Sampling from a Convex Body

Based on

1. Ravi Kannan's notes.
2. Jonathan Kelner's lecture notes.
3. "Techniques in Optimization and Sampling" by Yin Tat Lee and Santosh Vempala.
4. "Algorithmic Convex Geometry" by Santosh Vempala.
5. "Geometric Random Walks" by Santosh Vempala.

Computing a point in K

Given a separation oracle for a convex body K, and $R \in \mathbb{R}_+$ such that $\mathcal{B}(y, 1) \subseteq K \subseteq \mathcal{B}(0, R)$ (y is unknown), compute a point in K.

- Goal: minimize number of calls to the separation oracle.
- Any deterministic algorithm needs $n \log_2 R$ separation oracle calls in the worst case.

Computing a point in K

Given a separation oracle for a convex body K, and $R \in \mathbb{R}_+$ such that $\mathcal{B}(y, 1) \subseteq K \subseteq \mathcal{B}(0, R)$ (y is unknown), compute a point in K.

- Goal: minimize number of calls to the separation oracle.
- Any deterministic algorithm needs $n \log_2 R$ separation oracle calls in the worst case.

Algorithm:

1. Initialize $P_1 = [-R, R]^n$, $z = 0$.
2. for $i = 1$ to N (to be fixed later.)
   2.1 Compute $z \in P_i$ (to be specified later).
   2.2 Call separation oracle on z. If $z \in K$, output z.
   2.3 Let $a^T x \leq b$ be the hyperplane returned by the separation oracle. Set $P_{i+1} := P_i \cap \left\{x \in \mathbb{R}^n : a^T x \leq a^T z\right\}$.
3. Output that K is empty.

Centroid

- If at the end of the algorithm we have $\mathrm{vol}(P_N) < \mathrm{vol}(\mathcal{B}(0,1))$, then K is empty.

Centroid

- If at the end of the algorithm we have $\text{vol}(P_N) < \text{vol}(\mathcal{B}(0,1))$, then K is empty.
- We would like to compute new z such that volume of P_i is guaranteed to shrink by at least a constant factor. Then $N = \Theta(n \log R)$ will suffice.

Centroid

- If at the end of the algorithm we have $\text{vol}(P_N) < \text{vol}(\mathcal{B}(0,1))$, then K is empty.
- We would like to compute new z such that volume of P_i is guaranteed to shrink by at least a constant factor. Then $N = \Theta(n \log R)$ will suffice.
- Choosing z to be the centroid guarantees this (Grunbaum's Theorem), but computing centroid is $\#P$-hard in general.

Centroid

- If at the end of the algorithm we have $\mathrm{vol}\,(P_N) < \mathrm{vol}\,(\mathcal{B}\,(0,1))$, then K is empty.
- We would like to compute new z such that volume of P_i is guaranteed to shrink by at least a constant factor. Then $N = \Theta(n \log R)$ will suffice.
- Choosing z to be the centroid guarantees this (Grunbaum's Theorem), but computing centroid is $\#P$-hard in general.

Sample m independent and uniform random points from P_i, denote them by $y_1, \ldots, y_m$. Set $z = (\sum_{i \in [m]} y_i)/m$.

Centroid

- If at the end of the algorithm we have $\mathrm{vol}\left(P_N\right) < \mathrm{vol}\left(\mathcal{B}\left(0,1\right)\right)$, then K is empty.
- We would like to compute new z such that volume of P_i is guaranteed to shrink by at least a constant factor. Then $N = \Theta(n \log R)$ will suffice.
- Choosing z to be the centroid guarantees this (Grunbaum's Theorem), but computing centroid is $\#P$-hard in general.

Sample m independent and uniform random points from P_i, denote them by $y_1, \ldots, y_m$. Set $z = (\sum_{i \in [m]} y_i)/m$.

Theorem

$$\mathbb{E}\left[\mathrm{vol}\left(P_{i+1}\right)\right] \leq \left(1 - \frac{1}{e} + \sqrt{\frac{n}{m}}\right) \mathrm{vol}\left(P_i\right).$$

With $m = \Theta(n)$, $N = \Theta(n \log R)$ will suffice.

