Lower Bounds of Stochastic Bandits

Dick Jessen William

October 31, 2021

Abstract

In this report, we will continue our discussion about Stochastic Bandits. On the contrary with the previous chapter, we will focus on the limitations of the bandit algorithms. This report follows from the chapter 2 of the book of Aleksandrs Slivkins [24] and assumes knowledge from the chapter 1 of this book.

1 Introduction

We will revisit the Stochastic Bandits problem in a different view. Instead on looking on one algorithm, we will look at all possible algorithm and prove that it cannot achieves some level of regret rate. Firstly, we prove all algorithm suffers $\Omega(\sqrt{KT})$ on some problem instance. After that, we will use a similar technique to strengthen the inequality for non-adaptive exploration. Finally, we discuss without proof some instance-dependent $\Omega \log T$ lower bounds.

First, lets state the $\Omega(\sqrt{KT})$ lower bound.

Theorem 1. *For any time horizon T and total number of arms K. For any bandit algorithm, there exist a problem instance such that $\mathbb{E}[R(T)] \geq \Omega\sqrt{KT}$.*

To prove this, we construct a family F of problem instances that can force the algorithm to productbad results. There are two ways to do this.

1. Prove any algorithm has high regret on some instance in F.
2. Define a distribution over F and prove that any algorithm has high regret expectation over this distribution.

To prove the theorem above, we consider 0-1 rewards and the following family of problem instances, with $\epsilon > 0$ to be adjusted. For $j = \{1, 2, \cdots, K\}$, define I_j as below.

$$I_j = \begin{cases} \mu_i = \frac{1+\epsilon}{2} & \text{if i = j} \\ \mu_i = \frac{1}{2} & \text{otherwise} \end{cases}$$

By noting that Successive Elimination would sample every suboptimal arm at most $O(\epsilon^{-2} \log^k \epsilon^{-2})$ times, we note that sampling each arm $O(\epsilon^{-2} \log^k \epsilon^{-2})$ times suffices for our regret bound. Our goal is to prove that sampling each arm $\Theta(\epsilon^{-2})$ is necessary to check whether the arm is good or not. Hence, the regret is $\Theta(K/\epsilon)$. Choosing $\epsilon = \Omega(\sqrt{K/T})$ finishes our proof. The next sections will explore the technical details of this computation.

2 KL-Divergence

Our proof will use the notion of KL-divergence on information theory. This section will introduce KL-divergence for finite sample spaces.

Condsider a finite sample space Ω and p, q be two probability distributionn on Ω. Then, define the KL-divergence as

$$KL(p,q) = \sum_{x \in \Omega} p(x) \ln \frac{p(x)}{q(x)} = \mathbb{E}_p \left[\ln \frac{p(x)}{q(x)} \right].$$

Note that KL is not symmetric, and does not satisfy triangle inequality.

Here are some important properties of KL-divergence. Let RC_ϵ denote a biased random coin with bias $\epsilon/2$ for a positive ϵ. Then, the following holds.

Theorem 2. *KL-divergence satisfies the following.*

1. *(Gibbs Inequaity) For any distribution p, q, we have $KL(p,q) \geq 0$. Equality holds iff $p = q$*

2. *(Chain Rule) Let the sample space be a product $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$. Let p, q be two distributions of Ω such that $p = p_1 \times p_2 \cdots \times p_n$ and $q = q_1 \times q_2 \times \cdots \times q_n$, with p_i, q_i are distributions on Ω_j for $j \in \{1, 2, \cdots, n\}$. Then, $KL(p, q) = \sum_{i=1}^{n} KL(p_i, q_i)$.*

3. *(Pinsker's Inequality) For any event $A \in \Omega$, we have $2(p(A) - q(A))^2 \leq KL(p, q)$.*

4. *(Random Coins) $KL(RC_\epsilon, RC_0) \leq 2\epsilon^2$ and $KL(RC_0, RC_\epsilon) \leq \epsilon^2$ for all $\epsilon \in (0, \frac{1}{2})$.*

We can prove some properties using these facts.

