CS599: Convex and Combinatorial Optimization
Fall 2013
Lecture 25: Unconstrained Submodular Function Minimization

Instructor: Shaddin Dughmi

Announcements

Outline

Recall: Optimizing Submodular Functions

- As our examples suggest, optimization problems involving submodular functions are very common
- These can be classified on two axes: constrained/unconstrained and maximization/minimization

	Maximization	Minimization
Unconstrained	NP-hard $\frac{1}{2}$ approximation	Polynomial time via convex opt
Constrained	Usually NP-hard $1 - 1/e$ (mono, matroid) $O(1)$ ("nice" constriants)	Usually NP-hard to apx. Few easy special cases

Recall: Optimizing Submodular Functions

- As our examples suggest, optimization problems involving submodular functions are very common
- These can be classified on two axes: constrained/unconstrained and maximization/minimization

	Maximization	Minimization
Unconstrained	NP-hard $\frac{1}{2}$ approximation	Polynomial time via convex opt
Constrained	Usually NP-hard $1 - 1/e$ (mono, matroid) $O(1)$ (“nice” constriants)	Usually NP-hard to apx. Few easy special cases

Problem Definition

Given a submodular function $f : 2^X \to \mathbb{R}$ on a finite ground set X,

$$\begin{array}{ll} \text{minimize} & f(S) \\ \text{subject to} & S \subseteq X \end{array}$$

- We denote $n = |X|$
- We assume $f(S)$ is a rational number with at most b bits

Problem Definition

Given a submodular function $f : 2^X \to \mathbb{R}$ on a finite ground set X,

$$\begin{array}{ll} \text{minimize} & f(S) \\ \text{subject to} & S \subseteq X \end{array}$$

- We denote $n = |X|$
- We assume $f(S)$ is a rational number with at most b bits

Representation

In order to generalize all our examples, algorithmic results are often posed in the value oracle model. Namely, we only assume we have access to a subroutine evaluating $f(S)$ in constant time.

Problem Definition

Given a submodular function $f : 2^X \to \mathbb{R}$ on a finite ground set X,

$$\begin{array}{ll} \text{minimize} & f(S) \\ \text{subject to} & S \subseteq X \end{array}$$

- We denote $n = |X|$
- We assume $f(S)$ is a rational number with at most b bits

Representation

In order to generalize all our examples, algorithmic results are often posed in the value oracle model. Namely, we only assume we have access to a subroutine evaluating $f(S)$ in constant time.

Goal

An algorithm which runs in time polynomial in n and b.

Problem Definition

Given a submodular function $f : 2^X \to \mathbb{R}$ on a finite ground set X,

$$\begin{array}{ll} \text{minimize} & f(S) \\ \text{subject to} & S \subseteq X \end{array}$$

- We denote $n = |X|$
- We assume $f(S)$ is a rational number with at most b bits

Representation

In order to generalize all our examples, algorithmic results are often posed in the value oracle model. Namely, we only assume we have access to a subroutine evaluating $f(S)$ in constant time.

Goal

An algorithm which runs in time polynomial in n and b.

Note: weakly polynomial. There are strongly polytime algorithms.

Examples

Minimum Cut

Given a graph $G = (V, E)$, find a set $S \subseteq V$ minimizing the number of edges crossing the cut $(S, V \setminus S)$.

- G may be directed or undirected.
- Extends to hypergraphs.

Examples

Minimum Cut

Given a graph $G = (V, E)$, find a set $S \subseteq V$ minimizing the number of edges crossing the cut $(S, V \setminus S)$.

- G may be directed or undirected.
- Extends to hypergraphs.

Densest Subgraph

Given an undirected graph $G = (V, E)$, find a set $S \subseteq V$ maximizing the average internal degree.

- Reduces to supermodular maximization via binary search for the right density.

Outline

Continuous Extensions of a Set Function

Recall

A set function f on $X = \{1, \ldots, n\}$ with can be thought of as a map from the vertices $\{0, 1\}^n$ of the n-dimensional hypercube to the real numbers.

Continuous Extensions of a Set Function

Recall

A set function f on $X = \{1, \ldots, n\}$ with can be thought of as a map from the vertices $\{0, 1\}^n$ of the n-dimensional hypercube to the real numbers.

We will consider extensions of a set function to the entire hypercube.

Extension of a Set Function

Given a set function $f : \{0, 1\}^n \to \mathbb{R}$, an extension of f to the hypercube $[0, 1]^n$ is a function $g : [0, 1]^n \to \mathbb{R}$ satisfying $g(x) = f(x)$ for every $x \in \{0, 1\}^n$.

