Sampling with Riemannian Hamiltonian Monte Carlo in a Constrained Space

Yunbum Kook[1], Yin Tat Lee[2,3], Ruoqi Shen[2], Santosh S. Vempala[1]

[1]Georgia Institute of Technology [2]University of Washington [3]Microsoft Research

Sampling from Constrained Distributions

- **CRHMC*:** A package to sample from ill-conditioned, non-smooth, constrained distributions up to very high dimension efficiently.
- Sample from

$$\frac{d\pi(x)}{dx} \propto \exp\big(-f(x)\big) \;\; s.t. \; \boxed{Ax = b, x \in K.}$$

Constrained space; Can be skewed

for convex f and convex body K.

Method: Run a Markov chain with stationary distribution π until convergence.

Challenges:
- A naïve algorithm will not maintain the constraints $Ax = b, x \in K$.
- The mixing time can depend on the condition number of K.
- How to leverage the sparsity inherent in A?

Example:
Human Metabolic Network with 8,399 reactions and 13,543 metabolites.

Results:
- Popular sampling packages such as STAN and Pyro cannot move at all.
- ACHR takes >3 years per sample.
- CHRR (Coordinate Hit-and-Run) takes 8 hours per sample.
- **Our algorithm takes 31 sec per sample.**

Metabolic networks

Our package has been incorporated into the COBRA toolbox.

Riemannian Hamiltonian Monte Carlo

Hamiltonian Monte Carlo

- Define the *Hamiltonian*: $H(x,v) = f(x) + \frac{1}{2}\|v\|^2$,
- Repeat the following until convergence:
  - Draw $v_0 \sim \mathcal{N}(0, I)$
  - Solve the *Hamiltonian Eqs.* (ODE) with initial condition (x, v_0) and step size h:

$$\frac{dx}{dt}(x,v) = \frac{\partial H}{\partial v}(x,v), \qquad \frac{dv}{dt}(x,v) = -\frac{\partial H}{\partial x}(x,v).$$

- The stationary distribution is $\propto \exp(-H(x,v))$.

Constrained Riemannian Hamiltonian Monte Carlo (CRHMC)

Use the local geometry of the density function

$$H(x,v) = f(x) + \frac{1}{2}v^\top M(x)^\dagger v + \frac{1}{2}\log \mathrm{pdet}\, M(x)\,.$$

Local metric

* In RHMC, the velocity v_0 is drawn from $\mathcal{N}(0, M(x))$.
Using a carefully chosen M, RHMC can maintain the constraints and achieve a mixing time independent of the condition number.

(1) Equality constraints

To maintain $c(x) = 0$ (ex. $c(x) = Ax - b$), we want

$$\frac{d}{dt}c(x_t) = Dc(x_t)\cdot\frac{dx_t}{dt} = Dc(x_t)\cdot\frac{\partial H(x_t,v_t)}{\partial v_t} = Dc(x_t)\cdot M(x)^\dagger v = 0.$$

A natural choice is the orthogonal projection to $\mathrm{Null}(Dc(x))$:

$$Q(x) = I - Dc(x)^\top (Dc(x)Dc(x)^\top)^{-1}Dc(x).$$

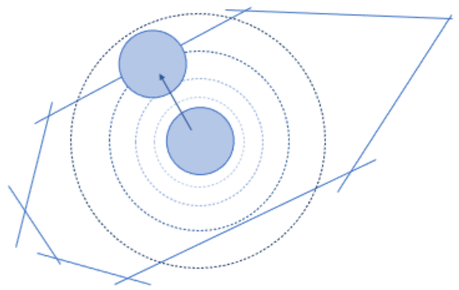
(2) Inequality constraints

For general convex body K, we can use a *self-concordant barrier*, a function defined on K such that $\phi(x)$ is self-concordant and $\phi(x) \to \infty$ as $x \to \partial K$.
Using the barrier ϕ, we employ the local metric defined by $g(x) = \nabla^2\phi(x)$.

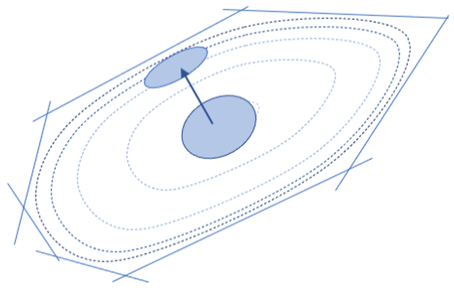
The choice of metric $\boldsymbol{M = Q(x)^\top g(x) Q(x)}$ satisfies both constraints.

Computation

- The computation of $\frac{\partial H}{\partial x}$ and $\frac{\partial H}{\partial v}$ may involve dense matrix inverse. We introduce an efficient algorithm for the problem when K is a product of convex bodies K_i , each with small dimension.
- Traditional integrators such as Leapfrog don't work. We use Implicit Midpoint Method.
  (See the paper and arXiv:2210.07219 for more details.)



