

Anticipating Concept Drift in Online Learning

Michał Dereziński (speaker), Badri Narayan Bhaskar

Online setting: predict with $\hat{\theta}_t \in \Theta$, get loss $f_t(\hat{\theta}_t)$.

Tracking Regret: compare losses to a *good* sequence θ_t :

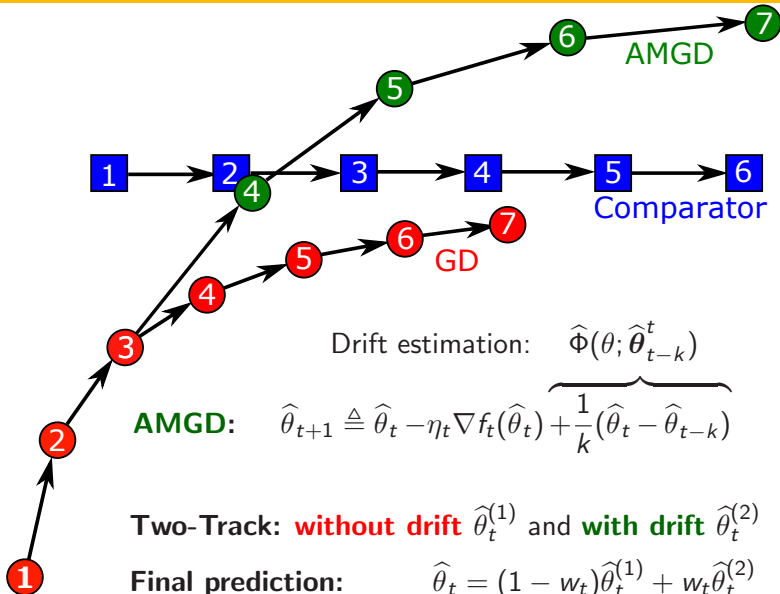
$$R_T(\theta) = \sum f_t(\hat{\theta}_t) - \sum f_t(\theta_t).$$

(Incurred Regret) \propto (Variability of θ_t)

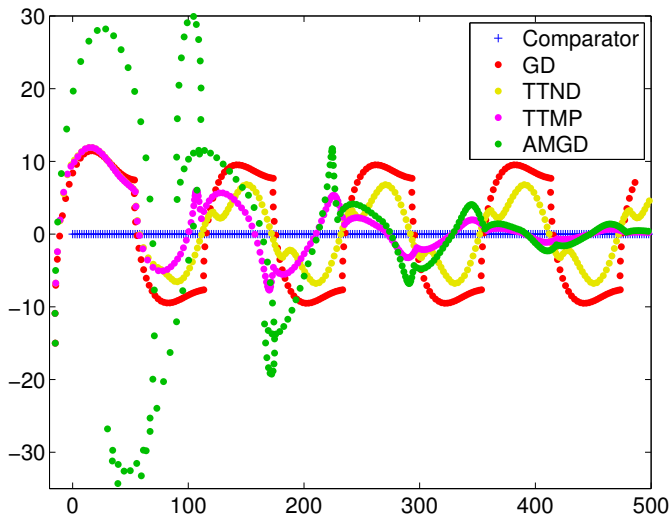
What if the drift trajectory of θ_t can be anticipated?

Use a window of past predictions $\hat{\theta}_{t-k}^t$ to get a drift estimate.

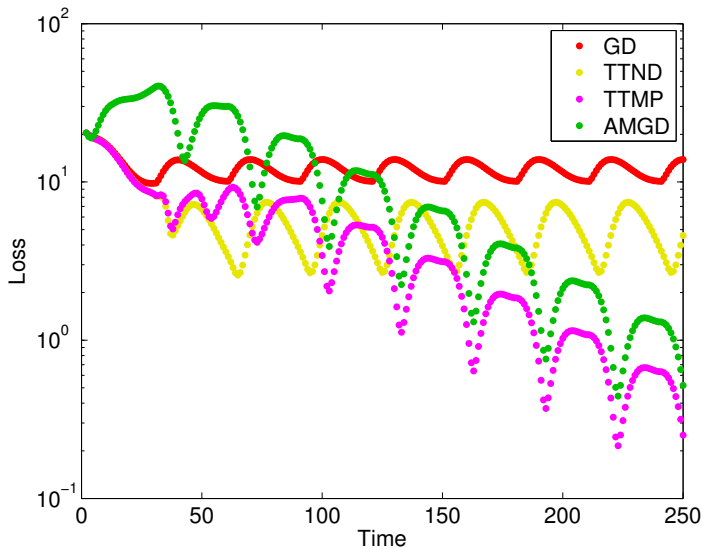
Linear Drift Model



Simulations: Trajectories



Simulations: Losses



Regret bounds for Two-Track:

$$O(\sqrt{T}(1 + V_{\hat{\Phi}}(\theta))) \quad \text{where} \quad V_{\hat{\Phi}}(\theta) \triangleq \sum_{t=1}^T \|\theta_{t+1} - \hat{\Phi}_t(\theta_t)\|$$

How good is the drift estimation $\hat{\Phi}$?

We show bounds for $V_{\hat{\Phi}}(\theta)$ against the optimal Φ^* .

Can we prove linear convergence observed in simulations?

We prove convergence for a special case of AMGD.

Many Open Questions

- ▶ Full regret and convergence **analysis**.
- ▶ Generalization to **nonlinear drift** models.
- ▶ How to select the **learning rate** η ?
- ▶ What if we use **time-stamps** instead of index t ?
- ▶ How to best **avoid instability** in AMGD?