Lemma 1. *Consider sample space $\Omega = \{0,1\}^n$ and two distributions on Ω, $p = RC_\epsilon^n$ and $q = RC_0^n$. Then, there exists $\epsilon > 0$ such that for all $A \in \Omega$, $|p(A) - q(A)| \leq \epsilon\sqrt{n}$.*

Proof. Consider the setting from Theorem 2-2 with n samples from 2 coins, with $p_j = RC_\epsilon$ is a biased random coin, and $q_k = RC_0$ is a fair random coin. Then,

$$\begin{aligned}
2(p(A) - q(A))^2 &\leq KL(p, q) && \text{(Pinsker's inequality)} \\
&= \sum_{i=1}^{n} KL(p_i, q_i) && \text{(Chain rule)} \\
&\leq n \times KL(RC_\epsilon, RC_0) \\
&\leq 2n\epsilon^2 && \text{(Theorem 2-4)}
\end{aligned}$$

Hence, taking the square root, we derived the lemma. □

3 Flipping One Coin

We will see how KL-divergence technique work in action. Consider a biased random coin with unknown mean $\mu \in [0, 1]$. Assume that $\mu \in \{\mu_1, \mu_2\}$ for known μ_1 and μ_2, with $\mu_1 > \mu_2$. Then, we flip it T times. We want to identify μ with a high probability.

More formally, define $\Omega = \{0,1\}^T$ as the sample space for the T coin tosses. We want to have a decision rule $Rule : \Omega \to \{HIGH, LOW\}$, and satisfies

$$P(Rule(Observations) = HIGH | \mu = \mu_1) \geq 0.99,$$

$$P(Rule(Observations) = LOW | \mu = \mu_2) \geq 0.99.$$

Now, how large T needs to be so that a rule exist. We already know that $\tilde{T(}\mu_1 - \mu_2)^{-2}$ is sufficient. We will prove the it is also necessary. First, lets solve the case when both means are close to $\frac{1}{2}$.

Lemma 2. *Let $\mu_1 = \frac{1+\epsilon}{2}$ and $\mu_2 = \frac{1}{2}$. With a decision rule like above, we have $T > \frac{1}{4\epsilon^2}$.*

Proof. Let $A_0 \subset \Omega$ denotes the event that the rule returns $HIGH$. Then, $P(A_0|\mu = \mu_1) - P(A_0|\mu = \mu_2) \geq 0.98$. Let $P_i(A) = P(A|\mu = \mu_i)$ for each event $A \subset \Omega$ and $i \in \{1, 2\}$. Then, $P_i = P_{i,1} \times \cdots \times P_{i,T}$, where $P_{i,t}$ denotes the distribution of the $t-th$ coin if $\mu = \mu_i$. Now, we apply Lemma 1, to get $|P_1(A) - P_2(A)| \leq \epsilon\sqrt{T}$. Now, plug A with A_0 and $T \leq \frac{1}{4\epsilon^2}$ to get $|P_1(A_0) - P_2(A_0)| < 0.5$, contradiction. □

4 Flipping Many Coins: Best-Arm Identification

We extend the idea before to multiple coins. Consider a bandit problem with K arms, with each arm is a biased coin with an unknown mean. After T rounds, the algorithm will output an arm y_T, a prediction of the most optimal arm. This version is called the best-arm identification. We will mainly focus on the quality of this prediction, not on the regret.

Define $[K]$ as the set of arms, $\mu(a)$ as the mean reward of arm a and a problem instance as a tuple $I = (\mu(a) : a \in [K])$. An algorithm is called good for best-arm identification if for all I, $P(y_T \text{ is the best arm}|T) \geq 0.99$. We will the family of problem instances discussed earlier with parameter ϵ to argue that $T \geq \Omega(\frac{K}{\epsilon^2})$ for any working algorithm. In fact, we prove a stronger statement here/

Lemma 3. *Consider a best arm identification problem with $T \leq \frac{cK}{\epsilon^2}$ for a small positive constant. For any fixed deterministic algorithm, there exists at least $\lceil K/3 \rceil$ arms such that for the problem instance in the earlier page I_a, we have $P(y_t = a|I_a) < 0.75$.*

