Geometric Data analysis

Random walks, Sampling, Volume

Ioannis Emiris

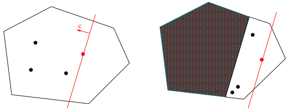
Dept Informatics & Telecoms, National Kapodistrian U. Athens
ATHENA Research & Innovation Center, Greece

May 20, 2022

Outline

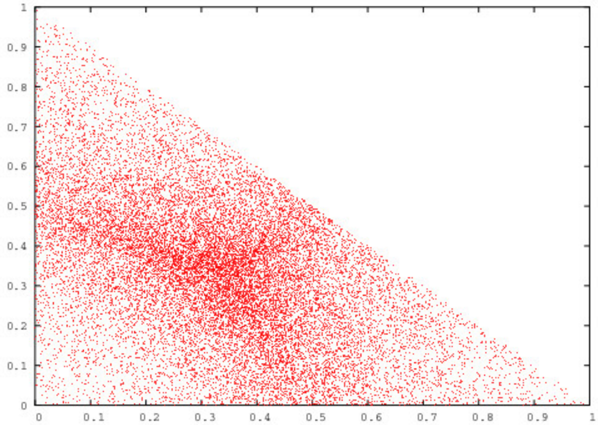
1. Sampling
2. Random walks
3. Convex Volumes
   - Poly-time approximation
   - V-polytopes
   - Nonlinear bodies
   - Oracles by ANN
4. Financial modeling

Motivation

- Monte Carlo Integration (which generalizes volume)
- Optimization 

- Sparse representation, check conjectures (# linear extensions)
- Contingency tables, underconstrained linear systems
- Systems biology [Chalkis et al.21], ...

Simplex sampling



Sample each coordinate uniformly and normalize is too naive.

Unit Simplex

Distinct uniform variables

1. Pick uniform distinct integers; then sort:
   $x_0 = 0 \leq x_1 < \cdots < x_d \leq x_{d+1} = M$.
2. Point $[y_i = (x_i - x_{i-1})/M \, : \, i = 1, \ldots, d]$ is uniform.

Complexity $= O(d \log d)$ [Smith,Tromble'04].
Fastest for $d < 80$ using Bloom filter (rather than hashing).

Exponential random variables

1. Pick uniform $x_i \in (0, 1)$; set $y_i = -\ln x_i$, $i = 1, \ldots, d+1$.
2. Let $T = \sum_{i=1}^{d+1} y_i$, then $[y_1/T, \ldots, y_d/T]$ is uniform.

Complexity $= O(d)$ [Rubinstein,Melamed'98].

Arbitrary with vertices v_i: $x \in$ unit simplex, $\sum_{i=1}^{d+1} x_i v_i$ is uniform.

Outline

1 Sampling

2 Random walks

3 Convex Volumes
- Poly-time approximation
- V-polytopes
- Nonlinear bodies
- Oracles by ANN

4 Financial modeling

Geometric Random walks

- Rejection shall not work: exponentially many points in bounding cube / simplex but outside P. Curse of dimensionality.
- Continuous (geometric) version of random walks on discrete structures (graphs).
- In arbitrary polytopes: Markov (memoryless) chains of points which "mix" to the desired distribution (typically uniform); complexity depends on (warm) start, roundedness of body.
- Each point generated with desired probability distribution after a number of steps: this number is the mixing time.
- Continuous uniform distribution: point in $A \subset P$ with probability $\text{vol}(A)/\text{vol}(P)$. Then, probability density function is $1/\text{vol}(P)$, and

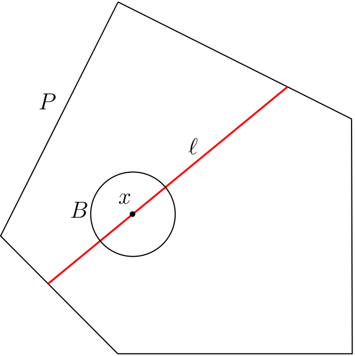
$$\int_P \frac{dv}{\text{vol}(P)} = 1.$$

Main existing walks

year	walk	mixing time	step cost
87	Coordinate HnR	?	m
06	Hit-and-Run	d^3	md
09	Dikin	md	md^2
14	Billiard	?	Rmd
16	Geodesic	$md^{3/4}$	md^2
17	Ball	$d^{2.5}$	md
17	Vaidya	$m^{1/2}d^{3/2}$	md^2
17	Riemmanian HMC	$md^{2/3}$	md^2
18	HMC w/reflections	?	md
19	sublinear Ball	$d^{2.5}$	m

dimension d, m facets, R bounds billiard reflections

Random Directions Hit-and-Run (RDHR)



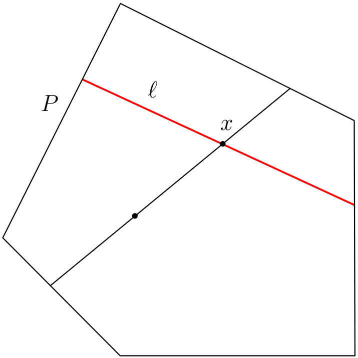
Input: point $x \in P$ and polytope $P \subset \mathbb{R}^d$

Output: a new point in P

1. line ℓ through x, uniform on $B(x, 1)$
2. new x uniform on $P \cap \ell$

Perform W steps, return x.

Random Directions Hit-and-Run (RDHR)



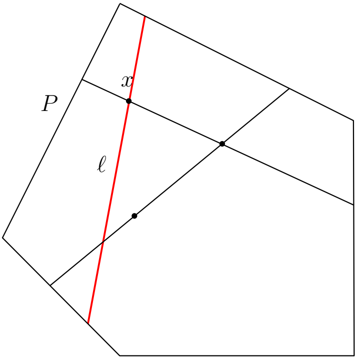
Input: point $x \in P$ and polytope $P \subset \mathbb{R}^d$

Output: a new point in P

1. line ℓ through x, uniform on $B(x, 1)$
2. new x uniform on $P \cap \ell$

Perform W steps, return x.

