A Polynomial Time Algorithm for Log-Concave Maximum Likelihood via Locally Exponential Families

Brian Axelrod[1], Ilias Diakonikolas[2], Anastasios Sidiropoulos[3], Alistair Stewart[4], Gregory Valiant[1]

[1]Stanford University, [2]UW Madison, [3]University of Illinois at Chicago, [4]Web3 Foundation

Result

Let $X_1, \ldots, X_n \in \mathbb{R}^d$. The log-concave MLE is the log-concave density which maximizes the likelihood of $X_1, \ldots, X_n$.

Result: We present an algorithm to compute the log-concave MLE in time $\mathrm{poly}(n, d)$.

At a high level, the algorithm mimicks the textbook first order algorithm for exponential family maximum likelihood.

Motivation

Learning Distributions:

Given $X_1, ..., X_n \overset{iid}{\sim} p$, find $\hat{p} \approx p$.

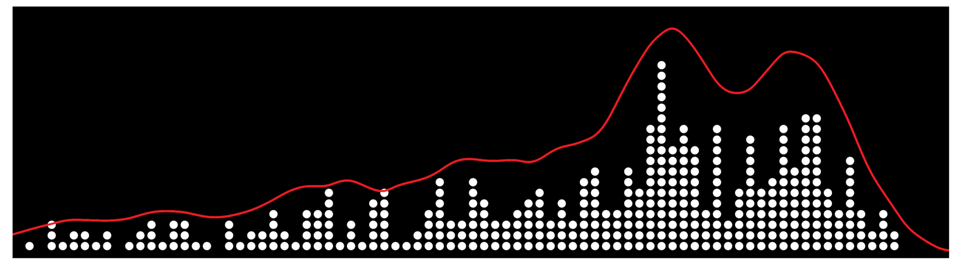

Figure 1:Figure source: https://mathisonian.github.io/kde/

Log-Concave Distributions:

Normal	Exponential	Uniform	Logistic
Wishart*	Gamma	Laplace	Chi
Chi-Square*	Beta*	Weibull*	Extreme Value

Log-concave MLE

- Non-parametric
- Parameter Free
- log-concavity is a [relatively] light assumption
- Near-optimal sample complexity
- Previous algorithm had a $n^{d/2}$ factor in it's runtime [CSS18]. Doesn't scale to medium dimensional regime.

Tent Distributions

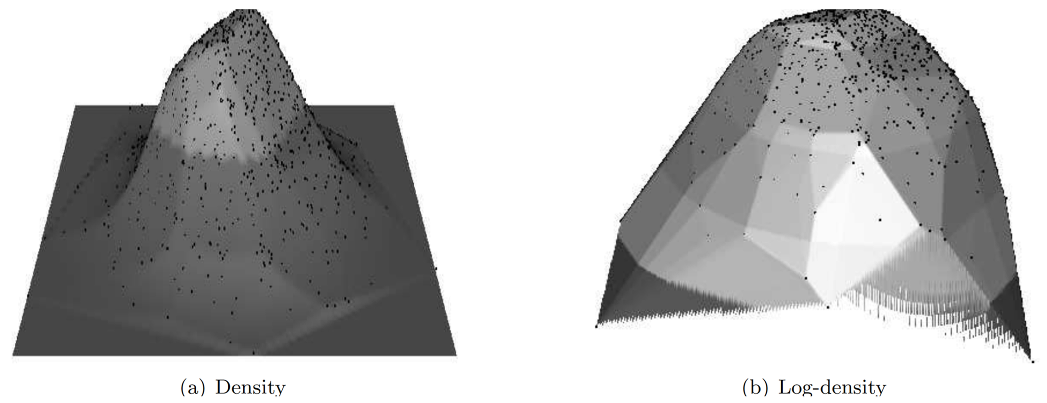

Figure 2:A figure originally presented in [1] of a tent density and log-density arising in a log-concave estimation problem

- Exp of polytope
- Solution to log-concave MLE is always tent with corners at the input points
- Fully defined by corners and their log-likelihoods

Our Algorithm

Algorithm 1 Compute log-concave MLE

$\theta \leftarrow \frac{1}{n}\mathbf{1}$
for Polynomially Many Iterations **do**
  $z_i \sim p_\theta$
  $\theta \leftarrow \theta + \eta_i(\frac{1}{n}\mathbf{1} - T_{X,\theta}(z_i))$
end for

