Online Convex Optimization: From Gambling to Minimax Theorems by Playing Repeated Games

Tim van Erven

Universiteit Leiden

Nederlands Mathematisch Congres, April 4, 2018

Example: Betting on Football Games

Precursor to modern football in China,
Han Dynasty (206 BC – 220 AD)

- Before every match t in the English Premier League, my PhD student Dirk van der Hoeven wants to predict the goal difference Y_t
- Given feature vector $\boldsymbol{X}_t \in \mathbb{R}^d$, he may predict $\hat{Y}_t = \boldsymbol{w}_t^\intercal \boldsymbol{X}_t$ with a linear model
- After the match: observe Y_t
- Measure loss by $f_t(\boldsymbol{w}_t) = (Y_t - \hat{Y}_t)^2$ and improve parameter estimates: $\boldsymbol{w}_t \to \boldsymbol{w}_{t+1}$

Example: Betting on Football Games

Precursor to modern football in China,
Han Dynasty (206 BC – 220 AD)

- Before every match t in the English Premier League, my PhD student Dirk van der Hoeven wants to predict the goal difference Y_t
- Given feature vector $\boldsymbol{X}_t \in \mathbb{R}^d$, he may predict $\hat{Y}_t = \boldsymbol{w}_t^\intercal \boldsymbol{X}_t$ with a linear model
- After the match: observe Y_t
- Measure loss by $f_t(\boldsymbol{w}_t) = (Y_t - \hat{Y}_t)^2$ and improve parameter estimates: $\boldsymbol{w}_t \to \boldsymbol{w}_{t+1}$

Goal: Predict almost as well as the best possible parameters $\boldsymbol{u}$:

$$\text{Regret}_T^{\boldsymbol{u}} = \sum_{t=1}^{T} f_t(\boldsymbol{w}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{u})$$

Online Convex Optimization

Parameters $\boldsymbol{w}$ take values in a convex domain $\mathcal{W} \subset \mathbb{R}^d$

for $t = 1, 2, \ldots, T$ **do**
  Learner estimates $\boldsymbol{w}_t \in \mathcal{W}$
  Nature reveals convex loss function $f_t : \mathcal{W} \to \mathbb{R}$
end for

Viewed as a **zero-sum game** against Nature:

$$V = \min_{\boldsymbol{w}_1} \max_{f_1} \min_{\boldsymbol{w}_2} \max_{f_2} \cdots \min_{\boldsymbol{w}_T} \max_{f_T} \max_{\boldsymbol{u} \in \mathcal{W}} \mathrm{Regret}_T^{\boldsymbol{u}}$$

Online Gradient Descent

$$\begin{aligned}\tilde{\boldsymbol{w}}_{t+1} &= \boldsymbol{w}_t - \eta_t \nabla f_t(\boldsymbol{w}_t) \\ \boldsymbol{w}_{t+1} &= \min_{\boldsymbol{w}\in\mathcal{W}} \|\tilde{\boldsymbol{w}}_{t+1} - \boldsymbol{w}\|\end{aligned}$$

Theorem (Zinkevich, 2003)

Suppose $\mathcal{W}$ compact with diameter at most D, and $\|\nabla f_t(\boldsymbol{w}_t)\| \leq G$. Then online gradient descent with $\eta_t = \frac{D}{G\sqrt{t}}$ guarantees

$$\mathrm{Regret}_T^{\boldsymbol{u}} \leq \frac{3}{2} GD\sqrt{T}$$

for **any** *choices of Nature.*

Without further assumptions, this is optimal (up to a constant factor).

Von Neumann's Minimax Theorem

A Minimax Theorem:

$$\inf_{a \in \mathcal{A}} \sup_{b \in B} f(a,b) = \sup_{b \in B} \inf_{a \in \mathcal{A}} f(a,b) \tag{*}$$

Von Neumann's Minimax Theorem:

- $f(a,b) = a^{\intercal} M b$ is the pay-off of a two-player zero-sum game, for an $m \times n$ pay-off matrix M.
- $a \in \Delta_m$ and $b \in \Delta_n$ are probability vectors that represent mixed strategies.

Classical proof by Nash requires Brouwer's fixed-point theorem.

Von Neumann's Minimax Theorem

A Minimax Theorem:

$$\inf_{a \in \mathcal{A}} \sup_{b \in B} f(a,b) = \sup_{b \in B} \inf_{a \in \mathcal{A}} f(a,b) \qquad (*)$$

Von Neumann's Minimax Theorem:

- $f(a,b) = a^\intercal M b$ is the pay-off of a two-player zero-sum game, for an $m \times n$ pay-off matrix M.
- $a \in \Delta_m$ and $b \in \Delta_n$ are probability vectors that represent mixed strategies.

