Convergence error analysis of reflected gradient Langevin dynamics for globally optimizing non-convex constrained problems
（非凸制約問題の大域最適化における反射勾配ランジュバン動力学の収束誤差解析）

Dept. of Mathematical Informatics 48206217 Kanji Sato
Supervisor Akiko Takeda, Professor

1 Introduction

There have been various kinds of non-convex problems in the real-world applications, but many of theoretical studies focus on proving convergence to stationary points. Meanwhile, gradient Langevin dynamics (GLD) and its variants have been used as a framework to provide global convergence guarantees under non-convex settings recently. The studies on GLD started from unconstrained convex problems and expanded to convex constrained non-convex problems very recently. This work proposes reflected gradient Langevin dynamics (RGLD) as an optimization algorithm for non-convex constrained problems and analyzes convergence rate to ϵ-sampling error. The obtained rate improves upon the previous work on convex constrained non-convex problems (see Table 1).

		objective function	
		conv	non-conv
unconstrained		[2] $\tilde{O}(\epsilon^{-2})$	[6] $\tilde{O}(d^4\epsilon^{-2})$
constrained	conv	[1] $\tilde{O}(d^{12}\epsilon^{-12})$	[3] $\tilde{O}(\mathrm{e}^d\epsilon^{-4})$
	non-conv	This work $\tilde{O}(\lambda_*^{-3}\epsilon^{-3})$	

Table 1: The iteration complexity of GLD to converge to the target distribution. $\tilde{O}$ is the order ignoring polylogarithmic factors. conv and non-conv stand for convex and non-convex respectively.

2 Problem Formulation

We solve the following (generally) non-convex problems over a domain $K \subset \mathbb{R}^d$ defined by

$$\min_{x\in\mathbb{R}^d} f(x)$$
$$\text{s.t. } x \in K.$$

2.1 Assumptions

The assumptions imposed on K and f are made here to state the convergence rate.

Assumption 2.1 (Projection-Friendliness)**.** *Projection onto K is efficiently computable by some oracle.*

The proposed algorithm includes reflection step, which is why K needs to be projection-friendly.

Assumption 2.2 (Interior Sphere Condition)**.** *$K \subset \mathbb{R}^d$ is a possibly non-convex open domain (hence connected) such that $0 \in K$ and, as a consequence, K contains a Euclidean ball of radius $r(>0)$.*

This assumption is necessary to guarantee the uniqueness of the invariant density of the reflected diffusion discussed in the following, which is a key component of global optimization of RGLD.

Assumption 2.3 (Smoothness of Boundary)**.** *$\partial K \in C^4$.*

Assumption 2.3 is a relatively strict condition because constrained problems may have multiple constraints and, thus, have several indifferentiable extreme points. Meanwhile, there are some examples which satisfy the assumption above, for instance, thick-walled sphere, Riemannian manifold such as Stiefel manifold and Grassmann manifold. l_q norm $(q < 1)$ constraint is one of the most popular non-convex constraint for sparse estimation but it violates Assumption 2.3 at the points with zero elements.

Assumption 2.4 (Bounded Domain)**.** *K is bounded, that is, K is included by the sphere of radius R centered at the origin.*

Assumption 2.5 (Smoothness of Objective Function)**.** *f is C^4 function and ∇f is accessible by some oracle.*

From these assumptions, we can conclude that f and ∇f are bounded. Moreover, M-smoothness is also easily established.

Proposition 2.6. *Under Assumptions 2.4 and 2.5, f is M-smooth, that is,*

$$\|\nabla f(x) - \nabla f(y)\| \le M\|x - y\|.$$

Finally, the existence of the optimal solution is assumed.

Assumption 2.7. *The above problem admits at least one optimal solution $x^* \in K$.*

2.2 Proposed Optimization Algorithm: RGLD

We propose reflected gradient Langevin dynamics (RGLD) as an optimization algorithm updated by

$$X_{k+1} = \mathcal{R}_K \left(X_k - \eta \boldsymbol{\nabla} f(X_k) + \sqrt{\frac{2\eta}{\beta}} \xi_{k+1} \right)$$

where η and β are step size and inverse temperature parameter respectively, and ξ_k is i.i.d. Rademacher random variables in $\mathbb{R}^d$. $\mathcal{R}_K$ is reflection operator to keep the trajectory inside the feasible region K, defined by $\mathcal{R}_K(x) = 2\mathcal{P}_K(x) - x$ where $\mathcal{P}_K(x) = \arg\min_{y \in K} \|y - x\|$.

