

---

# Lecture 6

Instructor: Haipeng Luo

---

## 1 Variation Bounds

Although Squint enjoys several nice properties simultaneously, there are still other “easy” cases that Squint does not cover. In this lecture we will talk about one of those where the “variation” of the experts’ losses is small in some senses.

Recall that the Squint’s regret bound is in terms of the quantity

$$V_T(i) = \sum_{t=1}^T (\langle p_t, \ell_t \rangle - \ell_t(i))^2,$$

where we use the loss of the algorithm  $\langle p_t, \ell_t \rangle$  as a benchmark to measure the performance of each expert. In general, for an arbitrary benchmark  $m_t \in [0, 1]^N$ , is it possible to obtain a bound that is in terms of the variation quantity

$$D_T(i) = \sum_{t=1}^T (m_t(i) - \ell_t(i))^2 \quad ?$$

The answer is obviously no if there are no restrictions on the benchmark  $m_t$  at all, because otherwise setting  $m_t = \ell_t$  would lead to zero regret. To see when this is possible, let’s first revisit the analysis of Squint (with a fixed  $\eta$ ) and see what needs to be changed accordingly to obtain such bounds. Naturally, we redefine the potential as  $\Phi_t = \mathbb{E}_{i \sim p_1} [\exp(\eta R_t(i) - \eta^2 D_t(i))]$ . We then have with  $d_t(i) = m_t(i) - \ell_t(i)$  and  $\eta < 1/2$ ,

$$\begin{aligned} & \Phi_t - \Phi_{t-1} \\ &= \mathbb{E}_{i \sim p_1} [\exp(\eta R_t(i) - \eta^2 D_t(i)) - \exp(\eta R_{t-1}(i) - \eta^2 D_{t-1}(i))] \\ &= \mathbb{E}_{i \sim p_1} [\exp(\eta R_{t-1}(i) - \eta^2 D_{t-1}(i)) (\exp(\eta r_t(i) - \eta^2 d_t^2(i)) - 1)] \\ &= \mathbb{E}_{i \sim p_1} [\exp(\eta R_{t-1}(i) - \eta^2 D_{t-1}(i)) e^{\eta r_t(i) - \eta^2 d_t^2(i)} (e^{\eta d_t(i) - \eta^2 d_t^2(i)} - e^{\eta d_t(i) - \eta r_t(i)})] \\ &= \mathbb{E}_{i \sim p_1} [\exp(\eta R_{t-1}(i) - \eta^2 D_{t-1}(i) + \eta \langle p_t, \ell_t \rangle - \eta m_t(i)) (e^{\eta d_t(i) - \eta^2 d_t^2(i)} - e^{\eta d_t(i) - \eta r_t(i)})] \\ &= \mathbb{E}_{i \sim p_1} [\exp(\eta R_{t-1}(i) - \eta^2 D_{t-1}(i) + \eta \langle p_t, \ell_t \rangle - \eta m_t(i)) (1 + \eta d_t(i) - e^{\eta d_t(i) - \eta r_t(i)})] \\ &\qquad (e^{x-x^2} \leq 1+x, \forall x \geq -\frac{1}{2}) \\ &= \mathbb{E}_{i \sim p_1} [\exp(\eta R_{t-1}(i) - \eta^2 D_{t-1}(i) + \eta \langle p_t, \ell_t \rangle - \eta m_t(i)) \eta r_t(i)]. \\ &\qquad (e^x \geq 1+x, \forall x) \end{aligned}$$

As discussed before, if we want the last term to be zero, the algorithm should play

$$p_t(i) \propto p_1(i) \exp(\eta R_{t-1}(i) - \eta^2 D_{t-1}(i) + \eta \langle p_t, \ell_t \rangle - \eta m_t(i)).$$

Noting that  $\eta \langle p_t, \ell_t \rangle$  is a constant for all experts, the above is equivalent to

$$p_t(i) \propto p_1(i) \exp(\eta R_{t-1}(i) - \eta^2 D_{t-1}(i) - \eta m_t(i)). \tag{1}$$

Now by the exact same argument as Squint, with an optimal tuned  $\eta$ , the fact  $\Phi_T \leq 1$  implies that for any  $q \in \Delta(N)$ ,

$$\mathbb{E}_{i \sim q}[R_T(i)] \leq 2\sqrt{\mathbb{E}_{i \sim q}[D_T(i)]\text{KL}(q, p_1)}.$$

Therefore, ignoring the learning rate tuning issue for a moment, we indeed obtain a bound that replaces  $V_T$  by  $D_T$  for any benchmark  $m_t$ , *as long as*  $m_t$  is available at the beginning of round  $t$  (that is, before seeing  $\ell_t$ ) so that Eq. (1) is a valid algorithm.

