Multi-Agent Online Optimization with Delays

Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, Panayotis Mertikopoulos (UGA)

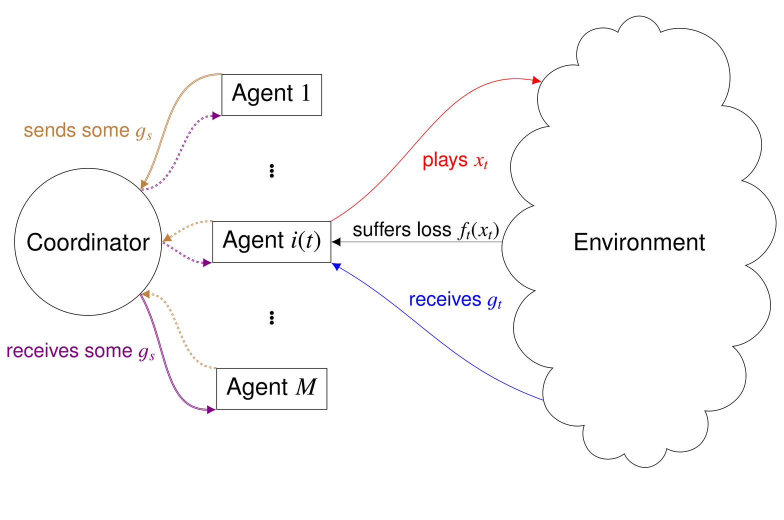
Summary. • Regret minimization in **multi-agent** environments with **delayed** feedback • Quantification of the impact of delay for an algorithm that aggregates the received feedback through **dual averaging** • Design of **adaptive** methods run exclusively based one **local information**

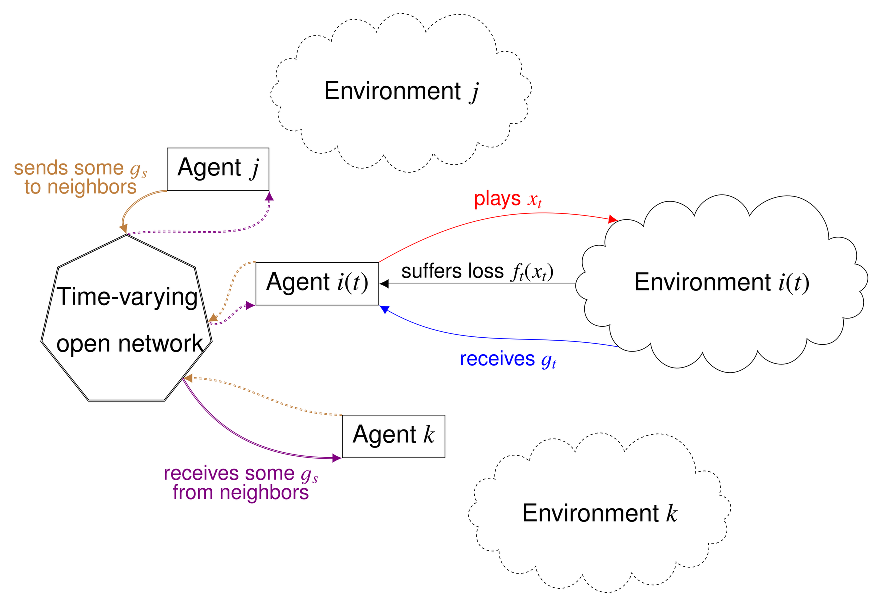
Motivation

- Increasing need for learning in a distributed fashion and in real time
  - Geographically distributed large-scale learning systems
  - Multi-agent systems deployed in a dynamic environment
- The feedback in the system is often delayed
  - inherent delays, computation delays, communication delays

