

# Linear Bandits

Dávid Pál

Google, New York

&

Department of Computing Science

University of Alberta

dpal@google.com



UNIVERSITY OF  
ALBERTA

November 2, 2011

joint work with Yasin Abbasi-Yadkori and Csaba Szepesvári

# Linear Bandits

In round  $t = 1, 2, \dots$

- ▶ Choose an action  $X_t$  from a set  $D_t \subset \mathbb{R}^d$ .
- ▶ Receive a reward

$$\langle X_t, \theta_* \rangle + \text{random noise}$$

- ▶ Weights  $\theta_*$  are unknown but fixed.
- ▶ Goal: Maximize total reward.

# Motivation

- ▶ exploration & exploitation with side information
- ▶ action = arm = ad = feature vector
- ▶ reward = click

# Outline

- ▶ Formal model & Regret
- ▶ Algorithm:  
Optimism in the Face of Uncertainty principle
- ▶ Confidence sets for Least Squares
- ▶ Sparse models: Online-to-Confidence-Set Conversion

# Formal model

Unknown but fixed weight vector  $\theta_* \in \mathbb{R}^d$ .

In round  $t = 1, 2, \dots$

- ▶ Receive  $D_t \subset \mathbb{R}^d$
- ▶ Choose an action  $X_t \in D_t$
- ▶ Receive a reward

$$Y_t = \langle X_t, \theta_* \rangle + \eta_t$$

Noise is conditionally  $R$ -sub-Gaussian i.e.

$$\forall \gamma \in \mathbb{R} \quad \mathbf{E}[e^{\gamma \eta_t} \mid X_{1:t}, \eta_{1:t-1}] \leq \exp\left(\frac{\gamma^2 R^2}{2}\right).$$

# Sub-Gaussianity

## Definition

Random variable  $Z$  is  $R$ -sub-Gaussian for some  $R \geq 0$  if

$$\forall \gamma \in \mathbb{R} \quad \mathbf{E}[e^{\gamma Z}] \leq \exp\left(\frac{\gamma^2 R^2}{2}\right).$$

The condition implies that

- ▶  $\mathbf{E}[Z] = 0$
- ▶  $\mathbf{Var}[Z] \leq R^2$

Examples:

- ▶ Zero-mean bounded in an interval of length  $2R$  (Hoeffding-Azuma)
- ▶ Zero-mean Gaussian with variance  $\leq R^2$

# Regret

- ▶ If we knew  $\theta_*$ , then in round  $t$  we'd choose action

$$X_t^* = \operatorname{argmax}_{x \in D_t} \langle x, \theta_* \rangle$$

- ▶ Regret is our reward in  $n$  rounds relative to  $X_t^*$ :

$$\operatorname{Regret}_n = \sum_{t=1}^n \langle X_t^*, \theta_* \rangle - \sum_{t=1}^n \langle X_t, \theta_* \rangle$$

- ▶ We want  $\operatorname{Regret}_n / n \rightarrow 0$  as  $n \rightarrow \infty$

# Optimism in the Face of Uncertainty

## Principle

- ▶ Maintain a confidence set  $C_t \subseteq \mathbb{R}^d$  such that  $\theta_* \in C_t$  with high probability.
- ▶ In round  $t$ , choose

$$(X_t, \tilde{\theta}_t) = \operatorname{argmax}_{(x, \theta) \in D_t \times C_{t-1}} \langle X_t, \theta_t \rangle$$

- ▶  $\tilde{\theta}_t$  is an “optimistic” estimate of  $\theta_*$ .
- ▶ UCB algorithm is a special case.

# Least Squares

- ▶ Data  $(X_1, Y_1), \dots, (X_n, Y_n)$  such that  $Y_t \approx \langle X_t, \theta_* \rangle$
- ▶ Stack them into matrices:  $\mathbf{X}_{1:n}$  is  $n \times d$  and  $\mathbf{Y}_{1:n}$  is  $n \times 1$
- ▶ Least squares estimate:

$$\hat{\theta}_n = (\mathbf{X}_{1:n} \mathbf{X}_{1:n}^T + \lambda I)^{-1} \mathbf{X}_{1:n}^T \mathbf{Y}_{1:n}$$

