Submodular Function Maximization in Parallel via the Multilinear Relaxation

Chandra Chekuri   Kent Quanrud

UIUC

Purdue Theory Seminar   November 30, 2018

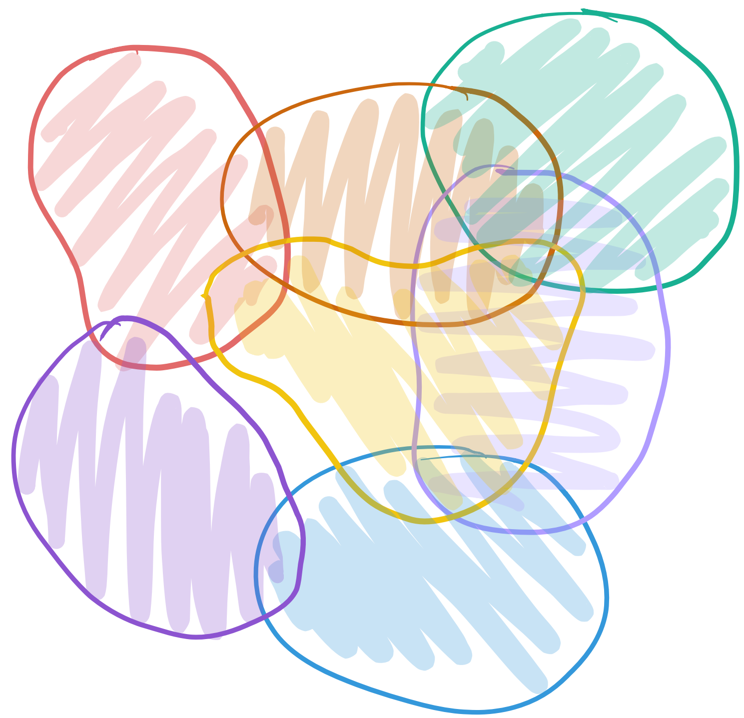
e.g. Maximum coverage

Input: n sets $S_1, S_2, \ldots, S_n$, cardinality k

Goal: choose k sets $S_{i_1}, S_{i_2}, \ldots, S_{i_k}$

maximizing their union $\left| \bigcup_{j=1}^{k} S_{i_j} \right|$

- NP-hard
- $1-\frac{1}{e}$ APX-hardness
- $1-\frac{1}{e}$ APX by greedy
- parallel?



GOAL:

maximize $f(S)$ s.t. $|S| \leq k$

in parallel in the oracle model

where

- $f: 2^N \to \mathbb{R}_{\geq 0}$ is a monotone submodular function
- $k \in \mathbb{N}$, alt: general packing constraints

Monotone submodular functions

set function $f: 2^N \to \mathbb{R}_{\geq 0}$

- $f(\emptyset) = 0$
- (increasing) $S \subseteq T \Rightarrow f(S) \leq f(T)$
- [decreasing marginal returns] denote $f_S(u) = f(S \cup u) - f(S)$

  = "marginal value of u to S"

