A Model for Optimizing Recalculation Schedules to Minimize Regret

Bethany Austhof(✉) and Lev Reyzin(✉)

Department of Mathematics, Statistics, and Computer Science,
University of Illinois at Chicago, Chicago, IL 60607, USA
{bausth2,lreyzin}@uic.edu

Abstract. In this paper we analyze online problems from the perspective of when to switch solutions when the cost is of doing so is high – we call such a solution change a "recalculation." We analyze this problem under the assumption we have algorithms that achieve per-round regret (which can be otherwise thought of as point-wise error, or other well-studied quantities) of the form $O(1/t^{\varepsilon})$ after seeing t data-points. We study schedules with a constant number and an increasing number of recalculations in the total number of datapoints, and we examine when achieving optimal cumulative regret is possible.

Keywords: Regret minimization · Recalculation schedules · Online learning

1 Introduction

In many online settings, one faces the problem on when to recalculate a given solution based on new data. This phenomenon occurs in various settings by different names: in the bandit framework, "sticky" decisions prevent the learner from switching arms without paying a cost [6,10,12], in the online clustering setting every time one switchs solutions, one may need to pay for new centers [3,4,8,14], in facility location, adding a new facility adds cost [7,13].

The specific details of different problems require the analysis of their respective structures and often have their own involved solutions. Here, we attempt to take a broader view, and we adopt the learning language of additive regret (instead of "competitive ratios," etc. from the streaming literature, though the results apply just as well there).

We assume we have a black-box algorithm $\mathcal{A}$ that produces an estimator with expected **per-round regret** of $\mathcal{R}_t(x) \in [0,1]$ on a new data-point x when given samples $S_t = \{x_1 \dots x_t\}$ and when point x is drawn from the same distribution as $x_1, \dots x_t$, which we will generically refer to as D. Given the assumption above, we will drop the argument from the function $\mathcal{R}$. We also define $\mathcal{R}_0 = 1$, which means that until samples are given to $\mathcal{A}$, full regret is suffered by its (lack of) solution.

R. P. Barneva et al. (Eds.): IWCIA 2024/ISAIM 2024, LNCS 14494, pp. 1–9, 2024.
https://doi.org/10.1007/978-3-031-63735-3_1

We call giving samples $\{x_1, \dots x_t\}$ to $\mathcal{A}$ a **recalculation** at round t. We work in the setting where calls to $\mathcal{A}$ are expensive and need to be traded off against the benefits to improving regret. A call to $\mathcal{A}$ involves recalculating, and possibly committing to, a new solution to a given problem. For example, in the case of facility location, it involves the costly opening of new facilities (or the moving of old ones).

Hence, we are interested in understanding the optimal recalculation strategy as minimize cumulative regret for various budgets on calls to $\mathcal{A}$. Given a strategy that recalculates at rounds $C = \{t_1, t_2, \dots\}$, where each t_i represents a sequential point drawn from the distribution, the **expected cumulative regret** $\mathcal{R}^T$ would be

$$\mathcal{R}^T = \sum_{i=1}^{T} \max_{t \in C \text{ s.t. } t<i} \mathcal{R}_t \tag{1}$$

In this paper, we will consider per-round regret guarantees of the form $t^{-\varepsilon}$

$$\mathcal{R}_t = O(1/t^{\varepsilon}), \tag{2}$$

which is known to be the asymptotically optimal regret for many problems in bandit and online learning in the stochastic setting [2]. Rephrasing this in our terminology, we have that for the recalculation schedule $C = \{1, \dots, T\}$, the expected cumulative regret for $0 \le \varepsilon < 1$

$$\mathcal{R}^T = \sum_{t=1}^{T} \mathcal{R}_{t-1} = O\left(\sum_{t=1}^{T} \frac{1}{t^{\varepsilon}}\right) = O(T^{1-\varepsilon}),$$

So even when recalculating at every round, a regret guarantee smaller than $O\left(T^{1-\varepsilon}\right)$ is not possible, and we therefore call this rate **optimal**. This is the quantity we will compare to while attempting to do many fewer recalculations.

