Machine Learning Theory (CS 6783)

Lecture 15: MD with Local norm bound, Learning With Bandit Feedback

1 Mirror Descent with Local Norms (full information case)

We have shown that if we are able to find a function R that is strongly convex w.r.t. some norm $\|\cdot\|$ then mirror descent algorithm with step size η using this function R has the following bound on regret:

$$\mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T) \leq \frac{\eta}{2} \sum_{t=1}^{T} \|\tilde{\nabla}_t\|_*^2 + \frac{1}{\eta} \sup_{f \in \mathcal{F}} \Delta_R(f|\hat{y}_1)$$

where $\|\cdot\|_*$ is the dual to the norm $\|\cdot\|$. We will now modify this result to replace the dual norm with a local norm. This will turn out to be useful to obtain bandit algorithms.

Recall the mirror descent algorithm:

$$\nabla R(\hat{\mathbf{y}}'_{t+1}) = \nabla R(\hat{\mathbf{y}}_t) - \eta \nabla_t \qquad \& \qquad \hat{\mathbf{y}}_{t+1} = \operatorname*{argmin}_{f \in \mathcal{F}} \Delta_R(f|\hat{\mathbf{y}}'_{t+1})$$

Assume that the function R is twice differentiable and let let $\nabla^2 R(f)$ denote the Hessian of the function at a point R. Now we prove the following claim.

Lemma 1. *For any twice differentiable convex R, if we run mirror descent using step size η, then*

$$n\mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T) \leq \frac{\eta}{2} \sum_{t=1}^{T} \|\tilde{\nabla}_t\|_{\nabla^2 R(z_t)^{-1}}^2 + \frac{1}{\eta} \sup_{f \in \mathcal{F}} \Delta_R(f|\hat{y}_1)$$

where z_t is some convex combination of $\hat{\mathbf{y}}_t$ and $\hat{\mathbf{y}}'_{t+1}$ (here matrix M, $\|x\|_M^2 = x^\top M x$)

Proof. We will recall the upper bound from the mirror descent proof of the form:

$$\left\langle \tilde{\nabla}_t, \hat{\mathbf{y}}_t - f^* \right\rangle \leq \left\langle \tilde{\nabla}_t, \hat{\mathbf{y}}_t - \hat{\mathbf{y}}'_{t+1} \right\rangle + \frac{1}{\eta} \left(\Delta_R(f^*|\hat{\mathbf{y}}_t) - \Delta_R(f^*|\hat{\mathbf{y}}_{t+1}) - \Delta_R(\hat{\mathbf{y}}'_{t+1}|\hat{\mathbf{y}}_t) \right)$$

Now the key trick is that we start with the definition of Bregman divergence and use Taylor's theorem. Note that:

$$\Delta_R(\hat{\mathbf{y}}'_{t+1}|\hat{\mathbf{y}}_t) = R(\hat{\mathbf{y}}'_{t+1}) - R(\hat{\mathbf{y}}_t) - \left\langle R(\hat{\mathbf{y}}_t), \hat{\mathbf{y}}'_{t+1} - \hat{\mathbf{y}}_t \right\rangle$$

Now using Taylor's theorem (+ intermediate value theorem) there exists a point z_t that is some convex combination of $\hat{\mathbf{y}}'_{t+1}$ and $\hat{\mathbf{y}}_t$ such that

$$R(\hat{\mathbf{y}}'_{t+1}) - R(\hat{\mathbf{y}}_t) - \left\langle R(\hat{\mathbf{y}}_t), \hat{\mathbf{y}}'_{t+1} - \hat{\mathbf{y}}_t \right\rangle = \frac{1}{2} (\hat{\mathbf{y}}'_{t+1} - \hat{\mathbf{y}}_t)^\top \nabla^2 R(z_t) (\hat{\mathbf{y}}'_{t+1} - \hat{\mathbf{y}}_t) = \frac{1}{2} \|\hat{\mathbf{y}}'_{t+1} - \hat{\mathbf{y}}_t\|_{\nabla^2 R(z_t)}^2$$

Hence using this we can conclude that

$$\left\langle \tilde{\nabla}_t, \hat{\mathbf{y}}_t - f^* \right\rangle \leq \left\langle \tilde{\nabla}_t, \hat{\mathbf{y}}_t - \hat{\mathbf{y}}'_{t+1} \right\rangle + \frac{1}{\eta} \left(\Delta_R(f^*|\hat{\mathbf{y}}_t) - \Delta_R(f^*|\hat{\mathbf{y}}_{t+1}) \right) - \frac{1}{2\eta} \|\hat{\mathbf{y}}'_{t+1} - \hat{\mathbf{y}}_t\|^2_{\nabla^2 R(z_t)}$$