Conductance

For a set of vertices $S \subset V$,

$$\phi(S) := \frac{\sum_{i \in S, j \in V \setminus S} \pi_i P_{ij}}{\min\{\pi(S), \pi(V \setminus S)\}}$$

and $\phi := \min_{S:0<\pi(S)<1} \phi(S)$.

Conductance

For a set of vertices $S \subset V$,

$$\phi(S) := \frac{\sum_{i \in S, j \in V \setminus S} \pi_i P_{ij}}{\min\{\pi(S), \pi(V \setminus S)\}}$$

and $\phi := \min_{S: 0 < \pi(S) < 1} \phi(S)$.
Markov Chain (informal definition)

- $(K, \mathcal{A})$ where K is the *state space* and $\mathcal{A}$ is a set of subsets of K that is closed under complements and countable unions.

Conductance

For a set of vertices $S \subset V$,

$$\phi(S) := \frac{\sum_{i \in S, j \in V \setminus S} \pi_i P_{ij}}{\min \{\pi(S), \pi(V \setminus S)\}}$$

and $\phi := \min_{S: 0 < \pi(S) < 1} \phi(S)$.
Markov Chain (informal definition)

- $(K, \mathcal{A})$ where K is the *state space* and $\mathcal{A}$ is a set of subsets of K that is closed under complements and countable unions.
- For each $u \in K$ and $A \in \mathcal{A}$, $P_u(A)$ is the probability of being in A after taking one step from u.

Conductance

For a set of vertices $S \subset V$,

$$\phi(S) := \frac{\sum_{i \in S, j \in V \setminus S} \pi_i P_{ij}}{\min\{\pi(S), \pi(V \setminus S)\}}$$

and $\phi := \min_{S : 0 < \pi(S) < 1} \phi(S)$.
Markov Chain (informal definition)

- $(K, \mathcal{A})$ where K is the *state space* and $\mathcal{A}$ is a set of subsets of K that is closed under complements and countable unions.
- For each $u \in K$ and $A \in \mathcal{A}$, $P_u(A)$ is the probability of being in A after taking one step from u.
- Given a starting distribution Q_0, w_0 is sampled from Q_0 and w_i is sampled from $P_{w_{i-1}}$.

- A distribution Q on $(K, \mathcal{A})$ is called stationary if one step from it gives the same distribution, i.e., for any $A \in \mathcal{A}$,

$$\int_K P_u(A) dQ(u) = Q(A).$$

- A distribution Q on $(K, \mathcal{A})$ is called stationary if one step from it gives the same distribution, i.e., for any $A \in \mathcal{A}$,

$$\int_K P_u(A) dQ(u) = Q(A).$$

- The conductance of a subset A is defined as

$$\phi(A) := \frac{\int_A P_u(K \setminus A) dQ(u)}{\min \{Q(A), Q(K \setminus A)\}}.$$

and the conductance of the Markov chain is $\phi := \min_A \phi(A)$.

- A distribution Q on $(K, \mathcal{A})$ is called stationary if one step from it gives the same distribution, i.e., for any $A \in \mathcal{A}$,

$$\int_K P_u(A) dQ(u) = Q(A).$$

- The conductance of a subset A is defined as

$$\phi(A) := \frac{\int_A P_u(K \setminus A) dQ(u)}{\min\{Q(A), Q(K \setminus A)\}}.$$

  and the conductance of the Markov chain is $\phi := \min_A \phi(A)$.
- A distribution Q is *atom-free* if there is no $x \in K$ with $Q(x) > 0$.
- P is said to be *M-warm* with respect to Q if

$$M = \sup_{A \in \mathcal{A}} \frac{P(A)}{Q(A)}.$$

Mixing Time

Let Q_t be the distribution of the random walk at time t.

Theorem

Let $M = \sup_A Q_0(A)/Q(A)$. *Then*

$$d_{\mathrm{TV}}\left(Q_t, Q\right) \leq \sqrt{M}\left(1 - \frac{\phi^2}{2}\right)^t.$$

Mixing Time

Let Q_t be the distribution of the random walk at time t.

Theorem

Let $M = \sup_A Q_0(A)/Q(A)$. Then

$$d_{\mathrm{TV}}\left(Q_t, Q\right) \leq \sqrt{M}\left(1 - \frac{\phi^2}{2}\right)^t.$$

Therefore, for

$$t = O\left(\frac{1}{\phi^2}\log\left(\frac{M}{\epsilon}\right)\right) \qquad \text{we have} \qquad d_{\mathrm{TV}}\left(Q_t, Q\right) \leq \epsilon.$$

Mixing Time

Let Q_t be the distribution of the random walk at time t.

