Uniform, Constrained, and Composite Sampling via Proximal Sampler

Jiaming Liang

Department of Computer Science & Goergen Institute for Data Science
University of Rochester

February 6, 2025

Joint work with Thanh Dang (UR, previously FSU)

Department of Mathematics, Florida State University

Uniform, Constrained, and Composite Sampling

- Uniform sampling

$$\mu(x) \propto \mathbf{1}_K(x) = \exp(-I_K(x)),$$

where

$$\mathbf{1}_K(x) = \begin{cases} 1, & \text{if } x \in K; \\ 0, & \text{otherwise}, \end{cases} \quad I_K(x) = \begin{cases} 0, & \text{if } x \in K; \\ +\infty, & \text{otherwise}. \end{cases}$$

- Constrained sampling

$$\nu(x) \propto \exp(-f(x))\mathbf{1}_K(x) = \exp\left(-f(x) - I_K(x)\right).$$

- Composite sampling (useful for Bayesian inference)

$$\tilde{\nu}(x) \;\propto\; \exp\left(-f(x) - h(x)\right).$$

Assume f, h, and K are all closed and convex, then all three problems are log-concave sampling.

Epigraph Transformation

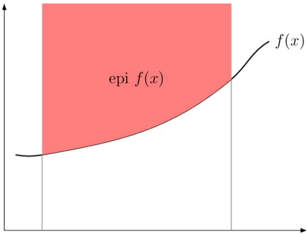
Motivation from optimization

$$\min\{f(x) : x \in K\},$$

reformulated as

$$\min\{t : (x,t) \in Q\}, \quad Q = \{(x,t) \in K \times \mathbb{R} : f(x) \le t\},$$

where Q is the *epigraph* of f restricted to K.



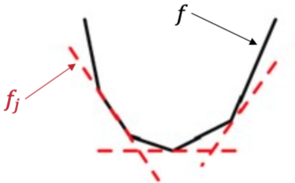
An Example in Optimization

Consider solving $\min\{f(x) : x \in \mathbb{R}^d\}$ by the proximal bundle method, a key subproblem

$$x_j = \underset{x\in\mathbb{R}^d}{\operatorname{argmin}} \left\{ f_j(x) + \frac{1}{2\eta}\|x-y\|^2 \right\},$$

where $f_j(x)$ is a cutting-plane model for $f(x)$

$$f_j(x) = \max_{0\le i\le j-1} \{f(x_i) + \langle f'(x_i), x - x_i\rangle\}.$$



Epigraph reformulation

$$x_j = \underset{x\in\mathbb{R}^d, t\in\mathbb{R}}{\operatorname{argmin}} \left\{ t + \frac{1}{2\eta}\|x-y\|^2 \right\},$$

s.t. $f_j(x) \le t$ or $f(x_i) + \langle f'(x_i), x - x_i\rangle \le t$, $0 \le i \le j-1$.

Epigraph Lifting for Constrained Sampling

Reformulating constrained sampling (similar to Kook and Vempala (2025))

$$\nu(x) \propto \exp\left(-f(x) - I_K(x)\right).$$

as a nearly uniform distribution in $\mathbb{R}^{d+1}$

$$\pi(x,t) \propto e^{-at}\mathbf{1}_Q(x,t) = \exp(-at - I_Q(x,t)).$$

with some scaling constant $a > 0$ and the lifted set

$$Q = \{(x,t) \in \mathbb{R}^d \times \mathbb{R} : x \in K, f(x) \leq at\}.$$

Integrating out t gives

$$\pi^X(x) = \int \pi(x,t)\mathrm{d}t \propto \left(\int_{f(x)/a}^{\infty} e^{-at}\mathrm{d}t\right)\mathbf{1}_K(x) = \frac{1}{a}e^{-f(x)}\mathbf{1}_K(x) \propto \nu(x).$$

Thus, if one can efficiently sample (x,t) from π over the lifted set Q, then discarding the t-coordinate produces samples from the target ν.

Double Lifting for Composite Sampling

Reformulating composite sampling $\tilde{\nu}(x) \;\propto\; \exp\left(-f(x) - h(x)\right)$ as a nearly uniform sampling in $\mathbb{R}^{d+2}$.