Exponential Family MLE

Algorithm 2 Compute Exponential Family MLE

$\theta \leftarrow \theta_{init}$
for Some Iterations **do**
  $z_i \sim p_\theta$
  $\theta \leftarrow \theta + \eta_i(\mu - T(z_i))$
end for

Optimization Formulation

$$\text{Original Formulation:} \quad \max_\theta \sum \log p_\theta(x_i) - \int p_\theta(x)dx$$

$$\text{New Formulation:} \quad \max_\theta \sum \log p_\theta(x_i) - n \log \int p_\theta(x)dx$$

- Equivalent to minizing $\log \int p_\theta(x)dx$ for a slightly more restrictive definition of p. Why?
  - Gradient norms are $O(1)$
  - Easy to find initialization close to opt
  - Can compute stochastic gradient in poly time
  - Connection to exponential family convex program

Computing $T_{X,\theta}$

- Let $X_1, \ldots, X_n$ be the corners of our tent density, and θ be the corresponding vector of log-likelihoods
- $p_\theta(y) = \exp(\theta^T T_{X,\theta}(y) - A(\theta))$ where $T_{X,\theta}(z)$ maps z to the corresponding face of the polytope, and expresses z as the convex combination of the corners of this face.
- [Note] On regions of X where the regular subdivision remains constant, tent distributions form exponential families!

