Invited Open Problem: Online Optimization of Piecewise-Lipschitz Functions with Applications to Data-Driven Algorithm Design

Maria-Florina Balcan NINAMF@CS.CMU.EDU
Carnegie Mellon University

Wesley Pegden WES@MATH.CMU.EDU
Carnegie Mellon University

Dravyansh Sharma DRAVY@TTIC.EDU
Toyota Technological Institute at Chicago, Northwestern University

Editors: Steve Hanneke and Tor Lattimore

Abstract

Classical online optimization theory focuses on regret guarantees for convex Lipschitz functions. However, online optimization problems motivated by machine learning for algorithm design fall outside this regime, since typically an algorithm's performance as a function of its hyperparameters is a highly volatile function. This has inspired recent work on online optimization of piecewise-Lipschitz functions with complex transition boundaries. We provide open questions in this direction. Resolving these questions would advance the learning-theoretic foundation for adaptive algorithm design by clarifying when desirable sublinear regret guarantees are possible for learning the algorithms from online problem instances.

1. Background and Motivation

Modern algorithms depend critically on tunable parameters whose optimal values vary across applications and over time. Traditionally, these parameters are selected via offline cross-validation or fixed heuristics that can be brittle and fail to adapt to changing data distributions. Online data-driven algorithm design instead adapts parameters on the fly with provable guarantees, but the key obstacle is that an algorithm's performance as a function of its parameters is highly non-smooth: infinitesimal parameter changes can often affect combinatorial decisions—like branching heuristics or cluster merges—producing discontinuous, non-concave utility (or non-convex loss) landscapes that place the problem outside classical online convex optimization. Typical algorithm design problems admit no sublinear regret guarantee against adversarial inputs without further structure (with a few exceptions, e.g., Balcan and Beyhaghi 2024). The dispersion framework of Balcan et al. (2018) and its extensions (Balcan et al., 2020b,c; Balcan and Sharma, 2021; Balcan et al., 2025a) show that strong regret guarantees become possible once discontinuities are suitably well-behaved. However, the current theory still relies on problem-specific structural arguments and often incurs prohibitively large dependence on key parameters (e.g., exponential in the polynomial degree of transition boundaries)—in sharp contrast to the statistical (i.i.d.) learning setting, where sample complexity is typically polynomial in both input complexity and the number of algorithm parameters (Balcan et al., 2021; Bartlett et al., 2022), even for complex settings like deep neural network hyperparameters (Balcan et al., 2025b). Understanding when learning with good regret guarantees is possible in the online setting remains a central challenge in the theory of online data-driven algorithm design.

Problem Setting. Let $\Theta \subset \mathbb{R}^p$ be a compact parameter space. At each round $t = 1, \ldots, T$, the adversary[1] chooses a function $u_t : \Theta \to [0, H]$ and the learner selects algorithm parameter $\theta_t \in \Theta$, and experiences utility $u_t(\theta_t)$ that measures the algorithm's performance (e.g., optimization objective, running time, memory, converted from loss to utility if needed). In the static full information setting, the learner then receives the feedback $u_t(.)$ for all $\theta \in \Theta$ in each round, and the performance is measured after T rounds using regret against the best algorithm parameter in hindsight. A common setting in data-driven algorithm design (Balcan, 2020) involves u_t that are piecewise-Lipschitz, with discontinuities defined by threshold conditions $\text{Disc}(u_t) \subseteq \cup_{j=1}^{m_t}\{\theta : \phi_{t,j}(\theta) = 0\}$, where $\{\phi_{t,j}\}_j$ are a collection of m_t boundary functions, e.g., multivariate polynomial functions.