First, apply the epigraph lifting to h, reducing to constrained sampling

$$\gamma(x,s) \;\propto\; \exp\left(-\tilde{f}(x,s) - I_{\widetilde{K}}(x,s)\right),$$

with some scaling constant $a > 0$ and

$$\widetilde{K} = \{(x,s) \in \mathbb{R}^d \times \mathbb{R} : \; h(x) \le as\}, \quad \tilde{f}(x,s) = f(x) + as.$$

Then, lift $\tilde{f}$ again, reducing to a nearly uniform distribution in $\mathbb{R}^{d+2}$

$$\tilde{\pi}(x,s,t) \;\propto\; \exp\left(-bt - I_{\widetilde{Q}}(x,s,t)\right)$$

with some scaling constant $b > 0$ and

$$\widetilde{Q} = \{(x,s,t) : \; (x,s) \in \widetilde{K}, \; \tilde{f}(x,s) \le bt\}.$$

Integrating out (s,t) shows that $\tilde{\pi}^X = \tilde{\nu}$, since

$$\int_{h(x)/a}^{\infty} \int_{\tilde{f}(x,s)/b}^{\infty} e^{-bt} \mathrm{d}t\mathrm{d}s = \frac{1}{ab} \exp(-f(x) - h(x)) \,.$$

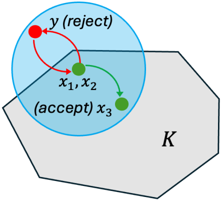
Outline

1. Uniform Sampling
2. Constrained Sampling
3. Composite Sampling

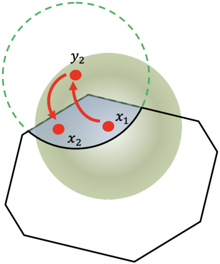
Outline

Uniform Sampling

Classical ideas are based on geometric random walks, for example, ball walk, hit-and-run, Dikin walk.



Recent work based on diffusion processes: In-and-Out by Kook, Vempala, and Zhang (2024), and our work



Proximal Sampler

Given a step size $\eta > 0$ and a target $\pi(x) \propto \exp(-f(x))$, the proximal sampler performs Gibbs sampling on the augmented distribution

$$\Pi^{X,Y}(x,y) \;\propto\; \exp\left(-f(x) - \frac{1}{2\eta}\|x-y\|^2\right),$$

whose X-marginal is exactly π. Reversible Markov chain with stat. distr. $\Pi^{X,Y}$ $\implies$ unbiasedness for the target $\pi(x)$.

Algorithm Proximal Sampler Lee, Shen, and Tian (2021)

1. Sample $y_k \sim \pi^{Y|X}(y \mid x_k) \propto \exp\left(-\frac{1}{2\eta}\|x_k - y\|^2\right)$ forward heat flow
2. Sample $x_{k+1} \sim \pi^{X|Y}(x \mid y_k) \propto \exp\left(-f(x) - \frac{1}{2\eta}\|x - y_k\|^2\right)$ backward flow

Each iteration alternates between sampling $y \sim \Pi^{Y|X}$, a Gaussian distribution, and sampling $x \sim \Pi^{X|Y}$, known as the restricted Gaussian oracle (RGO).

In general, the RGO is unavailable or difficult to implement.

Mixing Time Guarantees

Chen, Chewi, Salim, and Wibisobo (2024) give general mixing time analysis of the proximal sampler.
Extending Lee, Shen, and Tian (2021) from (strongly) log-concave cases to distributions satisfying log-Sobloev, Poincaré inequality.
Kook, Vempala, and Zhang (2024) further extend the resutls to uniform sampling, where discontinuity on the boundary is a challenge.

Theorem

Let $K \subset \mathbb{R}^d$ be a convex set. Assume PS starts from an M-warm distribution μ_0^X, i.e., $\frac{\mathrm{d}\mu_0}{\mathrm{d}\pi^X} \leq M$. Let $\epsilon > 0$. Denote C_{PI} and C_{LSI} respectively the Poincaré constant and the log Sobolev constant of the uniform distribution π^X on K.

