Cálculo del Volumen en Dimensión Alta

Santosh Vempala, Georgia Tech

Un problema muy antiguo

Given set K in n-dimensional space, estimate its volume.

E.g.,

- Piramides, barriles de vino, …

- Polytopes
- Intersection of a polytope with ellipsoid(s)
- Section of the semidefinite cone

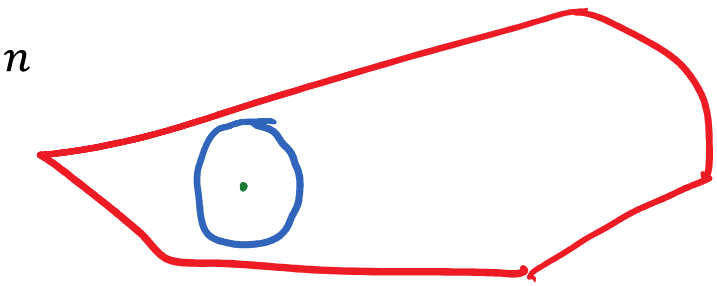
Fundamental, *computational*: how much?

Modelo de computación

Well-guaranteed Membership oracle:

Compact set K is given by

- a membership oracle: answers YES/NO to “$x \in K?$"
- a point $x_0 \in K$
- Numbers r, R s.t. $x_0 + rB^n \subseteq K \subseteq RB^n$



Well-guaranteed Function oracle

- An oracle that returns $f(x)$ for any $x \in R^n$
- A point x_0 with $f(x_0) \geq \beta$
- Numbers r, R s.t.

$$x_0 + rB^n \subset L_f\left(\frac{1}{8}\right) \text{ and } R^2 = E_f(\left|\left|X - \bar{X}\right|\right|^2)$$

Problemas en Dimensión Alta

Input:

- A set of points S in n-dimensional space R^n
  or a distribution in R^n
- A function f that maps points to real values (could be the indicator of a set)

Problemas en Dimensión Alta

- What is the complexity of computational problems as the dimension grows?

- Dimension = number of variables

- Typically, size of input is a function of the dimension.

Muestreo al azar y Integración

- Numerous applications in diverse areas: statistics, networking, biology, computer vision, privacy, operations research etc.

- This course: mathematical and algorithmic foundations of sampling and integration.

Problem 1: Optimización

Input: function f: $R^n \to R$ specified by an oracle,

point x, error parameter ε .

Output: point y such that

$$f(y) \geq \max f - \epsilon$$

Volume Computation is the special case when f is the indicator of a compact set.

Problem 2: Integración

Input: function f: $R^n \rightarrow R$ specified by an oracle,

point x, error parameter ε .

Output: number A such that:

$$(1-\epsilon)\int f \leq A \leq (1+\epsilon)\int f$$

Problem 3: Muestreo al azar

Input: function f: $R^n \rightarrow R$ specified by an oracle,

point x, error parameter ε.

Output: A point y from a distribution within distance ε of distribution with density proportional to f.

Problem 4: Redondear (Rounding)

Input: function f: $R^n \rightarrow R$ specified by an oracle,

point x, error parameter ε.

Output: An affine transformation that approximately "rounds" f, between two balls.

Problem 5: Aprendizaje

Input: i.i.d. points with labels from an unknown distribution, error parameter ε.

Output: A rule to correctly label $1 - \varepsilon$ of the input distribution.

Problemas en Dimensión Alta

- Integration (volume)
- Optimization
- Learning
- Rounding
- Sampling

All intractable in general, even to approximate.

Algoritmos Clasicos

P1. Optimization. Find minimum of f over the set S.

Ellipsoid algorithm [Yudin-Nemirovski; Shor;Khachiyan;GLS]

S is a convex set and f is a convex function.

P2. Integration. Find the integral of f.

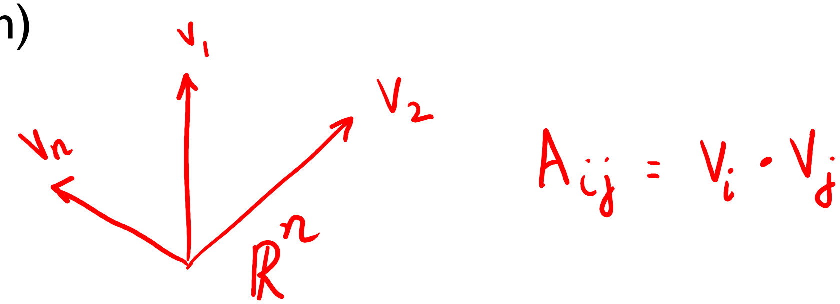
Dyer-Frieze-Kannan algorithm

f is the indicator function of a convex set.