  $$S \subseteq T \Rightarrow f_S(u) \geq f_T(u)$$

- Oracle: given S, returns $f(S)$

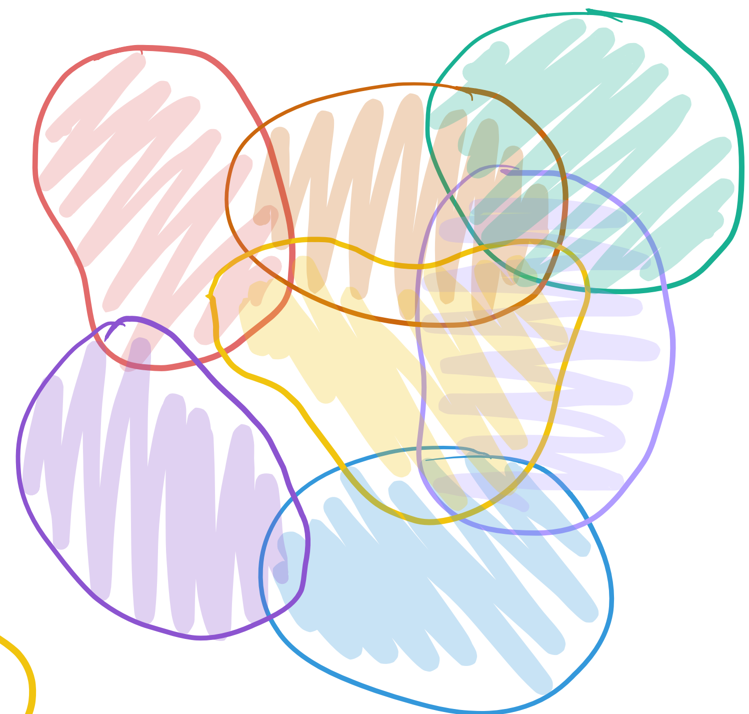
e.g. Maximum coverage

Input: n sets $S_1, S_2, \ldots, S_n$, cardinality k

we have a submodular function

$f: 2^{N} \to \mathbb{R}_{\geq 0}$ where

- $N = [n]$
- $f(I) = \left| \bigcup_{i \in I} S_i \right|$



(here "sets" get a little confusing)

Packing constraints

$$|S| \le k \longrightarrow A\,\mathbb{1}_S \le 1$$

$$A \in [0,1]^{m \times n}$$

e.g.

* Knapsack constraint

costs $a_1, a_2, \ldots, a_n$ $A = [a_1\ a_2\ \cdots\ a_n]$

* matchings * partition matroids

* laminar matroids * intersections of the above

Greedy algorithm [Nemhauser, Wolsey]

Input: $f: 2^N \to \mathbb{R}_{\geq 0}$, N, k

1. $S \leftarrow \emptyset$
2. while $|S| < k$
   a. $e \leftarrow \underset{e \in N}{\operatorname{argmax}} f_S(e)$ $\quad (= f(S+e) - f(S))$
   b. $S \leftarrow S + e$
3. return S

Greedy algorithm [Nemhauser, Wolsey]

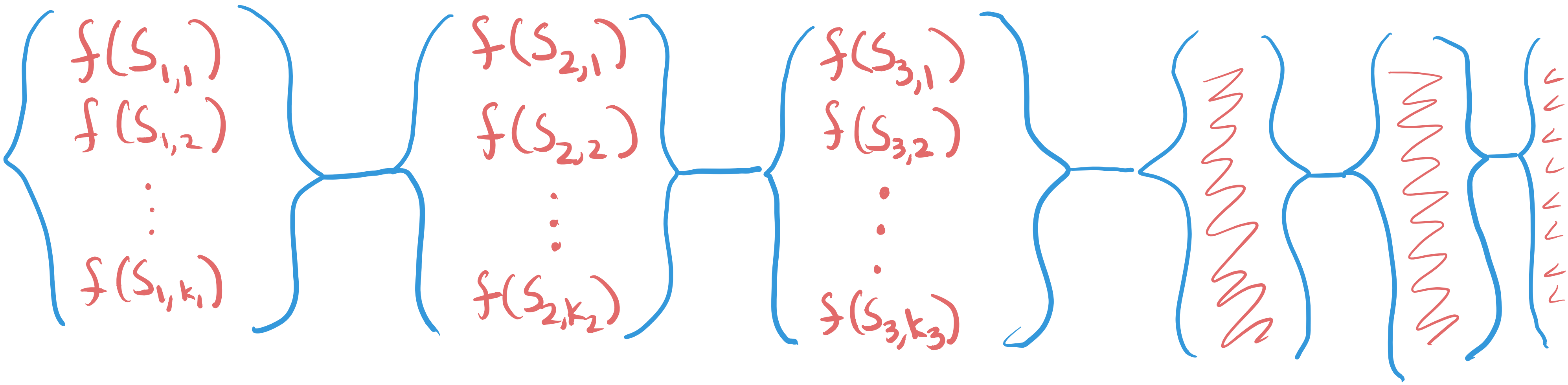
Input: $f: 2^N \to \mathbb{R}_{\geq 0}$, N, k

1. $S \leftarrow \emptyset$
2. while $|S| < k$
   a. $e \leftarrow \underset{e \in N}{\operatorname{argmax}} f_S(e)$
   b. $S \leftarrow S + e$
3. return S

- simple
- Optimal $1 - \frac{1}{e}$ APX
- Sequential: reevaluates margins after each iteration

Adaptivity and parallelization

- Queries to f divided into "adaptive rounds"
- Choice of queries can only depend on queries to f of previous rounds



- e.g. greedy: nk total queries
  k adaptive rounds

Q: how much adaptivity/depth needed to maximize $f(S)$ s.t. $|S| \le k$?

