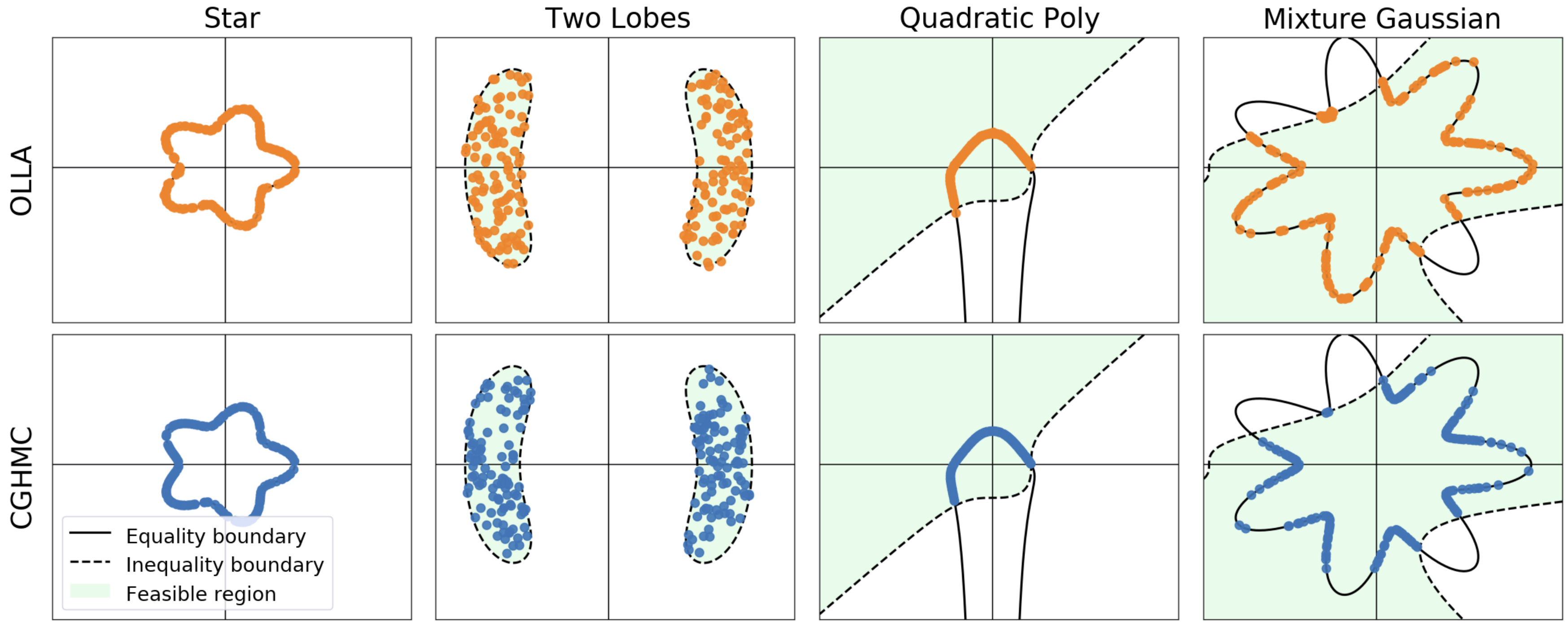
Fast Non-Log-Concave Sampling under Nonconvex Equality and Inequality Constraints with Landing

Kijung Jeon[1], Michael Muehlebach[2], Molei Tao[1]

[1]Georgia Institute of Technology, [2]MPI-IS

Circumstance: Want to sample **(fast)** from unnormalized density $\rho_\Sigma(x)$ with equality $h(x)$ and/or inequality $g(x)$ constraints.



39th Conference on Neural Information Processing Systems (NeurIPS 2025)

Motivation

Goal: sample from the target distribution supported on a (possible nonconvex) feasible set:

$$\Sigma := \{x \in \mathbb{R}^d \mid h(x) = 0,\ g(x) \leq 0\}, \qquad \rho_\Sigma(x) \propto e^{-f(x)} d\sigma_\Sigma(x)$$

where $h: \mathbb{R}^d \to \mathbb{R}^m$, $g: \mathbb{R}^d \to \mathbb{R}^l$ are smooth constraints, $d\sigma_\Sigma$ is the induced Hausdorff measure on Σ.

