

# Minimax Fixed-Design Linear Regression

Peter L. Bartlett, Wouter M. Koolen, **Alan Malek**,  
Eiji Takimoto, Manfred Warmuth



Conference on Learning Theory  
Paris, France  
July 5th, 2015

## Context: Linear regression

- ▶ We have data  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_T, y_T)$
- ▶ Offline linear regression: predict  $\hat{y} = \theta_T^T \mathbf{x}$ , where

$$\theta_T = (X^T X)^{-1} X^T Y.$$

- ▶ Online linear regression:
  1. We see  $\mathbf{x}_1, \dots, \mathbf{x}_T$  before hand
  2. Need to predict  $\hat{y}_t$  before seeing  $y_t$

# Protocol

Given:  $\mathbf{x}_1, \dots, \mathbf{x}_T \in \mathbb{R}^d$

For  $t = 1, 2, \dots, T$ :

- ▶ Learner predicts  $\hat{y}_t \in \mathbb{R}$ ,
- ▶ Adversary reveals  $y_t \in \mathbb{R}$ ,
- ▶ Learner incurs loss  $(\hat{y}_t - y_t)^2$ .

Figure: Fixed-design protocol

# Minimax

Our goal is to find a strategy that achieves the minimax regret:

$$\min_{\hat{y}_1} \max_{y_1} \cdots \min_{\hat{y}_T} \max_{y_T} \underbrace{\sum_{t=1}^T (\hat{y}_t - y_t)^2}_{\text{algorithm}} - \underbrace{\min_{\theta \in \mathbb{R}^d} \sum_{t=1}^T (\theta^\top \mathbf{x}_t - y_t)^2}_{\text{best linear predictor}}$$

# The Minimax Strategy

- ▶ Is linear

$$\hat{y}_t = \mathbf{s}_{t-1}^\top \mathbf{P}_t \mathbf{x}_t \quad \text{where} \quad \mathbf{s}_t = \sum_{q=1}^t \mathbf{x}_q y_q,$$

- ▶ with coefficients:

$$\mathbf{P}_t^{-1} = \underbrace{\sum_{q=1}^t \mathbf{x}_q \mathbf{x}_q^\top}_{\text{least squares}} + \underbrace{\sum_{q=t+1}^T \frac{\mathbf{x}_q^\top \mathbf{P}_q \mathbf{x}_q}{1 + \mathbf{x}_q^\top \mathbf{P}_q \mathbf{x}_q} \mathbf{x}_q \mathbf{x}_q^\top}_{\text{re-weighted future instances}}.$$

- ▶ Cheap recursive calculation, can be done before seeing  $y_t$ s.
- ▶ Minimax under alignment condition and  $|y_t| \leq B$

# Guarantees

- ▶ If the adversary plays  $y_t$  with

$$\sum_{t=1}^T y_t^2 \mathbf{x}_t^T \mathbf{P}_t \mathbf{x}_t = R,$$

we are minimax against all  $y_t$ s in this set

- ▶ Minimax strategy does not depend on  $R$
- ▶ We achieve regret exactly  $R = O(\log T)$
- ▶ Visit us at the poster session!