Continuous Extensions of a Set Function

Recall

A set function f on $X = \{1, \ldots, n\}$ with can be thought of as a map from the vertices $\{0, 1\}^n$ of the n-dimensional hypercube to the real numbers.

We will consider extensions of a set function to the entire hypercube.

Extension of a Set Function

Given a set function $f : \{0, 1\}^n \to \mathbb{R}$, an extension of f to the hypercube $[0, 1]^n$ is a function $g : [0, 1]^n \to \mathbb{R}$ satisfying $g(x) = f(x)$ for every $x \in \{0, 1\}^n$.

Long story short. . .

We will exhibit an extension which is convex when f is submodular, and can be minimized efficiently. We will then show that minimizing it yields a solution to the submodular minimization problem.

The Convex Closure

Convex Closure

Given a set function $f : \{0,1\}^n \to \mathbb{R}$, the convex closure $f^- : [0,1]^n \to \mathbb{R}$ of f is the point-wise greatest convex function under-estimating f on $\{0,1\}^n$.

The Convex Closure

Convex Closure

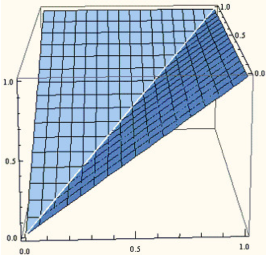
Given a set function $f : \{0,1\}^n \to \mathbb{R}$, the convex closure $f^- : [0,1]^n \to \mathbb{R}$ of f is the point-wise greatest convex function under-estimating f on $\{0,1\}^n$.

Geometric Intuition

What you would get by placing a blanket under the plot of f and pulling up.

$$f(\emptyset) = 0$$
$$f(\{1\}) = f(\{2\}) = 1$$
$$f(\{1,2\}) = 1$$

$$f^-(x_1, x_2) = \max(x_1, x_2)$$



The Convex Closure

Convex Closure

Given a set function $f : \{0,1\}^n \to \mathbb{R}$, the convex closure $f^- : [0,1]^n \to \mathbb{R}$ of f is the point-wise greatest convex function under-estimating f on $\{0,1\}^n$.

Claim

The convex closure exists for any set function.

Proof

- If $g_1, g_2 : [0,1]^n \to \mathbb{R}$ are convex under-estimators of f, then so is $\max \{g_1, g_2\}$
- Holds for infinite set of convex under-estimators
- Therefore $f^- = \max \{g : g$ is a convex underestimator of $f\}$ is the point-wise greatest convex underestimator of f.

Claim

The value of the convex closure at $x \in [0,1]^n$ is the solution of the following optimization problem:

$$\begin{array}{ll} \text{minimize} & \sum_{y \in \{0,1\}^n} \lambda_y f(y) \\ \text{subject to} & \sum_{y \in \{0,1\}^n} \lambda_y y = x \\ & \sum_{y \in \{0,1\}^n} \lambda_y = 1 \\ & \lambda_y \geq 0, \qquad\qquad \text{for } y \in \{0,1\}^n . \end{array}$$

Interpretation

- The minimum expected value of f over all distributions on $\{0,1\}^n$ with expectation x.
- Equivalently: the minimum expected value of f for a random set $S \subseteq X$ including each $i \in X$ with probability x_i.
- The upper bound on $f^-(x)$ implied by applying Jensen's inequality to every convex combination $\{0,1\}^n$.

Claim

The value of the convex closure at $x \in [0,1]^n$ is the solution of the following optimization problem:

$$\begin{array}{lll} \text{minimize} & \sum_{y \in \{0,1\}^n} \lambda_y f(y) & \\ \text{subject to} & \sum_{y \in \{0,1\}^n} \lambda_y y = x & \\ & \sum_{y \in \{0,1\}^n} \lambda_y = 1 & \\ & \lambda_y \geq 0, & \text{for } y \in \{0,1\}^n . \end{array}$$

Implication

- f^- is a convex extension of f.
- $f^-(x)$ has no "integrality gap"
  - For every $x \in [0,1]^n$, there is a random integer vector $y \in \{0,1\}^n$ such that $\mathbf{E}_y \, f(y) = f^-(x)$.
  - Therefore, there is an integer vector y such that $f(y) \leq f^-(x)$.

