A Short Note on the Relationship of Information Gain and Eluder Dimension

Kaixuan Huang [* 1]   **Sham M. Kakade** [* 2 3]   **Jason D. Lee** [* 1]   **Qi Lei** [* 1]

Abstract

Eluder dimension and information gain are two widely used methods of complexity measures in bandit and reinforcement learning. Eluder dimension was originally proposed as a general complexity measure of function classes, but the common examples of where it is known to be small are function spaces (vector spaces). In these cases, the primary tool to upper bound the eluder dimension is the elliptic potential lemma. Interestingly, the elliptic potential lemma also features prominently in the analysis of linear bandits/reinforcement learning and their nonparametric generalization, the information gain. We show that this is not a coincidence – eluder dimension and information gain are equivalent in a precise sense for reproducing kernel Hilbert spaces.

1. Introduction

Eluder dimension is first proposed by Russo & Van Roy (2013) to analyze the regret upper bounds of *upper confidence bound* (UCB) algorithms and *Thompson sampling* (TS) algorithms. This notion captures the worst-case sample complexity that is required to infer the values of unobserved points using the observed samples, which measures the degree of dependence among a function class. The regret of UCB algorithms and TS algorithms can be upper bounded by $\widetilde{\mathcal{O}}(\sqrt{\dim_{\mathrm{E}} \cdot \log N \cdot T})$, where $\dim_{\mathrm{E}}$ and N are the eluder dimension and the covering number of $\mathcal{F}$, respectively. Eluder dimensions are also used to establish regret bounds for learning contextual bandits and MDPs with function approximation (See reference therein (Osband & Van Roy, 2014; Du et al., 2020; Wang et al., 2020; Foster et al., 2020)).

Prior literature further investigated some necessary condi-

---
[*]Equal contribution   [1]Princeton University   [2]University of Washington   [3]Microsoft Research.   Correspondence to: Kaixuan Huang <kaixuanh@princeton.edu>, Sham M. Kakade <sham@cs.washington.edu>, Jason D. Lee <jasonlee@princeton.edu>, Qi Lei <qilei@princeton.edu>.

tions for a small or finite eluder dimension. In the original work, Russo & Van Roy (2013); Osband & Van Roy (2014) provided upper bounds of the eluder dimension for (generalized) linear models with bounded norm and finite dimension and for function classes with bounded domain size. Specifically, all these results have an explicit dependence on the dimension d of the input space or size of the domain. For other function classes, the understanding of the eluder dimension is limited, and thus bounded eluder dimension is considered as a strong assumption (Foster & Rakhlin, 2020). Recently, Li et al. (2021) analyzed the eluder dimension through the lens of *generalized-rank*, and calculated the eluder dimension of several interesting function classes including ReLU networks.

In these cases, the primary tool to upper bound the eluder dimension is the elliptic potential lemma (e.g. (Russo & Van Roy, 2013; Dani et al., 2008; Du et al., 2021)). Interestingly, the elliptic potential lemma also features prominently in the analysis of linear bandits/reinforcement learning and their nonparametric generalization, the *information gain*, which guided another line of studies on the complexity measure for bandit or RL (Russo & Van Roy, 2016; Srinivas et al., 2009):

Information gain is defined as the mutual information between the prior distribution and the noisy observations, which characterizes the reduction of the uncertainty of the underlying function after observing noisy measurements. It has an analytic formula when the prior and the noise distributions are Gaussian. Srinivas et al. (2009) analyzed the GP-UCB algorithm, a Bayesian variant of UCB algorithms for *Gaussian Process* (GP) bandits, and provided a dimension-free regret bounds in terms of maximum information gain. Specifically, they established the regret bounds of $\widetilde{\mathcal{O}}(\sqrt{T}(B\sqrt{\gamma_T} + \gamma_T))$, where γ_T is the maximum information gain for T samples and B is the upper bound of the RKHS norm of the ground-truth reward function. Afterwards, several studies improved the dependency on γ_T for the regret bounds for GP-UCB algorithms (Chowdhury & Gopalan, 2017) and extended to GP-TS algorithms (Russo & Van Roy, 2016). Du et al. (2021) provided regret bounds for learning bilinear MDPs in terms of *critical information gain*.

To achieve sublinear regrets, one requires the maximum

information gain to grow mildly in the number of samples. For finite-dimensional domains, the maximum information gain only grows logarithmically (Srinivas et al., 2009; Du et al., 2021). For infinite-dimensional RKHSs, the growth of the maximum information gain depends on the eigendecay of the reproducing kernels, and specific rates for different kernels can be established (Srinivas et al., 2009; Vakili et al., 2021).

With the existing two lines of work, some natural questions arise:

Is one complexity measure strictly tighter than the other? Or namely, what is the relationship between using the two complexity measures?

Contribution. Therefore in this work, we study the relationship between eluder dimension and information gain, aiming at bridging the two series of work. Our work shows that when the function class is a ball of an RKHS, the critial information gain and the eluder dimension of an action set are equivalent. (Jin et al., 2021) showed concurrently one direction of the equivalence that low critical information gain implies low eluder dimension.