To simplify notation further, we will henceforth use $\mathcal{R}$ for expected regret. Therefore, all of our results will be on bounding regret in expectation.

Before proceeding, we note that another interesting setting $\varepsilon = 1$ for $\mathcal{R}_t = O(1/t)$, which is, for example, relevant for problems of estimation of parameters to minimize mean squared error (MSE), where the squared error of the empirical average of t samples versus the true estimate scales as σ^2/t [5]. Here, we would have

$$\mathcal{R}^T = O\left(\sum_{t=1}^{T} \frac{1}{t}\right) = O(\log T).$$

2 Warm up

We begin by analyzing the special case of $\mathcal{R}_t = 1/\sqrt{t}$ The argument in the previous section shows that the optimum regret in this situation is $O(\sqrt{T})$. However, this was the setting in which we recalculated every time we drew a point, which in the online setting isn't always practicable. Thus, inducing a trade-off between the amount of times we recalculate and the regret formed in our calculation.

Proposition 1. *Let D be a probability distribution from which T data points are sampled online and let $\mathcal{R}_t = O(1/\sqrt{t})$. Using one recalculation, one can achieve an expected regret of $O(T^{2/3})$.*

Proof. We first begin by calculating the optimal expected regret of calculating the algorithm just once. Let c be a constant such that $0 \leq c \leq 1$. We have that the expected regret must be:

$$\sum_{t=1}^{T^c} \mathcal{R}_0 + \sum_{t=T^c+1}^{T} \mathcal{R}_{T^c} = O\left(T^c + T^{1-c/2}\right).$$

We note that this is just a simple optimization of the right hand side, so we equate the exponents and get $c = 2/3$. This recovers the well-known bound of ε-first, (ε-greedy, and epoch-greedy) sampling [11,15,16].

3 Uniform Recalculations

One naive idea is to create a sampling schedule that solves our problem is to sample uniformly, so after observing a set ℓ amount of points we sample again and continue in this manner. However, we find this to at best require $O(T^\varepsilon)$ recalculations to get asymptotically optimum regret.

Lemma 1. *Let D be a probability distribution from which T data points are sampled online and let $\mathcal{R}_t = O(1/t^\varepsilon)$ for $0 \leq \varepsilon < 1$. If we recalculate uniformly after observing every ℓ points (and therefore recalculate T/ℓ times), then we incur an expected regret of $O\left(\ell + T^{1-\varepsilon}\right)$.*

Proof. We calculate the expected regret for recalculating after every ℓ points, we carry out the calculation by using the general formula for a uniform schedule.

$$\begin{aligned}\sum_{i=0}^{T/\ell}\sum_{t=1}^{\ell} \mathcal{R}_{\ell i} &= \ell + O\left(\sum_{i=1}^{T/\ell}\sum_{t=1}^{\ell} \frac{1}{(i\ell)^\varepsilon}\right) \\ &= \ell + O\left(\sum_{i=1}^{T/\ell} \frac{\ell^{1-\varepsilon}}{i^\varepsilon}\right) \\ &= O\left(\ell + T^{1-\varepsilon}\right),\end{aligned}$$

which completes the proof. □

We note briefly, that this proof holds in the $\varepsilon = 1$ case, we just get the result $\mathcal{R} = O(\ell + \log(T))$.

As the lemma above shows, a constant number of recalculations would imply linearly many recalculations between rounds and therefore linear regret. Therefore, we need a smarter recalculation strategy if we want to recalculate only a constant number of times. This is presented in the following section.

We see in the previous section that uniform recalculation scheduling fails to guarantee small expected regret with a small number of recalculations. So we pivot to non-uniform schedules. Taking, first, a look at what we can do with a constant number of recalculations.

4 Constant Number of Recalculations

We now consider the case when we are allowed a constant number of recalculations, but they need not be uniformly spaced.