Claim

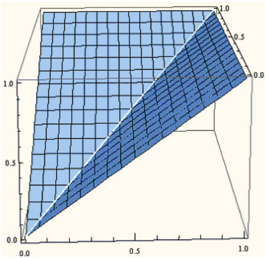
The value of the convex closure at $x \in [0,1]^n$ is the solution of the following optimization problem:

$$\begin{array}{ll} \text{minimize} & \sum_{y \in \{0,1\}^n} \lambda_y f(y) \\ \text{subject to} & \sum_{y \in \{0,1\}^n} \lambda_y y = x \\ & \sum_{y \in \{0,1\}^n} \lambda_y = 1 \\ & \lambda_y \geq 0, \qquad \text{for } y \in \{0,1\}^n . \end{array}$$

$f(\emptyset) = 0$
$f(\{1\}) = f(\{2\}) = 1$
$f(\{1,2\}) = 1$

When $x_1 \leq x_2$

$$\begin{aligned} f^-(x_1, x_2) &= x_1 f(\{1,2\}) \\ &+ (x_2 - x_1) f(\{2\}) \\ &+ (1 - x_2) f(\emptyset) \end{aligned}$$



Claim

The value of the convex closure at $x \in [0,1]^n$ is the solution of the following optimization problem:

$$\begin{array}{lll} \text{minimize} & \sum_{y \in \{0,1\}^n} \lambda_y f(y) & \\ \text{subject to} & \sum_{y \in \{0,1\}^n} \lambda_y y = x & \\ & \sum_{y \in \{0,1\}^n} \lambda_y = 1 & \\ & \lambda_y \geq 0, & \text{for } y \in \{0,1\}^n . \end{array}$$

Proof

- $OPT(x)$ is at least $f^{-}(x)$ for every x: By Jensen's inequality

Claim

The value of the convex closure at $x \in [0,1]^n$ is the solution of the following optimization problem:

$$\begin{array}{lll} \text{minimize} & \sum_{y \in \{0,1\}^n} \lambda_y f(y) & \\ \text{subject to} & \sum_{y \in \{0,1\}^n} \lambda_y y = x & \\ & \sum_{y \in \{0,1\}^n} \lambda_y = 1 & \\ & \lambda_y \geq 0, & \text{for } y \in \{0,1\}^n . \end{array}$$

Proof

- $OPT(x)$ is at least $f^-(x)$ for every x: By Jensen's inequality
- To show that $OPT(x)$ is equal to $f^-(x)$, suffices to show that is a convex under-estimate of f

Claim

The value of the convex closure at $x \in [0,1]^n$ is the solution of the following optimization problem:

$$\begin{array}{lll} \text{minimize} & \sum_{y \in \{0,1\}^n} \lambda_y f(y) & \\ \text{subject to} & \sum_{y \in \{0,1\}^n} \lambda_y y = x & \\ & \sum_{y \in \{0,1\}^n} \lambda_y = 1 & \\ & \lambda_y \geq 0, & \text{for } y \in \{0,1\}^n . \end{array}$$

Proof

- $OPT(x)$ is at least $f^-(x)$ for every x: By Jensen's inequality
- To show that $OPT(x)$ is equal to $f^-(x)$, suffices to show that is a convex under-estimate of f
- Under-estimate: $OPT(x) = f(x)$ for $x \in \{0,1\}^n$

Claim

The value of the convex closure at $x \in [0,1]^n$ is the solution of the following optimization problem:

$$\begin{array}{lll} \text{minimize} & \sum_{y \in \{0,1\}^n} \lambda_y f(y) & \\ \text{subject to} & \sum_{y \in \{0,1\}^n} \lambda_y y = x & \\ & \sum_{y \in \{0,1\}^n} \lambda_y = 1 & \\ & \lambda_y \geq 0, & \text{for } y \in \{0,1\}^n . \end{array}$$

Proof

- $OPT(x)$ is at least $f^-(x)$ for every x: By Jensen's inequality
- To show that $OPT(x)$ is equal to $f^-(x)$, suffices to show that is a convex under-estimate of f
- Under-estimate: $OPT(x) = f(x)$ for $x \in \{0,1\}^n$
- Convex: The value of a minimization LP is convex in its right hand side constants (check)

Using the Convex Closure

Fact

The minimum of f^- is equal to the minimum of f, and moreover is attained at minimizers $y \in \{0,1\}^n$ of f.

Proof

Using the Convex Closure

Fact

The minimum of f^- is equal to the minimum of f, and moreover is attained at minimizers $y \in \{0,1\}^n$ of f.

Proof

- $f^-(y) = f(y)$ for every $y \in \{0,1\}^n$
- Therefore $\min_{x \in [0,1]^n} f^-(x) \leq \min_{y \in \{0,1\}^n} f(y)$

Using the Convex Closure

Fact

The minimum of f^- is equal to the minimum of f, and moreover is attained at minimizers $y \in \{0,1\}^n$ of f.

