Appendix of *Smooth Bandit Optimization: Generalization to Hölder Space*

A Auxiliary proofs for the main document

A.1 Proof of Lemma 2

Proof. Recall the definition of Hölder smoothness: $|f(x)-T_y^l(x)| \le L\|x-y\|_\infty^\alpha$. For a hypercube B, $\|x-y\|_\infty \le \Delta^{\frac{1}{d}}, \forall x,y \in B$. By definition, when the function smoothness exponent $\alpha \in (1,2]$, $l=1$. Notice that the Taylor polynomial of degree $l=1$ around y is a linear [12] function of x: $T_y^{(l=1)}(x) = f(y)+\frac{\partial f}{\partial x_1}(y)(x_1-y_1)+\frac{\partial f}{\partial x_2}(y)(x_2-y_2)+\ldots\frac{\partial f}{\partial x_d}(y)(x_d-y_d) = \langle\theta,x\rangle$. When $\alpha>2$, the Taylor polynomial can still be written as a linear function but of higher-dimensional feature map of x: $\phi:[0,1]^d \to [0,1]^{d(\alpha)}$ which contains exponentiations of x, using the operations defined for definition 1, $\phi(x)=\{x^s, \forall s, s.t. |s| \le l\}$. Dimension of the feature satisfies

$$d(\alpha) = |\{s: 1 \le |s| \le l\}| = \sum_{1\le j\le l} \binom{j+d-1}{d-1} = \mathcal{O}(d^l). \tag{10}$$

When $l=1$, $\phi(x)=x$. The parameter θ is determined by the derivatives of f at y and the value of y. Therefore, we know locally there exists a linear parameter in dimension $\theta^* = \arg\min_\theta \|f-\phi(x)^T\theta\|_\infty, x\in B$, such that $\|f-\langle\theta^*,\phi(x)\rangle\|_\infty \le \epsilon = L\Delta^{\frac{\alpha}{d}}, \forall x\in B$. Also, note that $\|\phi(x)\|_2^2 \le d(\alpha)^2$ according to definition. When the exponent $\alpha\in(0,1]$, l is 0 and the Taylor polynomial is simply a constant. Therefore the same argument holds for θ^* for example when $\theta_1^*,\ldots,\theta_d^*=0$ (a constant function). □

A.2 Proof of Theorem 3

Proof. Throughout this proof, we assume that the assumptions A1∼3 hold. This proof is modified from that in Dani et al. (2008). Some techniques are from Abbasi-Yadkori et al. (2011). We only present the parts which we change. First we proof the following bound on simple regret at each step:

$$r_t \le 2\sqrt{\beta_t}\|A_t^{-1/2}x\| + 2\epsilon\sum_{\tau=1}^{t-1}\|x^TA_t^{-1}x_\tau\|. \tag{11}$$

And then we will bound the sum of these two terms separately. In order to proof inequality 11, we start from an important auxiliary theorem of confidence bound on θ^*, Theorem 9.

Theorem 9. *Let $\beta_t = C\sigma^2 d\ln(1+t\kappa^2/d)\ln(\frac{2t^2}{\delta})$ $\left(=\mathcal{O}(d\ln(t)\ln(\frac{t^2}{\delta}))\right)$ for a sufficiently large constant C, then with probability $1-\delta$, θ^* is contained in the confidence set:*

$$\tilde{C}_t = \{\hat{\theta}_t + \sqrt{\beta_t}A_t^{-1/2}z_d - A_t^{-1}(\sum_{s=1}^{t-1}b_sx_s),\ \|z_d\|_2\le 1\},$$

and as a result,

$$\langle x,\theta^*\rangle \le \langle x,\hat{\theta}_t\rangle + \sqrt{\beta_t}\|A_t^{-1/2}x\| + \epsilon\sum_{s=1}^{t-1}|x^TA_t^{-1}x_s|.$$

The proof of Theorem 9 is in Appendix A.2.1. Now, if $\theta^*\in\tilde{C}_t$, we have

$$\begin{aligned} r_t &= \langle x^*,\theta^*\rangle - \langle x_t,\theta^*\rangle \\ &\le \langle x^*,\theta^*\rangle - UCB_t(x^*) + UCB_t(x_t) - \langle x_t,\theta^*\rangle \\ &\le UCB_t(x_t) - \langle x_t,\theta^*\rangle \\ &\le 2\sqrt{\beta_t}\|A_t^{-1/2}x_t\| + 2\epsilon\sum_{s=1}^{t-1}|x_t^TA_t^{-1}x_s|. \end{aligned}$$

[12] We slightly abuse the notation and define short-hand notation $\langle\theta,x\rangle := \theta_0 + \sum_{i=1}^{d(\alpha)}\theta_ix_i$.

The first inequality is because our algorithm will only choose x_t when $UCB_t(x_t) \geq UCB_t(x^*)$. The last inequality holds because

$$\begin{aligned}\langle x, \theta^* \rangle &\geq \langle x, \hat{\theta}_t \rangle + \min_{z_d \in B_2^d} \sqrt{\beta_t} \langle x, A_t^{-1/2} z_d \rangle - \sum_{s=1}^{t-1} b_s x^T A_t^{-1} x_s \\ &\geq \langle x, \hat{\theta}_t \rangle - \sqrt{\beta_t} \|A_t^{-1/2} x\| - \sum_{s=1}^{t-1} b_s x^T A_t^{-1} x_s \\ &\geq UCB_t(x) - 2\sqrt{\beta_t} \|A_t^{-1/2} x\| - 2\epsilon \sum_{s=1}^{t-1} |x^T A_t^{-1} x_s|.\end{aligned}$$

By assumption on the mean reward function value, the absolute value of instant pseudo-regret $|r_t|$ is bounded by $1 + \epsilon$. Therefore, combining inequality (11) and $r_t \leq 2 + 2\epsilon$, we have that[13]

$$\begin{aligned}r_t &\leq (2 + 2\epsilon) \wedge \left(2\sqrt{\beta_t} \|A_t^{-1/2} x_t\| + 2\epsilon \sum_{\tau=1}^{t-1} \|x_t^T A_t^{-1} x_\tau\| \right) \\ &\leq 2 \underbrace{\left(1 \wedge \sqrt{\beta_t} \|A_t^{-1/2} x_t\| \right)}_{\#1} + 2 \underbrace{\epsilon \sum_{\tau=1}^{t-1} \|x_t^T A_t^{-1} x_\tau\|}_{\#2} + 2\epsilon.\end{aligned} \tag{12}$$

Sum of term #1 is bounded using bound (28) and Cauchy Schwartz inequality:

$$2 \sum_{t=1}^{T} (1 \wedge \sqrt{\beta_t} \|A_t^{-1/2} x_t\|) \leq 2 \sqrt{T \beta_T \sum_{t=1}^{T} (1 \wedge \|x_t^T A_t^{-1} x_t\|)} = \sqrt{8 d \beta_T T \ln(1 + T\kappa^2/d)}. \tag{13}$$