Proof. We prove the case for two arms here. Let $(r_t(a) : a \in [K].t \in [T])$ be a tuple of mutually independent Bernoulli random variables such that $\mathbb{E}[r_t(a)] = \mu(a)$. Call this tuple a rewards table. Here, $r_t(a)$ is the reward of choosing the $a-th$ arm for the $t-th$ time. Note that $\Omega = \{0,1\}^{K \times T}$, where each outcome is a realization og the reward table. Any event about the algorithm is interpreted as a subset of Ω. Each problem instance I_j defines distribution P_j on Ω :

$$P_j(A) = P(A|I_j)\forall A \subset \Omega.$$

Let $P_j^{a,t}$ be the distribution of $r_t(a)$ under I_j so that $P_j = \prod_{a\in[K],t\in[T]} P_J^{a,t}$. We only need to prove the inequality holds for at least one of the arms. Assume both do not for the sake of contradiction. Let $A = \{y_T = 1\} \subset \Omega$ be the event that the algorithm predicts arm 1. Then, $P_1(A) \geq 0.75$ and $P_2(A) \leq 0.25$. Hence, $P_1(A)-P_2(A) \geq 0.5$. By KL-Divergence,

$$\begin{aligned} 2(P_1(A) - P_2(A))^2 &\leq KL(P_1, P_2) && \text{(Pinsker's Inequality)} \\ &= \sum_{a=1}^{K}\sum_{t=1}^{T} KT(P_1^{a,t}, P_2^{a,t}) && \text{(Chain Rule)} \\ &\leq 2T \times 2\epsilon^2 && \text{(Theorem 2-4)} \end{aligned}$$

We can use Theorem 2-4 because for each arm a and each round t, one of $P_1^{a,t}$ and $P_2^{a,t}$ is a fair coin. Simplifying, we get that $P_1(A) - P_2(A) \leq \epsilon\sqrt{2T} < 0.5$, whenever $T \leq 1/16\epsilon^2$. Contradiction. □

Corollary 1. *Assuming T as above, if we fix any algorithm for best arm identification and we choose an arm a uniformly at random then running the algorithm on instance I_a, then $P(y_T \neq a) > \frac{1}{12}$.*

Proof. This can be seen from Lemma 3 for deterministic algorithm. Any randomized algorithm can be expressed as a distribution over deterministic algorithms. □

Theorem 3. *Fix time horizon T, number of arms K and a bandit algorithm. Run the algorithm on an instance I_a. Then, $\mathbb{E}[R(T)] \geq \Omega(\sqrt{KT})$.*

Proof. Fix $\epsilon > 0$ and assume that $T \leq \frac{cK}{\epsilon^2}$, with c being constant from Lemma 3. Fix a round t. Interpret the algorithm as a best arm algorithm, with prediction equal to the chosen arm in the round. Using Corollary 1, we have $P(a_t \neq a) \geq \frac{1}{12}$. In other words, the probability that the algorithm chooses an non-optimal arm is at least $\frac{1}{12}$. Recall that for each I_a, the gap is $\epsilon/2$ whenever a non-optimal arm is chosen. Hence, $\mathbb{E}[\Delta(a_t)] \geq \frac{1}{12} \times \epsilon/2 = \epsilon/24$. Hence, summing over all rounds, $\mathbb{E}[R(T)] \geq \epsilon T/24$. Substituting $\epsilon = \sqrt{cK/T}$ finishes the proof. □

5 The General Case

Now, we turn our attention to the general case. Note that using the proof for the case $K = 2$ only works if $T \leq c/\epsilon^2$, and this gives a lower bound of $\Omega(T)$. Hence, we need a better analysis. For this analysis, consider an additional problem instance $I_0 = \{\mu_i = \frac{1}{2}$ for all arms $i\}$. Denote $\mathbb{E}[\cdot]$ be the expectation given this problem instance. Denote T_a be the total number of times arm a is played.