Theorem

Let $M = \sup_A Q_0(A)/Q(A)$. Then

$$d_{\mathrm{TV}}(Q_t, Q) \leq \sqrt{M}\left(1 - \frac{\phi^2}{2}\right)^t.$$

Therefore, for

$$t = O\left(\frac{1}{\phi^2}\log\left(\frac{M}{\epsilon}\right)\right) \qquad \text{we have} \qquad d_{\mathrm{TV}}(Q_t, Q) \leq \epsilon.$$

To bound mixing time, sufficient to prove lower bound on ϕ.

Bounding Conductance

Fix a set $S \subset K$ and consider the Grid-walk or ball-walk where the walk is currently in S.

Bounding Conductance

Fix a set $S \subset K$ and consider the Grid-walk or ball-walk where the walk is currently in S. In one step, walk has a "large" probability of going to $K \setminus S$ if walk is currently "close" to the boundary of S and $K \setminus S$.

Bounding Conductance

Fix a set $S \subset K$ and consider the Grid-walk or ball-walk where the walk is currently in S. In one step, walk has a "large" probability of going to $K \setminus S$ if walk is currently "close" to the boundary of S and $K \setminus S$.

For $S_1, S_2 \subseteq K$,

$$d(S_1, S_2) := \inf \{\|u - v\| : u \in S_1,\ v \in S_2\}.$$

Bounding Conductance

Fix a set $S \subset K$ and consider the Grid-walk or ball-walk where the walk is currently in S. In one step, walk has a "large" probability of going to $K \setminus S$ if walk is currently "close" to the boundary of S and $K \setminus S$.

For $S_1, S_2 \subseteq K$,

$$d(S_1, S_2) := \inf \{\|u - v\| : u \in S_1,\ v \in S_2\}.$$

Theorem

Let S_1, S_2, S_3 be a partition into measurable sets of a convex body K of diameter D. Then,

$$\mathrm{vol}(S_3) \geq \frac{2d(S_1, S_2)}{D} \min \{\mathrm{vol}(S_1), \mathrm{vol}(S_2)\}.$$

Bounding Conductance

Fix a set $S \subset K$ and consider the Grid-walk or ball-walk where the walk is currently in S. In one step, walk has a "large" probability of going to $K \setminus S$ if walk is currently "close" to the boundary of S and $K \setminus S$.

For $S_1, S_2 \subseteq K$,

$$d(S_1, S_2) := \inf \{\|u - v\| : u \in S_1,\ v \in S_2\}.$$

Theorem

Let S_1, S_2, S_3 be a partition into measurable sets of a convex body K of diameter D. Then,

$$\mathrm{vol}(S_3) \geq \frac{2d(S_1, S_2)}{D} \min\{\mathrm{vol}(S_1), \mathrm{vol}(S_2)\}.$$

- Not true for non-convex bodies (e.g. dumbell graph).

Bounding Conductance

Fix a set $S \subset K$ and consider the Grid-walk or ball-walk where the walk is currently in S. In one step, walk has a "large" probability of going to $K \setminus S$ if walk is currently "close" to the boundary of S and $K \setminus S$.

For $S_1, S_2 \subseteq K$,

$$d(S_1, S_2) := \inf \left\{ \|u - v\| : u \in S_1, \ v \in S_2 \right\}.$$

Theorem

Let S_1, S_2, S_3 be a partition into measurable sets of a convex body K of diameter D. Then,

$$\mathrm{vol}\left(S_3\right) \geq \frac{2d(S_1, S_2)}{D} \min \left\{ \mathrm{vol}\left(S_1\right), \mathrm{vol}\left(S_2\right) \right\}.$$

- Not true for non-convex bodies (e.g. dumbell graph).
- Dependance on D unavoidable (e.g. cylinder verify this).

