Submodular Maximization via Taylor Series Approximation

Gözde Özcan*, Armin Moharrer, Stratis Ioannidis

Department of Electrical and Computer Engineering

April 29 – May 1, 2021

Northeastern University
Center for Signal Processing, Imaging, Reasoning, and Learning

siam 2021 International Conference on Data Mining

Northeastern University

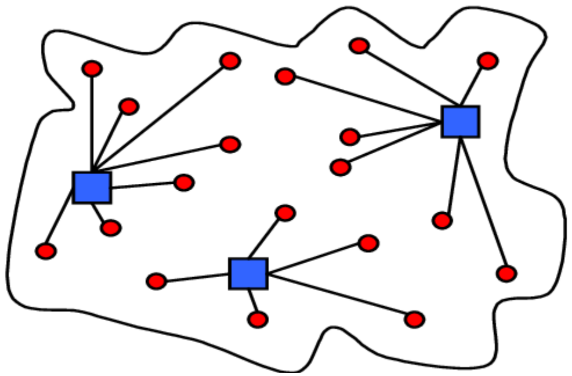
Submodular Function Maximization

- *Submodular function:* set function with diminishing returns

$$\forall A \subseteq B \subseteq V,\ e \in V$$

$$f(B \cup \{e\}) - f(B) \leq f(A \cup \{e\}) - f(A)$$

- *Examples:* facility location, document summarization, influence maximization maximum coverage etc.



LVX VERITAS VIRTVS

Maximum Coverage

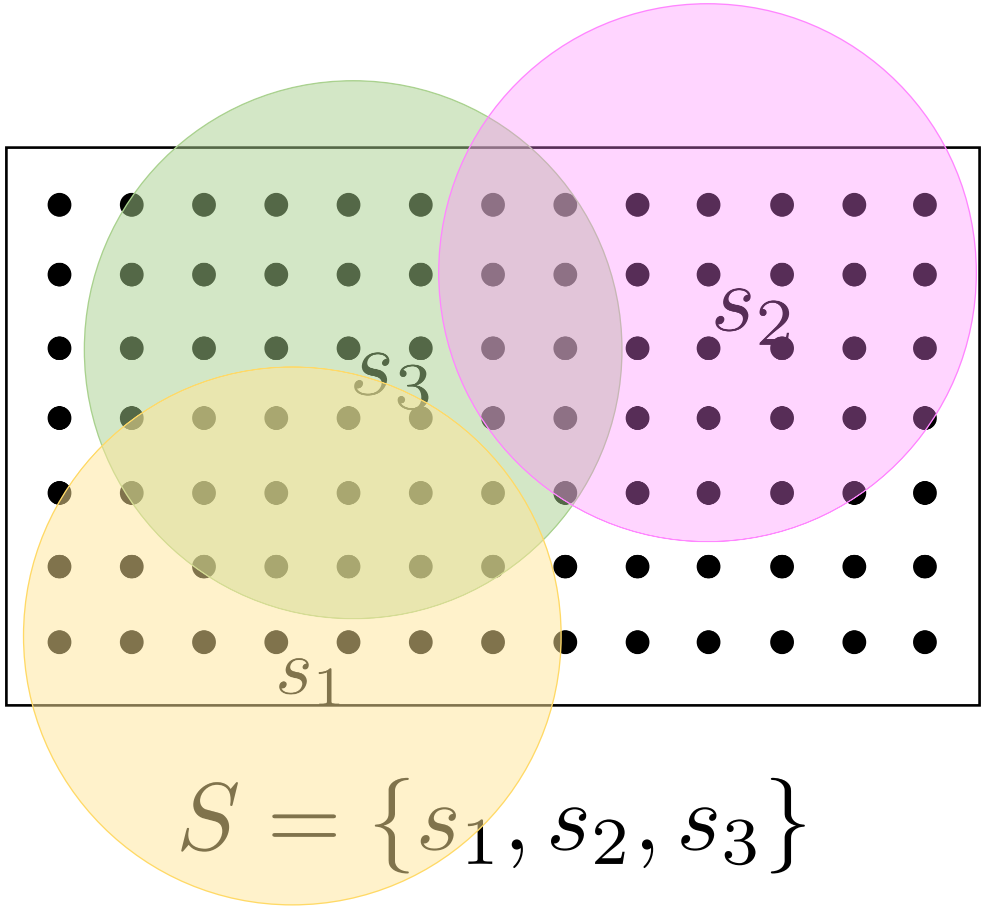
- Finite set $V = \{1,\ 2 \dots, N\}$
- Set function $f\colon\ 2^V \ \to \mathbb{R}$
- $f(S)$ is # locations covered by having sensors at S
- Diminishing returns