Consider the algorithm performance on I_0, and look at arms j that is not picked by the algorithm much and also not likely to be picked for the guess y_T. Formally, observe that

1. There are at least $2K/3$ arms j such that $\mathbb{E}_0(T_j) \leq 3T/K$.
2. There are at least $2K/3$ arms j such that $P_0(y_T = j) \leq 3/K$.

By Markov's Inequality, $\mathbb{E}_0(T_j) \leq 3T/K \implies P(T_j \leq 24T/K) \geq 7/8$. Because there are at least $K/3$ arms satisfying these two properties, we conclude that there are at least $K/3$ arms j such that $P(T_j \leq 24T/K) \geq 7/8$ and $P_0(y_T = j) \leq 3/K$.

Now, lets redefine the sample space. For each arm a, define the $t-$round sample space $\Omega_a^t = \{0, 1\}^t$, with each outcome is a realization of tuple $(r_s(a) : s \in [t])$. Then, the sample space we considered before is $\Omega = \prod_{a\in[K]} \Omega_a^T$. Fix an arm j satisfying the inequality above. We will prove that $P_j[Y_T = j] \leq 1/2$. After this, we can easily finish the proof. Consider the sample space which j is played only $\min(T, 24T/K)$ times $\Omega^* = \Omega_j^m \times \prod_{a\neq j} \Omega_a^T$. Then, for each problem I_l, define the distribution P_l^* on Ω^* as $P_l^*(A) = P(A|I_l)\ \forall A \subset \Omega^*$.

Using KL-divergence argument, for each $A \subset \Omega^*$, we have

$$\begin{aligned} 2(P_0^*(A) - P_j^*(A))^2 &\leq KL(P_0^*, P_j^*) && \text{(Pinsker's Inequality)} \\ &= \sum_{a}\sum_{t=1}^{T} KL(P_0^{a,t}, P_j^{a,t}) && \text{(Chain Rule)} \\ &= \sum_{a\neq j}\sum_{t=1}^{T} KL(P_0^{a,t}, P_j^{a,t}) + \sum_{t=1}^{m} KL(P_0^{j,t}, P_j^{j,t}) && \\ &\leq m + 2\epsilon^2 && \text{(Theorem 2.d)} \end{aligned}$$

Hence, if $T \leq \frac{cK}{\epsilon^2}$ with small c, we have that $|P_0^*(A) - P_j^*(A)| \leq \epsilon\sqrt{m} < \frac{1}{8}$ for all $A \subset \Omega^*$. To use this, we check first that A is indeed in Ω^*. In particular, we cannot use $A = \{y_T = j\}$ because this event can depend on more than m samples of j. Instead, we divide A into events $A = \{y_T = jT_j \leq m\}$ and $A' = \{T_j > m\}$. Recall that we see these event as subsets of $\{0,1\}^{K \times T}$. Note that $A, A' \subset \Omega^*$. Finally,

$$\begin{aligned} P_j(A) &\geq \frac{1}{8} + P_0(A) \\ &\geq \frac{1}{8} + P_0(y_t = j) \\ &\geq \frac{1}{4} \end{aligned}$$

$$\begin{aligned} P_j(A') &\geq \frac{1}{8} + P_0(A') \\ &\geq \frac{1}{4} \end{aligned}$$

$$P_j(Y_T = j) \leq P_j^*(Y_T = jT_j \leq m) + P_j^*(T_j > m) = P_j(A) + P_j(A') \leq \frac{1}{2}.$$

Hence, the fact is proven, and it finishes our proof.

6 Non-adaptive Exploration

The information theoretic approach implies stronger bounds for non-adaptive exploration. Firstly, the following holds.

Theorem 4. *For any non-adaptive exploration, if we fix T and K with $K < T$. Then, there exists a problem instance such that $\mathbb{E}[R(T)] \geq \Omega(T^{2/3}K^{1/3})$.*

Also, we can rule out logarithmic upper bound such as the one in the Successive Elimination. The statement is more nuanced, with a sense that the algorithm must not perform very poorly in the worst case.