Now note that for any invertible matrix M, $\|\cdot\|_{M^{-1}}$ is the dual norm to the norm $\|\cdot\|_M$ and hence using the fact (as we did in the earlier mirror descent proof) that

$$\left\langle \tilde{\nabla}_t, \hat{\mathbf{y}}_t - \hat{\mathbf{y}}'_{t+1} \right\rangle \leq \frac{\eta}{2} \|\tilde{\nabla}_t\|^2_{\nabla^2 R(z_t)^{-1}} + \frac{1}{2\eta} \|\hat{\mathbf{y}}'_{t+1} - \hat{\mathbf{y}}_t\|^2_{\nabla^2 R(z_t)}$$

we conclude that

$$\left\langle \tilde{\nabla}_t, \hat{\mathbf{y}}_t - f^* \right\rangle \leq \frac{\eta}{2} \|\tilde{\nabla}_t\|^2_{\nabla^2 R(z_t)^{-1}} + \frac{1}{\eta} \left(\Delta_R(f^*|\hat{\mathbf{y}}_t) - \Delta_R(f^*|\hat{\mathbf{y}}_{t+1}) \right)$$

Summing over t and simplifying the telescoping sum over the Bregman divergences we we obtain that

$$\begin{aligned} n\mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T) &\leq \frac{\eta}{2} \sum_{t=1}^{T} \|\tilde{\nabla}_t\|^2_{\nabla^2 R(z_t)^{-1}} + \frac{1}{\eta} \left(\Delta_R(f^*|\hat{\mathbf{y}}_1) - \Delta_R(f^*|\hat{\mathbf{y}}_{n+1}) \right) \\ &\leq \frac{\eta}{2} \sum_{t=1}^{T} \|\tilde{\nabla}_t\|^2_{\nabla^2 R(z_t)^{-1}} + \frac{1}{\eta} \sup_{f \in \mathcal{F}} \Delta_R(f|\hat{\mathbf{y}}_1) \end{aligned}$$

□

2 Online Linear Optimization With Bandit Feedback

The bandit linear online learning problem is as follows:

For $t = 1$ to T

1. Learner picks $\hat{\mathbf{y}}_t \in \mathcal{F}$
2. Adversary picks ∇_t simultaneously
3. Learner observes and suffers loss $\hat{\mathbf{y}}_t^\top \nabla_t$

End For

Goal: Minimize expected regret

$$\mathrm{Reg}_T = \sum_{t=1}^{T} \hat{\mathbf{y}}_t^\top \nabla_t - \inf_{f \in \mathcal{F}} \sum_{t=1}^{T} f^\top \nabla_t$$

Main Idea:

1. Obtain $\hat{\mathbf{y}}_t$ from a full information algorithm
2. Randomize move such that $\mathbb{E}[\hat{y}_t] = \hat{\mathbf{y}}_t$
3. Play $\hat{y}_t$ and receive feedback $\hat{y}_t^\top \nabla_t$

4. Build unbiased estimate of ∇_t based on feedback as $\mathbb{E}[\tilde{\nabla}_t] = \nabla_t$
5. Feed $\tilde{\nabla}_t$ to a full information online linear algorithms

For bandit algorithms, adaptive vs oblivious adversaries make a difference. Adaptive adversaries are ones that know the internal randomization of the learner up to given point while oblivious adversaries only know the learning algorithms but not the random bits produced by the adversary. That is, we can think of adversary (knowing the learning algorithm) first prefixing $\nabla_1, \ldots, \nabla_n$ and producing them one by one. In this case, note that:

