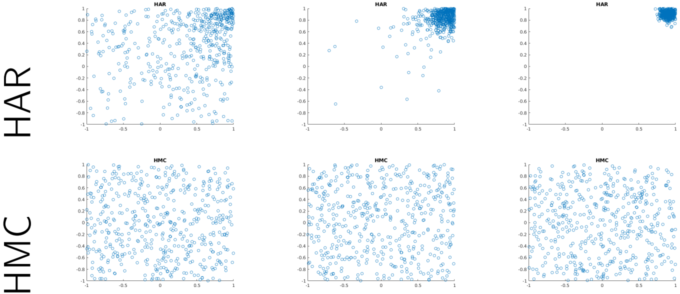
Randomized algorithms for volume/density of states calculations in high-dimensional spaces



dim. = 5   *dim.* = 10   *dim.* = 50

Augustin Chevallier, Sylvain Pion, Frédéric Cazals

Inria, France

Randomized algorithms for volume / DoS calculations

Motivation: volume and DoS calculations

Background: polytope volume calculations

Background: HMC

Novel HMC random walk

Experiments

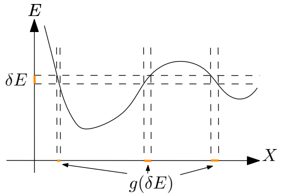
Randomized algorithms for volume / DoS calculations

Free energy, density of states, and volume calculations:

why are these questions so difficult?

▷ Partition function and density of states:

$$Z = \sum_{x_i:\text{state}} e^{-\beta E(x_i)} = \sum_{j:\text{energy level}} g(E_j) e^{-\beta E_j}$$



▷ Computing the volume of a slice in phase space:
- Best possible method in simple cases, e.g. when the region is a polytope?
- Connection to randomized algorithms (MCMC) used in statistical physics? Connection to random walks?

▷ Volume of polytopes, hardness: (Thm.) The volume of a polytope $K \subset \mathbb{R}^d$ can be approximated within a relative error ε with probability $1-\delta$ using $O^\star(n^4)$ oracle calls.

▷Ref: Dyer, Freeze, Kannan, J. ACM 38(1), 1991
▷Ref: Lovász, Vempala, J. Comput. Syst. Sci., 71(2), 2006

Randomized algorithms for volume / DoS calculations

Volume of polytopes: definition and hardness

▷ Polytope K in $\mathbb{R}^n$:

- H-polytope: (bounded) intersection of half-spaces
- V-polytope: convex hull of its vertices
- NB: for h facets, the max number of faces of all dimensions is $O(h^{\lfloor n/2 \rfloor})$

▷ Polytope oracle:

- Membership oracle: answers $p \in K$ (+ returns a linear separator when $p \notin K$)
- Boundary oracle: intersecton between a line l and the boundary

▷ Volume computation hardness:

Any polynomial time computing an upper and lower bounds of the volume of a convex body $K \in \mathbb{R}^n$, $\exists$ constant $c > 0$ such that

$$\frac{\overline{Vol}(K)}{\underline{Vol}(K)} \leq \left(\frac{cn}{\log n}\right)^n \tag{1}$$

cannot be guaranteed for $n \geq 2$.

▷`Ref: Bollobás, in` *`Favors of geometry`*`, 1997`

Randomized algorithms: typical results

▷ ε-approximation of the volume: for any parameter $\varepsilon > 0$, a number V

$$(1-\varepsilon)\mathrm{Vol}(K) \leq V \leq (1+\varepsilon)\mathrm{Vol}(K).$$

▷ (ε, δ)-approximation algorithm: algorithm returning an ε-approximation with a probability at least $1-\delta$.