3 Main Theorem

In this work, we derive a convergence rate of RGLD to ensure small expected excess risk of the constrained problem, that is,

$$\mathbb{E}\left[f(X_k)\right] - \min_{x \in K} f(x) \leq \epsilon + C_\beta,$$

where X_k is the solution found in the k-th iteration by RGLD and C_β is some constant depending on a hyperparameter of RGLD and the dimensionality of the problem. More technically, we analyze the convergence rate to ϵ-sampling error of the target distribution, that is,

$$\mathbb{E}\left[f(X_k)\right] - \mathbb{E}_\pi f \leq \epsilon,$$

where $\mathbb{E}_\pi$ denotes taking expectation with respect to the target distribution π.

To state the main theorem, we also define $[n] = \{0, 1, \ldots, n-1\}$. Then, the following notation is introduced for simplicity of arguments on convergence rate.

Definition 3.1.

$$f \preceq g \triangleq \exists C > 0,\ f \leq Cg,$$
$$f \sim g \triangleq f \preceq g \wedge f \succeq g.$$

Now we state our main theorem.

Theorem 3.2 (Main Theorem)**.** *For any $\epsilon \preceq 1$, by setting the inputs of algorithm as*

$$\beta \succeq 1,$$
$$\eta \preceq \min\left\{\frac{\lambda_*^2 \epsilon^2}{\beta^3}, \frac{\lambda_* \epsilon}{\sqrt{\beta} d}\right\},$$
$$N \succeq \frac{1}{\lambda_* \epsilon \eta} \succeq \max\left\{\frac{\beta^3}{\lambda_*^3 \epsilon^3}, \frac{\sqrt{\beta} d}{\lambda_*^2 \epsilon^2}\right\},$$

we have

$$\mathbb{E}\left[\min_{k \in [N]} f(X_k)\right] - \min f \leq \epsilon + \frac{d \log \beta}{\beta}.$$

The proof is deferred to our thesis.

4 Conclusion

In this work, we prove sub-linear rate convergence of RGLD under constrained non-convex problems. Additionally, the obtained rate is sharper than that of [3]. However, the order of the spectral gap remains to be specified because the smallest eigenvalue problems with Neumann boundary conditions have not been studied well. Moreover, the smoothness assumption on the boundary of the domain is essential in the analysis but it is not satisfied in many applications. The future works will relax the assumptions to more general unbounded non-smooth constraints. One promising approach is to assume dissipativity condition, as imposed in [4, 5]. We are also curious about the theoretical relationships with mirrored Langevin dynamics and Riemannian Langevin dynamics as solvers for constrained optimization.

References

[1] S. Bubeck, R. Eldan, and J. Lehec. Sampling from a log-concave distribution with projected Langevin Monte Carlo. *Discrete & Computational Geometry*, 59(4):757–783, 2018.

[2] A. Dalalyan. Further and stronger analogy between sampling and optimization: Langevin Monte Carlo and gradient descent. In S. Kale and O. Shamir, editors, *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 678–689. PMLR, 07–10 Jul 2017.

[3] A. Lamperski. Projected stochastic gradient langevin algorithms for constrained sampling and non-convex learning. In M. Belkin and S. Kpotufe, editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 2891–2937. PMLR, 15–19 Aug 2021.

[4] M. Raginsky, A. Rakhlin, and M. Telgarsky. Non-convex learning via stochastic gradient Langevin dynamics: a nonasymptotic analysis. In *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 1674–1703, 2017.

[5] P. Xu, J. Chen, D. Zou, and Q. Gu. Global convergence of Langevin dynamics based algorithms for nonconvex optimization. In *Advances in Neural Information Processing Systems 31*, pages 3122–3133. 2018.

[6] D. Zou, P. Xu, and Q. Gu. Faster convergence of stochastic gradient Langevin dynamics for non-log-concave sampling. In C. de Campos and M. H. Maathuis, editors, *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pages 1152–1162. PMLR, 27–30 Jul 2021.