Faster Differentially Private Samplers via Rényi Divergence Analysis of Discretized Langevin MCMC

Arun Ganesh

Kunal Talwar

Motivation

Common primitive in differential privacy: Given database d, want to output some sensitivity-1 statistic x, where the loss of outputting x is $f(x; d)$.

Exponential mechanism [MT07]: Output x with probability proportional to

$$\exp(-f(x; d))$$

Great utility and privacy guarantees! But hard to sample from in general.

For what families of f can we efficiently implement the exponential mechanism?

Langevin Dynamics

Overdamped Langevin dynamics: Given distribution $p(x) \propto \exp(-f(x))$,the following SDE converges to p under mild assumptions:

$$dx_t = -\nabla f(x_t)dt + \sqrt{2}dB_t.$$

Easier-to-implement discretization:

$$x_{t+\eta} = -\nabla f(x_t)\eta + N(0,2\eta I_d).$$

Continuous known to converge quickly in Rényi divergence [VW19] if f is smooth, which implies (ϵ, δ)-DP [M17], but discrete chain only converges under additional assumptions.

Discrete chain known to converge quickly in measures like Wasserstein distance, total variation distance, KL-divergence, but these only imply privacy when these measures are negligible.

Our Results

Theorem (Informal): If f is strongly convex and smooth, P is $\exp(-f)$, and Q is the distribution of discretized overdamped Langevin dynamics after $\tilde{O}\left(\frac{d \ln\frac{1}{\delta}}{\epsilon^4}\right)$ iterations, then

$$D_\infty^\delta(P||Q), D_\infty^\delta(Q||P) \leq \epsilon.$$

The same set of techniques also bounds the discretization error of in Rényi divergence:

- The *underdamped* Langevin dynamics for strongly convex and smooth f
- The overdamped Langevin dynamics for *Lipschitz* and smooth f

Techniques are simple! Almost no stochastic calculus knowledge needed.

Summary of Analysis

Both the discrete and continuous chain add the same random Gaussian noise.

So, the α-Rényi divergence between these chains increases in a length η interval of time $[t, t+\eta]$ by at most

$$D_\alpha\big(N(0, 2\eta I_d), N(\Delta, 2\eta I_d)\big)$$

Where Δ is

$$\int_t^{t+\eta} \|\nabla f(x_s) - \nabla f(x_t)\|\, ds$$

For the continuous chain x_t.

By smoothness:

$$\int_t^{t+\eta} \|\nabla f(x_s) - \nabla f(x_t)\|\, ds \leq L \int_t^{t+\eta} \|x_s - x_t\|\, ds$$
$$\leq L \max_{s\in[t,t+\eta]} ||x_s - x_t||\, ds$$

We can tail bound the final quantity: Gradient descent by itself effectively decays $\nabla f(x_t)$ exponentially, so a tail bound on Brownian motion implies a tail bound on Δ.

Tail bound on Δ gives a bound on the divergence with multiplicative dependence on $\ln\left(\frac{1}{\delta}\right)$ between the two chains conditioned on a probability $1-\delta$ event.

Technical lemma: Conditional bounds on α-Rényi divergence with $\ln\left(\frac{1}{\delta}\right)$ dependence give unconditional bounds on, say, $\alpha/2$-Rényi divergence.

This argument bounds the discretization error – we can then use results from [VW19] to give the main theorem.

Techniques generalize to any chain that is a drift + Brownian motion, if we can tail bound the discretization error of drift appropriately!

Future Directions

Possible to improve dependence on ϵ? ϵ is usually a small constant in DP applications, but may be useful for other applications of Langevin dynamics.

What other popular variants of Langevin dynamics and similar sampling techniques can our analysis generalize to?

We “bypass” bounding the error of the discrete chain’s stationary distribution, so it remains an open problem.

References

[MT04] Frank McSherry, Kunal Talwar. Mechanism Design via Differential Privacy. In FOCS 2007.

[M14] Ilya Mironov. Rényi Differential Privacy. In CSF 2017.

[VW19] Santosh Vempala, Andre Wibisono. Rapid Convergence of the Unadjusted Langevin Algorithm: Isoperimetry Suffices. In NeurIPS 2019.