Submodular Functions are Noise Stable

Mahdi Cheraghchi
CMU

Adam Klivans
UT Austin

Pravesh Kothari
UT Austin

Homin K. Lee
UT Austin

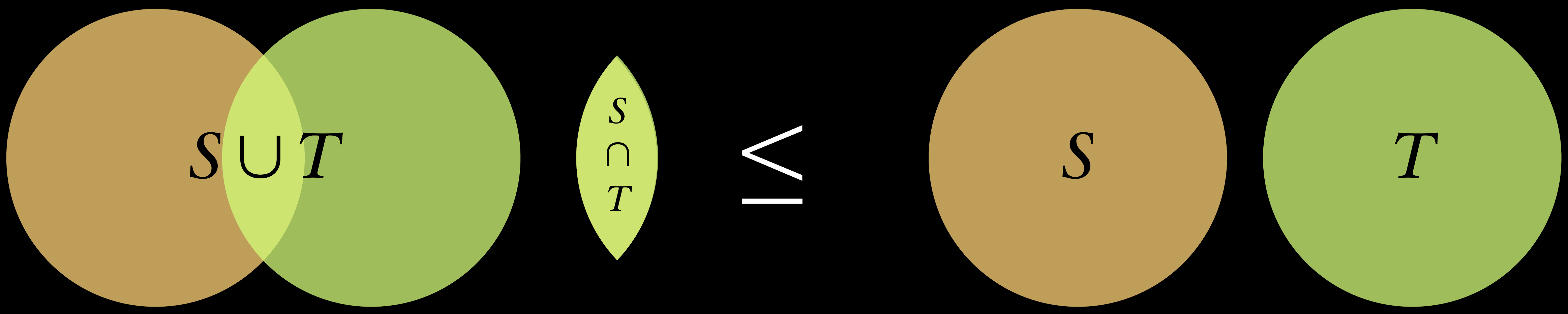
Overview

- Submodular Functions Are Noise Stable
- Application 1: Agnostic Learning
- Application 2: Differential Privacy

Submodular Functions

A function $f : 2^{[n]} \rightarrow R$ is *submodular* if $\forall S,T \subseteq [n]$:

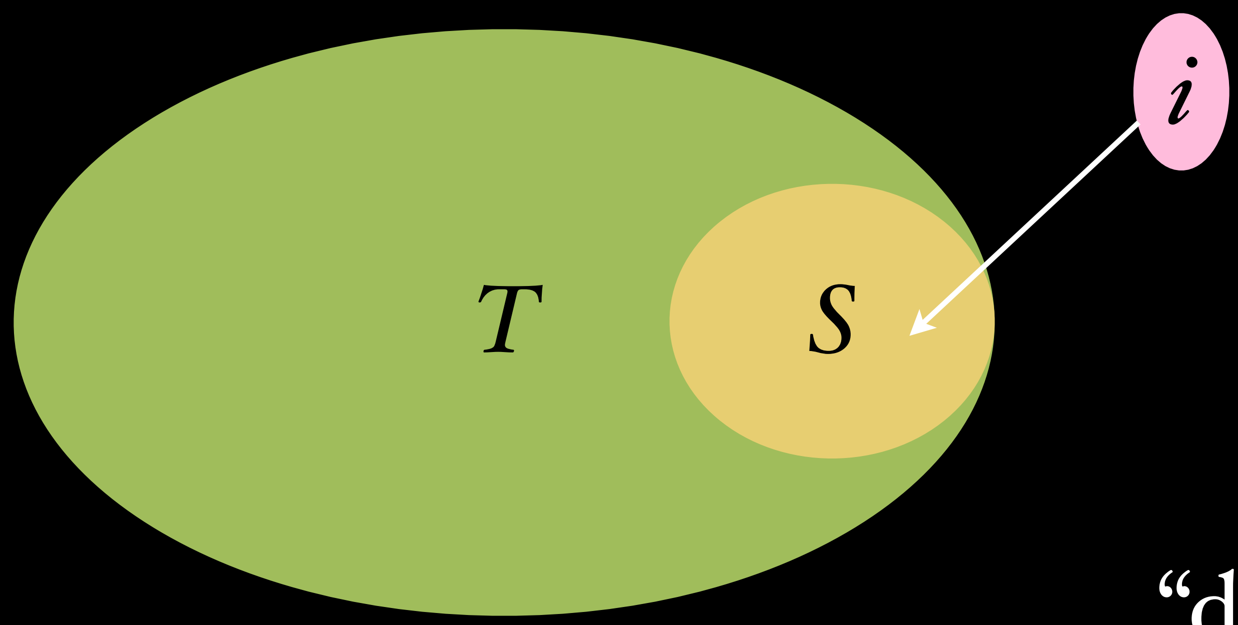
$$f(S \cup T) + f(S \cap T) \leq f(S) + f(T)$$



Submodular Functions

Equivalently, $\forall S \subseteq T \subseteq [n],\ i \in [n] \setminus T$:

$$f(T \cup \{i\}) - f(T) \ \leq\ f(S \cup \{i\}) - f(S)$$



"decreasing marginal returns"

Submodular Functions

$$f(T \cup \{i\}) - f(T) \leq f(S \cup \{i\}) - f(S)$$

"decreasing marginal returns"

Submodular Functions

$$f(T \cup \{i\}) - f(T) \leq f(S \cup \{i\}) - f(S)$$

$+$

"decreasing marginal returns"

Submodular Functions

$$f(T \cup \{i\}) - f(T) \leq f(S \cup \{i\}) - f(S)$$

$+$

"decreasing marginal returns"

Submodular Functions

$$f(T \cup \{i\}) - f(T) \leq f(S \cup \{i\}) - f(S)$$

 +  

"decreasing marginal returns"

Submodular Functions

$$f(T \cup \{i\}) - f(T) \leq f(S \cup \{i\}) - f(S)$$

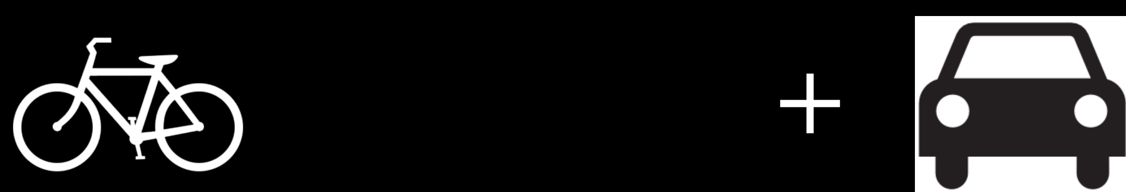
 +  

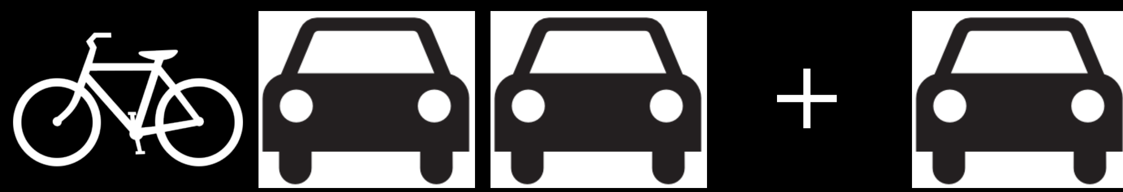
 + 

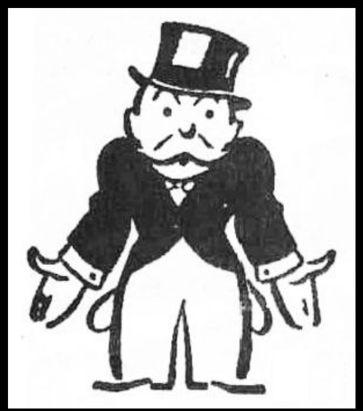
"decreasing marginal returns"