- ▶ Let  $V_n = \mathbf{X}_{1:n} \mathbf{X}_{1:n}^T + \lambda I$

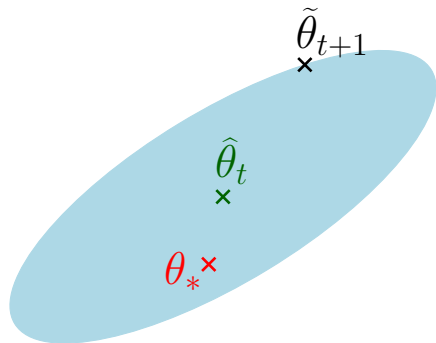
## Theorem

If  $\|\theta_*\|_2 \leq S$ , then with probability at least  $1 - \delta$ , for all  $t$ ,  $\theta_*$  lies in

$$C_t = \left\{ \theta : \|\hat{\theta}_t - \theta\|_{V_t} \leq R \sqrt{2 \ln \left( \frac{\det(V_t)^{1/2}}{\delta \det(\lambda I)^{1/2}} \right)} + S\sqrt{\lambda} \right\}$$

where  $\|v\|_A = \sqrt{v^T A v}$  is the matrix  $A$ -norm.

## Confidence Set $C_t$



- ▶ Least squares solution  $\hat{\theta}_t$  is the center of  $C_t$
- ▶  $\theta_*$  lies somewhere in  $C_t$  w.h.p.
- ▶ Next action  $\tilde{\theta}_{t+1}$  is on the boundary of  $C_t$

# Comparison with Previous Confidence Sets

- ▶ Our bound:

$$\|\hat{\theta}_t - \theta_*\|_{V_t} \leq R \sqrt{2 \ln \left( \frac{\det(V_t)^{1/2}}{\delta \det(\lambda I)^{1/2}} \right)} + S\sqrt{\lambda}$$

- ▶ [Dani et al.(2008)] If  $\|\theta_*\|_2, \|X_t\|_2 \leq 1$  then for a specific  $\lambda$

$$\|\hat{\theta}_t - \theta_*\|_{V_t} \leq R \max \left\{ \sqrt{128d \ln(t) \ln(t^2/\delta)}, \frac{8}{3} \ln(t^2/\delta) \right\}$$

- ▶ [Rusmevichientong and Tsitsiklis(2010)] If  $\|X_t\|_2 \leq 1$

$$\|\hat{\theta}_t - \theta_*\|_{V_t} \leq 2R\kappa \sqrt{\ln t} \sqrt{d \ln t + \ln(t^2/\delta)} + S\sqrt{\lambda}$$

where  $\kappa = 3 + 2 \ln((1 + \lambda d)/\lambda)$ .

Our bound doesn't depend on  $t$ .

# Regret of the Bandit Algorithm

## Theorem ([Dani et al.(2008)])

*If  $\|\theta_*\|_2 \leq 1$  and  $D_t$ 's are subsets of the unit 2-ball with probability at least  $1 - \delta$*

$$\text{Regret}_n \leq O(Rd\sqrt{n} \cdot \text{polylog}(n, d, 1/\delta))$$

We get the same result with smaller  $\text{polylog}(n, d, 1/\delta)$  factor.

# Sparse Bandits

What if  $\theta_*$  is sparse?

- ▶ Not good idea to use least squares.
- ▶ Better use e.g.  $L_1$ -regularization.
- ▶ How do we construct confidence sets?

Our new technique: *Online-to-Confidence-Set Conversion*

- ▶ Similar to Online-to-Batch Conversion, but very different
- ▶ We start with an online prediction algorithm.

# Online Prediction Algorithms

In round  $t$

- ▶ Receive  $X_t \in \mathbb{R}^d$
- ▶ Predict  $\hat{Y}_t \in \mathbb{R}$
- ▶ Receive correct label  $Y_t \in \mathbb{R}$
- ▶ Suffer loss  $(Y_t - \hat{Y}_t)^2$

No assumptions whatsoever on  $(X_1, Y_1), (X_2, Y_2), \dots$

There are heaps of algorithms of this structure:

- ▶ online gradient descent [Zinkevich(2003)]
- ▶ online least-squares [Azoury and Warmuth(2001), Vovk(2001)]
- ▶ exponentiated gradient [Kivinen and Warmuth(1997)]
- ▶ online LASSO (??)
- ▶ SeqSEW [Gerchinovitz(2011), Dalalyan and Tsybakov(2007)]

