From Convex Analysis to Learning, Prediction, and Elicitation*
Lecture 8: Regularity Lemma

Lunjia Hu

As a powerful application of the no-regret online learning framework, in this lecture we prove the Trevisan-Tulsiani-Vadhan regularity lemma [TTV09], a.k.a., Frieze-Kannan weak regularity lemma.

Suppose we have a class F consisting of functions $f : X \to [-1, 1]$ on an arbitrary domain X. To understand the regularity lemma, we should think of these functions as "simple" functions, or "low-complexity" functions. We also have another *ground-truth* function $g^* : X \to [-1, 1]$ that may have "high-complexity".

At a high level, the regularity lemma states that we can find a *model* function $g : X \to [-1, 1]$ such that

1. **(Low complexity)** g has complexity roughly as low as the functions in F;
2. **(Indistinguishability)** g is *indistinguishable* from g^* w.r.t. F.

Theorem 1 (TTV Regularity)**.** *Let X be an arbitrary domain, and let D be a probability distribution on X. Let F be a finite class of functions $f : X \to [-1, 1]$. For every ground-truth function $g^* : X \to [-1, 1]$ and every $\varepsilon \in (0, 1/2)$, there exists a model $g : X \to [-1, 1]$ with the following properties:*

1. ***(Low complexity)*** *There exist $T = O(1/\varepsilon^2)$ functions $f_1, \ldots, f_T \in F$ and an $O(T)$-time post-processing algorithm A such that for every $x \in X$, $A(f_1(x), \ldots, f_T(x))$ correctly computes $g(x)$.*
2. ***(Indistinguishability)*** *For every $f \in F$,*

$$|\mathbb{E}_{x\sim D}[(g(x) - g^*(x))f(x)]| \leq \varepsilon.$$

1 Proof of TTV Regularity via No-Regret Online Learning

We consider an online learning problem with T rounds, where in each round, the learner chooses $g_t : X \to [-1, 1]$, and the adversary reveals $f_t \in F \cup (-F)$. The loss incurred by the learner in round t is

$$L(g_t, f_t) := \mathbb{E}_{x\sim D}[(g_t(x) - g^*(x))f_t(x)].$$

*https://lunjiahu.com/convex-analysis/

Suppose the learner uses an algorithm that guarantees average regret at most ε (regardless of the adversary's actions $f_1, \dots, f_T$):

$$\frac{1}{T}\sum_{t=1}^{T} L(g_t, f_t) \le \inf_{g' : X \to [-1,1]} \frac{1}{T}\sum_{t=1}^{T} L(g', f_t) + \varepsilon. \tag{1}$$

Note that $L(g^*, f) = 0$ for every $f \in F$, so we have

$$\inf_{g' : X \to [-1,1]} \frac{1}{T}\sum_{t=1}^{T} L(g', f_t) \le \frac{1}{T}\sum_{t=1}^{T} L(g^*, f_t) = 0.$$

Combining the two inequalities above, we get

$$\frac{1}{T}\sum_{t=1}^{T} L(g_t, f_t) \le \varepsilon.$$

Thus there exists $t^* \in \{1, \dots, T\}$ such that

$$L(g_{t^*}, f_{t^*}) \le \varepsilon. \tag{2}$$

Note that this holds regardless of the adversary's strategy of choosing $f_1, \dots, f_T$. Now we let the adversary to use the "best response" strategy:

$$f_t := \operatorname*{arg\,max}_{f \in F \cup (-F)} L(g_t, f) \quad \text{for every } t = 1, \dots, T.$$

Plugging it into (2), we get

$$\max_{f \in F \cup (-F)} L(g_{t^*}, f) \le \varepsilon.$$

This proves that g_{t^*} satisfies the indistinguishability requirement of Theorem 1.