Submodular Functions

$$f(T \cup \{i\}) - f(T) \leq f(S \cup \{i\}) - f(S)$$







“decreasing marginal returns”

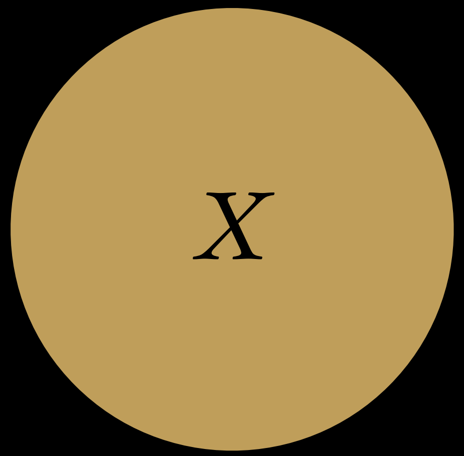
Submodular Functions

- Submodular functions are used to model:
  - bundle pricing,
  - bidders' valuations in combinatorial auctions,
  - welfare maximation.
- Submodular funcs can be used to train SVMs.
- Rank functions of matroids are submodular:
  - cut functions of graphs
  - dimension of a vector space

Noise Stability

$f: 2^{[n]} \to R$

D, a product distribution over $2^{[n]}$ (uniform)



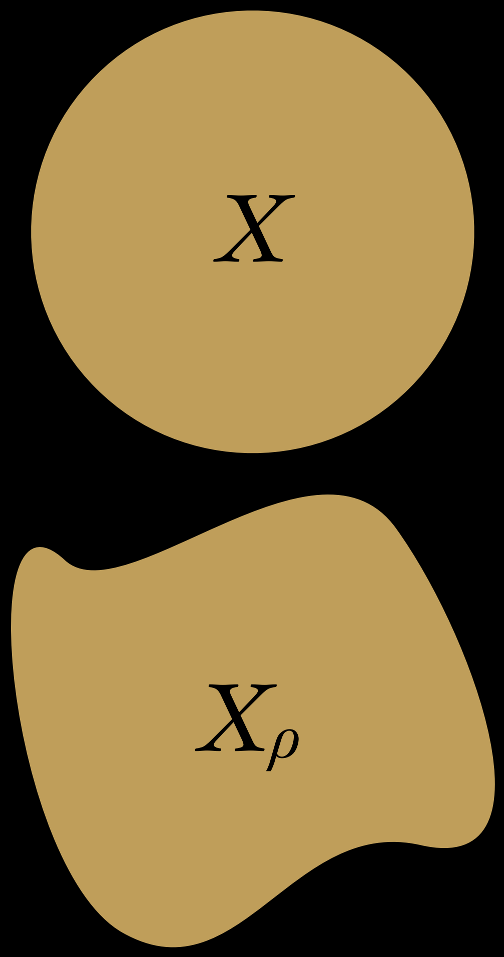
Noise Stability

$f: 2^{[n]} \to R$

D, a product distribution over $2^{[n]}$ (uniform)

X_ρ noisy version of X

(each i resampled with prob. 1-ρ)



Noise Stability

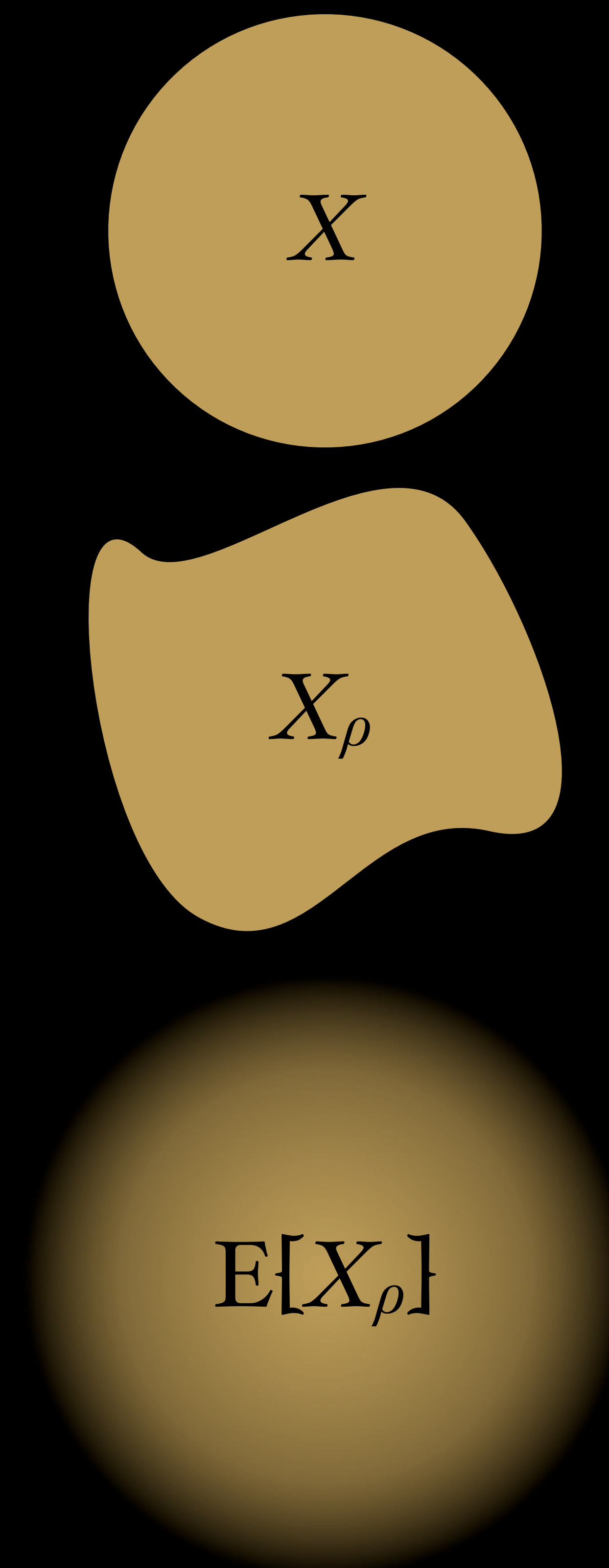
$f: 2^{[n]} \to R$

D, a product distribution over $2^{[n]}$ (uniform)

X_ρ noisy version of X

(each i resampled with prob. $1\text{-}\rho$)

$T_\rho f(X) = \mathrm{E}[f(X_\rho)]$ (noise operator)



Noise Stability

$f: 2^{[n]} \to R$

D, a product distribution over $2^{[n]}$ (uniform)