- greedy needs k

Parallel greedy???

Input: $f: 2^N \to \mathbb{R}_{\geq 0}$, N, k (large)

1. $S \leftarrow \emptyset$
2. while $|S| < k$

   /* need to take many (e.g. $\frac{k}{\text{polylog}(n)}$) elements.
   but elements negate each other */

   a. $S \leftarrow S + ???$

   (e.g. overlapping sets in max coverage)

3. return S

Q: how much adaptivity/depth needed to maximize $f(S)$ s.t. $|S| \le k$?

- greedy needs k
- [Balkanski & Singer] $\Omega\left(\frac{\log n}{\log\log n}\right)$ necessary

  $\frac{1}{3}$-APX w/ $O(\log n)$ rounds
- [Balkanski, Rubinstein, Singer] [Ene, Nguyen] $(1-\frac{1}{e}-\varepsilon)$ w/ $O\left(\frac{\log n}{\text{poly}(\varepsilon)}\right)$ rounds
- [Chekuri, Q.]: for general packing constraints

- [Balkanski, Rubinstein, Singer] [Ene, Nguyen] $(1-\frac{1}{e}-\varepsilon)$ w/ $O\left(\frac{\log n}{\text{poly}(\varepsilon)}\right)$ rounds

- [Chekuri, Q.]: for general packing constraints
  - different (simpler?) algorithm for cardinality constraints ("parallel-greedy") via multilinear extension of f
  - deterministic for some explicit cases of f
  - Cardinality ⇒ Knapsack ⇒ general packing via parallel multiplicative weight updates

multilinear extension of $f: 2^N \to \mathbb{R}_{\geq 0}$

$F: [0,1]^N \longrightarrow \mathbb{R}_{\geq 0}$ defined by

$F(x) = E[f(S)]$, where $\left\{ \begin{array}{l} S \text{ samples } e \in N \\ \text{independently with} \\ \text{probability } x_e \end{array} \right\}$

- multilinear, monotone

- monotone-concave: for $x \in [0,1]^N$, $v \in \mathbb{R}^N_{\geq 0}$, $\delta > 0$

  $F(x + \delta v)$ is concave in δ

- concentrated $\Rightarrow$ easy to estimate, round

GOAL:

maximize $f(S)$ s.t. $|S| \leq k$

in parallel in the oracle model

equivalent to

maximize $F(x)$ over $x \geq 0$ s.t. $\sum_i x_i \leq k$

in parallel in the oracle model

Continuous Greedy

[Calinescu, Chekuri, Pál, Vondrák]

1. $x \leftarrow \mathbb{0}$
2. continuously from $t=0$ to k
   a. $v \leftarrow \underset{v \in P}{\operatorname{argmax}} \langle F'(x), v \rangle$

      where $P = \{ v \in [0,1]^N : \langle v, \mathbb{1} \rangle \leq 1 \}$

   b. $\frac{dx}{dt} = v$
3. return x

- $F'(x)$ = continuous analogue of marginal values
- greedily select fractional point v w/ maximum marginal value
- easy to discretize

- works for any downward closed polytope (e.g. matroids)

Continuous greedy in parallel

1. $x \leftarrow 0$

2. continuously from $t=0$ to k

   a. $\frac{dx}{dt} = \frac{S}{|S|}$, where $S = \{e : F'_e(x) \geq (1-\varepsilon) \max_d F'_d(x)\}$

3. return x

- S gathers all $(1-\varepsilon)$-OPT points in polytope

- increase x along all of S uniformly (instead of just the best point).