Proof

- $f^-(y) = f(y)$ for every $y \in \{0,1\}^n$
- Therefore $\min_{x \in [0,1]^n} f^-(x) \leq \min_{y \in \{0,1\}^n} f(y)$
- For every x, $f^-(x)$ is the expected value of $f(y)$, for a random variable $y \in \{0,1\}^n$ with expectation x.
- Therefore, $\min_{x \in [0,1]^n} f^-(x) \geq \min_{y \in \{0,1\}^n} f(y)$

Using the Convex Closure

Fact

The minimum of f^- is equal to the minimum of f, and moreover is attained at minimizers $y \in \{0,1\}^n$ of f.

Good News?

We reduced minimizing set function f to minimizing a convex function f^- over a convex set $[0,1]^n$. Are we done?

Using the Convex Closure

Fact

The minimum of f^- is equal to the minimum of f, and moreover is attained at minimizers $y \in \{0, 1\}^n$ of f.

Good News?

We reduced minimizing set function f to minimizing a convex function f^- over a convex set $[0, 1]^n$. Are we done?

Problem

In general, it is hard to evaluate f^- efficiently, let alone its derivative. This is indispensible for convex optimization algorithms.

Using the Convex Closure

Fact

The minimum of f^- is equal to the minimum of f, and moreover is attained at minimizers $y \in \{0,1\}^n$ of f.

Good News?

We reduced minimizing set function f to minimizing a convex function f^- over a convex set $[0,1]^n$. Are we done?

Problem

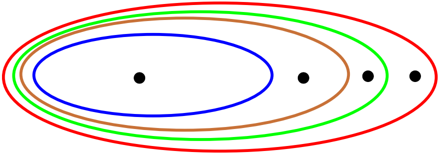
In general, it is hard to evaluate f^- efficiently, let alone its derivative. This is indispensible for convex optimization algorithms.

We will show that, when f is submodular, f^- is in fact equivalent to another extension which is easier to evaluate.

Chain Distributions

Chain Distribution

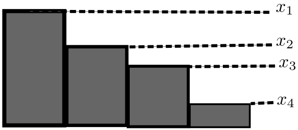
A chain distribution on the ground set X is a distribution over $S \subseteq X$ who's support forms a chain in the inclusion order.

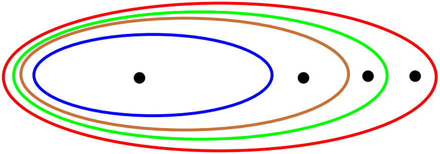


Chain Distributions

Chain Distribution with Given Marginals

Fix the ground set $X = \{1, \ldots, n\}$. The chain distribution with marginals $x \in [0,1]^n$ is the unique chain distribution $D^{\mathcal{L}}(x)$ satisfying $\mathbf{Pr}_{S \sim D^{\mathcal{L}}(x)}[i \in S] = x_i$ for all $i \in X$.

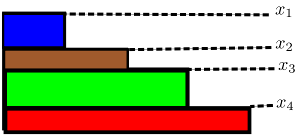


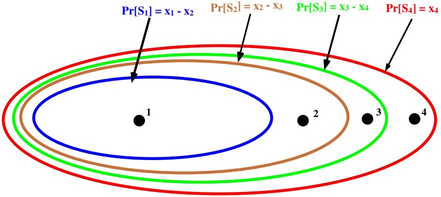


Chain Distributions

Chain Distribution with Given Marginals

Fix the ground set $X = \{1, \ldots, n\}$. The chain distribution with marginals $x \in [0,1]^n$ is the unique chain distribution $D^{\mathcal{L}}(x)$ satisfying $\mathbf{Pr}_{S \sim D^{\mathcal{L}}(x)}[i \in S] = x_i$ for all $i \in X$.

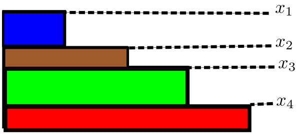


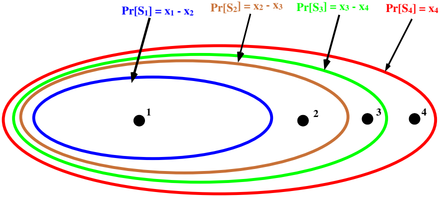


Chain Distributions

Chain Distribution with Given Marginals

Fix the ground set $X = \{1, \ldots, n\}$. The chain distribution with marginals $x \in [0,1]^n$ is the unique chain distribution $D^{\mathcal{L}}(x)$ satisfying $\mathbf{Pr}_{S \sim D^{\mathcal{L}}(x)}[i \in S] = x_i$ for all $i \in X$.





