INSTITUTE FOR COMPUTATIONAL ENGINEERING & SCIENCES ODEN

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

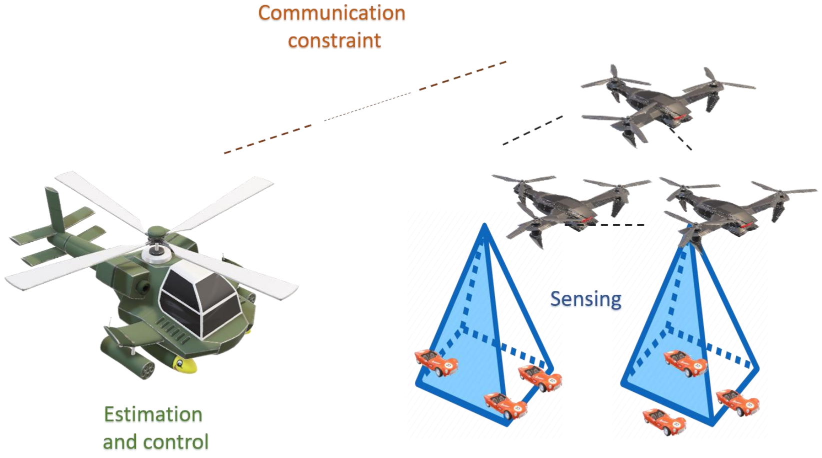
On the Performance-Complexity Tradeoff in Stochastic Greedy Weak Submodular Optimization

Abolfazl Hashemi, Haris Vikalo, Gustavo de Veciana

International Conference on Acoustics, Speech and Signal Processing (ICASSP)
6-11 June, 2021

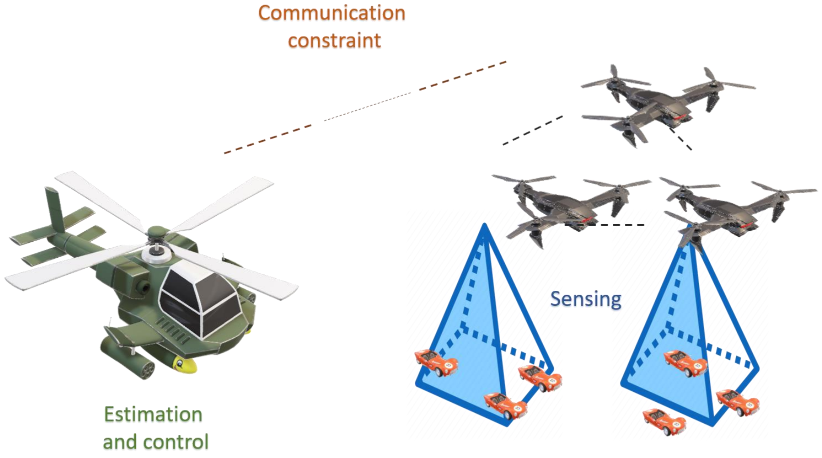
Supported in part by NSF ECCS grant 1809327

Subset Selection and Information Gathering



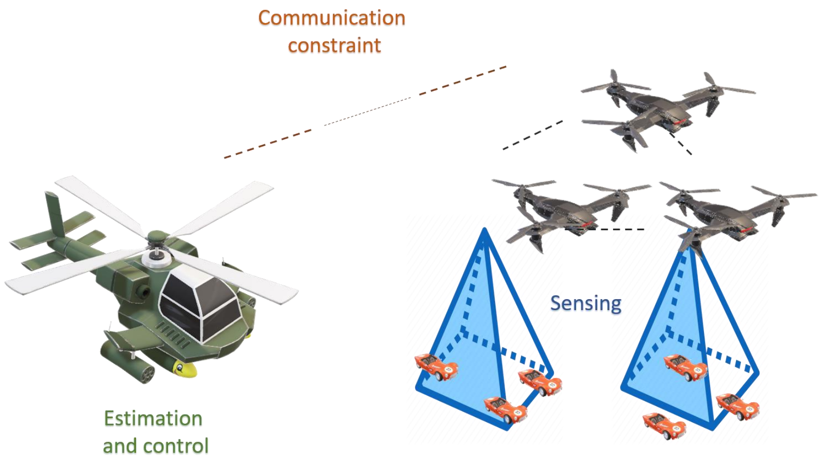
Subset Selection and Information Gathering