GOAL: discretize continuous algo to have low depth

Parallel greedy

1. $x \leftarrow 0$, $\lambda \leftarrow OPT$
2. while $\langle x, \mathbb{1} \rangle \leq k$ and $\lambda \geq e^{-1} OPT$
   a. $S = \{ e \in N : F_e'(x) \geq (1-\varepsilon)\lambda / k \}$
   b. while $S \neq \emptyset$
      (i) $x \leftarrow x + \delta S$ for greedy step size $\delta > 0$
   c. $\lambda \leftarrow (1-\varepsilon)\lambda$
3. return x

- λ = upper bound on margin of any competing set
- $S = \{$good coordinates w/r/t threshold $\lambda\}$
- $\delta > 0$ chosen large as possible s.t. S remains $(1-O(\varepsilon))$-good: max δ s.t. $F_x(x+\delta S) \geq (1-\varepsilon)^2 \frac{\delta |S|}{k} \lambda$

Parallel greedy

1. $x \leftarrow \mathbb{0}$, $\lambda \leftarrow OPT$
2. while $\langle x, \mathbb{1} \rangle \leq k$ and $\lambda \geq e^{-1} OPT$
   a. $S = \{ e \in N : F_e'(x) \geq (1-\varepsilon)\lambda / k \}$
   b. while $S \neq \emptyset$
      (i) $x \leftarrow x + \delta S$ for greedy step size $\delta > 0$
   c. $\lambda \leftarrow (1-\varepsilon)\lambda$
3. return x

δ is:

- small enough to keep directional deriv $\langle F'(x+\delta S), S \rangle$ good, $\Rightarrow$ continuous greedy
- big enough to decrease directional deriv enough that ε-fraction of S drops out

"greedy step size" is defined by $(\lambda \approx OPT - F(x))$

$$\max \delta \text{ s.t. } F(x+\delta S) - F(x) \geq (1-\varepsilon)^2 \frac{\delta|S|}{k} \lambda$$

$$\frac{F(x+\delta S) - F(x)}{\delta|S|} = \frac{\text{"bang"}}{\text{"buck"}}$$

(a) $$\lim_{\delta \downarrow 0} \frac{F(x+\delta S) - F(x)}{\delta|S|} = \frac{\langle F'(x), S \rangle}{|S|} \geq (1-\varepsilon) \frac{\lambda}{k}$$

since $S = \{ i : F_i'(x) \geq (1-\varepsilon) \lambda/k \}$

(b) $$\frac{F(x+\delta S) - F(x)}{\delta|S|}$$ is decreasing in δ as $F'(x+\delta S)$ is decreasing

Analysis: (easy?)

① Approximation factor

(or how we choose δ small enough to approximately follow continuous greedy)

② Depth

(or how we choose δ big enough to kill off a large fraction of the good points S)

APX analysis

"greedy step size" is defined by

$$\max \delta \text{ s.t. } F(x+\delta S) - F(x) \geq (1-\varepsilon)^2 \frac{\delta |S|}{k} \lambda$$

let $t = \sum_i x_i$ = frac. capacity used. $(\lambda \approx OPT - F(x))$

dF = change in F (as function of δ)

dt = change in t

rewrite greedy step size as

$$\max \delta \text{ s.t. } dF \geq (1-\varepsilon)^2 \frac{\lambda}{k} dt$$

rearranging:

$$\boxed{\frac{dF}{dt} \geq (1-\varepsilon)^2 \frac{\lambda}{k}}$$

same as Greedy and cont. greedy

"greedy step size" is defined by

APX analysis

max δ s.t. $\frac{dF(x)}{dt} \geq (1-\varepsilon)^2 \frac{OPT - F(x)}{k}$ $\quad (t = \sum_i x_i)$

intially $F(x) = 0$ when $x = 0$, $t = 0$

diff. system $\Rightarrow$

$$OPT - F(x) \leq e^{-(1-\varepsilon)^2 t/k} OPT \quad \forall t > 0$$

$$\Rightarrow F(x) \geq (1 - O(\varepsilon))(1 - e^{-1}) OPT$$

at the end, when $t = \sum_i x_i = k$

Analysis: (easy?)