Classical proof by Nash requires Brouwer's fixed-point theorem.

Theorem (Variant of Freund, Schapire, 1999, Cesa-Bianchi, Lugosi, 2006)

(*) *holds if:*

- $f(a,b)$ *convex in* a, *concave in* b;
- $\mathcal{A} \subset \mathbb{R}^m$ *compact and convex;* $\mathcal{B} \subset \mathbb{R}^n$ *convex;*
- $\|\nabla_a f(a,b)\| \leq G < \infty$ *for all* a, b;
- $\sup_b f(a,b) < \infty$ *for all* a

An Elementary Proof Using OCO, Part I

i.) $\inf_{a \in \mathcal{A}} \sup_{b \in B} f(a, b) \geq \sup_{b \in B} \inf_{a \in \mathcal{A}} f(a, b)$: Moving second gives advantage.

An Elementary Proof Using OCO, Part I

i.) $\inf_{a \in \mathcal{A}} \sup_{b \in B} f(a,b) \geq \sup_{b \in B} \inf_{a \in \mathcal{A}} f(a,b)$: Moving second gives advantage.

ii.) $\inf_{a \in \mathcal{A}} \sup_{b \in B} f(a,b) \leq \sup_{b \in B} \inf_{a \in \mathcal{A}} f(a,b)$:

Lemma

There exist $a_1, \ldots, a_T$ and $b_1, \ldots, b_T$ such that:

$$\sum_{t=1}^{T} f(a_t, b_t) \leq \inf_a \sum_{t=1}^{T} f(a, b_t) + c\sqrt{T}$$

$$f(a_t, b_t) \geq \sup_b f(a_t, b) - \frac{1}{T}$$

Proof.

- Select a_t depending on $b_1, \ldots, b_{t-1}$ using online gradient descent on $f_t(a) = f(a, b_t)$.
- Let b_t be the worst response to a_t up to $\epsilon = 1/T$. □

An Elementary Proof Using OCO, Part II

$$\inf_{a \in \mathcal{A}} \sup_b f(a, b) \leq \sup_b f\left(\frac{1}{T}\sum_{t=1}^{T} a_t, b\right) \leq \sup_b \frac{1}{T}\sum_{t=1}^{T} f\left(a_t, b\right)$$

An Elementary Proof Using OCO, Part II

$$\begin{aligned}\inf_{a \in \mathcal{A}} \sup_b f(a, b) &\leq \sup_b f\left(\frac{1}{T}\sum_{t=1}^{T} a_t, b\right) \leq \sup_b \frac{1}{T}\sum_{t=1}^{T} f\left(a_t, b\right) \\ &\leq \frac{1}{T}\sum_{t=1}^{T} \sup_b f\left(a_t, b\right) \leq \frac{1}{T}\sum_{t=1}^{T} f\left(a_t, b_t\right) + \frac{1}{T}\end{aligned}$$

An Elementary Proof Using OCO, Part II

$$
\begin{aligned}
\inf_{a \in \mathcal{A}} \sup_b f(a, b) &\leq \sup_b f\left(\frac{1}{T}\sum_{t=1}^{T} a_t, b\right) \leq \sup_b \frac{1}{T}\sum_{t=1}^{T} f\left(a_t, b\right) \\
&\leq \frac{1}{T}\sum_{t=1}^{T} \sup_b f\left(a_t, b\right) \leq \frac{1}{T}\sum_{t=1}^{T} f\left(a_t, b_t\right) + \frac{1}{T} \\
&\leq \inf_a \frac{1}{T}\sum_{t=1}^{T} f\left(a, b_t\right) + \frac{c\sqrt{T}}{T} + \frac{1}{T}
\end{aligned}
$$

An Elementary Proof Using OCO, Part II

$$
\begin{aligned}
\inf_{a \in \mathcal{A}} \sup_b f(a, b) &\leq \sup_b f\left(\frac{1}{T}\sum_{t=1}^{T} a_t, b\right) \leq \sup_b \frac{1}{T}\sum_{t=1}^{T} f\left(a_t, b\right) \\
&\leq \frac{1}{T}\sum_{t=1}^{T} \sup_b f\left(a_t, b\right) \leq \frac{1}{T}\sum_{t=1}^{T} f\left(a_t, b_t\right) + \frac{1}{T} \\
&\leq \inf_a \frac{1}{T}\sum_{t=1}^{T} f\left(a, b_t\right) + \frac{c\sqrt{T}}{T} + \frac{1}{T} \\
&\leq \inf_a f\left(a, \frac{1}{T}\sum_{t=1}^{T} b_t\right) + \frac{c\sqrt{T}}{T} + \frac{1}{T}
\end{aligned}
$$