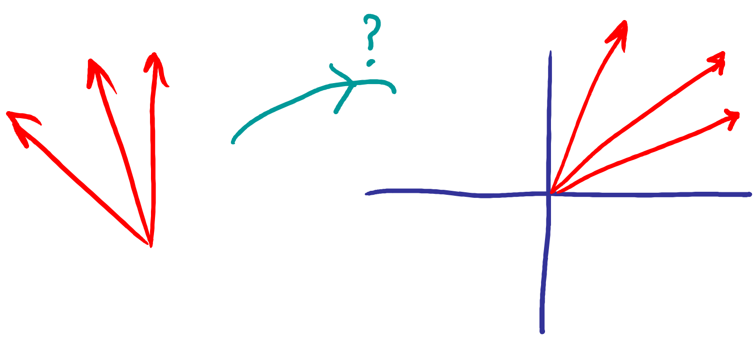
La frontera de complejidad

1. Are the entries of a given matrix inner products of a set of vectors?

   $A = BB^T$?   (semidefinite program)



2. Are they inner products of a set of nonnegative vectors?



Is $A = BB^T$, $B \geq 0$?   (completely positive)

Estructura

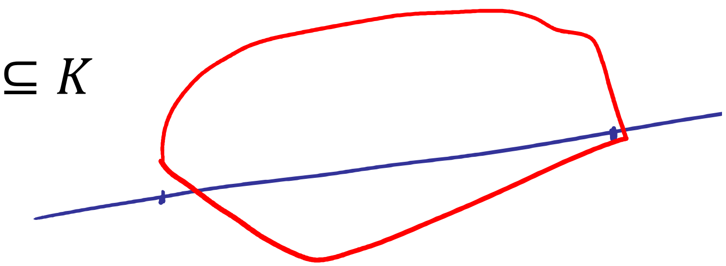
Q. What structure makes high-dimensional problems computationally tractable? (i.e., solvable with polynomial complexity)

- **Convexity** and its extensions appear to be the frontier of polynomial-time solvability.

Convexidad

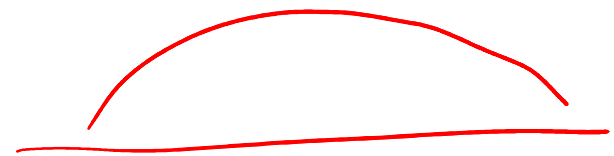
(Indicator functions of) Convex sets:

$\forall x, y \in R^n, \lambda \in [0,1], x, y \in K \Rightarrow \lambda x + (1-\lambda)y \subseteq K$



Concave functions:

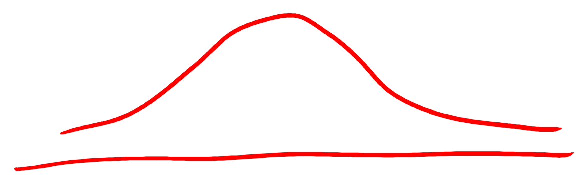
$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$



Logconcave functions:

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

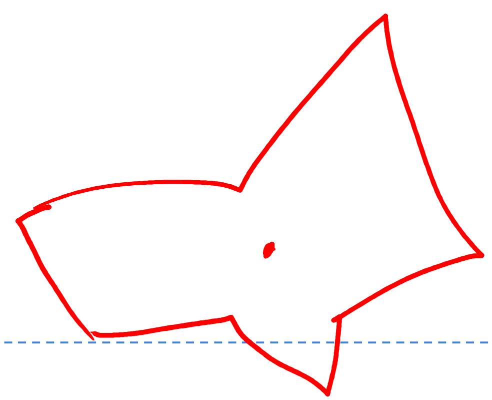
Quasiconcave functions:



$f(\lambda x + (1-\lambda)y) \geq \min \ f(x), f(y)$

Star-shaped sets:

$\exists x \in S \; s.t. \forall y \in S, \lambda x + (1-\lambda)y \in S$



Como especificar un conjunto convexo?