2. Preliminaries

Notation. Throughout the paper, we will use the following notation:

- $\mathcal{X}$: the action set.

- $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$: the function class.

- $\mathcal{H}$: the *reproducing kernel Hilbert space* (RKHS). Its formal definition is deferred in the next paragraph.

- $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$: the corresponding kernel function [1].

- $\mathcal{B}(S)$: the ball of radius S in $\mathcal{H}$, i.e.,

$$\mathcal{B}(S) = \{f \mid f(x) = \langle \theta_f, x \rangle, \|\theta_f\| \leq S\}.$$

For an $x \in \mathcal{X}$, we use $x^\top$ to represent the adjoint (transpose) of $K(\cdot, x) \in \mathcal{H}$. Therefore, $xx^\top$ is a self-adjoint linear operator from $\mathcal{H}$ to itself. We now formally introduce the key concepts used in this paper.

Reproducing Kernel Hilbert Space.

Definition 2.1 (Reproducing Kernel Hilbert Space)**.** *For any set $\mathcal{X}$, $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ is said to be a reproducing kernel Hilbert space (RKHS) with respect to the kernel function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, if $\mathcal{H}$ is a Hilbert space equipped with the*

---

[1] We do not distinguish between $x \in \mathcal{X}$ and $K(\cdot, x) \in \mathcal{H}$.

inner product $\langle \cdot, \cdot \rangle$. Furthermore, for any $x \in \mathcal{X}$ we have $K(\cdot, x) \in \mathcal{H}$ and

$$f(x) = \langle f, K(\cdot, x) \rangle, \;\; for\; all\; f \in \mathcal{H}.$$

Recent years have seen a tremendous studies on nonlinear reinforcement learning via eluder dimension and kernel methods, due to their potential for expressive function approximation. We are interested in analyzing the connection between eluder dimension and information gain when the function class falls into some RKHS. We then introduce the formal definition of eluder dimension and information gain respectively.

Eluder Dimension.

Definition 2.2 (ϵ-Dependence)**.** *An element $x \in \mathcal{X}$ is ϵ-dependent on a subset $\mathcal{S} = \{x_1, \ldots x_n\} \subseteq \mathcal{X}$ with respect to the function class $\mathcal{F}$, if any pair of functions $f, \tilde{f} \in \mathcal{F}$ satisfying $\sum_{i=1}^n (f(x_i) - \tilde{f}(x_i))^2 \leq \epsilon^2$ also satisfies $f(x) - \tilde{f}(x) \leq \epsilon$. We say x is ϵ-independent of $\mathcal{S}$ with respect to $\mathcal{F}$, if x is not ϵ-dependent on $\mathcal{S}$.*

The definition of ϵ-dependence generalizes linear dependence (which corresponds to $\epsilon = 0$). In particular, if $x \in \mathcal{X}$ is ϵ-dependent on the dataset $\mathcal{S}$, then knowing the values of $f \in \mathcal{F}$ on $\mathcal{S}$ allows for a good estimate of $f(x)$ up to ϵ-accuracy.

Definition 2.3 (Eluder Dimension)**.** *The ϵ-eluder dimension $\dim_{\mathrm{E}}(\mathcal{F}, \mathcal{X}, \epsilon)$ is the length d of the longest sequence of elements in $\mathcal{X}$ such that, for some $\epsilon' \geq \epsilon$, every element x_i is ϵ'-independent of its predecessors $(x_1, \ldots, x_{i-1})$ with respect to to $\mathcal{F}$.*

Notice that the eluder dimension is defined in a sequential manner. The eluder dimension of a set $\mathcal{X}$ with respect to a function class $\mathcal{F}$ corresponds to the number of samples needed to adaptively infer information of $\mathcal{F}$ among $\mathcal{X}$ in the worst case.

Information Gain.

Definition 2.4 (Information Gain)**.** *Assume that $\mathcal{X}$ is associated with a kernel K and a corresponding RKHS $\mathcal{H}$. For any $\lambda > 0$, the information gain after observing $x_1, \ldots, x_T \in \mathcal{X}$ is defined as*

$$\gamma(\lambda; x_1, \ldots, x_T) = \log \det \left(I + \frac{1}{\lambda} \sum_{i=1}^T x_i x_i^\top \right).$$

Remark 2.1. *One can show that the information gain can be equivalently written as*

$$\gamma(\lambda; x_1, \ldots, x_T) = \log \det \left(I + \frac{1}{\lambda} K_T \right),$$

where $(K_T)_{i,j} = \langle x_i, x_j \rangle = K(x_i, x_j)$ is the Gram matrix.

