Stochastic Submodular Maximization via Polynomial Estimators

Gözde Özcan, Stratis Ioannidis
Department of Electrical and Computer Engineering

May 26, 2023

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

PAKDD 2023

Northeastern University

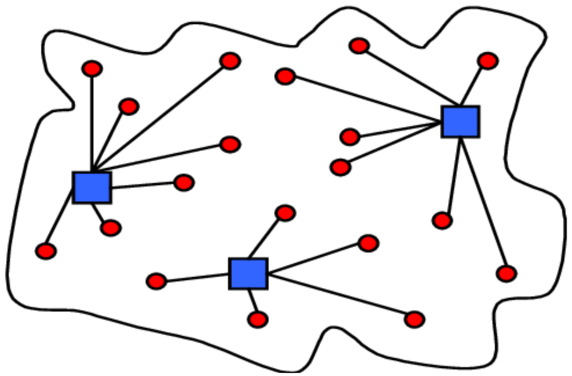
Submodular Function Maximization

- *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.

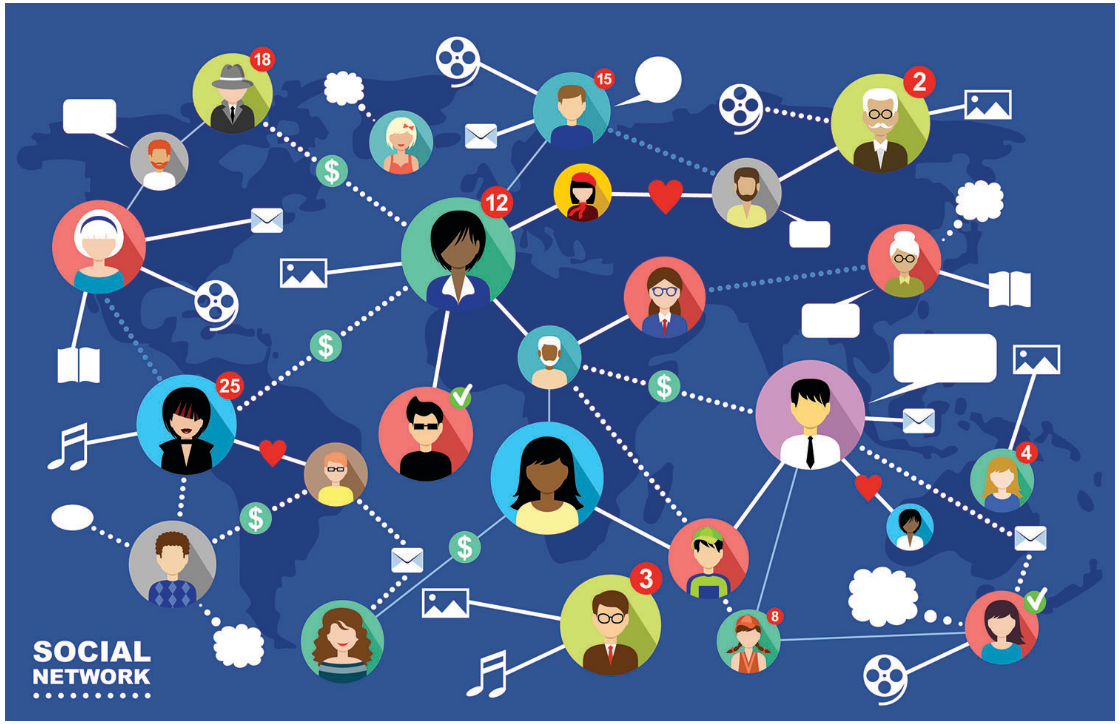


Stochastic Submodular Maximization

- No access to the function oracle

$$f(S) = \mathbb{E}_{z \sim P}[f_z(S)]$$

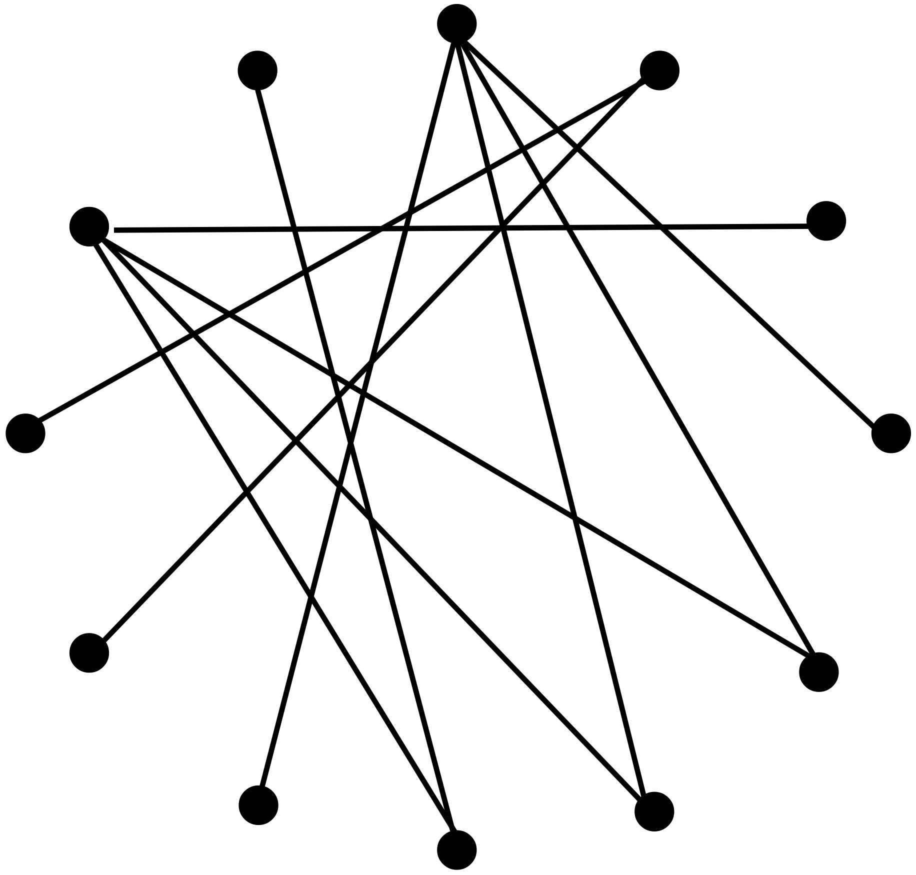
- Only sample a random $f_z(\cdot)$ at a time



Influence Maximization

$$f(S) = \mathbb{E}_{z\sim P}[f_z(S)]$$

- $G = (V, E)$
- Finite set $\mathrm{V} = \{1,\ 2 \ldots, \mathrm{n}\}$
- Independent Cascades or Linear Threshold [Kempe et al., 2003]
- $f_z\ (S)$ is the fraction of nodes that are reachable by infecting the seeds $S \subseteq V$



Approximation Algorithms for Stochastic Submodular Maximization

Concave relaxation method achieves $1 - 1/e$ approximation ratio only for coverage functions *[Karimi et al., 2017].*

Projected gradient methods for general submodular problems achieve $1/2$ *approximation ratio* [Hassani et al., 2017].

The *stochastic continuous greedy* algorithm improves this and achieves $1 - 1/e$ approximation ratio on general matroids in poly-time *[Mokhtari et al., 2020].*

Challenges

- Stochastic Continuous Greedy (SCG) [Mokhtari et al., 2020] uses the *multilinear relaxation*

