

Federated Learning: N users with privacy-sensitive data collaboratively learn a model

$$\min_w \frac{1}{N} \sum_{i=1}^N E_{x_i \sim p_i} [f(w, x_i)]$$

Challenge: improving performances of non-conformant users

Solution: Superquantile Federated Learning

$$\min_w \max_{\gamma \in \Delta} E_{x_i \sim p_i} [f(w, x_i)]$$

Algorithm ≈ FedAvg + filtering step
 same properties as FedAvg (unique, privacy, publication...)

Illustration: Classification task by CourNet ENIST dataset (← 1930 users 179 images/user)
 histogram over users of test misclassification error
 FedAvg vs SFL ← our robust result

Multi-agent Online Optimization with Delays Asynchronicity and Adaptivity

general framework multi-agent

work of Yu-Guan He (MIAI PhD) published in JMLR 2022 (Journal of Machine Learning Research)

multi-agent online learning with delays ← computation overload communication latency

2 challenges:
 ① no global info (e.g. no time)
 ② non-monotone feedback sequences

Performance of learning algorithms measured by regret $Reg_T(w) = \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(w)$

Contrib #1: optimal regret is obtained by Nesterov dual averaging

Contrib #2: learning rate policy η_t that adapts to data & delays

at time t , agent $i(t)$ computes $x_t = \arg \min_{x \in \mathcal{X}} \langle g_{i(t)}, x \rangle + h(x)$
 local averaging

Adaptive Sparsification of Communications in Distributed Optimization

work of Dmitry Grishchenko (PhD) published in SIAM Maths of Data Science 2017

Volume of communication is a bottleneck of distributed optim/learning (e.g. PL)
 ⇒ many compression schemes

Contribution: adaptive compression with nonsmooth regularization

$$\min_w \sum_{i=1}^n f_i(w) + R(w)$$

nonsmooth ⇒ structure-enforcing ⇒ we can be compressed
 For free! automatic!
 ex: $R(w) = \|w\|_1 \Rightarrow w$ sparse ⇒ we can be sparsified

Difficulty: sparsify/compress updates without losing information
 take the structure + add random entries (take the support)

we get a concrete proof with biased estimator and asynchronous comm

Collaborations

International collaborations:

- Yurii Nesterov (Louvain-la-Neuve, Belgium)
- Zaid Harchaoui (Univ. Washington, US)

Collaborations with companies: ST Microelectronics, Criteo
 with other MIAI chairs: towards more data efficiency (Julien Mairal) & ML for materials (Massih Amini) & stats by convex optimization (Anatoli Juditsky)

Publications

With at least 2 members of the chair – over the last 2 years

- NeurIPS x6 -- ICML x6
- ICLR x2 -- COLT x3
- JMLR -- MLR
- Maths of OR, Math Programming x3
- SIAM journals: optimization, matrix analysis, data science
- AiStats, CDCx2
- IEEE Trans on Signal Processing