For sum of term #2, we first have

$$\begin{aligned}\sum_{\tau=1}^{t-1} x_t^T A_t^{-1} x_\tau &\leq \sqrt{t \sum_{\tau=1}^{t-1} x_t^T A_t^{-1} x_\tau x_\tau^T A_t^{-1} x_t} \\ &= \sqrt{t x_t^T A_t^{-1} (\sum_{\tau=1}^{t-1} x_\tau x_\tau^T) A_t^{-1} x_t} \\ &\leq \sqrt{t x_t^T A_t^{-1} (\sum_{\tau=1}^{t-1} x_\tau x_\tau^T) A_t^{-1} x_t + x_t^T A_t^{-1} A_t^{-1} x_t} \\ &= \sqrt{t x_t^T A_t^{-1} (\sum_{\tau=1}^{t-1} x_\tau x_\tau^T + I_d) A_t^{-1} x_t} = \sqrt{t x_t^T A_t^{-1} x_t}.\end{aligned}$$

Then the sum $\sum_{t=1}^{T} (\sum_{\tau=1}^{t-1} x_t^T A_t^{-1} x_\tau)$ can be bounded by:

$$\begin{aligned}\sum_{t=1}^{T} (\sum_{\tau=1}^{t-1} x_t^T A_t^{-1} x_\tau) &\leq \sum_{t=1}^{T} (\sqrt{t x_t^T A_t^{-1} x_t}) \\ &\leq \sqrt{(\sum_{t=1}^{T} t)(\sum_{t=1}^{T} x_t^T A_t^{-1} x_t)}.\end{aligned}$$

Now, we need to bound $\sum_{t=1}^{T} x_t^T A_t^{-1} x_t$ with inequality (28). We know that A_t^{-1} is a full-rank matrix. Therefore,

[13] $a \wedge b = \min(a, b)$

denote its eigenvalues and eigenvectors as $\lambda_1 \dots \lambda_d, v_1 \dots v_d$. Then[14]

$$\begin{aligned} x_t^T A_t^{-1} x_t &= (c_1 v_1 + \cdots + c_d v_d)^T A_t^{-1} (c_1 v_1 + \cdots + c_d v_d) \\ &= c_1^2 \lambda_1 + \cdots + c_d^2 \lambda_d \\ &\le \lambda_{\max}(A_t^{-1}) \|x_t\|_2^2 = \frac{\kappa^2}{\lambda_{\min}(A_t)} \\ &\le \frac{\kappa^2}{\lambda_{\min}(I_d) + \lambda_{\min}(X_t^T X_t)} \le \kappa^2. \end{aligned}$$

The second last inequality holds due to Weyl's inequality. Therefore,

$$\begin{aligned} \sum_{t=1}^T x_t^T A_t^{-1} x_t &\le \kappa^2 \sum_{t=1}^T (x_t^T A_t^{-1} x_t \wedge 1) \\ &\le \kappa^2 (2d \ln(1 + T\kappa^2/d)). \end{aligned}$$

Putting the above together,

$$\begin{aligned} \sum_{t=1}^T \left(2\epsilon \sum_{\tau=1}^{t-1} x_t^T A_t^{-1} x_\tau \right) &\le 2\epsilon \sqrt{(\sum_{t=1}^T t)(\sum_{t=1}^T x_t^T A_t^{-1} x_t)} \\ &\le 2\epsilon T\kappa \sqrt{2d \ln(1 + T\kappa^2/d)}. \end{aligned} \tag{14}$$

Finally, plugging in $\kappa^2 = d$ gives the final results. □

A.2.1 Proof of Theorem 9

Proof. Let $\hat{\theta}_t = A_t^{-1} X_t^T y$ denote the regularized least square estimator at time t. Matrix X_t has dimension $(t-1) \times d$, where each row is a past action (until time t). We first define an unobserved variable $\tilde{\theta}_t$:

$$\tilde{\theta}_t = A_t^{-1} X_t^T (X_t \theta^* + \eta_t) = \hat{\theta}_t - A_t^{-1} X_t^T b_t, \tag{15}$$

here we abuse the notations and let η_t and b_t be the $(t-1) \times 1$ vector containing noise and bias of each time. Then we define the following confidence ellipsoid centered at $\tilde{\theta}_t$:

$$C_t = \{\theta : (\theta - \tilde{\theta}_t)^T A_t (\theta - \tilde{\theta}_t) \le \beta_t\}, \tag{16}$$

and prove the following lemma as an analog to Theorem 5 of Dani et al. (2008):

Lemma 10. *The true linear parameter θ^* is contained in ellipsoid C_t, specifically, $\mathbb{P}(\forall t, \theta^* \in C_t) \ge 1 - \delta$.*

The proof is in Appendix section A.2.2. However, we do not observe the vector b_t, so we cannot calculate C_t in our algorithm. So instead, we define a larger $\tilde{C}_t$ that contains C_t, which will naturally contains θ^* with high probability. To construct $\tilde{C}_t$, we first re-write C_t as

$$C_t = \{\tilde{\theta}_t + \sqrt{\beta_t} A_t^{-1/2} z_d, \|z_d\|_2 \le 1\}, \tag{17}$$

then plug in equation (15) to yield:

$$\begin{aligned} \tilde{\theta}_t + \sqrt{\beta_t} A_t^{-1/2} z &= \hat{\theta}_t + \sqrt{\beta_t} A_t^{-1/2} z - A_t^{-1} X_t^T b_t \\ &= \hat{\theta}_t + \sqrt{\beta_t} A_t^{-1/2} z - A_t^{-1} (\sum_{s=1}^{t-1} b_s x_s). \end{aligned} \tag{18}$$

Therefore, we know that with high probability,

$$\theta^* \in \tilde{C}_t = \{\hat{\theta}_t + \sqrt{\beta_t} A_t^{-1/2} z_d - A_t^{-1} (\sum_{s=1}^{t-1} b_s x_s)\}. \tag{19}$$

[14] This proof is extracted from a remark in proof of Theorem 3 in Abbasi-Yadkori et al. (2011)

Therefore, we have a computable confidence bound for x:

$$\begin{aligned} UCB_t(x) &= \max_{\theta \in \tilde{C}_t} \langle x, \theta \rangle \\ &= \langle x, \hat{\theta}_t \rangle + \max_{z_d \in B_2^d} \sqrt{\beta_t} \langle x, A_t^{-1/2} z_d \rangle - \sum_{s=1}^{t-1} b_s x^T A_t^{-1} x_s \\ &\leq \langle x, \hat{\theta}_t \rangle + \sqrt{\beta_t} \|A_t^{-1/2} x\| - \sum_{s=1}^{t-1} b_s x^T A_t^{-1} x_s \\ &\leq \langle x, \hat{\theta}_t \rangle + \sqrt{\beta_t} \|A_t^{-1/2} x\| + \epsilon \sum_{s=1}^{t-1} |x^T A_t^{-1} x_s|. \end{aligned} \tag{20}$$