This structure arises naturally in many algorithmic settings. For example, in hierarchical clustering one first builds a binary tree by sequentially merging pairs of clusters using a linkage criterion (single, complete, and average linkage induce different orderings of candidate cluster-pair merges based on inter-cluster distances), and then prunes the tree using dynamic programming. Recent work (Balcan et al., 2017, 2020a, 2025a) proposes several natural algorithm families that interpolate popular linkage algorithms and distance metrics, and the merge tree changes across polynomial or exponential boundaries in the parameter space. Other examples include tuning regularization parameters in regression (Balcan et al., 2022), graph algorithms in semi-supervised learning (Balcan and Sharma, 2021; Sharma and Jones, 2023; Du et al., 2025) and setting multi-armed bandit algorithm hyperparameters (Sharma and Suggala, 2025; Blum et al., 2026).

In all these examples, the utility can be expressed as a piecewise function whose boundaries are defined by multivariate polynomial equations. Classical online optimization theory fails to apply in this setting, and in the worst-case, linear regret is typically unavoidable, even in one dimension. Cohen-Addad and Kanade (2017) show sufficient conditions to overcome this in one-dimensional piecewise constant functions (with perturbed discontinuities). Prior work (Balcan et al., 2018, 2020b,c, 2025a) gives a powerful recipe for online learning algorithm parameters with piecewise-Lipschitz utility functions under sufficient "smoothness" conditions called *dispersion* (Definition 1).

2. Open Questions

A sufficient condition for online optimization of piecewise-structured functions is given by β-*dispersion* (Balcan et al., 2018, 2020b,c). Informally, a sequence $(u_1, \ldots, u_T)$ is β-dispersed if the number of discontinuities in any ε-ball, for all $\varepsilon \geq T^{-\beta}$, scales roughly as $O(\varepsilon T)$.

Definition 1 *A deterministic sequence of utility functions $(u_t)_{t\geq 1}$ is β-dispersed for the Lipschitz constant L if, for all $T \geq 1$ and for all $\varepsilon \geq T^{-\beta}$, at most $\tilde{O}(\varepsilon T)$ functions[2] are not L-Lipschitz in any ℓ_2-ball $\mathbf{B}(\theta, \varepsilon)$ of radius ε contained in the parameter domain Θ.*

$$\max_{\theta\in\Theta} \left|\{t \mid u_t \text{ not } L\text{-Lipschitz in } \mathbf{B}(\theta, \varepsilon)\}\right| \leq \tilde{O}(\varepsilon T).$$

Further, the random process generating them is said to be β-dispersed if the expectation of the displayed maximum satisfies the same bound.

Intuitively, dispersion requires that discontinuities are not overly concentrated in small regions of the parameter space, and β-dispersion implies existence of online learners with $\tilde{O}(T^{1-\beta})$ expected

1. The adversary may not choose a fully adversarial sequence, but will have some mild "smoothness" restrictions in order to make sublinear regret possible.
2. The soft-O notation is used to suppress logarithmic terms and dependence on parameters other than the time horizon T and the dimension of the parameter space p.

regret in the static full-information setting. Balcan et al. (2018) show several examples where dispersion holds with $\beta = \frac{1}{2}$ (which corresponds to $\tilde{O}(\sqrt{T})$ regret). Balcan et al. (2020b) provide a general methodology for proving dispersion for structured discontinuities, summarized below.

1. **Characterize discontinuities.** Show that discontinuities of each u_t occur at solutions of equations $\{\phi_{t,j}(\theta) = 0\}_j$, where each $\phi_{t,j}$ has a known functional form (e.g., linear or polynomial).
2. **Smoothed instances imply random coefficients.** Show that the coefficients of $\phi_{t,j}$ are random variables with bounded range and probability density. Often this can be obtained by making reasonable assumptions on the problem instances. For example, in the linkage clustering problem described above, it is sufficient to assume that pairwise distances between points come from a bounded-density distribution (Balcan et al. 2020a).
3. **Anti-concentration bound.** Prove that for any interval I of width ε (or any axis-aligned line segment of length ε in higher dimensions), $\Pr(\exists \theta \in I : \phi_{t,j}(\theta) = 0) \leq C \cdot \varepsilon$, for C that is independent of t, j (but may depend on other problem parameters). This implies that over T rounds, at most $O(T(\max_t m_t)C\varepsilon)$ discontinuities fall in any interval of width ε in expectation.
4. **Dispersion.** Shattering and VC-dimension based arguments for the function class of $\phi_{t,j}$ (polynomial, exponential, Pfaffian, etc.) to convert the anti-concentration results to dispersion bounds.

