Title

Potential Bounds for Confidence Radii in Bandits and Reinforcement Learning

ISBN

9798247993094

Author

He, Zhenyu

Publication Date

2026-06-09

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Potential Bounds for Confidence Radii in Bandits and Reinforcement Learning

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Statistics

by

Zhenyu He

2026

ABSTRACT OF THE THESIS

Potential Bounds for Confidence Radii in Bandits and Reinforcement Learning

by

Zhenyu He
Master of Science in Statistics
University of California, Los Angeles, 2026
Professor Ying Nian Wu, Chair

Many regret analyses for bandit and reinforcement learning (RL) algorithms contain two steps. The first is to establish confidence bounds for unknown reward functions. The second is a deterministic accounting step that bounds the cumulative contribution of the resulting confidence radii. This thesis develops a unified framework for this second step. The framework treats both full sums over all rounds and selected-subsequence sums, as well as both unclipped and clipped sums. The results are stated in a general reproducing kernel Hilbert space (RKHS) setting and then specialized to finite-dimensional weighted linear designs.

The thesis of Zhenyu He is approved.

Yuhua Zhu

Drago Plecko

Ying Nian Wu, Committee Chair

University of California, Los Angeles

2026

Contents

Chapter 1

Introduction

In bandits and reinforcement learning (RL), regret measures the performance loss from not always selecting the optimal action, and is widely used as a key measure of learning efficiency. There is a common two-step method of regret analysis. The first step is to establish high-probability confidence events. The second step is to upper-bound the cumulative uncertainty terms, such as the sums of the resulting confidence radii, for a fixed confidence event. This thesis abstracts this second step and develops a mathematical framework for this summation problem.

A classical example is the elliptical-potential lemma in the analysis of OFUL, an algorithm for linear bandits, by Abbasi-Yadkori et al. [1]. In this setting, the sequence of actions is given and represented by feature vectors $x_t \in \mathbb{R}^d$. OFUL maintains the regularized design matrix

$$V_t = \lambda I + \sum_{s<t} x_s x_s^\top$$

before round t. The uncertainty of every selected action is measured by the self-normalized radius

$$\|x_t\|_{V_t^{-1}} .$$

The elliptical-potential lemma bounds cumulative uncertainty terms such as

$$\sum_{t=1}^{T} \min\{1, \|x_t\|^2_{V_t^{-1}}\}$$

by a log-determinant ratio.

The same type of potential argument also appears beyond finite-dimensional linear models. In kernelized bandits and Gaussian process bandit optimization, the unknown reward function is assumed to belong to, or be modeled through, a reproducing kernel Hilbert space (RKHS). In the RKHS setting, each action x_t is represented by the feature map $\phi(x_t)$. The uncertainty radius is defined through an RKHS design operator, while the resulting potential is represented through the Gram matrix. In the analyses of GP-UCB and related kernelized bandit algorithms, the accumulated uncertainty is controlled through log-determinants of Gram matrices and related complexity measures, such as information gain [6, 7, 4].

The uncertainty terms are summed along the same sequence that updates the design object in the examples above. However, in some refined analyses, the uncertainty terms are summed over a selected subsequence. One motivating example is the gap-dependent analysis of LSVI-UCB++, where weighted regression matrices are updated from all observed episodes, while certain bonus sums are taken only over selected episodes [5, 8]. Abstractly, if $\mathcal{J} = \{\tau_1 < \cdots < \tau_m\} \subseteq [T]$ denotes the selected subsequence, the quantity of interest may take the form

$$\sum_{i=1}^{m} \min\left\{C, \beta \left\|x_{\tau_i}\right\|_{V_{\tau_i}^{-1}}\right\}, \qquad V_{\tau_i} = \lambda I + \sum_{s<\tau_i} q_s^{-1} x_s x_s^\top.$$

The summation only ranges over the selected indices τ_i, but each V_{τ_i} contains all updates before τ_i. Thus, the selection of the subsequence only affects the sum of uncertainty terms, but not the design matrix.

The truncation reflects the fact that a regret or bonus contribution is often bounded by a fixed scale C, even when the corresponding confidence radius is large. While the

elliptical-potential lemma is stated for squared radii, regret decompositions often involve linear confidence widths. For this reason, both squared potential sums and linear sums are discussed in this thesis.

The examples above lead to the same mathematical task: given a design object that is updated over time, bound sums of self-normalized radii such as $\|\phi(x)\|_{V^{-1}}$. The goal of the thesis is to provide bounding techniques that can be used whenever a regret analysis involves this kind of summation.

This thesis starts from the most general setting considered here: selected-subsequence potential bounds in an RKHS. This setting covers two common special cases. First, the full-sequence case is obtained by taking the selected set to be $[T]$. Second, the finite-dimensional linear setting is recovered by taking the RKHS to be $\mathbb{R}^d$ with feature map $\phi(x) = x$. After establishing the RKHS selected-subsequence bound, the thesis studies additional assumptions that lead to more explicit bounds on various types of summation. These results provide reusable tools for the deterministic step in regret analyses, rather than new regret bounds for a specific algorithm.

The rest of the thesis is organized as follows. Chapter 2 presents the results in the general RKHS settings. Chapter 3 specializes these results to finite-dimensional cases for weighted linear designs and gives more explicit bounds under additional conditions. Chapter 4 discusses applications of our framework to analysis of several bandit and RL algorithms.

Chapter 2

Potential Bounds in RKHSs

This chapter sets up the main mathematical problem in a general RKHS framework, without referring to any specific algorithm. Since the bounds are taken along a selected subsequence, we introduce a surrogate design S_i that only uses the selected past updates. Although the RKHS feature space may be infinite-dimensional, the selected surrogate design changes only through the selected feature vectors. Hence, the relevant information increments lie in the finite-dimensional span generated by these vectors. This allows the normalized surrogate radii to be controlled by an information-gain potential written in terms of the selected Gram matrix. Using this reduction, we prove the main theorem bounding the summation of the uncertainty terms of interest, together with several related consequences.