The first inequality is derived by Cauchy Schwartz inequality and the fact that z_d is in unit ball. □

A.2.2 Proof of Lemma 10

Proof. Lemma 10 is a parallel to Theorem 5 in Dani et al. (2008), with the difference of sub-gaussian noise, ellipsoid centre $\tilde{\theta}_t$ and misspecification in observation. The key idea is the same, namely to use induction to bound the growth of $Z_t = (\theta^* - \tilde{\theta}_t)^T A_t (\theta^* - \tilde{\theta}_t)$ and proof that $Z_t \leq \beta_t$, i.e. the θ^* is contained in C_t, at each time step t. The following analysis used the same notations and definitions as section 5.2 in Dani et al. (2008) unless otherwise specified. Under Lemma 10's definition of confidence set C_t, we have that:

$$H_t = A_t(\tilde{\theta}_t - \theta^*) = X_t^T \eta_t - \theta^*, \tag{21}$$

$$Z_t = (\theta^* - \tilde{\theta}_t)^T A_t (\theta^* - \tilde{\theta}_t) = H_t^T A_t^{-1} H_t. \tag{22}$$

Equation 21 holds because of this key property:

$$\tilde{\theta}_t : A_t \tilde{\theta}_t = X_t^T X_t \theta^* + X_t^T \eta_t. \tag{23}$$

And the rest of the proof in Dani et al. (2008) should go through by substituting Y_t with H_t (defined above) and $\hat{\mu}$ with our definition of $\tilde{\theta}$ (centre of the confidence ellipsoid). Except, to accommodate the sub-gaussian noise assumption that replaces their bounded noise assumption, we have to make two changes in the proof. Both are in analyzing the growth of Z_t in the induction. Recall that Dani et al. (2008) proved this relation:

$$Z_t \leq Z_1 + 2 \sum_{\tau=1}^{t-1} \eta_t \frac{x_t^T (\tilde{\theta}_t - \theta^*)}{1 + w_t^2} + \sum_{\tau=1}^{t-1} \eta_\tau^2 \frac{w_\tau^2}{1 + w_\tau^2}. \tag{24}$$

We first look at the concentration of the sum of martingale difference sequence that makes up Z_t: same with Dani et al. (2008), define $M_t = 2\eta_t \frac{x_t^T (\tilde{\theta}_t - \theta^*)}{1 + w_t^2}$ where $w_t \triangleq \sqrt{x_t^T A_t^{-1} x_t}$. According to our assumption, the noise sequence is a sub-gaussian martingale difference sequence with parameter σ^2. Therefore, M_t is a sub-gaussian martingale difference sequence. Specifically, we know that the square of subgaussian parameter is $4\sigma^2 (\frac{|x_t^T (\tilde{\theta}_t - \theta^*)|}{1 + w_t^2})^2$. By definitions we know that $M_t \mid \mathcal{H}_t$ is $(\nu_t^2 = 4\sigma^2 (\frac{|x_t^T (\tilde{\theta}_t - \theta^*)|}{1 + w_t^2})^2, a_t = 0)$ sub-exponential(definition 2.7 in Wainwright (2019)) and therefore the sum $\sum_{\tau=1}^t M_\tau$ is also sub-exponential, with parameters $(\sqrt{\sum_{\tau=1}^t \nu_\tau^2}, a = \max_\tau a_\tau = 0)$(Theorem 2.19 (1) in Wainwright (2019)). The following inequality is conditioned on the fact that from time

$\tau = 1 \dots t$, θ^* is contained in C_τ (by the induction).

$$\begin{aligned}
\sum_{\tau}^{t} \nu_t^2 &= 4\sigma^2 \sum_{\tau=1}^{t} (\frac{|x_\tau^T(\tilde{\theta}_\tau - \theta^*)|}{1+w_\tau^2})^2 \\
&\le 4\sigma^2 \sum_{\tau=1}^{t} (\frac{\sqrt{\beta_\tau} w_\tau}{1+w_\tau^2})^2 \\
&\le 4\sigma^2 \sum_{\tau=1}^{t} \beta_\tau (\min(1/2, w_\tau))^2 \\
&\le 4\sigma^2 \sum_{\tau=1}^{t} \beta_\tau \min(1/4, w_\tau^2) \\
&\le 4\sigma^2 \beta_t \sum_{\tau=1}^{t} \min(1, w_\tau^2) \\
&\le 4\sigma^2 \beta_t \left(2d \ln(1 + t\kappa^2/d)\right) \text{ See bound 28} \\
&= 8\sigma^2 d \beta_t \ln(1 + t\kappa^2/d).
\end{aligned}$$

The proof for the first three inequalities is the same as Lemma 7 and section 5.2.1 in Dani et al. (2008). Then we apply a Bernstein-type concentration bound for sub-exponential martingale difference sequence (Theorem 2.19 (2) in Wainwright (2019)). Plugging in the values of a and $\sum_{\tau=1}^{t} \nu_\tau^2$, we have that

$$\begin{aligned}
\mathbb{P}(|\sum_{\tau=1}^{t-1} M_\tau| \ge s) &\le 2\exp(\frac{-s^2}{2\sum_{\tau=1}^{t-1} \nu_t^2}) \\
&\le 2\exp(\frac{-s^2}{16\sigma^2 d \beta_t \ln(1 + (t-1)\kappa^2/d)}) \\
&\overset{s=\frac{\beta_t}{2}}{=} 2\exp\left(\frac{-\beta_t}{64\sigma^2 d \ln(1 + (t-1)\kappa^2/d)}\right) \\
&\le \frac{\delta}{2t^2} \text{ (Needed for union bound over all times).}
\end{aligned} \tag{25}$$

Therefore, as long as β_t is larger or equal to $64\sigma^2 d \ln(1 + (t-1)\kappa^2/d) \ln(\frac{4t^2}{\delta})$, $\sum_{\tau=1}^{t-1} M_\tau \le \frac{\beta_t}{2}$ with probability larger or equal to $1 - \frac{\delta}{2t^2}$.