▷ Complexity, the $O^\star(n)$otation:

- $O(n^4)$: upper bound as a function of the dimension n
- $O^\star(n^4)$: term in $\log n, \varepsilon, \delta$ removed; focus on the dimension solely

Randomized algorithms: complexity

▷ Thm. For a convex body K given by a membership oracle, and such that $B \subset K \subset RB$, an $(\varepsilon, \delta)-$ approximation can be obtained in time

$$O(\frac{n^4}{\varepsilon^2} \log^9 \frac{n}{\varepsilon\delta} + n^4 \log^8 \frac{n}{\delta} \log R) = O^\star(n^4) \tag{2}$$

▷ Volume estimated from:

$$\mathrm{Vol}(K) = \int_K f_0(x)dx \frac{\int_K f_1(x)dx}{\int_K f_0(x)dx} \cdots \frac{\int_K dx}{\int_K f_{m-1}(x)dx} \equiv \int_K f_0(x)dx \prod_{i=1,\dots,m} R_i \tag{3}$$

▷ Cooling schedule:

- f_0: sharply peaked in K
- f_{m-1}: essentially the uniform distribution

▷ Complexity, overview: $m = O^\star(\sqrt{n})$ functions used. Each ratio in the telescoping product is estimated (with guarantees) using $O^\star(\sqrt{n})$ samples. The complexity of generating a given sample being $O^\star(n^3)$, the overall algorithm has complexity $O^\star(n^4)$.

▷Ref: Lovász, Vempala, J Comp. Syst. Sciences, 2006

▷Ref: Cousins, Vempala, SIAM J. Comp., 2018

A practical algorithm: outline

▷ Method: multi-phase Monte-Carlo using $m = O(\sqrt{n})$ logconcave[1] functions $\{f_0, \dots, f_{m-1}\}$

(Exponential) $f_i(x) = \frac{e^{-a_i^T x}}{\int_K e^{-a_i^T y} dy}$ (Gaussian) $f_i(x) = exp(-a_i \|x\|^2)$

Volume(K, ε): Convex body K, error parameter ε.

- $T =$ **Round**(body: K, steps: $8n^3$), set $K' = T \cdot K$.
- $\{a_0, \dots, a_m\} =$ **GetAnnealingSchedule**(body: K').
- Set x to be random point from $f_0 \cap K'$, $\varepsilon' = \varepsilon/\sqrt{m}$.
- For $i = 1, \dots, m$,
  - Set $k = 0, x_0 = x, converged = false, W = 4n^2 + 500$.
  - While $converged = false$,
    - $k = k + 1$.
    - x_k=**HitAndRun**(body: K, target distribution: f_{i-1}, current point: x_{k-1}).
    - Set
    $$r_k = \frac{1}{k} \sum_{j=1}^{k} \frac{f_i(x_j)}{f_{i-1}(x_j)}.$$
    - Set $W_{max} = \max\{r_{k-W+1}, \dots, r_k\}$ and $W_{min} = \min\{r_{k-W+1}, \dots, r_k\}$.
    - If $W_{max} - W_{min} \le \varepsilon'/2 \cdot W_{max} \to converged = true$.
  - Set $R_i = r_k, x = x_k$.
- Return $volume = |T| \cdot (\pi/a_0)^{n/2} \cdot R_1 \dots R_m$.

▷Ref: Cousins and Vempala, Math. Prog. Comp., 2016

[1] $\log f(\alpha x + (1-\alpha)y) \ge \alpha \log f(x) + (1-\alpha) \log f(y)$

Ingredient: importance sampling

▷ Classical Monte Carlo integration: using iid $X_i \sim p$, estimate

$$\mathbb{E}_p[f] = \int_{\mathbb{R}^n} f(x)p(x)dx. \tag{4}$$

as

$$Z_N = \frac{1}{N}\sum_{i=1}^{N} f(X_i). \tag{5}$$

▷ Importance sampling: using iid RV $Y_i \sim q$, estimate

$$\mathbb{E}_p[f] = \int_{\mathbb{R}^n} f(x)p(x)dx = \int_{\mathbb{R}^n} \frac{f(x)p(x)}{q(x)} q(x)dx = \mathbb{E}_q[fp/q]$$

as

$$Z'_N = \frac{1}{N}\sum_{i=1}^{N} \frac{f(Y_i)p(Y_i)}{q(Y_i)}. \tag{6}$$

▷ Benefit: variance reduction

▷Ref: `Brooks et al, Handbook of MCMC`

Application to the ratios R_i

- Recall $R_i = \int_K \frac{f_i(x)}{f_{i-1}(x)} \frac{f_{i-1}(x)}{\int_K f_{i-1}(x)dx} dx$
- Define $Y = f_i(X)/f_{i-1}(X)$, with $X \sim f_{i-1}(X)$.
- One has

$$\mathbb{E}[Y] = \int_K \frac{f_i(x)}{f_{i-1}(x)} \frac{f_{i-1}(x)}{\int_K f_{i-1}(x)dx} dx = \frac{\int_K f_i(x)dx}{\int_K f_{i-1}(x)dx}. \tag{7}$$