1. ✓ Approximation factor

(or how we choose δ small enough to approximately follow continuous greedy)

2. Depth

(or how we choose δ big enough to kill off a large fraction of the good points S)

Depth

"greedy step size" is defined by

$$\max \delta \text{ s.t. } F(x+\delta S) - F(x) \geq (1-\varepsilon)^2 \frac{\delta|S|}{k} \lambda$$

$(\lambda \approx OPT - F(x))$

$$\frac{F(x+\delta S) - F(x)}{\delta |S|} = \frac{\text{"bang"}}{\text{"buck"}} \geq (1-\varepsilon)\frac{\lambda}{k} \text{ at } \delta \approx 0$$

since $\uparrow$ is decreasing in δ, greedy step size $\equiv$

$$\min \delta \text{ s.t. } \frac{F(x+\delta S) - F(x)}{\delta |S|} \leq (1-\varepsilon)^2 \frac{\lambda}{k}$$

* average margin of element in S has decreased

greedy step size $\Rightarrow \dfrac{F(x+\delta S)-F(x)}{\delta|S|} \leq (1-\varepsilon)^2 \dfrac{\lambda}{k}$

Depth

$$\delta\langle F'(x+\delta S), S\rangle \overset{①}{\leq} \int_0^\delta \langle F'(x+tS), S\rangle \, dt$$

$$= F(x+\delta S)-F(x)$$

$$\overset{②}{\leq} (1-\varepsilon)^2 \frac{\lambda}{k}\delta|S|$$

① $F'(x+\delta S)$ decreasing in δ by submodularity

② by greedy choice of δ

Depth

we now have

$$\langle F'(x+\delta S), S\rangle \le (1-\varepsilon)^2 \frac{\lambda}{k} |S|$$

$(S = \{i : F'_i(x) \ge (1-\varepsilon)\lambda\})$

also $$[\cdots] \ge \underbrace{\left|\left\{i : F'_i(x+\delta S) \ge (1-\varepsilon)\frac{\lambda}{k}\right\}\right|}_{=|S| \text{ in the next iteration}} \cdot (1-\varepsilon)\frac{\lambda}{k}$$

$\Rightarrow$ $|S|$ decreases by $(1-\varepsilon)$-mult. factor.

$\Rightarrow \cdots \Rightarrow \cdots \Rightarrow$ poly$(\log n, \frac{1}{\varepsilon})$ depth

Analysis: (easy?)

1. Approximation factor

(or how we choose δ small enough to approximately follow continuous greedy)

2. Depth

(or how we choose δ big enough to kill off a large fraction of the good points S)

$\blacksquare$

Parallel greedy

1. $x \leftarrow \mathbb{0}$, $\lambda \leftarrow \mathrm{OPT}$
2. while $\langle x, \mathbb{1} \rangle \leq k$ and $\lambda \geq e^{-1}\mathrm{OPT}$
   a. $S = \{ e \in N : F_e'(x) \geq (1-\varepsilon)\lambda / k \}$
   b. while $S \neq \emptyset$
      (i) $x \leftarrow x + \delta S$ for greedy step size $\delta > 0$
   c. $\lambda \leftarrow (1-\varepsilon)\lambda$
3. return x

"primal-dual" approach

- continuous greedy (along S) ensures $(1-\frac{1}{e})$ APX
- greedy step size drives down margins and limits # iterations

Parallel greedy

1. $x \leftarrow \mathbb{0}$, $\lambda \leftarrow \mathrm{OPT}$
2. while $\langle x, \mathbb{1} \rangle \leq k$ and $\lambda \geq e^{-1}\mathrm{OPT}$
   a. $S = \{e \in N : F_e'(x) \geq (1-\varepsilon)\lambda/k\}$
   b. while $S \neq \emptyset$
      (i) $x \leftarrow x + \delta S$ for greedy step size $\delta > 0$
   c. $\lambda \leftarrow (1-\varepsilon)\lambda$
3. return x

Q: is there an equally simple but combinatorial analog?