Proposition 2. *Let D be a probability distribution from which T data points are sampled online and let $\mathcal{R}_t = O(1/t^{\varepsilon})$ for $0 \leq \varepsilon \leq 1$. Using one recalculation, one can achieve an expected cumulative regret of*

$$\mathcal{R}^T = O\left(T^{1/(1+\varepsilon)}\right).$$

Proof. Consider recalculating after T^c rounds for $c = \frac{1}{1+\varepsilon}$. We suffer $O\left(T^{1/(1+\varepsilon)}\right)$ cumulative regret on those rounds. For the remaining $T - T^c \leq T$ rounds, we suffer at most

$$\begin{aligned} O\left(T/T^{c\varepsilon}\right) =& O\left(T^{1-c\varepsilon}\right) \\ =& O\left(T^{1-\varepsilon/(1+\varepsilon)}\right) \\ =& O\left(T^{1/(1+\varepsilon)}\right) \end{aligned}$$

regret. □

Proposition 3. *Let D be a probability distribution from which T data points are sampled online and let $\mathcal{R}_t = O(1/t^{\varepsilon})$ for $0 \leq \varepsilon \leq 1$. Using two recalculations, one can achieve an expected cumulative regret of*

$$\mathcal{R}^T = O\left(T^{1/(1+\varepsilon+\varepsilon^2)}\right).$$

Proof. Consider recalculating after T^{c_1} and T^{c_2} for $0 \leq c_1 \leq c_2 \leq 1$. Extending the argument in Lemma 2, we incur a total regret of $\mathcal{R}^T \leq T^{c_1} + T^{c_2 - c_1\varepsilon} + T^{1-c_2\varepsilon}$. Equalizing the three terms we get the equations

$$c_1 = c_2 - c_1\varepsilon \text{ and } c_1 = 1 - c_2\varepsilon.$$

Solving the above yields $c_1 = \frac{1}{1+\varepsilon+\varepsilon^2}$ and yields an expected cumulative regret of $O(T^{c_1})$. □

Theorem 1. *Let D be a probability distribution from which T data points are sampled online and let $\mathcal{R}_t = O(1/t^{\varepsilon})$ for $0 \leq \varepsilon < 1$. Using n recalculations, one can achieve an expected cumulative regret of*

$$\mathcal{R}^T = O\left(nT^{\frac{1-\varepsilon}{1-\varepsilon^{n+1}}}\right).$$

Proof. Generalizing from the above, we divide into n recalculations and bound $\mathcal{R}^T \leq \sum_{i=1}^{n+1} T^{c_i - c_{i-1}\varepsilon}$ setting $c_0 = 0$ and $c_{n+1} = 1$.

We argue that solving for the system of equations leads to the following recursive formula:

$$c_j = c_{j+1} \frac{\sum_{i=0}^{j-1} \varepsilon^i}{\sum_{i=0}^{j} \varepsilon^i}.$$

We proceed inductively, we first see that $c_1 = c_2 - \varepsilon c_1$, which gets us $c_1 = c_2(1+\varepsilon)^{-1}$, as desired. Moving on, we assume this holds for c_{j-1} and solve for c_j.

$$\begin{aligned} c_j - \varepsilon c_{j-1} &= c_{j+1} - \varepsilon c_j \\ c_j(1+\varepsilon) &= c_{j+1} + \varepsilon c_{j-1} \\ c_j(1+\varepsilon) &= c_{j+1} + c_j \varepsilon \frac{\sum_{i=0}^{j-2} \varepsilon^i}{\sum_{i=0}^{j-1} \varepsilon^i} \\ c_j \left(1 + \varepsilon - \frac{\sum_{i=1}^{j-1} \varepsilon^i}{\sum_{i=0}^{j-1} \varepsilon^i}\right) &= c_{j+1} \\ c_j \left(1 + \varepsilon - 1 + \frac{1}{\sum_{i=0}^{j-1} \varepsilon^i}\right) &= c_{j+1} \\ c_j \frac{\sum_{i=0}^{j} \varepsilon^i}{\sum_{i=0}^{j-1} \varepsilon^i} &= c_{j+1} \\ c_j = c_{j+1} \frac{\sum_{i=0}^{j-1} \varepsilon^i}{\sum_{i=0}^{j} \varepsilon^i}&. \end{aligned}$$