- *In terms of the Rényi divergence $\mathcal{R}_q$ and $q \geq 1$, the algorithm can achieve ϵ-accuracy within $O\left(\frac{1}{\eta} C_{\mathrm{LSI}} q \log\left(2 \frac{\log M}{\epsilon}\right)\right)$ iterations, where $C_{\mathrm{LSI}} = \mathcal{O}(D^2)$ in general and $C_{\mathrm{LSI}} = \mathcal{O}(D)$ if π^X is isotropic;*
- *In terms of the χ^2-divergence, the algorithm reaches ϵ-accuracy in $O\left(\frac{1}{\eta} C_{\mathrm{PI}} \log\left(2 \frac{M^2+1}{\epsilon}\right)\right)$ iterations, where $C_{\mathrm{PI}}(\pi) = \mathcal{O}\left(\|\mathrm{Cov}(\pi)\|_{\mathrm{op}} \log d\right)$ in general and $C_{\mathrm{PI}}(\pi) = \mathcal{O}\left(\log d\right)$ if π^X is isotropic.*

Uniform Sampling via Proximal Sampler

Recall that $\pi^X \propto \exp(-I_K(x)) = \mathbf{1}_K(x)$ so $f(x) = I_K(x)$. Hence, RGO is truncated Gaussian

$$\mathcal{N}(y, \eta I_d)|_K \propto \exp\left(-\frac{1}{2\eta}\|x-y\|^2 - I_K(x)\right) = \exp\left(-\frac{1}{2\eta}\|x-y\|^2\right)\mathbf{1}_K(x).$$

It is easy to figure out $\mathcal{N}(y, \eta I_d)|_K$ concentrates at

$$\operatorname{proj}_K(y) = \operatorname*{argmin}_{x\in\mathbb{R}^d}\left\{\Theta_y^{\eta,K}(x) := I_K(x) + \frac{1}{2\eta}\|x-y\|^2\right\}.$$

Algorithm Projection oracle-based implementation of RGO

1. Generate $X \sim \mathcal{N}\left(\operatorname{proj}_K(y), \eta I_d\right)$ and $U \sim \mathcal{U}[0,1]$.
2. If

$$U \le \exp\left(-\frac{1}{\eta}\langle X - \operatorname{proj}_K(y), \operatorname{proj}_K(y) - y\rangle\right)\mathbf{1}_K(X),$$

then accept X; otherwise, reject X and go to step 1.

Rejection Sampling Complexity

Assumptions

(A1) K is a non-empty, closed, and convex set in $\mathbb{R}^d$ such that $B(0,1) \subseteq K \subseteq B(0,R)$ for some $R > 1$, where $B(0,R)$ denotes the Euclidean ball centered at the origin with radius R.

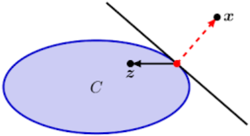
(A2) the initial distribution μ_0 is M-warm with respect to the uniform distribution π on K, i.e., $\frac{\mathrm{d}\mu_0}{\mathrm{d}\pi} \le M$.

Theorem

Assume conditions (A1) and (A2) hold, and consider the proximal sampler with stepsize $\eta = 1/d^2$. Then, the average number of proposals in RGO implementation is bounded by $M(\sqrt{2\pi e} + 1)$.

Separation Oracle

A separation oracle either certifies $x \in K$ or, if $x \notin K$, returns $g : \mathbb{R}^d \to \mathbb{R}^d$ such that $\langle g(x), x - y \rangle \geq 0$ for all $y \in K$; in particular, it subsumes membership.



Using a separation oracle on K to find an approximate solution $\hat{x}$ of $\min_{x \in \mathbb{R}^d} \Theta_y^{\eta, K}(x)$.

Using $\hat{x}$, we do rejection sampling with the proposal

$$\nu(x) \; \propto \; \exp\left(-\frac{1}{2\eta} \left[\|x - \hat{x}\|^2 - 2\sqrt{\frac{2\eta}{d}} \|x - \hat{x}\| \right] \right).$$

RGO Based on Separation Oracle

Algorithm Separation oracle-based implementation of RGO

1. Compute a $(1/d)$-solution $\hat{x}$ of $\min_{x \in \mathbb{R}^d} \Theta_y^{\eta,K}(x)$ via the Cutting Plane method by Jiang, Lee, Song, Wong (2020).
2. Generate $U \sim \mathcal{U}[0,1]$ and $X \sim \nu(x)$ via a subroutine below.
3. If

$$U \leq \exp\left(\mathcal{P}_2(X) - \Theta_y^{\eta,K}(X)\right), \tag{1}$$

then accept X; otherwise, reject X and go to step 2. The function $\mathcal{P}_2$ is defined below.