2.1 Setup

Let $\mathcal{X}$ be an input domain, $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric positive semidefinite kernel, and $\mathcal{H}$ be the associated reproducing kernel Hilbert space (RKHS) with canonical feature map $\phi : \mathcal{X} \to \mathcal{H}$ [2, 3], so that

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}} .$$

For $u, v \in \mathcal{H}$, write $u \otimes v$ for the rank-one operator

$$(u \otimes v)f = \langle v, f \rangle_{\mathcal{H}} u, \qquad f \in \mathcal{H}.$$

Fix $T \geq 1$, a sequence $x_1, \ldots, x_T \in \mathcal{X}$, positive scale parameters $q_1, \ldots, q_T$, and a regularization parameter $\lambda > 0$. Define the full design operator by

$$V_t = \lambda I_{\mathcal{H}} + \sum_{s<t} q_s^{-1} \phi(x_s) \otimes \phi(x_s), \qquad t = 1, \ldots, T+1. \tag{2.1}$$

Let

$$\mathcal{J} = (\tau_1, \ldots, \tau_m), \qquad 1 \leq \tau_1 < \cdots < \tau_m \leq T,$$

be a selected subsequence of $[T]$. The full-history design uses every observation before time t. In contrast, the selected surrogate design uses only the previously selected observations:

$$S_i = \lambda I_{\mathcal{H}} + \sum_{j<i} q_{\tau_j}^{-1} \phi(x_{\tau_j}) \otimes \phi(x_{\tau_j}), \qquad i = 1, \ldots, m+1. \tag{2.2}$$

Since $V_t, S_i \succeq \lambda I_{\mathcal{H}}$, the inverses are bounded self-adjoint operators. Define

$$r_i^2 = \left\langle \phi(x_{\tau_i}), V_{\tau_i}^{-1} \phi(x_{\tau_i}) \right\rangle_{\mathcal{H}}, \qquad \tilde{r}_i^2 = \left\langle \phi(x_{\tau_i}), S_i^{-1} \phi(x_{\tau_i}) \right\rangle_{\mathcal{H}}. \tag{2.3}$$

Here r_i is measured using the full-history design, while $\tilde{r}_i$ is measured using the selected surrogate design. We normalize and clip the surrogate leverage by setting

$$u_i = q_{\tau_i}^{-1} \tilde{r}_i^2, \qquad v_i = \min\{1, u_i\}. \tag{2.4}$$

For the selected subsequence, define the Gram matrix and the diagonal scale matrix by

$$[K_{\mathcal{J}}]_{ab} = k(x_{\tau_a}, x_{\tau_b}), \qquad Q_{\mathcal{J}} = \operatorname{diag}(q_{\tau_1}, \ldots, q_{\tau_m}). \tag{2.5}$$

With these matrices, define the selected information-gain potential

$$\Gamma_{\mathcal{J}} = \log\det\left(I_m + \lambda^{-1} Q_{\mathcal{J}}^{-1/2} K_{\mathcal{J}} Q_{\mathcal{J}}^{-1/2}\right). \tag{2.6}$$

2.2 Selected Surrogate Design

The quantity r_i is measured with the full-history design V_{τ_i}, which contains all updates before time τ_i. To obtain a determinant telescope along the selected subsequence, we instead use the surrogate design S_i, which only keeps the previously selected updates. The next lemma relates these two designs and shows that the surrogate radius $\tilde{r}_i$ upper bounds the full-history radius r_i.

Lemma 2.1. *For each $i = 1, \dots, m$,*

$$S_i \preceq V_{\tau_i}.$$

Consequently,

$$V_{\tau_i}^{-1} \preceq S_i^{-1}, \qquad r_i \le \tilde{r}_i.$$

Proof. The difference between the full design at time τ_i and the selected surrogate design before the i-th selected point consists exactly of the unselected updates before τ_i:

$$\begin{aligned} V_{\tau_i} - S_i &= \sum_{t<\tau_i} q_t^{-1} \phi(x_t) \otimes \phi(x_t) - \sum_{j<i} q_{\tau_j}^{-1} \phi(x_{\tau_j}) \otimes \phi(x_{\tau_j}) \\ &= \sum_{\substack{t<\tau_i: \\ t \notin \{\tau_1, \dots, \tau_{i-1}\}}} q_t^{-1} \phi(x_t) \otimes \phi(x_t). \end{aligned}$$

Each summand is positive semidefinite, since for any $f \in \mathcal{H}$,

$$\langle f, (\phi(x_t) \otimes \phi(x_t)) f \rangle_{\mathcal{H}} = |\langle \phi(x_t), f \rangle_{\mathcal{H}}|^2 \ge 0.$$

Thus $V_{\tau_i} - S_i \succeq 0$, and hence $S_i \preceq V_{\tau_i}$.

Let

$$C = S_i^{-1/2}(V_{\tau_i} - S_i)S_i^{-1/2} \succeq 0.$$

Then

$$V_{\tau_i} = S_i^{1/2}(I_{\mathcal{H}} + C)S_i^{1/2}.$$

Since $C \succeq 0$, we have $(I_{\mathcal{H}} + C)^{-1} \preceq I_{\mathcal{H}}$. Hence

$$V_{\tau_i}^{-1} = S_i^{-1/2}(I_{\mathcal{H}} + C)^{-1}S_i^{-1/2} \preceq S_i^{-1}.$$

Evaluating both sides at $\phi(x_{\tau_i})$ gives $r_i^2 \leq \tilde{r}_i^2$, and therefore $r_i \leq \tilde{r}_i$. □

The next identity is the reason for introducing the surrogate design. Although $\mathcal{H}$ may be infinite-dimensional, the surrogate sequence is generated only by the selected feature vectors. Hence its determinant potential can be evaluated on their finite-dimensional span.