So we have a recursive formula and now we can see that if we terminate after n recalculations we have a telescoping product, and we have that

$$c_1 = \frac{1}{\sum_{i=0}^{n} \varepsilon^i} = \frac{1-\varepsilon}{1-\varepsilon^{n+1}}.$$

Since the $n+1$ phases have equal regret, the total regret is $O(nT_1^c)$, producing the bound of $\mathcal{R}^T = O\left(nT^{\frac{1-\varepsilon}{1-\varepsilon^{n+1}}}\right)$. This bound tends to T^ε for arbitrarily high constant n. □

We observe that this proof does hold for $\varepsilon = 1$, we just leave c_1 unsimplified and have that the expected cumulative regret in this case is $O(nT^{1/n})$.

5 Schedules with Increasing Recalculations in T

After seeing some extreme examples and results on constants, we wish to discover how to improve the amount of times needed to recalculate while still maintaining the optimum expected regret. By employing a "doubling trick" argument, we see that we can easily reduce to $\log(T)$ recalculations. We note that other settings have also been studied that achieve optimal regret using logarithmically many policy "switches," [1,9].

Lemma 2. *Let D be a probability distribution from which T points are sampled online and let $\mathcal{R}_t = O(1/t^\varepsilon)$ for $0 \leq \varepsilon < 1$. There exists a sequence of* $\log(T)$ *recalculations for which we can achieve optimal expected regret of $O(T^{1-\varepsilon})$.*

Proof. Let $T = 2^m$, with m sufficiently large, and let $\mathcal{L} = \{2^0, 2^1, \ldots 2^{m-1}\}$. If we recalculate the algorithm after seeing each point in $\mathcal{L}$, then we have the following expected cost:

$$\begin{aligned}
\sum_{i=0}^{m} \sum_{t=1}^{2^i} \mathcal{R} &\leq O\left(\sum_{i=0}^{m} \sum_{t=1}^{2^i} \left(\frac{1}{2^{(i-1)\varepsilon}}\right)\right) \\
&= O\left(\sum_{i=0}^{m} \left(2^{i-(i-1)\varepsilon}\right)\right) \\
&= O\left(\sum_{i=0}^{m} \left(2^{i(1-\varepsilon)}\right)\right) \\
&= O\left(2^{(1-\varepsilon)m}\right) \\
&= O\left(T^{(1-\varepsilon)}\right).
\end{aligned}$$

This establishes that we can recalculate $\log(T)$ times and achieve optimal expected regret of $T^{1-\varepsilon}$. □

We observe that if the regret is of the form $\mathcal{R}_t = O(1/t)$ (the $\varepsilon = 1$ case), then the regret will be bounded by $O(\log T)$ in the above schedule.

We've established that we can achieve optimal expected regret with only $\log(T)$ recalculations, we go on to establish that we cannot improve this to $\log\log(T)$, by showing that the expected regret when taking $\log\log(T)$ recalculations is $O(\log\log(T)\sqrt{T})$.

Lemma 3. *Let D be a probability distribution from which T are sampled online and let $\mathcal{R}_t = O(1/t^\varepsilon)$ for $0 < \varepsilon < 1$, where $c = 1/\varepsilon$. There exists a sequence of* $\log\log(T)$ *recalculations such that we can achieve an expected regret of* $O(\log\log(T)T^{1-\varepsilon})$.

Proof. Let $T = c^{c^m}$, with m sufficiently large. Let $\mathcal{L} = \{\ell_0, \ell_1, \ldots \ell_m\}$. We recalculate after seeing each point in $\mathcal{L}$.