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Reg}_T\right] &= \mathbb{E}\left[\sum_{t=1}^{T} \hat{y}_t^\top \nabla_t\right] - \inf_{f\in\mathcal{F}} \sum_{t=1}^{T} f^\top \nabla_t \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \hat{y}_t^\top \nabla_t\right] - \inf_{f\in\mathcal{F}} \sum_{t=1}^{T} f^\top \mathbb{E}\left[\tilde{\nabla}_t\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \hat{y}_t^\top \nabla_t\right] - \mathbb{E}\left[\inf_{f\in\mathcal{F}} \sum_{t=1}^{T} f^\top \tilde{\nabla}_t\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}[\hat{y}_t]^\top \nabla_t\right] - \mathbb{E}\left[\inf_{f\in\mathcal{F}} \sum_{t=1}^{T} f^\top \tilde{\nabla}_t\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \hat{\mathbf{y}}_t^\top \nabla_t\right] - \mathbb{E}\left[\inf_{f\in\mathcal{F}} \sum_{t=1}^{T} f^\top \tilde{\nabla}_t\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \hat{\mathbf{y}}_t^\top \mathbb{E}[\tilde{\nabla}_t]\right] - \mathbb{E}\left[\inf_{f\in\mathcal{F}} \sum_{t=1}^{T} f^\top \tilde{\nabla}_t\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \hat{\mathbf{y}}_t^\top \tilde{\nabla}_t] - \inf_{f\in\mathcal{F}} \sum_{t=1}^{T} f^\top \tilde{\nabla}_t\right] \\
&= \mathbb{E}\left[\mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T)\right]
\end{aligned}$$

Hence overall we conclude that for this procedure:

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \mathbb{E}\left[\mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T)\right] \tag{1}$$

The above basically tells us that we have a reduction from bandit algorithm to full information algorithm. Hence, using this unbiased gradient estimate trick, we can apply a full information algorithm on the estimates and the expected regret of this full information algorithm will be the bound for our bandit algorithm. A word of caution though: typically our full information algorithms depend on norms of gradient being bounded under some appropriate norm. Eg. exponential weights algorithm on ℓ_∞ norm and gradient descent on ℓ_2 norm. However, while ∇_t's might have this norm bounded, our estimates can have very large norms and this can cause our bounds to blow up. To this end, we have two options: either we modify our full information algorithm (albeit at the cost of worse bounds) so that the estimates have smaller norms or alternatively we are more careful to get an adaptive bound for our full information algorithm to get tighter bounds in expectation. To deal with this, we will use the MD with the local norm analysis and show that the expected local norm is small.

3 Example: Multi-armed Bandit

For this example we will use the exponential weights algorithm as the full information algorithm. Recall that

$$R(f) = \sum_{i=1}^{N} f[i]\log(f[i]) + \log(N)$$

and note that

$$\widehat{\mathbf{y}'}_{t+1}[j] = \hat{\mathbf{y}}_t[j] \times \exp(-\eta \nabla_t[j])$$

For a given $\hat{\mathbf{y}}_t$ we will simply draw $i_t \sim \hat{\mathbf{y}}_t$ and play $\hat{y}_t = e_{i_t}$ (that is expert i_t) and note that clearly $\mathbb{E}[\hat{y}_t] = \hat{\mathbf{y}}_t$. Further, for the unbiased estimate of loss, we use:

$$\tilde{\nabla}_t = \frac{e_{i_t}^\top \nabla_t}{q_t(i_t)} e_{i_t}$$

So that: $\mathbb{E}[\tilde{\nabla}_t] = \sum_{i=1}^{d} \hat{\mathbf{y}}_t(i) \frac{\nabla_t[i]}{\hat{\mathbf{y}}_t(i)} e_i = \nabla_t$

Now note that:

$$\nabla^2 R(f) = \begin{bmatrix} \frac{1}{f[1]} & 0 & 0 & 0 \\ 0 & \frac{1}{f[2]} & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \frac{1}{f[N]} \end{bmatrix}$$

Hence note that

$$\nabla^2 R(f)^{-1} = \begin{bmatrix} f[1] & 0 & 0 & 0 \\ 0 & f[2] & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & f[N] \end{bmatrix}$$

Hence using this in the local norm lemma we have the bound:

$$\begin{aligned} \mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T) &\le \frac{\eta}{2}\sum_{t=1}^{T} \|\tilde{\nabla}_t\|^2_{\nabla^2 R(z_t)^{-1}} + \frac{1}{\eta}\sup_{f\in\mathcal{F}} \Delta_R(f|\hat{y}_1) \\ &\le \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{N} \tilde{\nabla}_t^2[i] z_t[i] + \frac{1}{\eta}\log(N) \end{aligned}$$