$$\forall A \subseteq B \subseteq V,\, e \in V$$

$$f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B)$$

$$(f(\emptyset \cup \{s_3\}) - f(\emptyset)) \geq (f(\{s_1, s_2\} \cup \{s_3\}) - f(\{s_1, s_2\}))$$

$$40 \geq 25$$



$$S = \{s_1, s_2, s_3\}$$

$$f(S) = 73$$

Maximum Coverage subject to Matroid Constraints

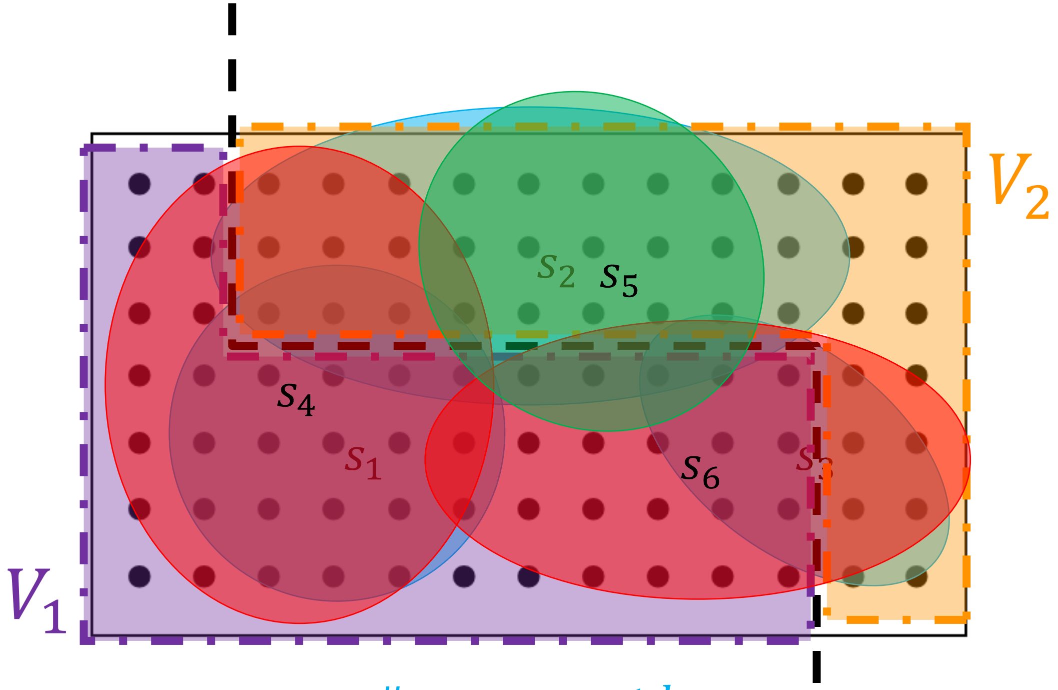
- Cardinality constraints

$$\max_{|S| \leq k} f(S)$$

- Matroid constraints

$$\max_{|S \cap V_1| \leq k_1,\, |S \cap V_2| \leq k_2} f(S)$$

- NP-hard
- Poly-time approximation algorithms exist



sensors $\leq k$

sensors in V_1 $\leq k_1$

sensors in V_2 $\leq k_2$

Approximation Algorithms

Greedy algorithm achieves $1 - 1/e$ approximation ratio on uniform matroids *[Nemhauser and Wolsey, 1978].*

This approximation ratio drops down to $\frac{1}{2}$ for matroids *[Vondrák, 2008].*

The *continuous greedy* algorithm improves this and achieves $1 - 1/e$ approximation ratio on general matroids in poly-time *[Calinescu et al., 2011].*

