Junta Approximations for Submodular, XOS and Self-Bounding Functions

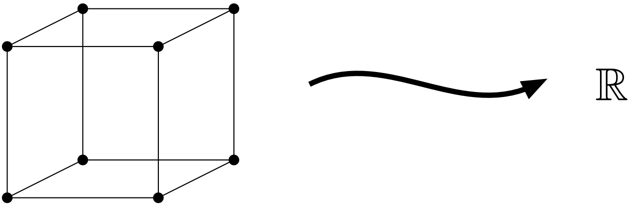
Vitaly Feldman   Jan Vondrák

IBM Almaden Research Center

Simons Institute, Berkeley, October 2013

Junta approximations

How can we simplify a function $f : \{0,1\}^n \to \mathbb{R}$?



In this talk:

- How well can we approximate f by a function g of few variables?

Def.: g approximates f within ϵ in L_p, if

$$\|f - g\|_p = (\mathbb{E}[|f(x) - g(x)|^p])^{1/p} \leq \epsilon$$

(in this talk, the uniform distribution)

Friedgut's Theorem

Definition (Average sensitivity)

The average sensitivity, or total influence, of $f : \{0,1\}^n \to \{0,1\}$ is

$$\mathrm{Infl}(f) = \sum_{i=1}^{n} \Pr_{x \in \{0,1\}^n}[f(x) \neq f(x \oplus e_i)].$$

Theorem (Friedgut '98)

For any function $f : \{0,1\}^n \to \{0,1\}$ of average sensitivity $\mathrm{Infl}(f)$ and every $\epsilon > 0$, there is a function g depending on $2^{O(\mathrm{Infl}(f)/\epsilon)}$ variables such that $\|f - g\|_1 \leq \epsilon$.

Junta approximations of real-valued functions

We investigate classes of **real-valued functions** $f : \{0,1\}^n \to [0,1]$:

- submodular functions: $f(x \vee y) + f(x \wedge y) \leq f(x) + f(y)$
- XOS functions: $f(x) = \max_i \sum_j a_{ij} x_j$
- subadditive functions: $f(x \vee y) \leq f(x) + f(y)$
- self-bounding functions: $f(x) \geq \sum_i (f(x) - f(x \oplus e_i))_+$

Junta approximations of real-valued functions

We investigate classes of **real-valued functions** $f : \{0,1\}^n \to [0,1]$:

- submodular functions: $f(x \vee y) + f(x \wedge y) \leq f(x) + f(y)$
- XOS functions: $f(x) = \max_i \sum_j a_{ij} x_j$
- subadditive functions: $f(x \vee y) \leq f(x) + f(y)$
- self-bounding functions: $f(x) \geq \sum_i (f(x) - f(x \oplus e_i))_+$

Why these classes?

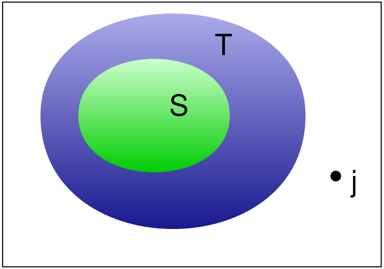
- Nice mathematical properties
- Role in game theory as *valuation functions* on bundles of goods

[Balcan-Harvey '11] Can we learn valuations from random examples?

Submodular functions

Submodularity = property of *diminishing returns*.

Let the *marginal value* of element j be $\partial_j f(S) = f(S \cup \{j\}) - f(S)$.
(we identify $f(S) = f(\mathbf{1}_S)$)



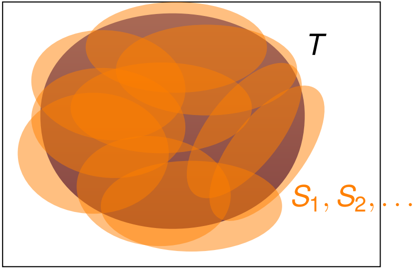
Definition: f is submodular, if for $S \subset T$ j cannot add more value to T than S.

$$\partial_j f(S) \geq \partial_j f(T)$$

Equivalently: $f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$.

Subadditive and Fractionally Subadditive functions

Definition: f is subadditive, if $f(A \cup B) \leq f(A) + f(B)$ for all A, B.

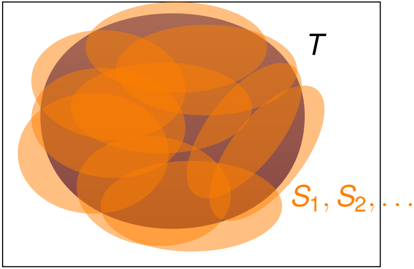


Definition: f is fractionally subadditive,

if $f(T) \leq \sum \alpha_i f(S_i)$
whenever $\mathbf{1}_T \leq \sum \alpha_i \mathbf{1}_{S_i}$.