More general random walks

Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an integrable function. It defines a measure π_f on measurable subsets of $\mathbb{R}^n$.

$$\pi_f(A) = \frac{\int_A f(x)dx}{\int_{\mathbb{R}^n} f(x)dx}.$$

More general random walks

Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an integrable function. It defines a measure π_f on measurable subsets of $\mathbb{R}^n$.

$$\pi_f(A) = \frac{\int_A f(x)dx}{\int_{\mathbb{R}^n} f(x)dx}.$$

Ball walk with Metropolis filter (δ, f)

1. Pick a uniform random point in the ball of radius δ centered at the current point x.
2. Move to y with probability $\min\{1, f(y)/f(x)\}$, stay at x with remaining probability.

More general random walks

Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an integrable function. It defines a measure π_f on measurable subsets of $\mathbb{R}^n$.

$$\pi_f(A) = \frac{\int_A f(x)dx}{\int_{\mathbb{R}^n} f(x)dx}.$$

Ball walk with Metropolis filter (δ, f)

1. Pick a uniform random point in the ball of radius δ centered at the current point x.
2. Move to y with probability $\min\{1, f(y)/f(x)\}$, stay at x with remaining probability.

- Ball walk has π_f as stationary distribution.

More general random walks

Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an integrable function. It defines a measure π_f on measurable subsets of $\mathbb{R}^n$.

$$\pi_f(A) = \frac{\int_A f(x)dx}{\int_{\mathbb{R}^n} f(x)dx}.$$

Ball walk with Metropolis filter (δ, f)

1. Pick a uniform random point in the ball of radius δ centered at the current point x.
2. Move to y with probability $\min\{1, f(y)/f(x)\}$, stay at x with remaining probability.

- Ball walk has π_f as stationary distribution.
- When does this have "large" conductance?

Logconcave functions

A function $f : \mathbb{R}^n \to \mathbb{R}_+$ is said to be *logconcave* if for any $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \geq f(x)^{\lambda} f(y)^{1-\lambda}.$$

Logcconcave functions

A function $f : \mathbb{R}^n \to \mathbb{R}_+$ is said to be *logconcave* if for any $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}.$$

Examples (verify)

1. For a convex body K, let $f(x) = \begin{cases} 1 & \text{if } x \in K \\ 0 & \text{otherwise} \end{cases}$.
2. $f(x) = e^{-\|x\|^2}$.
3. Product of two logconcave functions.

Centroid

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be a logconcave density function, and let H be any halfspace containing its centroid. Then

$$\int_H f(x)dx \geq \frac{1}{e}.$$

Centroid

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be a logconcave density function, and let H be any halfspace containing its centroid. Then

$$\int_H f(x)dx \geq \frac{1}{e}.$$

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be a logconcave density function, z be the average of m independent random points from π_f, and let H be any halfspace containing z. Then

$$\mathbb{E}\left[\pi_f(H)\right] \geq \frac{1}{e} - \sqrt{\frac{n}{m}}.$$

Conductance

Theorem

Let f be a logconcave density on $\mathbb{R}^n$ whose support has diameter D. Then for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3

$$\pi_f(S_3) \geq \frac{2d(S_1, S_2)}{D} \min\left\{\pi_f(S_1), \pi_f(S_2)\right\}.$$

Conductance

Theorem

Let f be a logconcave density on $\mathbb{R}^n$ whose support has diameter D. Then for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3

$$\pi_f(S_3) \geq \frac{2d(S_1, S_2)}{D} \min \left\{ \pi_f(S_1), \pi_f(S_2) \right\}.$$

Let z_f be the centroid of f and let $M(f) := \mathbb{E}_{x \sim \pi_f} \|x - z_f\|$.

Conductance

Theorem

Let f be a logconcave density on $\mathbb{R}^n$ whose support has diameter D. Then for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3

$$\pi_f(S_3) \geq \frac{2d(S_1, S_2)}{D} \min \{\pi_f(S_1), \pi_f(S_2)\}.$$

Let z_f be the centroid of f and let $M(f) := \mathbb{E}_{x \sim \pi_f} \|x - z_f\|$.