Large amount of high-dimensional data



Subset Selection and Information Gathering

Large amount of high-dimensional data



Resource-constrained decision making

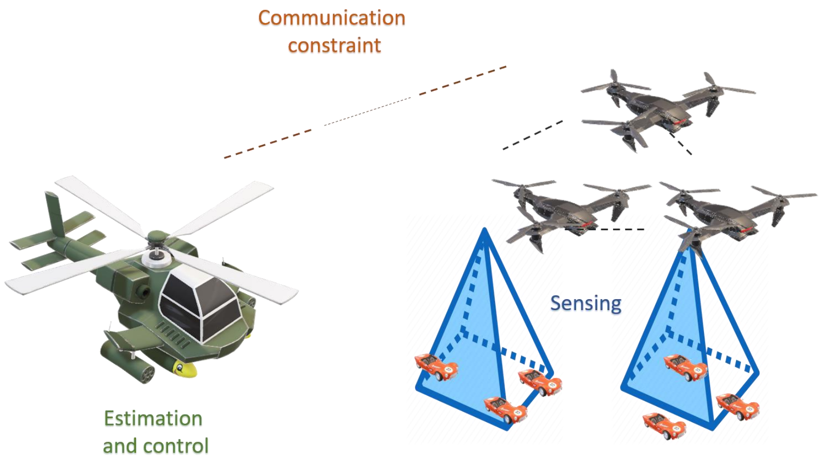
Krause et al., 2013

Subset Selection and Information Gathering

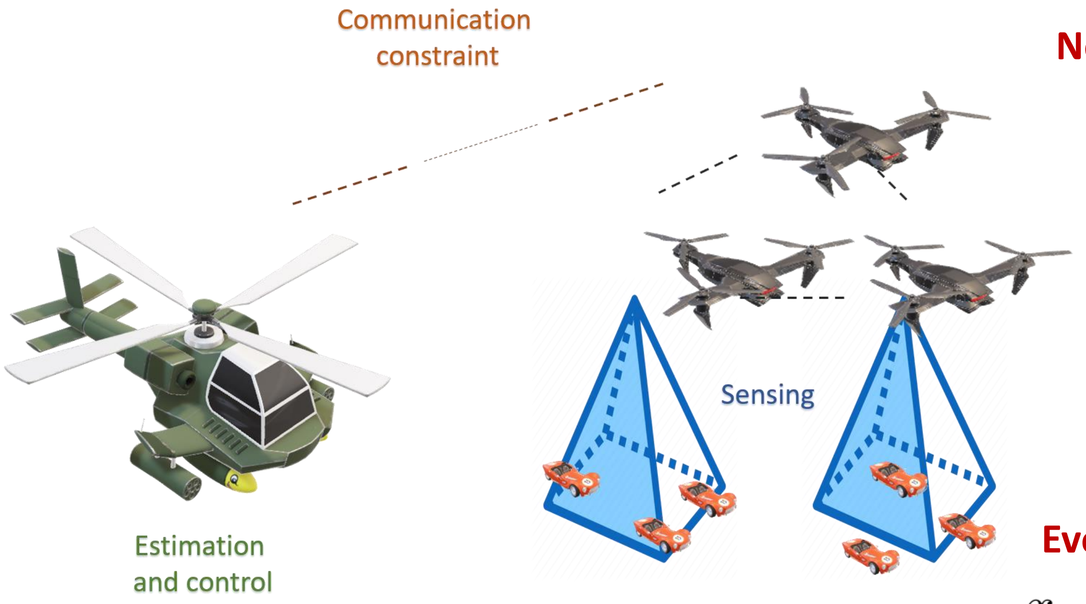
Large amount of high-dimensional data

Resource-constrained decision making

NP-hard, combinatorial optimization tasks



Subset Selection for Sensing Networks



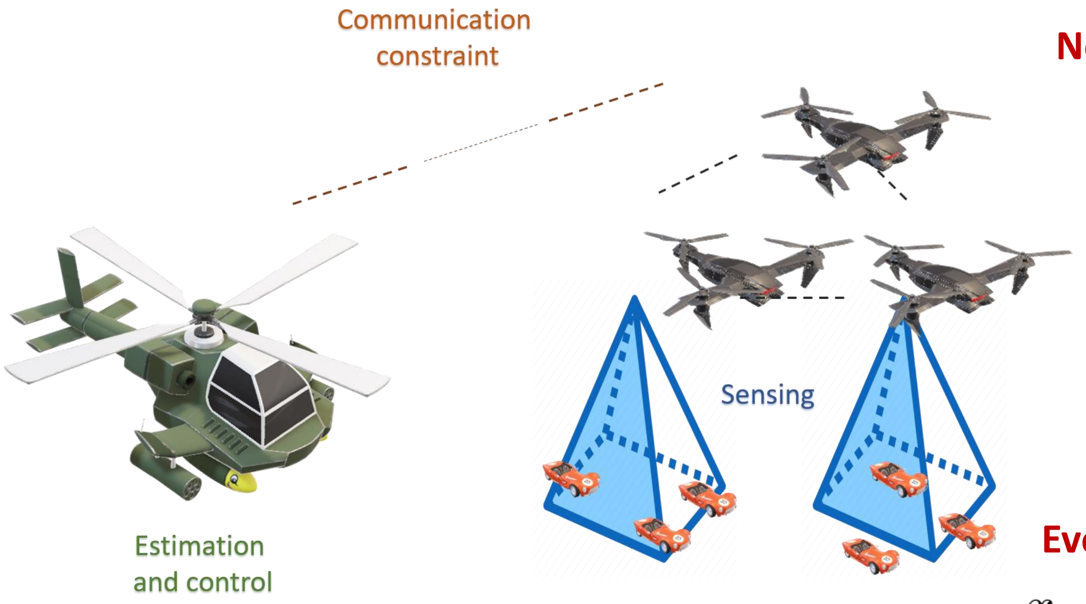
Noisy and time-varying observation model

$$y_t = h(x_t) + \nu_t$$

Evolving unknown states

$$x_t = g(x_{t-1}, u_t) + v_t$$

Subset Selection for Sensing Networks



Noisy and time-varying observation model

$$y_t = h(x_t) + \nu_t$$

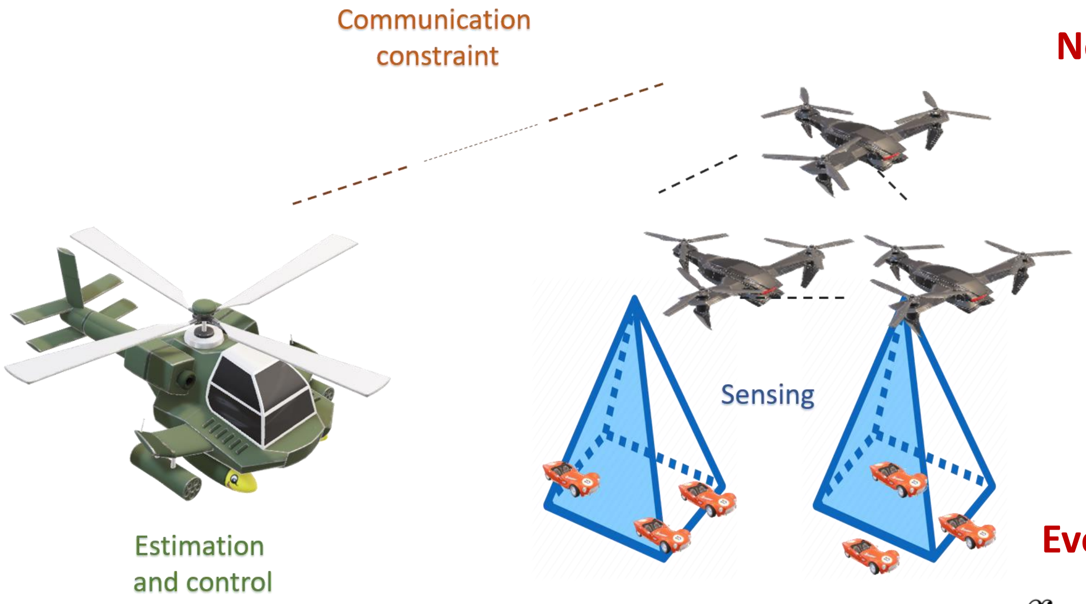
Evolving unknown states

$$x_t = g(x_{t-1}, u_t) + v_t$$

Observation selection and sharing for low-MSE estimation

$$\mathrm{MSE}_t(S) = \|\hat{x}_t(S) - x_t\|^2$$

Subset Selection for Sensing Networks



Noisy and time-varying observation model

$$y_t = h(x_t) + \nu_t$$

Evolving unknown states

$$x_t = g(x_{t-1}, u_t) + v_t$$

Observation selection and sharing for low-MSE estimation

$$\mathrm{MSE}_t(S) = \|\hat{x}_t(S) - x_t\|^2$$

<u>Question</u>

- How can we perform the selection efficiently and with guaranteed performance?

Observation Selection Criteria

Scalar functions of the predicted error covariance matrix $f(\mathrm{P}_\mathrm{t}(\mathrm{S}))$

Observation Selection Criteria

Scalar functions of the predicted error covariance matrix $f(\mathrm{P_t}(\mathrm{S}))$

Constrained combinatorial optimization

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

NP-hard

Krause, 2011

Observation Selection Criteria

Scalar functions of the predicted error covariance matrix $f(\mathrm{P_t(S)})$

Constrained combinatorial optimization

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

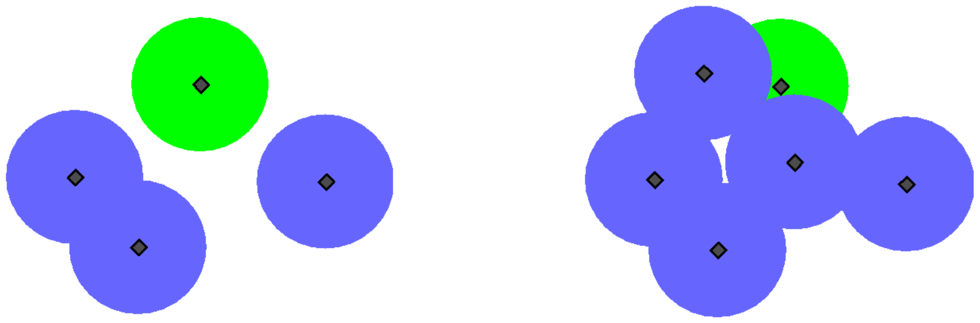
NP-hard

Krause, 2011

f(S) has nice properties:

- Monotonicity
- Weak-submodularity

Hashemi et al., 2018



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

Observation Selection Criteria

Scalar functions of the predicted error covariance matrix $f\left(\mathrm{P}_t(\mathrm{S})\right)$

Constrained combinatorial optimization

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

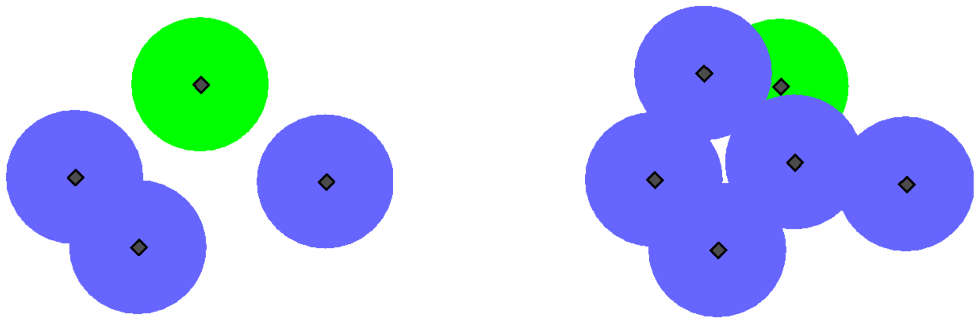
NP-hard

Krause, 2011

f(S) has nice properties:

- Monotonicity
- Weak-submodularity

Hashemi et al., 2018



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

Approximate greedy solution

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

$$f(\hat{S}) \geq \left(1 - e^{-\alpha}\right) f(S^\star)$$

Optimal approximation guarantee

Krause, 2011

Subset Selection in Large-Scale Settings

Greedy selection → Tight approximation guarantee

Greedy selection → Prohibitive computational cost

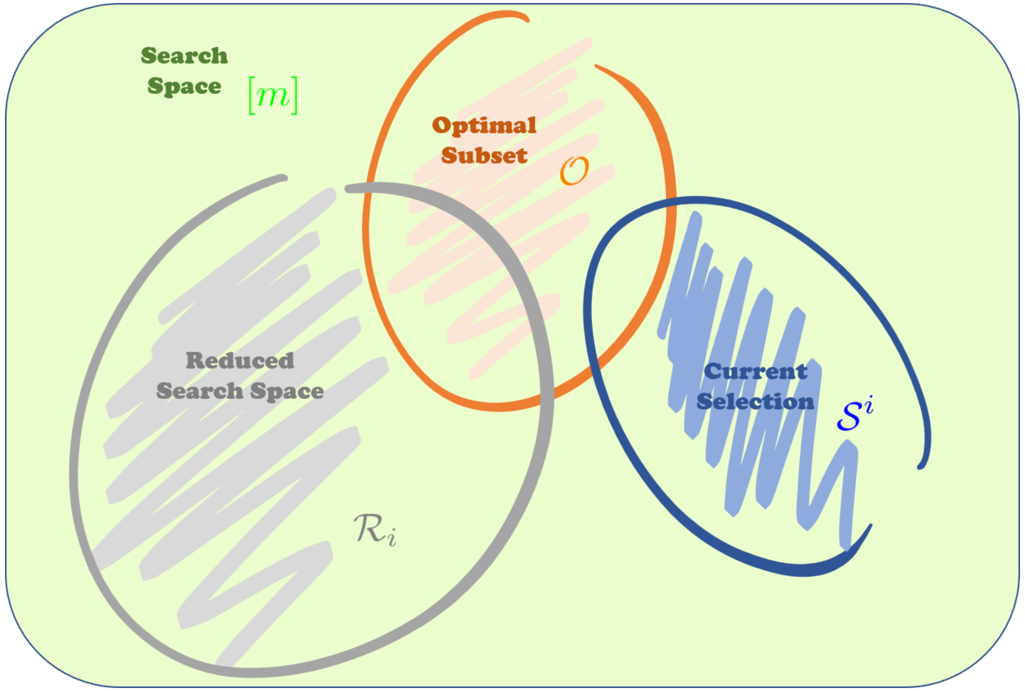
Subset Selection in Large-Scale Settings

Greedy selection → Tight approximation guarantee

Greedy selection → Prohibitive computational cost

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

Mirzasoleyman et al. 2015,
Hashemi et al. 2018



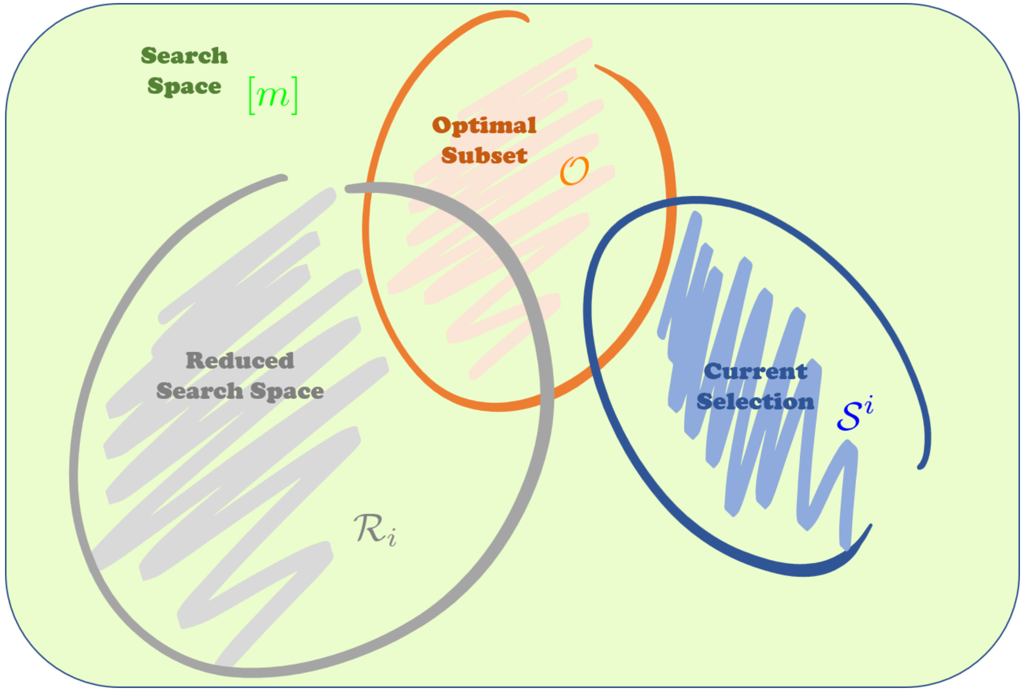
Subset Selection in Large-Scale Settings

Greedy selection → Tight approximation guarantee

Greedy selection → Prohibitive computational cost

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

Mirzasoleyman et al. 2015,
Hashemi et al. 2018



How to construct R_i

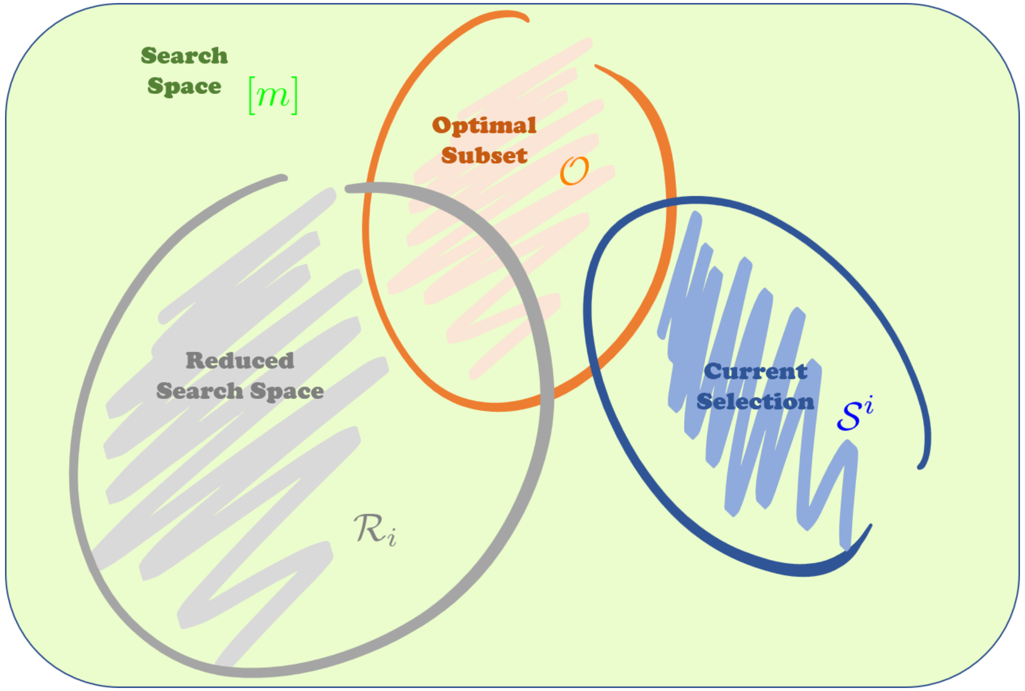
Subset Selection in Large-Scale Settings