Random Directions Hit-and-Run (RDHR)



Input: point $x \in P$ and polytope $P \subset \mathbb{R}^d$

Output: a new point in P

1. line ℓ through x, uniform on $B(x, 1)$
2. new x uniform on $P \cap \ell$

Perform W steps, return x.

- x is uniformly distributed in P after $W \sim 10^{11} d^3$ steps [LV'06].

Sample distribution

p_u: distribution on taking one step from u: $A \subset P$ reached w/prob. $p_u(A)$

Theorem

For $u \in P$, the pdf of point $v \in P$ at next step is

$$f_u(v) = \frac{2}{\mathrm{vol}_{d-1}(S_d)} \; \frac{1}{\ell(u,v)|v-u|^{d-1}}$$

where $\ell(u,v) =$ length of chord through u, v, sphere $S_d \subset \mathbb{R}^d$.

Proof. It suffices to prove $p_u(A) = \frac{2}{\mathrm{vol}_{d-1}(S_d)} \int_A \frac{dv}{\ell(u,v)|v-u|^{d-1}}$ for infinitesimally small A: $\ell(u,v) \approx \ell$, $\forall v \in A$; $|v-u| \approx t$. Given chord L through u, $\mathrm{Prob}[v \in A] = \mathrm{vol}_1(A \cap L)/\ell$. Now $p_u(A) =$ average over all L:

$$\mathbb{E}_L\left(\frac{\mathrm{vol}_1(A \cap L)}{\ell}\right) = \frac{2}{\mathrm{vol}(S_d)\, t^{d-1}} \; \frac{\mathrm{vol}(A)}{\ell} = \frac{2}{\mathrm{vol}(S_d)} \int_A \frac{1}{\ell t^{d-1}} dv$$

because $\mathrm{vol}(S_d)t^{d-1} = \mathrm{vol}(t\text{-sphere})$ counts directions of L.

Stationary distribution

- Recall p_u is distribution obtained on taking one step from $u \in P$: $A \subset P$ is reached with probability $p_u(A)$, and $p_u(P) = 1$.
- Distribution Q on P is stationary if one step gives same distribution:

$$\int_P p_u(A) dQ(u) = Q(A), \quad \text{for any } A \subset P.$$

- Symmetry/reversibility: $f_u(v) = f_v(u)$.

If Q is uniform on P then, $Q(A) = \text{vol}(A)/\text{vol}(P)$, and:

$$\int_P p_u(A) dQ(u) = \int_P \int_A f_u(v) dQ(v) dQ(u) = \int_A \int_P f_v(u) dQ(u) dQ(v) =$$
$$= \int_A p_v(P) dQ(v) = \int_A \frac{dv}{\text{vol}(P)} = \frac{\text{vol}(A)}{\text{vol}(P)} = Q(A).$$

- Hence the uniform distribution is stationary. Is it unique?

Uniform distribution

Theorem (Smith'86)

Any symmetric (has the reversibility property) random walk with positive transition pdf converges to the uniform distribution, and it is the unique such distribution.
Examples: RDHR, Billiard walk.

Similarly for non-negative transition pdf, e.g. CDHR.

Mixing time

- Q_T : distribution after T steps.
- Mixing time: T steps s.t. $\|Q_T - Q\| \leq \epsilon$, for $\epsilon \to 0^+$.

Theorem

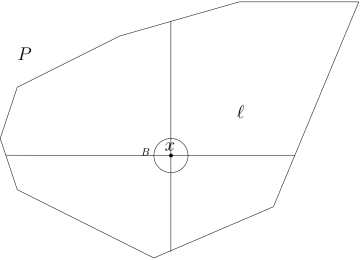
$T \approx 10^{11} d^3$ *for RDHR and uniform distribution* Q.

Proof

$T = O(1/\phi^2)$, where ϕ is the conductance of a (geometric) random walk, defined as:

$$\phi = \min_{0 \leq Q(A) \leq 1/2} \frac{\int_A p_u(P \setminus A)\, dQ(u)}{Q(A)}, \quad \text{out of some } A \subset P.$$

Coordinate Directions Hit-and-Run (CDHR)

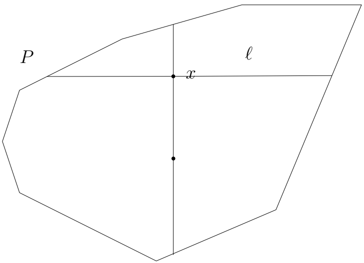


Input: point $x \in P$.

Output: a new point in P.

1. line ℓ through x, uniform on $\{e_1, \dots, e_d\}$, $e_i = (\dots, 0, 1, 0, \dots)$
2. x uniformly $\in P \cap \ell$.

Coordinate Directions Hit-and-Run (CDHR)

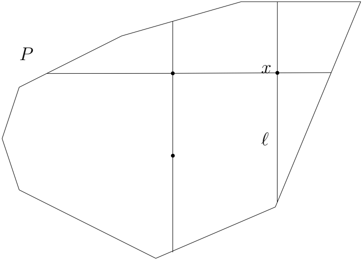


Input: point $x \in P$.

Output: a new point in P.

1. line ℓ through x, uniform on $\{e_1, \dots, e_d\}$, $e_i = (\dots, 0, 1, 0, \dots)$
2. x uniformly $\in P \cap \ell$.

Coordinate Directions Hit-and-Run (CDHR)



Input: point $x \in P$.

Output: a new point in P.

1. line ℓ through x, uniform on $\{e_1, \ldots, e_d\}$, $e_i = (\ldots, 0, 1, 0, \ldots)$
2. x uniformly $\in P \cap \ell$.

Perform W steps, return x.

"Continuous" grid walk: Converges to uniform, mixing $= O(d^{11} R^2)$ [2020].

Boundary oracle

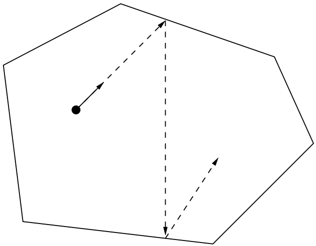
Compute intersection of line ℓ with boundary ∂P, given m hyperplanes:

- RDHR step in $O(md)$.
- CDHR $= O(m)$ per step: solve 1d (linear) problem per facet.
- Duality reduces oracle to farthest point search (max inner product) among m points: same asymptotics, practical if large m (16-dim cross-polytope: $m = 2^{16}$, 40x speedup).