$$G(\mathbf{y}) = \mathbb{E}_{S\sim\mathbf{y}}[f(S)] = \mathbb{E}_{S\sim\mathbf{y}}\left[\mathbb{E}_{z\sim P}[f_z(S)]\right]$$

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

- Sampling S to compute the multilinear relaxation (computationally expensive)
- Having two sources of randomness leads to high variance

Submodular Maximization via Taylor Series Approximation [Özcan et al., 2021]

- Multilinear relaxation of a polynomial function is still a polynomial function.
- $G(\boldsymbol{y})$ is computed efficiently is f is well approximated by a polynomial.

polynomial   polynomial

$$\mathbb{E}_{\mathbf{x}\sim\mathbf{y}}[p(\mathbf{x})] = \mathbb{E}_{\mathbf{x}\sim\mathbf{y}}[\dot{p}(\mathbf{x})] = \dot{p}(\mathbf{y})$$

does not require sampling!

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

LVX VERITAS VIRTVS

Our Contributions

- We extend the notion of polynomial estimators of multilinear relaxations [Özcan et al., 2021] to stochastic submodular optimization
  - No sampling
  - One source of randomness
- Theoretical guarantees
  - Bias/variance trade off

Outline

- Stochastic Setting
- Polynomial Estimator & Main Theorem
- Conclusion & Future Work

Stochastic Setting

- Discrete stochastic submodular maximization
- Set function $f\colon 2^V \to \mathbb{R}_+$ of the form:

$$f(S) = \mathbb{E}_{z\sim P}[f_z(S)]$$

where $S \subseteq V$, where z is the realization of the random variable Z is drawn from a distribution P.