Problem Setup

Online learning in multi-agent systems





- Agents $\mathcal{M} = \{1, ..., M\}$; shared constrained set $\mathcal{X}$
- At each time slot $t = 1, 2, \ldots$, an agent $i(t) \in \mathcal{M}$ becomes active
  - The agent is requested to make a prediction $x_t \in \mathcal{X}$
  - The agent incurs a loss $f_t(x_t)$ for some convex f_t
- Regret:

$$\mathrm{Reg}_T(u) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(u)$$

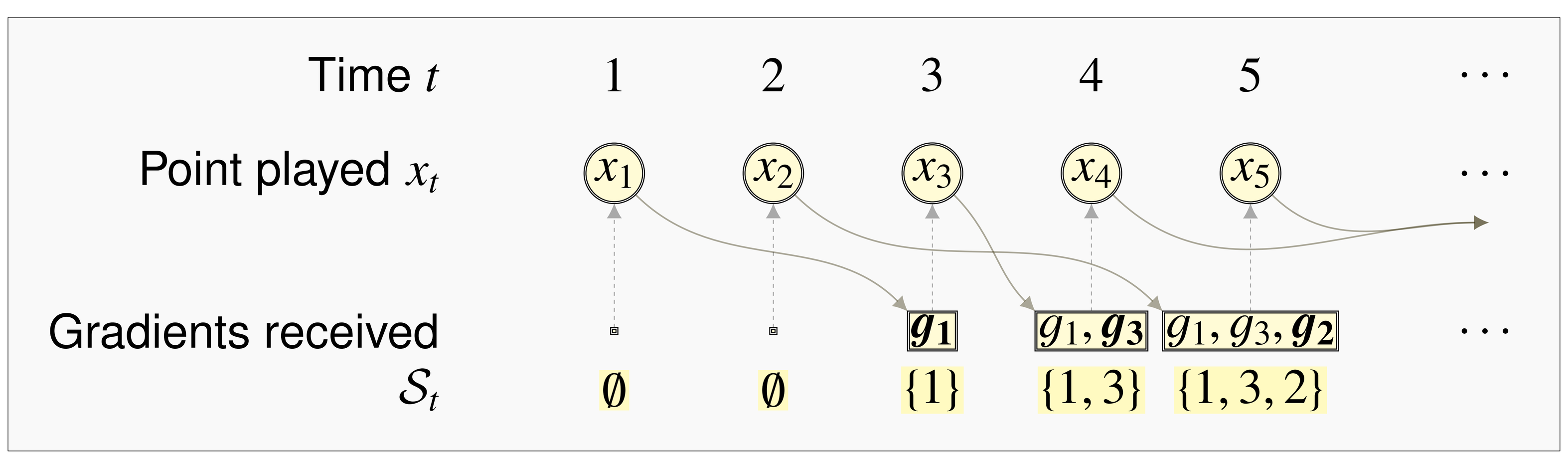
Delayed feedback. Feedback $g_t \in \partial f_t(x_t)$ received by the agents after some agent-dependent delay

- $\mathcal{S}_t^i \subseteq \{1, \ldots, t-1\}$ is the set of gradient timestamps that are available to agent i at time t
- The active agent $i(t)$ can only compute x_t based on $\{g_s : s \in \mathcal{S}_t^{i(t)}\}$
- Let $\mathcal{S}_t = \mathcal{S}_t^{i(t)}$ and $\mathcal{U}_t = \{1, \ldots, t-1\} \setminus \mathcal{S}_t$.