Challenge

- Continuous greedy uses *multilinear relaxation*

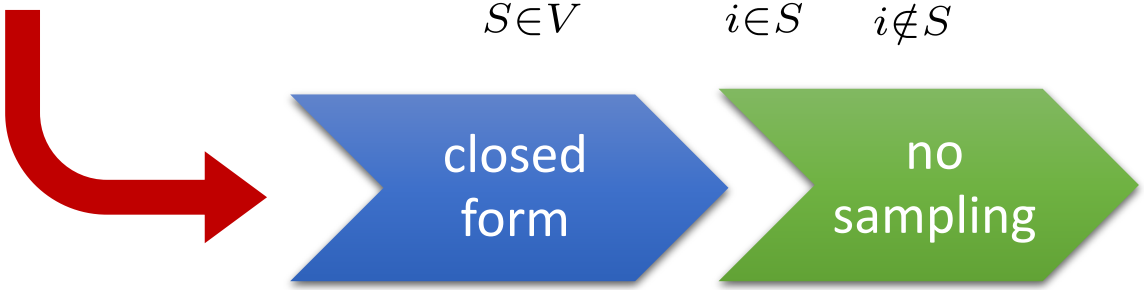
$$G(\mathbf{y}) = \mathbb{E}_{\mathbf{y}}[f(S)] = \sum_{S \in V} f(S) \prod_{i \in S} y_i \prod_{i \notin S} (1 - y_i)$$

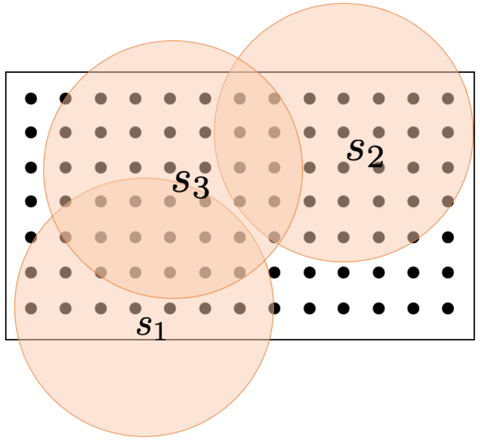
$$P(i \in S) = y_i, \text{where } y_i \in [0, 1]$$

- Exponential number of terms ($2^{|V|}$)
- Current methods often rely on sampling, computationally expensive (e.g $|V|^8$)

Coverage Function

$$G(\mathbf{y}) = \mathbb{E}_{\mathbf{y}}[f(S)] = \sum_{S \in V} f(S) \prod_{i \in S} y_i \prod_{i \notin S} (1 - y_i)$$



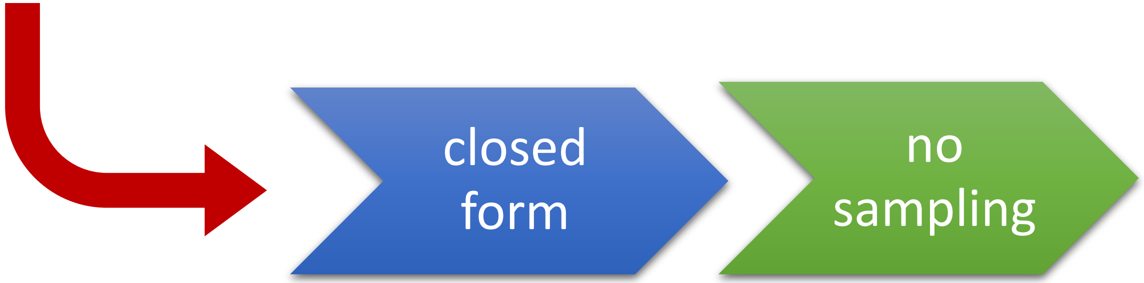


$S = \{s_1, s_2, s_3\}$

$f(S) = \#$ locations covered

- *Coverage function is multilinear.*

$$G(\mathbf{y}) = \mathbb{E}_{\mathbf{y}}[\text{multilinear}(S)]$$



This Paper

- We extend this notion of multilinearity and apply it to the functions which are compositions of multilinear functions.

$$f(S) = h(\text{multilinear}(S))$$

multilinear relaxation

$$G(\mathbf{y}) = \mathbb{E}_{\mathbf{y}}[h(\text{multilinear}(S))]$$

still has a closed form!

no sampling!

Our Approach

- We avoid sampling by replacing h with its *Taylor approximation.*

not multilinear

$$f(S) = h(\text{multilinear}(S))$$

Taylor approximation

$$\hat{f}(S) = \hat{h}(\text{multilinear}(S))$$

multilinear

multilinear relaxation

$$\hat{G}(\mathbf{y}) = \mathbb{E}_{\mathbf{y}}[\hat{h}(\text{multilinear}(S))]$$

no sampling!