- Associated estimator: with X_i a set of k iid RV $\sim {f_{i-1}(x)}/{\int_K f_{i-1}(y)dy}$:

$$\tilde{R}_i = \frac{1}{k} \sum_j \frac{f_i(X_j)}{f_{i-1}(X_j)}. \tag{8}$$

- Importance sampling in disguise: R_i has the form $\int {f(x)p(x)}/{q(x)} q(x)dx$ with

$$p(x) = 1/Vol(K), \quad f(x) = \frac{f_i(x)Vol(K)}{\int_K f_{i-1}(y)dy} \quad q(x) = \frac{f_{i-1}(x)}{\int_K f_{i-1}(y)dy}. \tag{9}$$

Ingredient: sampling with random walks

▷ Typical problem:

- design random walk via Markov chain, with prescribed limit distribution
- iterate sufficiently many times (**a polynomial number**), and return the endpoint

▷ Examples: walking on a grid, ball walk, hit-and-run, billiard walk, ...

▷ Convergence assessment: distance between distribution after m steps and the limit distribution

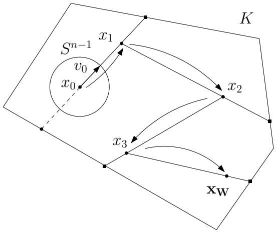
▷ Def. Let f and g be two probability measures on a state space S – or a Markov chain. The total variation distance

$$d_{TV}(f,g) = \sup_{A \subset S} \mid f(A) - g(A) \mid . \tag{10}$$

Random walk: hit-and-run

▷ Goal: sample point in K according to a prescribed density f

▷ (Random-direction) hit-and-run: random point x_W after W steps



▷ Iteratively:

- pick a random vector
- move to random point on the chord $l \cap K$, chosen from the distribution induced by f on l

▷ Comments:

- risk of being trapped near a vertex
- large W helps *forgetting* the origin x_0

▷ Thm (Berbee et al) The limit distribution induced by HR is uniform in K.

▷ Thm (Lovász) Let r and R denote the radii of the largest inscribed and circumscribed balls for K. One sample generation: $O^{\star}(n^3)$.

▷ NB: precise statement in terms of total variation distance omitted

▷`Ref: Berbee et al, Math. Prog., 1987`

▷`Ref: Lovász, Math. Prog. Ser. A, 1999`

▷`Ref: Lovász, Vempala, SIAM J Comp., 2006`

Convergence of HR to the uniform distribution π_K: details

$\triangleright$ Thm. Let K such that $rB \subset K \subset RB$. Let σ be a starting distribution and σ^m the distrib. after m steps of HR. Let $\varepsilon > 0$ and suppose that ${}^{d\sigma}/_{d\pi_K}$ is bounded by M except on a set S with $\sigma(S) \leq \varepsilon/2$. For

$$m > 10^{10} \frac{n^2 R^2}{r^2} \ln \frac{M}{\varepsilon} = O^\star(n^2),$$

one has $d_{TV}(\sigma^m, \pi_K) \leq \varepsilon$.

$\triangleright$ Thm. Under the same hypothesis, suppose that the starting distribution σ is concentrated on a point in K at distance d from the boundary. For

$$m > 10^{10} \frac{n^3 R^2}{r^2} \ln \frac{M}{d\varepsilon} = O^\star(n^3)$$

one has $d_{TV}(\sigma^m, \pi_K) \leq \varepsilon$.