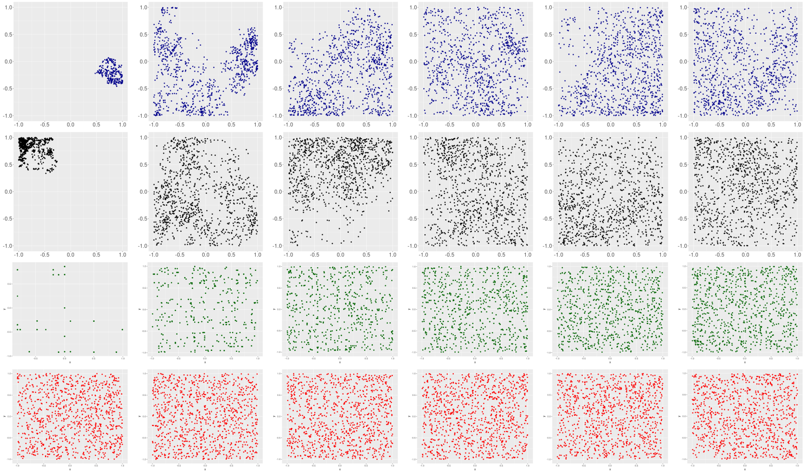
Billiard walk

BW-step (polytope P, point p_i, real τ, integer R) [Polyak'14]

1. Set length of trajectory $L = -\tau \ln \eta$, for random $\eta \sim U(0,1)$.
2. Pick uniform direction v to start the trajectory at p_i.
3. When trajectory meets ∂P with inner normal s, $\|s\| = 1$, the direction changes to $v - 2\langle v, s\rangle s$.
4. **return** the end of trajectory as p_{i+1}.
   If number of reflections exceeds R then **return** $p_{i+1} = p_i$.



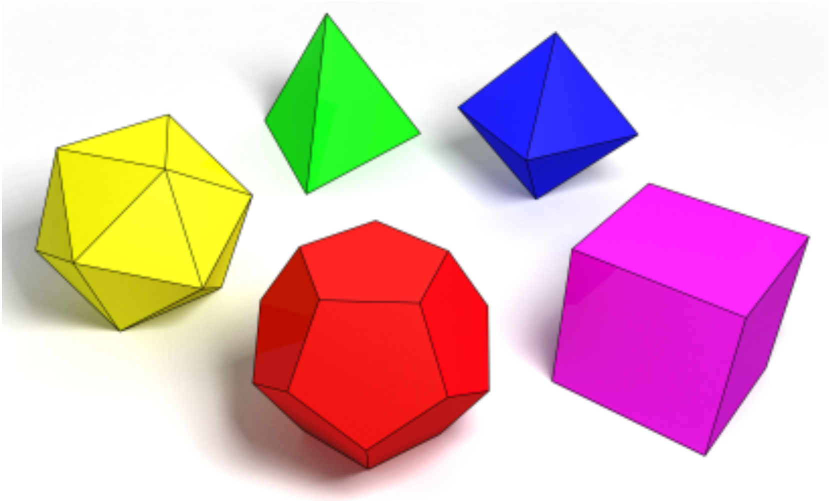
Experimental comparison



Sampling the 100d cube with Ball Walk, RDHR, CDHR, Billiard walk.
Walk length = 1,20,40,60,80,100.

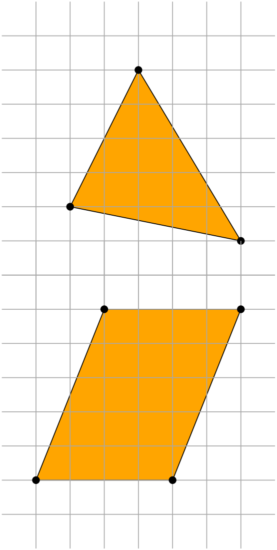
Outline

Famous polytopes



Easy cases

Some elementary polytopes have determinantal formulas.



$$\begin{vmatrix} 1 & 2 & 1 \\ 3 & 6 & 1 \\ 6 & 1 & 1 \end{vmatrix} /2! = 11$$

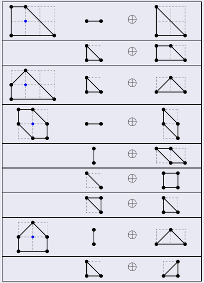
$$\begin{vmatrix} 2 & 5 \\ 4 & 0 \end{vmatrix} = 20$$

Convex polytope

- Convex polytopes are defined by
  – the set of all convex combinations of a finite set of points (V-rep):
    easy point generation, membership requires LP;
  – the intersection of a finite number of halfspaces (H-rep):
    easy membership, ray-shooting reduces to F linear systems.
- Further representations include Minkowski (vector) sums:

  – of a finite number of polytopes,
  – of segments v_i: zonotope (Z-rep)
  "generated" as follows:

$$\sum_{i=1}^{t} \lambda_i v_i, \quad 0 \leq \lambda_i \leq 1.$$



Hardness

IN: H-polytope $P := \{x \in \mathbb{R}^d \mid Ax \leq b,\ A \in \mathbb{R}^{m \times d},\ b \in \mathbb{R}^m\}$, which has m linear inequalities (maybe some redundant).

V-polytope defined by points (vertices) $v_i \in \mathbb{R}^d$:
$P := \{\lambda_1 v_1 + \cdots + \lambda_n v_n \in \mathbb{R}^d \mid \sum_i \lambda_i = 1, \lambda_i \geq 0\}$

OUT: Euclidean volume of P.