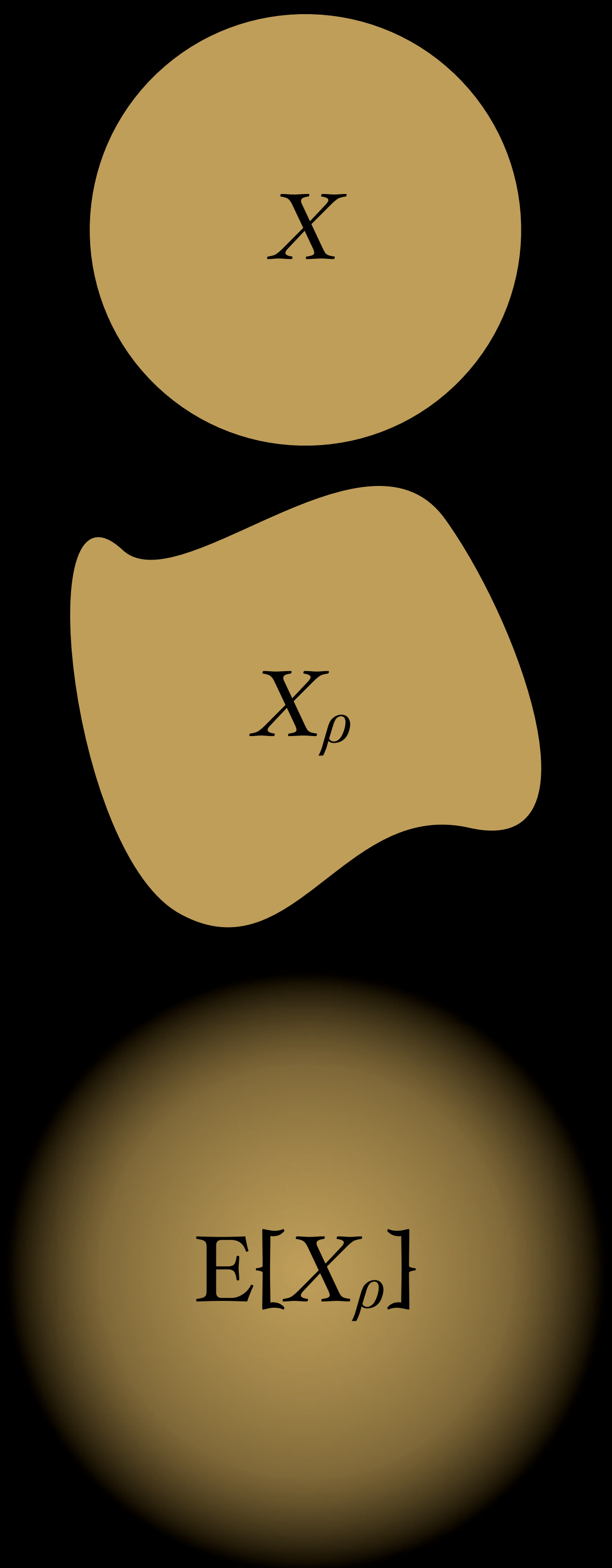
X_ρ noisy version of X

(each i resampled with prob. 1-ρ)

$T_\rho f(X) = \mathrm{E}[f(X_\rho)]$ (noise operator)

The ***noise stability*** of f at noise rate ρ

$S_\rho(f) = \mathrm{E}[f \cdot T_\rho f]$



Noise Stability

$$T_\rho f(X) = \mathrm{E}[f(X_\rho)]$$

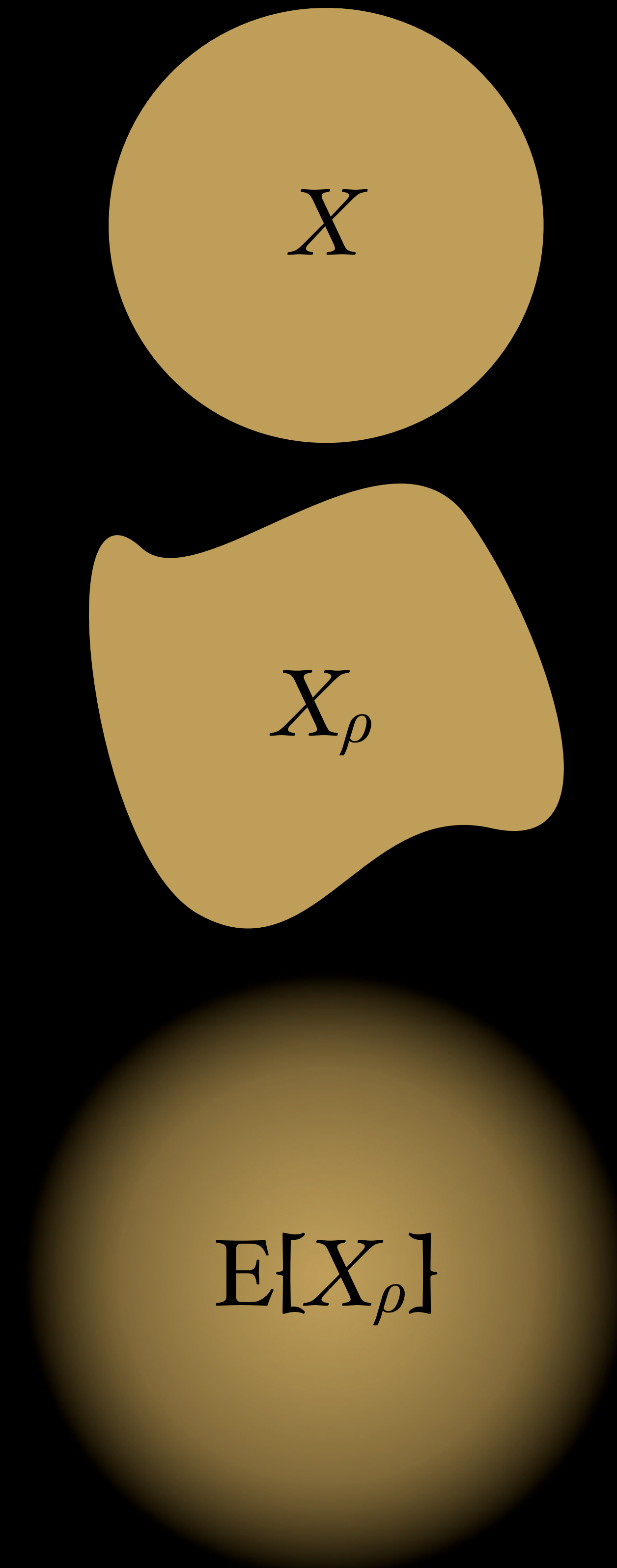
$$S_\rho(f) = \mathrm{E}[f \cdot T_\rho f]$$

Constant Functions:

$$f = c$$

$$T_\rho f(X) = c$$

$$S_\rho(f) = c^2$$



Noise Stability

$T_\rho f(X) = \mathrm{E}[f(X_\rho)]$

$S_\rho(f) = \mathrm{E}[f \cdot T_\rho f]$

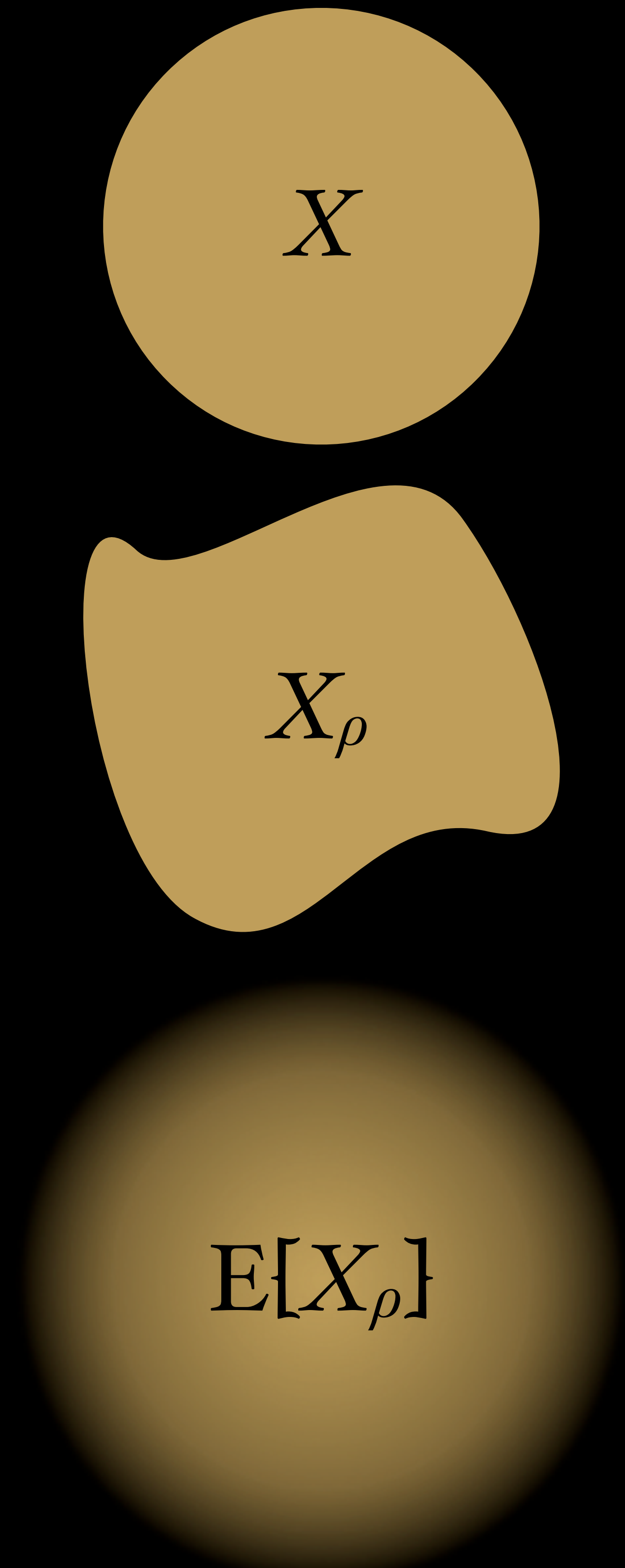
Dictator Functions:

$X_i = 1$ if $i \in X$, -1 o.w.