Prior Works

1. Barrier method or soft-penalty Langevin may distort the stationary distribution, and mirror Langevin requires convex geometry of Σ.

2. Prior constrained samplers (e.g., CLangevin, CHMC, CGHMC) often require projections (costly / intractable on nonconvex Σ).

3. Current landing-based samplers lack exponential convergence rate guarantee and focus only on either equality / inequalities (not both).

Position of this paper

A projection-free Langevin framework (OLLA) unifying equality and inequality constraints.

Construction of OLLA

Intuition: Find drift q, diffusion Q closest to Euclidean Langevin while satisfying the target property.

Lemma 1 (Exponential decay of constraint functions). *The following properties hold almost surely for* $\forall i \in [m], \forall j \in I_{X_0}$:

$$h_i(X_t) = h_i(X_0)e^{-\alpha t}, \quad t \geq 0 \tag{7}$$

and

$$\begin{cases} g_j(X_t) = -\epsilon + (g_j(X_0)+\epsilon)e^{-\alpha t}, & t \leq \frac{1}{\alpha}\ln\left(\frac{g_j(X_0)+\epsilon}{\epsilon}\right) \\ g_j(X_t) \leq 0, & t \geq \frac{1}{\alpha}\ln\left(\frac{g_j(X_0)+\epsilon}{\epsilon}\right) \end{cases}$$

with $g(X_t) \leq 0, \forall t \geq 0$ *for* $j \notin I_{X_0}$, *where* $I_x := \{k \in [l] \mid g_k(x) \geq 0\}$ *is the index set of active inequality constraints.*

[Lemma 1] Target property for constraints

1. **(Landing)** Equality constraint h should converge to 0 exponentially fast.
2. **(Landing &Repulsion)** Inequality constraint g should satisfy within finite time, after which violation is not permitted.

Proposition 1 (Construction of OLLA and its closed form SDE). *Consider the following SDE:*

$$dX_t = q(X_t)dt + Q(X_t)dW_t \tag{2}$$

where

$$Q := \underset{\bar{Q} \in \mathbb{R}^{d\times d}}{\operatorname{argmin}} \|\sqrt{2}I - \bar{Q}\|_F^2 \quad s.t \quad \begin{cases} \bar{Q}\nabla h_i = 0, \ \forall i \in [m], \\ \bar{Q}\nabla g_j = 0, \ \forall j \in I_x, \end{cases}$$

$$q := \underset{\bar{q} \in \mathbb{R}^{d}}{\operatorname{argmin}} \|\bar{q} + \nabla f\|_2^2 \quad s.t \quad \begin{cases} \nabla h_i^T \bar{q} + \frac{1}{2}\mathrm{Tr}\left(\nabla^2 h_i Q Q^T\right) + \alpha h_i = 0, & \forall i \in [m], \\ \nabla g_j^T \bar{q} + \frac{1}{2}\mathrm{Tr}\left(\nabla^2 g_j Q Q^T\right) + \alpha (g_j + \epsilon) = 0, & \forall j \in I_x. \end{cases}$$

Then, there exists a closed form SDE (OLLA) of (2) given by:

$$dX_t = -[\Pi(X_t)\nabla f(X_t) + \alpha \nabla J(X_t)^T G^{-1}(X_t) J(X_t)]dt + \mathcal{H}(X_t)dt + \sqrt{2}\Pi(X_t)dW_t \tag{3}$$

where

$$\mathcal{H} := -\nabla J^T G^{-1}\left[\mathrm{Tr}\left(\nabla^2 h_1 \Pi\right), \dots, \mathrm{Tr}\left(\nabla^2 h_m \Pi\right), \mathrm{Tr}\left(\nabla^2 g_{i_1} \Pi\right), \dots, \mathrm{Tr}\left(\nabla^2 g_{i_{|I_x|}} \Pi\right)\right]^T \tag{4}$$

is the associated mean curvature correction term of $\Sigma_{I_x} := \{x \in \mathbb{R}^d \mid h(x) = 0, g_{I_x}(x) = 0\}$.