The second change is for the third quantity that makes up Z_t: $\sum_{\tau=1}^{t-1} \eta_\tau^2 \frac{w_\tau^2}{1+w_\tau^2}$. We need to bound $\max_{\tau \le t-1} \eta_\tau^2$ with high probability. By algebra calculations, we know that η_τ^2 is sub-exponential with parameters ($\nu = 32\sigma^4, a = 4\sigma^2$)[15]. We can apply union bound with the tail bound of sub-exponential variables:

$$\begin{aligned}
\mathbb{P}(\max_{\tau \le t-1} (\eta_\tau^2 - \mathbb{E}[\eta^2]) \ge z) &\le \sum_{\tau=1}^{t-1} \mathbb{P}((\eta_\tau^2 - \mathbb{E}[\eta^2]) \ge z) \\
&\le (t-1)\exp(-\frac{z}{2a}) \text{ (Proposition 2.9 in Wainwright (2019))} \\
&\le \frac{\delta}{2t^2} \text{ (Needed for union bound over all times).}
\end{aligned}$$

Set $z = 8\sigma^2 \ln(\frac{2t^3}{\delta})$ so that $\mathbb{P}(\max_{\tau \le t-1} \eta_\tau^2 - \mathbb{E}[\eta^2] \le z) = \mathbb{P}(\max_{\tau \le t-1} \eta_\tau^2 \le z + \mathbb{E}[\eta^2]) \ge 1 - \frac{\delta}{2t^2}$. By the fact that $\mathbb{E}[\eta] = 0, \mathbb{E}[\eta^2] = \mathrm{Var}(\eta) \le \sigma^2$, which is a property of subgaussian variables. So $\mathbb{P}(\max_{\tau \le t-1} \eta_\tau^2 \le z + \sigma^2) \ge 1 - \frac{\delta}{2t^2}$.

[15] For this part, we borrowed the proof from Example 2.8 in Wainwright (2019) and `http://proceedings.mlr.press/v33/honorio14-supp.pdf`

The following holds with probability larger than $1 - \frac{\delta}{2t^2}$:

$$\begin{aligned}
\sum_{\tau=1}^{t-1} \eta_\tau^2 \frac{w_\tau^2}{1+w_\tau^2} &\le (\max_{\tau \le t-1} \eta_\tau^2) \sum_{\tau=1}^{t-1} \min(w_\tau^2, 1) \\
&\le (\max_{\tau \le t-1} \eta_\tau^2) 2d \ln(1 + t\kappa^2/d) \\
&= (8\sigma^2 \ln(\frac{2t^3}{\delta}) + \sigma^2) 2d \ln(1 + (t-1)\kappa^2/d) \\
&= 8\sigma^2 (\ln(\frac{2t^3}{\delta}) + \frac{1}{8}) 2d \ln(1 + (t-1))\kappa^2/d) \\
&= 16\sigma^2 d \ln(1 + (t-1)\kappa^2/d) \left(\ln(\frac{2t^3}{\delta}) + \frac{1}{8} \right).
\end{aligned}$$

Except the two changes above, one last thing to note is the quantity Z_1 analyzed at the end of proof of Lemma 12 in Dani et al. (2008). In our assumption of the reward function value, we conclude that

$$\begin{aligned}
Z_1 &= (\theta^* - 0)^T I (\theta^* - 0) = \|\theta^*\|_2^2 \\
&= \sum_{i=1}^{d} (e_i^T \theta^*)^2 \ (e_i \text{ is base vector of dimension } i, \text{ note that } e_i \in \mathcal{X}) \\
&\le d(1+\epsilon)^2.
\end{aligned}$$

As a result, if it is satisfied that $Z_t \le Z_1 + \beta_t/2 + 16\sigma^2 d \ln(1 + (t-1)\kappa^2/d)(\ln(\frac{2t^3}{\delta}) + \frac{1}{8}) \le \beta_t$, which enables the induction in Lemma 14 in Dani et al. (2008), then the rest of the proof should go through smoothly. We argue that setting $\beta_t = C\sigma^2 d \ln(t) \ln(\frac{4t^2}{\delta})$ for a large enough constant C suffices. This is under the reasonable assumption that ϵ is $\mathcal{O}(1)$ and σ is a constant[16].

It is worth mentioning[17] that Dani et al. (2008) requires the relationship between t and δ to be approximately $0 < 1.05\delta \le t^2$, hence their requirement[18] of "for sufficiently large T" in Theorem 1 and 2. This is because of the last step of their induction proof for Theorem 5 requires: $Z_t \le d + \beta_2/2 + 2d\ln(t) \le \beta_t$. In our setting, the requirement in induction translates to this (second) constraint(plugging in $\kappa^2 = d$): $\beta_t \ge 2d(1+\epsilon)^2 + 32\sigma^2 d \ln(t)(\ln(\frac{2t^3}{\delta} + \frac{1}{8}))$. Recall the first constraint on β_t is $\beta_t \ge 64\sigma^2 d \ln(t) \ln(\frac{4t^2}{\delta})$, from bound (25). Therefore, C should first satisfy $C \ge 64$ and for the second constraint we need[19]: $C \ge \frac{3(1+\epsilon)^2}{4(\ln(2))^2\sigma^2} + \frac{3}{2\ln(2)} + 48$. Therefore, the lower bound of C should depend on values of ϵ and σ^2. The choice of $C = 128$ in the main theorem is an example that requires approximately $\frac{1+\epsilon}{\sigma} \le 7$. □

A.3 Proof of Theorem 4

Let us treat the number of bins/local algorithms n as the input parameter to the algorithm. The regret bound of UCB-Meta (equation 4) should be independent of the input dimension d, given the dimension of the linear model $d(\alpha)$. Therefore, throughout this proof we will abuse the notations and let d denote the linear model dimension for simplicity.

Proof. First, we define the "good event" E_{good} as an event where all confidence bound holds for all bins at all times. For a fixed bin, if $\mathbb{P}(\theta^* \notin \tilde{C}_t, \exists t) \le \delta/n$, as set in the algorithm, where $\tilde{C}_t = \{\hat{\theta}_t + \sqrt{\beta_t} A_t^{-1/2} z_d - A_t^{-1}(\sum_{s=1}^{t-1} b_s x_s)\}$ (Theorem 9), then by union bound, $\mathbb{P}(\theta_k^* \notin \tilde{C}_{k,t}, \exists k) \le \delta$, where $\tilde{C}_{k,t}$ is the confidence ellipsoid of bin k at time t. The good event is $E_{good} = \{\forall t, \forall k \in [n], \theta_k^* \in \tilde{C}_{k,t}\}$. It happens with probability $\mathbb{P}(E_{good}) \ge 1 - \delta$, and the following proof will condition on it.

[16] Recall that according to Lemma 2, ϵ is bounded by the Lipschitz constant L and is therefore $\mathcal{O}(1)$

[17] This remark is made by Abbasi-Yadkori et al. (2011).

[18] However, we believe that this should not translate to a constraint on t, but on δ instead. Because $Z_t \le \beta_t$ is required for every step t to complete the induction, so if it only holds for large t then the induction will fail as well.

[19] This is from the second constraint: $C\sigma^2 d \ln(t) \ln(\frac{4t^2}{\delta}) \ge \frac{2}{3} C\sigma^2 d \ln(t) \ln(\frac{2t^3}{\delta}) \ge 2d(1+\epsilon)^2 + 32\sigma^2 d \ln(t)(\ln(\frac{2t^3}{\delta} + \frac{1}{8}))$.