This excludes the choice of  $m_t = \ell_t$  (which makes sense), but there are still several interesting valid choices. For example, setting  $m_t(i) = \mu_{t-1}(i)$  where  $\mu_t(i)$  is the empirical average loss of expert  $i$  at round  $t$  (that is  $\mu_t(i) = \frac{1}{t} \sum_{\tau=1}^t \ell_\tau(i)$  and  $\mu_0(i) = 0$ ), we have

$$\begin{aligned} D_T(i) &= \sum_{t=1}^T (\ell_t(i) - \mu_{t-1}(i))^2 \\ &= \sum_{t=1}^T (\ell_t(i) - \mu_t(i))^2 + (2\ell_t(i) - \mu_t(i) - \mu_{t-1}(i))(\mu_t(i) - \mu_{t-1}(i)) \\ &= \sum_{t=1}^T (\ell_t(i) - \mu_t(i))^2 + (2\ell_t(i) - \mu_t(i) - \mu_{t-1}(i)) \frac{\ell_t(i) - \mu_{t-1}(i)}{t} \\ &\leq \sum_{t=1}^T (\ell_t(i) - \mu_t(i))^2 + \frac{2}{t} \\ &\leq \sum_{t=1}^T (\ell_t(i) - \mu_T(i))^2 + \mathcal{O}(\ln T), \end{aligned}$$

where the last step is actually due to the BTL lemma (hint:  $\mu_t(i) = \operatorname{argmin}_\mu \sum_{\tau=1}^t (\ell_\tau(i) - \mu)^2$ ). Therefore, in this case  $D_T(i)$  is essentially the empirical variance of expert  $i$ 's losses (up to  $\mathcal{O}(\ln T)$ ), and the bound indicates that small variance implies small regret.

Another example is to let  $m_t(i) = \ell_{t-1}(i)$  for  $t \neq 1$  and  $m_1(i) = 0$ , so that  $D_T(i) = \sum_{t=1}^T (\ell_t(i) - \ell_{t-1}(i))^2$  measures the variation of expert  $i$ 's losses over time, which is sometimes called “path-length” bound. In fact, path-length is also bounded by the variance up to constants:

$$\begin{aligned} \sum_{t=1}^T (\ell_t(i) - \ell_{t-1}(i))^2 &\leq 3 \sum_{t=1}^T ((\ell_t(i) - \mu_t(i))^2 + (\ell_{t-1}(i) - \mu_{t-1}(i))^2 + (\mu_t(i) - \mu_{t-1}(i))^2) \\ &\leq 6 \sum_{t=1}^T (\ell_t(i) - \mu_T(i))^2 + 3 \sum_{t=1}^T \frac{1}{t^2} \\ &\leq 6 \sum_{t=1}^T (\ell_t(i) - \mu_T(i))^2 + 6. \end{aligned}$$

However, one can construct an example where the path-length is a constant while the variance (or even  $\sum_{t=1}^T (\ell_t(i) - \mu_{t-1}(i))^2$ ) is of order  $\Omega(T)$ . (Simply think about the case where  $\ell_t(i) = 0$  for all  $t \leq T/2$  and  $\ell_t(i) = 1$  otherwise.)

Coming back to the learning rate tuning issue, while one would naturally imagine that using the similar idea of putting a prior on  $\eta$  as Squint could possibly solve the problem, it actually does not work here in general. The reason is that according to discussions from last lecture the strategy should become

$$p_t(i) \propto p_1(i) \mathbb{E}_\eta [\eta \exp (\eta R_{t-1}(i) - \eta^2 D_{t-1}(i) + \eta \langle p_t, \ell_t \rangle - \eta m_t(i))],$$

and critically the term  $\eta \langle p_t, \ell_t \rangle$  cannot be factored out and removed anymore. This makes the algorithm invalid because it depends on  $\ell_t$  which is of course not available before playing  $p_t$ . The only exception is to have  $m_t(i) = \langle p_t, \ell_t \rangle$  so that the last two terms in the exponent cancel with each other, which simply just recovers Squint. It is in fact still not clear whether there is a parameter-free

algorithm that achieves the bound  $\sqrt{\mathbb{E}_{i \sim q}[D_T(i)]\text{KL}(q, p_1)}$  simultaneously for all  $q$ , even for the two special choices of  $m_t$  discussed previously.

Nevertheless, we remark that obtaining a weaker bound such as

$$R_T(i^*) \leq \mathcal{O} \left( \sqrt{\left( \max_{i \in [N]} D_T(i) \right) \ln N} \right)$$

with a parameter-free algorithm is possible via simple techniques such as the doubling trick (details omitted). While the bound is now in terms of the variation of all experts instead of the best expert, it could still be much smaller than the worse case  $\mathcal{O}(\sqrt{T \ln N})$ .