We present open questions around whether we can extend these results, with a focus on Step 3.

Anti-concentration bounds. Balcan et al. (2020b) show that for random degree-d polynomials for which the coefficients are bounded and have a bounded joint density function, the desired anti-concentration bound in step (3) holds. The following result is stated for $p = 1$ because there is a known argument to convert the bound for one dimensional parameter θ to one for high dimensional θ (Balcan et al., 2020b) by incurring an additional $O(p)$ factor. The key idea is that in order to establish dispersion, it is sufficient to bound expected number of discontinuities along axis-aligned paths between two parameter points, each segment of which is effectively the one-dimensional problem.

Theorem 2 *(Balcan et al. 2020b, Theorem 18) For every random polynomial $\phi(\theta) = \theta^d + \alpha_{d-1}\theta^{d-1} + \cdots + \alpha_0$, with coefficients $(\alpha_{d-1}, \ldots, \alpha_0) \in [-R, R]^d$ and joint density bounded by κ, and every bounded interval I, $\Pr(\exists \theta \in I : \phi(\theta) = 0) \leq \kappa(2R)^{d-1}\left(d + \frac{Rd(d-1)}{2}\right)|I|$.*

Another natural condition, namely if the coefficients $(\alpha_{d-1}, \ldots, \alpha_0)$ are bounded and have bounded marginal densities, is *not* a sufficient condition for any finite bound of this form. On the other hand, if we additionally assume independence of coefficients (which is too strong an assumption for most applications), bounded marginals also imply a bound on the joint density, and we can apply Theorem 2 to get finite anti-concentration bounds. To understand when low-regret online optimization is possible in this setting, it would be helpful to determine natural sufficient conditions that lead to finite (and ideally polynomial in R and d) root anti-concentration bounds, as formalized below.

Open Question 1. Consider parameter dimension $p = 1$. Suppose the discontinuities of u_t are given by roots of a polynomial equation $\phi_\alpha(\theta) = \theta^d + \alpha_{d-1}\theta^{d-1} + \cdots + \alpha_0 = 0$ with degree d and random coefficients $\alpha := (\alpha_{d-1}, \ldots, \alpha_0)$ supported on $[-R, R]^d$. For a coefficient distribution class $\mathfrak{D}$, define the root anti-concentration constant $C_\mathfrak{D}$ as

$$C_\mathfrak{D} = \sup_{\mu \in \mathfrak{D}} \sup_{\text{interval } I \subseteq \Theta, |I| > 0} \frac{\Pr_{\alpha \sim \mu}[\exists \theta \in I : \phi_\alpha(\theta) = 0]}{|I|}.$$

Under what natural necessary and sufficient conditions on the coefficient distribution class $\mathfrak{D}$ can we guarantee that $C_\mathfrak{D}$ is (a) finite, or (b) polynomial in d and R? Partial progress would include interesting sufficient conditions that improve or give new regret bounds via the above recipe for one of the applications previously studied in the literature, e.g., linear regression (Balcan et al., 2022), linkage clustering (Balcan et al., 2017), semi-supervised learning (Balcan and Sharma, 2021), or low-rank approximation (Bartlett et al., 2022). Theorem 2 gives a finite upper bound for the class of distributions $\mathfrak{D}$ supported on $[-R, R]^d$ with joint density at most κ.