Subadditive and Fractionally Subadditive functions

Definition: f is subadditive, if $f(A \cup B) \leq f(A) + f(B)$ for all A, B.



Definition: f is fractionally subadditive,

if $f(T) \leq \sum \alpha_i f(S_i)$
whenever $\mathbf{1}_T \leq \sum \alpha_i \mathbf{1}_{S_i}$.

Definition: f is an XOS function,
if f is a maximum over linear functions: $f(x) = \max_i \sum_j a_{ij} x_j \quad (a_{ij} \geq 0)$

Fact (for monotone functions with $f(\emptyset) = 0$)

Submodular $\subset$ Fract. Subadditive = XOS $\subset$ Subadditive Functions

Self-bounding functions

Definition: A function $f : D^n \to \mathbb{R}_+$ is a-self-bounding, if

$$\sum_{i=1}^{n} (f(x) - \min_{y_i \in D} f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)) \leq af(x).$$

Theorem: [Boucheron, Lugosi, Massart, 2000]
1-Lipschitz 1-self-bounding functions under product distributions are concentrated around $\mathbb{E}[f]$ with standard deviation $O(\sqrt{\mathbb{E}[f]})$ and
$\Pr[f(X) < \mathbb{E}[f] - t] < e^{-t^2/2\mathbb{E}[f]}, \quad \Pr[f(X) > \mathbb{E}[f] + t] > e^{-t^2/(2\mathbb{E}[f]+t)}$.

Self-bounding functions

Definition: A function $f : D^n \to \mathbb{R}_+$ is a-self-bounding, if

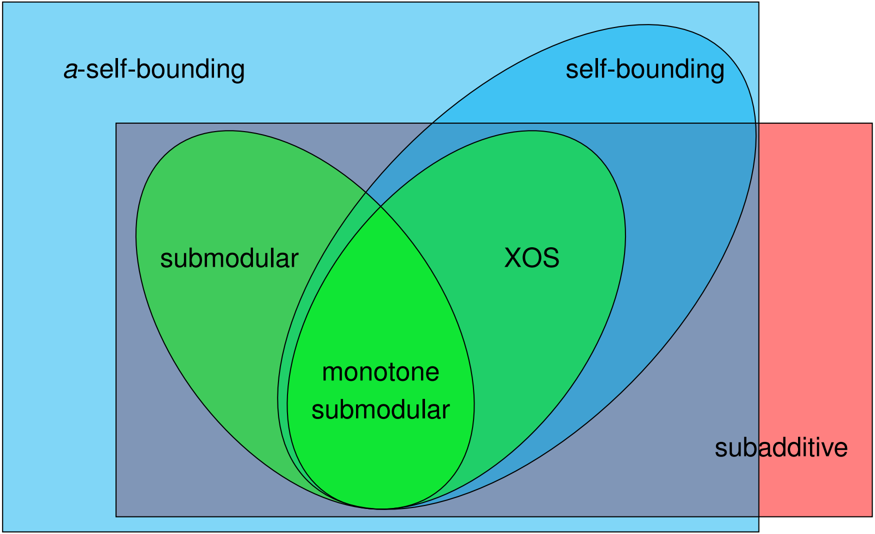
$$\sum_{i=1}^{n} (f(x) - \min_{y_i \in D} f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)) \leq af(x).$$

Theorem: [Boucheron, Lugosi, Massart, 2000]
1-Lipschitz 1-self-bounding functions under product distributions are concentrated around $\mathbb{E}[f]$ with standard deviation $O(\sqrt{\mathbb{E}[f]})$ and
$\Pr[f(X) < \mathbb{E}[f] - t] < e^{-t^2/2\mathbb{E}[f]}, \quad \Pr[f(X) > \mathbb{E}[f] + t] > e^{-t^2/(2\mathbb{E}[f]+t)}$.

Fact

XOS $\subset$ 1-Self-bounding functions.
Submodular $\subset$ 2-Self-bounding functions.

Overview of our function classes



Related work (learning of submodular functions)

[Balcan-Harvey '11]

- initiated the study of *learning of submodular functions*
- gave a learning algorithm for product distributions, using concentration properties of Lipschitz submodular functions
- proved a negative result for general distributions (no efficient learning within factors better than $n^{1/3}$)

Related work (learning of submodular functions)

[Balcan-Harvey '11]

- initiated the study of *learning of submodular functions*
- gave a learning algorithm for product distributions, using concentration properties of Lipschitz submodular functions
- proved a negative result for general distributions (no efficient learning within factors better than $n^{1/3}$)

[Gupta-Hardt-Roth-Ullman '11]

- learning of submodular fn. with applications in differential privacy
- decomposition into $n^{O(1/\epsilon^2)}$ ϵ-Lipschitz functions.