Definition 2.5 (Maximum Information Gain). *For any $\lambda > 0$, $T \geq 1$, the maximum information gain for T observations is defined as*

$$\gamma_T(\lambda; \mathcal{X}) = \max_{x_i \in \mathcal{X}, 1 \leq i \leq T} \gamma(\lambda; x_1, \ldots, x_T).$$

The maximum information gain is analogous to the log of the maximum volume of the ellipsoid generated by T points in $\mathcal{X}$, which captures the geometric structure of $\mathcal{X}$.

Definition 2.6 (Critical Information Gain). *For a fixed constant $c > 0$, the critical information gain is defined as:*

$$\tilde{\gamma}(\lambda, c; \mathcal{X}) = \min\{k \mid \gamma_k(\lambda; \mathcal{X}) \leq ck\}.$$

The critical information gain is the minimal T that $\gamma_T(\lambda; \mathcal{X})$ fails to grow linearly, which can be viewed as the effective dimension of $\mathcal{X}$ (Du et al., 2021).

3. Main Results

With the formal definition of information gain as well as eluder dimension, we now present our main result on connecting them.

Suppose the function class $\mathcal{F}$ is in the RKHS $\mathcal{H}$ over the action set $\mathcal{X}$. We prove that under some mild assumptions, the critical information gain and eluder dimension are equivalent up to constant factors. We drop the dependence on $\mathcal{X}$ and use $\dim_{\mathrm{E}}(\mathcal{F}, \epsilon)$ and $\tilde{\gamma}(\lambda, c)$ to represent the eluder dimension and the critical information gain, respectively.

Theorem 3.1. *Assume that $\mathcal{F} \subseteq \mathcal{B}(S)$. Set $\lambda = \left(\frac{\epsilon}{2S}\right)^2$, we have*

$$\dim_{\mathrm{E}}(\mathcal{F}, \epsilon) < \tilde{\gamma}\left(\lambda, \log \frac{3}{2}\right).$$

Theorem 3.2. *Assume that $\|x\|_2 \leq B$ for all $x \in \mathcal{X}$ and $\mathcal{F} \supseteq \mathcal{B}(S)$. Then for any $c > \log 2$, $\epsilon < 2BS$, by setting $\lambda = \left(\frac{\epsilon}{2S}\right)^2$, we have*

$$\dim_{\mathrm{E}}(\mathcal{F}, \epsilon) > (\tilde{\gamma}(\lambda, c) - 1) \cdot \frac{c - \log 2}{\log(1 + \frac{B^2}{\lambda}) - \log 2}.$$

Together we have that, when $\mathcal{F}$ is the ball of radius S in the RKHS $\mathcal{H}$ over the action set $\mathcal{X}$, its eluder dimension and critical information gain with proper λ are equivalent (up to constant multiplicative factors).

We defer the complete proof to the appendix and provide the proof sketch as follows.

3.1. Proof sketchs

First, we introduce the notion of *increment of information gain* and show that it is closely related to ϵ-dependence.

Definition 3.3 (Increment of Information Gain). *For any $\lambda > 0$, the increment of information gain of $x \in \mathcal{X}$ on $\mathcal{S} = \{x_1, \ldots, x_n\} \subseteq \mathcal{X}$ is defined as*

$$\Delta\gamma(\lambda; x|\mathcal{S}) = \gamma(\lambda; x, x_1, \ldots, x_n) - \gamma(\lambda; x_1, \ldots, x_n).$$

The next two lemmas characterize the relationships between large increment of information gain and ϵ-independence. Intuitively, if x is ϵ-independent of $\mathcal{S}$, then the increment of information gain of x on $\mathcal{S}$ will be large and vice versa.

Lemma 3.4. *Assume that $\mathcal{F} \subseteq \mathcal{B}(S)$. If x is ϵ'-independent of $\mathcal{S}$ with respect to $\mathcal{F}$ for some $\epsilon' \geq \epsilon$, then*

$$\Delta\gamma((\epsilon/2S)^2; x|\mathcal{S}) > \log \frac{3}{2}.$$

Lemma 3.5. *Assume that $\mathcal{F} \supseteq \mathcal{B}(S)$. If*

$$\Delta\gamma((\epsilon/2S)^2; x|\mathcal{S}) > \log 2.$$

then x is ϵ-independent of $\mathcal{S}$ with respect to $\mathcal{F}$.

By Lemma 3.4, a sequence whose elements are ϵ-independent of their predecessors implies a sequence with large increments of information gain. Therefore we can lower bound the critical information gain by eluder dimension. Next we provide the proof of Theorem 3.1.

proof of Theorem 3.1. Assume that $\dim_{\mathrm{E}}(\mathcal{F}, \epsilon) = d$ and $\{x_1, \ldots, x_d\}$ is the longest sequence such that there exists an $\epsilon' \geq \epsilon$, for each i, x_i is ϵ'-independent of $\{x_1, \ldots, x_{i-1}\}$. Then by Lemma 3.4, we have for $\lambda = (\epsilon/2S)^2$,

$$\Delta\gamma(\lambda; x_i|\{x_1, \ldots, x_{i-1}\}) > \log \frac{3}{2}.$$

This means for $k = 1, \ldots, d$,

$$\gamma_k(\lambda; \mathcal{X}) \geq \gamma(\lambda; x_1, \ldots, x_k) > k \log \frac{3}{2}.$$

Therefore $\dim_{\mathrm{E}}(\mathcal{F}, \epsilon) < \tilde{\gamma}(\lambda, \log \frac{3}{2}; \mathcal{X})$, which is $\min\{k \mid k \log \frac{3}{2} \geq \gamma_k(\lambda; \mathcal{X})\}$. $\square$

Next lemma shows that the increment of information gain can be arranged to be monotonically decreasing, which stems from the submodularity of the log-determinant function.

