Submodular Optimization with Submodular Cover and Submodular Knapsack Constraints

Rishabh Iyer and Jeff Bilmes

University of Washington, Seattle

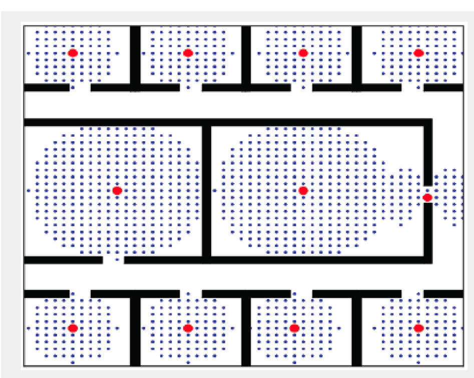
Overview

- Introduce two new problems:

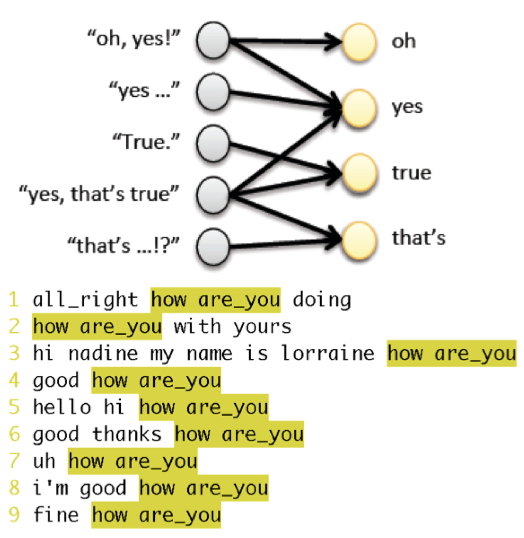
$$\text{(SCSC): } \min\{f(X) \mid g(X) \geq c\} \quad (1) \qquad \text{(SCSK): } \max\{g(X) \mid f(X) \leq b\} \quad (2)$$

- Formally show how they are closely related.
- Provide an algorithmic framework subsuming many common algorithms.
- Scalable approximation algorithms and hardness results.

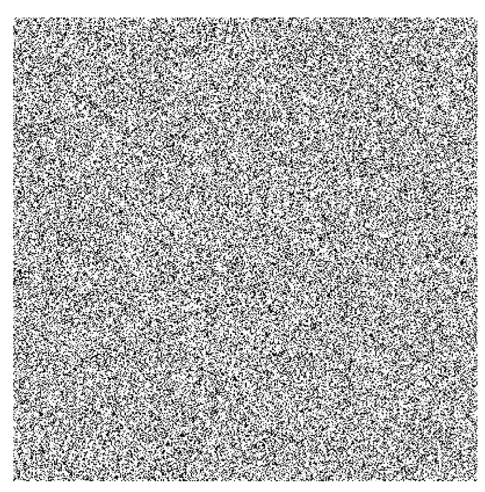
Motivation



Sensor Placement with Submodular Costs (I-Bilmes'12)



Limited vocabulary and accoustically diverse speech corpus selection (Lin-Bilmes'11, Wei et al'13)



Privacy preserving communication (I-Bilmes'13)

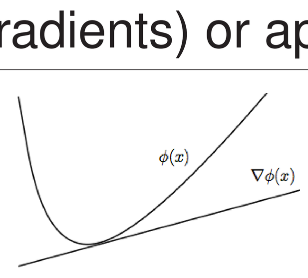
Algorithmic Framework

Algorithm 1 General algorithmic framework for Problems 1 and 2

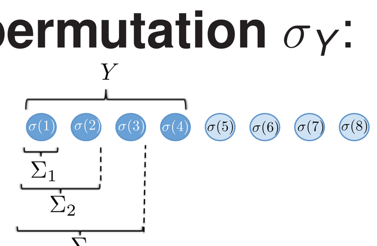
1: **for** $t = 1, 2, \cdots, T$ **do**
2: Choose surrogate functions $\hat{f}_t$ and $\hat{g}_t$ for f and g respectively, tight at X^{t-1}.
3: Obtain X^t as the optimizer of Problem 1 or 2 with $\hat{f}_t$ and $\hat{g}_t$ instead of f and g.
4: **end for**

- The Algorithm monotonically improves objective at every iteration.
- Surrogate functions are modular upper bounds (super-gradients), modular lower bounds (sub-gradients) or approximations.