Related work (learning of submodular functions)

[Balcan-Harvey '11]

- initiated the study of *learning of submodular functions*
- gave a learning algorithm for product distributions, using concentration properties of Lipschitz submodular functions
- proved a negative result for general distributions (no efficient learning within factors better than $n^{1/3}$)

[Gupta-Hardt-Roth-Ullman '11]

- learning of submodular fn. with applications in differential privacy
- decomposition into $n^{O(1/\epsilon^2)}$ ϵ-Lipschitz functions.

[Cheraghchi-Klivans-Kothari-Lee '12]

- learning based on Fourier analysis of submodular functions
- submodular fns are ϵ-approximable by polynomials of degree $1/\epsilon^2$
- learning for uniform distributions, using $n^{O(1/\epsilon^2)}$ *random examples*

Related work (cont'd)

[Rashodnikova-Yaroslavtsev'13, Blais-Onak-Servedio-Yaroslavtsev'13]

- learning/testing of *discrete* submodular functions (k possible values), using $k^{O(k \log k/\epsilon)} poly(n)$ value queries
- ϵ-approximation by a junta of size $(k \log \frac{1}{\epsilon})^{\tilde{O}(k)}$

Related work (cont'd)

[Rashodnikova-Yaroslavtsev'13, Blais-Onak-Servedio-Yaroslavtsev'13]

- learning/testing of *discrete* submodular functions (k possible values), using $k^{O(k \log k/\epsilon)} poly(n)$ value queries
- ϵ-approximation by a junta of size $(k \log \frac{1}{\epsilon})^{\tilde{O}(k)}$

[Feldman-Kothari-V. '13]

- ϵ-approximation of submodular functions by *decision trees* of depth $O(1/\epsilon^2)$, and hence *juntas* of size $2^{O(1/\epsilon^2)}$
- PAC-learning using $2^{poly(1/\epsilon)} poly(n)$ random examples (vs. [CKKL'12] $n^{poly(1/\epsilon)}$ examples but in the agnostic setting)

Related work (cont'd)

[Rashodnikova-Yaroslavtsev'13, Blais-Onak-Servedio-Yaroslavtsev'13]

- learning/testing of *discrete* submodular functions (k possible values), using $k^{O(k \log k/\epsilon)} poly(n)$ value queries
- ϵ-approximation by a junta of size $(k \log \frac{1}{\epsilon})^{\tilde{O}(k)}$

[Feldman-Kothari-V. '13]

- ϵ-approximation of submodular functions by *decision trees* of depth $O(1/\epsilon^2)$, and hence *juntas* of size $2^{O(1/\epsilon^2)}$
- PAC-learning using $2^{poly(1/\epsilon)} poly(n)$ random examples (vs. [CKKL'12] $n^{poly(1/\epsilon)}$ examples but in the agnostic setting)

QUESTIONS:

- Why is this restricted to submodular functions?
- Is the junta of size $2^{O(1/\epsilon^2)}$ related to Friedgut's Theorem?
- What are the best juntas that we can achieve?

Our results
[to appear in FOCS'13]

Result 1:

- XOS and self-bounding functions can be ϵ-approximated in L_2 by juntas of size $2^{O(1/\epsilon^2)}$
- This follows from a "real-valued Friedgut's theorem"

Our results
[to appear in FOCS'13]

Result 1:

- XOS and self-bounding functions can be ϵ-approximated in L_2 by juntas of size $2^{O(1/\epsilon^2)}$
- This follows from a "real-valued Friedgut's theorem"

Result 2:

- Submodular fns can be ϵ-approximated by $O(\frac{1}{\epsilon^2}\log\frac{1}{\epsilon})$-juntas
- Proof avoids Fourier analysis, uses concentration properties + "boosting lemma" from [Goemans-V. '04]

Our results
[to appear in FOCS'13]

Result 1:

- XOS and self-bounding functions can be ϵ-approximated in L_2 by juntas of size $2^{O(1/\epsilon^2)}$
- This follows from a "real-valued Friedgut's theorem"

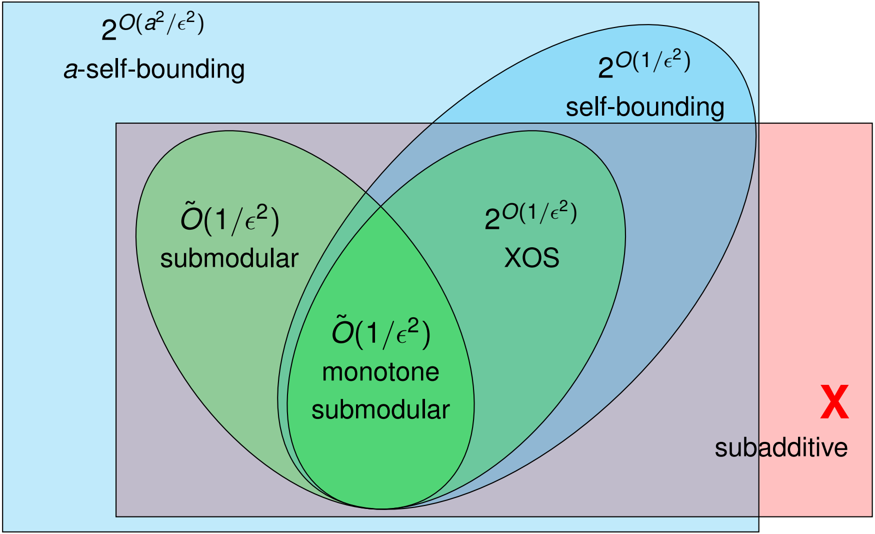
Result 2:

- Submodular fns can be ϵ-approximated by $O(\frac{1}{\epsilon^2}\log\frac{1}{\epsilon})$-juntas
- Proof avoids Fourier analysis, uses concentration properties + "boosting lemma" from [Goemans-V. '04]

Applications to learning:

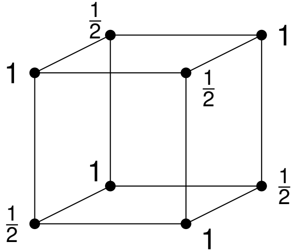
- Submodular, XOS and monotone self-bounding functions can be PAC-learned in time $2^{poly(1/\epsilon)} poly(n)$ within L_2-error ϵ
- Submodular functions can be "PMAC"-learned in time $2^{poly(1/\epsilon)} poly(n)$ within *multiplicative error* $1+\epsilon$

Overview of our junta approximations



No junta approximation for subadditive functions

Example: any function $f : \{0,1\}^n \to \{\frac{1}{2}, 1\}$ is subadditive.



$$\forall A, B;\ f(A \cup B) \leq 1 \leq f(A) + f(B)$$

Therefore, we can encode any function whatsoever,
e.g. a parity function, which cannot be approximated by a junta.

Plan

Remaining plan of the talk:

1. Friedgut's theorem for real-valued functions
2. $\Rightarrow$ junta approximations for XOS and self-bounding functions
3. Improved junta approximation for submodular functions
4. Conclusions

Friedgut's Theorem

Friedgut's Theorem: for boolean functions $f : \{0,1\}^n \to \{0,1\}$
average sensitivity $\mathrm{Infl}(f) \Rightarrow \epsilon$*-approx. by a junta of size* $2^{O(\mathrm{Infl}(f)/\epsilon)}$

$\mathrm{Infl}(f) = \sum_{i=1}^{n} \Pr_{x\in\{0,1\}^n}[f(x) \neq f(x \oplus e_i)] = \sum_{S\subseteq[n]} |S|\hat{f}^2(S)$

What should it say for real-valued functions?

Friedgut's Theorem

Friedgut's Theorem: for boolean functions $f : \{0,1\}^n \to \{0,1\}$
average sensitivity $\mathrm{Infl}(f) \Rightarrow \epsilon$*-approx. by a junta of size* $2^{O(\mathrm{Infl}(f)/\epsilon)}$

$\mathrm{Infl}(f) = \sum_{i=1}^n \Pr_{x\in\{0,1\}^n}[f(x) \neq f(x \oplus e_i)] = \sum_{S\subseteq[n]} |S|\hat{f}^2(S)$

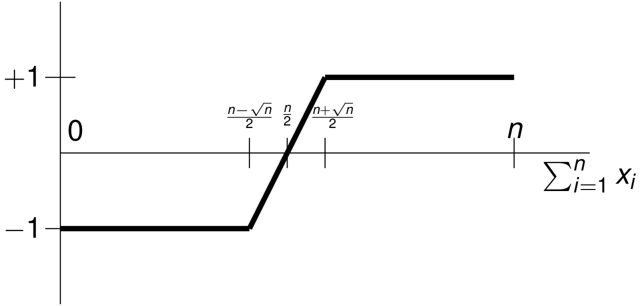
What should it say for real-valued functions?

Natural extension of average sensitivity:
$\mathrm{Infl}^2(f) = \sum_{S\subseteq[n]} |S|\hat{f}^2(S) = \sum_{i=1}^n \mathbb{E}_{x\in\{0,1\}^n}[(f(x) - f(x \oplus e_i))^2]$

But Friedgut's Theorem for this notion of average sensitivity is FALSE!
as observed by [O'Donnell-Servedio '07]

Counterexample to Friedgut's Theorem?

Counterexample for $f : \{0,1\}^n \to [-1,1]$: (from [O'Donnell-Servedio '07])



- $\forall x, i; |f(x) - f(x \oplus e_i)| \leq \frac{1}{\sqrt{n}}$
- $\mathrm{Infl}^2(f) = \sum_{i=1}^{n} \mathbb{E}[(f(x) - f(x \oplus e_i)^2)] = O(1)$
- so there should be an ϵ-approximate junta of size $2^{O(1/\epsilon)}$?
- but we need $\Omega(n)$ variables to approximate within a constant ϵ

How to fix Friedgut's Theorem?