- For each realization of $z \sim P$, $f_z\colon 2^V \to \mathbb{R}_+$ is monotone and submodular.
- Objective is:

$$\max_{S\in\mathcal{I}} f(S) = \max_{S\in\mathcal{I}} \mathbb{E}_{z\sim P}[f_z(S)],$$

where $\mathcal{I}$ is a general matroid constraint.

Binary Notation

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

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

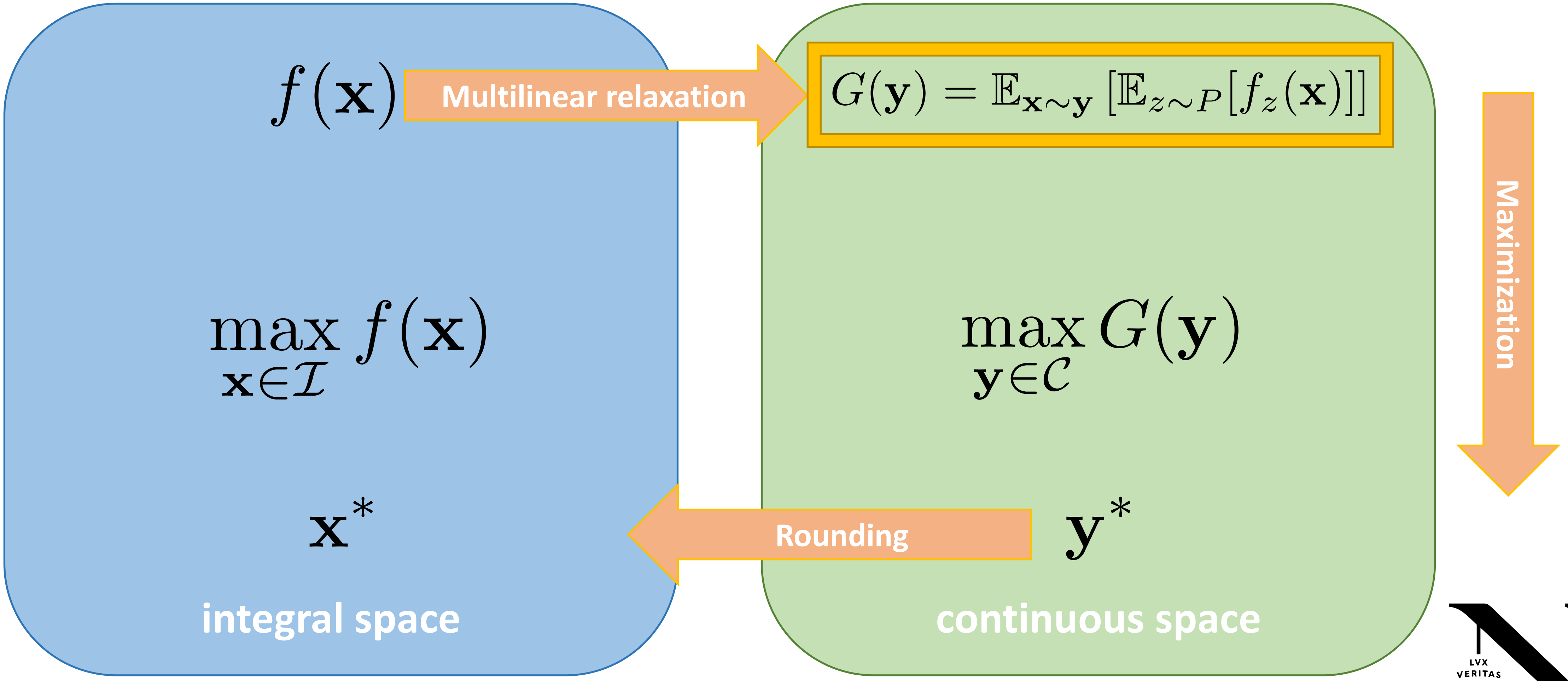
$$i \in S \iff x_i = 1$$

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

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

Stochastic Continuous Greedy Algorithm

[Mokhtari et al., 2020]



Outline

- Stochastic Setting
- Polynomial Estimator & Main Theorem
- Conclusion & Future Work

Assumptions

1) Function $f: \{0,1\}^N \rightarrow \mathbb{R}_+ : f(S) = \mathbb{E}_{z \sim P}[f_z(S)]$ is monotone and submodular.
2) Elements in the constraint set C are uniformly bounded, i.e.,

$$||y|| \leq D.$$

3) For all $x \in \{0,1\}^n$, there exists $\varepsilon_z(L) \geq 0$ such that $\lim_{L \rightarrow \infty} \varepsilon_z(L) = 0$ and

$$|f_z(\mathbf{x}) - \hat{f}_z^L(\mathbf{x})| \leq \underbrace{\varepsilon_z(L)}_{\text{bias}}.$$