Our Contributions

- A novel polynomial estimator of the multilinear relaxation
    - No sampling
- Theoretical guarantees
- Multiple applications
    - Data summarization, influence maximization, facility location, cache networks
    - Performance improvement (74% lower error in 89% less time)

Outline

- Polynomial Estimator
- Experiments
- Conclusion & Future Work

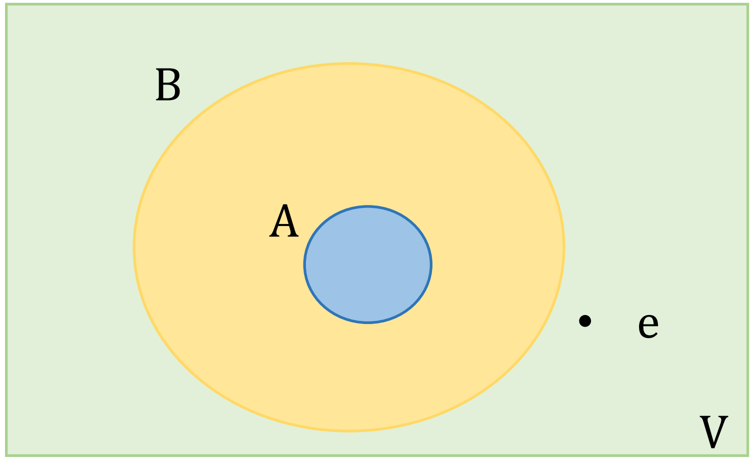
Monotone Submodular Functions

- **Submodularity:** A function $f: 2^V \to \mathbb{R}$ is *submodular* if for every $A \subseteq B \subseteq V$, and $e \in V$ / B it holds that

$$f(B \cup \{e\}) - f(B) \leq f(A \cup \{e\}) - f(A).$$

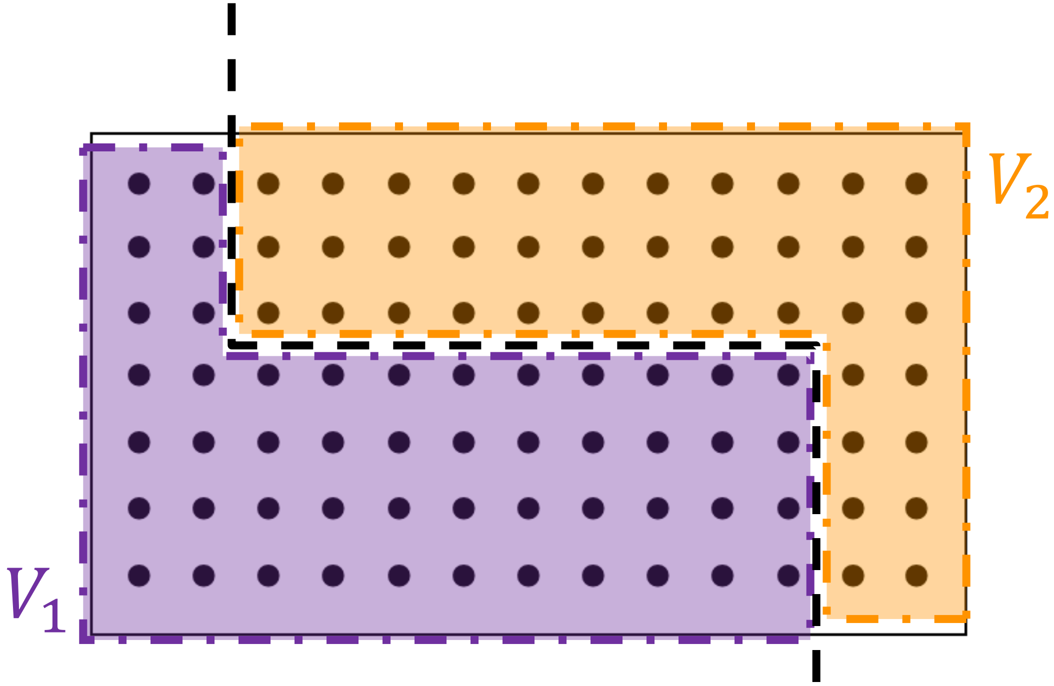
- **Monotonicity:** A function $f: 2^V \to \mathbb{R}$ is *monotone* if for every $A \subseteq B \subseteq V$,

$$f(A) \leq f(B).$$



Matroids

- **Matroids:** Given a ground set V, a matroid is a pair $\mathcal{M} = (V, I)$, where $I \subseteq 2^V$ is a collection of *independent sets*, for which
    1. If $B \in I$ and $A \subset B$, then $A \in I$.
    2. If $A, B \in I$ and $|A| < |B|$, there exists $x \in B \setminus A$ such that $A \cup \{x\} \in I$.