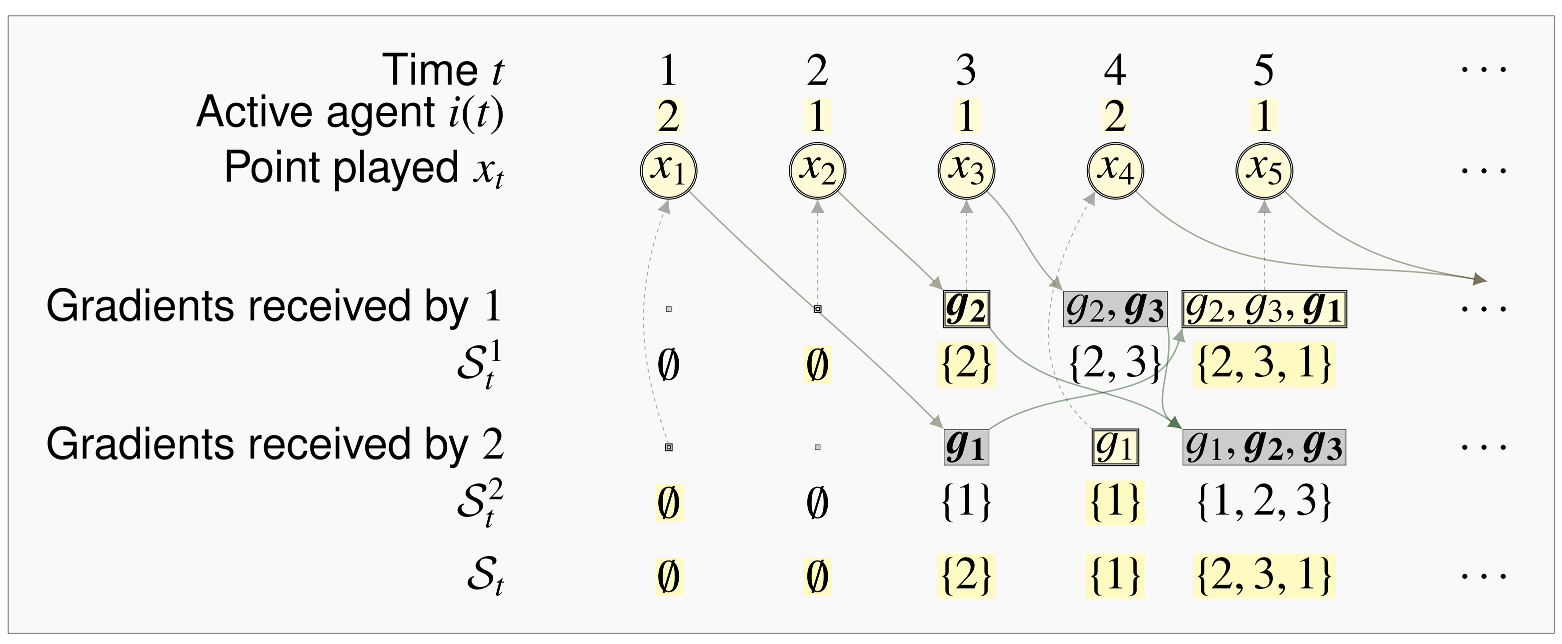
Challenges

- The feedback sequence is non-monotone
- Global information (such as t) is not known by individual agents

Single-agent ($M = 1$)



Multi-agent ($M = 2$)



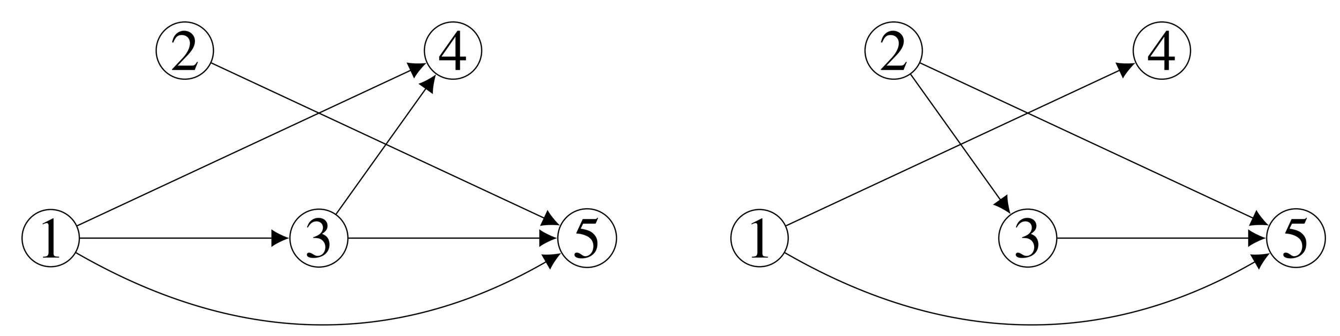
Delayed Dual Averaging

Let $h\colon \mathcal{X} \to \mathbb{R}$ be a regularizer and $Q(y) = \arg\max_{x\in\mathcal{X}}\{\langle y, x\rangle - h(x)\}$ be the mirror map induced by h. The feedback is aggregated with $\eta_t > 0$.