$$\mathcal{P}_2(x) := \frac{1}{2\eta}\left(\|x-\hat{x}\|^2 + \|\hat{x}-y\|^2 - 2\sqrt{\frac{2\eta}{d}}\left(\|x-\hat{x}\| + \|\hat{x}-y\|\right) - \frac{12\eta}{d}\right)$$

satisfies that

$$\mathcal{P}_2(x) \leq \Theta_y^{\eta,K}(x), \quad \forall x \in \mathbb{R}^d,$$

so the acceptance test (1) is well-defined

RGO Complexities

Theorem

Assume conditions (A1), (A2), and the stepsize $\eta = 1/d^2$.

a) *The average number of proposals in Algorithm 3 is no more than* $\sqrt{2\pi}M\exp\left(\frac{13}{4}+\frac{20}{d}\right)+M\exp\left(\frac{9}{4}+\frac{12}{d}\right)$.

b) *There are at most* $\mathcal{O}\left(d\log\frac{d\gamma}{\alpha}\right)$ *queries to the separation oracle on* K, *where* $\gamma = \frac{R}{\textit{minwidth}(K)}$ *and* $\alpha \in (0,1)$ *satisfies the concentration inequality* $\Pr\left(\alpha \le \frac{2}{7d^3R^2}\right) \le 4\exp\left(-\frac{d^2R^2}{8}\right)$.

Algorithm Sample $X \sim \exp\left(-\frac{1}{2\eta}\left(\|x-\hat{x}\|^2 - 2\sqrt{\frac{2\eta}{d}}\|x-\hat{x}\|\right)\right)$

---

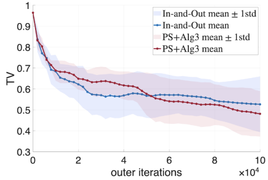
1. Generate $W \sim \mathcal{N}(0,I)$ and set $\theta = W/\|W\|$.
2. Generate $r \propto r^{d-1}\exp\left(-\frac{(r-b)^2}{2\eta}\right)$ by Adaptive Rejection Sampling for one-dimensional log-concave distribution by Gilks and Wild (1992).
3. Output $X = \hat{x} + r\theta$.

Experiments

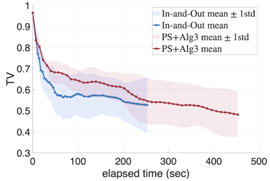
Uniform sampling on ense Z-polytope

$$Z = \{x = Vt : \|t\|_\infty \le 1\} = \left\{\sum_{j=1}^{m} t_j v_j : t_j \in [-1,1]\right\}.$$

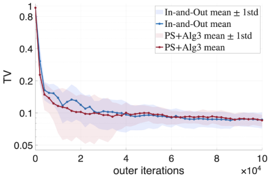
$d = 40$ and $m = 80$, and run $n = 10^5$ outer iterations per seed over 6 independent seeds



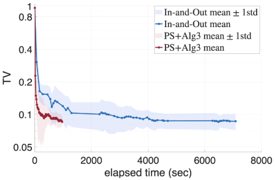
(a) Exp. A (η_A): TV vs. iterations



(b) Exp. A (η_A): TV vs. elapsed time



(c) Exp. B (η_B): TV vs. iterations



(d) Exp. B (η_B): TV vs. elapsed time

Outline

Going Back to Constrained Sampling

Recall the lifted sampling problem as a nearly uniform sampling

$$\pi(x,t) \propto \exp(-at - I_Q(x,t)), \quad Q = \{(x,t) \in \mathbb{R}^d \times \mathbb{R} : \ x \in K, \ f(x) \le at\}.$$

(B1) $K \subset \mathbb{R}^d$ is nonempty, closed, convex, with $\mathrm{B}_d(0) \subseteq K \subseteq R\mathrm{B}_d(0)$;

(B2) f is convex and Lipschitz continuous on K with constant $L > 0$;

(B3) there exists an initial distribution ν_0 that is M-warm w.r.t. ν;

(B4) there exists a subgradient oracle for f and a separation oracle for K.

The augmented distribution in the proximal sampler is

$$\Pi((x,t),(y,s)) \propto \exp\left(-I_Q(x,t) - at - \frac{1}{2\eta}\left\|(x,t)-(y,s)\right\|^2\right).$$

Fixing $(y,s) \in \mathbb{R}^{d+1}$, the above potential (i.e., negative log-density) function becomes

$$\Theta^{\eta,Q}_{y,s}(x,t) = I_Q(x,t) + \frac{\|(x,t)-(y,s-a\eta)\|^2}{2\eta} - \frac{a^2\eta}{2} + as.$$

Proximal Sampler in the Lifted Space

Note that

$$\Pi^{Y,S|X,T}(y,s|x,t) = \mathcal{N}\left((x,t), \eta I_{d+1}\right),$$
$$\Pi^{X,T|Y,S}(x,t|y,s) = \mathcal{N}\left((y,s-a\eta), \eta I_{d+1}\right)|_Q.$$

Thus, the proximal sampler for the target $\pi(x,t)$ is as follows.