2.1. Beyond polynomials: Pfaffian transition boundaries

The polynomial anti-concentration question above concerns transition boundaries of the form $\phi(\theta) = 0$, where ϕ is a random polynomial with bounded-density coefficients. A natural next step is to ask for analogous guarantees for natural, richer function classes, such as Pfaffians (Khovanskii, 1991), which include exponential, logarithmic, and polynomial functions. Consider a one-dimensional parameter interval J, and let $\phi_\alpha(\theta) = \sum_{i=1}^N \alpha_i F_i(\theta) = \langle \alpha, F(\theta) \rangle$, where $F = (F_1, \ldots, F_N)$ is a vector of Pfaffian functions on J, and $\alpha \in [-R, R]^N$ is drawn from a distribution with bounded density. One might hope for a uniform anti-concentration bound (similar to Theorem 2) of the form $\Pr\big(\exists \theta \in I : \phi_\alpha(\theta) = 0\big) \leq C^{\mathrm{Pf}}(q, \Delta, M, N, R, A)\,|I|$, where q, Δ, M are the Pfaffian complexity parameters (see Balcan et al. 2025a for definitions), $I \subseteq J$ is bounded, and $A = (2R)^N \kappa$ is the normalized joint density bound on the coefficients α. However, such a bound cannot hold with C^{Pf} polynomial in these formal complexity parameters alone.

Counter-example 1. Fix $\delta \in (0, 1]$, set $J = [-1, 1]$, and take $F_1(\theta) = 1$, $F_2(\theta) = \theta/\delta$. These are degree-one polynomials, hence Pfaffian with constant complexity, regardless of δ. Let α_1, α_2 be independent uniform random variables on $[-1, 1]$. For $I = [0, \varepsilon]$ with $\varepsilon \in (0, \delta]$, the equation $\phi_\alpha(\theta) = \alpha_1 + \alpha_2\theta/\delta = 0$ has a root in I exactly when α_1, α_2 have opposite signs and $\delta|\alpha_1| \leq \varepsilon|\alpha_2|$, giving $\Pr\big(\exists\, \theta \in [0, \varepsilon] : \phi_\alpha(\theta) = 0\big) = \frac{\varepsilon}{4\delta}$. Since δ can be taken arbitrarily small while the formal Pfaffian degrees and the coefficient distribution of α remain fixed, any bound of the form $\Pr(\exists\, \theta \in I : \phi_\alpha(\theta) = 0) \leq C^{\mathrm{Pf}}|I|$ must have $C^{\mathrm{Pf}} \geq 1/(4\delta)$, so no bound polynomial in the formal complexity parameters and the coefficient distribution alone is possible.

Open Question 2. Theorem 2 shows that for random polynomials, after normalizing so that the leading coefficient is fixed to 1, anti-concentration holds with a constant that is finite in d, R, and κ. What is the analogous normalization condition for a Pfaffian vector $F = (F_1, \ldots, F_N)$ on a domain Θ—replacing the role played by "fixing the leading coefficient"—under which one can guarantee, for the coefficient distribution class $\mathfrak{D}$ over $[-R, R]^N$ consisting of distributions with bounded joint density, that the Pfaffian root anti-concentration constant

$$C_{\mathfrak{D}}^{\mathrm{Pf}} = \sup_{\mu \in \mathfrak{D}} \sup_{\text{interval } I \subseteq \Theta, |I| > 0} \frac{\Pr_{\alpha \sim \mu}[\exists \theta \in I : \langle \alpha, F(\theta) \rangle = 0]}{|I|}.$$

is finite? In particular, if F_i are all polynomials, then we should recover the bound in Theorem 2.

We conjecture that a simple condition—e.g., bounding the Lipschitz constant of the normalized function $F(\theta)/\|F(\theta)\|_2$—may suffice to play the role that fixing the leading coefficient plays for polynomials. Resolving this open question, however, requires understanding when this normalized Lipschitz constant (or some other natural conditioning parameter) is itself polynomially bounded in the relevant instance-complexity parameters, since the counter-example above shows that boundedness cannot be taken for granted.

More generally, it would be interesting to give other sufficient conditions beyond dispersion (Balcan et al., 2018, 2020b) for online optimization of piecewise-Lipschitz functions u_t.