Greedy selection → Tight approximation guarantee

Greedy selection → 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 $\frac{r}{m} \leq k^{\beta - 1}$, then

$$\limsup_{m,k\to\infty} \Pr\left(\mathcal{S}_{alg}^k = \mathcal{S}^\star\right) = 0.$$

2. Relative regime: $\beta_1 \in (0, 1)$ such that $\frac{r}{m} \leq \beta_1$, then

$$0 < \delta_1 < \limsup_{m,k\to\infty} \Pr\left(\mathcal{S}_{alg}^k = \mathcal{S}^\star\right) < \delta_2 < 0.63.$$

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 $\frac{r}{m} \leq k^{\beta - 1}$, then

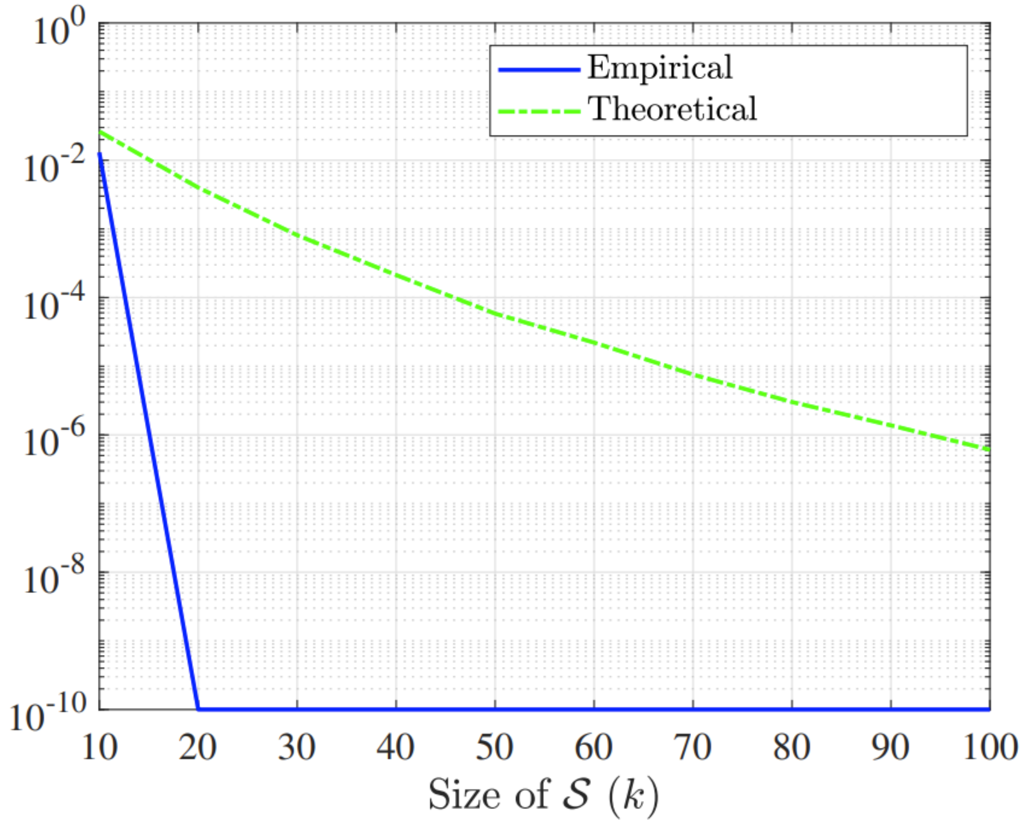
$$\limsup_{m,k\to\infty} \Pr\left(\mathcal{S}^k_{alg} = \mathcal{S}^\star\right) = 0.$$

2. Relative regime: $\beta_1 \in (0, 1)$ such that $\frac{r}{m} \leq \beta_1$, then

$$0 < \delta_1 < \limsup_{m,k\to\infty} \Pr\left(\mathcal{S}^k_{alg} = \mathcal{S}^\star\right) < \delta_2 < 0.63.$$

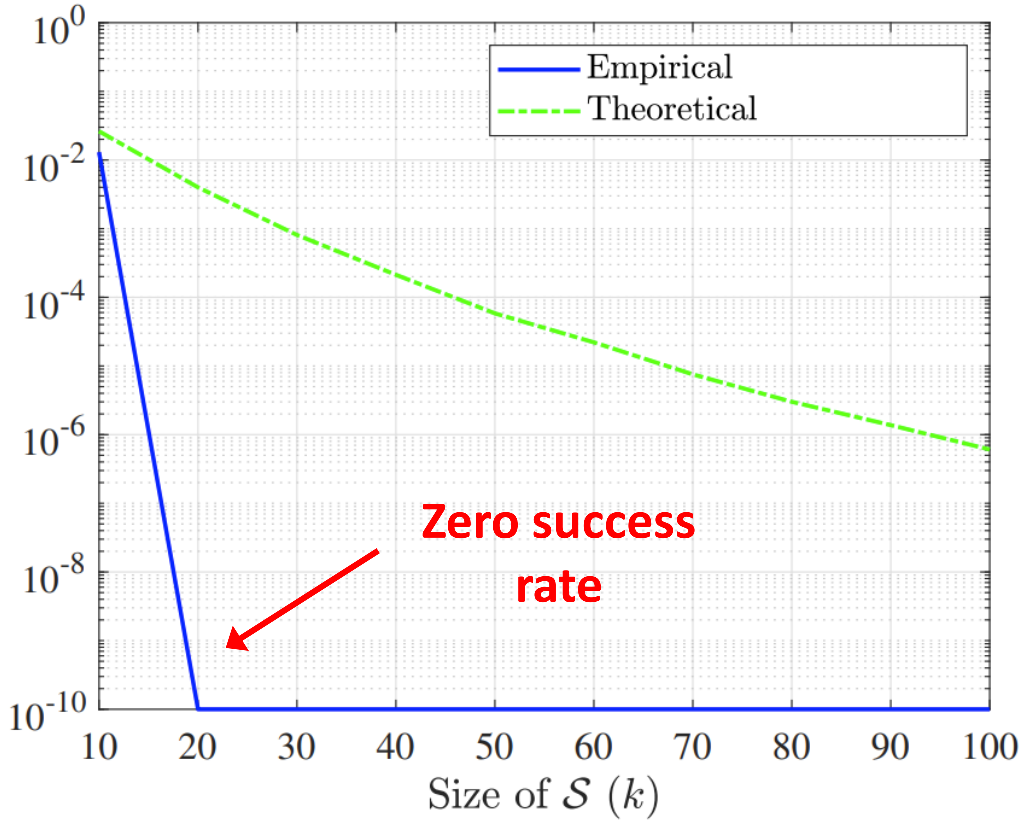
Fixed schedules fail!