$\triangleright$`Ref: Lovász, Vempala, SIAM J Comp., 2006`

Randomized algorithms for volume / DoS calculations

Motivation: volume and DoS calculations

Background: polytope volume calculations

Background: HMC

Novel HMC random walk

Experiments

Hamiltonian Monte Carlo (HMC)

▷ Hamiltonian: $H(p, q) = U(q) + K(p)$

▷ Hamiltonian dynamics:

- ODE

$$\begin{cases} \dfrac{dq_i}{dt} = \dfrac{\partial H}{\partial p_i}, \forall i = 1, \ldots, n \\ \dfrac{dp_i}{dt} = -\dfrac{\partial H}{\partial q_i}, \forall i = 1, \ldots, n \end{cases}$$

- Flow $\Phi_t q, p$ solution at time t with initial condition (q, p)

▷ Key properties preserved by the flow:

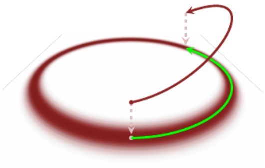
- Hamiltonian
- Lesbesgue measure of phase space
- Any measure of the form $\mu(q, p) = f(H(q, p))$
  - In particular, Boltzman measure $\pi(q, p) = \exp(-H(q, p)/k_B T)$ is preserved

▷`Ref: Neal, in Handbook of MCMC`

Using HMC to sample a distribution

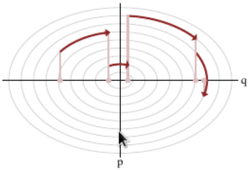
▷ Goal: sample a distribution $\pi(q)$

- Define $U(q) = -log(\pi(q))$ and $K(p) = 1/2\|p\|^2$ (Nb: unit masses)
- $H(p,q) = U(q) + K(p)$
- Invariant measure used: $\mu(q,p) = \exp(-H(q,p)) = \pi(q)\exp(-K(p))$, with $\pi(q) = \exp(-U(q))$



▷ Sampling with HMC: algorithm

- fix travel time $L > 0$
- Iterate
  - resample $p \sim \mathcal{N}(0, I_n)$
  - $(q^{(t+1)}, p^{(t+1)}) = \Phi_L(q^{(t)}, p)$

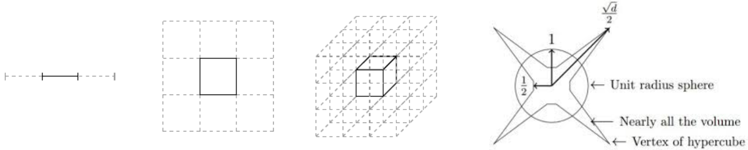


▷ Rmk: resampling p changes the energy level

▷`Ref: Betancourt, ArXiv, 2018`

Concentration in high-dimensional spaces
The hidden strength of HMC

- Example: volume of the cube



- Exploration: whole space vs typical set

Useless excursions, myopia

HMC glides around the typical set

▷Ref: `Betancourt, ArXiv,2018`

The Gaussian annulus theorem

▷ Density of the isotropic Gaussian:

$$f_G(X) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{x_1^2+x_2^2+\cdots+x_d^2}{2}}. \tag{11}$$

▷ Expectation of $\|X\|^2$:

$$\mathbb{E}[\|X\|^2] = \mathbb{E}[\sum_{i=1,\ldots,d} x_i^2] = \sum_{i=1,\ldots,d} \mathbb{E}[x_i^2] = d\mathbb{E}[x_1^2] = d. \tag{12}$$

▷ Thm. Consider an isotropic d dimensional Gaussian with $\sigma = 1$. For any $\beta \leq \sqrt{d}$, consider the annulus defined by

$$\mathcal{A} = \{X \text{ such that } \sqrt{d} - \beta \leq \|X\| \leq \sqrt{d} + \beta\}. \tag{13}$$

There exists a fixed positive constant c such that

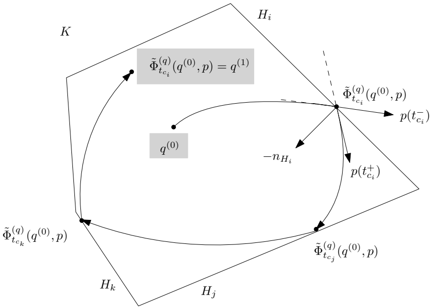
$$\mathbb{P}(\mathcal{A}^c) \leq 3e^{-c\beta^2}. \tag{14}$$

▷ Rmk: how come the mass concentrates around $\sqrt{d}$?