[Proposition 1] Construction of OLLA

1. Apply Ito's lemma on diffusion process
2. Kill the noise (two constraints).
3. Equate deterministic parts to accomplish target properties (two constraints).

Convergence of OLLA

Exponential convergence of OLLA on all scenarios [Theorem 1, 2, 3]

With α = landing rate, λ_{LSI} = LSI constant of ρ_Σ on Σ, and ρ_t = law of X_t following OLLA dynamics.

1. **(Equality-only)**

$$W_2(\rho_t, \rho_\Sigma) = \mathcal{O}(e^{-\alpha t}) + \mathcal{O}\left(e^{-\lambda_{LSI} t}\right)$$

2. **(Inequality-only)**

$$W_2(\rho_t, \rho_\Sigma) = \mathcal{O}\left(e^{-\lambda_{LSI} t}\right)$$

3. **(Mixed-case)**

$$W_2(\rho_t, \rho_\Sigma) = \mathcal{O}(e^{-\alpha t}) + \mathcal{O}\left(e^{-\lambda_{LSI} t}\right)$$

for $t \geq t_{cut}$, where t_{cut} depends on the landing rate α and the repulsion rate ϵ.

Remark on assumptions [Appendix A]

(Equality-only) and (Inequality-only) requires compact Σ, boundness of constraints at $t = 0$, and LICQ.
(Mixed-case) further requires stronger regularity conditions of Σ.

Computational gain of OLLA

Implementation of OLLA: Euler-Maruyama with Hutchinson estimator for $\mathcal{H}$ (OLLA-H).

Computational complexity of baselines

$$\{\mathrm{CLangevin}, \mathrm{CHMC}\} = \mathcal{O}\left(N_{\mathrm{newton}} \cdot (m + l)^3\right), \qquad \mathrm{CGHMC}^* = \mathcal{O}\left(N_{\mathrm{newton}} \cdot m^3\right)$$

(Note*: the sampling efficiency of CHHMC ∝ acceptance rate; hence efficiency degrades as constraint becomes complicated)

[Notations]

- m = # of equalities
- l = # of inequalities
- N= # of Hutchinson probes
- N_{newton} = # of Newton iterations
- $|I_x|$ = # of active inequalities

Computational complexity of OLLA-H

$$\mathrm{OLLA-H} = \mathcal{O}\left(N \cdot (m + |I_x|)^3\right)$$

In particular, when $\boldsymbol{N \ll N_{\mathrm{newton}}}$ and $\boldsymbol{|I_x| \ll l}$, OLLA-H significantly accelerate the sampling.

Experiment results (Synthetic 2D)

Setup: Equality-only (Star), Inequality-only (Two Lobes), Mixed (Quadratic Poly, Mixture Gaussian)