Theorem 5. *Using the setup on theorem 4, if $\mathbb{E}[R(T)] \leq CT^\gamma$ for all problem instances, with $2/3 \geq \gamma < 1$. Then, for any problem instance, a random arms satisfies that $\mathbb{E}[R(T)] \geq \Omega(C^{-2}T^{2-2\gamma}\sum_a \Delta(a))$.*

In particular, if an algorithm has regret $\mathbb{E}[R(T)] \leq O(T^{2/3}K^{1/3} \operatorname{polylog} T^{2/3}K^{1/3})$ over all problems, like Explore-First or Epsilon-Greedy algorithm, this algorithm has a similar regret for any problem instance, if the arms are randomly permuted. Note that $\mathbb{E}[R(T)] \leq \Gamma(\Delta T^{2/3}K^{1/3} \operatorname{polylog} \Delta T^{2/3}K^{1/3})$, with Δ is a minimum gap. This can be seen by taking $C = O(K^{1/3} \operatorname{polylog} K^{1/3})$.

7 Instance-dependent Lower Bounds

The other fundamental lower bounds states that $\Omega(\log T)$ regret with an instance-dependent constant and applies to every problem instance. The lower bound can be used to combine the $\log T$ upper bound in UCB1 and Successive Elimination algorithms.

We first focus on 0-1 rewards. For a problem instance, we want to check the growth of $\mathbb{E}[R(t)]$ with respect of t. First, we have this result.

Theorem 6. *No algorithm can have regret $\mathbb{E}[R(t)] = o(c_I \log t)$ for all problem instance I, for some constant c_I which depends on I but not t.*

Hence, we have a guarantee that there is a problem instance which an algorithm has a high regret. We now want to have a stronger lower bound which ensures high regret for every problem instance. However, this is impossible. To see this, consider an algorithm that always picks arm 1. If arm 1 is optimal, this algorithm has regret 0 for this problem instance. To prevent this counter example, we require an algorithm to perform good enough across every problem instance.

Theorem 7. *For a fixed K, consider an algorithm such that $\mathbb{E}[R(t)] \leq O(C_{I,\alpha}t^\alpha), \forall I, \alpha > 0$. Here, $C_{I,\alpha}$ depends on I, α, but not on t. Now, fix a problem instance I. For this I, there exists t_0 such that $\forall t \geq t_0, \mathbb{E}[R(t)] \geq C_I \ln t$, with C_I depends on I but not t.*

The theorem above can be sharpened to specify how C_I can be chosen. Recall that $\Delta(a) = \mu^* - \mu(a)$.

Theorem 8. *Keep the setup from the previous theorem. For any I and algorithm that satisfies the previous theorem,*

1. *The bound works with $C_I = \sum_{\Delta(a)>0} \frac{\mu^*(1-\mu^*)}{\Delta(a)}$.*

2. *For each $\epsilon > 0$, the bound holds with $C_I = \sum_{\Delta(a)>0} \frac{\Delta(a)}{KL(\mu(a),\mu^*)} - \epsilon$.*

8 Literature Review and Discussion

The $\Omega(\sqrt{KT})$ lower bound on regret is from Auer et. al. [3]. The theory of KL-divergence is a branch of Information Theory, which came from textbooks such as the book by Cover and Thomas [10]. The outlines of the technical details in this report is based on Kleinberg's lecture notes in 2007 [2]. However, the proof in this paper is notably simpler, where we replace the KL chain rule with the special case of independent distributions. This special case is much easier to formulate and use. The proof of lemma 3 for $K > 2$ is altered accordingly. In particular, our definition of reduced sample space Ω^* with only small number of samples from bad arm j and use KL-divergence argument to define the events.

Lower bounds for the non-adaptive exploration has been studied a lot. The first version is done by Babaioff et. al. (2014) [4]. In this paper, the version of non-adaptive explanation and the lower bounds are similar, but has different technical setting.

The logarithmic lower bound is due to Lai and Robbins [19], with proof using the KL-divergence technique. It is also found on the paper of Bubeck and Cesa-Bianchi (2012) [8].