It remains to show that there exists an algorithm for the learner that achieves the low-regret guarantee (1) while ensuring that g_{t^*} has low complexity. Using the definition of L, inequality (1) is equivalent to

$$\sum_{t=1}^{T} \mathbb{E}_{x \sim D}[g_t(x) f_t(x)] - \inf_{g' : X \to [-1,1]} \sum_{t=1}^{T} \mathbb{E}_{x \sim D}[g'(x) f_t(x)] \le \varepsilon T.$$

A sufficient condition for the condition above is that for every $x \in X$,

$$\sum_{t=1}^{T} g_t(x) f_t(x) - \inf_{g'(x) \in [-1,1]} \sum_{t=1}^{T} g'(x) f_t(x) \le \varepsilon T. \tag{3}$$

It thus suffices to achieve the regret guarantee (3) for every fixed $x \in X$. This can be done via a standard (one-dimensional) FTRL algorithm with regularizer $\varphi : [-1, 1] \to \mathbb{R}$ and learning rate $\eta > 0$:

For every $x \in X$:

- Initialize $h_1(x) = 0$;

- In each round $t = 1, \ldots, T$,
  1. play
$$g_t(x) \leftarrow \arg\min_{v \in [-1,1]} (\varphi(v) - v \cdot h_t(x)), \tag{4}$$
  2. observe $f_t(x) \in [-1, 1]$, and
  3. update $h_{t+1}(x) \leftarrow h_t(x) - \eta f_t(x)$.

We simply choose φ to be the quadratic function $\varphi(v) = v^2/2$. This allows us to compute (4) easily: for every $z \in \mathbb{R}$,

$$\arg\min_{v \in [-1,1]} (\varphi(v) - vz) = \mathsf{proj}_{[-1,1]}(z) = \begin{cases} z, & \text{if } z \in [-1, 1]; \\ -1, & \text{if } z < -1; \\ 1, & \text{if } z > 1. \end{cases}$$

Therefore, when t is small, g_t always has low complexity relative to F:

$$g_t(x) = \arg\min_{v \in [-1,1]} (\varphi(v) - v \cdot h_t(x)) = \mathsf{proj}_{[-1,1]}(h_t(x)) = \mathsf{proj}_{[-1,1]}(-\eta(f_1(x) + \cdots + f_{t-1}(x))).$$

It remains to prove that we achieve the regret bound (3) in $T = O(1/\varepsilon^2)$ rounds. It is easy to verify that φ is 1-strongly convex and has range $[0, 1/2]$ on domain $[-1, 1]$. From what we have learned in previous lectures,

$$\begin{aligned} \eta \left(\sum_{t=1}^{T} g_t(x) f_t(x) - \sum_{t=1}^{T} g'(x) f_t(x) \right) &\le \Gamma_{\varphi,\psi}(g'(x), h_1(x)) + \sum_{t=1}^{T} \Gamma_{\varphi,\psi}(g_t(x), h_{t+1}(x)) \\ &\le 1/2 + T\eta^2/2. \end{aligned}$$

Choosing $\eta = 1/\sqrt{T}$, we get the regret bound

$$\sum_{t=1}^{T} g_t(x) f_t(x) - \sum_{t=1}^{T} g'(x) f_t(x) \le \frac{1}{2\eta} + \frac{T\eta}{2} = \sqrt{T}.$$

Thus (3) holds for $T = O(1/\varepsilon^2)$, as desired.

Remark 1 (Early stop). *To find the model g_{t^*}, we don't need to always finish all $T = O(1/\varepsilon^2)$ rounds of the FTRL algorithm. We can stop after round t^* as long as (2) is satisfied.*

2 Potential Function Analysis

To be continued.

Structure VS Pseudorandomness Dichotomy.

References

[TTV09] Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Regularity, boosting, and efficiently simulating every high-entropy distribution. In *2009 24th Annual IEEE Conference on Computational Complexity*, pages 126–136, 2009. URL: `https://people.seas.harvard.edu/~salil/research/regularity-ccc09.pdf`, `doi:10.1109/CCC.2009.41`.