Lemma 2.2. *With u_i and $\Gamma_{\mathcal{J}}$ defined in* (2.4) *and* (2.6),

$$\Gamma_{\mathcal{J}} = \sum_{i=1}^{m} \log(1 + u_i). \tag{2.7}$$

Proof. Set

$$\psi_i = q_{\tau_i}^{-1/2}\phi(x_{\tau_i}), \qquad E = \operatorname{span}\{\psi_1, \dots, \psi_m\} \subseteq \mathcal{H}.$$

For $i = 0, \dots, m$, define the operator on E

$$A_i = S_{i+1}|_E = \left(\lambda I_{\mathcal{H}} + \sum_{j=1}^{i} \psi_j \otimes \psi_j\right)\Bigg|_E.$$

Then

$$A_0 = \lambda I_E, \qquad A_i = A_{i-1} + \psi_i \otimes \psi_i.$$

The rank-one determinant update on E gives

$$\det_E(A_i) = \det_E(A_{i-1}) \left(1 + \langle \psi_i, A_{i-1}^{-1} \psi_i \rangle_{\mathcal{H}}\right).$$

Since both E and $E^\perp$ are invariant under S_i, the inverse of $S_i|_E$ is the restriction of S_i^{-1} to E. Therefore

$$\langle \psi_i, A_{i-1}^{-1} \psi_i \rangle_{\mathcal{H}} = \langle \psi_i, S_i^{-1} \psi_i \rangle_{\mathcal{H}} = q_{\tau_i}^{-1} \left\langle \phi(x_{\tau_i}), S_i^{-1} \phi(x_{\tau_i}) \right\rangle_{\mathcal{H}} = u_i.$$

Thus,

$$\det_E(A_i) = \det_E(A_{i-1})(1 + u_i).$$

Multiplying over $i = 1, \ldots, m$ gives

$$\det_E(\lambda^{-1} A_m) = \prod_{i=1}^{m} (1 + u_i). \tag{2.8}$$

Now define $\Psi : \mathbb{R}^m \to E$ by

$$\Psi a = \sum_{i=1}^{m} a_i \psi_i.$$

Then

$$A_m = \lambda I_E + \Psi \Psi^*, \qquad \Psi^* \Psi = Q_{\mathcal{J}}^{-1/2} K_{\mathcal{J}} Q_{\mathcal{J}}^{-1/2}.$$

By Sylvester's determinant identity,

$$\det_E(\lambda^{-1} A_m) = \det \left(I_m + \lambda^{-1} Q_{\mathcal{J}}^{-1/2} K_{\mathcal{J}} Q_{\mathcal{J}}^{-1/2}\right).$$

Combining this identity with (2.8) and taking logarithms yields

$$\Gamma_{\mathcal{J}} = \sum_{i=1}^{m} \log(1 + u_i).$$

□

2.3 Potential Bounds

We now record a few consequences of Lemma 2.2. We first obtain a bound on the clipped quantities $v_i = \min\{1, u_i\}$.

Lemma 2.3. *For $v_i = \min\{1, u_i\}$,*

$$\sum_{i=1}^{m} v_i \le \frac{\Gamma_{\mathcal{J}}}{\log 2}. \tag{2.9}$$

Proof. Use the elementary inequality

$$\min\{1, u\} \le \frac{\log(1+u)}{\log 2}, \qquad u \ge 0.$$

Applying this inequality to u_i and summing over i, then using Lemma 2.2, gives

$$\sum_{i=1}^{m} v_i = \sum_{i=1}^{m} \min\{1, u_i\} \le \frac{1}{\log 2} \sum_{i=1}^{m} \log(1+u_i) = \frac{\Gamma_{\mathcal{J}}}{\log 2}.$$

□

When the surrogate leverage is uniformly bounded, the same potential also controls the untruncated sum.

Lemma 2.4. *Assume that $0 \le u_i \le U$ for each i, where $U > 0$. Then*

$$\sum_{i=1}^{m} u_i \le \frac{U}{\log(1+U)} \Gamma_{\mathcal{J}}. \tag{2.10}$$

Proof. For $0 \le u \le U$, the elementary inequality

$$u \le \frac{U}{\log(1+U)} \log(1+u)$$

holds. Applying it to u_i and using Lemma 2.2 gives

$$\sum_{i=1}^{m} u_i \le \frac{U}{\log(1+U)} \sum_{i=1}^{m} \log(1+u_i) = \frac{U}{\log(1+U)} \Gamma_{\mathcal{J}}.$$

□

A uniform bound on u_i follows, for example, from bounded kernel values and a positive lower bound on the selected scales.

Corollary 2.5. *Assume that for each $i = 1, \dots, m$, $k(x_{\tau_i}, x_{\tau_i}) \le L^2$ and $q_{\tau_i} \ge q_{\min} > 0$. Let*

$$U = \frac{L^2}{\lambda q_{\min}}.$$

If $U > 0$, then

$$\sum_{i=1}^{m} q_{\tau_i}^{-1} \tilde{r}_i^2 = \sum_{i=1}^{m} u_i \le \frac{U}{\log(1+U)} \Gamma_{\mathcal{J}}. \tag{2.11}$$

If, in addition, $q_{\tau_i} \le q_{\max}$ for each $i = 1, \dots, m$, then

$$\sum_{i=1}^{m} \tilde{r}_i^2 \le q_{\max} \frac{U}{\log(1+U)} \Gamma_{\mathcal{J}}. \tag{2.12}$$

Proof. Since $S_i \succeq \lambda I_{\mathcal{H}}$,

$$\tilde{r}_i^2 = \left\langle \phi(x_{\tau_i}), S_i^{-1} \phi(x_{\tau_i}) \right\rangle_{\mathcal{H}} \le \lambda^{-1} \left\| \phi(x_{\tau_i}) \right\|_{\mathcal{H}}^2 = \lambda^{-1} k(x_{\tau_i}, x_{\tau_i}) \le \frac{L^2}{\lambda}.$$

Thus $u_i = q_{\tau_i}^{-1} \tilde{r}_i^2 \le L^2/(\lambda q_{\min}) = U$. Apply Lemma 2.4. If $q_{\tau_i} \le q_{\max}$, then $\tilde{r}_i^2 = q_{\tau_i} u_i \le q_{\max} u_i$, which gives (2.12). □

The same potential also bounds the number of selected indices with large radii.