Lemma 3.6. *For any $\{x_1, \ldots, x_T\}$ and any $\lambda > 0$, we can arrange the order of the T points, such that for $i = 1, \ldots, T - 1$,*

$$\Delta\gamma(\lambda, x_i|x_1, \ldots, x_{i-1}) \geq \Delta\gamma(\lambda, x_{i+1}|x_1, \ldots, x_i).$$

By Lemma 3.6, the first few elements of the maximizers of $\gamma_T(\lambda; \mathcal{X})$ will have large increments of information gain, which, by Lemma 3.5, means that each element is ϵ-independent of its predecessors. Therefore we can lower bound the eluder dimension by critical information gain. Next we provide the proof of Theorem 3.2.

proof of Theorem 3.2. Let $d = \tilde{\gamma}(\lambda, c) - 1$. Then $\gamma_d(\lambda) > cd$. Then we can find $\{x_1, \ldots, x_d\} \subseteq \mathcal{X}$ such that $\gamma(\lambda; x_1, \ldots, x_d) > cd$. Denote $\Delta_i = \Delta\gamma(\lambda; x_i | x_1, \ldots, x_{i-1})$. By Lemma 3.6, we can arrange those x_i's such that Δ_i's are monotonically non-increasing. Let $k = \max\{i \mid \Delta_i > \log 2\}$, then by Lemma 3.5, we know that $\{x_1, \ldots, x_k\}$ statisfies the definition in the eluder dimension, so $\dim_{\mathrm{E}}(\mathcal{F}, \epsilon) \geq k$.

Since $\|x_i\| \leq B$ for all i, we have

$$\gamma_k(\lambda) \leq k \log(1 + \frac{B^2}{\lambda}).$$

Therefore we get

$$cd < \gamma(\lambda; x_1, \ldots, x_d)$$
$$= \sum_{i=1}^{k} \Delta_i + \sum_{i=k+1}^{d} \Delta_i$$
$$\leq \gamma_k(\lambda) + (d - k) \log 2$$
$$\leq k \log(1 + \frac{B^2}{\lambda}) + (d - k) \log 2.$$

By the assumption that $\epsilon < 2BS$, we have $B^2/\lambda = (2BS)^2/\epsilon^2 > 1$, then we have

$$k > d \cdot \frac{c - \log 2}{\log(1 + \frac{B^2}{\lambda}) - \log 2}.$$

This finishes the proof. $\qquad\square$

4. Case Study: Comparison of Two Regret Bounds

In this section, we briefly discuss two existing results on learning episodic MDPs under the completeness assumption with respect to an RKHS (Yang et al., 2020; Wang et al., 2020). These results base on different techniques and provide the regret guarantees in terms of maximum information gain and eluder dimension, respectively. Using our result on the equivalence of the eluder dimension and the critical information gain, we characterize how the eluder dimension grows in T when the function class is an RKHS-norm ball, and provide a comparison of the two regret bounds in this setting.

Setting. We consider learning an episodic MDP $(\mathcal{S}, \mathcal{A}, P, H, r, \mu)$ with function approximation. In RL with function approximation, the input domain of the function class $\mathcal{F}$ is $\mathcal{X} = \mathcal{S} \times \mathcal{A}$, containing all the state-action pairs $x = (s, a)$. We assume that there exists a kernel $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ and a corresponding RKHS $\mathcal{H}$. Therefore, every $x = (s, a)$ can be viewed as an element in $\mathcal{H}$. We assume that the function class $\mathcal{F} = \mathcal{B}(R_Q H)$ is the RKHS

norm ball with radius $R_Q H$. We also assume that $\|x\| \leq 1$ for all $x = (s, a) \in \mathcal{S} \times \mathcal{A}$. We make the following completeness assumption on the function class $\mathcal{F}$ as in (Yang et al., 2020; Wang et al., 2020).