References

Maria-Florina Balcan. Data-Driven Algorithm Design (book chapter). In *Beyond Worst-Case Analysis of Algorithms, Tim Roughgarden (Ed)*. Cambridge University Press, 2020.

Maria-Florina Balcan and Hedyeh Beyhaghi. New guarantees for learning revenue maximizing menus of lotteries and two-part tariffs. *Transactions on Machine Learning Research (TMLR)*, 2024.

Maria-Florina Balcan and Dravyansh Sharma. Data driven semi-supervised learning. *Advances in Neural Information Processing Systems*, 34:14782–14794, 2021.

Maria-Florina Balcan, Vaishnavh Nagarajan, Ellen Vitercik, and Colin White. Learning-theoretic foundations of algorithm configuration for combinatorial partitioning problems. In *Conference on Learning Theory (COLT)*, pages 213–274. PMLR, 2017.

Maria-Florina Balcan, Travis Dick, and Ellen Vitercik. Dispersion for data-driven algorithm design, online learning, and private optimization. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 603–614. IEEE, 2018.

Maria-Florina Balcan, Travis Dick, and Manuel Lang. Learning to link. In *International Conference on Learning Representations (ICLR)*, 2020a.

Maria-Florina Balcan, Travis Dick, and Wesley Pegden. Semi-bandit optimization in the dispersed setting. In *Conference on Uncertainty in Artificial Intelligence*, pages 909–918. PMLR, 2020b.

Maria-Florina Balcan, Travis Dick, and Dravyansh Sharma. Learning piecewise Lipschitz functions in changing environments. In *International Conference on Artificial Intelligence and Statistics*, pages 3567–3577. PMLR, 2020c.

Maria-Florina Balcan, Dan DeBlasio, Travis Dick, Carl Kingsford, Tuomas Sandholm, and Ellen Vitercik. How much data is sufficient to learn high-performing algorithms? generalization guarantees for data-driven algorithm design. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 919–932, 2021.

Maria-Florina Balcan, Mikhail Khodak, Dravyansh Sharma, and Ameet Talwalkar. Provably tuning the ElasticNet across instances. *Advances in Neural Information Processing Systems*, 35:27769–27782, 2022.

Maria-Florina Balcan, Anh Tuan Nguyen, and Dravyansh Sharma. Algorithm configuration for structured Pfaffian settings. *Transactions on Machine Learning Research (TMLR)*, 2025a.

Maria-Florina Balcan, Anh Tuan Nguyen, and Dravyansh Sharma. Sample complexity of data-driven tuning of model hyperparameters in neural networks with structured parameter-dependent dual function. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b.

Peter Bartlett, Piotr Indyk, and Tal Wagner. Generalization bounds for data-driven numerical linear algebra. In *Conference on Learning Theory*, pages 2013–2040. PMLR, 2022.

Avrim Blum, Marten Garicano, Kavya Ravichandran, and Dravyansh Sharma. Algorithm design and stronger guarantees for the improving multi-armed bandits problem. *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2026. To appear.

Vincent Cohen-Addad and Varun Kanade. Online optimization of smoothed piecewise constant functions. In *Artificial Intelligence and Statistics (AISTATS)*, pages 412–420. PMLR, 2017.

Ally Yalei Du, Eric Huang, and Dravyansh Sharma. Tuning algorithmic and architectural hyperparameters in graph-based semi-supervised learning with provable guarantees. In *Conference on Uncertainty in Artificial Intelligence*, pages 1084–1111. PMLR, 2025.

Askold G Khovanskii. *Fewnomials*, volume 88. American Mathematical Soc., 1991.

Dravyansh Sharma and Maxwell Jones. Efficiently learning the graph for semi-supervised learning. In *Uncertainty in Artificial Intelligence*, pages 1900–1910. PMLR, 2023.

Dravyansh Sharma and Arun Suggala. Offline-to-online hyperparameter transfer for stochastic bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 20362–20370, 2025.