Lemma 2.6. *For $a > 0$,*

$$\#\{i : q_{\tau_i}^{-1/2} \tilde{r}_i \ge a\} \le \frac{\Gamma_{\mathcal{J}}}{\log(1+a^2)}. \tag{2.13}$$

If $q_{\tau_i} \le q_{\max}$ *for each* $i = 1, \dots, m$, *then*

$$\#\{i : r_i \ge a\} \le \frac{\Gamma_{\mathcal{J}}}{\log(1 + a^2/q_{\max})}. \tag{2.14}$$

Proof. The event $q_{\tau_i}^{-1/2}\tilde{r}_i \ge a$ is exactly the event $u_i \ge a^2$. Hence each counted index contributes at least $\log(1 + a^2)$ to $\sum_{i=1}^{m} \log(1 + u_i)$. By Lemma 2.2,

$$\sum_{i=1}^{m} \log(1 + u_i) = \Gamma_{\mathcal{J}},$$

and therefore

$$\#\{i : q_{\tau_i}^{-1/2}\tilde{r}_i \ge a\} \le \frac{\Gamma_{\mathcal{J}}}{\log(1 + a^2)}.$$

For the count involving the full-history radii, Lemma 2.1 gives $r_i \le \tilde{r}_i$. If $r_i \ge a$, then $\tilde{r}_i \ge a$, and hence

$$u_i = q_{\tau_i}^{-1}\tilde{r}_i^2 \ge a^2/q_{\tau_i} \ge a^2/q_{\max}.$$

Similarly, each such index contributes at least $\log(1+a^2/q_{\max})$ to $\Gamma_{\mathcal{J}}$, which proves (2.14). □

2.4 Main Theorem and Bounds on the Selected Potential

In related regret analyses, cumulative effects of uncertainty usually appear in terms of the actual full-history radii r_i, may be weighted by deterministic coefficients attached to the selected indices, and are often clipped at a fixed scale. Our goal is to bound this cumulative quantity by the selected potential $\Gamma_{\mathcal{J}}$, which leads to the main theorem of this chapter. We then further upper bound $\Gamma_{\mathcal{J}}$ by the standard maximum information gain.

For the main theorem, we use the scale condition

$$q_{\tau_i} \le B_i + \kappa r_i.$$

In Chapter 4, we will see that natural choices of B_i and κ give useful bounds for the concrete algorithms considered there.

Theorem 2.7. *Assume that for all $i = 1, \dots, m$, there are nonnegative numbers B_i and a constant $\kappa \geq 0$ such that*

$$q_{\tau_i} \leq B_i + \kappa r_i, \qquad i = 1, \dots, m. \tag{2.15}$$

Let $c_i \geq 0$ and $\eta_i \geq 0$. Define

$$D_{\max} = \max_{1 \leq i \leq m} c_i \max\{1, \kappa\eta_i\}. \tag{2.16}$$

Then

$$\sum_{i=1}^{m} c_i \min\{1, \eta_i r_i\} \leq \frac{D_{\max}}{\log 2}\Gamma_{\mathcal{J}} + \sqrt{\frac{\Gamma_{\mathcal{J}}}{\log 2}\sum_{i=1}^{m} c_i^2 \eta_i^2 B_i}. \tag{2.17}$$

Proof. By Lemma 2.1, $r_i \leq \tilde{r}_i$. The condition (2.15) therefore implies

$$q_{\tau_i} \leq B_i + \kappa\tilde{r}_i. \tag{2.18}$$

Using $\tilde{r}_i^2 = q_{\tau_i} u_i$, we get

$$\tilde{r}_i^2 = q_{\tau_i} u_i \leq u_i(B_i + \kappa\tilde{r}_i). \tag{2.19}$$

Solving this quadratic inequality in the nonnegative variable $\tilde{r}_i$ gives

$$\begin{aligned} \tilde{r}_i &\leq \frac{\kappa u_i + \sqrt{\kappa^2 u_i^2 + 4B_i u_i}}{2} \\ &\leq \kappa u_i + \sqrt{B_i u_i}. \end{aligned} \tag{2.20}$$

We now prove a pointwise bound. If $u_i < 1$, then $v_i = u_i$, and

$$\begin{aligned} c_i \min\{1, \eta_i r_i\} &\leq c_i \eta_i r_i \leq c_i \eta_i \tilde{r}_i \\ &\leq c_i \kappa \eta_i v_i + c_i \eta_i \sqrt{B_i v_i} \leq D_{\max} v_i + c_i \eta_i \sqrt{B_i v_i}. \end{aligned}$$

If $u_i \geq 1$, then $v_i = 1$, and

$$c_i \min\{1, \eta_i r_i\} \leq c_i \leq D_{\max} v_i \leq D_{\max} v_i + c_i \eta_i \sqrt{B_i v_i}.$$

Both cases give

$$c_i \min\{1, \eta_i r_i\} \leq D_{\max} v_i + c_i \eta_i \sqrt{B_i v_i}. \tag{2.21}$$

Summing (2.21) over i yields

$$\sum_{i=1}^{m} c_i \min\{1, \eta_i r_i\} \leq D_{\max} \sum_{i=1}^{m} v_i + \sum_{i=1}^{m} c_i \eta_i \sqrt{B_i v_i}. \tag{2.22}$$

The second term is bounded by Cauchy–Schwarz Inequality:

$$\begin{aligned} \sum_{i=1}^{m} c_i \eta_i \sqrt{B_i v_i} &= \sum_{i=1}^{m} (c_i \eta_i \sqrt{B_i}) \sqrt{v_i} \\ &\leq \sqrt{\left(\sum_{i=1}^{m} c_i^2 \eta_i^2 B_i \right) \left(\sum_{i=1}^{m} v_i \right)}. \end{aligned} \tag{2.23}$$

Finally apply Lemma 2.3, namely $\sum_{i=1}^{m} v_i \leq \Gamma_{\mathcal{J}} / \log 2$, to (2.22) and (2.23). This proves (2.17). □

The theorem above leaves the bound in terms of $\Gamma_{\mathcal{J}}$. Next, we further bound $\Gamma_{\mathcal{J}}$ in terms of maximum information gain. For $\eta > 0$, define

$$\gamma_m^{\eta}(k, \mathcal{X}) = \frac{1}{2} \sup_{z_1, \dots, z_m \in \mathcal{X}} \log \det \left(I_m + \eta^{-1} K(z_1, \dots, z_m) \right), \tag{2.24}$$

where $K(z_1, \dots, z_m) = [k(z_a, z_b)]_{a,b=1}^{m}$.

A lower bound on the selected scales is enough to compare $\Gamma_{\mathcal{J}}$ with this quantity.