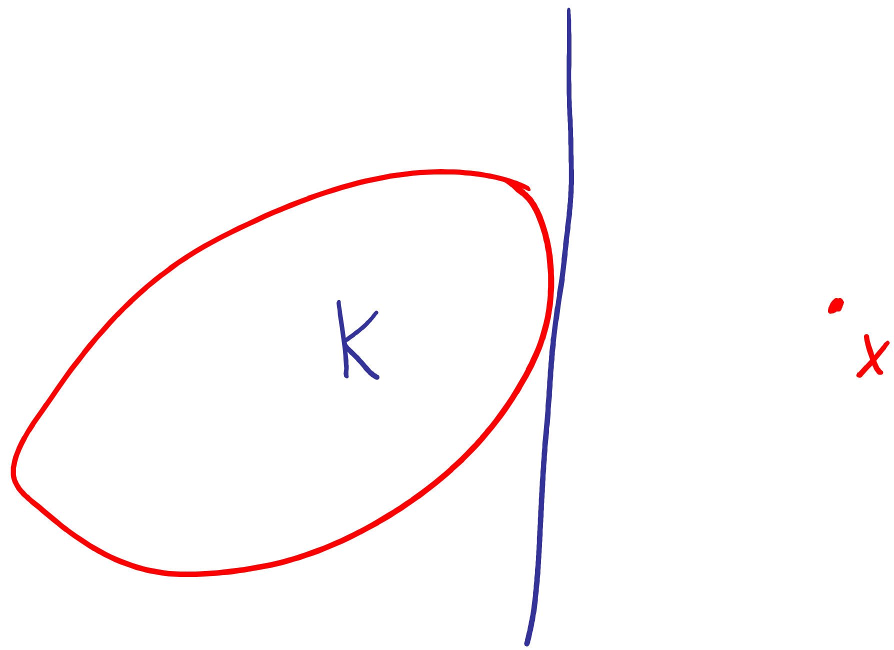
- Explicit list of constraints, e.g., a linear program:

  $$Ax \leq \mathrm{b}$$

- What about the set of positive semidefinite matrices?

- Or the set of vectors on the edges of a graph that have weight at least one on every cut?

Estructura I: Oraculo de la separacion



Either $x \in K$ or there is a halfspace containing K and not x.

Conjuntos convexos tienen oraculos de separacion

- If x is not in K, let y be the point in K that is closest to x.

- y is unique: If y_1, y_2 are both closest, then $(y_1 + y_2)/2$ is closer.

- Take the hyperplane normal to (x-y):

$$\{z : (x-y)^T z \leq (x-y)^T y\}$$

Oraculos de separacion

- For an LP, simply check all the linear constraints
- For a ball or ellipsoid, find the tangent plane
- For the SDP cone, check if the eigenvalues are all nonnegative; if not eigenvector gives a separating hyperplane.
- For cut example, find mincut to check if all cuts are at least 1.

Ejemplo: Aprendizaje por Muestreo

Sequence of points $X_1, X_2, \ldots,$
Unknown $-1/1$ function f

We get X_i and have to guess $f(X_i)$

Goal: minimize number of wrong guesses.

Aprender semi-espacios

Unknown -1/1 function f

f(X) = 1 if $w^T x > 0$ and f(X) = -1 otherwise

For an unknown vector w, with each component w_i being a b-bit integer.

What is the minimum number of mistakes?

Algoritmo de Mayoria

After $X_1, X_2, \ldots, X_k$

the set of consistent functions f correspond to

$S_k = \{w : w^T(\text{sign}(X_i)X_i) > 0 \text{ for } i = 1,2,\ldots, k\}$

Guess $f(X_{k+1})$, to be the majority of the predictions of each w in S_k

Claim. Number of wrong guesses $\leq (b+1)n$

But how to compute majority?? $|S_k|$ could be $2^{(b+1)n}$!

Algoritmo Aleatorio

- Pick random w in S_k
- Guess $w^T X$

Algoritmo Aleatorio

- Pick random w in S_k
- Guess $w^T X_{k+1}$

Lemma 1. E(#wrong guesses) ≤ 2bn.

Proof idea. Every time random guess is wrong, majority algorithm has probability at least ½ of being wrong.

Exercise 1. Prove Lemma 1.

Aprendizaje por Muestreo

- How to pick random w in $S_{k?}$
- S_k is a convex set!
- It can be efficiently sampled.