Assumption 4.1 (Completeness). *For any $h \in [H]$ and any Q function $Q : \mathcal{S} \times \mathcal{A} \rightarrow [0, H]$, we have $\mathbb{T}_h(Q) \in \mathcal{F}$, where $\mathbb{T}_h$ is the Bellman operator at level h, defined as*

$$\mathbb{T}_h(Q)(s, a) = r_h(s, a) + \sum_{s' \in \mathcal{S}} P_h(s'|s, a) \max_{a' \in \mathcal{A}} Q(s', a')$$
$$\forall (s, a) \in \mathcal{S} \times \mathcal{A}.$$

First we state the two results from (Yang et al., 2020; Wang et al., 2020). Specifically, Theorem 4.1 provides the regret bound for KOVI algorithm in terms of the maximum information gain, while Theorem 4.2 provides the regret bound for $\mathcal{F}$-LSVI algorithm in terms of the eluder dimension.

Theorem 4.1 (Yang et al. (2020)). *Under Assumption 4.1, for KOVI, after interacting with the environment for T episodes ($E = TH$ steps), with probability $1 - E^{-2}$,*

$$\mathrm{Regret}(T) = \tilde{\mathcal{O}}\Big(H^2 \cdot \Big[\gamma_T(\lambda; \mathcal{X}) +$$
$$\max_{h \in [H]} \sqrt{\gamma_T(\lambda; \mathcal{X}) \cdot \log N_\infty^{(ucb)}(\epsilon^*, h, B_T)} \Big] \cdot \sqrt{T} \Big),$$

where $\lambda = 1 + \frac{1}{T}$, $\epsilon^* = H/T$.

Theorem 4.2 (Wang et al. (2020)). *Under Assumption 4.1, after interacting with the environment for $E = TH$ steps, with probability $1 - \delta$, $\mathcal{F}$-LSVI achieves a a regret bound of*

$$\mathrm{Regret}(T) \leq \sqrt{\iota H^2 E},$$

where

$$\iota \leq C \cdot \log^2(E/\delta) \cdot \dim_{\mathrm{E}}^2 (\mathcal{F}, \delta/E^3)$$
$$\cdot \log\left(N_\infty\left(\mathcal{F}, \delta/E^2\right)/\delta\right) \cdot \log(N_\infty(\mathcal{S} \times \mathcal{A}, \delta/E) \cdot E/\delta),$$

for some constant $C > 0$.

By our Theorems 3.1 and 3.2, the regret bound in Theorem 4.2 can be equivalently stated in terms of the critical information gain. In particular, if we set $\delta = E^{-2}$ and notice that $\mathcal{F} = \mathcal{B}(R_Q H)$, then we have

$$\tilde{\Omega}\Big(\tilde{\gamma}((4T^{10}H^{12}R_Q^2)^{-1}, c) \Big) \leq \dim_{\mathrm{E}}(\mathcal{F}, \delta/E^3)$$
$$\leq \tilde{\gamma}\Big((4T^{10}H^{12}R_Q^2)^{-1}, \log\frac{3}{2} \Big),$$

where $\tilde{\Omega}$ hides the constant factors and the log factors. Therefore we have the following corollary.

Corollary 4.3 (Wang et al. (2020)). *Under Assumption 4.1, after interacting with the environment for $E = TH$ steps,*

with probability $1 - E^{-2}$, $\mathcal{F}$-LSVI achieves a a regret bound of

$$\text{Regret}(T) \leq \sqrt{\iota H^2 E},$$

where

$$\iota \leq C \cdot \log^2(E^3) \cdot \tilde{\gamma}^2\left((4T^{10}H^{12}R_Q^2)^{-1}, \log\frac{3}{2}\right)$$
$$\cdot \log\left(N_\infty\left(\mathcal{F}, 1/E^4\right)E^2\right) \cdot \log(N_\infty(\mathcal{S} \times \mathcal{A}, 1/E^3) \cdot E^3),$$

for some constant $C > 0$.

Next we focus on the dependency of these regret bounds on T, ignoring all the log factors as well as all the other problem-specific constants, such as H and R_Q. Both Corollary 4.3 and Theorem 4.1 have an explicit $\sqrt{T}$ factor and a term characterizing how the information gain grows. To continue our comparison, we make the additional assumption on the eigenvalues of the kernel operator as in (Srinivas et al., 2009; Yang et al., 2020).

Assumption 4.2 (Polynomial decay). *Assume that the state-action set $\mathcal{X} = \mathcal{S} \times \mathcal{A}$ is a compact subset of $\mathbb{R}^d$. Furthermore, assume that the kernel function K admits the following orthonormal decomposition*

$$K(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(x'),$$

with $\|\phi_i(x)\|_\infty \leq C_\phi$ and the eigenvalues λ_i's of L_K satisfying β-polynomial decay: $\lambda_i \leq C_p i^{-\beta}$, for some $\beta > 2 + 1/d$.

Under Assumption 4.2, the regret bound of Theorem 4.1 can be reduced to the following when $d \geq \Omega(1), \beta \geq \Omega(1)$; see (Yang et al., 2020) for more details.

$$\text{Regret}(T) \leq \tilde{\mathcal{O}}(T^{\frac{d+1}{d+\beta}+1/2}). \tag{1}$$

Next we calculate the dependency on T of Corollary 4.3. We introduce the following lemma from the appendix of (Yang et al., 2020) (Lemmas D.2 and D.6).