While we have the both bounds for the basic version of multi-armed bandits, it is not enough for other versions. Some bandit problem has constraints such as Lipschitzness or linearity, and the construction of the lower bound need to take note of these. Such lower bounds do not depend on the number of actions. In other problem, the space complexity of algorithm is also a consideration. Hence, much stronger bounds may occur.

Some notable results are below.

1. In dynamic pricing ([15], [1]) and lipschitz bandits ([16] , [23], [17] . [18]), see chapter 4 of [24].

2. For linear bandits ([11], [20] , [22]), see chapter 7 of [24].

3. For pay-per-click ad auctions ([4], [12]). Ad auctions are parametized by click probabilities for any ad, which are unknown at first but can be learned over time by a bandit algorithm. The algorithm is constrained to be computable by the ad placer.

4. For dynamic pricing with limited supply ([1], [7]) and bandits with resource constraints ([5], [13] , [21]), see chapter 10 of [24].

5. For best arm identification ([14], [9]).

Some lower bounds in the literature are derived from first principles, such as the lower bounds in [15], [16], [6], [5]. Other bounds are derived by reduction, such as [1], [17].

References

[1]

[2] Cs683: Learning, games, and electronic markets, a class at cornell university.

[3] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 05 2002.

[4] Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. Characterizing truthful multi-armed bandit mechanisms. *SIAM J. Comput.*, 43:194–230, 2014.

[5] Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. *J. ACM*, 65(3), March 2018.

[6] Gábor Bartók, Dean Foster, Dávid Pál, Alexander Rakhlin, and Csaba Szepesvári. Partial monitoring—classification, regret bounds, and algorithms. *Mathematics of Operations Research*, 39:967–997, 11 2014.

[7] Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57:1407–1420, 12 2009.

[8] Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

[9] Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem, 2016.

[10] Cover and Thomas. *Elements of Information Theory, 2nd Edition.*

[11] Varsha Dani, Thomas P. Hayes, and Sham M. Kakade. Stochastic linear optimization under bandit feedback. In *COLT*, 2008.

[12] Nikhil R. Devanur and Sham M. Kakade. The price of truthfulness for pay-per-click auctions. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, EC '09, page 99–106, New York, NY, USA, 2009. Association for Computing Machinery.

[13] Nicole Immorlica, Karthik Abinav Sankararaman, Robert Schapire, and Aleksandrs Slivkins. Adversarial bandits with knapsacks. *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, Nov 2019.

[14] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best arm identification in multi-armed bandit models, 2016.

[15] R. Kleinberg and T. Leighton. The value of knowing a demand curve: bounds on regret for online posted-price auctions. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 594–605, 2003.

[16] Robert Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. 01 2004.

[17] Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Bandits and experts in metric spaces. *Journal of the ACM*, 66, 12 2013.

[18] Akshay Krishnamurthy, John Langford, Aleksandrs Slivkins, and Chicheng Zhang. Contextual bandits with continuous actions: Smoothing, zooming, and adapting. *Journal of Machine Learning Research*, 21(137):1–45, 2020.

[19] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules*1. *Advances in Applied Mathematics - ADVAN APPL MATH*, 6:4–22, 03 1985.

[20] Paat Rusmevichientong and John N. Tsitsiklis. Linearly parameterized bandits. *Math. Oper. Res.*, 35(2):395–411, May 2010.

[21] Karthik Abinav Sankararaman and Aleksandrs Slivkins. Bandits with knapsacks beyond the worst-case, 2020.

[22] Ohad Shamir. On the complexity of bandit linear optimization. In Peter Grünwald, Elad Hazan, and Satyen Kale, editors, *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 1523–1551, Paris, France, 03–06 Jul 2015. PMLR.

[23] Aleksandrs Slivkins. Contextual bandits with similarity information. *J. Mach. Learn. Res.*, 15(1):2533–2568, January 2014.

[24] Aleksandrs Slivkins. Introduction to multi-armed bandits. *CoRR*, abs/1904.07272, 2019.