Algorithm Proximal sampler for the target π

1: **Gaussian**: generate $(y_k, s_k) \sim \mathcal{N}\left((x_k, t_k), \eta I_{d+1}\right)$;
2: **RGO**: generate $(x_{k+1}, t_{k+1}) \sim \mathcal{N}\left((y_k, s_k - a\eta), \eta I_{d+1}\right)|_Q$.

Theorem

Assume (A1)-(A3) holds and $a = d$. Denote by $C_{\mathrm{PI}}(\pi)$ the Poincaré constant of π. It can be shown that $C_{\mathrm{PI}}(\pi) = \mathcal{O}\left(\log(d+1)\left(\|\mathrm{Cov}(\nu)\|_{\mathrm{op}} + 1\right)\right)$. Given $\epsilon > 0$, the iteration complexity in χ^2 divergence to reach ϵ-accuracy is $\mathcal{O}\left(\frac{C_{\mathrm{PI}}(\pi)}{\eta}\log\frac{M^2}{\epsilon}\right)$. Moreover, for $\mathcal{R}_q$ with $q \geq 2$, if $M \leq e^{1-1/q}$, then the iteration complexity in $\mathcal{R}_q$ to reach ϵ-accuracy is $\mathcal{O}\left(\frac{C_{\mathrm{PI}}(\pi)q}{\eta}\log\left(2\frac{\log M}{\epsilon}\right)\right)$.

RGO Implementation

Note that the RGO $\Pi^{X,T|Y,S}(x,t|y,s) = \mathcal{N}\left((y, s - a\eta), \eta I_{d+1}\right)|_Q$ is concentrated around

$$\underset{(x,t)\in\mathbb{R}^{d+1}}{\operatorname{argmin}}\ \Theta_{y,s}^{\eta,Q}(x,t) = \underset{w\in Q}{\operatorname{argmin}}\left\|w - z\right\|^2 = \operatorname{proj}_Q(z).$$

Using the separation oracle of Q and the cutting-plane method to find $\tilde{w}$ such that

$$\left\|\tilde{w} - \operatorname{proj}_Q(z)\right\| \leq \left(1 + \frac{L}{a}\right)\sqrt{\frac{2\eta}{d+1}}.$$

Construct the proposal potential

$$\mathcal{P}_1(w) = \frac{\|w-\tilde{w}\|^2 + \|\tilde{w}-z\|^2}{2\eta} - \frac{\sqrt{2}(L+d)}{a\sqrt{\eta(d+1)}}\left(\|w-\tilde{w}\| + \|\tilde{w}-z\|\right) - \frac{6(L+d)^2}{a^2(d+1)} + as - \frac{a^2\eta}{2}$$

Algorithm RGO implementation

1: Compute the approximate solution $\tilde{w}$ via the CP method by Jiang, Lee, Song, Wong (2020).
2: Generate $U \sim \mathcal{U}[0,1]$ and $w \sim \exp\left(-\mathcal{P}_1(w)\right)$ via a subroutine.
3: If $U \leq \exp\left(-\Theta_{y,s}^{\eta,Q}(w) + \mathcal{P}_1(w)\right)$, then accept w; otherwise, reject w and go to step 2.

Technical Results

Lemma

Evaluation and subgradient oracles of f, and membership and separation oracles of K give a valid separation oracle for Q.

By completing the square, sampling $w \sim \exp\left(-\mathcal{P}_1(w)\right)$ is equivalent to

$$w \sim \Lambda(w) \propto \exp\left(-\frac{1}{2\eta}\left(\|w-\tilde{w}\| - \sqrt{\frac{2\eta}{d+1}}\left(1+\frac{L}{a}\right)\right)^2\right).$$

This becomes one-dimensional sampling using pplar coordinate.