Lemma 4.4 ((Yang et al., 2020)). *Under Assumption 4.2, the maximum information gain can be bounded as*

$$\gamma_\lambda(T) \leq \tilde{\mathcal{O}}(\lambda^{-1/\beta} T^{(d+1)/(\beta+d)}).$$

Therefore, the critical information gain satisfies

$$\tilde{\gamma}(\lambda, c) = \tilde{\mathcal{O}}((1/\lambda)^{(\beta+d)/[\beta(\beta-1)]}).$$

Furthermore, the $\log$ of the L^∞ covering number of the RKHS ball of radius S can be bounded as

$$\log N_\infty\left(\mathcal{B}(S), 1/\epsilon\right) \leq \tilde{\mathcal{O}}((S/\epsilon)^{2/(\beta-1)}).$$

By Lemma 4.4, the regret bound in Corollary 4.3 (Wang et al., 2020) is

$$\text{Regret}(T) \leq \tilde{\mathcal{O}}(T^{\frac{10(\beta+d)+7\beta}{\beta(\beta-1)}+1/2}). \tag{2}$$

By comparing Equation (1) and Equation (2), we conclude that when the kernel K satisfies polynomial decay (Assumption 4.2), Equation (1) is better than Equation (2) as long as $\beta \gg d$.

5. Conclusion

In this note, we clarified the relationship between two complexity measures used in bandits and reinforcement learning, the eluder dimension and information gain in RKHS function spaces. Eluder dimension is a general complexity measure defined for any function class; however, information gain as defined in (Srinivas et al., 2009) is only defined for RKHS. It is straightforward to extend the information gain in the Bayesian setting when there is a prior over the function class/prior. We conjecture that for function classes that have a natural vector space structure (i.e. convex subset with non-empty interior of a vector space), that the eluder dimension does not yield sublinear regret, unless information gain is sublinear in the sample size n. To our knowledge, all commonly encountered function classes with vector space structure do not have bounded eluder dimension, unless they are subsets of an RKHS. We leave to future work to identify complexity measures for sequential decision making that are more structure adaptive, such as sparsity, low-rank, or other measures of low-dimensionality, than eluder dimension and information gain which essentially only capture the extrinsic vector space dimensionality.

References

Chowdhury, S. R. and Gopalan, A. On kernelized multi-armed bandits. In *International Conference on Machine Learning*, pp. 844–853. PMLR, 2017.

Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic linear optimization under bandit feedback. 2008.

Du, S. S., Lee, J. D., Mahajan, G., and Wang, R. Agnostic Q-learning with function approximation in deterministic systems: Tight bounds on approximation error and sample complexity. *Advances in Neural Information Processing Systems*, 2020.

Du, S. S., Kakade, S. M., Lee, J. D., Lovett, S., Mahajan, G., Sun, W., and Wang, R. Bilinear classes: A structural framework for provable generalization in rl. *arXiv preprint arXiv:2103.10897*, 2021.

Foster, D. and Rakhlin, A. Beyond ucb: Optimal and efficient contextual bandits with regression oracles. In

International Conference on Machine Learning, pp. 3199–3210. PMLR, 2020.

Foster, D. J., Rakhlin, A., Simchi-Levi, D., and Xu, Y. Instance-dependent complexity of contextual bandits and reinforcement learning: A disagreement-based perspective. *arXiv preprint arXiv:2010.03104*, 2020.

Jin, C., Liu, Q., and Miryoosefi, S. Bellman eluder dimension: New rich classes of rl problems, and sample-efficient algorithms, 2021.

Li, G., Kamath, P., Foster, D. J., and Srebro, N. Eluder dimension and generalized rank. *arXiv preprint arXiv:2104.06970*, 2021.

Osband, I. and Van Roy, B. Model-based reinforcement learning and the Eluder dimension. In *Advances in Neural Information Processing Systems*, pp. 1466–1474, 2014.

Russo, D. and Van Roy, B. Eluder dimension and the sample complexity of optimistic exploration. In *NIPS*, pp. 2256–2264. Citeseer, 2013.

Russo, D. and Van Roy, B. An information-theoretic analysis of thompson sampling. *The Journal of Machine Learning Research*, 17(1):2442–2471, 2016.

Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.

Vakili, S., Khezeli, K., and Picheny, V. On information gain and regret bounds in gaussian process bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 82–90. PMLR, 2021.

Wang, R., Salakhutdinov, R. R., and Yang, L. Reinforcement learning with general value function approximation: Provably efficient approach via bounded eluder dimension. *Advances in Neural Information Processing Systems*, 33, 2020.

Yang, Z., Jin, C., Wang, Z., Wang, M., and Jordan, M. I. On function approximation in reinforcement learning: Optimism in the face of large state spaces. *arXiv preprint arXiv:2011.04622*, 2020.

A. Missing proofs of the lemmas

Additional notation. For any positive definite matrix A, we define $\|x\|_A^2 = x^\top A x$.

