#4658: On the Performance-Complexity Tradeoff in Stochastic Greedy Weak Submodular Optimization

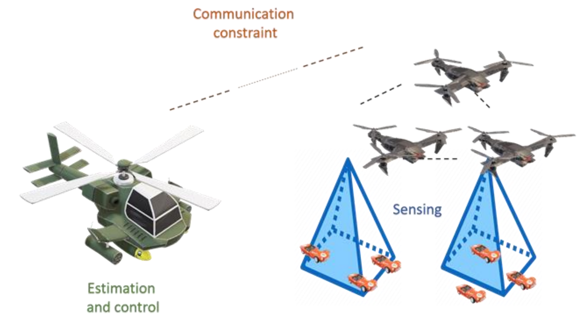
INSTITUTE FOR COMPUTATIONAL ENGINEERING & SCIENCES — ODEN

Abolfazl Hashemi, Haris Vikalo, Gustavo de Veciana

The University of Texas at Austin
Electrical and Computer Engineering
Cockrell School of Engineering

SUBSET SELECTION

Large amount of high-dimensional data



Resource-constrained decision making

NP-hard, combinatorial optimization tasks

WEAK SUBMODULARITY

Constrained combinatorial optimization

$$\hat{S} = \arg\max_{|S|\le k} f(S)$$

NP-hard

Krause, 2011

Weak-submodularity constant $0 < \alpha \le 1$

$$f(\hat{S}) \ge (1 - e^{-\alpha})\, f(S^*)$$

Optimal approximation guarantee

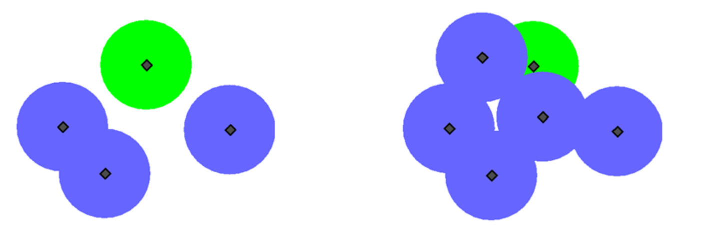
Krause, 2011

Approximate greedy solution

f(S) has nice properties:

- Monotonicity
- Weak-submodularity

Hashemi et al., 2018



$$f(A \cup \{d\}) - f(A) \;\ge\; f(B \cup \{d\}) - f(B)$$

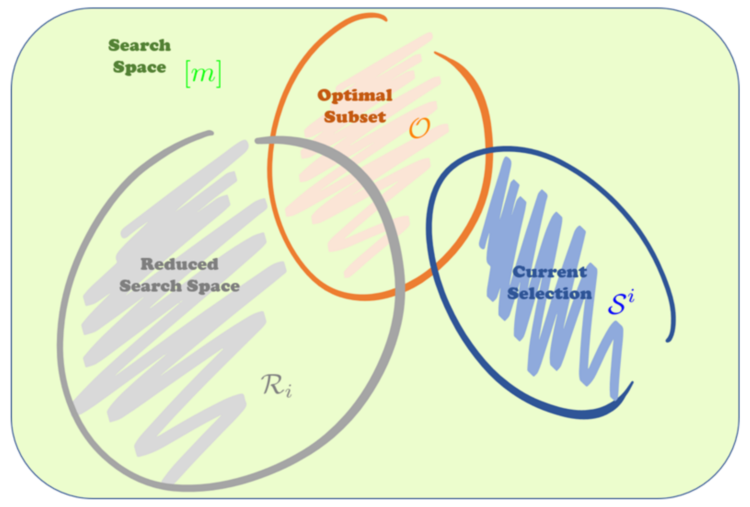
STOCHASTIC GREEDY METHODS

Greedy selection

- Tight approximation guarantee
- Prohibitive computational cost

Reduce the space of greedy by **random sampling**

Mirzasoleyman et al. 2015, Hashemi et al. 2018



How to construct R_i

The dilemma of greedy selection with restricted search space

Greedy with a restricted search space should succeed in each iteration

A fixed schedule: chance of sampling from optimal decreases given success in prior iterations

HOW TO SELECT THE SCHEDULE?