Euclidean Metric



Riemannian Metric

Experiments

Settings. We performed experiments on the Standard DS12 v2 model from MS Azure cloud, which has a 2.1GHz Intel Xeon Platinum 8171M CPU and 28GB memory.

Comparison. We used as a baseline the Coordinate Hit-and-Run (CHAR) implemented in two different languages. The former is Coordinate Hit-and-Run with Rounding (CHRR) written in MATLAB [1] and the latter is the same algorithm (CDHR) with an R interface and a C++ library, VolEsti [2]. Popular sampling packages such as STAN and Pyro were not included in the experiments as they do not support constrained-based models.

Bio Model	Vars (n)	nnz	CRHMC	CHRR	CDHR
ecoli	95	291	0.0098	0.0365	0.0022
cardiac_mit	220	228	0.0100	0.0059	0.0005
Aci_D21	851	1758	0.4257	0.6884	0.2974
Aci_MR95	994	2859	0.9624	2.0668	0.5237
Abi_49176	1069	2951	0.9608	1.9395	0.9622
Aci_20731	1090	2946	0.1540	2.3014	1.1086
Aci_PHEA	1561	4640	0.3701	12.06	-
iAF1260	2382	6368	4.4355	3687.2	-
iJO1366	2583	7284	4.1608	70.5	35.556
Recon1	3742	8717	0.7184	208.5	-
Recon2	7440	19791	2.6116	10445*	-
Recon3	13543	48187	31.114	29211*	-

LP Model	Vars (n)	nnz	CRHMC	CHRR	CDHR
israel	316	2519	0.1186	1.2224	0.4426
gfrd_pnc	1160	2393	0.2199	40.988	18.468
25fv47	1876	10566	0.8159	199.9	-
pilot_ja	2267	11886	1.3490	5059*	-
sctap2	2500	7334	0.6752	520.2	-
ship08l	4363	9434	0.6258	6512	-
cre_a	7248	17368	2.2205	30455*	-
woodw	8418	23158	2.0689	30307*	-
80bau3b	12061	22341	11.881	47432*	-
ken_18	154699	295946	1616.3	-	-

Figure 1: Sampling time per effective sample of CRHMC and the competitors.

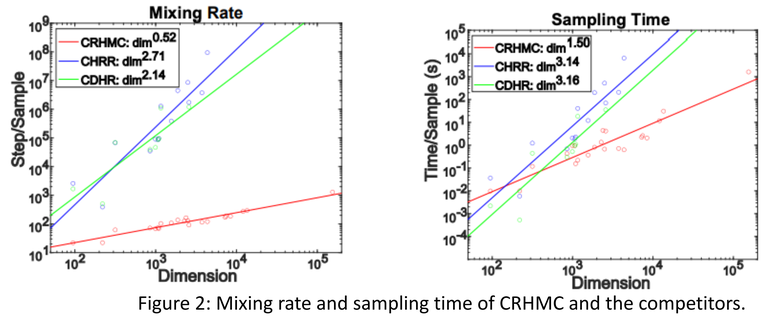


Figure 2: Mixing rate and sampling time of CRHMC and the competitors.

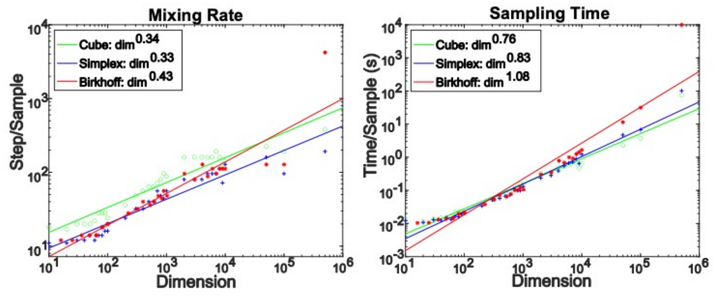


Figure 3: Mixing rate and sampling time on structured polytopes including hybercubes, simplices, and Birkhoff polytopes. **To the best of our knowledge, this is the first demonstration that it is possible to sample such a large model.**

Reference [1] Hulda S Haraldsdóttir, Ben Cousins, Ines Thiele, Ronan MT Fleming, and Santosh Vempala. Chrr: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. Bioinformatics, 33(11):1741–1743, 2017.

[2] Apostolos Chalkis and Vissarion Fisikopoulos. volEsti: Volume approximation and sampling for convex polytopes in R. arXiv preprint arXiv:2007.01578, 2020.

Acknowledgement This work was supported in part by NSF awards DMS-1839116, DMS-1839323, CCF-1909756, CCF-2007443 and CCF-2134105.