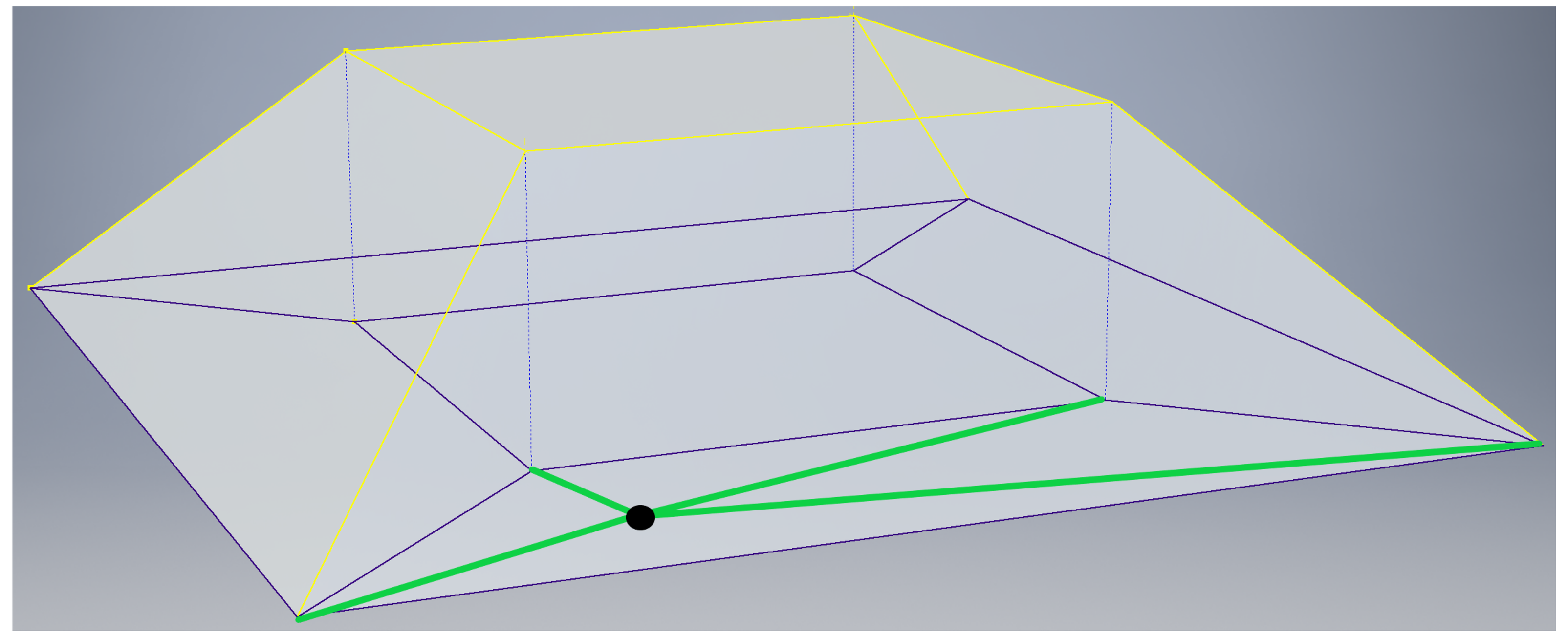

Figure 3: Computation of T

Sampling P_θ

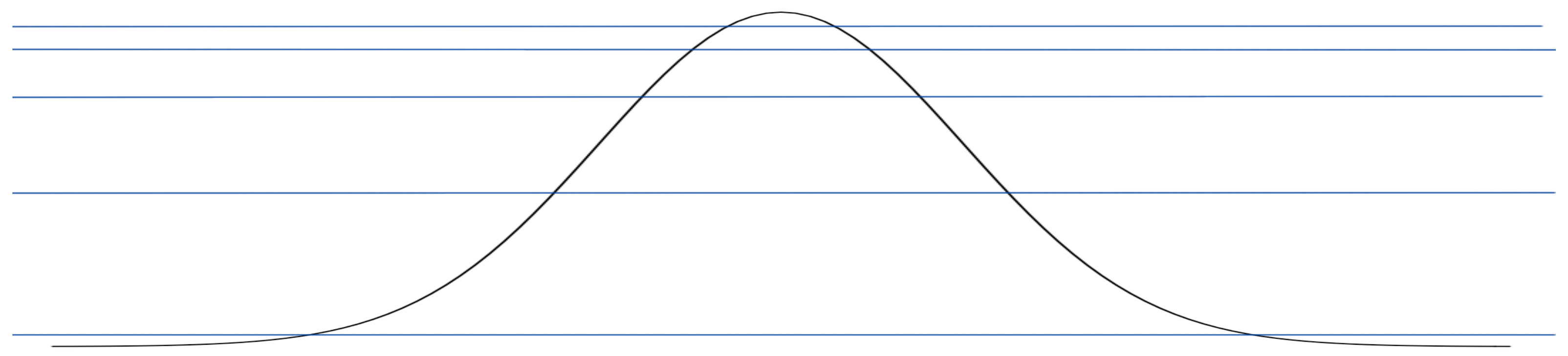

Figure 4: Computation of T

- Manually sample using the fact that level sets are polytopes with efficient membership oracles
- ULA, RHMC etc. analysis doesn't apply directly since density is *not*:
  - Smooth
  - Strongly log-concave
- Do first and second order sampling methods work here? *Open Question!*

Connection To Exponential Families

- Log-concave MLE has $\mathbb{E}[T[x]] = \frac{1}{n}\mathbf{1}_n$.
- Consider a parameter region where the faces of the polytope don't change (where the regular subdivision is constant).
  - Here, tent distributions are exponential families! Use textbook alg.
  - To show the proposed algorithm is correct for tent distributions in general, we must show that the log-partition function of tent distributions has many of the properties we expect from exponential families. We call distributions with these properties "locally" exponential.
- "locally" exponential family $p_\theta(x) = \exp(\theta^T T_\theta(x) - A(\theta))$ when A is convex and $E_{x \sim p_\theta}[T(x)] = \nabla_\theta A(\theta)$
- For these "locally" exponential families, the same algorithmic framework applies as for exponential families (albeit with a reduced convergence rate)

References

[1] Madeleine Cule, Richard Samworth, and Michael Stewart. Maximum likelihood estimation of a multi-dimensional log-concave density. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(5):545–607, 2010.

[2] Brian Axelrod, Ilias Diakonikolas, Anastasios Sidiropoulos, Alistair Stewart, and Gregory Valiant. A polynomial time algorithm for log-concave maximum likelihood via locally exponential families. *NeurIPS*, 2019.

[3] Madeleine Cule, Richard Samworth, et al. Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electronic Journal of Statistics*, 4:254–270, 2010.

[4] Timothy Carpenter, Ilias Diakonikolas, Anastasios Sidiropoulos, and Alistair Stewart. Near-optimal sample complexity bounds for maximum likelihood estimation of multivariate log-concave densities. In *Conference On Learning Theory*, pages 1234–1262, 2018.