Here are some useful notations that make the proof easier to read: let $N^k(t)$ denote the number of times base-algorithm $\mathcal{A}_k^{local}$ has been selected by(including) time t; let $k(t)$ denote the bin selected at time t; let x_t denote the action selected at time t; let $\{\beta_{k,\cdot}\}$, $\{A_{k,\cdot}\}$ and $\{\hat{\theta}_{k,\cdot}\}$ denote the set of parameters kept by that base-algorithm $\mathcal{A}_k^{local}$.

The upper confidence bound on value of the local linear function achieved by sub-algorithms at round t is defined as $UCB_{k(t),t}(x) = \langle x, \hat{\theta}_{k,N^k(t)}\rangle + \sqrt{\beta_{k,N^k(t)}}\|A_{k,N^k(t)}^{-1/2}x\| + \epsilon\sum_{\tau=1}^{N^k(t)-1}|x^TA_{N^k(t)}^{-1}x_\tau|$ for any action $x \in B_k$. Using the proof of Theorem 3, the good event hence indicates that for the base-algorithm selected at time t and any action $x \in B_{k(t)}$:

$$UCB_{k(t),t}(x) - 2\sqrt{\beta_{k,N^k(t)}}\|A_{k,N^k(t)}^{-1/2}x\| - 2\epsilon\sum_{\tau=1}^{N^k(t)-1}|x^TA_{N^k(t)}^{-1}x_\tau| \le \langle x, \theta_k^*\rangle \le UCB_{k(t),t}(x).$$

By Lemma 2, the expected local function value $f(x)$ is bounded by

$$UCB_{k(t),t}(x) - 2\sqrt{\beta_{k,N^k(t)}}\|A_{k,N^k(t)}^{-1/2}x\| - 2\epsilon\sum_{\tau=1}^{N^k(t)-1}|x^TA_{N^k(t)}^{-1}x_\tau| - \epsilon \le f(x) \le UCB_{k(t),t}(x) + \epsilon.$$

A common way to bound pseudo regret for stochastic bandit is via Wald's equality: $R_T = \sum_{k=1}^n \Delta_k\mathbb{E}[\tau_k(T)]$ where $\tau_k(T)$ is the number of times arm k gets pulled until time T, and Δ_k is the reward gap. We cannot trivially follow this, because the rewards of each bins are no longer i.i.d. Instead, we use this gap-independent decomposition for each bin k:

$$\begin{aligned}
R_k &= \sum_{t:\mathrm{bin}_t=k}(f^* - f_{x_t\in B_k}(x_t)) \\
&= \sum_{t:\mathrm{bin}_t=k}\big(f^* - UCB_{\mathcal{A}_{k(t)},t} + UCB_{\mathcal{A}_{k(t)},t} - f(x_t)\big) \\
&= \sum_{t:\mathrm{bin}_t=k}\big(f^* - UCB_{\mathcal{A}_{k(t)},t} + UCB_{k(t),t}(x_t) + \epsilon - f(x_t)\big) \\
&\le \sum_{t:\mathrm{bin}_t=k}\big(UCB_{k(t),t}(x_t) + \epsilon - f(x_t)\big) \qquad (26)\\
&\le \sum_{t:\mathrm{bin}_t=k}\left(2\sqrt{\beta_{k,N^k(t)}}\|A_{k,N^k(t)}^{-1/2}x_t\| + 2\epsilon\sum_{\tau=1}^{N^k(t)-1}|x_t^TA_{N^k(t)}^{-1}x_\tau| + 2\epsilon\right) \\
&= \sum_{s=1}^{N^k(T)}\left(2\sqrt{\beta_{k,s}}\|A_{k,s}^{-1/2}x_t\| + 2\epsilon\sum_{\tau=1}^{s-1}|x_t^TA_s^{-1}x_\tau| + 2\epsilon\right).
\end{aligned}$$

The first inequality holds because of the algorithm's bin selection rule: if bin B_k is chosen then $f^* \le UCB_{k*,t} \le UCB_{k(t)}$. By the bounded function value assumption, $f^* - f_{x_t\in B_k}(x_t) \le 2$, therefore:

$$\begin{aligned}
R_k &\le \sum_{s=1}^{N^k(T)}\left(2\sqrt{\beta_{k,s}}\|A_{k,s}^{-1/2}x_t\| + 2\epsilon\sum_{\tau=1}^{s-1}|x_t^TA_s^{-1}x_\tau| + 2\epsilon\right)\wedge 2 \\
&\le \sum_{s=1}^{N^k(T)}\left(\underbrace{2\left(\sqrt{\beta_{k,s}}\|A_{k,s}^{-1/2}x_t\|\wedge 1\right)}_{\#1} + \underbrace{2\epsilon\sum_{\tau=1}^{s-1}|x_t^TA_s^{-1}x_\tau|}_{\#2}\right) + 2\epsilon N^k(T). \qquad (27)
\end{aligned}$$

A.3.1 High probability regret bound part I (term #1)

First we establish this bound the same way as Dani et al. (2008). Namely, for any local misspecified linear bandit algorithm that is ran T times with data $(x_t, y_t)_{t=1\dots T}$,

$$\begin{aligned}\sum_{t=1}^{T}\|x_t^T A_t^{-1}x_t\| \wedge 1 &\le 2\ln(\prod_{t=1}^{T}(1+x_t^T A_t^{-1}x_t))\\ &= 2\ln(\prod_{t=1}^{T}\frac{\det(A_{t+1})}{\det(A_t)})\\ &= 2\ln(\frac{\det A_{T+1}}{\det A_1}) \le 2\ln((1+T\kappa^2/d)^d)\\ &= 2d\ln(1+T\kappa^2/d),\end{aligned} \tag{28}$$

where we used Lemma 11. Now we can bound term #1 using bound (28).

$$\begin{aligned}&\sum_{s=1}^{N^k(T)} 2(\sqrt{\beta_{k,s}}\|A_{k,s}^{-1/2}x_t\| \wedge 1)\\ &\le \sqrt{N^k(T)\sum_{s=1}^{N^k(T)} 4(\beta_{k,s}\|x_{k,s}^T A_{k,s}^{-1}x_{k,s}\| \wedge 1)}\\ &\le \sqrt{4\beta_{k,N^k(T)}N^k(T)\sum_{s=1}^{N^k(T)}\|x_{k,s}^T A_{k,s}^{-1}x_{k,s}\| \wedge 1}\\ &= \sqrt{4\beta_{k,N^k(T)}N^k(T)2\ln\left(\prod_{s=1}^{N^k(T)}(1+x_{k,s}^T A_{k,s}^{-1}x_{k,s})\right)}\\ &= \sqrt{4\beta_{k,N^k(T)}N^k(T)2\ln\left(\frac{\det(A_{N^k(T)+1})}{\det(A_1)}\right)}\\ &= \sqrt{8d\beta_{k,N^k(T)}N^k(T)\ln(1+N^k(T)\kappa^2/d)}\\ &\overset{\kappa^2=d}{=} \sqrt{8d\beta_{k,N^k(T)}N^k(T)\ln(1+N^k(T))}.\end{aligned}$$

