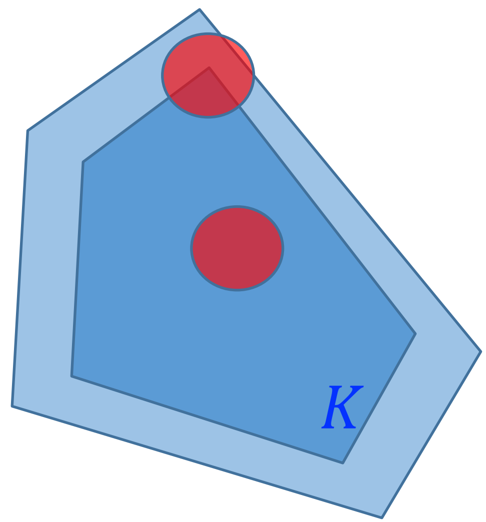
Sampling from Log-Concave Distributions with Infinity-Distance Guarantees



Oren Mangoubi
WPI

Nisheeth Vishnoi
Yale

NeurIPS 2022

Problem setting

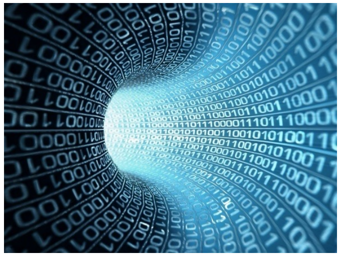
Input: Polytope $K := \{\theta \in R^d : A\theta \leq b\}, \quad A \in R^{m\times d}, b \in R^m$,
$B(0,r) \subseteq K \subseteq B(0,R)$ some $R, r > 0$
Convex function $f: K \to R$

Goal: Generate a sample from a distribution ν which is ε-close to $\pi(\theta) \propto e^{-f(\theta)}$ in **infinity distance:** $d_\infty(\pi,\nu) := \sup_{\theta\in K} \left|\log\frac{\nu(\theta)}{\pi(\theta)}\right| < \varepsilon$

We consider the setting where f is L-Lipschitz for some $L > 0$

- Sampling from log-concave distributions on K is a fundamental problem in Computer Science and Machine learning, with many applications to optimization, integration, Bayesian inference, etc.
- Infinity-distance gives stronger guarantees: Implies bounds on weaker metrics, including Total Variation (TV), KL-divergence, and α-Renyi divergence distances for $\alpha > 0$
- Many applications to differentially private (DP) optimization, e.g.
  - Convex DP empirical risk minimization
  - DP matrix approximation and PCA

Differentially private optimization



Applications:
- Medical data
- Census data
- etc.

- Data may contain sensitive information about individuals
- Many examples of privacy breaches (e.g. Netflix problem)
- Need algorithms which output ML model parameters θ that hide private information, while still allowing researchers to learn from the data

---

Pure ε-differential privacy (DP) [Dwork, '06]:
Given $\varepsilon > 0$, a randomized mechanism $\mathcal{A}$ is ε-DP if for any neighboring datasets $x, x' \in D$, $\mathbb{P}[\mathcal{A}(x) \in S] \leq \exp(\varepsilon) \times \mathbb{P}[\mathcal{A}(x') \in S]$

(x, x' are "neighbors" if they differ by at most one datapoint)

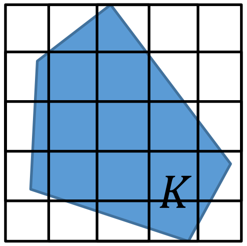
DP optimization problems can be reduced to sampling from exponential mechanism of [McSherry, Talwar, '07]:
Sample from $\pi(\theta) \propto e^{-\frac{\varepsilon}{\Delta} f(\theta;x)}$, where $\Delta := \sup_{,\theta,x,x'} |f(\theta;x) - f(\theta;x')|$

Need to sample from exponential mechanism with infinity distance error $O(\varepsilon)$ to ensure output is pure ε-DP

Previous work

Sampling from log-concave π, within TV distance $\varepsilon > 0$:

- *Hit-and-run Markov chain* [Lovasz, Vempala '06]: $O(d^{4.5} \log \frac{d}{\varepsilon r})$ calls to oracle for f and membership oracle for convex body K, from a cold start
- *Dikin walk Markov chain* [Narayanan, Rakhlin '17]: $O(md^{+\omega} + md^{2+\omega} L^2 R^2) \log \frac{w}{\varepsilon})$ arithmetic operations, from a w-warm start, if K polytope given by m inequalities ($\omega \leq 2.31$ is matrix multiplication exponent)



Sampling from log-concave π, within infinity distance $\varepsilon > 0$:

- [Bassily, Smith, Thakurta '14], using grid walk of [Applegate, Kannan '91]: $\tilde{O}\left(\frac{1}{\varepsilon^2} d^{10} + d^6 LR\right)$ calls to oracle for f and membership oracle for K
  - $\frac{1}{\varepsilon^2}$ *dependence because need grid of size* $\frac{1}{\varepsilon}$ *to sample within infinity distance* $O(\varepsilon)$

*Can one use a continuous-space Markov chain to sample within $O(\varepsilon)$ infinity distance of π, with runtime **logarithmic** in $\frac{1}{\varepsilon}$?*

Main Results

Theorem: There is an algorithm which, given $\varepsilon, L, R > 0$, a polytope $K \subseteq B(0, R)$ given by m inequalities, and an oracle for an L-Lipschitz $f: K \to R$, returns a point $O(\varepsilon)$-close in infinity distance to $\pi \propto e^{-f}$, in $O(T \times md^{\omega-1})$ arithmetic operations plus $O(T)$ evaluations of f, where $T = (m^2d^3 + m^2dL^2R^2)\times[LR + d \log\frac{LRD}{r\varepsilon}]$.

- *Improves by* $\frac{1}{m^3\varepsilon^2}d^{8-\omega}$ *on the* $\tilde{O}\left(\frac{1}{\varepsilon^2}d^{11} + d^7LR\right)$ *runtime of* *[Bassily, Smith, Thakurta '14]* *for sampling within* $O(\varepsilon)$ *infinity distance of log-Lipschitz* π *on polytope* K.

In particular, improves runtime from polynomial-in-$\frac{1}{\varepsilon}$ *to* ***logarithmic****-in-*$\frac{1}{\varepsilon}$.

- **Corollary (ε-DP empirical risk minimization (ERM)):** Plugging into exponential mechanism, get algorithm for minimizing $\sum_{i=1}^{n} f(\theta; x_i)$ *under* ε*-DP for convex L-Lipschitz* f *on polytope* K with optimal excess utility $\mathrm{E}_{\hat{\theta}}[\mathrm{f}(\hat{\theta}, x) - \mathrm{f}(\theta, x)] \leq O\left(\frac{dLR}{\varepsilon}\right)$, *in* $(m^2d^3 + m^2dn^2\varepsilon^2)\times(\varepsilon n + d)\log^2(\frac{nRd}{r\varepsilon}) \times md^{\omega-1}$ *artihmetic operations*
  - *Improves by* $\frac{d^{8-\omega}}{\varepsilon^2m^2}$ *on runtime of* *[Bassily, Smith, Thakurta '14]*, *for* **ε-DP** *convex* L*-Lipschitz empirical risk minimization on polytope* K
- **Corollary (ε-DP** low rank approximation **):** Plugging into mechanism of [Leake, McSwiggen, Vishnoi '21], get algorithm for ε-DP rank-k matrix approximation with best-known utility, with *logarithmic-in-*$\frac{1}{\varepsilon}$ runtime (improving on *polynomial-in-*$\frac{1}{\varepsilon}$ for the previous implementation of their mechanism).