We first introduce the next lemma.

Lemma A.1. *Assume that $x \in \mathcal{X}$ and $\mathcal{S} = \{x_1, \ldots, x_n\} \subseteq \mathcal{X}$. For any $\lambda > 0$, denote $V = \lambda I + \sum_{i=1}^n x_i x_i^\top$, then we have the following holds*

$$\Delta\gamma(\lambda; x|\mathcal{S}) = \log(1 + \|x\|_{V^{-1}}^2).$$

Proof. By definition we have

$$\Delta\gamma(\lambda; x|\mathcal{S}) = \gamma(\lambda; x, x_1, \ldots, x_n) - \gamma(\lambda; x_1, \ldots, x_n) = \log\det\left(\frac{1}{\lambda}(V + xx^\top)\right) - \log\det\left(\frac{1}{\lambda}V\right).$$

By the Matrix determinant lemma, we have

$$\begin{aligned}
\det\left(\frac{1}{\lambda}(V + xx^\top)\right) &= \det\left[\frac{1}{\lambda}V \cdot (I + V^{-1}xx^\top)\right] \\
&= \det\left(\frac{1}{\lambda}V\right) \cdot \det(I + V^{-1}xx^\top)] \\
&= \det\left(\frac{1}{\lambda}V\right) \cdot \det(I + x^\top V^{-1}x)] \\
&= \det\left(\frac{1}{\lambda}V\right) \cdot (1 + \|x\|_{V^{-1}}^2).
\end{aligned}$$

$\square$

Proofs of the Lemmas

proof of Lemma 3.4. Let $\mathcal{S} = \{x_1, \ldots, x_n\}$, $\lambda = (\epsilon/2S)^2$, and $V = \lambda I + \sum_{i=1}^n x_i x_i^\top$. By definition, if x is ϵ'-independent of $x_1, \ldots, x_n$ for some $\epsilon' \geq \epsilon$, then there exist $\theta_1 \in \mathcal{H}$ and $\theta_2 \in \mathcal{H}$ such that $f_i(x) := \langle\theta_i, x\rangle \in \mathcal{F}$ statisfies

$$\langle\theta_1, x\rangle - \langle\theta_2, x\rangle > \epsilon', \tag{3}$$

and

$$\|\theta_1\| \leq S, \ \|\theta_2\| \leq S, \ \sum_{i=1}^n (\langle\theta_1, x_i\rangle - \langle\theta_2, x_i\rangle)^2 \leq (\epsilon')^2. \tag{4}$$

Then we have

$$\begin{aligned}
\|\theta_1 - \theta_2\|_V^2 &= (\theta_1 - \theta_2)^\top\left(\lambda I + \sum_{i=1}^n x_i x_i^\top\right)(\theta_1 - \theta_2) \\
&= (\theta_1 - \theta_2)^\top\left(\frac{\epsilon^2}{(2S)^2}I + \sum_{i=1}^n x_i x_i^\top\right)(\theta_1 - \theta_2) \\
&= \frac{\epsilon^2}{(2S)^2}\|\theta_1 - \theta_2\|^2 + \sum_{i=1}^n (\theta_1 - \theta_2)^\top x_i x_i^\top (\theta_1 - \theta_2) \\
&\leq \epsilon^2 + (\epsilon')^2,
\end{aligned}$$

where the last inequality follows from Equation (4).

By Equation (3), we have

$$\begin{aligned}
\epsilon' &< \langle\theta_1 - \theta_2, x\rangle \\
&\leq \|\theta_1 - \theta_2\|_V \cdot \|x\|_{V^{-1}} \qquad\qquad \text{(Cauchy Inequality)} \\
&\leq \sqrt{\epsilon^2 + (\epsilon')^2} \cdot \|x\|_{V^{-1}}.
\end{aligned}$$

Therefore,

$$\|x\|_{V^{-1}}^2 > \frac{(\epsilon')^2}{\epsilon^2 + (\epsilon')^2} \geq \frac{1}{2}.$$

Finally, by Lemma A.1, we have

$$\Delta\gamma((\epsilon/2S)^2; x|\mathcal{S}) = \log(1 + \|x\|_{V^{-1}}^2) > \log\frac{3}{2}.$$

$\square$

proof of Lemma 3.5. Let $\mathcal{S} = \{x_1, \ldots, x_n\}$, $\lambda = (\epsilon/2S)^2$, and $V = \lambda I + \sum_{i=1}^{n} x_i x_i^\top$. By Lemma A.1, we know that $\Delta\gamma((\epsilon/2S)^2; x|\mathcal{S}) > \log 2$ implies $\|x\|_{V^{-1}} > 1$.