## 2 Optimistic FTRL

In this section we explore a different technique that allows one to obtain similar bounds but with a very important difference. The new algorithm is simply to remove the term  $\eta^2 D_{t-1}(i)$  in Eq. (1). However, it is easier to interpret and analyze the algorithm if we write it in an FTRL format:

$$p_t = \operatorname{argmin}_{p \in \Delta(N)} \langle p, L_{t-1} + m_t \rangle + \frac{1}{\eta} \psi(p). \quad (2)$$

where  $\psi(p)$  is the negative entropy. One can verify that this is exactly equivalent to  $p_t(i) \propto \exp(-\eta(L_{t-1}(i) + m_t(i)))$ . Compared to the standard FTRL, the only difference here is the term  $m_t$ . One way to interpret the algorithm is that we are optimistic that the loss for round  $t$  will be close to  $m_t$ , and therefore we incorporate  $m_t$  into the cumulative loss and treat  $L_{t-1} + m_t$  as a proxy for  $L_t$ . This is called optimistic FTRL [Rakhlin and Sridharan, 2013, Syrgkanis et al., 2015].

We prove the following theorem for this algorithm. Note that while we restrict to the expert setting with entropy regularizer, similar results hold for the general OCO setting with any decision space and any strongly convex regularizer.

**Theorem 1.** *Optimistic FTRL (2) ensures for any  $q \in \Delta(N)$ ,*

$$\sum_{t=1}^T \langle p_t - q, \ell_t \rangle \leq \frac{2 + \ln N}{\eta} + \eta \sum_{t=1}^T \|\ell_t - m_t\|_\infty^2 - \frac{1}{4\eta} \sum_{t=1}^T \|p_{t+1} - p_t\|_1^2.$$

Note that the term  $\sum_{t=1}^T \|\ell_t - m_t\|_\infty^2 = \sum_{t=1}^T \max_i (\ell_t(i) - m_t(i))^2$  is worse than  $\mathbb{E}_{i \sim q}[D_T(i)]$  or even  $\max_i D_T(i)$  discussed in the last section. However, the negative term is crucial and turns out to be very useful. At first glance the negative term is a bit counter-intuitive since it is suggesting that less stability leads to smaller regret. We defer the discussion to the next lecture where we will see why this makes sense in the context of game theory.

*Proof.* Let  $p'_t = \operatorname{argmin}_p \langle p, L_{t-1} \rangle + \frac{1}{\eta} \psi(p)$  be the regular FTRL strategy. The regret can be decomposed as

$$\sum_{t=1}^T \langle p_t - q, \ell_t \rangle = \sum_{t=1}^T \langle p_t - p'_{t+1}, \ell_t - m_t \rangle + \sum_{t=1}^T (\langle p_t - p'_{t+1}, m_t \rangle + \langle p'_{t+1} - q, \ell_t \rangle).$$

The first summation can be bounded in a similar way as in the stability lemma of FTRL, which we summarize again in Lemma 1. Using the lemma with  $L = L_{t-1} + m_t$  and  $L' = L_t$ , we have

$$\|p_t - p'_{t+1}\| \leq \eta \|\ell_t - m_t\|_\infty,$$

and therefore  $\langle p_t - p'_{t+1}, \ell_t - m_t \rangle \leq \|p_t - p'_{t+1}\| \|\ell_t - m_t\|_\infty \leq \eta \|\ell_t - m_t\|_\infty^2$ .

It remains to bound the second summation. Intuitively, since the optimization in calculating  $p_t$  takes  $m_t$  into account while the one for  $p'_{t+1}$  does not, the term  $\langle p_t - p'_{t+1}, m_t \rangle$  should be small. On the other hand, the other term  $\langle p'_{t+1} - q, \ell_t \rangle$  is basically the regret for BTL, which also should be small. In fact, we will prove the following stronger statement

$$\sum_{t=1}^T (\langle p_t - p'_{t+1}, m_t \rangle + \langle p'_{t+1} - q, \ell_t \rangle) \leq \frac{\ln N + \psi(q) - A_T}{\eta}, \quad (3)$$

where  $A_T = \frac{1}{2} \sum_{t=1}^T (\|p_t - p'_{t+1}\|^2 + \|p_t - p'_t\|^2)$ . This will finish the proof since  $\psi(q) \leq 0$  and

