Dimensionally Tight Bounds for Second-Order Hamiltonian Monte Carlo

Given a gradient oracle for $F: \mathbb{R}^d \rightarrow \mathbb{R}$, sample from the Gibbs distribution: $\pi(x) \propto e^{-F(x)}$.

**Applications:** Optimization (via annealing), computing integrals/volumes, Bayesian inference, molecular dynamics

Some Markov chains used for sampling

- # of gradient evaluations to sample from smooth, strongly logconcave $\pi$ (for smoothness/convexity parameters $= \Theta(1)$):
  - Random Walk Metropolis: $d^2$ (conj. $d$) [Gelman et al. ’97]
  - Unadjusted Langevin: $d$ [Durmus, Moulines, ’16]
  - Undamped Langevin: $d^{1/2}$ [Cheng et al. ’17]

**Main result**

**Assume:**
1. $F$ is $m$-strongly convex and $M$-smooth, and let $\kappa = M/m$.
2. Lipschitz condition for $L_0, r > 0$, $X_0, \ldots, X_T \in \mathbb{R}^d$.

Then: $\hat{O}(\max \{ \frac{1}{2} (\frac{1}{2} \kappa^{2.75} + \frac{r}{\kappa^{2.25}} (L_0)^{1/2}) \}^{-1/2})$ gradients are sufficient for 2nd-order HMC to obtain a sample $\epsilon$-close to $\pi$, from a warm start ($\text{We obtain slightly weaker bounds from a cold start}$)

**Application to Bayesian logistic “ridge” regression**

- Given data $(X_i, Y_i)$, sample from $\pi(x) \propto e^{-\frac{1}{2} \| \gamma \|_2^2}$, $f_i(x) = Y_i \log(\sigma(x^T X_i)) + (1 - Y_i) \log(\sigma(-x^T X_i))$, where $\sigma(s) = (1 + e^{-s})^{-1}$.
- For logistic regression, $L_0 = \nabla c$, coherence $C = \max_{i \neq j} \frac{\|X_i X_j\|_2}{\|X_i\|_2 \|X_j\|_2}$.
- For example, if $r = d$ and $X_1, \ldots, X_r \sim \text{uniform}(\mathbb{R}^d)$, # of gradient calls is $\hat{O}(d^{3/2} \epsilon^{-1/2})$ from a warm start

**Simulations**

- Simulations performed on logistic regression, $X_1, \ldots, X_r \sim \text{uniform}(\mathbb{R}^d)$ suggest that 2nd order HMC (UHMC) has faster autocorrelation time than Metropolis HMC and Langevin in this setting (Fig. 1).

Proof highlights

- For simplicity, let $M, m = \Theta(1), \epsilon \leq 1, r = d$. We couple our HMC chain $X$ to an “idealized” HMC chain $Y$ with exact solutions by giving their trajectories the same initial velocity (Fig. 2).
- [Mangoubi, Smith ’17] show that exact solutions with same initial velocity contract by a constant factor for $T = \Theta(1)$. We extend to 2nd order HMC by showing it approximates exact trajectories with error $O(\epsilon)$:
  - We bound (inductively on $j$) the errors $\| x_j - x(\eta_j) \|_2$ and $\| y_j - v(\eta_j) \|_2$ by $O(\eta_j)$, where $(x(t), v(t))$ is the continuous solution to Hamilton’s equations with initial conditions $(X_0, V_0)$.
  - The invariance property of Hamiltonian mechanics implies $v$ is roughly $N(0, I_d)$ at every point on the exact trajectory if HMC has a warm start (Fig. 3). Thus, $\| x(T) \|_2 = O(\log(d))$ w.h.p., since by inductive assumption $\| y_j - v(\eta_j) \|_2 = O(\eta_j) = O(1)$.
- After $T/\eta$ iterations, the errors sum to $\hat{O}(\eta^2 L_0 \| \nabla c \|_2^2)$. Choosing $\eta$ to have error $\epsilon$, # of gradients is $T/\eta = \hat{O}(\epsilon^{-1/2} d^{3/2} L_0^{1/2})$.

Conclusion and future directions

- First faster-than-$\sqrt{d}$ bound for sampling from a large class of logconcave distributions, including logistic regression posteriors.
- Can we improve dependence on parameters $\kappa$ and $\kappa$?
- Can we generalize to nonconvex $F$ and higher-order integrators?