- #-P hard for vertex, halfspace representations [Dyer,Frieze'88]
- Open if both vertex & halfspace representations are available.
- APX-hard in oracle model: deterministic poly-time approximations have exponential error [Elekes'86]

Volume Approximation (H-rep)

- Curse of dimensionality:
  – Triangulation is exponential in d.
  – V(unit ball) $= \pi^{d/2}/\Gamma(1+d/2) = \Theta((2\pi e/d)^{d/2}/\sqrt{d}) = O((1/d)^d)$
  Hence rejection sampling does not scale.
- det. poly-time approximation with error $\leq d!$ [Betke,Henk'93]
- Fully Poly-time Randomized Approx. Scheme: arbitrarily small error with high probability; grid random walk, telescoping sphere sequence [D,F,Kannan'91] in $O^*(d^{23})$.
- Ball walk [K,Lovász,Simonovits'97] $O^*(d^5)$.
  $O^*(d^4 m)$ [LVempala'04] by simulated annealing, Hit-and-Run.
  If rounded $O^*(d^3 F)$ [CousinsV'14] by Gaussian cooling.
  Hamiltonian walk $O^*(d^{2/3} F)$ [LeeV'18].

Implementations

Exact: `VINCI` [Bueler et al'00], `Latte` [deLoera et al], `Qhull` [Barber et al]

- too slow in high dimensions (e.g. > 20)

Randomized for H-polytopes:

- [Lovász,Deák'12] only in ≤ 10 dimensions.
- Zonotopes via LP oracles, shake-and-bake [Fukuda et al.]
- Ours: based on Sampling [DFK'91], [Kannan,Lovász,Simonovits'97]; few hrs for few hundred dimensions.
- Matlab code by Cousins & Vempala based on [LV04], needs #facets.
- Hit-and-run in non-convex regions [Abbasi-Yadkori et al.'17]

Outline

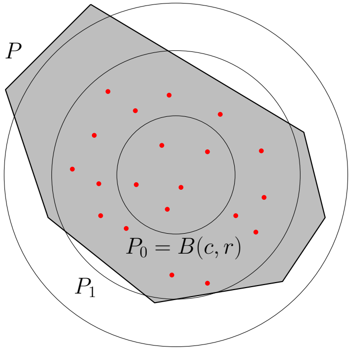
Algorithmic ingredients

✓ Sampling by Hit-and-Run

- Telescoping (multiphase) sequence of balls;

- Sandwiching input P between balls;
- Rounding input P.

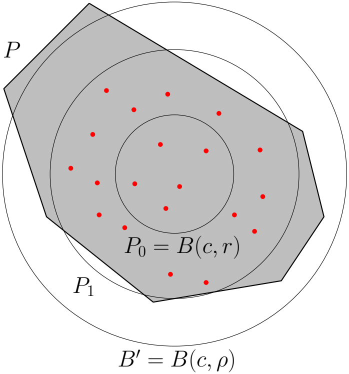
Ball sequence



- Cocentric ball $B(c, 2^{i/d})$ sequence: centered at point $c \in P$, sequence of radii $r, 2r, \ldots, \rho$, for $i = \lfloor d \log r \rfloor, \ldots, \lceil d \log \rho \rceil$ s.t. $B(c, r) \subset P \subsetneq B(c, \rho)$.
- Define convex $P_i := P \cap B(c, 2^{i/d})$.

$$vol(P) = vol(P_{d \log r}) \prod_{i=\lfloor d \log r \rfloor + 1}^{\lceil d \log \rho \rceil} \frac{vol(P_i)}{vol(P_{i-1})} \qquad \text{[DFK91]}$$

Multiphase Monte Carlo



The P_i's are sampled uniformly.

Partial inverse point generation:

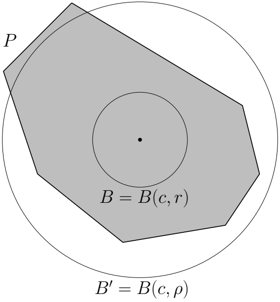
1. Let N uniform points in P_i.
2. Count (+ keep) ν in P_{i-1}.
3. Sample $N - \nu$ in P_{i-1}.

$$vol(P) = vol(P_{d \log r}) \prod_{i=\lfloor d \log r \rfloor + 1}^{\lceil d \log \rho \rceil} \frac{vol(P_i)}{vol(P_{i-1})}.$$

where each ratio is approximated by rejection sampling (step 2).

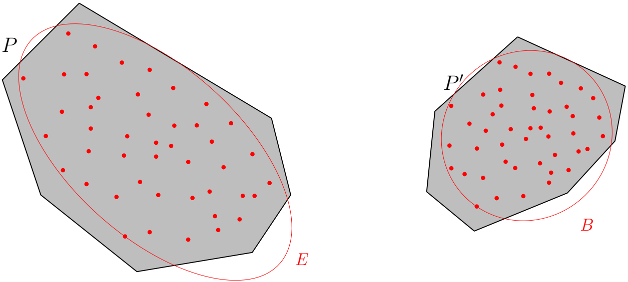
Sandwiching (Schedule)

- compute max inscribed ball $B(c, r)$ of P, by LP:
  $\max r : A_i c + r\|A_i\|_2 \leq b_i,\ i = 1, \ldots, m.$
- get uniformly distributed $p \in B(c, r)$; sample N uniform points $\in P$
- $\rho =$ max distance between c and N points: $P \subsetapprox B(c, \rho)$



Well-Rounding

1. given set S of s uniformly distributed points $\in P$
2. compute (approximate) min-volume ellipsoid E covering S:
   $S \subset E = \{x : (x-c)^T L^T L(x-c) \leq 1\}$
3. compute L mapping E to unit ball B: apply L to P



Iterate till ratio of max over min ellipsoid axes reaches threshold.
Note: Isotropic position (identity covarince) implies well-rounded.

Complexity

Theorem (Kannan,Lovász,Simonovits'97; Lovász'99)

Let a polytope P be well-rounded: $B(c, r = 1) \subseteq P \subseteq B(c, \rho)$, for $c \in P$. The algorithm computes, with probability $\geq 3/4$, an estimate of $vol(P)$ in $[(1-\epsilon)vol(P), (1+\epsilon)vol(P)]$, by

$$O^*(d^4\rho^2) = O^*(d^5)$$

oracle calls, with probability $\geq 9/10$, where $\rho = O^*(\sqrt{d})$ by isotropic sandwiching, and $\epsilon > 0$ is fixed.