$$x_t = \arg\min_{x\in\mathcal{X}} \left\{ \sum_{s\in\mathcal{S}_t} \langle g_s, x\rangle + \frac{h(x)}{\eta_t} \right\} = Q\left(-\eta_t \sum_{s\in\mathcal{S}_t} g_s \right) \qquad \text{(DDA)}$$

Dependency graph and faithful permutation

- We view each timestamp as a node and include a directed edge from s to t if and only if $s \in \mathcal{S}_t$
- A permutation σ of $\{1, 2, \ldots, T\}$ is faithful if and only if $\sigma(1), \ldots, \sigma(T)$ is a topological ordering of $\mathcal{G}$, i.e., $s \in \mathcal{S}_t$ implies $\sigma^{-1}(s) < \sigma^{-1}(t)$



Impact of Delays

If σ is faithful and $\eta_{\sigma(t+1)} \le \eta_{\sigma(t)}$, the algorithm enjoys the regret bound

$$\mathrm{Reg}_T(u) \le \frac{h(u)}{\eta_{\sigma(T)}} + \frac{1}{2} \sum_{t=1}^{T} \eta_{\sigma(t)} \left(\|g_{\sigma(t)}\|_*^2 + 2\|g_{\sigma(t)}\|_* \sum_{s\in\mathcal{U}_t^\sigma} \|g_s\|_* \right)$$

where $\mathcal{U}_t^\sigma = \{\sigma(1), \ldots, \sigma(t)\} \setminus \mathcal{S}_{\sigma(t)}$

Ideal regret bound

- Maximum delay τ is the longest wait to receive an element of feedback: $\tau = \min\{\tau : \{1, \ldots, t-\tau-1\} \subseteq \mathcal{S}_t$ for all $t \in \{1, \ldots, T\}\}$.
- Maximum unavailability is $\nu = \max_{t\in\{1,\ldots,T\}} \mathrm{card}(\mathcal{U}_t) \le \tau$.
- Cumulative unavailability is $D_T^\sigma = \sum_{s=1}^{T} \mathrm{card}(\mathcal{U}_s^\sigma) \le \nu t$.
- Lag contains pairing terms of $\{\sigma(1), \ldots, \sigma(t)\}$ that are not adjacent to each other in the dependency graph. $\Lambda_t^\sigma = \Lambda_t^{\mathrm{id}}$ if σ is faithful.

$$\Lambda_t^\sigma = \sum_{s=1}^{t} \left(\|g_{\sigma(s)}\|_*^2 + 2\|g_{\sigma(s)}\|_* \sum_{l\in\mathcal{U}_s^\sigma} \|g_l\|_* \right)$$

Proposition. With suitably tuned constant learning rate, we get regret in $\mathcal{O}(\sqrt{\Lambda_T})$, which is in $\mathcal{O}(\sqrt{D_T})$ if feedback is bounded.

Main result: Adaptive learning rate

- Problem: Λ_t^σ is not known at time $\sigma(t)$ due to delays.
- Let $\mathcal{F}_t^i$ be the set of all feedback received before g_t by agent i.
  We approximate Λ_t^σ by $\Gamma_{\sigma(t)}$, where for all t we define

$$\Gamma_t = \sum_{s\in\mathcal{S}_t} \left(\|g_s\|_*^2 + 2\|g_s\|_* \sum_{l\in\mathcal{F}_s^{i(t)}\setminus\mathcal{S}_s} \|g_l\|_* \right).$$

Theorem. If feedback is bounded and $\mathcal{S}_t \subseteq \mathcal{F}_t^i$, i.e., an agent receives a subgradient only after receiving all the subgradients used to compute it, then (DDA) with $\eta_t = 1/\sqrt{\Gamma_t + \beta}$ for some $\beta > 0$ guarantees $\mathrm{Reg}_T(u) = \mathcal{O}(\sqrt{\Lambda_T} + \tau^2)$.