Now recall that z_t is of the form $z_t = \alpha_t \hat{\mathbf{y}}_t + (1-\alpha_t)\hat{\mathbf{y}}'_{t+1}$ for some $\alpha_t \in [0,1]$ and so

$$z_t[j] = \alpha_t \hat{\mathbf{y}}_t[j] + (1-\alpha)\hat{\mathbf{y}}_t[j] \times \exp(-\eta\nabla_t[j]) \le \hat{\mathbf{y}}_t[j](1+\exp(\eta))$$

Hence we conclude that

$$\mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T) \le \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{N} \tilde{\nabla}_t^2[i]\hat{\mathbf{y}}_t[i] + \frac{1}{\eta}\log(N)$$

Now plugging in the form of $\tilde{\nabla}_t = \frac{e_{i_t}^\top \nabla_t}{\hat{\mathbf{y}}_t(i_t)} e_{i_t}$ we get,

$$\mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T) \le \frac{\eta}{2}\sum_{t=1}^{T} \frac{\nabla_t^2[i_t]}{\hat{\mathbf{y}}_t[i_t]} + \frac{1}{\eta}\log(N)$$

Now since

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \mathbb{E}\left[\mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T)\right]$$

Taking expectation over the draw of actions we get:

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Reg}_T\right] &\leq \mathbb{E}\left[\frac{\eta}{2}\sum_{t=1}^{T}\frac{\nabla_t^2[i_t]}{\hat{\mathbf{y}}_t[i_t]}\right] + \frac{1}{\eta}\log(N)\\
&\leq \frac{\eta}{2}\sum_{t=1}^{T}\mathbb{E}\left[\frac{\nabla_t^2[i_t]}{\hat{\mathbf{y}}_t[i_t]}\right] + \frac{1}{\eta}\log(N)\\
&= \frac{\eta}{2}\sum_{t=1}^{T}\mathbb{E}\left[\sum_{j=1}^{N}\mathbf{y}_t[j]\frac{\nabla_t^2[j]}{\hat{\mathbf{y}}_t[j]}\right] + \frac{1}{\eta}\log(N)\\
&= \frac{\eta}{2}\sum_{t=1}^{T}\mathbb{E}\left[\sum_{j=1}^{N}\nabla_t^2[j]\right] + \frac{1}{\eta}\log(N)\\
&\leq \frac{\eta}{2}NT + \frac{1}{\eta}\log(N)
\end{aligned}$$

Setting $\eta = \sqrt{2\log(N)/NT}$ we conclude that for the bandit algorithm, the expected regret is bounded as:

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \sqrt{2N\log(N)T}$$

4 Example: Multi-armed Bandit Different Algorithm

Let us this time we shall use MD with function R as

$$R(f) = -\sum_{i=1}^{N}\log(f[i]) = -\sum_{i=1}^{N-1}\log(f[i]) - \log\left(1 - \sum_{i=1}^{N-1}f[i]\right)$$

and using $\hat{y}_1 = 1/N\mathbf{1}$ Note that $\nabla R(f) = -[1/f[1], \ldots, 1/f[N]]^\top$ For a given $\hat{\mathbf{y}}_t$ we will simply draw $i_t \sim \hat{\mathbf{y}}_t$ and play $\hat{y}_t = e_{i_t}$ (that is expert i_t) and note that clearly $\mathbb{E}[\hat{y}_t] = \hat{\mathbf{y}}_t$. Further, for the unbiased estimate of loss, we use:

$$\tilde{\nabla}_t = \frac{e_{i_t}^\top \nabla_t}{q_t(i_t)} e_{i_t}$$

So that: $\mathbb{E}[\tilde{\nabla}_t] = \sum_{i=1}^{d}\hat{\mathbf{y}}_t(i)\frac{\nabla_t[i]}{\hat{\mathbf{y}}_t(i)}e_i = \nabla_t$ just like in previous section.

Now note that:

$$\nabla^2 R(f) = \begin{bmatrix} \frac{1}{f^2[1]} & 0 & 0 & 0 \\ 0 & \frac{1}{f^2[2]} & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \frac{1}{f^2[N]} \end{bmatrix}$$

Hence note that

$$\nabla^2 R(f)^{-1} = \begin{bmatrix} f^2[1] & 0 & 0 & 0 \\ 0 & f^2[2] & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & f^2[N] \end{bmatrix}$$