An Elementary Proof Using OCO, Part II

$$
\begin{aligned}
\inf_{a\in\mathcal{A}} \sup_b f(a,b) &\le \sup_b f\left(\frac{1}{T}\sum_{t=1}^{T} a_t, b\right) \le \sup_b \frac{1}{T}\sum_{t=1}^{T} f\left(a_t, b\right) \\
&\le \frac{1}{T}\sum_{t=1}^{T} \sup_b f\left(a_t, b\right) \le \frac{1}{T}\sum_{t=1}^{T} f\left(a_t, b_t\right) + \frac{1}{T} \\
&\le \inf_a \frac{1}{T}\sum_{t=1}^{T} f\left(a, b_t\right) + \frac{c\sqrt{T}}{T} + \frac{1}{T} \\
&\le \inf_a f\left(a, \frac{1}{T}\sum_{t=1}^{T} b_t\right) + \frac{c\sqrt{T}}{T} + \frac{1}{T} \\
&\le \sup_b \inf_a \; f\left(a, b\right) + \frac{c\sqrt{T}}{T} + \frac{1}{T}
\end{aligned}
$$

and let $T \to \infty$. $\square$

Online Portfolio Selection

Investing without a stochastic model:

- Sequential investment in d assets
- $x_{t,i} \geq 0$: ratio between closing and opening price for i-th asset in trading period t
- Reinvest fraction $w_{t,i}$ of money in asset i
- Trader's wealth grows by factor $\boldsymbol{w}_t^\intercal \boldsymbol{x}_t$
- $f_t(\boldsymbol{w}) = -\log(\boldsymbol{w}^\intercal \boldsymbol{x}_t)$

The Bitcoin (XBT) to EUR exchange rate crashing (again) after China announces trading restrictions. (Figure from www.xe.com.)

Online Portfolio Selection

Investing without a stochastic model:

- Sequential investment in d assets
- $x_{t,i} \geq 0$: ratio between closing and opening price for i-th asset in trading period t
- Reinvest fraction $w_{t,i}$ of money in asset i
- Trader's wealth grows by factor $\boldsymbol{w}_t^\intercal \boldsymbol{x}_t$
- $f_t(\boldsymbol{w}) = -\log(\boldsymbol{w}^\intercal \boldsymbol{x}_t)$

The Bitcoin (XBT) to EUR exchange rate crashing (again) after China announces trading restrictions. (Figure from www.xe.com.)

Theorem (Cover,1991)

There exists an algorithm with runtime $O(T^d)$ that guarantees

$$\text{Regret}_T^{\boldsymbol{u}} = O(d \log T)$$

for **any** *asset prices $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$. This is optimal.*

Online Portfolio Selection

Investing without a stochastic model:

- Sequential investment in d assets
- $x_{t,i} \geq 0$: ratio between closing and opening price for i-th asset in trading period t
- Reinvest fraction $w_{t,i}$ of money in asset i
- Trader's wealth grows by factor $\boldsymbol{w}_t^\intercal \boldsymbol{x}_t$
- $f_t(\boldsymbol{w}) = -\log(\boldsymbol{w}^\intercal \boldsymbol{x}_t)$

The Bitcoin (XBT) to EUR exchange rate crashing (again) after China announces trading restrictions. (Figure from www.xe.com.)

Theorem (Cover,1991)

There exists an algorithm with runtime $O(T^d)$ that guarantees

$$\text{Regret}_T^{\boldsymbol{u}} = O(d \log T)$$

*for **any** asset prices $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$. This is optimal.*

run-time:	$O(T)$	$O(T^2)$	$O(T^3)$	$O(T^4)$
max. data size:	10^{10} (Google)	10^5 (big data)	2000 (data)	300 (small data)

Open Problem (for 27 years)

Is there an algorithm for online portfolio selection with $O(T^2)$ (or preferably $O(T)$) runtime that also guarantees $O(d \log T)$ regret?