$\hat{f}_z^{\,L}(\cdot)$: polynomial estimator of $f_z(\cdot)$

L: degree of the polynomial

Variance of the Gradients

- Variance of the unbiased gradients [Mokhtari et al., 2020]:

$$\sigma^2 = \sup_{\mathbf{y}\in\mathcal{C}} \mathbb{E}\left[\|\widehat{\nabla G_z(\mathbf{y})} - G(\mathbf{y})\|^2\right]$$

- Variance of the biased gradients:

$$\sigma_0^2 = \sup_{\mathbf{y}\in\mathcal{C}} \mathbb{E}_{z\sim P}\left[\|\nabla G_z(\mathbf{y}) - \nabla G(\mathbf{y})\|^2\right]$$

- Mean bias:

$$\varepsilon(L) = \mathbb{E}_{z\sim P}\left[\varepsilon_z(L)\right]$$

Main Theorem

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

$$\mathbb{E}[G(\mathbf{y}_T)] \geq (1 - 1/e)OPT - \frac{15DK}{T^{1/3}} - \frac{f_{\max} r D^2}{2T}$$

Unbiased version: $K \approx 4\sigma + \sqrt{3r} f_{\max} D$

Biased version: $K \approx \sqrt{16\sigma_0^2 + 224\sqrt{n}\varepsilon(L)} + 2\sqrt{r} f_{\max} D$

Examples

	Input	$g_z : \{0,1\}^{\|V\|} \to [0,1]$ $\mathbf{x} \to g_z(\mathbf{x})$	$f_z : \{0,1\}^{\|V\|} \to \mathbb{R}_+$ $\mathbf{x} \to f_z(\mathbf{x})$	$\hat{f}_z^L : \{0,1\}^{\|V\|} \to \mathbb{R}_+$ $\mathbf{x} \to \hat{f}_z^L(\mathbf{x})$	Bias $\varepsilon(L)$
SM	Weighted bipartite graph $G = (V \cup P)$ weights $\mathbf{r}_z \in \mathbb{R}_+^n$, and $\sum_{i=1}^n r_{i,z} = 1$	$\sum_{i \in V \cap P_j} r_{i,z} x_i$	$\sum_{j=1}^J h(g_z(\mathbf{x}))$, where $h(s) = \log(1+s)$	$\hat{h}^L(g_z(\mathbf{x}))$	$\frac{1}{(L+1)2^{L+1}}$
IM	Instances $G = (V, E)$ of a directed graph, partitions $P_v^z \subset V$	$\sum_{i \in V} \frac{1}{N}\left(1 - \prod_{u \in P_i^z}(1 - x_u)\right)$	$h(g_z(\mathbf{x}))$ where $h(s) = \log(1+s)$	$\hat{h}^L(g_z(\mathbf{x}))$	$\frac{1}{(L+1)2^{L+1}}$
FL	Complete weighted bipartite graph $G = (V \cup V')$ weights $w_{i_\ell,z} \in [0,1]^{N \times \|z\|}$	$\sum_{\ell=1}^{N}(w_{i_\ell,z} - w_{i_{\ell+1},z})\left(1 - \prod_{k=1}^{\ell}(1 - x_{i_k})\right)$	$h(g_z(\mathbf{x}))$ where $h(s) = \log(1+s)$	$\hat{h}^L(g_z(\mathbf{x}))$	$\frac{1}{(L+1)2^{L+1}}$
CN	Graph $G = (V, E)$, service rates $\mu \in \mathbb{R}_+^{\|z\|}$, requests $r \in \mathcal{R}$, P_z path of r, arrival rates $\lambda \in \mathbb{R}_+^{\|\mathcal{R}\|}$	$\frac{1}{\mu_z}\sum_{r \in \mathcal{R}: z \in p^r} \lambda^r \prod_{k'=1}^{k_{p^r}(v)}(1 - x_{p_k^r, i^r})$	$h(g_z(\mathbf{0})) - h(g_z(\mathbf{x}))$ where $h(s) = s/(1-s)$	$\hat{h}^L(g_z(\mathbf{x}))$	$\frac{\bar{s}^{L+1}}{1-\bar{s}}$

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

Outline

- Stochastic Setting
- Polynomial Estimator & Main Theorem
- Conclusion & Future Work

Conclusion & Future Work

- We propose an alternative to the sampling estimator
  - polynomial,
  - achieves $1 - 1/e$ approximation ratio,
  - with only one source of randomness.

- A similar extension can be made to the online setting with regret analysis.

our paper

CCF-1750539

Thank you!

Northeastern University