Algorithm Sample $w \sim \Lambda(w)$

1. Generate $W \sim \mathcal{N}(0, I_{d+1})$ and set $\theta = W/\|W\|$;
2. Generate $r \propto r^d \exp\left(-\frac{1}{2\eta}\left(r - \sqrt{\frac{2\eta}{d+1}}\left(1+\frac{L}{a}\right)\right)^2\right)$ by Adaptive Rejection Sampling for one-dimensional log-concave distribution by Gilks and Wild (1992).
3. Output $w = \tilde{w} + r\theta$.

RGO Complexities

Theorem

Assume conditions (B1)-(B4) hold, $a = d$, and $\eta = 1/d^2$.

(a) ***(oracle complexity)*** *There are at most $\mathcal{O}\left(d \log \frac{d\gamma}{\alpha}\right)$ calls to the separation oracle of K and to the subgradient oracle of f. Here $\gamma = R/\text{minwidth}(K)$. Moreover, it holds that $\Pr\left(\alpha \leq \frac{B}{(d+1)^3}\right) \leq 6 \exp\left(-\frac{d^2}{8}\right)$ for $B := \min\left\{\frac{1}{2R(1+3R)}, \frac{1}{12R^2 \max\{L,1\} \max\{\|f\|_\infty,1\}}\right\}$.*

(b) ***(proposal complexity)*** *The average number of proposals is no more than*

$$M \exp\left(\frac{2(L+d)^2}{d^2} + \frac{16(L+d)^2}{d^2(d+1)} + \frac{3C_L(L+d)}{d^{5/2}}\right) \cdot \left[\frac{\sqrt{2\pi}C_L}{d} \exp\left(\frac{C_L^2}{2d^2}\right) + 1\right],$$

where $C_L = \sqrt{a^2 + L^2} + LR + d$. Moreover, if we assume $L = \mathcal{O}(\sqrt{d})$, $R = \mathcal{O}(\sqrt{d})$, and $M = \mathcal{O}(1)$, then the bound simplifies to $\mathcal{O}(1)$.

As an example, it is easy to verify that $K = [-1,1]^d \subseteq \sqrt{d}\,\mathrm{B}_d(0)$ and hence $R = \sqrt{d}$. Also for $f(x) = \|x\|_1$, we can show $\|f'(x)\|_2 \leq \sqrt{d}$ for every $x \in \mathbb{R}^d$, and hence $L = \sqrt{d}$.

Advantages of This Approach

- Algorithm design is conceptually simple. Reduction to uniform sampling via proximal sampler. Only need to implement the RGO.
- Unlike existing constrained samplers that depend on projections, reflections, barrier functions, or mirror maps, this approach enforces feasibility using only separation oracle for Q, which can be implemented by subgradient of f and separation of K.

Outline

An Analogy of Constrained Sampling

Recall the lifted problem in $\mathbb{R}^{d+2}$

$$\tilde{\pi}(x,s,t) \propto \exp\left(-bt - I_{\widetilde{Q}}(x,s,t)\right), \quad \widetilde{Q} = \{(x,s,t) : (x,s) \in \widetilde{K},\ \tilde{f}(x,s) \le bt\}.$$

The proximal sampler construct the augmented distribution

$$\widetilde{\Pi}\left((x,s,t),(y,u,v)\right) \propto \exp\left(-\widetilde{\Theta}^{\eta,\widetilde{Q}}_{y,u,v}(x,s,t)\right)$$

where

$$\widetilde{\Theta}^{\eta,\widetilde{Q}}_{y,u,v}(x,s,t) = I_{\widetilde{Q}}(x,s,t) + \frac{1}{2\eta}\|(x,s,t) - (y,u,v-\eta b)\|^2 + bv - \frac{\eta b^2}{2}.$$

Algorithm Proximal sampler for the target $\tilde{\pi}$

1: **Gaussian**: Generate $(y_k,u_k,v_k) \sim \mathcal{N}\left((x_k,s_k,t_k),\eta I_{d+2}\right)$;
2: **RGO**: Generate $(x_{k+1},s_{k+1},t_{k+1}) \sim \mathcal{N}\left((y_k,u_k,v_k - b\eta),\eta I_{d+2}\right)|_{\widetilde{Q}}$.

Assumptions

(C1) f and h are closed and convex functions on $\mathbb{R}^d$;

(C2) f and h are Lipschitz continuous with constants L_f and L_h, respectively;

(C3) there exists an initial distribution $\tilde{\nu}_0$ that is M-warm with respect to $\tilde{\nu}$ i.e., $\frac{d\tilde{\nu}_0}{d\tilde{\nu}} \leq M$;

(C4) there exists a separation oracle for $\widetilde{Q}$.