<u>Randomized-Parallel-Greedy</u>

("combinatorial version")

1. $Q \leftarrow \emptyset$, $\lambda \leftarrow \mathrm{OPT}$
2. while $|Q| \leq k$ and $\lambda \geq e^{-1}\mathrm{OPT}$
   a. $S = \{e \in N : f_Q(e) \geq (1-\varepsilon)\lambda/k\}$
   b. while $S \neq \emptyset$
      (i) $Q \leftarrow Q \cup R$ for R sampling δS w/r/t greedy step size $\delta > 0$
   c. $\lambda \leftarrow (1-\varepsilon)\lambda$
3. return Q

---

maintain discrete set Q by immediately rounding δS in each iteration

Beyond cardinality

- From continuous viewpoint,

  cardinality $\approx$ knapsack $\overset{①}{\approx}$ linear packing constraints

① via MWU:

Integrate submodular techniques w/ "parallel MWU" techniques [Young '02]

- From continuous viewpoint,

cardinality $\approx$ knapsack $\approx$ linear packing constraints

* For knapsack, "good coordinates" defined by

"bang-for-buck ratio" $\dfrac{F_e'(x)}{c_e}$ [where c_e = cost of item e]

* MWU uses Lagrangian relaxations to collapse to a sequence of knapsack problems.

Matroids

$\max f(S)$ s.t. S independent in some matroid $M = (N, I)$

e.g. suppose N = edges in a graph:

$\max f(S)$ s.t. S being a forest

cont. greedy $\Rightarrow (1-e^{-1})$-APX in sequential setting.

Parallel?

Why matroids?

* matroids are very general, help unify many ideas in combinatorial optimization

* $(1-e^{-1})$-APX for monotone submodular max led to many new insights and techniques

* Adds another layer of "sequential"-ness, as elements depend on each other opaquely w/r/t feasibility as well

Greedy ($\frac{1}{2}$-APX) [Fisher, Nemhauser, Wolsey]

Input: $f : N \to \mathbb{R}_{\geq 0}$, matroid $\mathcal{M} = (N, \mathcal{I})$

1. $S \leftarrow \emptyset$
2. while S does not span N
   a. $e \leftarrow \underset{d \in N \setminus S}{\text{argmax}} \{ f_S(d) \mid S + d \text{ feasible} \}$
   b. $S \leftarrow S + e$
3. return S

---

sequential due to (a) reevaluating margins
(b) feasibility w/r/t matroid

Q: can we maximize submodular function s.t. matroid constraint w/ low depth?

We get following APX w/ poly$(\log n, \frac{1}{\varepsilon})$ depth

	monotone f	generally nonnegative f
discrete solution	$(1-\varepsilon)\frac{1}{2}$	$\Omega(1)$
fractional solution	$(1-\varepsilon)(1-e^{-1})$	$(1-\varepsilon)e^{-1}$

* some overlapping results w/ [Balkanski Rubinstein Singer] [Ene Nguyen Vladu]

Parallel Greedy($f: \mathcal{N} \to \mathbb{R}_{\geq 0}$, $\mathcal{M} = (\mathcal{N}, \mathcal{I})$)

1. $I \leftarrow \emptyset$, $R \leftarrow \emptyset$, $\lambda \leftarrow \max_e f(e)$
2. while $\lambda \geq OPT/\text{poly}(n)$
   a. while $S = \{e \in \mathcal{N} \setminus \text{span}(R) : f_R(e) \geq (1-\varepsilon)\lambda\} \neq \emptyset$
      (i) let Q sample S indep. w/ prob. δ for "greedy step size" δ // sample
      (ii) $J \leftarrow \{e \in Q$ s.t. $e \notin \text{span}(R+Q-e)\}$ // filter
      (iii) $I \leftarrow I + J$, $R \leftarrow R + Q$
   b. $\lambda \leftarrow (1-\varepsilon)\lambda$
3. return I

a. while $S = \{e \in N \setminus span(R) : f_R(e) \geq (1-\varepsilon)\lambda\} \neq \emptyset$