Lemma 11. *For $t \ge 1$, $1 + x_t^T A_t^{-1}x_t = \det(A_{t+1})/\det(A_t)$. Also, $\det(A_t) \le (1+(t-1)\kappa^2/d)^d$.*

Proof of Lemma 11.

$$\begin{aligned}\det(A_{t+1}) &= \det(A_t(I_d + A_t^{-1}x_t x_t^T)) = \det(A_t)\det(I_d + A_t^{-1}x_t x_t^T)\\ &= \det(A_t)\det(I_1 + x_t^T A_t^{-1}x_t) = det(A_t)(1 + x_t^T A_t^{-1}x_t).\end{aligned}$$

The third equation uses Sylvester's determinant theorem: $\det(I_m + A_{m\times n}B_{n\times m}) = \det(I_n + B_{n\times m}A_{m\times n})$. The trace of a matrix is the product of its eigenvalues and the determinant is the sum of eigenvalues, and for the trace of the positive definite matrix A_t we have,

$$\mathrm{tr}(A_t) = \mathrm{tr}(I + \sum_{\tau}^{t-1} x_\tau x_\tau^T) = d + \sum_{\tau}^{t-1}\|x_\tau\|_2^2 \le d + (t-1)\kappa^2.$$

Therefore, using the inequality of arithmetic and geometric mean, $\det(A_t) \le (1+(t-1)\kappa^2/d)^d$. □

Summing over all the suboptimal bins, we have that

$$\begin{aligned}
&\sum_{k=1}^{n-1}\sum_{s=1}^{N^k(T)} 2(\sqrt{\beta_{k,s}}\|A_{k,s}^{-1/2}x_{k,s}\| \wedge 1) \le \sum_{k=1}^{n}\sqrt{8d\beta_{k,N^k(T)}N^k(T)\ln\left(1+N^k(T)\right)} \\
&\le \sqrt{\sum_{k=1}^{n} N^k(T)\sum_{k=1}^{n} 8d\beta_{k,N^k(T)}\ln\left(1+N^k(T)\right)} \\
&= \sqrt{T\sum_{k=1}^{n} 8d\beta_{k,N^k(T)}\ln\left(1+N^k(T)\right)} \\
&\overset{N^k(T)\le T}{\le} \sqrt{8dTn\beta_T\ln\left(1+T\right)}.
\end{aligned} \tag{29}$$

A.3.2 High probability regret bound part II (term #2)

Here we directly call previous result in bound (14), but replace the total number of step with $N^k(T)$, the number of pulls for one fixed bin k. We have for term #2,

$$\sum_{s=1}^{N^k(T)} 2\epsilon\sum_{\tau=1}^{s-1}|x_{k,s}^T A_{k,s}^{-1}x_{k,\tau}| \le 2\epsilon N^k(T)d\sqrt{2\ln(1+N^k(T))}.$$

Summing over all suboptimal bins, we have that

$$\begin{aligned}
&\sum_{k=1}^{n} 2\epsilon N^k(T)d\sqrt{2\ln(1+N^k(T))} \\
&\overset{N^k(T)\le T}{\le} 2\epsilon d\sqrt{2\ln(1+T)}\sum_{k=1}^{n} N^k(T) \\
&= 2\epsilon dT\sqrt{2\ln(1+T)}.
\end{aligned} \tag{30}$$

A.3.3 Putting it together

Combining the decomposition in equation (27) and the results in subsections A.3.1 and A.3.2, we have a high probability regret bound for the UCB-Meta-algorithm:

$$\begin{aligned}
R_T &= \sum_{k=1}^{n} R_k \\
&\le \sqrt{8dTn\beta_T\ln\left(1+T\right)} + 2\epsilon dT\sqrt{2\ln(1+T)} + 2\epsilon T \\
&= \mathcal{O}(d\ln(T)\sqrt{Tn\ln(T^2n/\delta)} + \epsilon dT\sqrt{\ln(T)} + \epsilon T).
\end{aligned} \tag{31}$$

The last step plugs in $\beta_T = \mathcal{O}(d\ln(T)\ln(T^2n/\delta))$.

□

A.4 Proof of Theorem 5

Proof. Algorithm 3 executes Algorithm 2 for a sequence of pre-defined time periods, $\{T_i = 2^i, i = 0, 1, \dots N\}$. At the beginning of each period, the update history is cleared and the number of arms n is reset with respect to the current horizon T_i. However, since we would like to acquire a high-probability regret bound after applying the doubling trick, we need to set the fail probability of Meta-algorithms during period i to $\delta_i = 6\delta/\pi^2 i^2$. Using a union bound, we can conclude the following ($R_i(T_i)$ denotes the regret incurred in time period i of length T_i

only).

$$\begin{aligned}&\mathbb{P}(\forall i, \text{the bound hold for } R_i(T_i))\\&= 1 - \sum_i \mathbb{P}(\text{the bound does not hold for} R_i(T_i))\\&= 1 - \sum_i \frac{6\delta}{\pi^2 i^2} \approx 1 - \delta.\end{aligned}$$

In the last step we use the fact that the sum of sequence $\sum_i^\infty \frac{1}{i^2}$ converges to $\frac{\pi^2}{6}$.

Now, the total regret is simply a summation over i. The following holds with probability $1-\delta$,

$$\begin{aligned}R(T) &\le \sum_{i=1}^N R_i(T_i)\\&\le \sum_{i=1}^N \tilde{\mathcal{O}}(dT_i^a) = \tilde{\mathcal{O}}\left(d\sum_{i=1}^N 2^{ia}\right)\\&\le \tilde{\mathcal{O}}\left(d2^{a(N-1)}\right)\\&= \tilde{\mathcal{O}}(dT^a).\end{aligned} \tag{32}$$

At step 4, the number of time periods N is the smallest integer such that $\sum_{i=0}^N 2^i \ge T$, so $N = 1 + \lceil \log_2(T) \rceil$. The sum of geometric sequence is $2^{a\lceil \log_2(T)\rceil} = (2^{\log_2(T)+c})^a = T^a 2^{ca}$ for some constant c smaller than 1. Also, note that step 2 holds even though the fail probability is changed to $\delta_i = 6\delta/\pi^2 i$ is because as specified in Theorem 4, the term δ appears in a log term and the maximum value of $1/\delta$ is $1/\delta_N = \pi^2 \log_2(T)/6\delta$, therefore the extra factor caused by smaller δ to the regret is still a log term of T_i and omitted in the proof here.