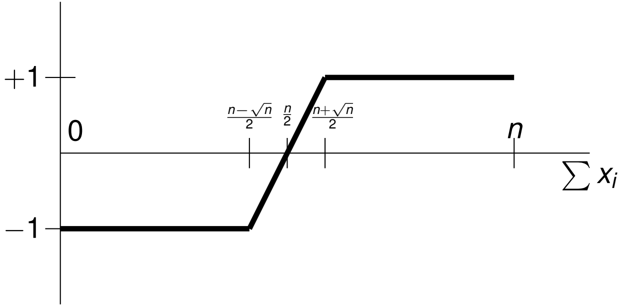
[O'Donnell-Servedio '07] prove a variant of Friedgut's theorem for discretized functions $f : \{0,1\}^n \to [-1,1] \cap \delta\mathbb{Z}$.

We don't know how to discretize while preserving submodularity etc.

How to fix Friedgut's Theorem?

[O'Donnell-Servedio '07] prove a variant of Friedgut's theorem for discretized functions $f : \{0,1\}^n \to [-1,1] \cap \delta\mathbb{Z}$.

We don't know how to discretize while preserving submodularity etc.



Note: If we define $\mathrm{Infl}^{\kappa}(f) = \sum_{i=1}^{n} \mathbb{E}[|f(x) - f(x \oplus e_i)|^{\kappa}]$, then

$$\mathrm{Infl}^{1}(f) = n \cdot \Theta(1/\sqrt{n}) = \Theta(\sqrt{n}).$$

Friedgut's Theorem for real-valued functions

Theorem

Let $f : \{0,1\}^n \to \mathbb{R}$. *Then there exists a polynomial g of degree* $O(\mathrm{Infl}^2(f)/\epsilon^2)$ *depending on* $2^{O(\mathrm{Infl}^2(f)/\epsilon^2)} \mathit{poly}(\mathrm{Infl}^1(f)/\epsilon)$ *variables such that* $\|f - g\|_2 \leq \epsilon$.

Friedgut's Theorem for real-valued functions

Theorem

Let $f : \{0,1\}^n \to \mathbb{R}$. Then there exists a polynomial g of degree $O(\mathrm{Infl}^2(f)/\epsilon^2)$ depending on $2^{O(\mathrm{Infl}^2(f)/\epsilon^2)} poly(\mathrm{Infl}^1(f)/\epsilon)$ variables such that $\|f - g\|_2 \leq \epsilon$.

Notes:

- we could replace $\mathrm{Infl}^1(f)$ by $\mathrm{Infl}^\kappa(f)$ for $\kappa < 2$, but not $\mathrm{Infl}^2(f)$
- for boolean functions, $\mathrm{Infl}^1(f) = \mathrm{Infl}^2(f)$, so it doesn't matter
- we will show that for submodular, XOS and self-bounding functions, $\mathrm{Infl}^1(f)$ and $\mathrm{Infl}^2(f)$ are small

Proof of Real-valued Friedgut

Follow Friedgut's proof:
Fourier analysis, hypercontractive inequality...

Let

- $d = 2\mathrm{Infl}^2(f)/\epsilon^2$
- $\alpha = (\epsilon^2(\kappa - 1)^d / \mathrm{Infl}^\kappa(f))^{\kappa/(2-\kappa)}$, $\kappa = 4/3$
- $J = \{i \in [n] : \mathrm{Infl}_i^\kappa(f) \geq \alpha\}$
- $\mathcal{J} = \{S \subseteq J, |S| \leq d\}$

Goal: $\sum_{S \notin \mathcal{J}} \hat{f}^2(S) \leq \epsilon^2$.
Then, $g(x) = \sum_{S \in \mathcal{J}} \hat{f}(S)\chi_S(x)$ is an ϵ-approximation to f.

Proof cont'd

The bound on $\sum_{S \notin \mathcal{J}} \hat{f}^2(S)$ has two parts:

1. $\sum_{|S|>d} \hat{f}^2(S) \leq \frac{1}{d} \sum |S| \hat{f}^2(S) = \frac{1}{d} \mathrm{Infl}^2(f) \leq \epsilon^2/2$
   — by the definition of d

Proof cont'd

The bound on $\sum_{S \notin \mathcal{J}} \hat{f}^2(S)$ has two parts:

1. $\sum_{|S|>d} \hat{f}^2(S) \leq \frac{1}{d} \sum |S| \hat{f}^2(S) = \frac{1}{d} \mathrm{Infl}^2(f) \leq \epsilon^2/2$
— by the definition of d
2. $\sum_{S \not\subseteq J, |S| \leq d} \hat{f}^2(S) \leq \sum_{i \notin J} \sum_{|S| \leq d, i \in S} \hat{f}^2(S)$
$\leq (\kappa - 1)^{1-d} \sum_{i \notin J} \| T_{\sqrt{\kappa-1}}(\partial_i f) \|_2^2$
— this part requires the hypercontractive inequality:

$$\| T_{\sqrt{\kappa-1}}(f) \|_2 \leq \| f \|_\kappa$$

where T_ρ is the noise operator ($\widehat{T_\rho f}(S) = \rho^{|S|} \hat{f}(S)$).