Runtime

- $N = 400d \log d/\epsilon^2 = O^*(d)$ random points per P_i,
- each point computed after $W \sim 10^{11}d^3$ walk steps.

[E,Fisikopoulos'14-18]

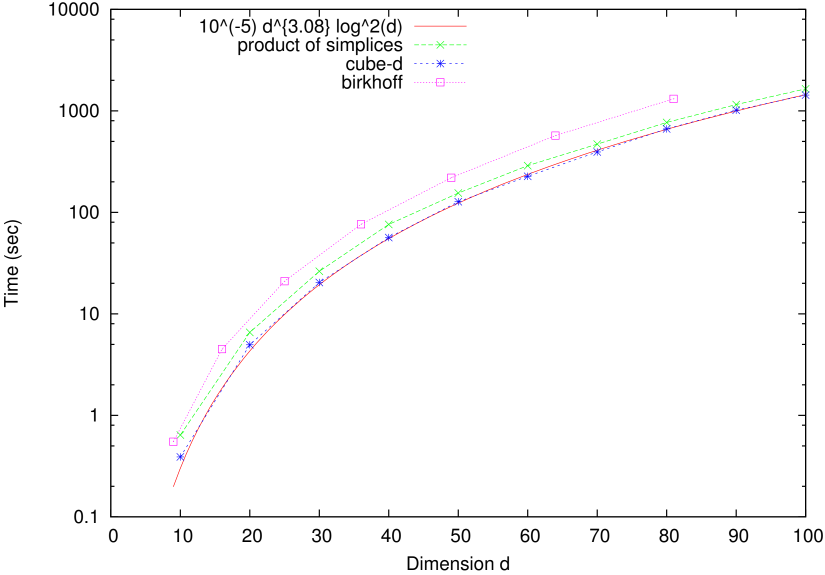
- CDHR: boundary oracle $= O(m)$.
- Set $W = \lfloor 10 + d/10 \rfloor$ walk steps, also [LovDeák]: achieves $< 1\%$ error in $d \leq 100$. Hence our algorithm takes $O^*(md^3)$ ops.
- sample partial generations of $\leq N$ points per ball $\cap P$, starting from largest; saves constant fraction per ball.
- rounding $= O^*(sd^2) = O^*(d^3)$ [Khachiyan'96]; k iterations in $O^*(k(md + d^3))$, typically $k = 1$.

- 2.5K lines C++, `github.com/GeomScale`
- CGAL for LP, min-ellipsoid; Eigen for linear algebra
- Google summer of code 2018: `R` interface [Chalkis]

Experimental results

- approximate the volume of polytopes (cubes, random, cross, Birkhoff) up to dimension 100 in < 2hrs with mean error $< 1\%$
- estimate always in $[(1-\epsilon)\text{vol}(P), (1+\epsilon)\text{vol}(P)]$, with $W = \Theta(d)$
- CDHR faster (and more accurate) than RDHR
- volume of Birkhoff polytopes $B_{11}, \ldots, B_{15}$ in few hrs; exact specialized software computed B_{10} in ~1 year [BeckPixton03]

Runtime vs. dimension



Birkhoff polytopes

$B_n = \{x \in \mathbb{R}^{n \times n} \mid x_{ij} \geq 0, \sum_i x_{ij} = 1, \sum_j x_{ij} = 1, 1 \leq i,j \leq n\}$:
perfect matchings of $K_{n,n}$, or Newton polytope of determinant.

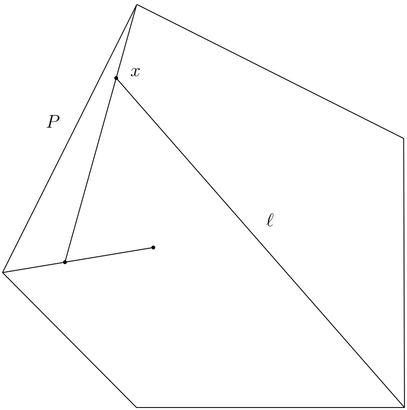
n	d	estimate	asymptotic [CanfieldMcKay09]	$\frac{\text{estimate}}{\text{asympt.}}$	exact	$\frac{\text{exact}}{\text{asympt.}}$
4	9	6.79E-002	7.61E-002	0.89194	6.21E-002	0.81593
5	16	1.41E-004	1.69E-004	0.83444	1.41E-004	0.83419
6	25	7.41E-009	8.62E-009	0.85987	7.35E-009	0.85279
7	36	5.67E-015	6.51E-015	0.87139	5.64E-015	0.86651
8	49	4.39E-023	5.03E-023	0.87295	4.42E-023	0.87786
9	64	2.62E-033	2.93E-033	0.89608	2.60E-033	0.88741
10	81	8.14E-046	9.81E-046	0.83052	8.78E-046	0.89555
11	100	1.40E-060	1.49E-060	0.93426	?	?
12	121	7.85E-078	8.38E-078	0.93705	?	?
13	144	1.33E-097	1.43E-097	0.93315	?	?
14	169	5.96E-120	6.24E-120	0.95501	?	?
15	196	5.70E-145	5.94E-145	0.95938	?	?

All volumes in few hrs; exact $V(B_{10})$ in ~1 year [BeckPixton03].