Bound (32) suffices to say that meta-algorithm with doubling trick has the same regret rate as meta-algorithm with known horizon, with some additional constant factors suffered from restarting. □

A.5 Proof of Theorem 6

Proof. Here we prove that Corral with smooth-wrapper is applicable to this task and achieves minimax expected regret rate apart from log factors. We directly use the proof of Theorem 5.3 in Pacchiano et al. (2020) and their notations. δ is the fail probability, M is the number of base-algorithms, ρ is the reciprocal of the smallest possibility for base-algorithms over the T rounds and η is the learning rate. $U(T,\delta)$ is the high probability bound of the selected base-algorithm. The regret of Corral with smooth wrapper is bounded by:

$$R(T) \le \mathcal{O}(\frac{M\ln(T)}{\eta} + T\eta) + \delta T + 8\sqrt{MT\log(\frac{4TM}{\delta})} - \mathbb{E}[\frac{\rho}{40\eta\ln(T)} - 2\rho U(T/\rho,\delta)\log(T)], \tag{33}$$

and we know from Theorem 3 in our paper that the base algorithm (Algorithm 1) that locates in the global maximum's bin has anytime high probability regret bound $U(T,\delta) = \tilde{\mathcal{O}}(\epsilon T d(\alpha) + c(\delta)d(\alpha)\sqrt{T})$, note that this is because the dimension of the local linear parameter is $d(\alpha)$. Therefore,

$$\begin{aligned}R(T) &= \tilde{\mathcal{O}}(\sqrt{MT} + \frac{M}{\eta} + T\eta) + \delta T - \mathbb{E}[\frac{\rho}{40\eta\ln(T)} - 2\rho\tilde{\mathcal{O}}(d(\alpha)\sqrt{T/\rho} + \frac{\epsilon d(\alpha)T}{\rho})])\\&= \tilde{\mathcal{O}}(\sqrt{MT} + \frac{M}{\eta} + T\eta) + \delta T + \tilde{\mathcal{O}}(\epsilon T d(\alpha)) + \mathbb{E}[\tilde{\mathcal{O}}(d(\alpha)\sqrt{T\rho} - \frac{\rho}{\eta})].\end{aligned} \tag{34}$$

Firstly, we set $\delta = 1/T$ so that $\delta T = \mathcal{O}(1)$. Then we maximize this formulation over ρ by setting $\rho = \tilde{\mathcal{O}}(\eta^2 d(\alpha)^2 T)$, yielding the following bound on expected regret.

$$\begin{aligned}&\tilde{\mathcal{O}}(\sqrt{MT} + \frac{M}{\eta} + T\eta + \epsilon T d(\alpha) + \eta d(\alpha)^2 T)\\&\overset{\substack{M=n,\\ \epsilon = n^{-\frac{\alpha}{d}}}}{=} \tilde{\mathcal{O}}(\sqrt{nT} + \frac{n}{\eta} + n^{-\frac{\alpha}{d}} T d(\alpha) + \eta d(\alpha)^2 T).\end{aligned} \tag{35}$$

We minimize this by setting the derivative w.r.t n and η to zero, i.e. $\eta = \frac{1}{d(\alpha)}\sqrt{\frac{n}{T}}$ and $n = \tilde{\mathcal{O}}(T^{\frac{d}{d+2\alpha}})$. As a result the rate comes to $\tilde{\mathcal{O}}(d(\alpha)T^{\frac{d+\alpha}{d+2\alpha}})$. □

A.6 Proof of Lemma 7

Proof. According to Theorem 4, the algorithm sets $n = T^{\frac{d}{d+2\alpha'}}/\ln(T)^{\frac{2d}{d+2\alpha'}}$ and $\epsilon = n^{\frac{-\alpha'}{d}}$. Note that we can only use the result in Theorem 4 if the high probability upper confidence bound defined in line 4 of sub-procedure Algorithm 2 holds honestly. When the input parameter α' is larger than α, the calculated misspecification error ϵ is smaller than the true $\epsilon^* = \tilde{\mathcal{O}}(T^{\frac{-\alpha}{d+2\alpha}})$, which invalidates the confidence bound. Therefore, the regret bound does not hold for when $\alpha' > \alpha$. When the input parameter is smaller than α, we can simply use the fact that functions that are α-Hölder smooth are also α'-Hölder smooth: $H(\alpha, L) \subset H(\alpha', L)$. Therefore, the regret of the algorithm with input parameter $\alpha' \leq \alpha$ is bounded by $R(T) \leq \tilde{\mathcal{O}}(d(\alpha')(\sqrt{Tn} + \epsilon T)) = \tilde{\mathcal{O}}(d(\alpha')T^{\frac{d+\alpha'}{d+2\alpha'}})$. □

A.7 Proof of Theorem 8

Proof. There exists an $\hat{\alpha} \in \mathcal{G}$, s.t. $\hat{\alpha} \leq \alpha \leq \hat{\alpha} + \frac{R}{\log(T)}$, for any true α in $(0, R]$. There are two sources that made up the cost of adaptation when using Corral. The first one is the cost of searching over a grid for the unknown point $\hat{\alpha}$. The second one is the cost of approximation, specifically the difference between the rates achieved for $\hat{\alpha}$ and the true α. We will first derive the cost of grid search.

As specified in the proof of Theorem 5.3 in Pacchiano et al. (2020), the following bound of regret of the Corral algorithm holds with respect to any of its base-algorithm with high probability regret bound $U(T, \delta)$. The notations were introduced in Appendix section A.5.

$$R(T) \leq \mathcal{O}(\frac{M\ln(T)}{\eta} + T\eta) - \mathbb{E}[\frac{\rho}{40\eta\ln(T)} - 2\rho U(T/\rho, \delta)\log(T)] + \delta T + 8\sqrt{MT\log(\frac{4TM}{\delta})}. \tag{36}$$

Plugging the regret rate of base-algorithm in Lemma 7, the expected pseudo-regret of Corral with smooth wrapper is therefore bounded by:

$$\begin{aligned} R(T) &\overset{\hat{\alpha}\leq\alpha}{\leq} \tilde{\mathcal{O}}(\frac{M}{\eta} + T\eta + \sqrt{MT}) + \delta T - \mathbb{E}[\frac{\rho}{40\eta\ln(T)} - 2\rho(\tilde{\mathcal{O}}(d(\frac{T}{\rho})^{\frac{d+\hat{\alpha}}{d+2\hat{\alpha}}}))\log(T)] \\ &\overset{\delta=1/T}{=} \tilde{\mathcal{O}}(\frac{M}{\eta} + T\eta + \sqrt{MT}) - \mathbb{E}[\tilde{\mathcal{O}}(\frac{\rho}{\eta} - \rho d(\frac{T}{\rho})^{\frac{d+\hat{\alpha}}{d+2\hat{\alpha}}})] \\ &= \tilde{\mathcal{O}}(\frac{M}{\eta} + T\eta + \sqrt{MT}) - \mathbb{E}[\tilde{\mathcal{O}}(\frac{\rho}{\eta} - dT^{\frac{d+\hat{\alpha}}{d+2\hat{\alpha}}}\rho^{\frac{\hat{\alpha}}{d+2\hat{\alpha}}})]. \end{aligned} \tag{37}$$