$f = X_i$

$T_\rho f(X) = \rho \cdot f(X)$

$S_\rho(f) = \rho \cdot \mathrm{E}[f^2] = \rho$



Noise Stability

$$T_\rho f(X) = \mathrm{E}[f(X_\rho)]$$

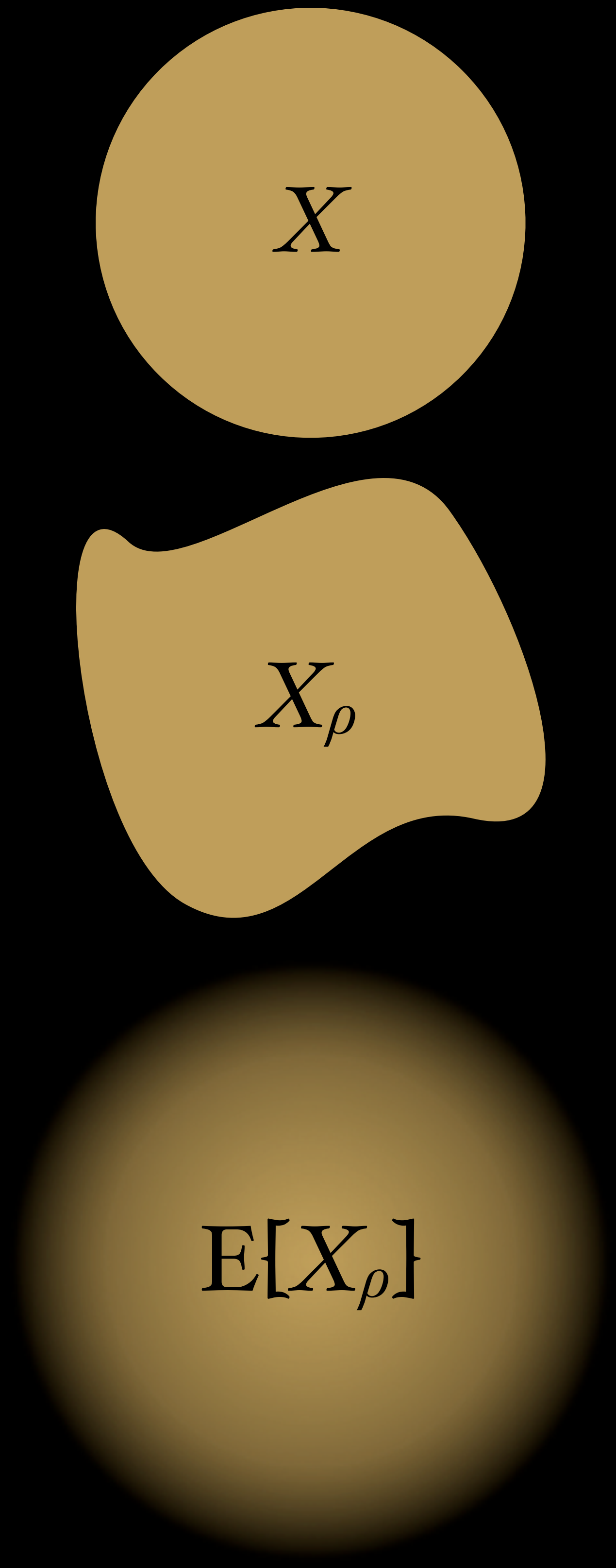
$$S_\rho(f) = \mathrm{E}[f \cdot T_\rho f]$$

Submodular Functions:

f non-negative

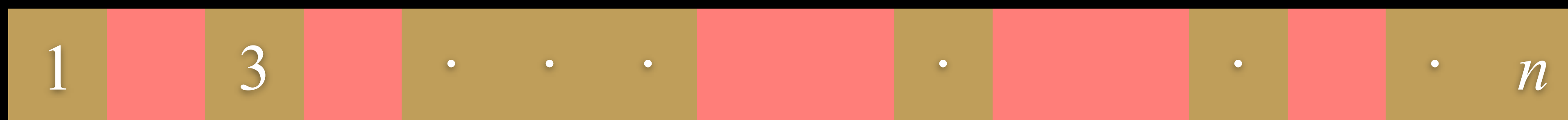
$T_\rho f(X) \geq \rho \cdot f(X)$

$S_\rho(f) \geq \rho \cdot \mathrm{E}[f^2]$



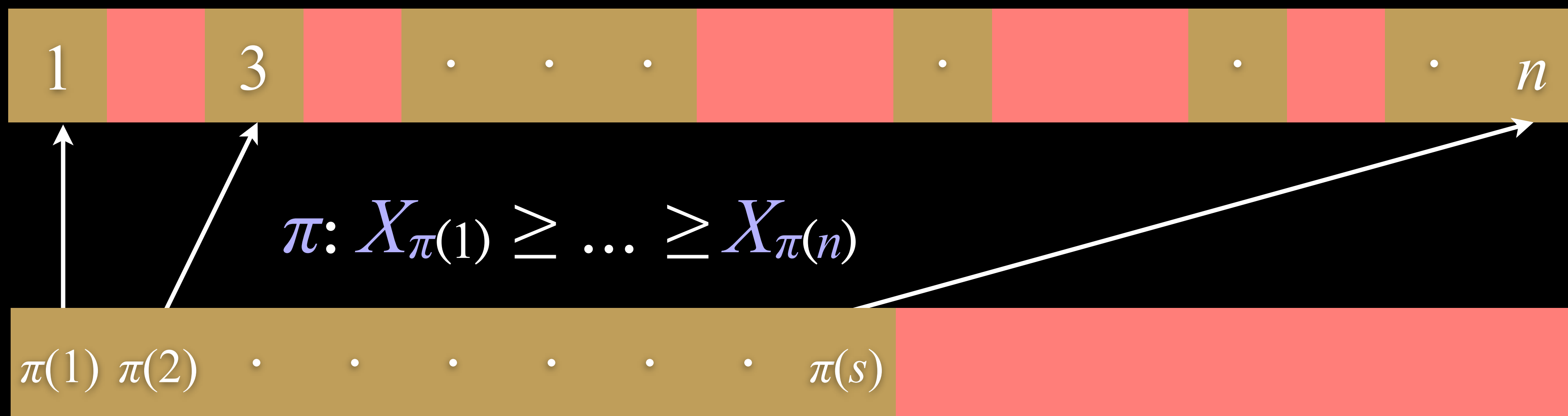
Noise Stability of Submodular Funcs

$f(X) = f(+ - + - + + - + -)$



Noise Stability of Submodular Funcs

$f(X) = f(+ - + - + + - + -)$

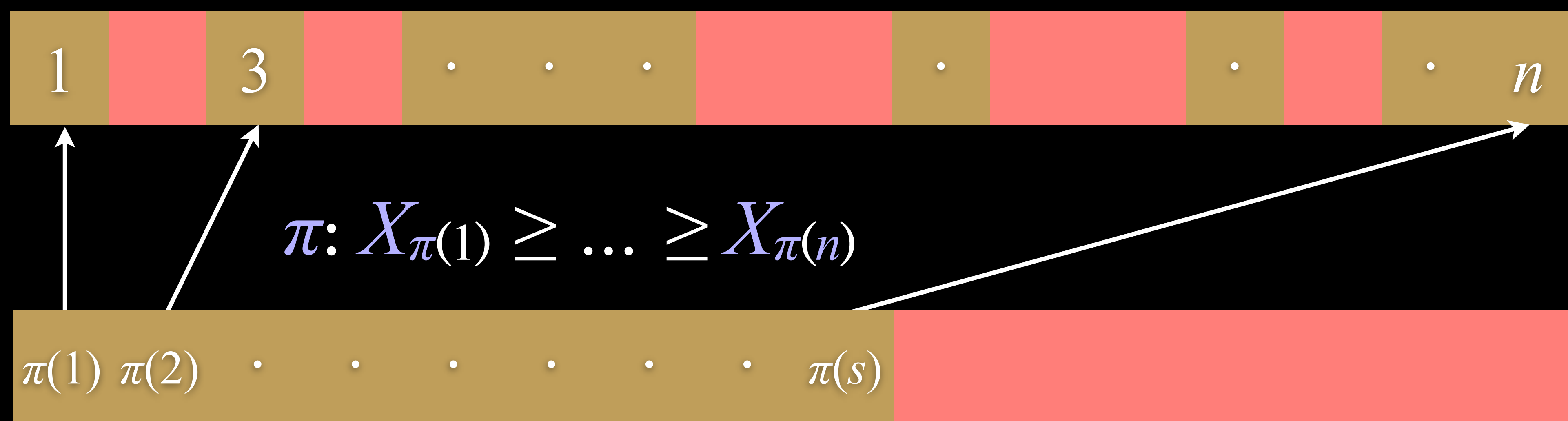


$X_{[j]} = \{\pi(1), \ldots, \pi(j)\}$

$X_{[0]} = \varnothing, X_{[n]} = [n]$

Noise Stability of Submodular Funcs

$f(X) = f(+ - + - + + - + -)$



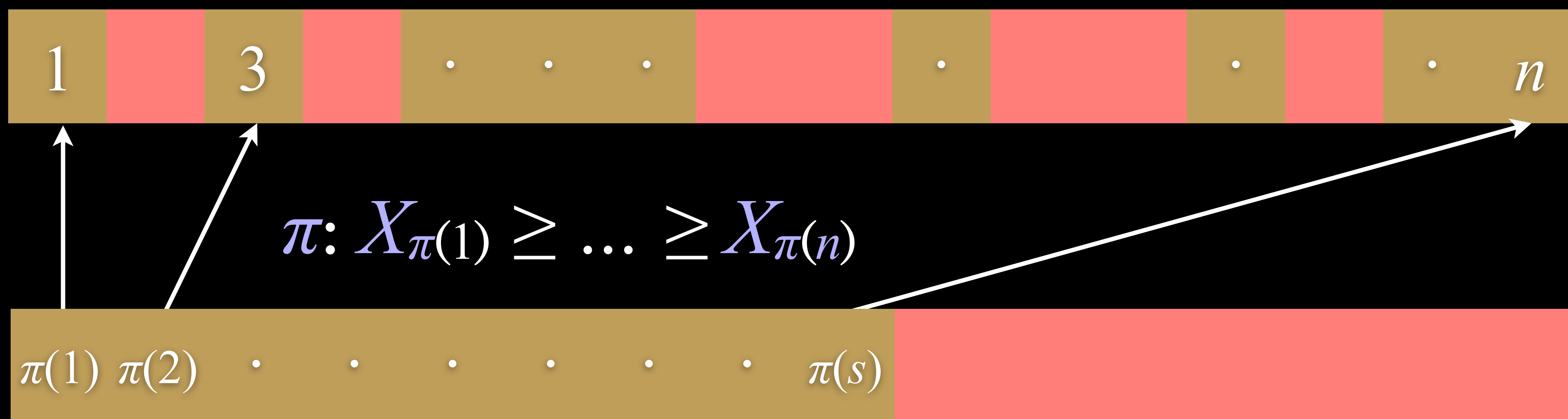
$X_{[j]} = \{\pi(1), \ldots, \pi(j)\}$ $\quad$ $X_{[0]} = \emptyset, X_{[n]} = [n]$

$$T_\rho f(X) = \mathrm{E}[f(X_\rho)]$$

$$= f(X_{[0]}) + \mathrm{E}\, \Sigma f(X_\rho \cap X_{[j]}) - f(X_\rho \cap X_{[j-1]})$$

Noise Stability of Submodular Funcs

$$f(X) = f(+ - + - + + - + -)$$



$$X_{[j]} = \{\pi(1), \ldots, \pi(j)\} \qquad X_{[0]} = \varnothing,\ X_{[n]} = [n]$$

$$T_\rho f(X) = \mathrm{E}[f(X_\rho)]$$

$$= f(X_{[0]}) + \mathrm{E}\ \boxed{\Sigma f(X_\rho \cap X_{[j]}) - f(X_\rho \cap X_{[j-1]})}$$

Noise Stability of Submodular Funcs

$\pi(1)\ \pi(2)\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \pi(s)$

$$T_\rho f(X) = f(X_{[0]}) + \mathrm{E}\,\Sigma\, f(X_\rho \cap X_{[j]}) - f(X_\rho \cap X_{[j\text{-}1]})$$

Noise Stability of Submodular Funcs

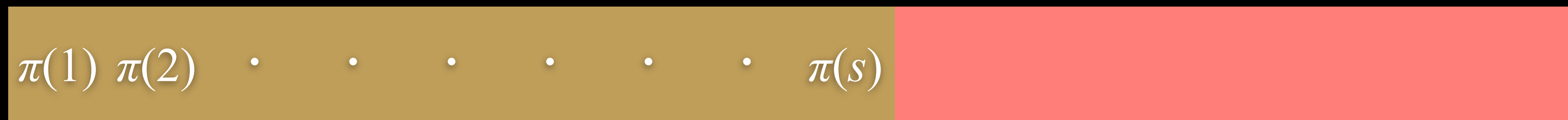
$\pi(1)$	$\pi(2)$	·	·	·	·	·	·	$\pi(s)$	

$$T_\rho f(X) = f(X_{[0]}) + \mathrm{E} \sum f(X_\rho \cap X_{[j]}) - f(X_\rho \cap X_{[j-1]})$$

submodularity

$$f(S) - f(S \cap T) \geq f(S \cup T) - f(T)$$

Noise Stability of Submodular Funcs



$$T_\rho f(X) = f(X_{[0]}) + \mathrm{E}\,\Sigma\, f(X_\rho \cap X_{[j]}) - f(X_\rho \cap X_{[j\text{-}1]})$$

submodularity

$$f(S) - f(S \cap T) \geq f(S \cup T) - f(T)$$

$S = X_\rho \cap X_{[j]}$ $T = X_{[j\text{-}1]}$

Noise Stability of Submodular Funcs

$\pi(1)\ \pi(2)\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \pi(s)$

$$T_\rho f(X) = f(X_{[0]}) + \mathrm{E}\sum f(X_\rho \cap X_{[j]}) - f(X_\rho \cap X_{[j-1]})$$

submodularity

$$f(S) - f(S \cap T) \geq f(S \cup T) - f(T)$$

$S = X_\rho \cap X_{[j]}$ $\qquad$ $T = X_{[j-1]}$

$S \cap T = X_\rho \cap X_{[j-1]}$

Noise Stability of Submodular Funcs

$\pi(1)\ \pi(2)\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \pi(s)$

$$T_\rho f(X) = f(X_{[0]}) + \mathrm{E}\,\Sigma\, f(X_\rho \cap X_{[j]}) - f(X_\rho \cap X_{[j-1]})$$

submodularity

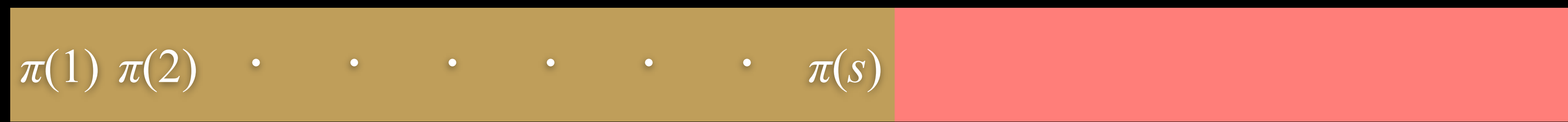
$$f(S) - f(S \cap T) \geq f(S \cup T) - f(T)$$