Outline

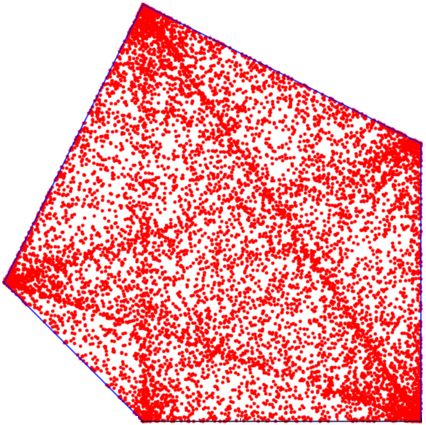
Open: V-polytopes

Given by optimization oracle



Open: V-polytopes

Given by optimization oracle

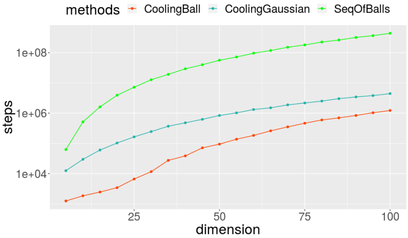


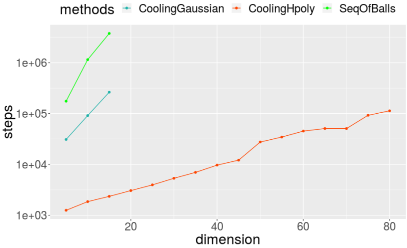
H-polytopes [E-Fisikopoulos14]

- CDHR amortized $O(1)$, $\lfloor 10 + d/10 \rfloor$ vs. $\simeq 10^{11} d^3$ random walks.
- $d \leq 100$: < 2hrs, $< 1\%$ error.

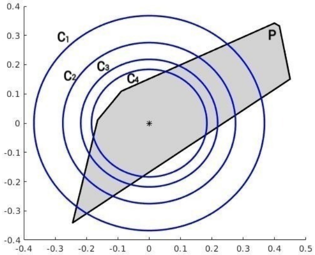
H/V-polytopes, zonotopes [Chalkis-E-Fisikopoulos'19]

- Sequence of convex bodies: good fit, easy sampling (rejection)
- Simulated annealing to construct sequence
- Statistical criterion of convergence





New Multiphase Monte Carlo



Convex $C_1 \supseteq \cdots \supseteq C_m$ intersect $P = P_0$, $P_i = C_i \cap P$, $i = 1, \ldots, m$:

$$\text{vol}(P) = \frac{\text{vol}(P_0)}{\text{vol}(P_1)} \cdots \frac{\text{vol}(P_{m-1})}{\text{vol}(P_m)} \cdot \frac{\text{vol}(P_m)}{\text{vol}(C_m)} \cdot \text{vol}(C_m),$$

is good sequence provided ratios computed fast, m small;
inner ratio may be approximated by rejection sampling.

Annealing schedule: body sequence

Employ (ideas of) simulated annealing to reduce length of sequence by adapting to the problem: non-deterministic, varying steps.

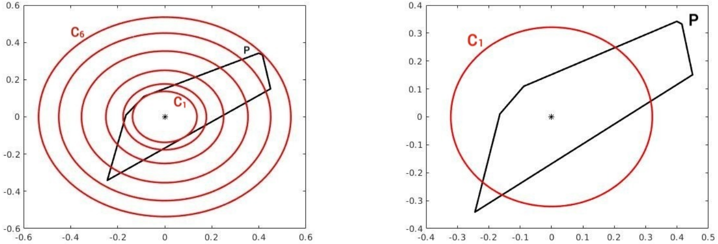
Input: Polytope P, error ϵ, cooling parameters $r, \delta > 0$ s.t. $0 < r + \delta \ll 1$.

Output: A sequence of convex bodies $C_1 \supseteq \cdots \supseteq C_m$ s.t.

$$\text{vol}(P_{i+1})/\text{vol}(P_i) \in [r, r+\delta] \text{ with high probability}$$

where $P_i = C_i \cap P,\ i = 1, \ldots, m$ and $P_0 = P$.

Annealing schedule: reduce number of phases



Six balls C_i (left), one by annealing r=0.25, δ=0.05 (right)

- Classic MMC [LKS97]: $\frac{\mathrm{vol}(C_2 \cap P)}{\mathrm{vol}(C_1 \cap P)} \cdots \frac{\mathrm{vol}(C_6 \cap P)}{\mathrm{vol}(C_5 \cap P)}\ \mathrm{vol}(C_1)$.

- Annealing schedule: $\frac{\mathrm{vol}(C_1 \cap P)}{\mathrm{vol}(C_1)} \cdot \frac{\mathrm{vol}(P)}{\mathrm{vol}(C_1 \cap P)} \cdot \mathrm{vol}(C_1)$.

Statistical tests to estimate volume ratio

Given $P_i \supseteq P_{i+1}$, $r, \delta > 0$, $0 < r + \delta \ll 1$, define null hypotheses H_0:

testLeft: $H_0 : \text{vol}(P_{i+1})/\text{vol}(P_i) \leq r + \delta$
testRight: $H_0 : \text{vol}(P_{i+1})/\text{vol}(P_i) \leq r$

1. Sample set of N points from P_i, repeat ν times.
2. $\forall$ set, binomial r.v. X counts points in P_{i+1}, success probability is unknown ratio $r_i = \text{vol}(P_{i+1})/\text{vol}(P_i)$.
3. Use $\hat{\mu}$ = mean of ν ratios.

Statistical tests

testL$(P_i, P_{i+1}, r, \delta)$:
$H_0:$ $\text{vol}(P_{i+1})/\text{vol}(P_i) \geq r + \delta$
Successful if we **reject** H_0

testR$(P_i, P_{i+1}, r, \delta)$:
$H_0:$ $\text{vol}(P_{i+1})/\text{vol}(P_i) \leq r$
Successful if we **reject** H_0

- If both successful then $r_i = \text{vol}(P_{i+1})/\text{vol}(P_i) \in [r, r + \delta]$ whp.

Statistical tests

testL$(P_i, P_{i+1}, r, \delta)$:
$H_0:$ vol$(P_{i+1})/$vol$(P_i) \geq r + \delta$
Successful if we **reject** H_0

testR$(P_i, P_{i+1}, r, \delta)$:
$H_0:$ vol$(P_{i+1})/$vol$(P_i) \leq r$
Successful if we **reject** H_0

- If both successful then $r_i = \text{vol}(P_{i+1})/\text{vol}(P_i) \in [r, r+\delta]$ whp.

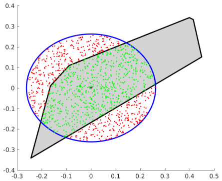


Figure: **testL**: succeeds, **testR**: fails

- Binary search a radius in $[r_{\max}, r_{\min}]$ until both tests are successful.