(i) let Q sample S indep. w/ prob. δ for "greedy step size" δ //sample

(ii) $J \leftarrow \{e \in Q$ s.t. $e \notin span(R+Q-e)\}$ //filter

(iii) $I \leftarrow I+J$, $R \leftarrow R+Q$

R+Q may be dependent,

Filtering out elements spanned by R+Q

$\Rightarrow$ I+J independent

a. while $S = \{e \in N \setminus span(R) : f_R(e) \geq (1-\varepsilon)\lambda\} \neq \emptyset$

(i) let Q sample S indep. w/ prob. δ for "greedy step size" δ // sample

(ii) $J \leftarrow \{e \in Q \text{ s.t. } e \notin span(R+Q-e)\}$ // filter

(iii) $I \leftarrow I+J$, $R \leftarrow R+Q$

choose δ large as possible s.t.*

* simplifying slightly

$$\underbrace{E[f(I+J)] - E[F(I)]}_{\text{"bang"}} \geq (1-\varepsilon)^2 \lambda \delta \underbrace{|S|}_{\text{"buck"}}$$

maximality of $\delta \Rightarrow |S|$ decreases in expectation

- Parallel greedy $\Rightarrow \frac{1-\varepsilon}{2}$-APX for monotone submodular f

- "Multilinear amplification" [Badanidiyuru, Vondrák]
  $\Rightarrow (1-\varepsilon-\frac{1}{e})$-APX fractional sol'n

- analysis extends to generally nonnegative submodular f

  (new) amplification framework
  $\Rightarrow \frac{1-\varepsilon}{e}$-APX fractional sol'n

Conclusion

- Recent interest in parallel submodular max: different constraints, nonnegative setting, better depth, better oracle complexity

[Balkanski Singer] [Ene Nguyen] [Balkanski Rubinstein Singer]

[Fahrback Mirrokni Zadimoghaddam] [Ene Nguyen Vladu] [Chen Feldman Karbasi]

(w/ multiplicities)

Unifying techniques

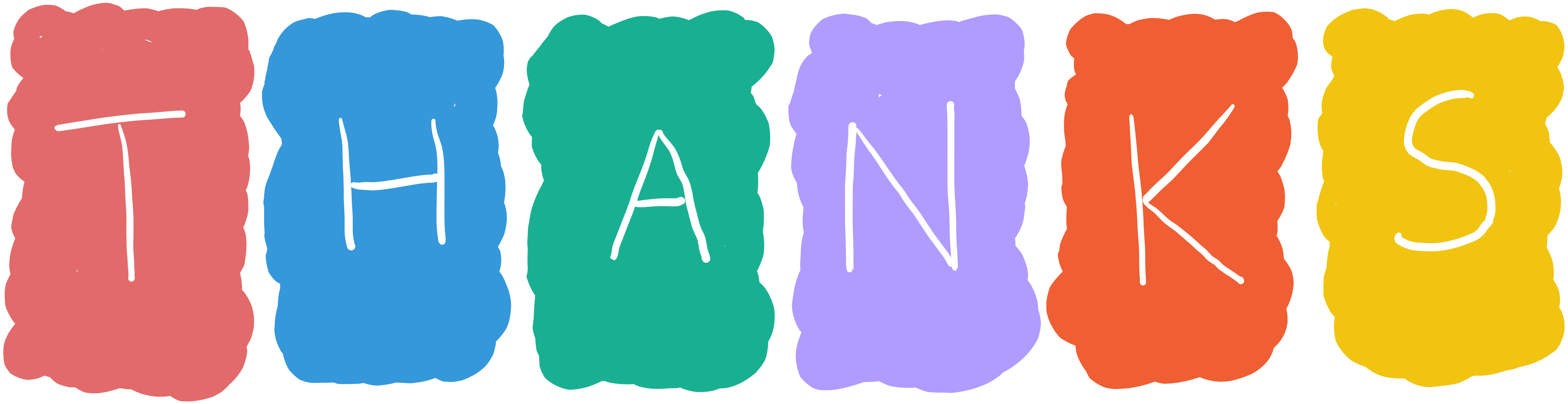
- [continuous greedy]
  + [bulk update]
  + [greedy step size]

  parallel greedy w/ low depth

- easily discretized
- extends to general packing via MWU

- matroids too by similar techniques

- Q: rounding?


THANKS