In practice, specific oracle of f and h (such as evaluation, subgradient, and proximal) are required to implement the separation oracle for $\widetilde{Q}$.

A proximal oracle for f returns

$$\mathrm{prox}_{\lambda f}(x) := \operatorname*{argmin}_{y \in \mathbb{R}^d} \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}$$

for a given pair (x, λ).

Separation Oracle for $\widetilde{Q}$

Lemma

Evaluation and proximal oracles of f, and membership and separation oracles of K give a valid separation oracle for Q.

- prox_f and prox_h are available.

  prox_h gives a projection oracle for $\mathrm{epi}(h)$, and hence a separation oracle for $\mathrm{epi}(h)$. By a simple scaling, this yields a separation oracle for $\widetilde{K}$. In addition, it is easy to show that prox_f yields a proximal oracle for $\tilde{f}(x, s)$. Applying the above lemma to $\tilde{f}$ and $\widetilde{K}$, we further have a separation oracle for $\widetilde{Q}$.

- f' and prox_h are available.

  The same argument as in the above case gives a separation oracle for $\widetilde{K}$. In addition, it is straightforward to see that f' yields a subgradient oracle for $\tilde{f}(x, s)$. Now, applying a previous lemma to $\widetilde{K}$ and $\tilde{f}$, we have a separation oracle for $\widetilde{Q}$.

RGO Complexities

Theorem

Assume (C1)-(C4) holds, $a = b = d$, and $\eta = 1/d^2$.

(a) ***(oracle complexity)*** *there are at most $\mathcal{O}\left((d+2)\log\frac{(d+2)\gamma}{\alpha}\right)$ calls to the separation oracle of $\widetilde{Q}$. Here γ is bounded from above and α is bounded from below w.h.p.*
Thus, $\mathcal{O}\left((d+2)\log\frac{(d+2)\gamma}{\alpha}\right) = \mathcal{O}\left(d\log d\right)$ in high probability.

(b) ***(proposal complexity)*** *the average number of proposals is at most*

$$M\exp\left(\frac{7}{4} + \frac{16}{d} + \frac{3\widetilde{C}_L}{d^{3/2}}\right)\left[\frac{\sqrt{2\pi}\widetilde{C}_L}{d}\exp\left(\frac{\widetilde{C}_L^2}{2d^2}\right) + 1\right],$$

where $\widetilde{C}_L = \sqrt{b^2 + a^2 + L_f^2} + \sqrt{a^2 + L_h^2}$. Moreover, if we assume $L_f = \mathcal{O}(d)$, $L_h = \mathcal{O}(d)$, and $M = \mathcal{O}(1)$, then the bound becomes $\mathcal{O}(1)$.

Conclusion

- A generalization hierachy in potential functions

  uniform sampling $I_K(x) \prec$ constrained sampling $f(x) + I_K(x) \prec$ composite sampling $f(x) + h(x)$

- A sequence of reductions using epigraph lifting

  composite sampling $\rightarrow$ constrained sampling $\rightarrow$ uniform sampling $\rightarrow$ proximal sampler

- Efficient RGO implementations via minimal oracle access (subgradient, separation)

References

- Gilks and Wild. Adaptive rejection sampling for gibbs sampling. Journal of the Royal Statistical Society: Series C. 1992
- Jiang, Lee, Song, and Wong. An improved cutting plane method for convex optimization, convex-concave games, and its applications. Symposium on Theory of Computing, 2020
- Lee, Shen, and Tian. Structured Logconcave Sampling with a Restricted Gaussian Oracle. Conference on Learning Theory 2021
- Chen, Chewi, Salim, and Wibisono. Improved Analysis for a Proximal Algorithm for Sampling. Conference on Learning Theory 2022
- Kook, Vempala, and Zhang. In-and-out: Algorithmic diffusion for sampling convex bodies. Advances in Neural Information Processing Systems, 2024
- Kook and Vempala. Sampling and integration of logconcave functions by algorithmic diffusion. Symposium on Theory of Computing, 2025
- Dang and Liang. Oracle-based Uniform Sampling from Convex Bodies. Preprint, 2025
- Dang and Liang. Constrained and Composite Sampling via Proximal Sampler. Upcoming, 2026

Thank you!