State of the Art

- $O(T)$ runtime, but $O(\sqrt{dT \log d})$ regret
- $O(T)$ runtime and $O(dG \log T)$ regret, but assumes bounded gradients $\|\nabla f_t(\boldsymbol{w}_t)\| = \frac{\|\boldsymbol{x}_t\|}{\boldsymbol{w}_t^\intercal \boldsymbol{x}_t} \leq G$ (cannot handle stocks crashing)

Open Problem (for 27 years)

Is there an algorithm for online portfolio selection with $O(T^2)$ (or preferably $O(T)$) runtime that also guarantees $O(d \log T)$ regret?

State of the Art

- $O(T)$ runtime, but $O(\sqrt{dT \log d})$ regret
- $O(T)$ runtime and $O(dG \log T)$ regret, but assumes bounded gradients $\|\nabla f_t(\boldsymbol{w}_t)\| = \frac{\|\boldsymbol{x}_t\|}{\boldsymbol{w}_t^\intercal \boldsymbol{x}_t} \leq G$ (cannot handle stocks crashing)

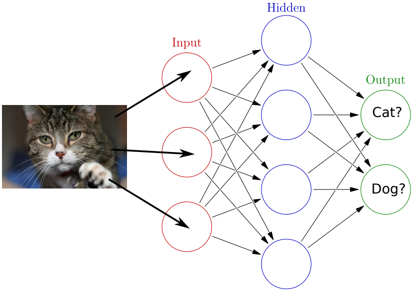
Our Progress (with Van der Hoeven, Koolen, Kotłowski)

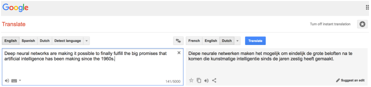
- Have simple proposed algorithm with $O(d^2 T^2)$ runtime: minimize $\phi_t(\boldsymbol{w}) = \sum_{s=1}^{t} f_s(\boldsymbol{w}) - \lambda \sum_{i=1}^{d} \log(\boldsymbol{w}^\intercal \boldsymbol{e}_i)$
- Using self-concordance techniques from interior point methods:

$$\mathrm{Regret}_T^{\boldsymbol{u}} = O\left(\sum_{t=1}^{T} g_t^2 + d \log T\right),$$

  where $g_t = \sqrt{\nabla f_t(\boldsymbol{w}_t)^\intercal \nabla^{-2} \phi_t(\boldsymbol{w}_t) \nabla f_t(\boldsymbol{w}_t)}$ measures gradient in **local norm**
- Local norms are always bounded and go to zero as we get more data
- This recovers $O(d \log T)$ in special cases, and implies $O((\log T)^d)$ in general...

Deep Neural Networks





Machine translation

Speech recognition

Self-driving cars

Class of **non-convex** functions parametrized by matrices $\boldsymbol{w} = (A_1, \ldots, A_m)$:

$$h_{\boldsymbol{w}}(\boldsymbol{x}) = A_m \sigma_{m-1} A_{m-1} \cdots \sigma_1 A_1 \boldsymbol{x},$$

where $\sigma_i(z) = \max\{0, z\}$ applied component-wise to vectors.

Deep Learning: the Big Question

Optimization

- Millions of images: too many to process all at once
- Process one image at a time using online learning algorithms:
  - Online gradient descent (OGD)
  - AdaGrad $=$ OGD with separate η_t per dimension

Deep Learning: the Big Question

Optimization

- Millions of images: too many to process all at once
- Process one image at a time using online learning algorithms:
  - Online gradient descent (OGD)
  - AdaGrad = OGD with separate η_t per dimension

High-dimensional Setting

- Still many more parameters than images (e.g. 25 times as many)
- Statistically obvious: we cannot estimate so many parameters unless we add constraints (e.g. restrict to L_p ball)

Deep Learning: the Big Question

Optimization

- Millions of images: too many to process all at once
- Process one image at a time using online learning algorithms:
  - Online gradient descent (OGD)
  - AdaGrad = OGD with separate η_t per dimension

High-dimensional Setting

- Still many more parameters than images (e.g. 25 times as many)
- Statistically obvious: we cannot estimate so many parameters unless we add constraints (e.g. restrict to L_p ball)
- But even if you disable all standard regularization, it still works! [Zhang,Bengio,Hardt,Recht,Vinyals,ICLR 2017]
- So how are the parameters restricted?