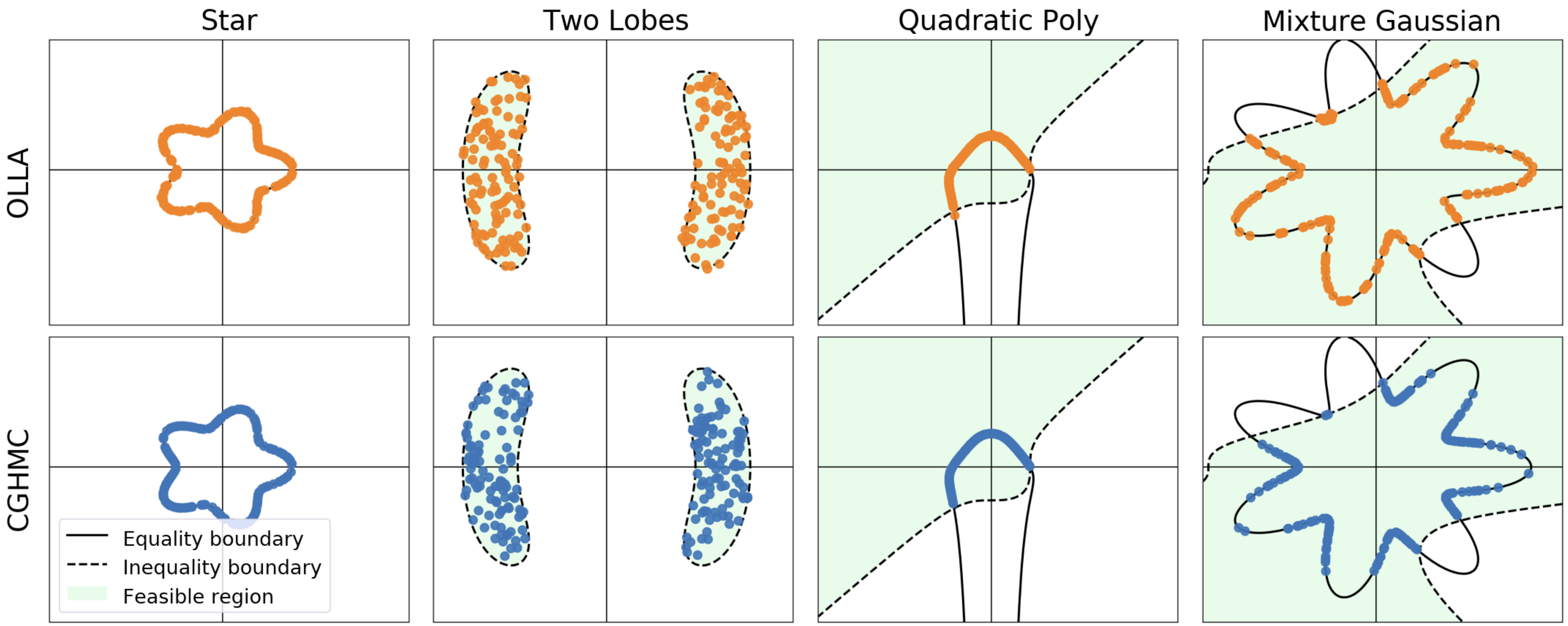


Table 1: Effect of α on $W_2^2, \mathbb{E}[|h|]$

α	W_2^2	$\mathbb{E}[\lvert h(x)\rvert]$
1	0.363 ± 0.064	0.682 ± 0.017
10	0.200 ± 0.035	0.130 ± 0.001
100	0.159 ± 0.032	0.017 ± 0.001
200	0.121 ± 0.019	0.008 ± 0.001

Table 2: Effect of ϵ on $W_2^2, \mathbb{E}[\max g^+]$

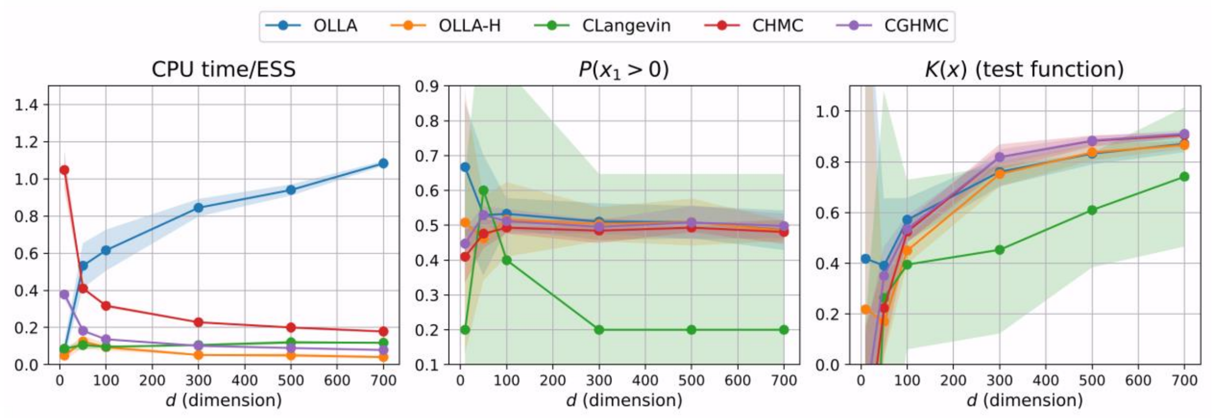
ϵ	W_2^2	$\mathbb{E}[\max g^+(x)]$
0.1	0.151 ± 0.026	0.082 ± 0.017
1	0.108 ± 0.011	0.067 ± 0.027
5	0.123 ± 0.018	0.040 ± 0.015
10	0.112 ± 0.034	0.019 ± 0.006