$\#\ sensors\ in\ V_1 \leq k_1$ $\quad$ $\#\ sensors\ in\ V_2 \leq k_2$

$$\mathcal{I} = \{S \subseteq 2^V \mid |S \cap V_\ell| \leq k_\ell,\ \text{for}\ \ell = 1, 2\}$$

Binary Notation

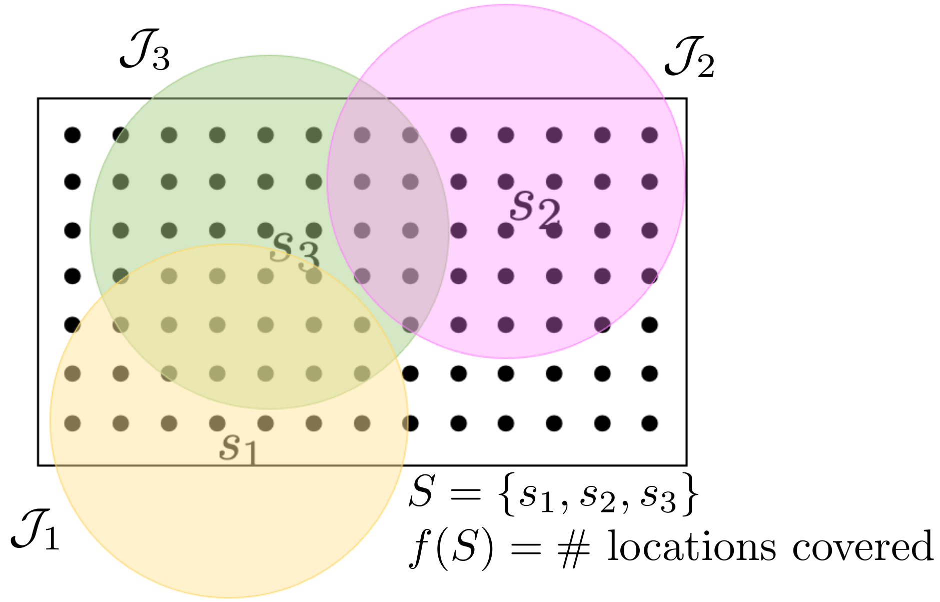
- Binary vector $\boldsymbol{x}$ whose support is S:

$$S \subset V \iff x \in \{0,1\}^N$$

$$i \in S \iff x_i = 1$$
$$i \notin S \iff x_i = 0$$

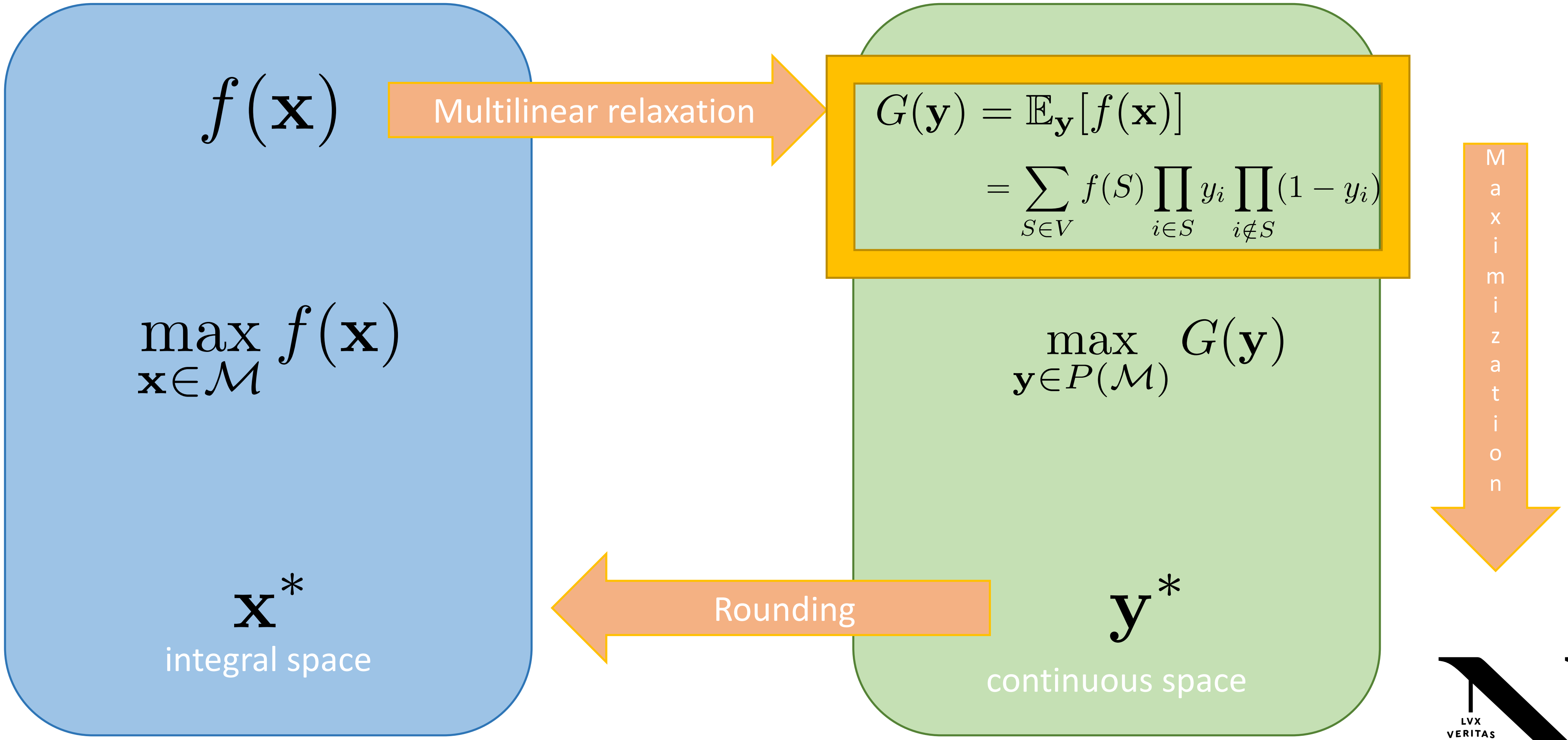
$$\boxed{\max_{S \in \mathcal{M}} f(S)} \iff \boxed{\max_{\mathbf{x} \in \mathcal{M}} f(\mathbf{x})}$$

- Example: coverage



$$f(\mathbf{x}) = \sum_{\ell=1}^{N} \left(1 - \prod_{i:\ell \in \mathcal{J}_i} (1 - x_i)\right)$$

Continuous Greedy Algorithm



Multilinear Functions

- A function $g\colon \mathbb{R}^N \to \mathbb{R}$ is multilinear if it is affine w.r.t. each of its coordinates:

$$g(\mathbf{x}) = \sum_{\ell \in \mathcal{I}} c_\ell \prod_{i \in \mathcal{J}_\ell} x_i$$

$c_\ell \in \mathbb{R}$

$\ell \in \text{some index set } \mathcal{I}$

$\text{subsets } \mathcal{J}_\ell \subseteq V$

- Multilinear relaxation of a multilinear function has a closed form:

$$G(\mathbf{y}) = \mathbb{E}_{\mathbf{y}}[g(\mathbf{x})] = g(\mathbf{y})$$

does not require sampling!

Polynomial Estimator

- Given g is a multilinear function:

not multilinear

$$f(\mathbf{x}) = \underbrace{h(g(\mathbf{x}))}$$

Taylor approximation of degree L

$$\hat{f}_L(\mathbf{x}) = \underbrace{\hat{h}_L(g(\mathbf{x}))}$$

multilinear

multilinear relaxation

L^{th} *degree polynomial estimator of* f

$$\hat{G}(\mathbf{y}) = \mathbb{E}_{\mathbf{y}}[\hat{f}_L(\mathbf{x})] = \hat{f}_L(\mathbf{y})$$

no sampling!