# Online Prediction Algorithms, cnt'd

- ▶ Regret with respect to a linear predictor  $\theta \in \mathbb{R}^d$

$$\rho_n(\theta) = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 - \sum_{t=1}^n (Y_t - \langle X_t, \theta \rangle)^2$$

- ▶ Prediction algorithms come with “regret bounds”  $B_n$ :

$$\forall n \quad \rho_n(\theta) \leq B_n$$

- ▶  $B_n$  depends on  $n, d, \theta$  and possibly  $X_1, X_2, \dots, X_n$  and  $Y_1, Y_2, \dots, Y_n$
- ▶ Typically,  $B_n = O(\sqrt{n})$  or  $B_n = O(\log n)$

# Online-to-Confidence-Set Conversion

- ▶ Data  $(X_1, Y_1), \dots, (X_n, Y_n)$  where  $Y_t = \langle X_t, \theta_* \rangle + \eta_t$  and  $\eta_t$  is conditionally  $R$ -sub-Gaussian.
- ▶ Predictions  $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n$
- ▶ Regret bound  $\rho(\theta_*) \leq B_n$

## Theorem (Conversion)

With probability at least  $1 - \delta$ , for all  $n$ ,  $\theta_*$  lies in

$$C_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^n (\hat{Y}_t - \langle X_t, \theta \rangle)^2 \leq 1 + 2B_n + 32R^2 \ln \left( \frac{R\sqrt{8} + \sqrt{1 + B_n}}{\delta} \right) \right\}$$

# Optimistic Algorithm with Conversion

## Theorem

*If  $|\langle x, \theta_* \rangle| \leq 1$  for all  $x \in D_t$  and all  $t$  then with probability at least  $1 - \delta$ , for all  $n$ , the regret of Optimistic Algorithm is*

$$\text{Regret}_n \leq O\left(\sqrt{dnB_n} \cdot \text{polylog}(n, d, 1/\delta, B_n)\right) .$$

# Bandits combined with SeqSEW

## Theorem ([Gerchinovitz(2011)])

If  $\|\theta\|_\infty \leq 1$  and  $\|\theta\|_0 \leq p$  then SEQSEW algorithm has regret bound

$$\rho_n(\theta) \leq B_n = O(p \log(nd)) .$$

Suppose  $\|\theta_*\|_2 \leq 1$  and  $\|\theta_*\|_0 \leq p$ . Via the conversion, the Optimistic Algorithm has regret

$$O(R\sqrt{pdn} \cdot \text{polylog}(n, d, 1/\delta))$$

which is better than  $O(Rd\sqrt{n} \cdot \text{polylog}(n, d, 1/\delta))$ .

# Open problems

- ▶ Confidence sets for batch algorithms e.g. offline LASSO.
- ▶ Adaptive bandit algorithm that doesn't need  $p$  upfront.

Questions?

Read papers at

<http://david.palenica.com/>

# References



Katy S. Azoury and Manfred K. Warmuth.  
Relative loss bounds for on-line density estimation with the exponential family of distributions.  
*Machine Learning*, 43:211–246, 2001.



Arnak S. Dalalyan and Alexandre B. Tsybakov.  
Aggregation by exponential weighting and sharp oracle inequalities.  
*In Proceedings of the 20th Annual Conference on Learning Theory*, pages 97–111, 2007.



Varsha Dani, Thomas P. Hayes, and Sham M. Kakade.  
Stochastic linear optimization under bandit feedback.  
*In Rocco Servedio and Tong Zhang, editors, Proceedings of the 21st Annual Conference on Learning Theory (COLT 2008)*, pages 355–366, 2008.



Sébastien Gerchinovitz.  
Sparsity regret bounds for individual sequences in online linear regression.

*In Proceedings of the 24th Annual Conference on Learning Theory (COLT 2011)*, 2011.



Jyrki Kivinen and Manfred K. Warmuth.  
Exponentiated gradient versus gradient descent for linear predictors.  
*Information and Computation*, 132(1):1–63, January 1997.



Paat Rusmevichientong and John N. Tsitsiklis.  
Linearly parameterized bandits.  
*Mathematics of Operations Research*, 35(2):395–411, 2010.



Vladimir Vovk.  
Competitive on-line statistics.  
*International Statistical Review*, 69:213–248, 2001.



Martin Zinkevich.  
Online convex programming and generalized infinitesimal gradient ascent.  
*In Proceedings of Twentieth International Conference on Machine Learning*, 2003.