Proof cont'd

The bound on $\sum_{S \notin \mathcal{J}} \hat{f}^2(S)$ has two parts:

1. $\sum_{|S|>d} \hat{f}^2(S) \leq \frac{1}{d} \sum |S| \hat{f}^2(S) = \frac{1}{d} \mathrm{Infl}^2(f) \leq \epsilon^2/2$
— by the definition of d
2. $\sum_{S \not\subseteq J, |S| \leq d} \hat{f}^2(S) \leq \sum_{i \notin J} \sum_{|S| \leq d, i \in S} \hat{f}^2(S)$
$\leq (\kappa - 1)^{1-d} \sum_{i \notin J} \| T_{\sqrt{\kappa-1}}(\partial_i f) \|_2^2$
— this part requires the hypercontractive inequality:

$$\| T_{\sqrt{\kappa-1}}(f) \|_2 \leq \| f \|_\kappa$$

where T_ρ is the noise operator ($\widehat{T_\rho f}(S) = \rho^{|S|} \hat{f}(S)$).

The difference: we need a bound on $\sum_{i \notin J} \| \partial_i f \|_\kappa^2 = \sum_{i \notin J} (\mathrm{Infl}_i^\kappa(f))^{2/\kappa}$, which does not follow from $\mathrm{Infl}^2(f)$ for real-valued functions.

Finishing the proof: $\sum_{i \notin J} (\mathrm{Infl}_i^\kappa(f))^{2/\kappa} \leq \alpha^{2/\kappa - 1} \sum \mathrm{Infl}_i^\kappa(f) \leq (\kappa - 1)^d \epsilon^2$.

Application to self-bounding functions

Recall: f is self-bounding if

$$\sum_{i=1}^{n}(f(x) - min_{x_i} f(x)) \leq f(x).$$

Application to self-bounding functions

Recall: f is self-bounding if

$$\sum_{i=1}^{n}(f(x) - min_{x_i} f(x)) \leq f(x).$$

By double counting,

$$\mathrm{Infl}^1(f) = \sum_{i=1}^{n} \mathbb{E}[|f(x) - f(x \oplus e_i)|] = 2\sum_{i=1}^{n} \mathbb{E}[f(x) - \min_{x_i} f(x)] \leq 2\mathbb{E}[f(x)].$$

For $f : \{0,1\}^n \to [0,1]$, we get $\mathrm{Infl}^2(f) \leq \mathrm{Infl}^1(f) = O(1)$.

Application to self-bounding functions

Recall: f is self-bounding if

$$\sum_{i=1}^{n}(f(x) - min_{x_i} f(x)) \leq f(x).$$

By double counting,

$$\text{Infl}^1(f) = \sum_{i=1}^{n} \mathbb{E}[|f(x) - f(x \oplus e_i)|] = 2\sum_{i=1}^{n} \mathbb{E}[f(x) - \min_{x_i} f(x)] \leq 2\mathbb{E}[f(x)].$$

For $f : \{0,1\}^n \to [0,1]$, we get $\text{Infl}^2(f) \leq \text{Infl}^1(f) = O(1)$.

Corollary (of real-valued Friedgut)

For any self-bounding (or submodular or XOS) function $f : \{0,1\}^n \to [0,1]$ and $\epsilon > 0$, there is a polynomial g of degree $d = O(1/\epsilon^2)$ over $2^{O(d)}$ variables such that $\|f - g\|_2 \leq \epsilon$.

Lower bound for XOS functions

Friedgut's Theorem: it is known that $2^{\Omega(1/\epsilon)}$ variables are necessary.