$$\begin{aligned} A_T &= \frac{1}{2} \sum_{t=1}^T (\|p_t - p'_{t+1}\|^2 + \|p_t - p'_t\|^2) \\ &\geq \frac{1}{2} \sum_{t=1}^T (\|p_t - p'_{t+1}\|^2 + \|p_{t+1} - p'_{t+1}\|^2) - \frac{1}{2} \|p_{T+1} - p'_{T+1}\|^2 \\ &\geq \frac{1}{4} \sum_{t=1}^T ((\|p_t - p'_{t+1}\| + \|p_{t+1} - p'_{t+1}\|)^2) - 2 \quad ((a+b)^2 \leq 2a^2 + 2b^2) \\ &\geq \frac{1}{4} \sum_{t=1}^T \|p_t - p_{t+1}\|^2 - 2 \quad (\text{by triangle inequality}) \end{aligned}$$

We use induction to prove Eq. (3). The base case when  $T = 0$  holds trivially since  $\psi(q) \geq -\ln N$ . Now assume we have for any  $q$ ,

$$\sum_{t=1}^{T-1} (\langle p_t - p'_{t+1}, m_t \rangle + \langle p'_{t+1} - q, \ell_t \rangle) \leq \frac{\ln N + \psi(q) - A_{T-1}}{\eta}. \quad (4)$$

We then have

$$\begin{aligned} &\sum_{t=1}^T (\langle p_t - p'_{t+1}, m_t \rangle + \langle p'_{t+1}, \ell_t \rangle) \\ &\leq \langle p_T - p'_{T+1}, m_T \rangle + \langle p'_{T+1}, \ell_T \rangle + \frac{\ln N + \psi(p'_T) - A_{T-1}}{\eta} + \langle p'_T, L_{T-1} \rangle \quad (\text{by setting } q = p'_T \text{ in Eq. (4)}) \\ &\leq \langle p_T - p'_{T+1}, m_T \rangle + \langle p'_{T+1}, \ell_T \rangle + \frac{\ln N + \psi(p_T) - A_{T-1} - \frac{1}{2} \|p_T - p'_T\|^2}{\eta} + \langle p_T, L_{T-1} \rangle \quad (\text{by Eq. (5)}) \\ &= \langle p'_{T+1}, \ell_T - m_T \rangle + \frac{\ln N + \psi(p_T) - A_{T-1} - \frac{1}{2} \|p_T - p'_T\|^2}{\eta} + \langle p_T, L_{T-1} + m_T \rangle \\ &\leq \langle p'_{T+1}, \ell_T - m_T \rangle + \frac{\ln N + \psi(p'_{T+1}) - A_T}{\eta} + \langle p'_{T+1}, L_{T-1} + m_T \rangle \quad (\text{by Eq. (5)}) \\ &= \frac{\ln N + \psi(p'_{T+1}) - A_T}{\eta} + \langle p'_{T+1}, L_T \rangle \\ &\leq \frac{\ln N + \psi(q) - A_T}{\eta} + \langle q, L_T \rangle. \quad (\text{by optimality of } p'_{T+1}) \end{aligned}$$

Rearranging finishes the induction and thus proves the theorem.  $\square$

**Lemma 1** (Stability). *If  $p_\star = \operatorname{argmin}_p \langle p, L \rangle + \frac{1}{\eta} \psi(p)$  and  $p'_\star = \operatorname{argmin}_p \langle p, L' \rangle + \frac{1}{\eta} \psi(p)$  for a 1-strongly convex regularizer  $\psi$  (with respect to a norm  $\|\cdot\|$ ) and some  $L$  and  $L'$ . Then*

$$\|p_\star - p'_\star\| \leq \eta \|L - L'\|_\star.$$

*Proof.* Let  $F(p; L) = \langle p, L \rangle + \frac{1}{\eta} \psi(p)$ . By strong convexity and first order optimality we have for any  $q$ ,

$$\begin{aligned} F(p_\star; L) &\leq F(q; L) + \langle \nabla F(p_\star, L); p_\star - q \rangle - \frac{1}{2\eta} \|p_\star - q\|^2 \\ &\leq F(q; L) - \frac{1}{2\eta} \|p_\star - q\|^2. \end{aligned} \quad (5)$$

Similar statement holds for  $p'_*$ . Therefore we have

$$\langle p_* - p'_*, L' - L \rangle = F(p'_*; L) - F(p_*; L) + F(p_*; L') - F(p'_*; L') \geq \frac{1}{\eta} \|p_* - p'_*\|^2,$$

and on the other hand by Hölder's inequality

$$\langle p_* - p'_*, L' - L \rangle \leq \|p_* - p'_*\| \|L - L'\|_\star.$$

Combining the two inequalities proves the lemma.  $\square$

## References

- Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems 26*, 2013.
- Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems 28*, 2015.