Empirical Evaluation: Sparse Linear Regression



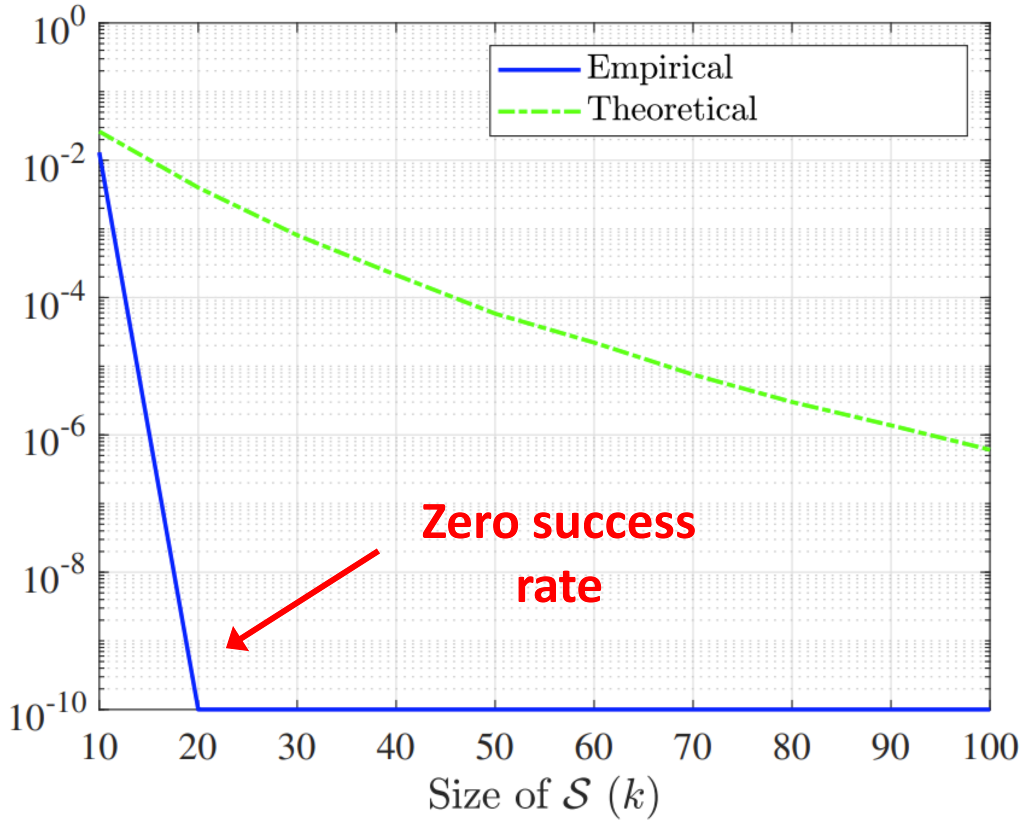
$$r = m/\sqrt{k}$$

Empirical Evaluation: Sparse Linear Regression

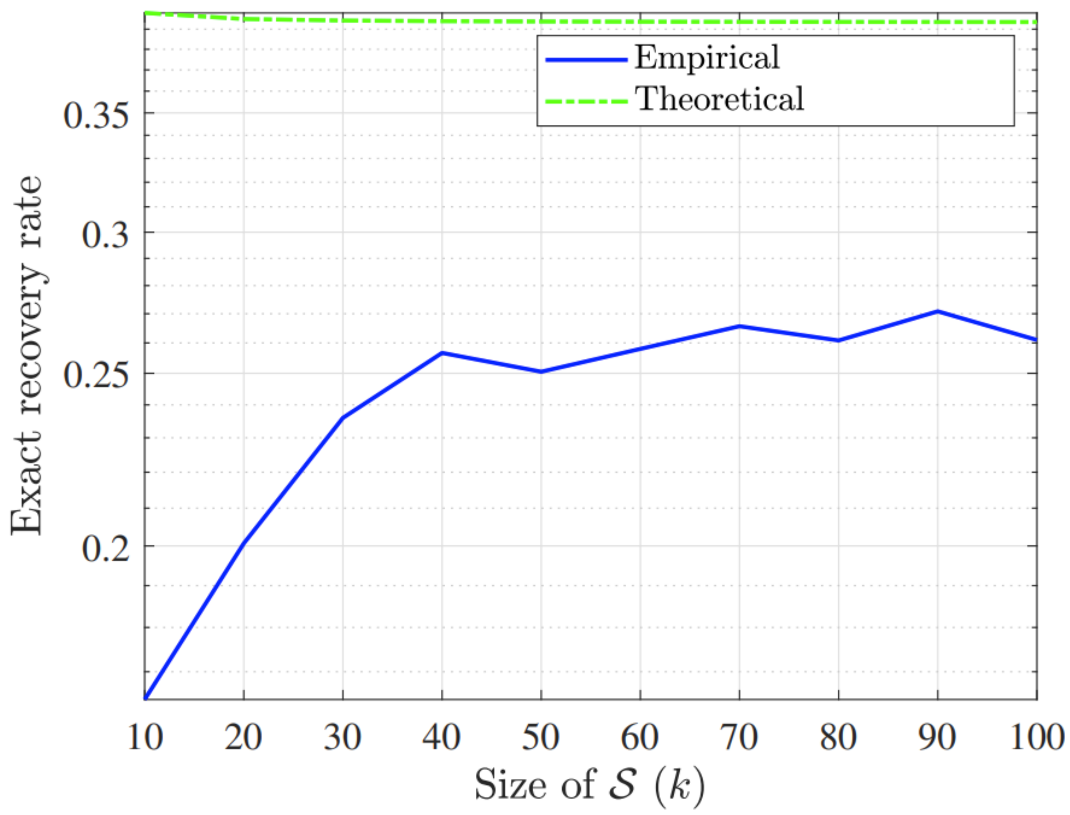


$$r = m/\sqrt{k}$$

Empirical Evaluation: Sparse Linear Regression

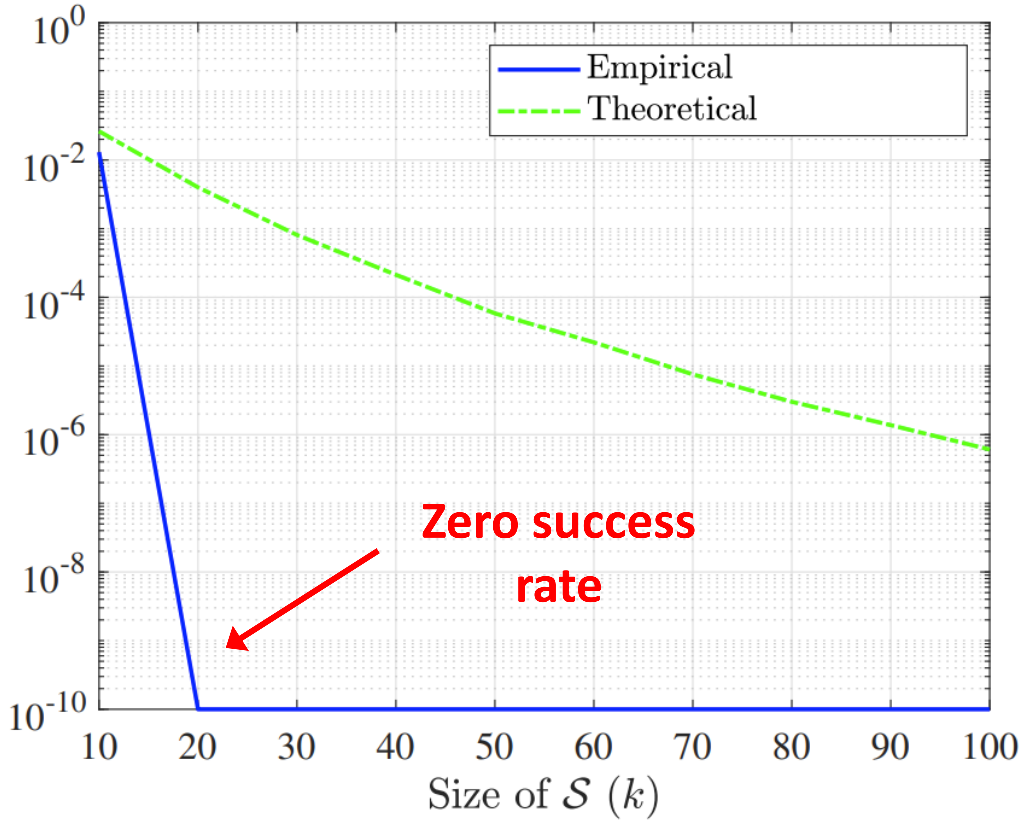


$r = m/\sqrt{k}$

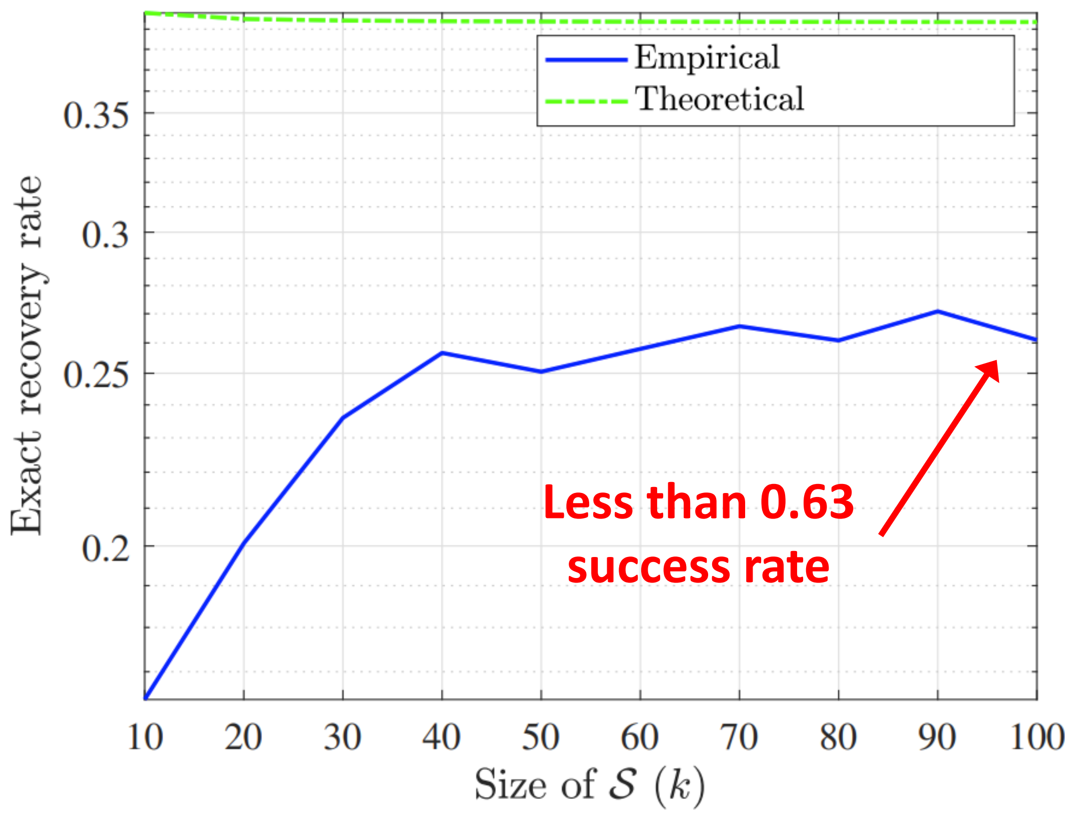


$r = m/2$

Empirical Evaluation: Sparse Linear Regression