- The sampling outcome matches the one from slack-extension of CGHMC (based on MH correction).
- Reasonably high α, ϵ tends to give better constraint satisfaction and sampling quality.

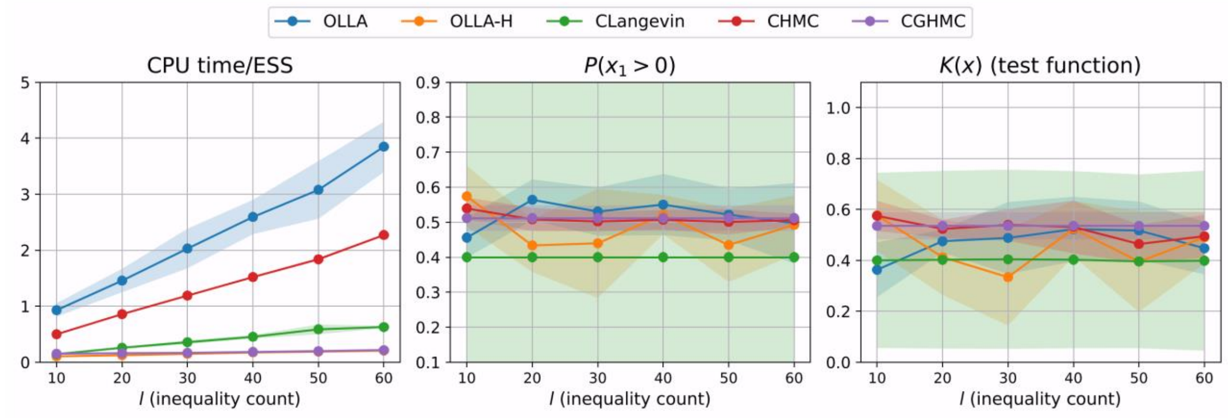
Experiment results (High-dimensional data)

Setup: $h_i(x) = a_i^T x - b_i,\ h_m(x) = \|x\|_2^2 - R^2,\ g_j(x) = r^2 - \left\|x - c_j\right\|_2^2$

with $a_i \sim N(0, I), b_i \sim N\left(0, 0.1^2\right)\ c_j \sim N\left(0, \sqrt{R/2I}\right)$ for fixed $R = 5, r = 1$ and for $i \in [m-1], j \in [l]$



(dimension scalability): d increases, $m = 5, l = 5$



(Inequality constraint scalability): $d = 100, m = 5, l$ increases

- OLLA-H is consistently faster than (projection-based) CLangevin, CHMC, and CGHMC.
- OLLA-H is effective for large dimension, inequality constraints; requires α tuning for large equality constraints **(limitation)**.

Experiment results (Molecular System)

Setup: (1) equality h for fixed bond lengths and angles, (2) inequality g for steric hindrance, and (3) potential f modeling Weeks-Chandler-Anderson (WCA) potential.