We wish to find a schedule that will cover all T with $\log\log(T)$ recalculations. We do this with the following schedule: at epoch i, once we have seen a total of $T^{1-c^{-i}}$ points we recalculate. We show that if we cover half of T at epoch $\log\log(T)$, then in the next epoch we will have seen all of T. So we establish this recalculation schedule achieves this. Let y be the amount of times needed to see half of T under the above scheduling scheme, so:

$$\begin{aligned}
1/c &= T^{-c^{-y}} \\
-\log_c(c) &= (-c^{-y})\log_c(T) \\
\frac{1}{\log_c(T)} &= c^{-y} \\
y &= \log_c \log_c(T).
\end{aligned}$$

We note that we found the exact point at which we would have half, and that this makes $\log\log(T)$ the minimum amount of rounds needed to cover all of T with this particular schedule. We can be sure of this since there were no assumptions made on the size of T, so we couldn't reduce to $\log\log\log(T)$ rounds. So now, we can sum our per-round cost over the amount of rounds we must run:

$$\begin{aligned}
\sum_{i=0}^{m}\sum_{t=1}^{T^{1-c^{-i}}} \mathcal{R}_{T^{1-c^{-(i-1)}}} &\leq \sum_{i=1}^{m}\sum_{t=1}^{T^{1-c^{-i}}} T^{-\frac{c^{i-1}-1}{c^{i}}} \\
&= \sum_{i=1}^{m} T^{1-c^{-i}-c^{-1}+c^{-i}} \\
&= \log\log(T)T^{1-c^{-1}} \\
&= \log\log(T)T^{1-\varepsilon}.
\end{aligned}$$

Computing the inner sum, we see that at each round, we will have $O(T^{1-\varepsilon})$ expected regret. □

We note this proof does not extend to the $\varepsilon = 1$ case, since the argument hinges on the expected regret being bounded by $O(T^{1-\varepsilon})$.

Now we establish a lower bound, in particular that we cannot achieve an expected regret of $O(T^{1-\varepsilon})$ with $\log\log(T)$ recalculations.

Lemma 4. *Any schedule that performs $O(\log\log T)$ recalculations using an algorithm that suffers per-round regret of $\mathcal{R} = O(1/t^{\varepsilon})$, $0 < \varepsilon < 1$, must suffer $\omega(T^{1-\varepsilon})$ cumulative regret.*

Proof. Given $\log\log(T)$ recalculations, we are lower-bounded by a regret of $O(T^{1-\varepsilon})$. Specifically, we show that we cannot achieve a regret of $O(T^{1-\varepsilon})$ given $\log\log(T)$ recalculations. So assume to the contrary, that there exists a schedule that does this. Now, to achieve this clearly the regret incurred at each recalculation has to be $O(T^{1-\varepsilon})$. Based on this we craft an inductive argument on the size of each recalculation round. At round $i-1$, the size, $T^{c_{i-1}}$ must be $O(T^{1-\varepsilon^{i}})$. As a base case, we see that $T^{c_0} = O(T^{1-\varepsilon})$. We move onto the inductive step: We note that the regret for round i is

$$\begin{aligned}
T^{c_i}T^{-\varepsilon c_{i-1}} &= O(T^{1-\varepsilon}) \\
T^{c_i} &= O(T^{1-\varepsilon+\varepsilon c_{i-1}}).
\end{aligned}$$

This implies,

$$c_i \leq 1 - \varepsilon + \varepsilon(1 - \varepsilon^{i-2})$$
$$c_i \leq 1 - \varepsilon^{i-1}.$$

This, then results in the total portion of T witnessed as $\sum_{n=1}^{\log\log(T)} T^{1-\varepsilon^n}$. Now, since

$$\sum_{n=1}^{\log\log(T)} T^{1-\varepsilon^n} \leq \log\log(T) T^{1-\varepsilon^{\log\log(T)}} = o(T),$$

we have failed to witness all of T in $\log\log(T)$ rounds and have reached a contradiction. □

We note this proof does not extend to the $\varepsilon = 1$ case, since the argument hinges on the expected regret being bounded by $O(T^{1-\varepsilon})$.