Example: *tribes function* $\rightarrow$ lower bound for XOS functions as well.

$$f(x) = \max\left\{ \frac{1}{|A_1|}\sum_{i\in A_1} x_i, \frac{1}{|A_2|}\sum_{i\in A_2} x_i, \ldots, \frac{1}{|A_b|}\sum_{i\in A_b} x_i \right\}.$$

- $b = 2^{1/\epsilon}$ disjoint blocks A_j of size $|A_j| = 1/\epsilon$
- any junta smaller than $2^{1/\epsilon-1}$ misses $2^{1/\epsilon-1}$ blocks
- and cannot approximate f within ϵ

Better juntas for submodular functions

Theorem

For every submodular function $f : \{0,1\}^n \to [0,1]$ *and* $\epsilon > 0$*, there is a function* g *depending on* $O(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon})$ *variables, such that* $\|f - g\|_2 \leq \epsilon$*.*

Better juntas for submodular functions

Theorem

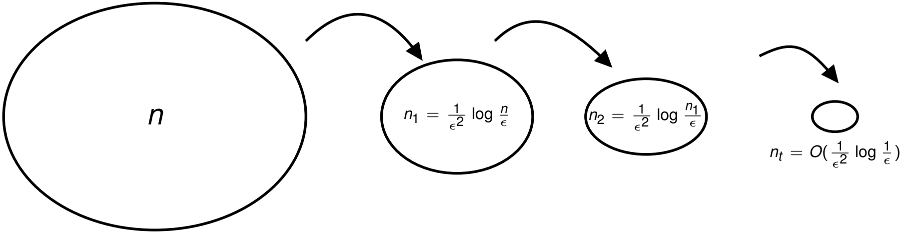
For every submodular function $f : \{0,1\}^n \to [0,1]$ and $\epsilon > 0$, there is a function g depending on $O(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon})$ variables, such that $\|f - g\|_2 \leq \epsilon$.

Notes:

- this is tight up to the log factor; consider $f(x) = \epsilon^2 \sum_{i=1}^{1/\epsilon^2} x_i$
- in this sense, submodular functions are close to linear functions, while XOS/self-bounding functions are "more complicated"

About the proof

Inductive step: submodular function f of n variables
$\longrightarrow$ a function of $O(\frac{1}{\epsilon^2}\log\frac{n}{\epsilon})$ variables, approximating f within $\frac{1}{2}\epsilon$



- the process stops when $n_t = O(\frac{1}{\epsilon^2}\log\frac{1}{\epsilon})$
- errors form a geometric series, converging to ϵ

How to reduce n to $|J| = \frac{1}{\epsilon^2} \log \frac{n}{\epsilon}$

For simplicity: assume $f : \{0,1\}^n \to [0,1]$ monotone submodular.

How to reduce n to $|J| = \frac{1}{\epsilon^2} \log \frac{n}{\epsilon}$

For simplicity: assume $f : \{0,1\}^n \to [0,1]$ monotone submodular.

Idea: build J by including variables $x_{i_1}, x_{i_2}, x_{i_3}, \ldots$ that contribute significantly to the current set:

$$\mathbb{E}_{x \in \{0,1\}^J}[\partial_i f(x)] = \mathbb{E}_{x \in \{0,1\}^J}[f(x \oplus e_i) - f(x)] > \alpha.$$

Hopefully:

1. we cannot include too many variables, because the function is bounded by $[0,1]$
2. f is α-Lipschitz in the variables that are not selected, and hence we can use concentration to argue that they can be ignored

How to reduce n to $|J| = \frac{1}{\epsilon^2} \log \frac{n}{\epsilon}$

For simplicity: assume $f : \{0,1\}^n \to [0,1]$ monotone submodular.

Idea: build J by including variables $x_{i_1}, x_{i_2}, x_{i_3}, \ldots$ that contribute significantly to the current set:

$$\mathbb{E}_{x \in \{0,1\}^J}[\partial_i f(x)] = \mathbb{E}_{x \in \{0,1\}^J}[f(x \oplus e_i) - f(x)] > \alpha.$$

Hopefully:

1. we cannot include too many variables, because the function is bounded by $[0,1]$
2. f is α-Lipschitz in the variables that are not selected, and hence we can use concentration to argue that they can be ignored

BUT: f is α-Lipschitz in the remaining variables only "in expectation"; we need a statement for most points in $\{0,1\}^J$ and for all $j \notin J$

The Boosting Lemma

Lemma (Goemans,V. 2004)

Let $\mathcal{F} \subseteq \{0,1\}^n$ be down-closed ($x \leq y \in \mathcal{F} \Rightarrow x \in \mathcal{F}$) and

$$\sigma(p) = \Pr_{x \sim \mu_p}[x \in \mathcal{F}] = \sum_{F \in \mathcal{F}} p^{|F|}(1-p)^{n-|F|}.$$

Then $\sigma(p) = (1-p)^{\phi(p)}$ where $\phi(p)$ is a non-decreasing function of p.