$D^{\mathcal{L}}(x)$ is the distribution given by the following process:

- Sort $x_1 \geq x_2 \ldots \geq x_n$
- Let $S_i = \{x_1, \ldots, x_i\}$
- Let $\mathbf{Pr}[S_i] = x_i - x_{i+1}$

The Lovasz Extension

Definition

The Lovasz extension of a set function f is defined as follows.

$$f^{\mathcal{L}}(x) = \mathop{\mathbf{E}}_{S \sim D^{\mathcal{L}}(x)} f(S)$$

i.e. the Lovasz extension at x is the expected value of a set drawn from the unique chain distribution with marginals x.

Observations

- $f^{\mathcal{L}}$ is an extension, since the chain distribution with marginals $y \in \{0,1\}^n$ is the point distribution at y.

The Lovasz Extension

Definition

The Lovasz extension of a set function f is defined as follows.

$$f^{\mathcal{L}}(x) = \mathop{\mathbf{E}}_{S \sim D^{\mathcal{L}}(x)} f(S)$$

i.e. the Lovasz extension at x is the expected value of a set drawn from the unique chain distribution with marginals x.

Observations

- $f^{\mathcal{L}}$ is an extension, since the chain distribution with marginals $y \in \{0,1\}^n$ is the point distribution at y.
- $f^{\mathcal{L}}(x)$ is the expected value of f on some distribution on $\{0,1\}^n$ with marginals x, therefore $f^{\mathcal{L}}(x) \geq f^{-}(x)$.

The Lovasz Extension

Definition

The Lovasz extension of a set function f is defined as follows.

$$f^{\mathcal{L}}(x) = \mathop{\mathbf{E}}_{S \sim D^{\mathcal{L}}(x)} f(S)$$

i.e. the Lovasz extension at x is the expected value of a set drawn from the unique chain distribution with marginals x.

Observations

- $f^{\mathcal{L}}$ is an extension, since the chain distribution with marginals $y \in \{0,1\}^n$ is the point distribution at y.
- $f^{\mathcal{L}}(x)$ is the expected value of f on some distribution on $\{0,1\}^n$ with marginals x, therefore $f^{\mathcal{L}}(x) \geq f^{-}(x)$.
- Together, those imply: if $f^{\mathcal{L}}$ is convex, then $f^{\mathcal{L}} = f^{-}$.

Equivalence of the Convex Closure and Lovasz Extension

Theorem

If f is submodular, then $f^{\mathcal{L}} = f^{-}$.

Converse holds: if f is not submodular, then $f^{\mathcal{L}}$ is not convex.

Equivalence of the Convex Closure and Lovasz Extension

Theorem

If f is submodular, then $f^{\mathcal{L}} = f^-$.

Converse holds: if f is not submodular, then $f^{\mathcal{L}}$ is not convex.

Intuition

- Recall: $f^-(x)$ evaluates f on the “lowest” distribution with marginals x
- It turns out that, when f is submodular, this lowest distribution is the chain distribution $D^{\mathcal{L}}(x)$.

Equivalence of the Convex Closure and Lovasz Extension

Theorem

If f is submodular, then $f^{\mathcal{L}} = f^{-}$.

Converse holds: if f is not submodular, then $f^{\mathcal{L}}$ is not convex.

Intuition

- Recall: $f^{-}(x)$ evaluates f on the "lowest" distribution with marginals x
- It turns out that, when f is submodular, this lowest distribution is the chain distribution $D^{\mathcal{L}}(x)$.
- Contingent on marginals x, submodularity implies that cost is minimized by "packing" as many elements together as possible
  - diminishing marginal returns
- This gives the chain distribution

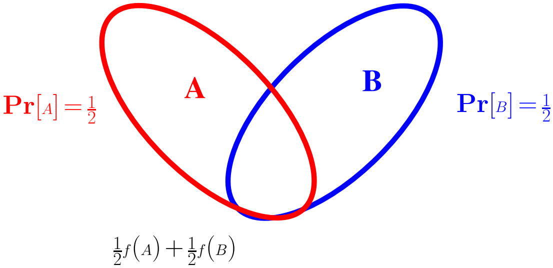
It suffices to show that the chain distribution with marginals x is in fact the "lowest" distribution with marginals x.

Proof (Special case)

It suffices to show that the chain distribution with marginals x is in fact the "lowest" distribution with marginals x.