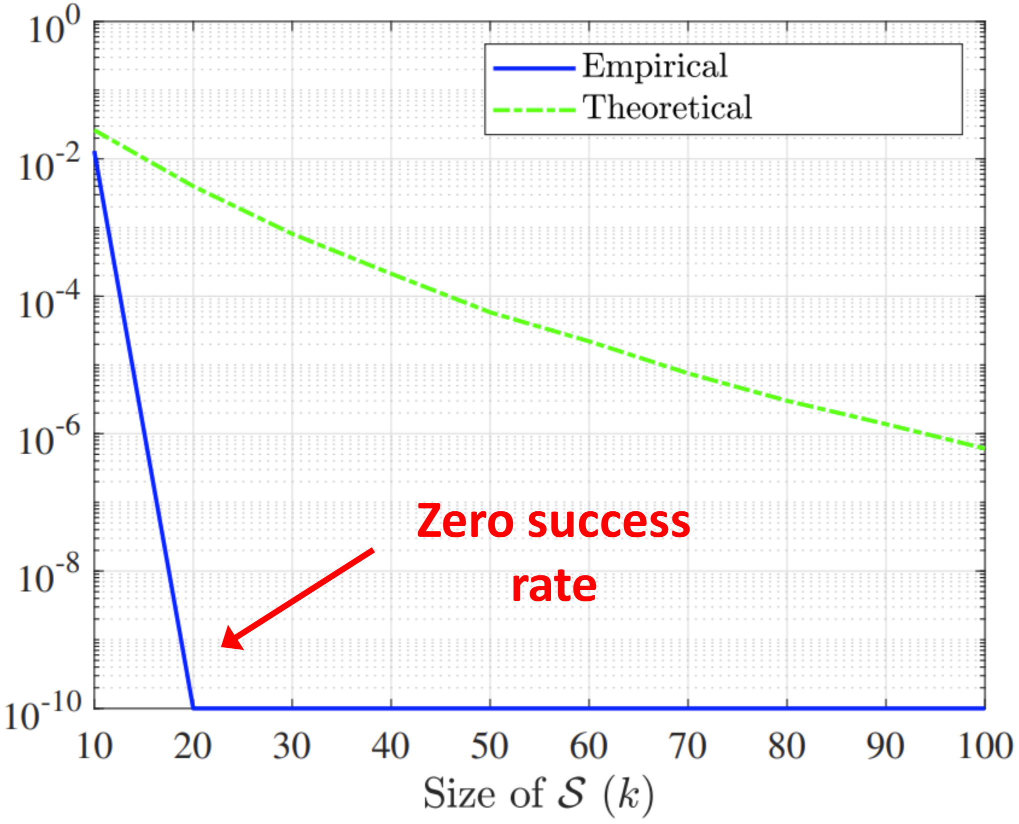


$r = m/\sqrt{k}$

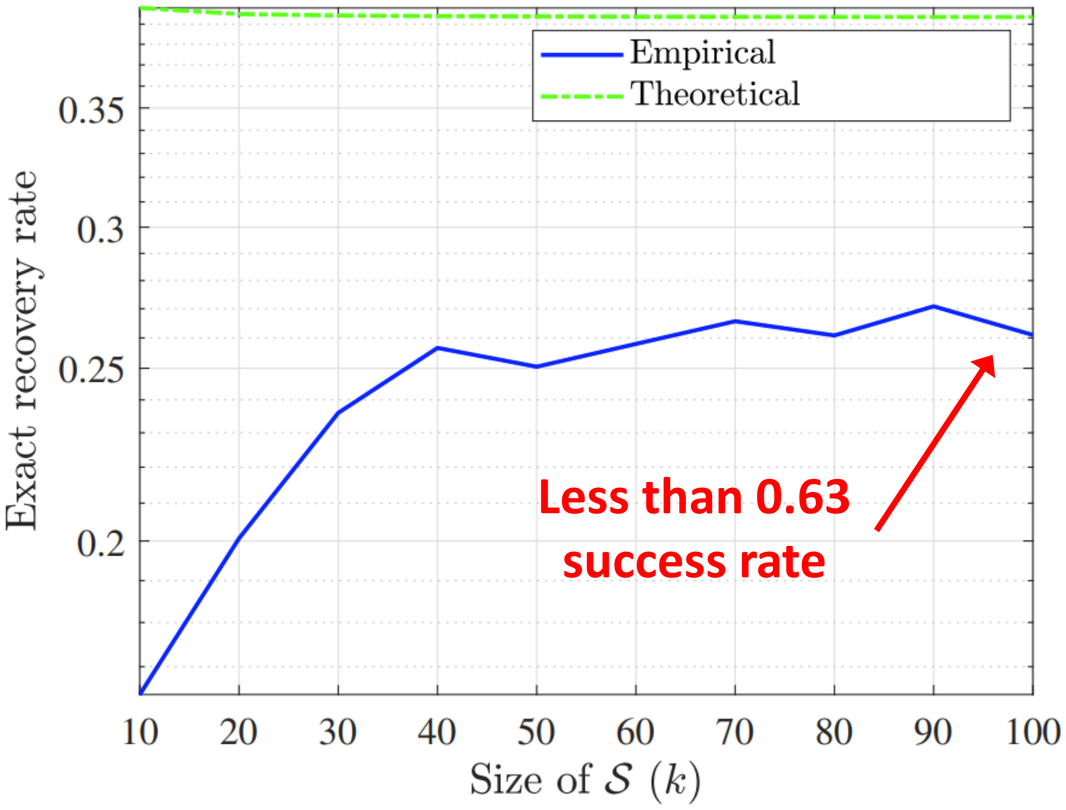


$r = m/2$

Empirical Evaluation: Sparse Linear Regression



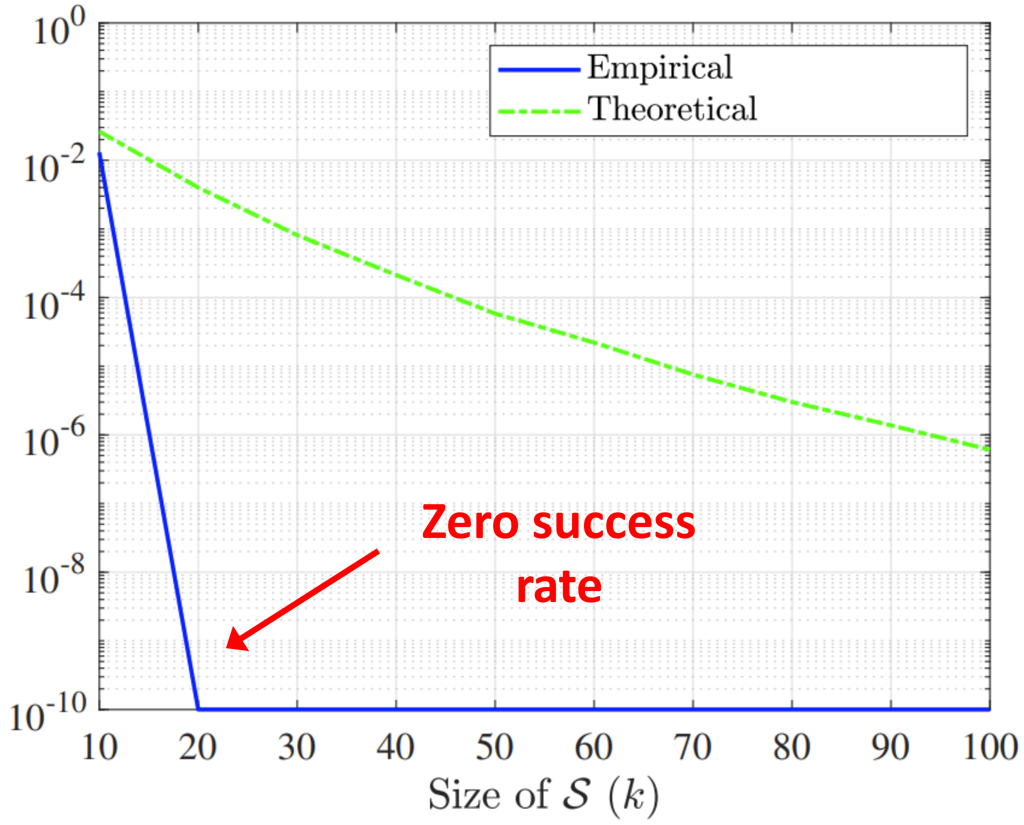
$r = m/\sqrt{k}$

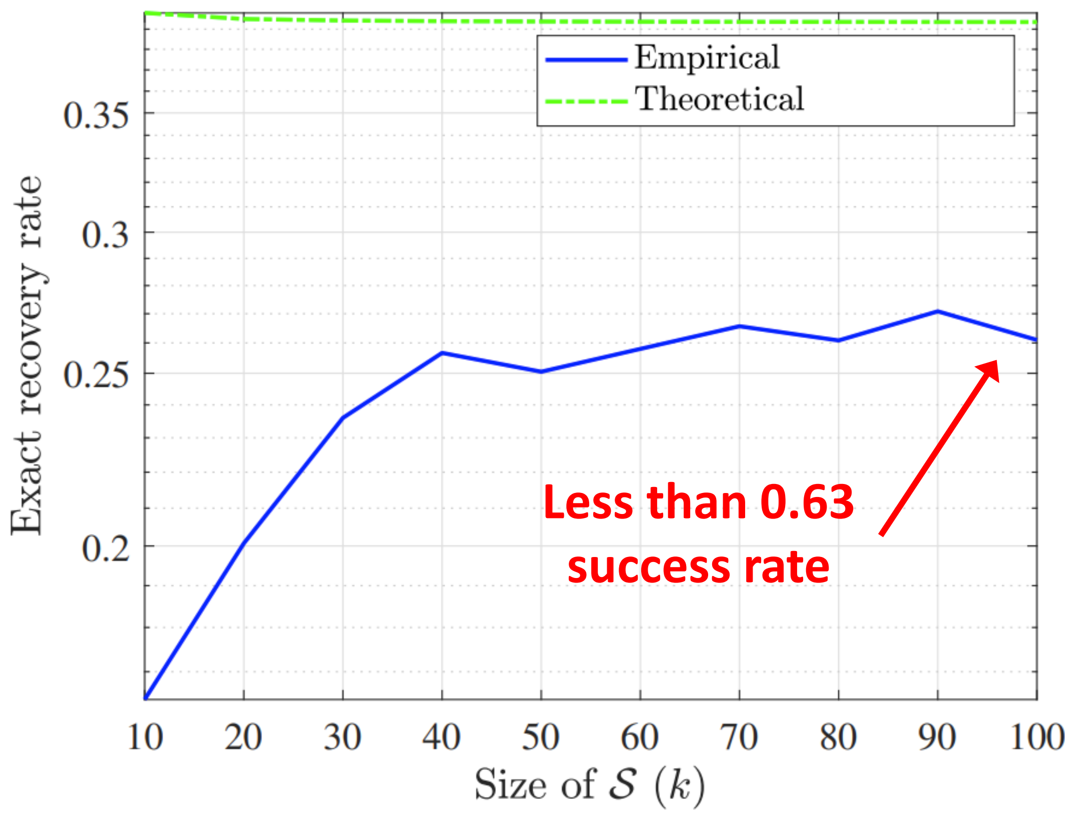


$r = m/2$

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

Empirical Evaluation: Sparse Linear Regression

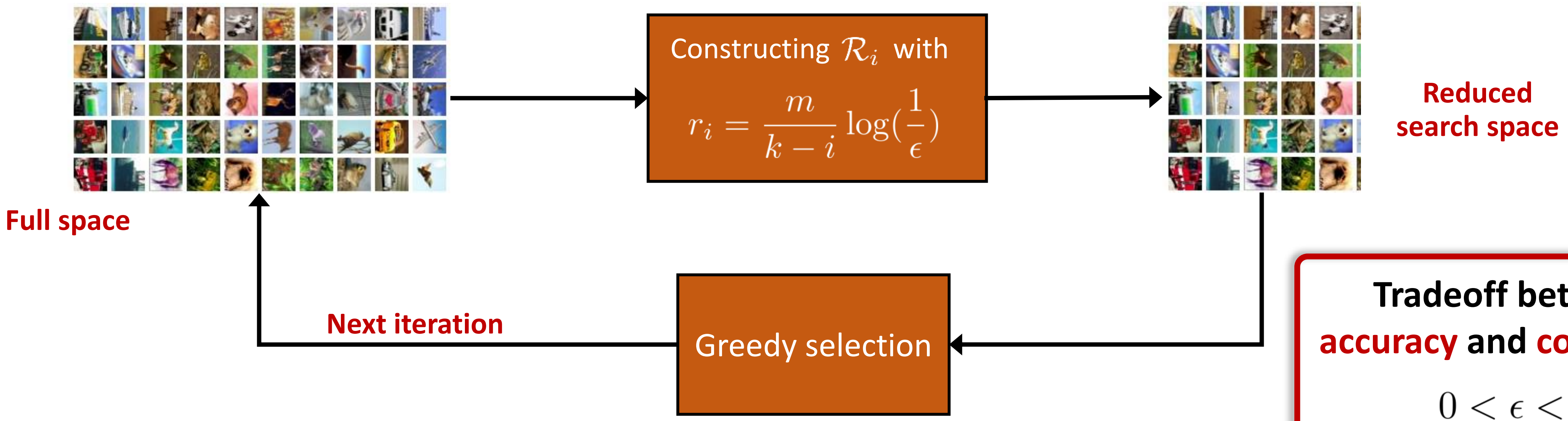




$r = m/\sqrt{k}$ $\qquad$ $r = m/2$