Define

$$\theta_1 = \frac{\epsilon}{2} \cdot \frac{V^{-1}x}{\sqrt{x^\top V^{-1}x}}$$

and $\theta_2 = -\theta_1$. In this case, we have

$$
\begin{aligned}
\|\theta_1\|^2 &= \frac{\epsilon^2}{4} \cdot \frac{(V^{-1/2}x)^\top V^{-1}(V^{-1/2}x)}{x^\top V^{-1}x} \\
&= \frac{\epsilon^2}{4} \cdot \frac{(V^{-1/2}x)^\top \left(\lambda I + \sum_{i=1}^{n} x_i x_i^\top\right)^{-1}(V^{-1/2}x)}{x^\top V^{-1}x} \\
&\leq \frac{\epsilon^2}{4} \cdot \frac{(V^{-1/2}x)^\top \left(\lambda I\right)^{-1}(V^{-1/2}x)}{x^\top V^{-1}x} \\
&= \frac{\epsilon^2}{4} \cdot \frac{(2S)^2}{\epsilon^2} \cdot \frac{(V^{-1/2}x)^\top(V^{-1/2}x)}{x^\top V^{-1}x} \\
&= S^2.
\end{aligned}
$$

Similarly we have $\|\theta_2\|^2 \leq S^2$. Therefore $f_i(x) := \langle \theta_i, x \rangle \in \mathcal{F}$ for $i = 1, 2$.

Next, we have

$$
\begin{aligned}
\sum_{i=1}^{n} (\langle \theta_1 - \theta_2, x_i \rangle)^2 &= \epsilon^2 \cdot \frac{(V^{-1}x)^\top \left(\sum_{i=1}^{n} x_i x_i^\top\right)(V^{-1}x)}{x^\top V^{-1}x} \\
&\leq \epsilon^2 \cdot \frac{(V^{-1}x)^\top \cdot V \cdot (V^{-1}x)}{x^\top V^{-1}x} \\
&= \epsilon^2,
\end{aligned}
$$

and

$$\langle \theta_1 - \theta_2, x \rangle = \epsilon\|x\|_{V^{-1}} > \epsilon.$$

Then we conclude that x is ϵ-independent of $\mathcal{S} = \{x_1, \ldots, x_n\}$.

$\square$

proof of Lemma 3.6. Denote $\Sigma_k = \sum_{i=1}^{k} x_i x_i^\top$. The re-arrangement is done by sequentially choosing x_i to be

$$x_1 = \underset{x_j : 1 \leq j \leq T}{\arg\max} \log \det(I + \frac{1}{\lambda}(x_j x_j^\top)),$$

$$x_2 = \underset{x_j : 2 \leq j \leq T}{\arg\max} \log \det(I + \frac{1}{\lambda}(\Sigma_1 + x_j x_j^\top)),$$

$$\cdots$$

$$x_i = \underset{x_j : i \leq j \leq T}{\arg\max} \log \det(I + \frac{1}{\lambda}(\Sigma_{i-1} + x_j x_j^\top)),$$

$$\cdots$$

where ties are broken arbitrarily. By our procedure of choosing x_i, we have

$$\log\det(I + \frac{1}{\lambda}(\Sigma_{i-1} + x_i x_i^\top)) \geq \log\det(I + \frac{1}{\lambda}(\Sigma_{i-1} + x_{i+1} x_{i+1}^\top)). \tag{5}$$

Since $\Sigma_i - \Sigma_{i-1}$ is positive semidefinite, we have

$$\begin{aligned}
&\log\det(I + \frac{1}{\lambda}(\Sigma_{i-1} + x_{i+1} x_{i+1}^\top)) - \log\det(I + \frac{1}{\lambda}\Sigma_{i-1})\\
=&\log(1 + \|x_{i+1}\|^2_{(\lambda I + \Sigma_{i-1})^{-1}})\\
\geq&\log(1 + \|x_{i+1}\|^2_{(\lambda I + \Sigma_i)^{-1}})\\
=&\log\det(I + \frac{1}{\lambda}\Sigma_{i+1}) - \log\det(I + \frac{1}{\lambda}\Sigma_i).
\end{aligned} \tag{6}$$

Then by Equations (5) and (6), for $i = 1,\ldots,T-1$

$$\begin{aligned}
&\log\det(I + \frac{1}{\lambda}\Sigma_i) - \log\det(I + \frac{1}{\lambda}\Sigma_{i-1})\\
=&\log\det(I + \frac{1}{\lambda}(\Sigma_{i-1} + x_i x_i^\top)) - \log\det(I + \frac{1}{\lambda}\Sigma_{i-1})\\
\geq&\log\det(I + \frac{1}{\lambda}(\Sigma_{i-1} + x_{i+1} x_{i+1}^\top)) - \log\det(I + \frac{1}{\lambda}\Sigma_{i-1})\\
\geq&\log\det(I + \frac{1}{\lambda}\Sigma_{i+1}) - \log\det(I + \frac{1}{\lambda}\Sigma_i).
\end{aligned}$$

This means

$$\Delta\gamma(\lambda, x_i | x_1,\ldots,x_{i-1}) \geq \Delta\gamma(\lambda, x_{i+1} | x_1,\ldots,x_i).$$

$\square$