Table 3: Estimates of radius of gyration squared (R_g^2) and total CPU time for sampling (in brackets) on the molecular system with realistic potential (complex Σ, small $|I_x|$, large N_{newton}). The average constraint violation for OLLA-H was below 0.007 (equality), while projection-based samplers maintained violations below 0.0001. Inequality violations were observed to be 0 for all algorithms. The (d, m, l) configurations are: $(15, 7, 6)$, $(30, 17, 36)$, $(45, 27, 91)$, $(60, 37, 171)$, $(90, 57, 406)$.

method / dim (d)	15	30	45	60	90
OLLA-H ($N = 0$)	1.392 ± 0.026 (**78**s)	5.414 ± 0.047 (**151**s)	12.240 ± 0.102 (**254**s)	21.820 ± 0.098 (**334**s)	49.080 ± 0.223 (**704**s)
OLLA-H ($N = 5$)	1.370 ± 0.032 (182s)	5.424 ± 0.045 (400s)	12.200 ± 0.155 (656s)	21.780 ± 0.098 (916s)	48.940 ± 0.273 (1732s)
CLangevin	1.396 ± 0.012 (421s)	5.526 ± 0.015 (3620s)	12.400 ± 0.000 (10940s)	22.140 ± 0.049 (22600s)	49.960 ± 0.049 (52220s)
CHMC	1.410 ± 0.000 (147s)	5.580 ± 0.000 (468s)	12.500 ± 0.000 (1012s)	22.200 ± 0.000 (1712s)	49.960 ± 0.049 (3660s)
CGHMC	1.410 ± 0.000 (135s)	5.580 ± 0.000 (282s)	12.500 ± 0.000 (467s)	22.200 ± 0.000 (652s)	49.940 ± 0.049 (1116s)

- OLLA-H (even at $N = 0$) shows comparable sampling quality with significantly lower computational cost and reasonable constraint satisfaction under many constraints.

Experiment results (Bayesian Logistic Regression)

Setup: (1) equality h for fairness constraint, (2) inequality g for monotonicity constraints, and (3) potential f for the posterior distribution.

Table 4: Test NLL and total CPU time for sampling (in brackets) on the Bayesian logistic regression with fairness and monotonicity constraints (high-dimensional Σ). The average constraint violations for OLLA-H were below 0.005 (equality) and 0.15 (inequality), while projection-based samplers (CLangevin, CHMC, CGHMC) maintained feasibility below 0.0008 (equality) and no inequality violation.

method / dim (d)	1986	4994	9986	49986	100002
OLLA-H ($N = 0$)	0.514 ± 0.013 (**63**s)	0.521 ± 0.008 (**70**s)	0.524 ± 0.014 (**70**s)	0.523 ± 0.011 (**81**s)	0.520 ± 0.015 (**82**s)
OLLA-H ($N = 5$)	0.520 ± 0.013 (159s)	0.524 ± 0.008 (180s)	0.505 ± 0.004 (205s)	0.517 ± 0.011 (189s)	0.516 ± 0.004 (197s)
CLangevin	0.573 ± 0.004 (1162s)	0.568 ± 0.013 (1176s)	0.564 ± 0.022 (1194s)	0.580 ± 0.005 (1428s)	0.570 ± 0.011 (1370s)
CHMC	0.599 ± 0.015 (526s)	0.595 ± 0.020 (532s)	0.599 ± 0.017 (561s)	0.606 ± 0.004 (586s)	0.605 ± 0.004 (611s)
CGHMC	0.600 ± 0.007 (76s)	0.600 ± 0.009 (77s)	0.606 ± 0.003 (82s)	0.598 ± 0.020 (83s)	0.601 ± 0.007 (88s)

- CGHMC suffers from high rejection rate at large step size, leading to low NLL loss.
- OLLA-H (even at $N = 0$) can reduce the computational cost while maintaining sampling stability.

Conclusion

1. OLLA provides a unified, projection-free framework for sampling under mixed equality and inequality constraints.

2. It achieves exponential convergence guarantees and practical efficiency through the landing dynamics and Hutchinson-based discretization.

3. The method scales to high-dimensional, non-log-concave targets and enables fast constrained sampling compared to prior projection-based samplers.

Code: https://github.com/KraitGit/OLLA || Paper: https://arxiv.org/abs/2510.22044 || Email: kjeon@gatech.edu