Proof (Special case)

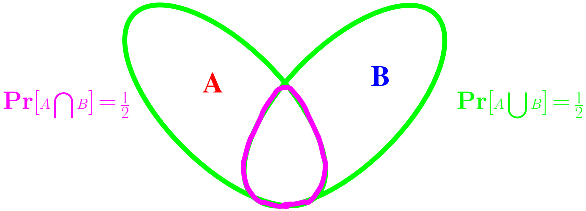
- Consider a distribution $\mathcal{D}$ on two "crossing" sets A and B, with probability 0.5 each.



It suffices to show that the chain distribution with marginals x is in fact the “lowest” distribution with marginals x.

Proof (Special case)

- Consider a distribution $\mathcal{D}$ on two “crossing” sets A and B, with probability 0.5 each.
- “uncrossing” implies that replacing them with $A \bigcap B$ and $A \bigcup B$, with probability 0.5 each, gives a chain distribution with lower expected value of f.

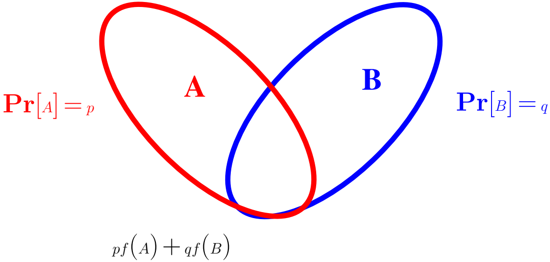


$$\frac{1}{2} f(A) + \frac{1}{2} f(B) \geq \frac{1}{2} f(A \bigcap B) + \frac{1}{2} f(A \bigcup B)$$

Proof (Slightly Less Special Case)

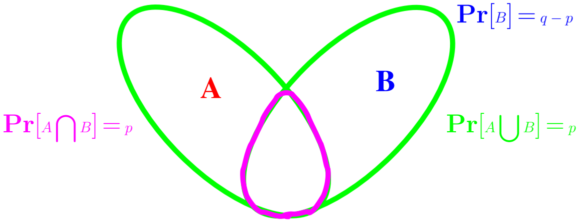
Proof (Slightly Less Special Case)

- Consider a distribution $\mathcal{D}$ on two “crossing” sets A and B, with probabilities $p \leq q$.



Proof (Slightly Less Special Case)

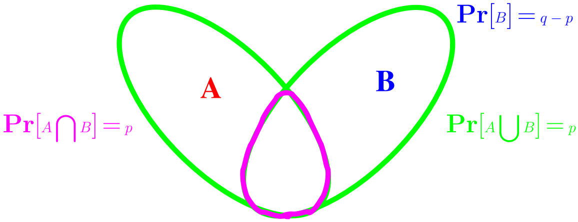
- Consider a distribution $\mathcal{D}$ on two "crossing" sets A and B, with probabilities $p \leq q$.
- Can "uncross" a probability mass of p of each, decreasing the expected value of f



$$pf(A) + qf(B) \geq pf(A \bigcap B) + pf(A \bigcup B) + (q-p)f(B)$$

Proof (Slightly Less Special Case)

- Consider a distribution $\mathcal{D}$ on two “crossing” sets A and B, with probabilities $p \leq q$.
- Can “uncross” a probability mass of p of each, decreasing the expected value of f
- Now a chain distribution

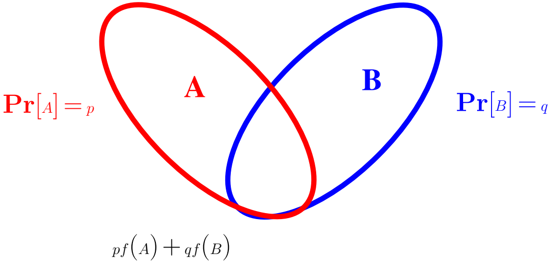


$$pf(A) + qf(B) \geq pf\left(A \bigcap B\right) + pf\left(A \bigcup B\right) + (q-p)f(B)$$

Proof (General Case)

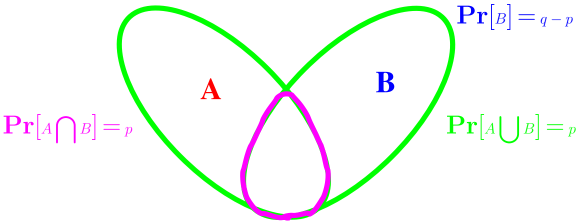
Proof (General Case)

- Consider a distribution $\mathcal{D}$ which includes two "crossing" sets A and B in its support



Proof (General Case)