$S = X_\rho \cap X_{[j]}$

$S \cap T = X_\rho \cap X_{[j-1]}$

$T = X_{[j-1]}$

Noise Stability of Submodular Funcs



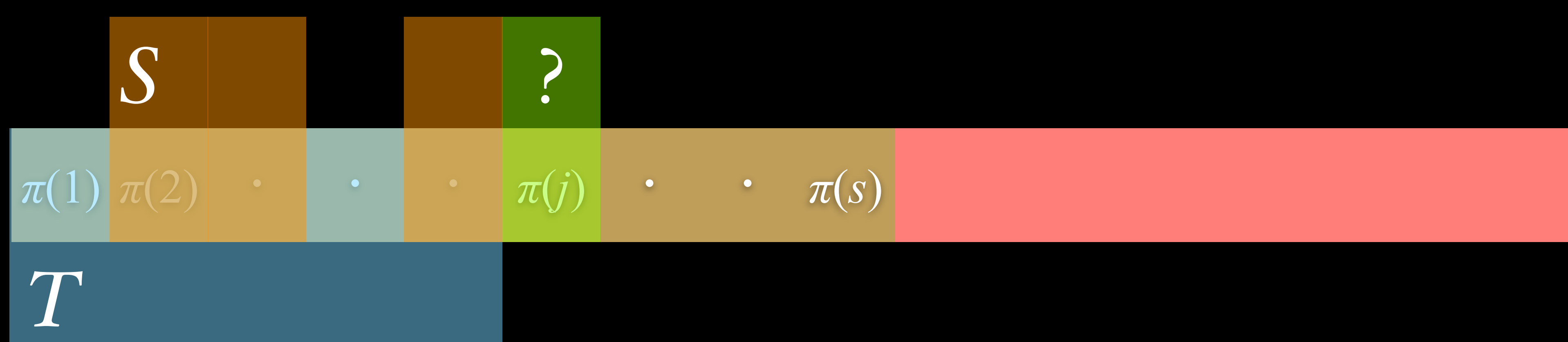
$$T_\rho f(X) = f(X_{[0]}) + \mathrm{E}\,\Sigma f(X_\rho \cap X_{[j]}) - f(X_\rho \cap X_{[j-1]})$$

submodularity

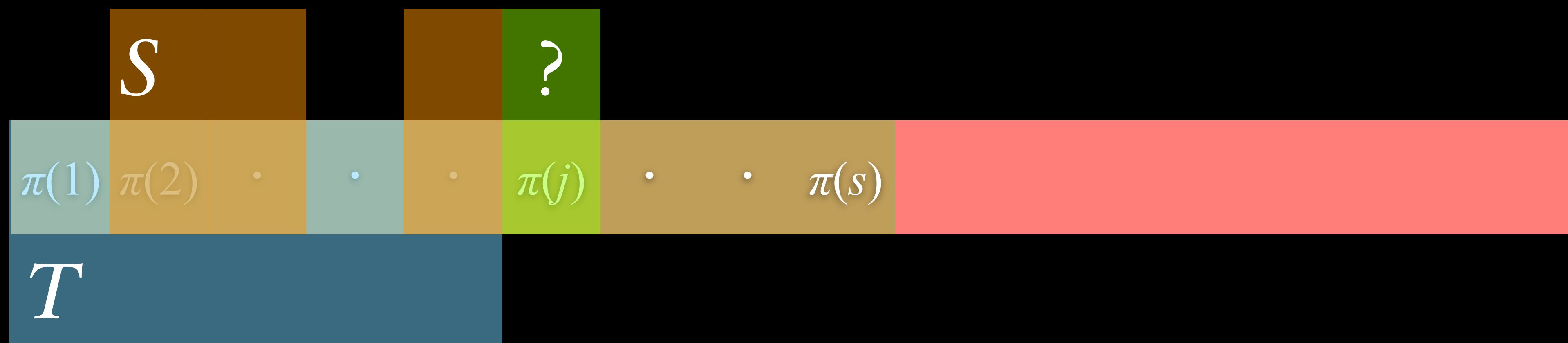
$$f(S) - f(S \cap T) \geq f(S \cup T) - f(T)$$

$S = X_\rho \cap X_{[j]}$ $T = X_{[j-1]}$

$S \cap T = X_\rho \cap X_{[j-1]}$ $S \cup T = (X_\rho \cap \{\pi(j)\}) \cup X_{[j-1]}$



Noise Stability of Submodular Funcs

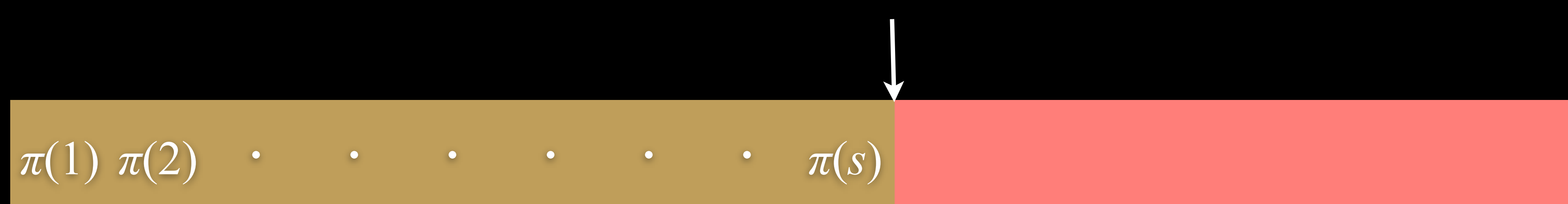


$$T_\rho f(X) \ge \mathrm{E}\, \Sigma f(S \cup T) - f(T)$$

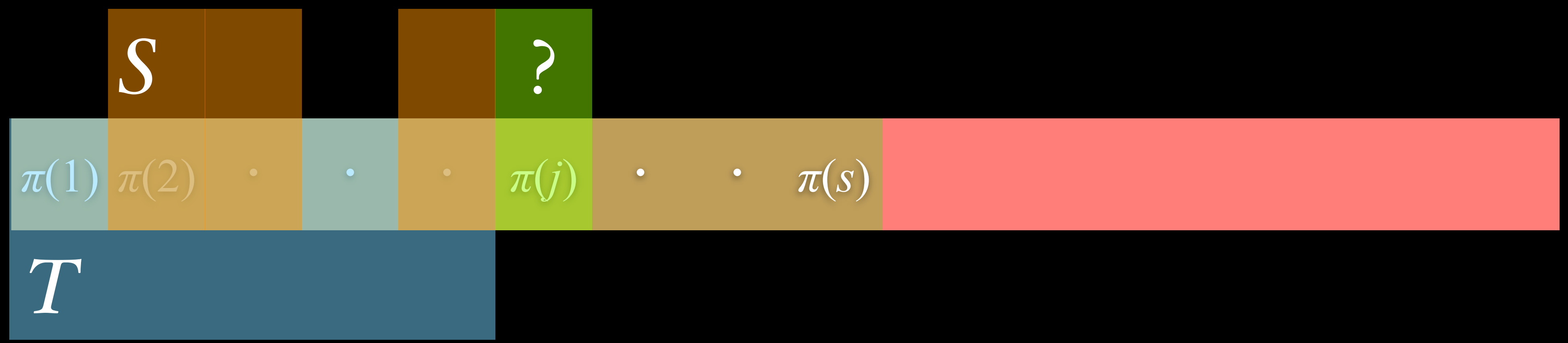
$$= \Sigma(f(X_{[j]}) - f(X_{[j-1]}))\,((1+\rho X_{\pi(j)})/2)$$