Corollary 2.8. *If $q_{\tau_i} \geq q_{\min} > 0$ for every $i = 1, \dots, m$, then*

$$\Gamma_{\mathcal{J}} \leq 2\gamma_m^{\lambda q_{\min}}(k, \mathcal{X}). \tag{2.25}$$

Proof. Since $Q_{\mathcal{J}}^{-1} \preceq q_{\min}^{-1} I_m$,

$$K_{\mathcal{J}}^{1/2} Q_{\mathcal{J}}^{-1} K_{\mathcal{J}}^{1/2} \preceq q_{\min}^{-1} K_{\mathcal{J}}.$$

Using Sylvester's determinant identity,

$$\begin{aligned}
\Gamma_{\mathcal{J}} &= \log\det\left(I_m + \lambda^{-1} Q_{\mathcal{J}}^{-1/2} K_{\mathcal{J}} Q_{\mathcal{J}}^{-1/2}\right) \\
&= \log\det\left(I_m + \lambda^{-1} K_{\mathcal{J}}^{1/2} Q_{\mathcal{J}}^{-1} K_{\mathcal{J}}^{1/2}\right) \\
&\leq \log\det\left(I_m + (\lambda q_{\min})^{-1} K_{\mathcal{J}}\right) \\
&\leq 2\gamma_m^{\lambda q_{\min}}(k, \mathcal{X}).
\end{aligned}$$

The last inequality follows from (2.24) with $\eta = \lambda q_{\min}$, applied to the selected points. □

Thus, whenever the selected scales have a positive lower bound $q_{\min}$, the main theorem reduces the remaining complexity control to known bounds on $\gamma_m^{\lambda q_{\min}}(k, \mathcal{X})$. For standard kernels, such bounds can be found, for example, in Srinivas et al. [6, Theorem 5].

Chapter 3

Potential Bounds in Finite-Dimensional Linear Spaces

The previous chapter gives potential bounds for general RKHS feature maps. This chapter specializes the feature space to a finite-dimensional linear space, such as $\mathbb{R}^d$, under which the same bounds apply directly. In this finite-dimensional setting, the selected potential also admits an explicit trace-based upper bound. Such finite-dimensional linear feature models are the standard setting for regret analyses in linear bandits and linear Markov Decision Processes (MDPs).

3.1 Finite-Dimensional Linear Setup

We specialize the RKHS setting of Chapter 2 to the finite-dimensional linear case. Throughout this chapter, we identify the feature space with $\mathbb{R}^d$, use the canonical linear feature map $\phi(x) = x$, and take the associated kernel to be

$$k(x, x') = x^\top x'.$$

Under this specialization, the RKHS design operators become ordinary $d \times d$ matrices.

Let $d \geq 1$, let $x_1, \dots, x_T \in \mathbb{R}^d$, $q_1, \dots, q_T > 0$, and $\lambda > 0$. The full finite-dimensional design becomes

$$V_t = \lambda I_d + \sum_{s<t} q_s^{-1} x_s x_s^\top. \tag{3.1}$$

For a selected subsequence $\mathcal{J} = (\tau_1, \dots, \tau_m)$, define the selected surrogate design

$$S_i = \lambda I_d + \sum_{j<i} q_{\tau_j}^{-1} x_{\tau_j} x_{\tau_j}^\top. \tag{3.2}$$

The corresponding full-history radius, selected surrogate radius, and selected leverage are

$$r_i^2 = x_{\tau_i}^\top V_{\tau_i}^{-1} x_{\tau_i}, \qquad \tilde{r}_i^2 = x_{\tau_i}^\top S_i^{-1} x_{\tau_i}, \qquad u_i = q_{\tau_i}^{-1} \tilde{r}_i^2. \tag{3.3}$$

Let

$$X_{\mathcal{J}} = [x_{\tau_1}, \dots, x_{\tau_m}] \in \mathbb{R}^{d \times m}, \qquad Q_{\mathcal{J}} = \operatorname{diag}(q_{\tau_1}, \dots, q_{\tau_m}).$$

The selected potential can be written in Gram form as

$$\Gamma_{\mathcal{J}} = \log\det\left(I_m + \lambda^{-1} Q_{\mathcal{J}}^{-1/2} X_{\mathcal{J}}^\top X_{\mathcal{J}} Q_{\mathcal{J}}^{-1/2}\right). \tag{3.4}$$

By Sylvester's determinant identity, the same quantity also has the d-dimensional form

$$\Gamma_{\mathcal{J}} = \log\det\left(I_d + \lambda^{-1} X_{\mathcal{J}} Q_{\mathcal{J}}^{-1} X_{\mathcal{J}}^\top\right). \tag{3.5}$$

Therefore, the results of Chapter 2 apply directly in this finite-dimensional linear setting. In particular,

$$S_i \preceq V_{\tau_i}, \qquad r_i \leq \tilde{r}_i, \qquad \Gamma_{\mathcal{J}} = \sum_{i=1}^{m} \log(1 + u_i).$$

3.2 Selected Potential Bounds

The following bounds are direct applications of Theorem 2.7 and Corollary 2.5 under the finite-dimensional linear specialization in this chapter.

Corollary 3.1. *Assume that for all* $i = 1, \dots, m$

$$q_{\tau_i} \le B_i + \kappa r_i, \qquad B_i \ge 0, \qquad \kappa \ge 0. \tag{3.6}$$

For nonnegative c_i *and* η_i*, define*

$$D_{\max} = \max_{1 \le i \le m} c_i \max\{1, \kappa\eta_i\}.$$

Then

$$\sum_{i=1}^{m} c_i \min\{1, \eta_i r_i\} \le \frac{D_{\max}}{\log 2}\Gamma_{\mathcal{J}} + \sqrt{\frac{\Gamma_{\mathcal{J}}}{\log 2}\sum_{i=1}^{m} c_i^2 \eta_i^2 B_i}. \tag{3.7}$$

Corollary 3.2. *Assume that* $\left\|x_{\tau_i}\right\|_2 \le L$ *and* $q_{\tau_i} \ge q_{\min} > 0$ *for all* $i = 1, \dots, m$*. Let*

$$U = \frac{L^2}{\lambda q_{\min}}.$$

If $U > 0$*, then*

$$\sum_{i=1}^{m} u_i \le \frac{U}{\log(1+U)}\Gamma_{\mathcal{J}}. \tag{3.8}$$

If, in addition, $q_{\tau_i} \le q_{\max}$ *for all* $i = 1, \dots, m$*, then*

$$\sum_{i=1}^{m} x_{\tau_i}^{\top} S_i^{-1} x_{\tau_i} \le q_{\max}\frac{U}{\log(1+U)}\Gamma_{\mathcal{J}}. \tag{3.9}$$

3.3 Bounding the Selected Potential by Trace

As in Chapter 2, the selected potential $\Gamma_{\mathcal{J}}$ can be bounded through the information-gain quantity. In finite-dimensional linear settings, we can also give a direct trace-based upper

bound through its log-determinant representation. Under standard boundedness conditions on the selected feature norms and scales, this trace bound further yields a coarser explicit bound, which is useful in regret analyses for linear bandit and linear MDP algorithms.