Theorem 1

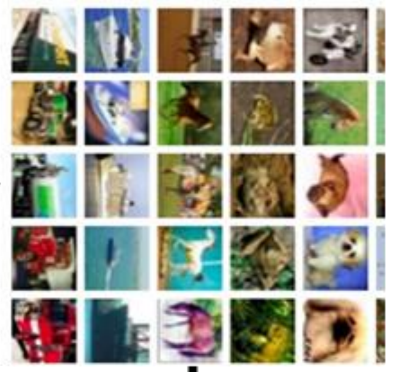
Let $\mathcal{P}(m,k)$ denote a series of subset selection problems where $m, k \to \infty$, $m > k$. Let ALG denote a variant of Greedy with a restricted uniform search space $\mathcal{R}_i \subset [m]$ having cardinality r. Then:

1. Vanishing regime: $\beta \in (0,1)$ such that $\frac{r}{m} \le k^{\beta-1}$, then
$$\limsup_{m,k\to\infty} \Pr\left(\mathcal{S}^k_{alg} = \mathcal{S}^\star\right) = 0.$$
2. Relative regime: $\beta_1 \in (0,1)$ such that $\frac{r}{m} \le \beta_1$, then
$$0 < \delta_1 < \limsup_{m,k\to\infty} \Pr\left(\mathcal{S}^k_{alg} = \mathcal{S}^\star\right) < \delta_2 < 0.63.$$

We need increasing $\{r_i\}$ that ultimately grows to $r_{i_m} = m$

Constructing $\mathcal{R}_i$ with

$$r_i = \frac{m}{k-i}\log\left(\frac{1}{\epsilon}\right)$$



Reduced search space

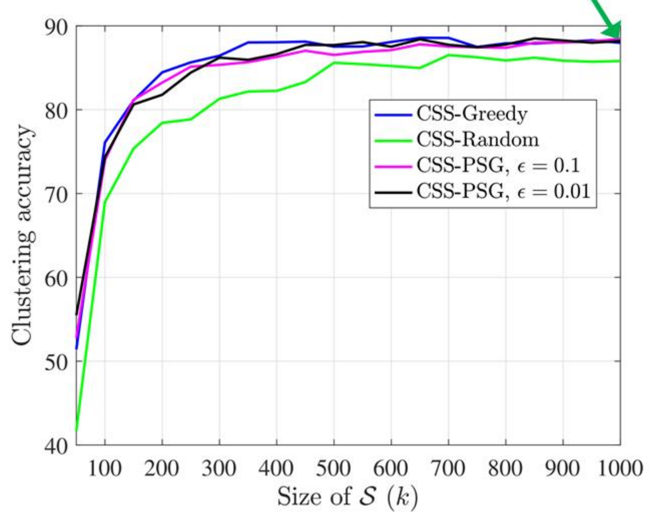
Full space

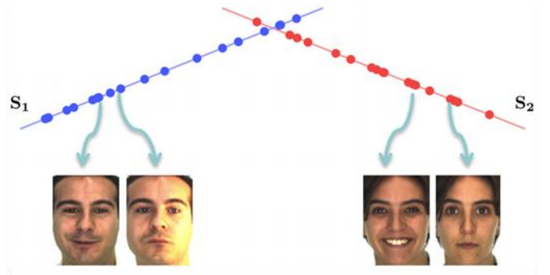
Next iteration

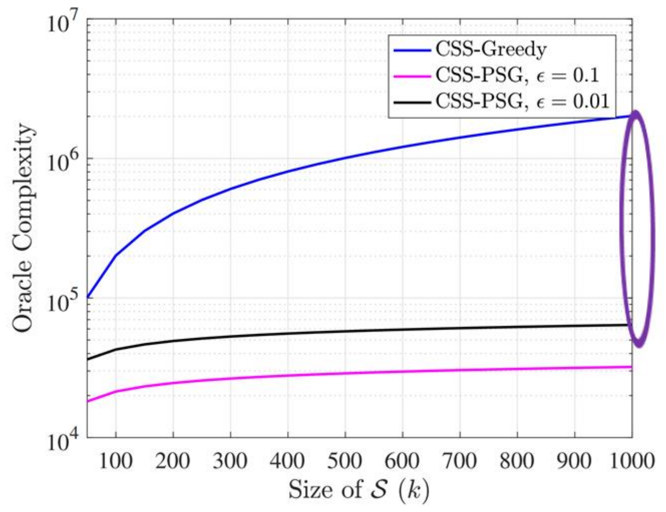
Greedy selection

RESULTS

Almost zero performance gap







Order of magnitude faster

EXPECTED PERFORMANCE

Theorem 2

Let $\mathcal{S}_{psg}$ denote the random subset selected by PSG. Then,

$$\mathbb{E}[f(\mathcal{S}_{psg})] \ge \left(1 - e^{-\alpha} - \alpha\epsilon^{\eta}\right) f(\mathcal{S}^\star),$$

where $\eta = 1 + \mathcal{O}(1/k)$.

CONCLUSION

- Asymptotic conditions for identification of the optimal subset in stochastic greedy weak submodular maximization
- A fixed schedule fails, an increasing schedule for high success probability
- PSG: a new scheme with near-optimal expected approximation factor
- **Future Directions**
- High probability guarantees vs expected guarantees
- Extensions to continuous weak submodular functions