$$\ge \Sigma(\rho/2)(X_{\pi(j)} - X_{\pi(j+1)}) f(X_{[j]})$$

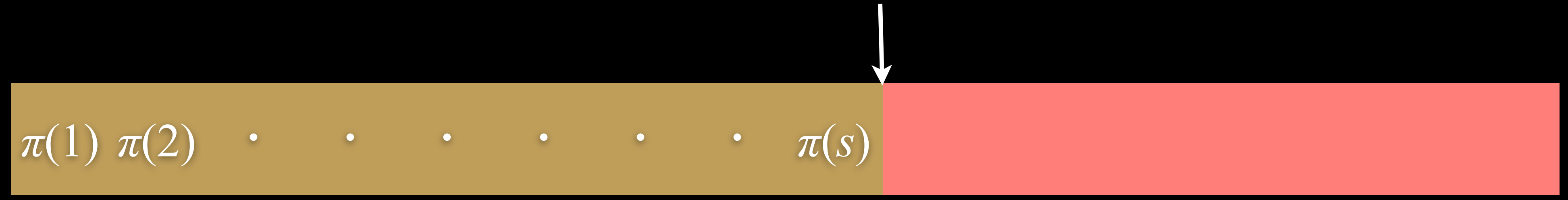
$$= \rho f(X)$$



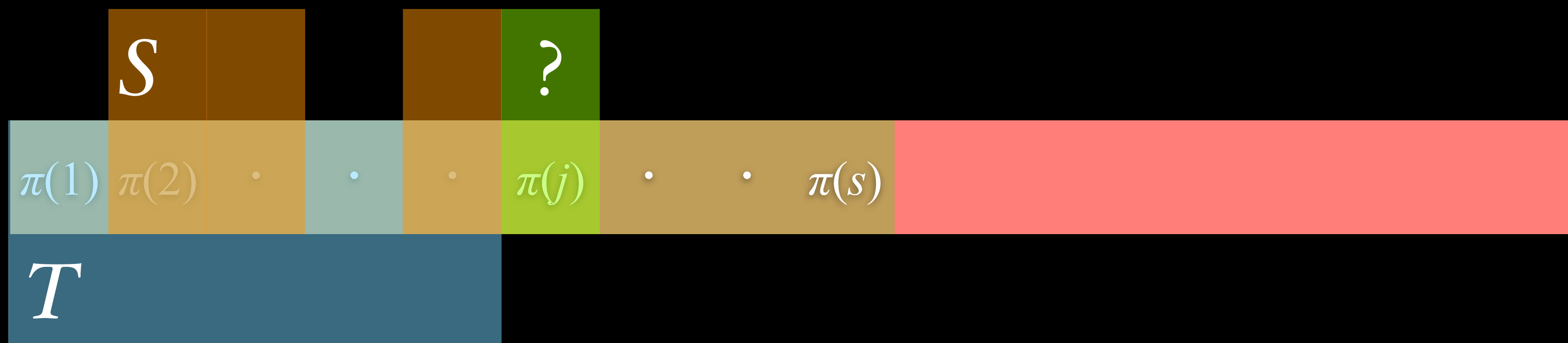
Noise Stability of Submodular Funcs



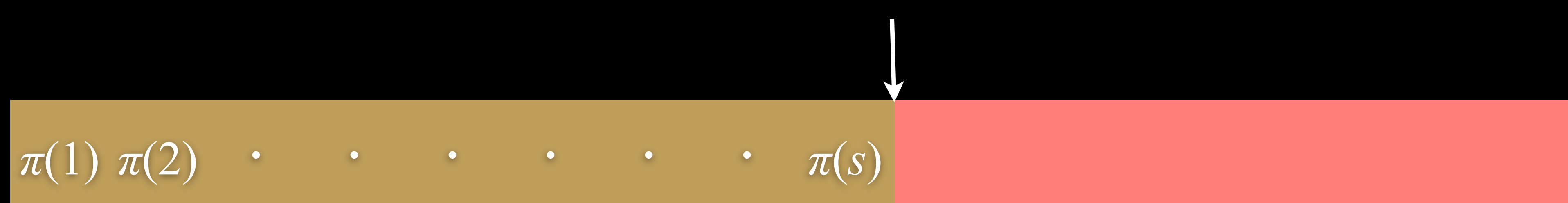
$$
\begin{aligned}
T_\rho f(X) &\geq \mathrm{E}\, \Sigma f(S \cup T) - f(T) \\
&= \Sigma\big(f(X_{[j]}) - f(X_{[j-1]})\big)\,\big((1+\rho X_{\pi(j)})/2\big) \\
&\geq \Sigma(\rho/2)(X_{\pi(j)} - X_{\pi(j+1)}) f(X_{[j]}) \\
&= \rho f(X)
\end{aligned}
$$



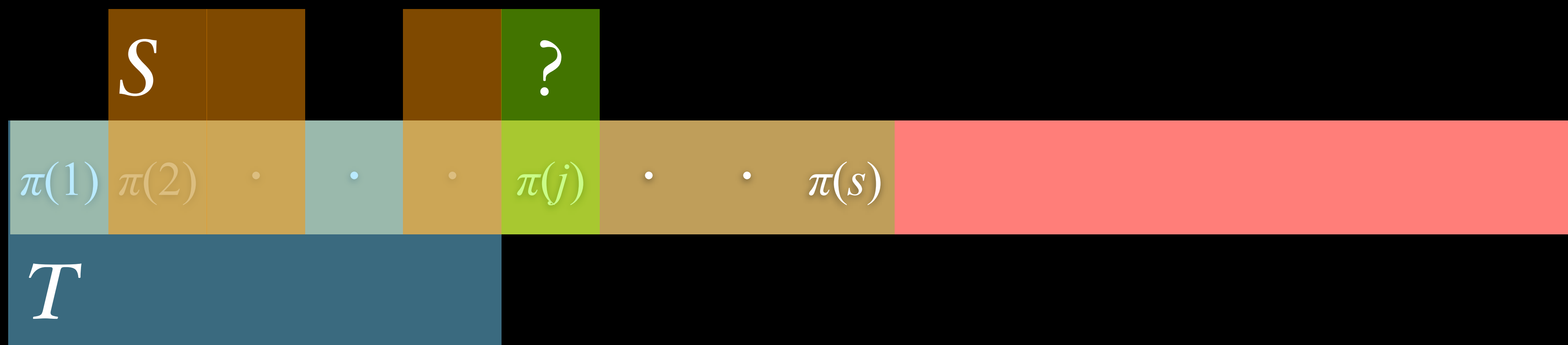
Noise Stability of Submodular Funcs



$$
\begin{aligned}
T_\rho f(X) &\geq \mathrm{E}\,\Sigma f(S \cup T) - f(T) \\
&= \boxed{\Sigma\left(f(X_{[j]}) - f(X_{[j-1]})\right)\left((1+\rho X_{\pi(j)})/2\right)} \\
&\geq \Sigma(\rho/2)(X_{\pi(j)} - X_{\pi(j+1)}) f(X_{[j]}) \\
&= \rho f(X)
\end{aligned}
$$