Algorithm: From TV-bounds to infinity-distance bounds

Input: membership oracle for convex body $K \subseteq B(0, R)$

Input: sampling oracle for continuous-space distribution μ on K s.t. $||\pi - \mu||_{TV} \leq \delta$

1. Sample a point $\theta \sim \mu$
2. Set $Z \leftarrow \theta + \Delta r \xi$, where $\xi \sim \mathrm{Unif}(\mathrm{B}(0,1))$

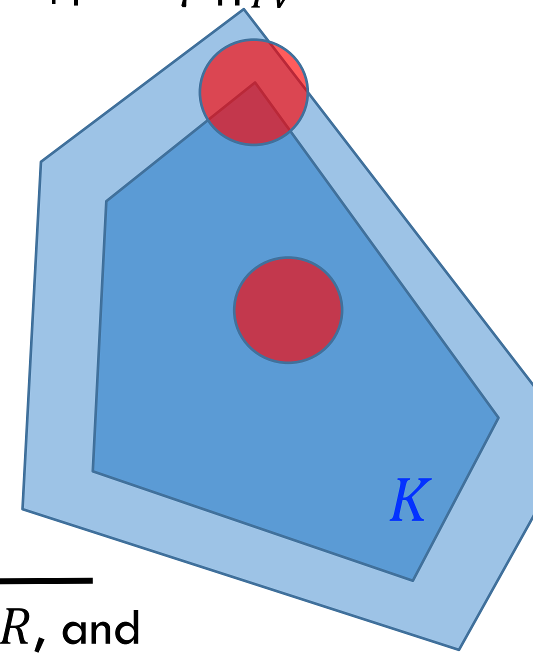
Smooth μ by convolving μ with uniform noise on a small ball

Points near corners of K are much less likely to be sampled, since (e.g., if K is a cube) only $\frac{1}{2^d}$ of ball near a corner falls inside K

3. Set $\hat{\theta} \leftarrow \frac{1}{1-\Delta} Z$

"Stretch" K to remove samples originating near boundary

4. If $\hat{\theta} \in K$, output $\hat{\theta}$ otherwise, go back to step 1



Main technical lemma: Given: $\varepsilon, L, R > 0$, L-Lipschitz $f: K \to R$, and

- a membership oracle for a convex body $K \subseteq B(0, R)$,
- an oracle which returns a sample from a continuous distribution μ on K within TV distance $\delta \leq \varepsilon \left(\frac{R(d+LR)^2}{\varepsilon r}\right)^{-d} e^{-LR}$ of $\pi \propto e^{-f}$,

our algorithm outputs point $O(\varepsilon)$-close in infinity distance to π, in $O(1)$ calls to oracles.

*Plug in sample $O(\delta)$-close to π in TV distance generated by continuous-space Markov chain with **logarithmic**-in-$\frac{1}{\delta}$ runtime (e.g. Dikin walk for polytopes; hit-and-run for more general convex bodies) to obtain point $O(\varepsilon)$-close to π in infinity distance, in **logarithmic**-in-$\frac{1}{\varepsilon}$ runtime.*

Conclusion

Introduced new method of converting TV-bounded samples from Lipschitz log-densities π on convex bodies, which can be generated by continuous-space Markov chains, into samples with bounded infinity distance to π:

- Improves runtime for generating sample $O(\varepsilon)$-close in infinity distance from Lipschitz logconcave distribution π on polytope K by $\frac{1}{m^3\varepsilon^2} d^{8-\omega}$
  - *In particular, improves runtime from polynomial-in-$\frac{1}{\varepsilon}$ to **logarithmic**-in-$\frac{1}{\varepsilon}$*
  - Can also get *logarithmic-in-$\frac{1}{\varepsilon}$ runtime* for general convex bodies K
- Application to differentially private optimization:
  - Improves *by factor of $\frac{d^{8-\omega}}{\varepsilon^2 m^2}$ the runtime to obtain optimal utility for **ε**-DP convex L-Lipschitz empirical risk minimization (ERM) on polytope.*
  - *Also obtain improved runtimes for **ε**-DP ERM on more general convex bodies, and for **ε**-DP matrix approximation problems*

Open problem: Can one sample within $O(\varepsilon)$ infinity distance of any log-concave distribution on K with runtime independent of Lipschitz constant L?

Thanks!