- Concentration thm: the mass concentrates near $\sqrt{\mathbb{E}[\|X\|^2]} = \sqrt{d}$
- The density f_G is max. at the origin; but integrating over the unit ball ... no mass since the volume of the unit ball tends to 0. (prop. seen earlier.)
- In going well beyond $\sqrt{d}$: the density f_G gets too small.

HMC in a polytope: a curved billiard walk

- Method:
  - HMC with $U(q) = \exp(-a \|q\|^2)$
  - Reflexions on boundaries of K
  - Analytical solutions for trajectories: harmonic oscillator
- Parameters:
  - Travel time L
  - Max number of reflexions Max_{reflex} should be large for the RW to forget its origin and mix



Robust implementation based on multi-precision intervals

- Input convex K: in matrix form $AX \leq B$ (A and B: rational entries)
- Geometric operations:
  - (Predicate) Membership oracle: $q \in K$
  - (Construction) Intersection point trajectory $\cap$ hyper-plane H_i
  - (Construction) Main oracle: intersection point with **nearest** hyper-plane
- Numerically:
  - Implementation with doubles fails $\Rightarrow$ multi-precision needed
  - Using `iRRAM`:
    - real numbers represented as a sequence of intervals with rational endpoints i.e. $\{(l_0, r_0), (l_1, r_1), \dots\}$
    - bounds refined on demand to satisfy operations e.g. $x < y$
    - backend for l_i, r_i: multiple precision arithmetic from GMP or MPFR
    - n-dimensional points: nested boxes

▷`Ref: Müller, Computability and Complexity in Analysis, 2001`

HMC in a polytope: conservation properties

- Theorem: invariance of π
  - one step of HMC with reflections preserves π
  - detailed balance in space of positions (but not phase space)
- Theorem: convergence to π
  - Markov chain uniformly ergodic: $d_{TV}(P^t(x,.), \pi(.)) \leq (1-\epsilon)^t$
  - Proof ingredients: convex well connected; high initial velocities $\Rightarrow$ almost straight trajectories

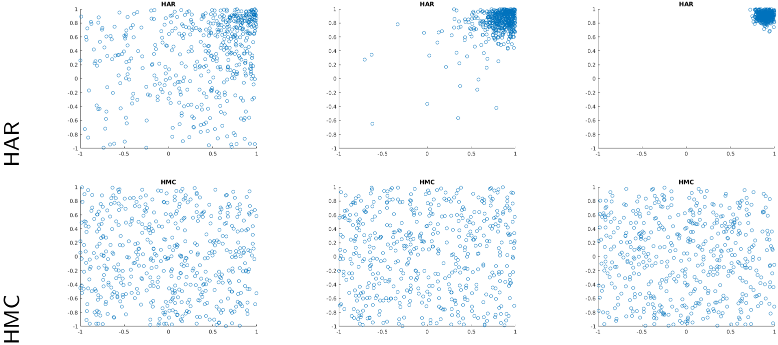
Randomized algorithms for volume / DoS calculations

Sampling a target distribution with HMC: illustration of mixing properties

▷ Setup:

- Cube $[-1,1]^n$, $n = 5, 10, 50$
- Target distribution $\pi(q)$: flat isotropic Gaussian ($\sigma_i^2 = 500$)
- Starting point $q^{(0)}$: $q_i^{(0)} = 0.9, \forall i$, return $q^{(10)}$
- Repeat 500 times
- Plots: projection i.e. first 2 coordinates

▷ HAR vs HMC



dim. $= 5$ *dim.* $= 10$ *dim.* $= 50$

Embedding HMC into the volume algorithm

Volume(K, ε): Convex body K, error parameter ε.