Tutorial outline

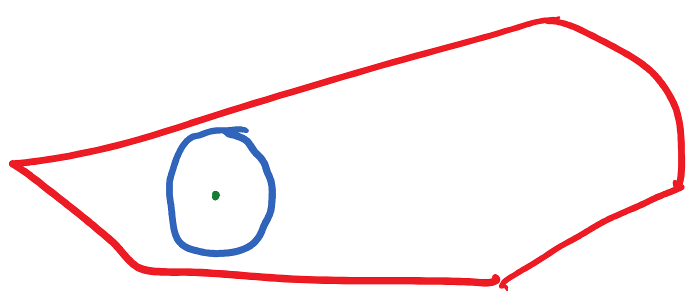
- Part 1. Intro to high dimension and convexity
  - Volume distribution, logconcavity
  - Ellipsoids
  - Lower bounds
- Part 2. Algorithms
  - Rounding
  - Volume/Integration
  - Optimization
  - Sampling
- Part 3. Probability
  - Markov chains
  - Conductance
  - Mixing of ball walk
  - Mixing of hit-and-run
- Part 4. Geometry
  - Isoperimetry
  - Concentration
  - Localization and applications
- Open problems

Un problema muy antiguo

Given a measurable, compact set K in n-dimensional space and $\epsilon > 0$, find a number A such that:

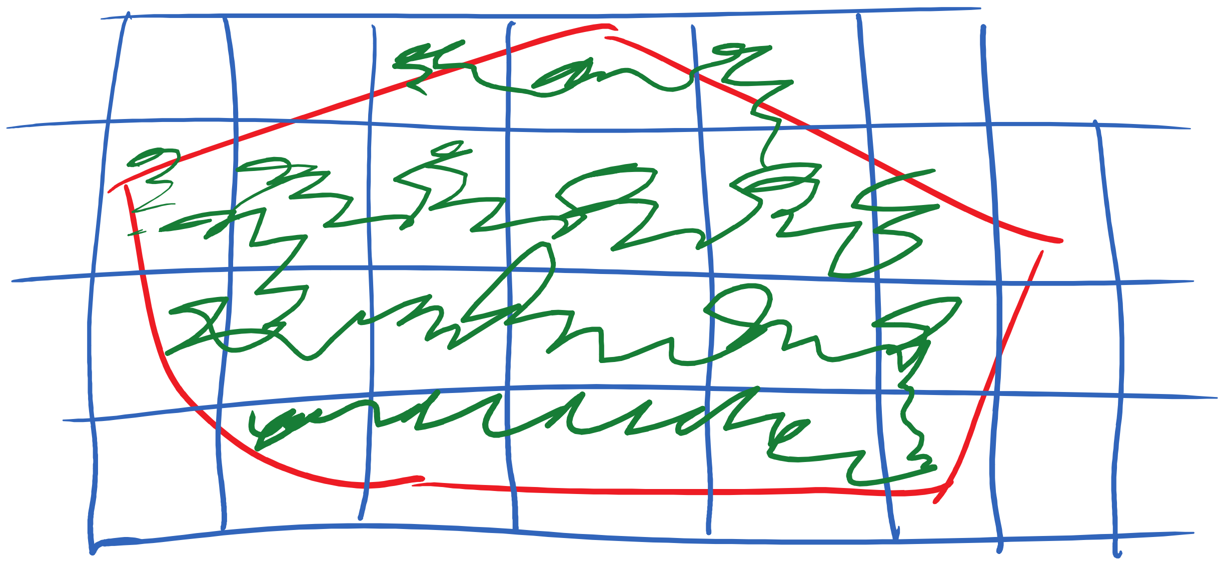
$$(1-\epsilon)\,\text{volume}(K) \;\le\; A \;\le\; (1+\epsilon)\,\text{volume}(K)$$

K is given by a well-guaranteed membership oracle.



Volumen: primer intento

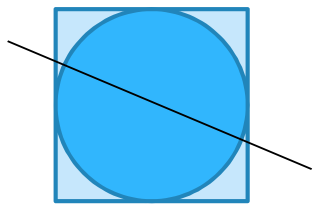
- Divide and conquer:



- Difficulty: number of parts grows exponentially in n.