Assumptions

1) Function $f: \{0,1\}^N \to \mathbb{R}_+$ is monotone and submodular.
2) Function $f: \{0,1\}^N \to \mathbb{R}_+$ has form

$$f(\mathbf{x}) = \sum_{j=1}^{M} w_j h_j(g_j(\mathbf{x}))$$

$M \in \mathbb{N}$

$w_j \in \mathbb{R}$

$h_j : [0,1] \to \mathbb{R}_+$

$g_j : [0,1]^N \to [0,1]$

$j \in \{1, \ldots, M\}$

$L \in \mathbb{N}$

a) multilinear ($g_j(\mathbf{x})$)

b) polynomial $\hat{h}_L$ of degree L exists (h_j)

Main Theorem

- Assume a function $f: \{0,1\}^N \rightarrow \mathbb{R}_+$ satisfies these assumptions. Then, consider the continuous greedy algorithm in which $G(\boldsymbol{y})$ is estimated via the polynomial estimator $\widehat{f_L}(\boldsymbol{y})$. Then,

$$G(\mathbf{y}) \geq \left(1 - \frac{1}{e}\right) G(\mathbf{y}^*) - O\left(\varepsilon(L)\right) - O\left(\frac{1}{K}\right)$$

Multilinear relaxation of f

Output of the continuous greedy algorithm

Multilinear relaxation of f

an optimal solution to $\max_{\mathbf{y} \in P(\mathcal{M})} G(\mathbf{y})$

bias of the polynomial estimator

# of iterations

Examples

	Input	$g_j : \{0,1\}^{\lvert V \rvert} \to [0,1]$ $\mathbf{x} \to g_j(\mathbf{x})$	$h_j : [0,1] \to \mathbb{R}_+$ $s \to h_j(s)$	$f : \{0,1\}^{\lvert V \rvert} \to \mathbb{R}_+$ $\mathbf{x} \to f(\mathbf{x})$	Bias $\varepsilon(L)$
SM	Partitions $\bigcup_{j=1}^{M} \{P_j\} = V$ weights $\mathbf{r} \in \mathbb{R}^N_+$ and $\sum_{i=1}^{N} r_i = 1$	$\sum_{i \in P_j} r_i x_i$	$\log(1+s)$	$\sum_{j=1}^{M} h(s_j)$	$\frac{M\sqrt{N}}{(L+1)2^L}$
IM	Instances $G = (V, E)$ of a directed graph, partitions $\{P^j\}_{j=1}^{N} \subset V$	$\sum_{i \in V} \frac{1}{N}\left(1 - \prod_{u \in P_i^j} (1 - x_u)\right)$	$\log(1+s)$	$\frac{1}{M}\sum_{j=1}^{M} h(s_j)$	$\frac{\sqrt{N}}{(L+1)2^L}$
FL	Complete weighted bipartite graph $G = (V \cup V')$ weights $w \in [0,1]^{N \times M}$	$\sum_{\ell=1}^{N} (w_{i_\ell,j} - w_{i_{\ell+1},j}) \left(1 - \prod_{k=1}^{\ell} (1 - x_{i_k})\right)$	$\log(1+s)$	$\frac{1}{M}\sum_{j=1}^{M} h(s_j)$	$\frac{\sqrt{N}}{(L+1)2^L}$
CN	Graph $G = (V, E)$, service rates $\mu \in \mathbb{R}^M_+$, requests $r \in \mathcal{R}$, P_j path of r, arrival rates $\lambda \in \mathbb{R}^{\lvert\mathcal{R}\rvert}_+$	$\frac{1}{\mu_j} \sum_{r \in \mathcal{R}: j \in p^r} \lambda^r \prod_{k'=1}^{k_{p^r}(v)} (1 - x_{p^r_{k'}, i^r})$	$\frac{s}{1-s}$	$\sum_{j=1}^{M} h(s_0) - \sum_{j=1}^{M} h(s_j)$	$2M\sqrt{\lvert V\rvert\lvert\mathcal{C}\rvert}\frac{\bar{s}^{L+1}}{1-\bar{s}}$