– $T = \textbf{Round}$(body: K, steps: $8n^3$), set $K' = T \cdot K$.
– $\{a_0, \ldots, a_m\} = \textbf{GetAnnealingSchedule}$(body: K').
– Set x to be random point from $f_0 \cap K'$, $\varepsilon' = \varepsilon/\sqrt{m}$.
– For $i = 1, \ldots, m$,
  – Set $k = 0, x_0 = x, converged = false$, $W = 4n^2 + 500$.
  – While $converged = false$,
    • $k = k + 1$.
    • $x_k =$ ~~**HitAndRun**~~ HMC (body: K, target distribution: f_{i-1}, current point: x_{k-1}).
    • Set

$$r_k = \frac{1}{k} \sum_{j=1}^{k} \frac{f_i(x_j)}{f_{i-1}(x_j)}.$$

    • Set $W_{max} = \max\{r_{k-W+1}, \ldots, r_k\}$ and $W_{min} = \min\{r_{k-W+1}, \ldots, r_k\}$.
    • If $W_{max} - W_{min} \leq \varepsilon'/2 \cdot W_{max} \to converged = true$.
  – Set $R_i = r_k, x = x_k$.
– Return $volume = |T| \cdot (\pi/a_0)^{n/2} \cdot R_1 \ldots R_m$.

Window sizes: $n^0, n^1, n\sqrt{n}, n^2$

▷ Stop condition: the window size W sets the stop criterion

▷ Stats monitored:

- Relative error $|\, V - \mathrm{Vol}(K) \,| / \mathrm{Vol}(K)$
- # calls to the oracle

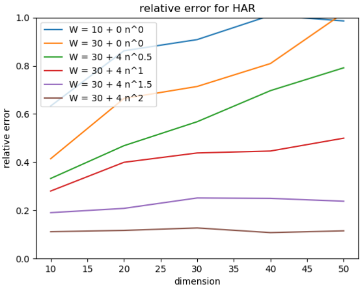
▷ Polytopes tested in $\mathbb{R}^n$, for $n = 10, \ldots, 50$:

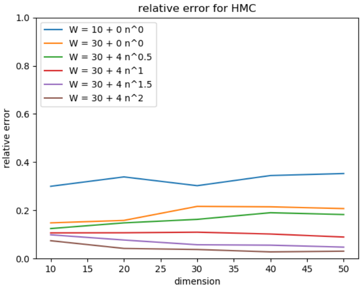
- Cube: a must
- Simplex: standard simplex,isotropic simplex
- Halfball, ellipsoid

▷Ref: `Cousins and Vempala, Math. Prog. Comp., 2016`

Volume calculation: relative error

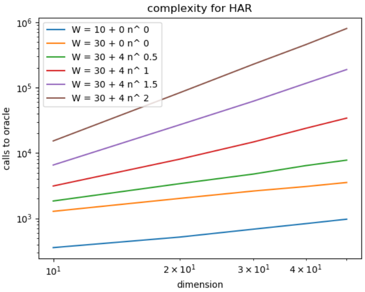
▷ **Relative errors on volume: HR (left) vs HMC (right)**

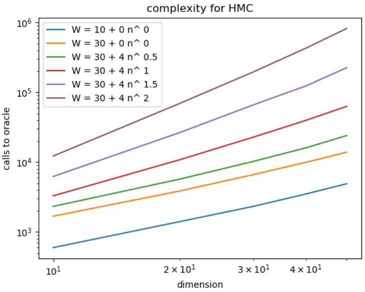




Volume calculation: number of calls to the oracle

- **Complexity i.e. number of calls to the oracle HR: (left) vs HMC (right)**





Software

- Structural Bioinformatics Library: `http://sbl.inria.fr`
  - Package on HMC: in preparation
- Other packages of interest, see `http://sbl.inria.fr/applications`
  - Landscape explorer
  - Energy landscape analysis
  - Energy landscape comparison
  - Molecular distances flexible
  - Structural motifs

Conclusion

- Hamiltonian Monte Carlo versus Hit-and-run:
  - Mixes faster, scales better
  - One volume calculation: $\sim$ minute
  - Oracle calls more expensive, but still beneficial
    - Reflexions instrumental to escape from corners
  - Multi-precision numbers mandatory – systematic failures otherwise
- Open problems, theory:
  - Role of travel time L and max. num. reflexions Max_{reflex}
  - Convergence analysis with reflections (current proof skips them ... loose bound)
  - Error bounds: scaling with dimension
- Open problems, applications:
  - Coupling polytope sampling to rejection sampling
  - Computing DoS and partition functions on a *per-basin* basis