The Boosting Lemma

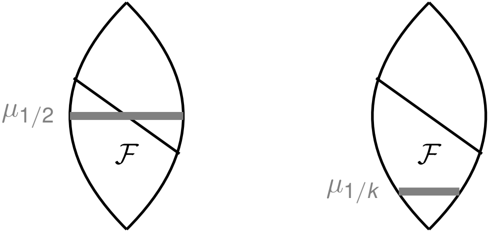
Lemma (Goemans,V. 2004)

Let $\mathcal{F} \subseteq \{0,1\}^n$ *be down-closed* ($x \leq y \in \mathcal{F} \Rightarrow x \in \mathcal{F}$) *and*

$$\sigma(p) = \Pr_{x \sim \mu_p}[x \in \mathcal{F}] = \sum_{F \in \mathcal{F}} p^{|F|}(1-p)^{n-|F|}.$$

Then $\sigma(p) = (1-p)^{\phi(p)}$ *where* $\phi(p)$ *is a non-decreasing function of* p.

Example:



$$\sigma(\tfrac{1}{2}) \geq \frac{1}{2^k} \Longrightarrow \sigma(\tfrac{1}{k}) \geq (1-\tfrac{1}{k})^k \simeq \frac{1}{e}$$

How we find the small junta

Algorithm: (for f monotone submodular)

- Initialize $J := \emptyset$, $\alpha \simeq \epsilon^2$, $\delta \simeq 1/\log \frac{n}{\epsilon}$.
- Let $J(\delta) =$ each element of J independently with prob. δ.
- As long as $\exists i \notin J$ such that

$$\Pr[\partial_i f(\mathbf{1}_{J(\delta)}) > \alpha] > 1/e,$$

include i in J.

Algorithm: (for f monotone submodular)

- Initialize $J := \emptyset$, $\alpha \simeq \epsilon^2$, $\delta \simeq 1/\log \frac{n}{\epsilon}$.
- Let $J(\delta) =$ each element of J independently with prob. δ.
- As long as $\exists i \notin J$ such that

$$\Pr[\partial_i f(\mathbf{1}_{J(\delta)}) > \alpha] > 1/e,$$

  include i in J.

Using the boosting lemma:
If we did not include i in the final set J, then $\Pr_{x \sim J(\delta)}[\partial_i f(x) > \alpha] \leq 1/e$, and hence $\Pr_{x \sim J(1/2)}[\partial_i f(x) > \alpha] \leq (1/2)^{1/\delta} \simeq \epsilon/n$.

How we find the small junta

Algorithm: (for f monotone submodular)

- Initialize $J := \emptyset$, $\alpha \simeq \epsilon^2$, $\delta \simeq 1/\log\frac{n}{\epsilon}$.
- Let $J(\delta) =$ each element of J independently with prob. δ.
- As long as $\exists i \notin J$ such that

$$\Pr[\partial_i f(\mathbf{1}_{J(\delta)}) > \alpha] > 1/e,$$

include i in J.

Using the boosting lemma:
If we did not include i in the final set J, then $\Pr_{x\sim J(\delta)}[\partial_i f(x) > \alpha] \leq 1/e$, and hence $\Pr_{x\sim J(1/2)}[\partial_i f(x) > \alpha] \leq (1/2)^{1/\delta} \simeq \epsilon/n$.
Union bound $\Rightarrow \Pr_{x\sim J(1/2)}[\exists i \notin J; \partial_i f(x) > \alpha] \leq \epsilon$.

Finishing the proof

Accuracy of the junta:

- We found a set J such that with probability $1-\epsilon$ over $x \in \{0,1\}^J$, the function $g_x(y) = f(x,y)$ is ϵ^2-Lipschitz in y
- By concentration, g_x is ϵ-approximated by its expectation.
- Hence, f is 2ϵ-approximated by its averaging-projection on $\{0,1\}^J$.

Finishing the proof

Accuracy of the junta:

- We found a set J such that with probability $1-\epsilon$ over $x \in \{0,1\}^J$, the function $g_x(y) = f(x,y)$ is ϵ^2-Lipschitz in y
- By concentration, g_x is ϵ-approximated by its expectation.
- Hence, f is 2ϵ-approximated by its averaging-projection on $\{0,1\}^J$.

Size of the junta:

- every time we include $i \in J$, we have $\mathbb{E}_{x \sim J(\delta)}[\partial_i f(x)] > \alpha/e$
- so we increase $\mathbb{E}[f(J(\delta))]$ by $\alpha\delta/e$
- this can repeat at most $O(\frac{1}{\alpha\delta}) = O(\frac{1}{\epsilon^2} \log \frac{n}{\epsilon})$ times.

Concluding comments and questions

- Self-bounding functions are ϵ-approximated by $2^{O(1/\epsilon^2)}$-juntas
- Submodular functions are ϵ-approximated by $\tilde{O}(1/\epsilon^2)$-juntas
- We also have a $(1+\epsilon)$*-multiplicative approximation except for* ϵ*-fraction of* $\{0,1\}^n$, for monotone submodular functions, by a junta of size $\tilde{O}(1/\epsilon^2)$.
- We don't know if such a junta exists for non-monotone submodular functions

More on our learning algorithms and results: Vitaly Feldman on Oct 30.