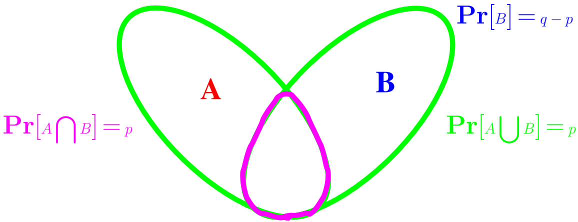
- Consider a distribution $\mathcal{D}$ which includes two "crossing" sets A and B in its support
- Can "uncross" a probability mass of $\min(\mathbf{Pr}[A], \mathbf{Pr}[B])$ of each, decreasing expected value of f



$$pf(A) + qf(B) \geq pf(A \bigcap B) + pf(A \bigcup B) + (q-p)f(B)$$

Proof (General Case)

- Consider a distribution $\mathcal{D}$ which includes two "crossing" sets A and B in its support
- Can "uncross" a probability mass of $\min(\mathbf{Pr}[A], \mathbf{Pr}[B])$ of each, decreasing expected value of f
- Decreases number of crossing pairs of sets in the support.
  - Closer to being a chain distribution.



$$pf(A) + qf(B) \geq pf(A \bigcap B) + pf(A \bigcup B) + (q-p)f(B)$$

Outline

Minimizing the Lovasz Extension

Because $f^{\mathcal{L}} = f^{-}$, we know the following:

Fact

The minimum of $f^{\mathcal{L}}$ is equal to the minimum of f, and moreover is attained at minimizers $y \in \{0,1\}^n$ of f.

Minimizing the Lovasz Extension

Because $f^{\mathcal{L}} = f^-$, we know the following:

Fact

The minimum of $f^{\mathcal{L}}$ is equal to the minimum of f, and moreover is attained at minimizers $y \in \{0,1\}^n$ of f.

Therefore, minimizing f reduces to the following convex optimization problem

Minimizing the Lovasz Extension

$$\begin{array}{ll} \text{minimize} & f^{\mathcal{L}}(x) \\ \text{subject to} & x \in [0,1]^n \end{array}$$

Recall: Solvability of Convex Optimization

Weak Solvability

An algorithm weakly solves our optimization problem if it takes in approximation parameter $\epsilon > 0$, runs in $\mathrm{poly}(n, \log \frac{1}{\epsilon})$ time, and returns $x \in [0,1]^n$ which is ϵ-optimal:

$$f^{\mathcal{L}}(x) \leq \min_{y \in [0,1]^n} f^{\mathcal{L}}(y) + \epsilon[\max_{y \in [0,1]^n} f^{\mathcal{L}}(y) - \min_{y \in [0,1]^n} f^{\mathcal{L}}(y)]$$

Recall: Solvability of Convex Optimization

Polynomial Solvability of CP

In order to weakly minimize $f^{\mathcal{L}}$, we need the following operations to run in $\mathrm{poly}(n)$ time:

1. Compute a starting ellipsoid $E \supseteq [0,1]^n$ with $\frac{\mathbf{vol}(E)}{\mathbf{vol}([0,1]^n)} = O(\exp(n))$.
2. A separation oracle for the feasible set $[0,1]^n$
3. A first order oracle for $f^{\mathcal{L}}$: evaluates $f^{\mathcal{L}}(x)$ and a subgradient of $f^{\mathcal{L}}$ at x.

Recall: Solvability of Convex Optimization

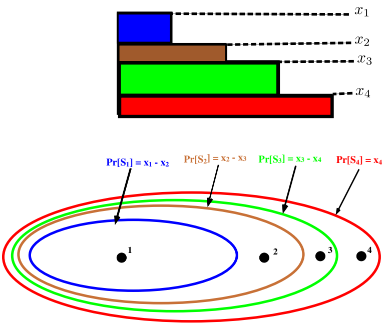
Polynomial Solvability of CP

In order to weakly minimize $f^{\mathcal{L}}$, we need the following operations to run in $\text{poly}(n)$ time:

1. Compute a starting ellipsoid $E \supseteq [0,1]^n$ with $\frac{\mathbf{vol}(E)}{\mathbf{vol}([0,1]^n)} = O(\exp(n))$.
2. A separation oracle for the feasible set $[0,1]^n$
3. A first order oracle for $f^{\mathcal{L}}$: evaluates $f^{\mathcal{L}}(x)$ and a subgradient of $f^{\mathcal{L}}$ at x.

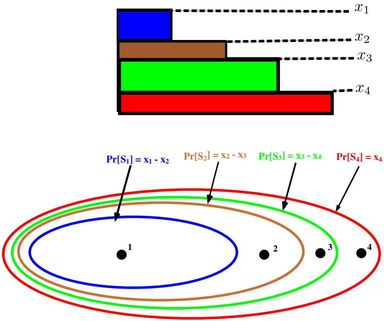
1 and 2 are trivial.