Noise Stability of Submodular Funcs

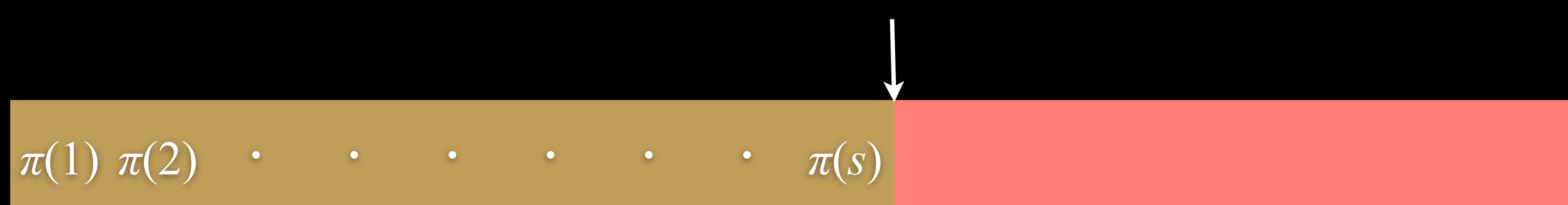


$$T_\rho f(X) \geq \mathrm{E}\,\Sigma f(S \cup T) - f(T)$$

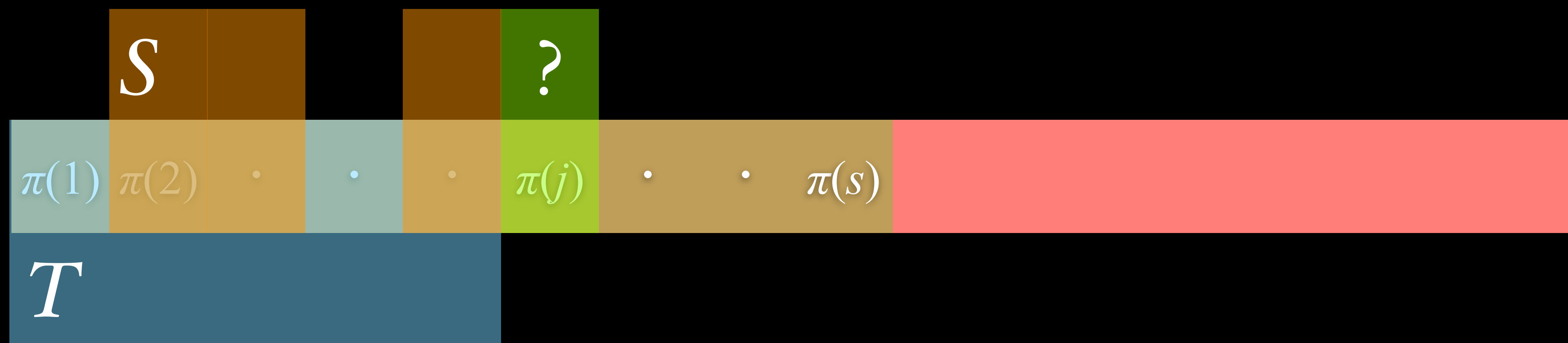
$$= \Sigma(f(X_{[j]}) - f(X_{[j-1]}))\,((1+\rho X_{\pi(j)})/2)$$

$$\geq \boxed{\Sigma(\rho/2)(X_{\pi(j)} - X_{\pi(j+1)}) f(X_{[j]})}$$

$$= \rho f(X)$$



Noise Stability of Submodular Funcs



$$T_\rho f(X) \geq \mathrm{E}\, \Sigma\, f(S \cup T) - f(T)$$

$$= \Sigma(f(X_{[j]}) - f(X_{[j-1]}))\, ((1+\rho X_{\pi(j)})/2)$$

$$\geq \Sigma(\rho/2)(X_{\pi(j)} - X_{\pi(j+1)}) f(X_{[j]})$$

$$= \rho f(X)$$

Submodular Functions Are Noise Stable

Let $f : 2^{[n]} \to R^+$ be a submodular function.

Then for all $\rho \in [0,1]$:

$$\boxed{S_\rho(f) \geq \rho \cdot \mathrm{E}[f^2]}$$

[MadimanTetali-10, Vondrák-09]

Application 1

Submodular functions are *agnostically learnable* using *statistical queries*.

Agnostic Learning

C a class of concepts, $c : 2^{[n]} \longrightarrow R$

D a (product) distribution over $2^{[n]}$

f $2^{[n]} \longrightarrow R$: an arbitrary target function

opt $\min_{c \in C} \mathrm{E}_D[|c(x) - f(x)|]$

An ***agnostic learner*** is given access to f and w.h.p. outputs h s.t.

$$\mathrm{E}_D[|h(x) - f(x)|] \leq opt + \varepsilon$$

Levels of Access

		Query	Answer	
Strong	VQ_f	x	$f(x)$	query access
	EX_f	n/a	$x \sim D, f(x)$	random examples
Weak	SQ_f	s, τ	$E_D[s(x, f(x))] \pm \tau$	s: an efficiciently computable statistic

Proof

1. Submodular functions are noise stable.

Proof

1. Submodular functions are noise stable.
2. Noise stable functions are well-approximated by low-degree multilinear polynomials. [KOS-02]

Proof

1. Submodular functions are noise stable.
2. Noise stable functions are well-approximated by low-degree multilinear polynomials. {KOS-02}
3. The {KKMS-05} algorithm agnostically learns low-degree multilinear polynomials using statistical queries.

Our Result

There is an SQ algorithm that can agnostically learn the class of non-negative submodular functions with bounded *l*2-norm in time poly($n^{O(1/\varepsilon^2)}$).

Previous Results

	Caveat	Noise	Access	Approx.
BalcanHarvey 2011	monotone non-negative Lipschitz min-val m.	None	EX	Mult. $\log(1/\varepsilon)/m$ w.p. $1-\varepsilon$

Previous Results

	Caveat	Noise	Access	Approx.
BalcanHarvey 2011	monotone non-negative Lipschitz min-val m.	None	EX	Mult. $\log(1/\varepsilon)/m$ w.p. $1-\varepsilon$
GHRU 2011	non-negative bounded range	Small	VQ	Add. ε w.p. $1-\delta$

Previous Results

	Caveat	Noise	Access	Approx.
BalcanHarvey 2011	monotone non-negative Lipschitz min-val m.	None	EX	Mult. $\log(1/\varepsilon)/m$ w.p. $1\text{-}\varepsilon$
GHRU 2011	non-negative bounded range	Small	VQ	Add. ε w.p. $1\text{-}\delta$
This Work	non-negative bounded $l2$ norm	Adversarial	SQ	$l1$-error ε

Previous Results

	Caveat	Noise	Access	Approx.
BalcanHarvey 2011	monotone non-negative Lipschitz min-val m.	None	EX	Mult. $\log(1/\varepsilon)/m$ w.p. $1-\varepsilon$
GHRU 2011	non-negative bounded range	Small	VQ	Add. ε w.p. $1-\delta$
This Work	non-negative bounded $l2$ norm	Adversarial	SQ	$l1$-error ε
BalcanHarvey 2011	monotone non-negative *any* D	None	EX	Mult. $O(\sqrt{n})$ w.p. $1-\varepsilon$

Application 2: Privacy

There is an ε-differentially private algorithm that (α,β)-releases the class of all Boolean conjunctions of width w in time $w^{O(\log 1/\alpha)}$.

There is an ε-differentially private algorithm that (α,β) releases the class of all halfspaces in d dimensions in time $d^{O(1/\alpha 4)}$.

Thanks!

And remember...

1. Submodular functions are as noise stable as dictator functions.
2. Noise stability is more important than Lipschitzness over product distributions.