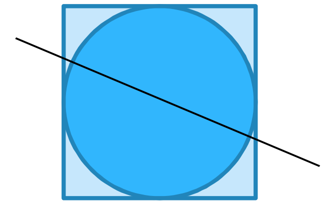
Distribucion del volumen

- Volume(unit cube) = 1
- Volume(unit ball) $\sim \left(\frac{c}{n}\right)^{\frac{\mathrm{n}}{2}}$ drops exponentially with n.

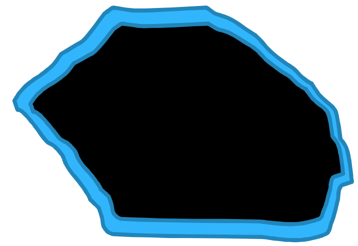


- For any central hyperplane, most of the mass of the unit ball is within distance $\frac{1}{\sqrt{n}}$. Section volume decays as $\left(1-t^2\right)^{\frac{n-1}{2}}$ at distance t from the origin.

Distribucion del volumen



- Volume(unit cube) = 1
- Volume(unit ball) $\sim \left(\frac{c}{n}\right)^{\frac{n}{2}}$ drops exponentially with n.
- For any central hyperplane, most of the mass of the unit ball is within distance $\frac{1}{\sqrt{n}}$.



- Most of the volume is near the boundary:

$$\mathrm{vol}((1-\varepsilon)K) = (1-\varepsilon)^n \mathrm{vol}(K)$$

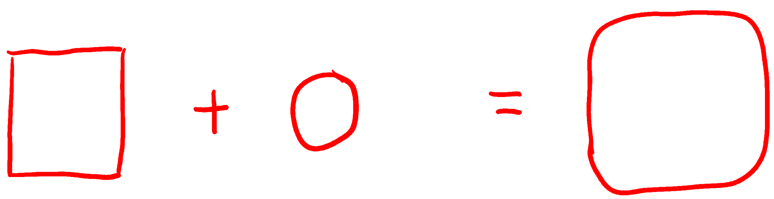
So, $\mathrm{vol}(K) - \mathrm{vol}\big((1-\varepsilon)K\big) \geq (1 - e^{-\varepsilon n})\mathrm{vol}(K)$

- "Everything interesting for a convex body happens near its boundary" --- Imre Bárány.

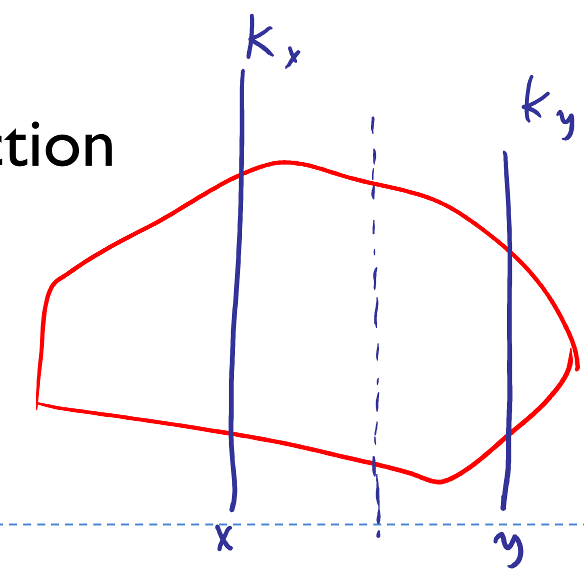
Distribucion del volumen

A,B sets in R^n, their Minkowski sum is:

$$A + B = \{x + y : x \in A, y \in B\}$$



For a convex body, the hyperplane section at $\frac{x+y}{2}$ contains $\frac{A_x + A_y}{2}$.



What is the volume distribution?

Brunn-Minkowski inequality

Thm. A, B compact sets in R^n, $\forall\lambda \in [0,1]$,

$$vol(\lambda A + (1-\lambda)B)^{\frac{1}{n}} \geq \lambda vol(A)^{\frac{1}{n}} + (1-\lambda)vol(B)^{\frac{1}{n}}$$

Suffices to prove

$$vol(A+B)^{\frac{1}{n}} \geq vol(A)^{\frac{1}{n}} + vol(B)^{\frac{1}{n}}$$

by taking the sets to be $\lambda A, (1-\lambda)B$

Brunn-Minkowski inequality

Thm. A, B: compact sets in R^n

$$vol(A+B)^{\frac{1}{n}} \geq vol(A)^{\frac{1}{n}} + vol(B)^{\frac{1}{n}}$$

Proof. First take A, B to be cuboids, i.e.,

$$A = [0,a_1] \times [0,a_2] \times \ldots \times [0,a_n]$$
$$B = [0,b_1] \times [0,b_2] \times \ldots \times [0,b_n]$$