Hence using this in the local norm lemma we have the bound:

$$\sum_{t=1}^{T} \tilde{\nabla}_t^\top \hat{\mathbf{y}}_t - \sum_{t=1}^{T} \tilde{\nabla}_t^\top f \le \frac{\eta}{2} \sum_{t=1}^{T} \|\tilde{\nabla}_t\|^2_{\nabla^2 R(z_t)^{-1}} + \frac{1}{\eta} \Delta_R(f|\hat{y}_1)$$

But note that:

$$\begin{aligned}\Delta_R(f|\hat{y}_1) &= R(f) - R(\hat{y}_1) - \nabla R(\hat{y}_1)(f - \hat{y}_1)\\ &= -\sum_{i=1}^{N} \log(f[i]) - N\log(N) \le N \max_{i\in[N]} \log(1/Nf[i]) + N\end{aligned}$$

Now the R we picked is what is called a self-concordant barrier function for the simplex, Fur such functions one can show the property that: there exists a constant c such that

$$\|\cdot\|^2_{\nabla^2 R(z_t)^{-1}} \le c\|\cdot\|^2_{\nabla^2 R(\hat{y}_t)^{-1}}$$

and so we get that:

$$\sum_{t=1}^{T} \tilde{\nabla}_t^\top \hat{\mathbf{y}}_t - \sum_{t=1}^{T} \tilde{\nabla}_t^\top f \le \frac{\eta}{2} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{\nabla}_t^2[i] \hat{\mathbf{y}}_t^2[i] + \frac{N}{\eta} \max_{i\in[N]} \left(\log(1/Nf[i]) + 1\right)$$

Now plugging in the form of $\tilde{\nabla}_t = \frac{e_{i_t}^\top \nabla_t}{\hat{\mathbf{y}}_t(i_t)} e_{i_t}$ we get,

$$\sum_{t=1}^{T} \tilde{\nabla}_t^\top \hat{\mathbf{y}}_t - \sum_{t=1}^{T} \tilde{\nabla}_t^\top f \le \frac{\eta}{2} \sum_{t=1}^{T} \nabla_t^2[i_t] + \frac{N}{\eta} \max_{i\in[N]} \left(\log(1/Nf[i]) + 1\right)$$

Now one thing to note, if we take $f = e_i$, RHS in the above bound blows up. Hence, we instead take $f_i = (1 - 1/T)e_i + \mathbf{1}/NT$ (that is best expert with a small amount of mixing of uniform distribution). The idea is that comparing with this expert is almost same as comparing with best expert with an additional additive term of $+1$ in our bound. Hence,

$$\sum_{t=1}^{T} \tilde{\nabla}_t^\top \hat{\mathbf{y}}_t - \min_{i\in[N]} \sum_{t=1}^{T} \tilde{\nabla}_t^\top e_i \le \frac{\eta}{2} \sum_{t=1}^{T} \nabla_t^2[i_t] + \frac{N}{\eta} \left(\log(T) + 1\right) + 1$$

Now since

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le \mathbb{E}\left[\mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T)\right]$$

Taking expectation over the draw of actions we get:

$$\begin{aligned}\mathbb{E}\left[\mathrm{Reg}_T\right] &\le \mathbb{E}\left[\frac{\eta}{2} \sum_{t=1}^{T} \nabla_t^2[i_t]\right] + \frac{N}{\eta} \left(\log(T) + 1\right) + 1\\ &\le \frac{\eta}{2} \sum_{t=1}^{T} \mathbb{E}\left[\nabla_t[i_t]\right] + \frac{N}{\eta} \left(\log(T) + 1\right) + 1\\ &= \frac{\eta}{2} \sum_{t=1}^{T} \mathbb{E}\left[\nabla_t^\top \hat{y}_t\right] + \frac{N}{\eta} \left(\log(T) + 1\right) + 1\end{aligned}$$

Hence we have shown that for any $i \in [N]$,

$$(1-\eta/2)\mathbb{E}\left[\sum_{t=1}^{T} \nabla_t^\top \hat{y}_t\right] - \sum_{t=1}^{T} \nabla_t[i] \le \frac{N}{\eta}\left(\log(T)+1\right)+1$$

or rewriting:

$$\mathbb{E}\left[\sum_{t=1}^{T} \nabla_t^\top \hat{y}_t\right] - \sum_{t=1}^{T} \nabla_t[i] \le \frac{\eta}{2(1-\eta/2)}\sum_{t=1}^{T} \nabla_t[i] + \frac{1}{(1-\eta/2)}\frac{N}{\eta}\left(\log(T)+1\right)+1$$