Statistical tests

testL$(P_i, P_{i+1}, r, \delta)$:
H_0 : $\text{vol}(P_{i+1})/\text{vol}(P_i) \geq r + \delta$
Successful if we **reject** H_0

testR$(P_i, P_{i+1}, r, \delta)$:
H_0 : $\text{vol}(P_{i+1})/\text{vol}(P_i) \leq r$
Successful if we **reject** H_0

- If both successful then $r_i = \text{vol}(P_{i+1})/\text{vol}(P_i) \in [r, r + \delta]$ whp.

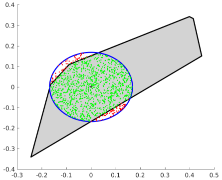


Figure: **testL**: fails, **testR**: succeeds

- Binary search a radius in $[r_{\max}, r_{\min}]$ until both tests are successful.

Statistical tests

testL$(P_i, P_{i+1}, r, \delta)$:
H_0 : $\text{vol}(P_{i+1})/\text{vol}(P_i) \geq r + \delta$
Successful if we **reject** H_0

testR$(P_i, P_{i+1}, r, \delta)$:
H_0 : $\text{vol}(P_{i+1})/\text{vol}(P_i) \leq r$
Successful if we **reject** H_0

- If both successful then $r_i = \text{vol}(P_{i+1})/\text{vol}(P_i) \in [r, r + \delta]$ whp.

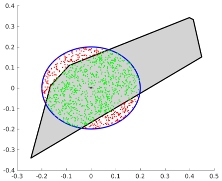


Figure: **testL**: succeeds, **testR**: succeeds

- Binary search a radius in $[r_{\max}, r_{\min}]$ until both tests are successful.

Statistical tests

Given convex bodies $P_i \supseteq P_{i+1}$, we define two statistical tests:

$\textbf{testL}(P_i, P_{i+1}, r, \delta)$:	$\textbf{testR}(P_i, P_{i+1}, r, \delta)$:
$H_0: \text{vol}(P_{i+1})/\text{vol}(P_i) \geq r + \delta$	$H_0: \text{vol}(P_{i+1})/\text{vol}(P_i) \leq r$
Successful if we **reject** H_0	**Successful** if we **reject** H_0

- If both successful then $r_i = \text{vol}(P_{i+1})/\text{vol}(P_i) \in [r, r+\delta]$ whp.

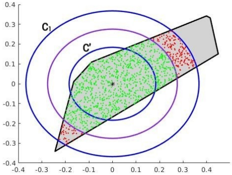


Figure: **testL**: succeeds, **testR**: succeeds

- Binary search a radius in $[r_{\max}, r_{\min}]$ until both tests are successful.

Bound #phases

- The annealing schedule terminates with constant probability.
- #phases $m = O\Big(log(\text{vol}(P)/\text{vol}(C' \cap P))\Big)$.
- If the body we use in MMC is a "good fit" to P, then $\text{vol}(C' \cap P)$ increases and m decreases.

Outline

Nonlinear bodies

For ellipsoids we generalized:

- Boundary oracle: univariate quadratic equation.
- Compute internal point, inscribed ball, enclosing ball.
- Sequence of concentric balls: Stop when all rays first hit inscribed ball

Internal point

- Transform ellipsoid to sphere H_0, transform simplex similarly.
- Find $B(p, r)$ of $\max$ radius r, satisfying constraints:

$$\text{dist}(p, H_i) \geq r \Leftrightarrow a_i^T p + b_i \geq r\|a_i\|,$$

$$\text{dist}(p, H_0) \geq r \Leftrightarrow \|p - c_0\| \leq r_0 - r.$$

  This is a Second Order Cone Program. In general, polytope intersection with $O(1)$ balls.
- Solved by SDP / interior-point method in poly-time.
- Inverse transform yields inscribed ellipsoid, maybe not max. Center is good internal point.
- Get max inscribed ball by taking distance of p to H_i's.

Outline

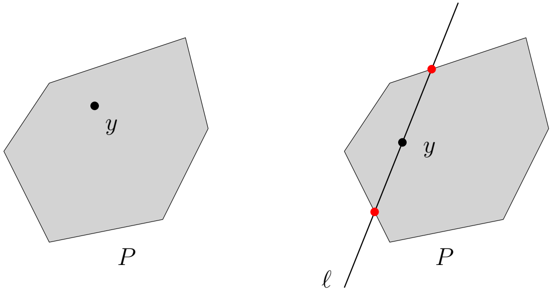
Polytope Oracles

Membership oracle

Given point $y \in \mathbb{R}^d$, return yes if $y \in P$ otherwise return no.

Boundary oracle

Given $y \in P$, ray ℓ through y, return points $\ell \cap \partial P$.



Approximation

Given is polytope $P \subset \mathbb{R}^d$ and approximation parameter $\epsilon \in (0, 1)$:

Definition (Approximate Polytope Membership)

Preprocess P into data-structure so that, given query point q, decide whether $q \in P$ or not. If $d(q, \partial P) \leq \epsilon \cdot diam(P)$ the data structure can answer either way.

Definition (Approximate Polytope Boundary)

Preprocess P into data-structure so that given query ray r emanating from $y \in P$, compute point r^*, s.t.

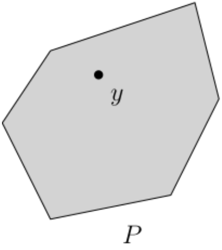
$$r^* \in r \text{ and } d(r^*, \partial P) \leq \epsilon \cdot diam(P).$$

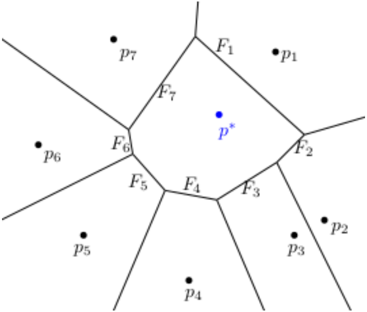
Previous approaches have complexity exponential in d.