$A+B = [0,a_1+b_1] \times [0,a_2+b_2] \ldots \times [0,a_n+b_n]$.

$$\frac{vol(A)^{\frac{1}{n}} + vol(B)^{\frac{1}{n}}}{vol(A+B)^{\frac{1}{n}}} \leq \left(\prod_i \frac{a_i}{a_i+b_i}\right)^{\frac{1}{n}} + \left(\prod_i \frac{b_i}{a_i+b_i}\right)^{\frac{1}{n}}$$
$$\leq \frac{1}{n}\sum_i \frac{a_i}{a_i+b_i} + \frac{1}{n}\sum_i \frac{b_i}{a_i+b_i} = 1.$$

Brunn-Minkowski inequality

Thm. A, B: compact sets in R^n: $vol(A+B)^{\frac{1}{n}} \geq vol(A)^{\frac{1}{n}} + vol(B)^{\frac{1}{n}}$

Take A,B to be finite unions of disjoint cuboids: A = $\bigcup_i A_i$ and B = $\bigcup_i B_i$. Induction.
There exists an axis-parallel hyperplane s.t. at least one complete cuboid of A is on either side.
Partition into A^+, B^+, A^-, B^-.

Translate B s.t. $\frac{vol(A)}{vol(B)} = \frac{vol(A^+)}{vol(B^+)}$.

$$\begin{aligned}
vol(A+B) &\geq vol(A^+ + B^+) + vol(A^- + B^-) \\
&\geq \left(vol(A^+)^{\frac{1}{n}} + vol(B^+)^{\frac{1}{n}}\right)^n + \left(vol(A^-)^{\frac{1}{n}} + vol(B^-)^{\frac{1}{n}}\right)^n \quad \text{(induction)} \\
&\geq vol(A^+)\left(1 + \frac{vol(B^+)^{\frac{1}{n}}}{vol(A^+)^{\frac{1}{n}}}\right)^n + vol(A^-)\left(1 + \frac{vol(B^-)^{\frac{1}{n}}}{vol(A^-)^{\frac{1}{n}}}\right)^n \\
&\geq \left(vol(A^+) + vol(A^-)\right)\left(1 + \frac{vol(B)^{\frac{1}{n}}}{vol(A)^{\frac{1}{n}}}\right)^n \geq \left(vol(A)^{\frac{1}{n}} + vol(B)^{\frac{1}{n}}\right)^n
\end{aligned}$$

Approximate to arbitrary accuracy by a finite union of cuboids.

Logconcave functions

- $f: R^n \to R$ is concave if for any $x, y \in R^n$,

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y)$$

- $f: R^n \to R_+$ is logconcave if for any $x, y \in R^n$,

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

i.e., f is nonnegative and its logarithm is concave.

Logconcave functions

- $f: R^n \to R_+$ is logconcave if for any $x, y \in R^n$,
$$f(\lambda x + (1-\lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}$$
- Examples:
  - Indicator functions of convex sets are logconcave
  - Gaussian density function,
  - exponential function
- Level sets, $L_f(t) = \{x : f(x) \geq t\}$, are convex.
- Many other useful geometric properties

Properties of logconcave functions

For two logconcave functions f and g

- Their sum might not be logconcave

- But their product $h(x) = f(x)g(x)$ is logconcave
- And so is their minimum $h(x) = \min f(x), g(x)$.

Prekopa-Leindler inequality

Prekopa-Leindler: $f, g, h: R^n \to R_+ \; s.t.$

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

then

$$\int h \geq \left(\int f\right)^\lambda \left(\int g\right)^{1-\lambda}.$$

Functional version of [B-M], *equivalent to it.*

Properties of logconcave functions

- Convolution is logconcave

$$h(x) = \int_{R^n} f(y)g(x-y)dy$$

- E.g., any marginal:

$$h(x_1, x_2, \ldots, x_k) = \int_{R^{n-k}} f(x)dx_{k+1}dx_{k+2} \ldots dx_n$$

This follows from Prekopa-Leindler.