Say we know the value $L^* = \min_{i\in[N]} \sum_{t=1}^{T} \nabla_t[i]$ and we set $\eta = \min\{1, \sqrt{\frac{2N\log(eT)}{L^*}}\}$ then we have that:

$$\mathbb{E}\left[\sum_{t=1}^{T} \nabla_t^\top \hat{y}_t\right] - \min_{i\in[N]}\sum_{t=1}^{T} \nabla_t[i] \le O\left(\sqrt{N\log(T)\min_{i\in[N]}\sum_{t=1}^{T}\nabla_t[i]}+1\right)$$

5 Example: Linear Bandit on Euclidian Ball

In this example $\mathcal{F} = \{f : \|f\|_2 \le 1\}$. In this case we can use the logarithmic barrier

$$R(f) = -\log(1-\|f\|_2^2)$$

In this case, using the mirror descent update we get:

$$\mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T) \le \frac{\eta}{2}\sum_{t=1}^{T} \|\tilde{\nabla}_t\|^2_{\nabla^2 R(z_t)^{-1}} + \frac{1}{\eta}\sup_{f\in\mathcal{F}} \Delta_R(f|\hat{y}_1)$$

Again for a barrier function we can show that there is a constant c such that: $\|\cdot\|^2_{\nabla^2 R(z_t)^{-1}} \le c\|\cdot\|^2_{\nabla^2 R(\hat{\mathbf{y}}_t)^{-1}}$ and so we get:

$$\mathrm{Reg}_T(\tilde{\nabla}_1, \ldots \tilde{\nabla}_T) \le \frac{c\eta}{2}\sum_{t=1}^{T} \|\tilde{\nabla}_t\|^2_{\nabla^2 R(\hat{\mathbf{y}}_t)^{-1}} + \frac{1}{\eta}\sup_{f\in\mathcal{F}} \Delta_R(f|\hat{y}_1)$$

Now our randomized strategy we use is

$$\hat{y}_t = \hat{\mathbf{y}}_t + \epsilon_t(\lambda^t_{i_t})^{-1/2} v^t_{i_t}$$

where $i_t \sim \mathrm{unif}\{[d]\}$ and λ^t_i, v^t_i are the i'th eigenvalue and eigenvector of $\nabla^2 R^{-1}(\hat{\mathbf{y}}_t)$. We can then set

$$\tilde{\nabla}_t = d\ (\nabla_t^\top \hat{y}_t)\ \ \epsilon_t\sqrt{\lambda^t_{i_t}}v^t_{i_t}$$

First, you can easily verify that the above is an unbiased estimate. Next, note that:

$$
\begin{aligned}
\mathrm{Reg}_T(\tilde{\nabla}_1, \dots \tilde{\nabla}_T) &\leq \frac{c\eta}{2} \sum_{t=1}^{T} \|\tilde{\nabla}_t\|^2_{\nabla^2 R(\hat{\mathbf{y}}_t)^{-1}} + \frac{1}{\eta} \sup_{f \in \mathcal{F}} \Delta_R(f|\hat{y}_1) \\
&= \frac{cd^2\eta}{2} \sum_{t=1}^{T} (\nabla_t^\top \hat{y}_t)^2 \lambda_{i_t}^t (v_{i_t}^t)^\top \nabla^2 R(\hat{\mathbf{y}}_t)^{-1} v_{i_t}^t + \frac{1}{\eta} \sup_{f \in \mathcal{F}} \Delta_R(f|\hat{y}_1) \\
&= \frac{cd^2\eta}{2} \sum_{t=1}^{T} (\nabla_t^\top \hat{y}_t)^2 + \frac{1}{\eta} \sup_{f \in \mathcal{F}} \Delta_R(f|\hat{y}_1) \\
&\leq \frac{cd^2\eta T}{2} + \frac{1}{\eta} \sup_{f \in \mathcal{F}} \Delta_R(f|\hat{y}_1)
\end{aligned}
$$

In this case, if instead of all of the euclidean ball of radius one we consider ball of radius $1 - 1/T$ then $\sup_{f \in \mathcal{F}} \Delta_R(f|\hat{y}_1) \leq O(\log(T))$ and so optimizing over η we get

$$
\mathbb{E}[\mathrm{Reg}_T] \leq \mathbb{E}[\mathrm{Reg}_T(\tilde{\nabla}_1, \dots \tilde{\nabla}_T)] \leq O(d\sqrt{T \log(T)})
$$