Proposition 3.3. *For every selected subsequence $\mathcal{J}$,*

$$\Gamma_{\mathcal{J}} \leq d \log \left(1 + \frac{1}{\lambda d} \sum_{i=1}^{m} q_{\tau_i}^{-1} \left\| x_{\tau_i} \right\|_2^2 \right). \tag{3.10}$$

If $\left\| x_{\tau_i} \right\|_2 \leq L$ and $q_{\tau_i} \geq q_{\min} > 0$ for all $i = 1, \ldots, m$, then

$$\Gamma_{\mathcal{J}} \leq d \log \left(1 + \frac{mL^2}{\lambda d q_{\min}} \right). \tag{3.11}$$

Proof. Let

$$A = \lambda^{-1} X_{\mathcal{J}} Q_{\mathcal{J}}^{-1} X_{\mathcal{J}}^{\top} \succeq 0.$$

By (3.5), $\Gamma_{\mathcal{J}} = \log \det(I_d + A)$. Let $a_1, \ldots, a_d \geq 0$ be the eigenvalues of A. Then

$$\det(I_d + A) = \prod_{j=1}^{d} (1 + a_j).$$

By the arithmetic-geometric mean inequality,

$$\prod_{j=1}^{d} (1 + a_j) \leq \left(1 + \frac{1}{d} \sum_{j=1}^{d} a_j \right)^d.$$

Taking logarithms gives

$$\log \det(I_d + A) \leq d \log \left(1 + \frac{\operatorname{tr}(A)}{d} \right).$$

Finally,

$$\operatorname{tr}(A) = \lambda^{-1} \operatorname{tr}(X_{\mathcal{J}} Q_{\mathcal{J}}^{-1} X_{\mathcal{J}}^{\top}) = \lambda^{-1} \sum_{i=1}^{m} q_{\tau_i}^{-1} \left\| x_{\tau_i} \right\|_2^2.$$

This proves (3.10). The bound (3.11) follows by substituting $\left\|x_{\tau_i}\right\|_2^2 \leq L^2$ and $q_{\tau_i}^{-1} \leq q_{\min}^{-1}$. $\square$

Chapter 4

Potential Bounds in Existing Regret Analyses

This chapter connects the potential bounds developed in the previous chapters with the summation arguments that appear in existing regret analyses for bandit and RL algorithms. The goal is not to reprove the full regret bounds of these algorithms, but to isolate the common step of controlling cumulative confidence radii. We illustrate this connection through several examples, including OFUL, WeightedOFUL+, LSVI-UCB++, and GP-UCB.

4.1 OFUL

We begin with OFUL, a stochastic linear bandit algorithm proposed by Abbasi-Yadkori et al. [1]. Given a sequence of actions $x_1, \dots, x_T \in \mathbb{R}^d$, define

$$V_t = \lambda I_d + \sum_{s<t} x_s x_s^\top, \qquad \ell_t = x_t^\top V_t^{-1} x_t = \|x_t\|^2_{V_t^{-1}}.$$

A key step in their analysis is obtaining bounds on a clipped sum, given by Lemma 11 of Abbasi-Yadkori et al. [1].

$$\sum_{t=1}^{T} \min\{1, \ell_t\} \leq 2 \log \frac{\det(V_{T+1})}{\det(V_1)}. \tag{4.1}$$

Further, if $\|x_t\|_2 \leq L$ for all t, then

$$\sum_{t=1}^{T} \min\{1, \ell_t\} \leq 2d \log \left(1 + \frac{TL^2}{\lambda d}\right). \tag{4.2}$$

In the notation of Chapter 3, this problem is exactly the unweighted full-sequence case, with $q_t \equiv 1$ and $\mathcal{J} = (1, \ldots, T)$. Combining Lemma 2.3 with Proposition 3.3 recovers (4.1) and (4.2), with a slightly sharper coefficient.

Proposition 4.1.

$$\sum_{t=1}^{T} \min\{1, \ell_t\} \leq \frac{1}{\log 2} \log \frac{\det(V_{T+1})}{\det(V_1)}. \tag{4.3}$$

Further, if $\|x_t\|_2 \leq L$ *for all* t*, then*

$$\sum_{t=1}^{T} \min\{1, \ell_t\} \leq \frac{d}{\log 2} \log \left(1 + \frac{TL^2}{\lambda d}\right). \tag{4.4}$$

Proof. Take $\mathcal{J} = (1, \ldots, T)$ and $q_t = 1$, Lemma 2.3 gives

$$\sum_{t=1}^{T} \min\{1, \ell_t\} \leq \Gamma_{[T]} / \log 2,$$

and by (3.5),

$$\Gamma_{[T]} = \log \det \left(I_d + \lambda^{-1} \sum_{t=1}^{T} x_t x_t^\top\right) = \log \frac{\det(V_{T+1})}{\det(V_1)},$$

which proves (4.3). The explicit bound (4.4) follows from Proposition 3.3 with $q_{\min} = 1$ and $m = T$. □

4.2 WeightedOFUL+

We next consider WeightedOFUL+, a weighted variant of OFUL for heterogeneous linear bandits [9]. Let $\{x_k\}_{k=1}^K \subseteq \mathbb{R}^d$ be the action sequence. Now the design matrix is updated with adaptive weights

$$Z_1 = \lambda I_d, \qquad Z_{k+1} = Z_k + \frac{x_k x_k^\top}{\bar{\sigma}_k^2},$$

where

$$\bar{\sigma}_k = \max\left\{\sigma_k, \alpha, \gamma \|x_k\|_{Z_k^{-1}}^{1/2}\right\}, \qquad \sigma_k \geq 0, \qquad \alpha > 0, \qquad \gamma > 0.$$

The goal is to upper bound a weighted clipped sum $\sum_{k=1}^K \min\left\{1, \hat{\beta}_k \|x_k\|_{Z_k^{-1}}\right\}$, with nonnegative weights $\hat{\beta}_k$.

