A Proximal Algorithm for Sampling from Non-smooth Potentials

Georgia Statistics Day 2021

Jiaming Liang [1] Yongxin Chen [2]

[1]School of Industrial and Systems Engineering, Georgia Tech [2]School of Aerospace Engineering, Georgia Tech

An Overview

- **Problem**: Sampling from a distribution π on $\mathbb{R}^d$ proportional to $\exp(-f(x))$ where the potential f is convex and M-Lipschitz continuous.
- **Goal**: Design an algorithm to generate a point following a distribution within ε total variation distance to π.
- **Four Algorithmic Tools**:
  - the alternating sampling framework;
  - the restricted Gaussian oracle (RGO);
  - the proximal bundle method;
  - the rejection sampling algorithm.
- **Features**:
  - a polynomial complexity $\tilde{\mathcal{O}}(d\varepsilon^{-1})$ to obtain ε total variation distance to the target density π;
  - an efficient and implementable RGO for convex and Lipschitz continuous functions.

Main results

- **Total complexity for strongly convex potentials**

Theorem 1. Let $x^0 \in \mathbb{R}^d$, $\varepsilon > 0$, $\delta > 0$, $M > 0$ $\mu > 0$ and $\eta > 0$ satisfying

$$\frac{\delta}{M^2} \le \eta \le \min\left\{\frac{1}{64M^2 d}, \frac{1}{\mu}\right\} \tag{1}$$

be given. Let π be a distribution on $\mathbb{R}^d$ satisfying

$$\pi(x) \propto \exp(-g(x)) = \exp\left(-f(x) - \frac{\mu}{2}\|x - x^0\|^2\right).$$

Consider Algorithm 1 using Algorithm 4 as an RGO for step 1, initialized at a β-warm start, then the iteration-complexity bound for obtaining ε total tolerance to π in terms of total variation is

$$\tilde{\mathcal{O}}\left(\frac{M^2}{\mu\delta}\log\left(\frac{\log\beta}{\varepsilon}\right) + 1\right),$$

and each iteration queries one subgradient oracle of f and solves a quadratic programming problem. Moreover, the number of Gaussian distribution sampling queries in Algorithm 1 is

$$\Theta\left(\frac{1}{\eta\mu}\log\left(\frac{\log\beta}{\varepsilon}\right) + 1\right).$$

- **Total complexity for convex potentials**

Theorem 2. Let π be a distribution on $\mathbb{R}^d$ satisfying $\pi(x) \propto \exp(-f(x))$. Let $x^0 \in \mathbb{R}^d$ and $\varepsilon > 0$ be given and

$$\mu = \frac{\varepsilon}{\sqrt{\mathcal{M}_4} + \|x^0 - x^*\|^2}$$

where

$$\mathcal{M}_4 = \int_{x\in\mathbb{R}^d} \|x - x^*\|^4 d\pi(x), \quad x^* \in \mathrm{Argmin}\left\{f(x) : x \in \mathbb{R}^d\right\}.$$

Choose $\delta > 0$ and $\eta > 0$ such that (1) holds and consider Algorithm 1 using Algorithm 4 as an RGO for step 1, applied to

$$g = f + \frac{\mu}{2}\left\|\cdot - x^0\right\|^2,$$

and initialized at a β-warm start. Then, the iteration-complexity bound for obtaining ε total tolerance to π is

$$\tilde{\mathcal{O}}\left(\frac{M^2\left(\sqrt{\mathcal{M}_4} + \|x^0 - x^*\|^2\right)}{\varepsilon\delta}\log\left(\frac{\log\beta}{\varepsilon}\right) + 1\right).$$

Alternating sampling framework and RGO

Proximal mapping in optimization:

$$\mathrm{argmin}\left\{g(\cdot) + \frac{1}{2\eta}\|\cdot - y\|^2\right\}.$$

RGO in sampling:

$$\text{sample} \quad \exp\left(-g(\cdot) - \frac{1}{2\eta}\|\cdot - y\|^2\right).$$

Algorithm 1 Alternating Sampling Framework

0. sample $y \sim \pi_x(y) \propto e^{-\frac{1}{2\eta}\|x-y\|^2}$
1. sample $x \sim \pi_y(x) \propto e^{-g(x) - \frac{1}{2\eta}\|x-y\|^2}$

Theorem 3. Let π be a distribution on $\mathbb{R}^d$ with $\pi(x) \propto \exp\left(-f_{\text{oracle}}(x)\right)$ such that f_{oracle} is μ-strongly convex, and let $\varepsilon \in (0,1)$. Let $\eta > 0$,

$$T = \Theta\left(\frac{1}{\eta\mu}\log\frac{\log\beta}{\varepsilon}\right)$$

for some $\beta \ge 1$. Algorithm 1, initialized at a β-warm start, runs in T iterations, each querying RGO for f_{oracle} with parameter η a constant number of times, and obtais ε total variation distance to π.