Subgradients:



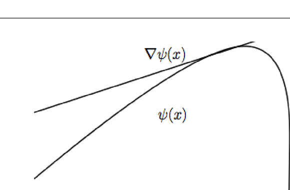
- Akin to convexity.
- Denote a **permutation** σ_Y:



- $h_Y(\sigma_Y(i)) = f(\Sigma_i) - f(\Sigma_{i-1})$
- Modular Lower bound:

$$m_{h_Y}(X) = f(Y) + h_Y(X) - h_Y(Y) \leq f(X)$$

Supergradients:



- Akin to concavity.
- Three specific supergradients:

$$\underbrace{\begin{aligned}\hat{g}_Y(j) &= f(j|V\backslash\{j\}) \\ \check{g}_Y(j) &= f(j|Y\backslash\{j\}) \\ \bar{g}_Y(j) &= f(j|V\backslash\{j\})\end{aligned}}_{\text{for } j \in Y} \qquad \underbrace{\begin{aligned}\hat{g}_Y(j) &= f(j|Y) \\ \check{g}_Y(j) &= f(j|\emptyset) \\ \bar{g}_Y(j) &= f(j|\emptyset)\end{aligned}}_{\text{for } j \notin Y.}$$

- Modular Upper bound:

$$m^{g_Y}(X) = f(Y) + g_Y(X) - g_Y(Y) \geq f(X).$$

- The Ellipsoidal approximation (Goemans et al, 2009; Iyer et al, 2013) provides the tightest bounds for these problems (though not practical).
- Define the **curvature** of a monotone submodular function κ_f as:

$$\kappa_f = 1 - \min_{j \in V} \frac{f(j|V\backslash j)}{f(j)} \qquad (1)$$

Relation between SCSC and SCSK

- **Bi-criterion guarantees:** $[\sigma, \rho]$ approx. for (1) $\implies$ a set $X : f(X) \leq \sigma f(X^*)$ and $g(X) \geq \rho c$. Similarly a $[\rho, \sigma]$ approx. for (2) $\implies$ a set $X : g(X) \geq \rho g(X^*)$ and $f(X) \leq \sigma b$.

Algorithm 2 Approx. algo. for SCSK using an approx. alg. for SCSC

1: **Input:** An SCSK instance with budget b, $[\sigma, \rho]$ approx. SCSC, $\epsilon \in [0, 1)$.
2: **Output:** $[(1-\epsilon)\rho, \sigma]$ approx. for SCSK.
3: $c \leftarrow g(V)$, $\hat{X}_c \leftarrow V$.
4: **while** $f(\hat{X}_c) > \sigma b$ **do**
5: $c \leftarrow (1-\epsilon)c$
6: $\hat{X}_c \leftarrow [\sigma, \rho]$ approx. for SCSC using c.
7: **end while**
8: Return $\hat{X}_c$

Algorithm 3 Approx. algo for SCSC using an approx. alg. for SCSK.

1: **Input:** An SCSC instance with cover c, $[\rho, \sigma]$ approx. SCSK, $\epsilon > 0$.
2: **Output:** $[(1+\epsilon)\sigma, \rho]$ approx. for SCSC.
3: $b \leftarrow \text{argmin}_j f(j)$, $\hat{X}_b \leftarrow \emptyset$.
4: **while** $g(\hat{X}_b) < \rho c$ **do**
5: $b \leftarrow (1+\epsilon)b$
6: $\hat{X}_b \leftarrow [\rho, \sigma]$ approx. for SCSK using b.
7: **end while**
8: Return $\hat{X}_b$.

Approx. Algorithms for SCSC

- **Submodular Set Cover (SSC):**
  - Special case of SCSC with f modular and g submodular.
  - Lemma: The greedy algorithm for SSC is special case of Algorithm 1 with g replaced by its modular lower bound (Approx. factor $\approx 1 + \log g(V)$).
  - Dual SSC: Obtain a bicriterion approximate solution for SSC using Algorithm 3 and Submodular Knapsack.
  - Lemma: Dual SSC obtains a $[1+\epsilon, 1 - 1/e]$ Bi-criterion Approximation.
- **Iterative Submodular Set Cover (ISSC):**
  - Choose surrogate functions $\hat{f}_t$ as modular upper bounds (supergradients).
  - Iteratively solve SSC.
  - Theorem: ISSC obtains an approximation factor of $\frac{K_g H_g}{1+(K_g-1)(1-\kappa_f)}$ where $K_g = 1 + \max\{|X| : g(X) < c\}$ and H_g is the approx. factor of SSC using g.
- **Ellipsoidal Approx. based Submodular Set Cover (EASSC):**
  - Choose surrogate functions $\hat{f}_t$ as Ellipsoidal Approximation.
  - Iteratively solve SSC.
  - Theorem: EASSC obtains an approximation factor of $O(\frac{\sqrt{n}\log n H_g}{1+(\sqrt{n}\log n - 1)(1-\kappa_f)})$ where H_g is the approximation factor of SSC using g.
  - A much simpler non-iterative algorithm acheives a factor of $O(\sqrt{n}\log n\sqrt{H_g})$.