Similarly, we first maximize over ρ by setting the derivative w.r.t ρ to zero by setting $\rho = \tilde{\mathcal{O}}(\eta^{\frac{d+2\hat{\alpha}}{d+\hat{\alpha}}} d^{\frac{d+2\hat{\alpha}}{d+\hat{\alpha}}} T)$. Then the above rate comes to

$$R(T) \leq \tilde{\mathcal{O}}(\frac{M}{\eta} + T\eta + \sqrt{MT} + d^{\frac{d+2\hat{\alpha}}{d+\hat{\alpha}}} T\eta^{\frac{\hat{\alpha}}{d+\hat{\alpha}}}). \tag{38}$$

However, since η is a parameter of the Corral algorithm which does not know $\hat{\alpha}$ or α, we will rely on the parameter R specified by the user. Let us set η with repsect to $\alpha = R$, i.e. $\eta = \tilde{\mathcal{O}}(d^{-1}T^{-\frac{d+R}{d+2R}})$, and plug in the number of grid points (base-algorithms) $M = \lceil\log(T)\rceil$.

$$\begin{aligned} &\tilde{\mathcal{O}}(\frac{M}{\eta} + T\eta + \sqrt{MT} + d^{\frac{d+2\hat{\alpha}}{d+\hat{\alpha}}} T\eta^{\frac{\hat{\alpha}}{d+\hat{\alpha}}}) \\ &= \tilde{\mathcal{O}}(dT^{\frac{d+R}{d+2R}} + d^{-1}T^{\frac{R}{d+2R}} + dT^{\frac{d^2+2Rd+R\hat{\alpha}}{(d+2R)(d+\hat{\alpha})}}) \\ &= \tilde{\mathcal{O}}(dT^{\frac{d+R}{d+2R}} + dT^{\frac{d^2+2Rd+R\hat{\alpha}}{(d+2R)(d+\hat{\alpha})}}). \end{aligned} \tag{39}$$

It is obvious that this rate is not the minimax optimal rate for class $\sum(\hat{\alpha})$, this gap shows the cost of grid search.

Next, let us consider the cost of approximation and how it is eliminated by using the linear grid (Hoffmann et al. (2011)). Namely, we show that adaptation for $\hat{\alpha}$ is equivalent to adaptation for α:

$$\tilde{\mathcal{O}}(dT^{\frac{d+R}{d+2R}} + dT^{\frac{d^2+2Rd+R\hat{\alpha}}{(d+2R)(d+\hat{\alpha})}}) = \tilde{\mathcal{O}}(dT^{\frac{d+R}{d+2R}} + dT^{\frac{d^2+2Rd+R\alpha}{(d+2R)(d+\alpha)}}). \tag{40}$$

The equality holds because $|\alpha - \hat{\alpha}| \leq \frac{R}{\log(T)}$. Let $J = \frac{d^2+2Rd+R\alpha}{(d+2R)(d+\hat{\alpha})}$ and $Q = \frac{d^2+2Rd+R\alpha}{(d+2R)(d+\alpha)}$, then $W \triangleq \frac{T^J}{T^Q} \leq T^{\frac{(d^2+2Rd+R\alpha)\frac{R}{\log(T)}}{(d+2R)(d+\alpha)(d+\hat{\alpha})}}$. Taking the log of W yields $\log(W) = R\frac{d^2+2Rd+R\alpha}{(d+2R)(d+\alpha)(d+\hat{\alpha})}$. Since both α and $\hat{\alpha}$ are bounded by a constant range $(0, 2]$, the term $\frac{d^2+2Rd+R\alpha}{(d+2R)(d+\alpha)(d+\hat{\alpha})} \leq C$ for some constant C, W is therefore $\mathcal{O}(1)$ as well.

Therefore, for functions with Hölder exponent $\alpha < R$, the second term in equation (40) is the dominant term and the expected regret rate is $\tilde{\mathcal{O}}(dT^{\frac{d^2+2Rd+R\alpha}{(d+2R)(d+\alpha)}})$. For functions with Hölder exponent $\alpha \geq R$, which essentially belongs to a subset of $\sum(R, L)$, they will all have the same rate which is $\tilde{\mathcal{O}}(dT^{\frac{d+R}{d+2R}})$. When $\alpha = R$, this matches the minimax rate for α. □

B Additional algorithms for the main document

B.1 Doubling procedure for Algorithm 2

Algorithm 3 Doubling procedure for Algorithm 2

Require: Meta-algorithm $\mathcal{A}^{global}$ (Algorithm 2), fail probability δ

1: **for** $i = 0 \ldots$ **do**
2: $T_i = 2^i$
3: Restart $\mathcal{A}^{global}$ with initialization parameters $n_i = \lfloor T_i^{\frac{d}{d+2\alpha}} / \ln(T_i)^{\frac{2d}{d+2\alpha}} \rceil$ and fail probability $\delta_i = 6\delta/\pi^2 i^2$
4: Run $\mathcal{A}^{global}$ for T_i steps.
5: **end for**

B.2 The Corral Master algorithm

For easier reference, we include the copy of Algorithm 7 in Pacchiano et al. (2020).

Algorithm 4 Corral Master (Algorithm 7 in Pacchiano et al. (2020))

Require: Base algorithms $\{\mathcal{B}_j\}_{j=1}^M$, learning rate η.

1: Initialize: $\gamma = 1/T, \beta = e^{\frac{1}{\ln(T)}}, \eta_{1,j} = \eta, \rho_1^j = 2M, \underline{p}_1^j = \frac{1}{\rho_1^j}, p_1^j = 1/M$ for all $j \in [M]$.
2: **for** $t = 1, \ldots, T$ **do**
3: Sample $i_t \sim p_t$.
4: Receive feedback r_t from base $\mathcal{B}_{i_t}$.
5: Update p_t, η_t and $\underline{p}_t$ to p_{t+1}, η_{t+1} and $\underline{p}_{t+1}$using r_t via Corral-Update (takes input η_t, p_t, β, lower bound $\underline{p}_t$ and current feedback r_t).
6: **for** j=1, …, M **do**
7: Set $\rho_{t+1}^j = \frac{1}{\underline{p}_{t+1}^j}$.
8: **end for**
9: **end for**

The corral update procedure is in Algorithm 5 and the smooth wrapper for the base-algorithms is in Algorithm 3 in Pacchiano et al. (2020).