RGO with an optimization oracle

Assume that f has a proximal mapping and let

$$x^* = \underset{x\in\mathbb{R}^d}{\mathrm{argmin}}\left\{g^\eta(x) := g(x) + \frac{1}{2\eta}\|x - y\|^2\right\}. \tag{2}$$

Lemma 1. Let $\eta_\mu := \eta/(1 + \eta\mu)$ and define

$$h_1 := \frac{1}{2\eta_\mu}\|\cdot - x^*\|^2 + g^\eta(x^*), \quad h_2 := \frac{1}{2\eta_\mu}\|\cdot - x^*\|^2 + 2M\|\cdot - x^*\| + g^\eta(x^*).$$

Then, for every $x \in \mathbb{R}^d$, we have

$$h_1(x) \le g^\eta(x) \le h_2(x).$$

Algorithm 2 Implementation of the RGO with an optimization oracle

1. Compute x^* as in (2);
2. Generate $X \sim \exp(-h_1(x))$;
3. Generate $U \sim \mathcal{U}[0,1]$;
4. If

$$U \le \frac{\exp(-g^\eta(X))}{\exp(-h_1(X))},$$

then accept $\tilde{X} = X$; otherwise, reject X and go to step 2.

Proposition 1. Let

$$p(x) = p(x|y) \propto \exp\left(-g(x) - \frac{1}{2\eta}\|x - y\|^2\right),$$

then $\tilde{X} \sim p(x)$. If $\eta_\mu \le 1/(16M^2 d)$, then the expected number of iteration in Algorithm 2 is at most 2.

Review of the proximal bundle method

Consider the subproblem

$$g_*^\eta := g^\eta(x^*) = \min\left\{g^\eta(x) := g(x) + \frac{1}{2\eta}\|x - y\|^2 : x \in \mathbb{R}^d\right\},$$

and we aim at obtaining a δ-solution (i.e., a point $\bar{x}$ such that $g^\eta(\bar{x}) - g_*^\eta \le \delta$) to the above subproblem.

Algorithm 3 Solving the Proximal Bundle Subproblem

0. Let y, $\eta > 0$ and $\delta > 0$ be given, and set $\tilde{x}_0 = y$, $C_1 = \{y\}$ and $j = 1$;
1. Update

$$f_j = \max\left\{f(x) + \langle f'(x), \cdot - x\rangle : x \in C_j\right\};$$

2. Define $g_j := f_j + \mu\|\cdot - x^0\|^2/2$ and compute

$$x_j = \underset{u\in\mathbb{R}^n}{\mathrm{argmin}}\left\{g_j^\eta(u) := g_j(u) + \frac{1}{2\eta}\|u - y\|^2\right\}, \qquad \tilde{x}_j \in \mathrm{Argmin}\left\{g^\eta(u) : u \in \{x_j, \tilde{x}_{j-1}\}\right\};$$

3. If $g^\eta(\tilde{x}_j) - g_j^\eta(x_j) \le \delta$, then **stop**; else, go to step 4;
4. Choose C_{j+1} such that

$$A_j \cup \{x_j\} \subset C_{j+1} \subset C_j \cup \{x_j\}$$

where

$$A_j := \left\{x \in C_j : f(x) + \langle f'(x), x_j - x\rangle = f_j(x_j)\right\};$$

5. Set $j \leftarrow j + 1$ and go to step 1.

Proposition 2. Algorithm 3 takes $\tilde{\mathcal{O}}(\eta_\mu M^2/\delta + 1)$ iterations to terminate, and each iteration solves a linearly constrained convex quadratic programming problem.

RGO without an optimization oracle

Define

$$h_1 := \frac{1}{2\eta_\mu}\|\cdot - x_j\|^2 + g^\eta(\tilde{x}_j) - \delta,$$

$$h_2 := \frac{1}{2\eta_\mu}\|\cdot - \tilde{x}_j\|^2 + \left(2M + \frac{\sqrt{2\delta}}{\sqrt{\eta_\mu}}\right)\|\cdot - \tilde{x}_j\| + g^\eta(\tilde{x}_j).$$

Algorithm 4 Implementation of the RGO without an optimization oracle

1. Compute x_j and $\tilde{x}_j$ as in Algorithm 3;
2. Generate $X \sim \exp(-h_1(x))$;
3. Generate $U \sim \mathcal{U}[0,1]$;
4. If

$$U \le \frac{\exp(-g^\eta(X))}{\exp(-h_1(X))},$$

then accept $\tilde{X} = X$; otherwise, reject X and go to step 2.

Lemma 2. For every $x \in \mathbb{R}^d$, we have $h_1(x) \le g^\eta(x) \le h_2(x)$.

Proposition 3. If

$$\eta_\mu \le \frac{1}{64M^2 d}, \quad \delta \le \frac{1}{32d},$$

then the expected number of iterations in the rejection sampling is at most 3.