Theorem

Let f be a logconcave density on $\mathbb{R}^n$. Then for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3

$$\pi_f(S_3) \geq \frac{\ln 2}{M(f)} d(S_1, S_2) \min \{\pi_f(S_1), \pi_f(S_2)\}.$$

Isotropic Densities

A density function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *isotropic* if

$$\mathbb{E}_{X \sim \pi_f}[X] = 0 \qquad \text{and} \qquad \mathbb{E}_{X \sim \pi_f}\left[XX^T\right] = I.$$

Isotropic Densities

A density function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *isotropic* if

$$\mathbb{E}_{X \sim \pi_f}[X] = 0 \qquad \text{and} \qquad \mathbb{E}_{X \sim \pi_f}\left[XX^T\right] = I.$$

For an isotropic density f,

$$M(f) = \mathbb{E}_{X \sim \pi_f} \|X\| \le \sqrt{\mathbb{E}_{X \sim \pi_f} \|X\|^2} = \sqrt{n}.$$

Isotropic Densities

A density function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *isotropic* if

$$\mathbb{E}_{X\sim\pi_f}[X] = 0 \qquad \text{and} \qquad \mathbb{E}_{X\sim\pi_f}\left[XX^T\right] = I.$$

For an isotropic density f,

$$M(f) = \mathbb{E}_{X\sim\pi_f}\|X\| \le \sqrt{\mathbb{E}_{X\sim\pi_f}\|X\|^2} = \sqrt{n}.$$

Theorem

For an isotropic logconcave density f,

$$\mathbb{P}\left[\|X\| \ge t\sqrt{n}\right] \le e^{-t}.$$

A density function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *C-isotropic* if

$$\mathbb{E}_{X \sim \pi_f}[X] = 0 \qquad \text{and} \qquad \frac{1}{C} I \preceq \mathbb{E}_{X \sim \pi_f}\left[XX^T\right] \preceq CI.$$

A density function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *C-isotropic* if

$$\mathbb{E}_{X\sim\pi_f}[X] = 0 \qquad \text{and} \qquad \frac{1}{C} I \preceq \mathbb{E}_{X\sim\pi_f}\left[XX^T\right] \preceq CI.$$

- Let $\Sigma = \mathbb{E}_{X\sim\pi_f}\left[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^T\right]$. Then $Y := \Sigma^{-1/2}(X-\mathbb{E}[X])$ is isotropic.

$$\begin{aligned}\mathbb{E}\left[YY^T\right] &= \mathbb{E}\left[\Sigma^{-1/2}(X-\mathbb{E}[X])\left(\Sigma^{-1/2}(X-\mathbb{E}[X])\right)^T\right]\\ &= \Sigma^{-1/2}\mathbb{E}\left[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^T\right]\Sigma^{-1/2} = I.\end{aligned}$$

A density function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *C-isotropic* if

$$\mathbb{E}_{X\sim\pi_f}[X] = 0 \qquad \text{and} \qquad \frac{1}{C} I \preceq \mathbb{E}_{X\sim\pi_f}\left[XX^T\right] \preceq CI.$$

- Let $\Sigma = \mathbb{E}_{X\sim\pi_f}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\right]$. Then $Y := \Sigma^{-1/2}(X - \mathbb{E}[X])$ is isotropic.

$$\begin{aligned}\mathbb{E}\left[YY^T\right] &= \mathbb{E}\left[\Sigma^{-1/2}(X - \mathbb{E}[X])\left(\Sigma^{-1/2}(X - \mathbb{E}[X])\right)^T\right] \\ &= \Sigma^{-1/2}\mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\right]\Sigma^{-1/2} = I.\end{aligned}$$

- For any linear transformation A, $\mathrm{vol}(AK) = \det(A) \cdot \mathrm{vol}(K)$.

A density function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *C-isotropic* if

$$\mathbb{E}_{X\sim\pi_f}[X] = 0 \qquad \text{and} \qquad \frac{1}{C} I \preceq \mathbb{E}_{X\sim\pi_f}\left[XX^T\right] \preceq CI.$$

- Let $\Sigma = \mathbb{E}_{X\sim\pi_f}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\right]$. Then $Y := \Sigma^{-1/2}(X - \mathbb{E}[X])$ is isotropic.

$$\begin{aligned}\mathbb{E}\left[YY^T\right] &= \mathbb{E}\left[\Sigma^{-1/2}(X - \mathbb{E}[X])\left(\Sigma^{-1/2}(X - \mathbb{E}[X])\right)^T\right] \\ &= \Sigma^{-1/2}\mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\right]\Sigma^{-1/2} = I.\end{aligned}$$

- For any linear transformation A, $\mathrm{vol}(AK) = \det(A) \cdot \mathrm{vol}(K)$.
- Σ can be estimated by sampling?