Approx. Algorithms for SCSK

- **Submodular Knapsack (SK):**
  - Special case of SCSK with f modular and g submodular.
  - Lemma: The greedy algorithm for SK is special case of Algorithm 1 with g replaced by its modular lower bound (Approx. factor $1 - 1/e$).
- **Greedy (Gr):**
  - A simple greedy algorithm (can be seen as special case of Algorithm 1).
  - Lemma: Gr obtains a worst case guarantee of $\frac{1}{\kappa_g}(1 - (\frac{K_f - \kappa_g}{K_f})^{k_f}) \geq \frac{1}{K_f}$, where $K_f = \max\{|X| : f(X) \leq b\}$ and $k_f = \min\{|X| : f(X) \leq b \;\&\; \forall j \in X, f(X \cup j) > b\}$.
- **Iterative Submodular Knapsack (ISSC):**
  - Choose surrogate functions $\hat{f}_t$ as modular upper bounds (supergradients).
  - Iteratively solve SK.
  - Theorem: ISK obtains a bicriterion approximation factor of $[1 - e^{-1}, \frac{K_f}{1+(K_f-1)(1-\kappa_f)}]$ where $K_f = \max\{|X| : f(X) \leq b\}$.
- **Ellipsoidal Approx. based Submodular Knapsack (EASK):**
  - Choose surrogate functions $\hat{f}_t$ as Ellipsoidal Approximation.
  - This is based on looking at its dual problem.
  - Approximation guarantee similar to EASSC.

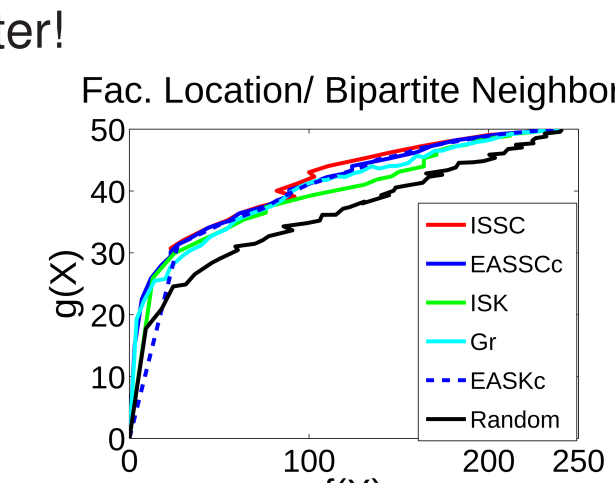
Hardness

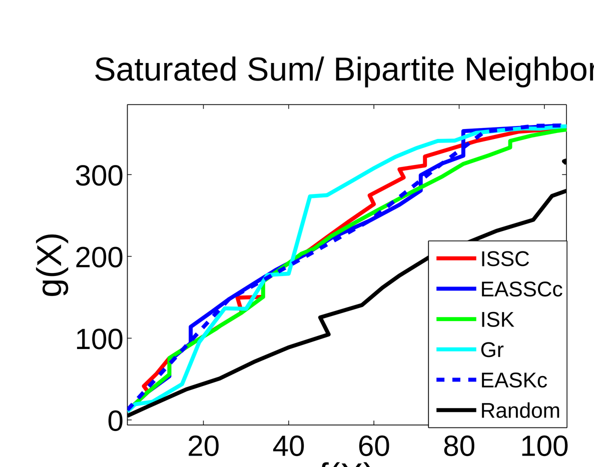
- We can show matching lower bounds:

Theorem: For any $\kappa > 0$, there exists submodular functions with curvature κ such that no polynomial time algorithm for Problems 1 and 2 achieves a bi-criterion factor better than $\frac{\sigma}{\rho} = \frac{n^{1/2-\epsilon}}{1+(n^{1/2-\epsilon}-1)(1-\kappa)}$ for any $\epsilon > 0$.

Experiments

- Compare our algorithms on data subset selection on TIMIT corpus.
- f as a bipartite neighborhood, and g as facility location and saturated graph cut respectively.
- ISSC, ISK and Gr compare in performance to EASSC/ EASK, though they are much faster!





Acknowledgements

Based upon work supported by National Science Foundation Grant No. IIS-1162606, and by a Google, a Microsoft, and an Intel research award.