Assuming $\|x_k\|_2 \leq A$ and $\iota = \log\left(1 + \frac{KA^2}{d\lambda\alpha^2}\right)$, Lemma 4.4 of Zhou and Gu [9] gives

$$\sum_{k=1}^K \min\left\{1, \hat{\beta}_k \|x_k\|_{Z_k^{-1}}\right\} \leq 2d\iota + 2\max_{k\in[K]} \hat{\beta}_k \gamma^2 d\iota + 2\sqrt{d\iota \sum_{k=1}^K \hat{\beta}_k^2(\sigma_k^2 + \alpha^2)}. \tag{4.5}$$

The following proposition gives a sharper corresponding bound using our results.

Proposition 4.2. *Under the setting above,*

$$\sum_{k=1}^K \min\left\{1, \hat{\beta}_k \|x_k\|_{Z_k^{-1}}\right\} \leq \frac{d\iota}{\log 2} \max\left\{1, \gamma^2 \max_{k\in[K]} \hat{\beta}_k\right\} + \sqrt{\frac{d\iota}{\log 2} \sum_{k=1}^K \hat{\beta}_k^2(\sigma_k^2 + \alpha^2)}. \tag{4.6}$$

Proof. Set

$$q_k = \bar{\sigma}_k^2, \qquad r_k = \|x_k\|_{Z_k^{-1}}.$$

Since

$$\bar{\sigma}_k = \max\{\sigma_k, \alpha, \gamma r_k^{1/2}\},$$

we have

$$q_k = \bar{\sigma}_k^2 \leq \sigma_k^2 + \alpha^2 + \gamma^2 r_k.$$

Applying Corollary 3.1 to the full sequence $\mathcal{J} = [K]$ with

$$c_k = 1, \qquad \eta_k = \hat{\beta}_k, \qquad B_k = \sigma_k^2 + \alpha^2, \qquad \kappa = \gamma^2$$

gives

$$\sum_{k=1}^{K} \min\left\{1, \hat{\beta}_k \left\|x_k\right\|_{Z_k^{-1}}\right\} \le \frac{\Gamma_{[K]}}{\log 2} \max\left\{1, \gamma^2 \max_{k\in[K]} \hat{\beta}_k\right\} + \sqrt{\frac{\Gamma_{[K]}}{\log 2} \sum_{k=1}^{K} \hat{\beta}_k^2(\sigma_k^2 + \alpha^2)}. \tag{4.7}$$

Moreover, because $q_k \ge \alpha^2$ and $\left\|x_k\right\|_2 \le A$, Proposition 3.3 gives

$$\Gamma_{[K]} \le d \log\left(1 + \frac{KA^2}{d\lambda\alpha^2}\right) = d\iota.$$

Substituting this bound into (4.7) proves (4.6). □

Compared with (4.5), Proposition 4.2 gives a somewhat sharper bound under the same normalization. The first two terms in (4.5) are combined into one maximum term, and the coefficients are also improved.

4.3 LSVI-UCB++

The next example is LSVI-UCB++, an optimism-based algorithm for linear MDPs. He et al. [5] show that LSVI-UCB++ achieves the nearly minimax worst-case regret bound $\widetilde{O}(d\sqrt{H^3K})$. Later, Zhang et al. [8] establish a gap-dependent bound for the same algorithm. For a fixed minimum suboptimality gap $\Delta_{\min} > 0$, their regret bound has only polylogarithmic dependence on K.

The worst-case analysis of He et al. [5] controls bonus and potential terms over the full sequence of episodes, in the same spirit as the full-sequence weighted estimate discussed in Section 4.2. In contrast, the gap-dependent analysis of Zhang et al. [8] requires bounds on partial sums of bonuses over selected episodes. We focus on this selected-sequence potential

estimate and compare it with Lemma 5.3 of Zhang et al. [8].

We now formulate the selected-sequence setting. Fix a stage h and suppress the stage index. Let $\{x_k\}_{k=1}^K \subseteq \mathbb{R}^d$ be the feature sequence, and let the full-history design matrices be

$$Z_1 = \lambda I_d, \qquad Z_{k+1} = Z_k + \frac{x_k x_k^\top}{\bar{\sigma}_k^2},$$

where

$$\bar{\sigma}_k = \max\left\{\sigma_k, \alpha, \gamma \|x_k\|_{Z_k^{-1}}^{1/2}\right\}.$$

Here $\lambda, \alpha, \gamma > 0$, $\sigma_k \geq 0$, and $\|x_k\|_2 \leq A$. Let

$$\mathcal{J} = \{\tau_1 < \tau_2 < \cdots < \tau_m\} \subseteq [K]$$

be a selected subsequence of episodes. The selected-sequence sum considered below is

$$\sum_{i=1}^m \min\left\{C, \beta \left\|x_{\tau_i}\right\|_{Z_{\tau_i}^{-1}}\right\},$$

where $C, \beta > 0$.

Under this setting, Lemma 5.3 of Zhang et al. [8] gives the following bound

$$\sum_{i=1}^m \min\left\{C, \beta \left\|x_{\tau_i}\right\|_{Z_{\tau_i}^{-1}}\right\} \leq 2Cd\iota + 2\beta\gamma^2 d\iota + 2\beta\sqrt{d\iota \sum_{i=1}^m (\sigma_{\tau_i}^2 + \alpha^2)}, \tag{4.8}$$

where $\iota = \log(1 + KA^2/(d\lambda\alpha^2))$.