Deep Learning: the Big Question

Optimization

- Millions of images: too many to process all at once
- Process one image at a time using online learning algorithms:
  - Online gradient descent (OGD)
  - AdaGrad = OGD with separate η_t per dimension

High-dimensional Setting

- Still many more parameters than images (e.g. 25 times as many)
- Statistically obvious: we cannot estimate so many parameters unless we add constraints (e.g. restrict to L_p ball)
- But even if you disable all standard regularization, it still works! [Zhang,Bengio,Hardt,Recht,Vinyals,ICLR 2017]
- So how are the parameters restricted? **By the behavior of the optimization algorithm!**

Deep Learning: the Big Question

Optimization

- Millions of images: too many to process all at once
- Process one image at a time using online learning algorithms:
  - Online gradient descent (OGD)
  - AdaGrad = OGD with separate η_t per dimension

High-dimensional Setting

- Still many more parameters than images (e.g. 25 times as many)
- Statistically obvious: we cannot estimate so many parameters unless we add constraints (e.g. restrict to L_p ball)
- But even if you disable all standard regularization, it still works! [Zhang,Bengio,Hardt,Recht,Vinyals,ICLR 2017]
- So how are the parameters restricted? **By the behavior of the optimization algorithm!**

Big Question: Can we characterize subspace searched by optimization methods (on realistic inputs) and prove it is small enough to generalize?

Beyond Adversarial Thinking: A Modern View

Applications Are Not Zero-sum Games:

1. Worst-case regret witnessed on data where even best parameters predict poorly. So no point in achieving small regret.
2. Nature is not trying to win (e.g. football teams do not fix results to make statistical analysis hard)

Beyond Adversarial Thinking: A Modern View

Applications Are Not Zero-sum Games:

1. Worst-case regret witnessed on data where even best parameters predict poorly. So no point in achieving small regret.
2. Nature is not trying to win (e.g. football teams do not fix results to make statistical analysis hard)

Theorem (Van Erven, Koolen, 2016)

The MetaGrad algorithm guarantees the following **data-dependent** *bound:*

$$\mathrm{Regret}_T^{\boldsymbol{u}} \leq \sum_{t=1}^{T} (\boldsymbol{w}_t - \boldsymbol{u})^\intercal \nabla f_t(\boldsymbol{w}_t) \preccurlyeq \begin{cases} \sqrt{T \ln \ln T} \\ \\ \sqrt{V_T^{\boldsymbol{u}}\, d \ln T} + d \ln T \end{cases}$$

where

$$V_T^{\boldsymbol{u}} = \sum_{t=1}^{T} ((\boldsymbol{u} - \boldsymbol{w}_t)^\intercal \nabla f_t(\boldsymbol{w}_t))^2.$$

Consequences

1. Non-stochastic adaptation:

Convex f_t	$\sqrt{T \ln\ln T}$
Exp-concave f_t	$d \ln T$
Fixed convex $f_t = f$	$d \ln T$

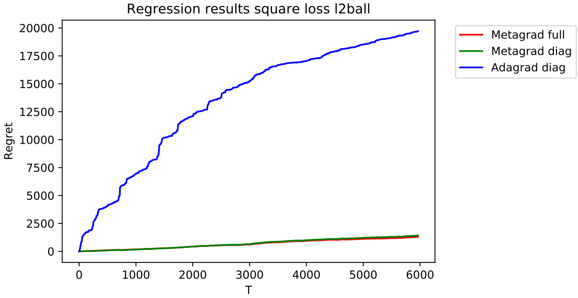
2. Stochastic without curvature

Suppose f_t i.i.d. with stochastic optimum $\boldsymbol{u}^* = \arg\min_{\boldsymbol{u} \in \mathcal{W}} \mathbb{E}_f[f(\boldsymbol{u})]$.
Then expected regret $\mathbb{E}[\mathrm{Regret}_T^{\boldsymbol{u}^*}]$:

Absolute loss* $f_t(w) = \lvert w - X_t \rvert$	$\ln T$
Hinge loss $\max\{0, 1 - Y_t \langle \boldsymbol{w}, \boldsymbol{X}_t \rangle\}$	$d \ln T$
(B, β)-**Bernstein**	$(Bd \ln T)^{1/(2-\beta)}\, T^{(1-\beta)/(2-\beta)}$

*Conditions apply

MetaGrad Football Experiments (Preliminary)



Dirk van der Hoeven
(my PhD student)

Raphaël Deswarte
(visiting PhD student)

- Predict difference in goals in 6000 football games in English Premier League (Aug 2000–May 2017).
- Square loss on Euclidean ball
- 37 features: running average of goals, shots on goal, shots over $m = 1, \ldots, 10$ previous games; multiple ELO-like models; intercept.