$$\lim_{L \to \infty} \varepsilon(L) \to 0$$

Outline

- A Polynomial Estimator
- Experiments
- Conclusion & Future Work

Experiments

$$\texttt{err} = (f(\mathbf{y}) - f^*)/f^*$$
$$f^* = \max f(\mathbf{y})$$

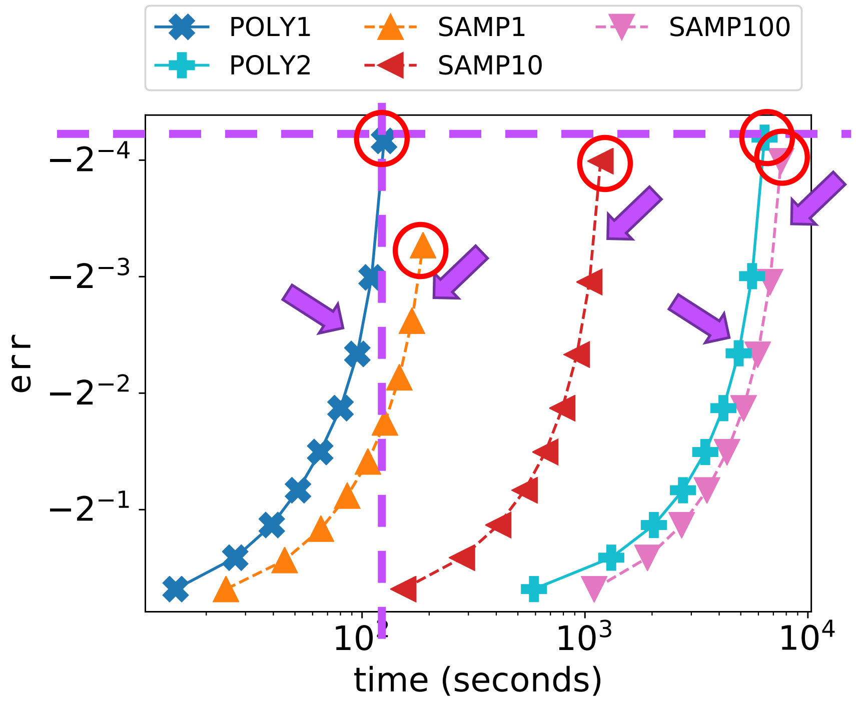


Figure 1: Trajectory of the `err` function as a function of time.

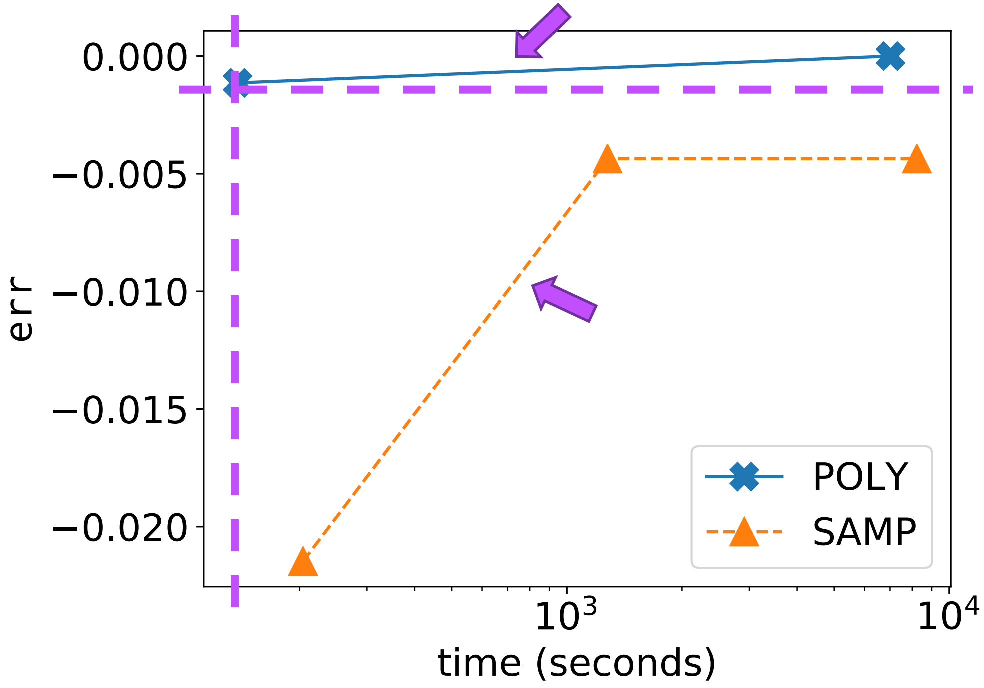


Figure 2: `err` function of the final results as a function of time.

Outline

- A Polynomial Estimator
- Experiments
- Conclusion & Future Work

Conclusion & Future Work

- We propose an alternative to the sampling estimator
  - polynomial,
  - achieves $1 - 1/e$ approximation ratio.

- In the future, this work might be extended to
  - online,
  - stochastic algorithms.

CCF-1750539

Thank you!

Northeastern University