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

Progressive Stochastic Greedy



Tradeoff between accuracy and complexity

$0 < \epsilon < 1$

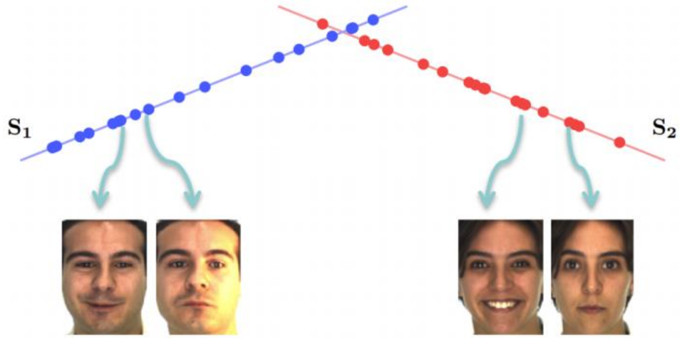
Theorem 2

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

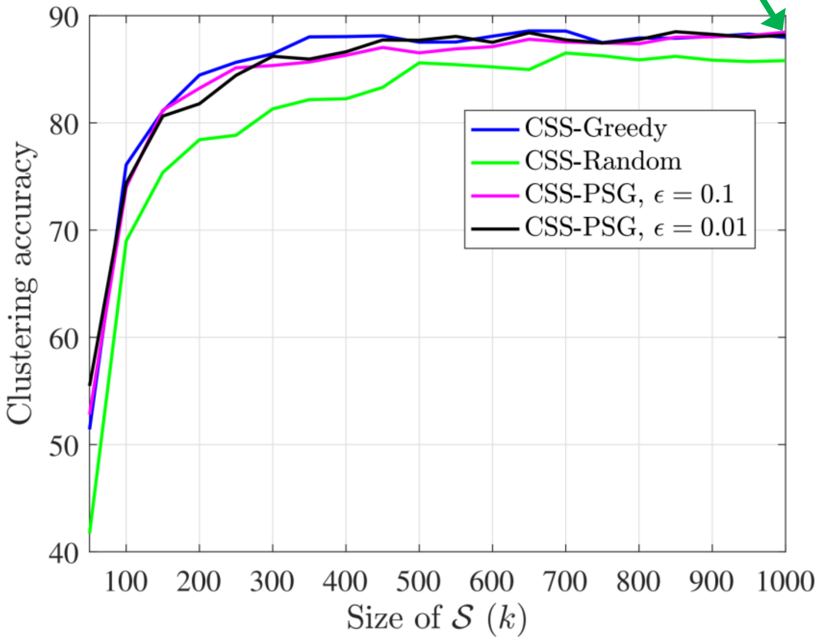
$$\mathbb{E}[f(\mathcal{S}_{psg})] \geq \left(1 - e^{-\alpha} - \alpha\epsilon^{\eta}\right) f(\mathcal{S}^{\star}),$$