This bound can be viewed as the selected-subsequence counterpart of (4.5). Similarly, applying our selected-subsequence bound yields the following sharper estimate

$$\sum_{i=1}^m \min\left\{C, \beta \left\|x_{\tau_i}\right\|_{Z_{\tau_i}^{-1}}\right\} \leq \frac{d\iota}{\log 2} \max\{C, \beta\gamma^2\} + \beta\sqrt{\frac{d\iota}{\log 2} \sum_{i=1}^m (\sigma_{\tau_i}^2 + \alpha^2)}. \tag{4.9}$$

4.4 GP-UCB

We finally turn to the Gaussian Process Upper Confidence Bound (GP-UCB) algorithm as an example in the RKHS setting. Srinivas et al. [6] analyze this algorithm and bound its cumulative regret in terms of the maximum information gain. Their probabilistic analysis provides confidence bounds for the unknown function. After these bounds are established, the core of the deterministic step is to control sums of posterior variances and their clipped variants by information gain.

Let k be the kernel and let $\mathcal{H}$ be its RKHS, with feature map $\phi(x) = k(x, \cdot)$. Let $\lambda > 0$ be the noise variance in the GP posterior. For a realized sequence of chosen points $x_1, \dots, x_T$, define

$$V_t = \lambda I_{\mathcal{H}} + \sum_{s<t} \phi(x_s) \otimes \phi(x_s), \qquad r_t^2 = \left\langle \phi(x_t), V_t^{-1} \phi(x_t) \right\rangle_{\mathcal{H}}.$$

The posterior variance satisfies

$$\sigma_{t-1}^2(x_t) = \lambda r_t^2.$$

Thus the normalized leverage score is

$$u_t = r_t^2 = \lambda^{-1} \sigma_{t-1}^2(x_t).$$

For the realized design $A_T = (x_1, \dots, x_T)$, let

$$K_T = [k(x_s, x_t)]_{s,t=1}^T,$$

and the information gain

$$I(y_{A_T}; f_{A_T}) = \frac{1}{2} \log \det \left(I_T + \lambda^{-1} K_T \right). \tag{4.10}$$

For $T \geq 1$, define the maximum information gain

$$\gamma_T = \max_{A:\,|A|=T} I(y_A; f_A).$$

To recover corresponding estimates in Srinivas et al. [6], we give an information-gain identity, and control sums of normalized posterior variances $u_t = \lambda^{-1}\sigma^2_{t-1}(x_t)$ and their clipped variants by information gain; see Lemmas 5.3, 5.4, and 7.1 of Srinivas et al. [6].

Proposition 4.3. *For the quantities defined above,*

$$\sum_{t=1}^{T} \log(1 + u_t) = 2I(y_{A_T}; f_{A_T}). \tag{4.11}$$

Moreover, for any $\alpha > 0$,

$$\sum_{t=1}^{T} \min\{u_t, \alpha\} \leq \frac{2\alpha}{\log(1+\alpha)} \gamma_T. \tag{4.12}$$

Further, if $k(x, x) \leq 1$ for all $x \in \mathcal{X}$, then

$$\sum_{t=1}^{T} u_t \leq \frac{2\lambda^{-1}}{\log(1+\lambda^{-1})} \gamma_T. \tag{4.13}$$

Proof. With $q_t \equiv 1$ and $\mathcal{I} = (1, \ldots, T)$, Lemma 2.2 gives

$$\sum_{t=1}^{T} \log(1 + u_t) = \log\det\left(I_T + \lambda^{-1} K_T\right).$$

Combining it with (4.10) proves (4.11).

For any $u \geq 0$,

$$\min\{u, \alpha\} \leq \frac{\alpha}{\log(1+\alpha)} \log(1 + u).$$

Applying this inequality to u_t and summing over t, then using (4.11) gives

$$\sum_{t=1}^{T} \min\{u_t, \alpha\} \leq \frac{\alpha}{\log(1+\alpha)} \sum_{t=1}^{T} \log(1+u_t) = \frac{2\alpha}{\log(1+\alpha)} I(y_{A_T}; f_{A_T}) \leq \frac{2\alpha}{\log(1+\alpha)} \gamma_T.$$

Now assume $k(x,x) \leq 1$. Since $V_t \succeq \lambda I_{\mathcal{H}}$, we have $V_t^{-1} \preceq \lambda^{-1} I_{\mathcal{H}}$. Hence

$$0 \leq u_t = \|\phi(x_t)\|^2_{V_t^{-1}} \leq \lambda^{-1} \|\phi(x_t)\|^2_{\mathcal{H}} \leq \lambda^{-1}.$$

Applying Lemma 2.4 with $U = \lambda^{-1}$ gives

$$\sum_{t=1}^{T} u_t \leq \frac{\lambda^{-1}}{\log(1+\lambda^{-1})} \sum_{t=1}^{T} \log(1+u_t) = \frac{2\lambda^{-1}}{\log(1+\lambda^{-1})} I(y_{A_T}; f_{A_T}) \leq \frac{2\lambda^{-1}}{\log(1+\lambda^{-1})} \gamma_T.$$

□

Bibliography

[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems 24*, pages 2312–2320, 2011. URL https://proceedings.neurips.cc/paper/2011/hash/e1d5be1c7f2f456670de3d53c7b54f4a-Abstract.html.

[2] Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950. doi: 10.2307/1990404.

[3] Alain Berlinet and Christine Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, Boston, 2004.

[4] Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 844–853. PMLR, 2017. URL https://proceedings.mlr.press/v70/chowdhury17a.html.

[5] Jiafan He, Heyang Zhao, Dongruo Zhou, and Quanquan Gu. Nearly minimax optimal reinforcement learning for linear markov decision processes. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 12790–12822. PMLR, 2023. URL https://proceedings.mlr.press/v202/he23d.html.

[6] Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design, 2010.

[7] Michal Valko, Nathaniel Korda, Rémi Munos, Ilias Flaounas, and Nello Cristianini. Finite-time analysis of kernelised contextual bandits. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pages 654–663, 2013. URL https://arxiv.org/abs/1309.6869.

[8] Haochen Zhang, Zhong Zheng, and Lingzhou Xue. Gap-dependent bounds for nearly minimax optimal reinforcement learning with linear function approximation, 2026. URL https://arxiv.org/abs/2602.20297.

[9] Dongruo Zhou and Quanquan Gu. Computationally efficient horizon-free reinforcement learning for linear mixture MDPs. In *Advances in Neural Information Processing Systems 35*, pages 36337–36349, 2022. URL https://openreview.net/forum?id=H4GmqyYMxFP.