Reduction

Exact setting [Aurenhammer'87]

Let $P \subset \mathbb{R}^d$ have n facets. $\forall p^* \in P \setminus \partial P$, compute set S of n points: membership of q reduces to finding its Nearest Neighbor in $S \cup \{p^*\}$





Approximate membership

Let $P^{-\epsilon} = \{x \in P \mid d(x, \partial P) > \epsilon \cdot diam(P)\} \neq \emptyset$.

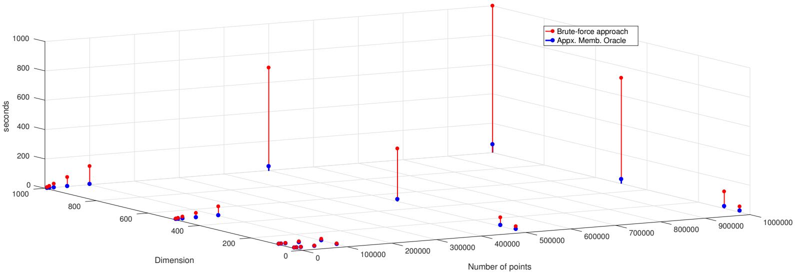
Approximate Membership reduces to ϵANN on $S \cup \{p^*\}$, $p^* \in P^{-\epsilon}$.

Theorem (Complexity)

We answer Approximate Membership queries in $O^*(dn^{\rho+o(1)})$, using $O^*(n^{1+\rho+o(1)} + dn)$ space, whp, where $\rho \leq 1/(1+4\epsilon^2) < 1$.

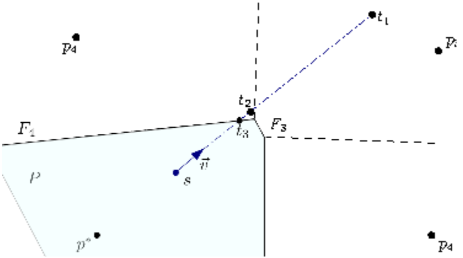
[Anagnostopoulos-E-Fisikopoulos'17]

Membership experiments



Approximate Boundary Oracle

1. Compute $t_1 \notin P$, $t_1 \in r$, where r is ray shooting query.
2. For $t_i \notin P$, compute t_{i+1} closer to apex: $p_i := \mathrm{NN}(t_i)$.
   - hyperplane H_i supports facet F_i defining p_i; $t_{i+1} := H_i \cap r$.
3. Terminate by checking (approximate) membership oracle.



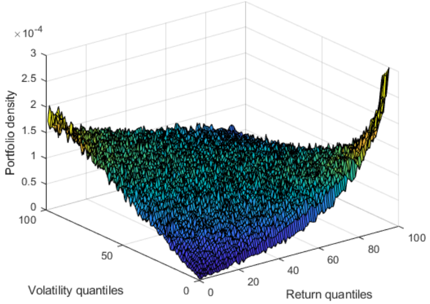
May get in local "optimum": If t_i does not decrease distance to apex, set $t_i := (t_{i-1} - r.apex) - r.unitdir \cdot \epsilon$.

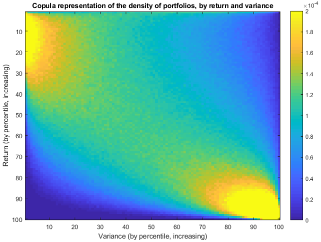
Outline

1. Sampling
2. Random walks
3. Convex Volumes
   - Poly-time approximation
   - V-polytopes
   - Nonlinear bodies
   - Oracles by ANN
4. Financial modeling

Stock markets exhibit 3 types of behavior:

- Normal: slightly positive returns, moderate volatility.
- Up-market (bubbles): high returns, low volatility.
- Crises: strongly negative returns, high volatility.





The copula is a volatility-return probability distribution.
Figure: up-market and crisis: bubble burst in Sep. 2000.

Financial modeling

- Portfolios of $d+1$ assets represented by simplex $\Delta^d \subset \mathbb{R}^{d+1}$.
- For portfolio $\omega \in \Delta^d$, returns $R \in \mathbb{R}^{d+1}$, total return $f(\omega, R) = R^T\omega$ is linear combination of returns.
- Cross-sectional score of portfolio ω^* is $vol(\Delta^*)/vol(\Delta^d)$ s.t.

$$\Delta^* = \{\omega \in \Delta^d \,:\, f(\omega, R) \leq f(\omega^*, R)\}.$$

  Score corresponds to cumulative distribution of $f(\omega, C)$.
- Volatility is quadratic form of returns.

Families of constraints

- Let R_{ij} be the return at day i of asset j. Consider compound returns over k days starting at day i: define $(d+1)$-vector v whose j-th coordinate, $j = 1, \ldots, d+1$, equals

$$v_j = (1 + R_{i,j})(1 + R_{i+1,j}) \cdots (1 + R_{i+k-1,j}) - 1.$$

  Normal vector v defines family of hyperplanes.
- Volatility requires estimation of the returns' variance – covariance matrix, yielding concentric ellipsoids.
- Copula populated by intersecting Δ^d along asset characteristics: Hyperplane families normal to two compound vectors, or to one vector and concentric ellipsoids.

Formula for single halfspace

Let $H : a^T x \leq a_0$, $a = (a_1, \dots, a_d)$, let S be the unit simplex.

1. Let $y_i = a_i - a_0$ if ≥ 0, $i = 1, \dots, K$,
   $x_i = a_i - a_0$ if < 0, $i = 1, \dots, J$, s.t. $J + K = d$.
2. Initialize $A_0 = 1, A_1 = \cdots = A_K = 0$.
3. For $j = 1, \dots, J$ do:

$$A_k \longleftarrow \frac{y_k A_k - x_j A_{k-1}}{y_k - x_j}, \quad k = 1, \dots, K.$$

For $j = J$,

$$A_K = vol(S \cap H) \, / \, vol(S).$$

Complexity $= O(d^2)$ [Varsi'73,Ali'73,Gerber'81].

Thank you!