First order Oracle for $f^{\mathcal{L}}$

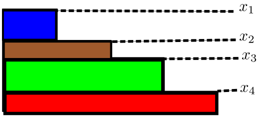


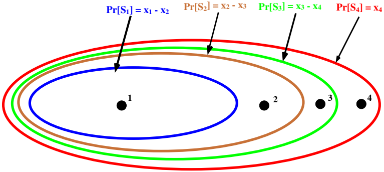
- Recall: the chain distribution with marginals x
  - Sort $x_1 \geq x_2 \ldots \geq x_n$
  - Let $S_i = \{x_1, \ldots, x_i\}$
  - Let $\mathbf{Pr}[S_i] = x_i - x_{i+1}$

First order Oracle for $f^{\mathcal{L}}$



- Recall: the chain distribution with marginals x
  - Sort $x_1 \geq x_2 \ldots \geq x_n$
  - Let $S_i = \{x_1, \ldots, x_i\}$
  - Let $\mathbf{Pr}[S_i] = x_i - x_{i+1}$
- Can evaluate $f^{\mathcal{L}}(x) = \sum_i f(S_i)(x_i - x_{i+1})$





- Recall: the chain distribution with marginals x
  - Sort $x_1 \geq x_2 \ldots \geq x_n$
  - Let $S_i = \{x_1, \ldots, x_i\}$
  - Let $\mathbf{Pr}[S_i] = x_i - x_{i+1}$
- Can evaluate $f^{\mathcal{L}}(x) = \sum_i f(S_i)(x_i - x_{i+1})$
- $f^{\mathcal{L}}$ is peicewise linear, so can compute a sub-gradient.

Recovering an Optimal Set

We can get an ϵ-optimal solution x^* to the optimization problem in $\mathrm{poly}(n, \log \frac{1}{\epsilon})$ time.

Minimizing the Lovasz Extension

$$\begin{array}{ll} \text{minimize} & f^{\mathcal{L}}(x) \\ \text{subject to} & x \in [0,1]^n \end{array}$$

Recovering an Optimal Set

We can get an ϵ-optimal solution x^* to the optimization problem in $\text{poly}(n, \log \frac{1}{\epsilon})$ time.

Minimizing the Lovasz Extension

$$\begin{array}{ll} \text{minimize} & f^{\mathcal{L}}(x) \\ \text{subject to} & x \in [0,1]^n \end{array}$$

- Set $\epsilon < 2^{-b}$, runtime is $\text{poly}(n, b)$.

Recovering an Optimal Set

We can get an ϵ-optimal solution x^* to the optimization problem in $\text{poly}(n, \log \frac{1}{\epsilon})$ time.

Minimizing the Lovasz Extension

$$\begin{array}{ll} \text{minimize} & f^{\mathcal{L}}(x) \\ \text{subject to} & x \in [0,1]^n \end{array}$$

- Set $\epsilon < 2^{-b}$, runtime is $\text{poly}(n, b)$.
- $\min_S f(S) \leq f^{\mathcal{L}}(x^*) < \min2_S f(S)$

Recovering an Optimal Set

We can get an ϵ-optimal solution x^* to the optimization problem in $\mathrm{poly}(n, \log \frac{1}{\epsilon})$ time.

Minimizing the Lovasz Extension

$$\begin{array}{ll} \text{minimize} & f^{\mathcal{L}}(x) \\ \text{subject to} & x \in [0,1]^n \end{array}$$

- Set $\epsilon < 2^{-b}$, runtime is $\mathrm{poly}(n, b)$.
- $\min_S f(S) \leq f^{\mathcal{L}}(x^*) < \min2_S f(S)$
- $f^{\mathcal{L}}(x^*)$ is the expectation f over a distribution of sets
  - It must include an optimal set in its support

Recovering an Optimal Set

We can get an ϵ-optimal solution x^* to the optimization problem in $\text{poly}(n, \log \frac{1}{\epsilon})$ time.

Minimizing the Lovasz Extension

$$\begin{array}{ll} \text{minimize} & f^{\mathcal{L}}(x) \\ \text{subject to} & x \in [0,1]^n \end{array}$$

- Set $\epsilon < 2^{-b}$, runtime is $\text{poly}(n, b)$.
- $\min_S f(S) \leq f^{\mathcal{L}}(x^*) < \min 2_S f(S)$
- $f^{\mathcal{L}}(x^*)$ is the expectation f over a distribution of sets
  - It must include an optimal set in its support
- We can identify this set by examining the chain distribution with marginals x^*