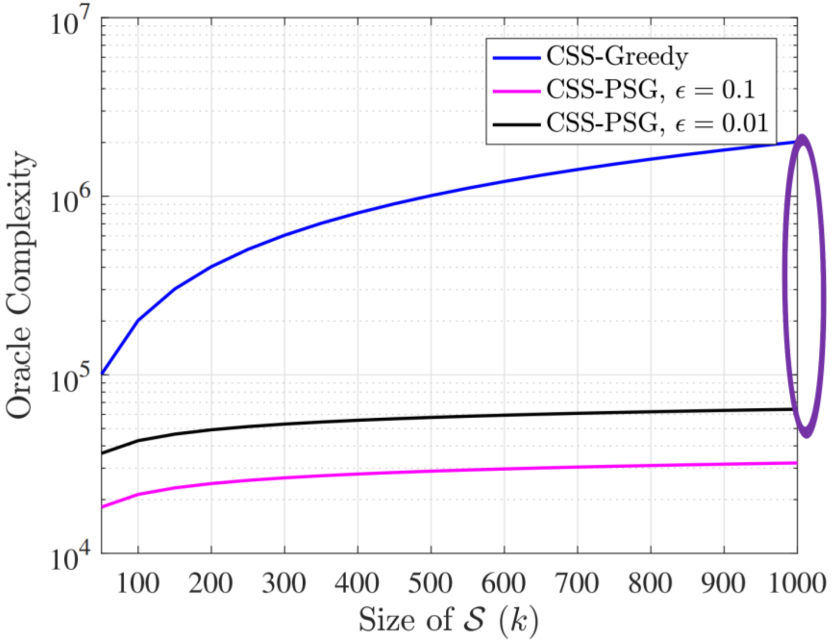
where $\eta = 1 + \mathcal{O}(1/k)$.

Applications in Clustering



Almost zero performance gap





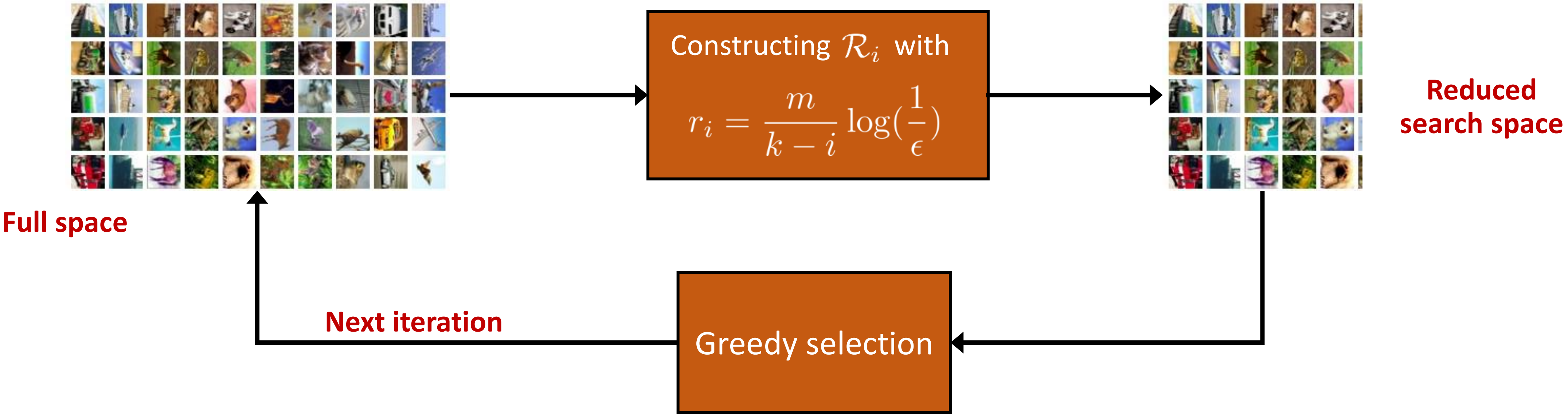
Order of magnitude faster

Conclusion and Future Work

- 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



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

Abolfazl Hashemi, Haris Vikalo, Gustavo de Veciana

Supported in part by NSF ECCS grant 1809327