Acknowledgments. This work was supported in part by National Science Foundation grant ECCS-2217023. We thank the anonymous reviewers of this paper for their helpful comments.

References

1. Abbasi-Yadkori, Y., Pál, D., Szepesvári, C.: Improved algorithms for linear stochastic bandits. In: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F.C.N., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held Granada, Spain, 12–14 December 2011, pp. 2312–2320 (2011)
2. Auer, P., Cesa-Bianchi, N., Freund, Y., Schapire, R.E.: The nonstochastic multiarmed bandit problem. SIAM J. Comput. **32**(1), 48–77 (2002)
3. Bhattacharjee, R., Imola, J., Moshkovitz, M., Dasgupta, S.: Online k-means clustering on arbitrary data streams. In: Agrawal, S., Orabona, F. (eds.) International Conference on Algorithmic Learning Theory, Singapore, 20–23 February 2023. Proceedings of Machine Learning Research, vol. 201, pp. 204–236. PMLR (2023)
4. Bhattacharjee, R., Moshkovitz, M.: No-substitution k-means clustering with adversarial order. In: Feldman, V., Ligett, K., Sabato, S. (eds.) Algorithmic Learning Theory, Virtual Conference, Worldwide. Proceedings of Machine Learning Research, 16–19 March 2021, vol. 132, pp. 345–366. PMLR (2021)
5. DeGroot, M.H.: Probability and Statistics (1986)
6. Dekel, O., Ding, J., Koren, T., Peres, Y.: Bandits with switching costs: $T^{2/3}$ regret. In: Shmoys, D.B. (ed.) Symposium on Theory of Computing, STOC 2014, New York, NY, USA, 31 May–03 June 2014, pp. 459–467. ACM (2014)
7. Fotakis, D.: On the competitive ratio for online facility location. Algorithmica **50**(1), 1–57 (2008)
8. Hess, T., Moshkovitz, M., Sabato, S.: A constant approximation algorithm for sequential random-order no-substitution k-median clustering. In: Ranzato, M., Beygelzimer, A., Dauphin, Y.N., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, 6–14 December 2021, Virtual, pp. 3298–3308 (2021)

9. Jaksch, T., Ortner, R., Auer, P.: Near-optimal regret bounds for reinforcement learning. J. Mach. Learn. Res. **11**, 1563–1600 (2010)
10. Kash, I.A., Reyzin, L., Yu, Z.: Slowly changing adversarial bandit algorithms are provably efficient for discounted mdps. CoRR arxiv:2205.09056 (2022)
11. Langford, J., Zhang, T.: The epoch-greedy algorithm for contextual multi-armed bandits. Adv. Neural. Inf. Process. Syst. **20**(1), 96–1 (2007)
12. Machado, M.C., Bellemare, M.G., Talvitie, E., Veness, J., Hausknecht, M.J., Bowling, M.: Revisiting the arcade learning environment: evaluation protocols and open problems for general agents (extended abstract). In: Lang, J. (ed.) Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, Stockholm, Sweden, 13–19 July 2018, pp. 5573–5577. ijcai.org (2018)
13. Meyerson, A.: Online facility location. In: 42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, Las Vegas, Nevada, USA, 14–17 October 2001, pp. 426–431. IEEE Computer Society (2001)
14. Moshkovitz, M.: Unexpected effects of online no-substitution k-means clustering. In: Feldman, V., Ligett, K., Sabato, S. (eds.) Algorithmic Learning Theory, Virtual Conference, Worldwide. Proceedings of Machine Learning Research, 16–19 March 2021, vol. 132, pp. 892–930. PMLR (2021)
15. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT press, Cambridge (2018)
16. Tran-Thanh, L., Chapman, A., De Cote, E.M., Rogers, A., Jennings, N.R.: Epsilon–first policies for budget–limited multi-armed bandits. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 24, pp. 1211–1216 (2010)