Massachusetts
Institute of
Technology

CSAIL

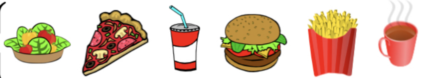
# Optimal approximation for unconstrained non-submodular minimization

Marwa El Halabi    Stefanie Jegelka

CSAIL, MIT

ICML 2020

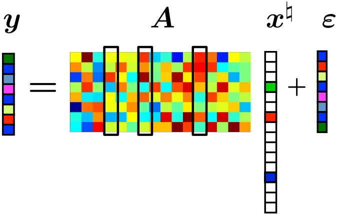
# Set function minimization

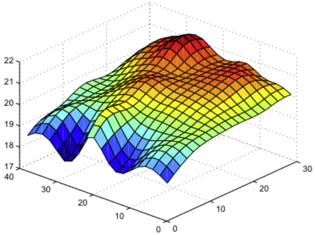
$V = \{$  $\}$

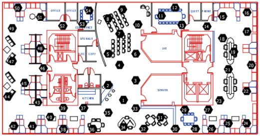
$H : 2^V \to \mathbb{R}$, $\quad H\Big($  $\Big)$ = cost of items

**Goal:** Select collection $S$ of items in $V$ that minimize cost $H(S)$

# Set function minimization in Machine learning



**Structured sparse learning** **Batch Bayesian optimization**

Figures from [Mairal et al., 2010, Krause et al., 2008]

# Set function minimization

Ground set $V = \{1, \cdots, d\}$, set function $H : 2^V \to \mathbb{R}$

$$\min_{S \subseteq V} H(S)$$

- Assume: $H(\emptyset) = 0$, black box oracle to evaluate H

# Set function minimization

Ground set $V = \{1, \cdots, d\}$, set function $H : 2^V \to \mathbb{R}$

$$\min_{S \subseteq V} H(S)$$

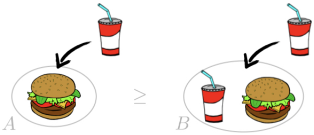
- Assume: $H(\emptyset) = 0$, black box oracle to evaluate H
- NP-hard to approximate in general

# Set function minimization

Ground set $V = \{1, \cdots, d\}$, set function $H : 2^V \to \mathbb{R}$

$$\min_{S \subseteq V} H(S)$$

- Assume: $H(\emptyset) = 0$, black box oracle to evaluate H
- NP-hard to approximate in general
- Submodularity helps: diminishing returns (DR) property



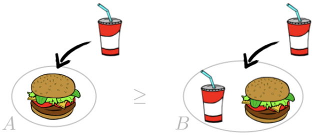

$$H(A \cup \{i\}) - H(A) \geq H(B \cup \{i\}) - H(B) \text{ for all } A \subseteq B$$

# Set function minimization

Ground set $V = \{1, \cdots, d\}$, set function $H : 2^V \to \mathbb{R}$

$$\min_{S \subseteq V} H(S)$$

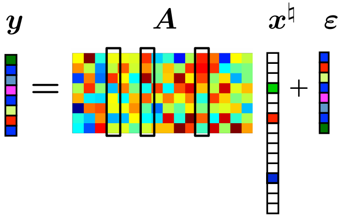
- Assume: $H(\emptyset) = 0$, black box oracle to evaluate H
- NP-hard to approximate in general
- Submodularity helps: diminishing returns (DR) property

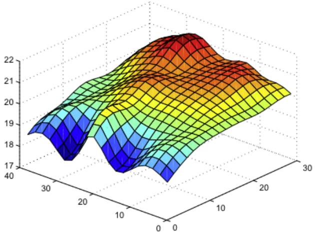



$$H(A \cup \{i\}) - H(A) \geq H(B \cup \{i\}) - H(B) \text{ for all } A \subseteq B$$

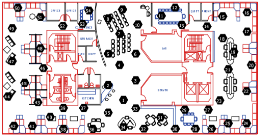
- Efficient minimization

# Set function minimization in Machine learning




**Structured sparse learning**

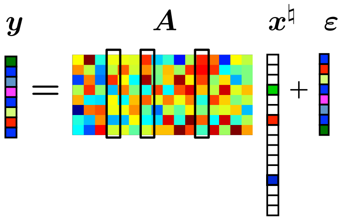


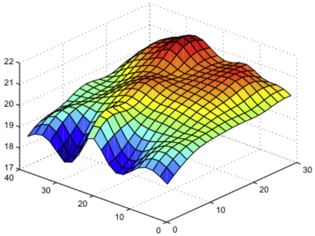
**Bayesian optimization**

**H is not submodular**

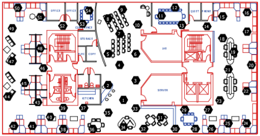
Figures from [Mairal et al., 2010, Krause et al., 2008]

# Set function minimization in Machine learning




**Structured sparse learning**



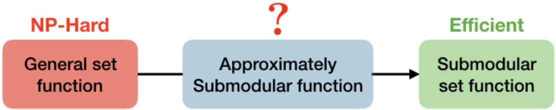
**Bayesian optimization**

**H is not submodular but it is "close" . . .**

---

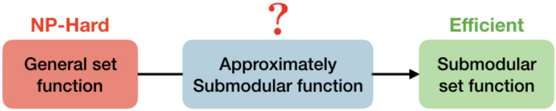
Figures from [Mairal et al., 2010, Krause et al., 2008]

# Approximately submodular functions




**What if the objective is not submodular, but "close"?**

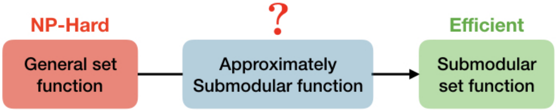
# Approximately submodular functions




**What if the objective is not submodular, but "close"?**

- Several works on non-submodular maximization [Das and Kempe, 2011, Bian et al., 2017, Kuhnle et al., 2018, Horel and Singer, 2016, Hassidim and Singer, 2018]
- Only constrained non-submodular minimization is studied [Wang et al., 2019, Bai et al., 2016, Qian et al., 2017, Sviridenko et al., 2017]

# Approximately submodular functions




Can submodular minimization algorithms extend to such non-submodular functions?

# Overview of main results

Can submodular minimization algorithms extend to such non-submodular functions? Yes!

- First approximation guarantee
- Efficient simple algorithm: Projected subgradient method
- Extension to noisy setting
- Matching lower-bound showing optimality

# Weakly DR-submodular functions

$H$ is $\alpha$-weakly DR-submodular [Lehmann et al., 2006], with $\alpha > 0$ if

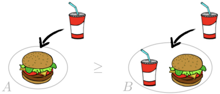
$$H(A \cup \{i\}) - H(A) \geq \alpha \Big( H(B \cup \{i\}) - H(B) \Big) \text{ for all } A \subseteq B$$

# Weakly DR-submodular functions

$H$ is $\alpha$-weakly DR-submodular [Lehmann et al., 2006], with $\alpha > 0$ if

$$H(A \cup \{i\}) - H(A) \geq \alpha\Big(H(B \cup \{i\}) - H(B)\Big) \text{ for all } A \subseteq B$$

- $H$ is submodular $\Rightarrow \alpha = 1$

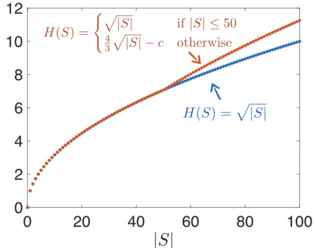


# Weakly DR-submodular functions

$H$ is $\alpha$-weakly DR-submodular [Lehmann et al., 2006], with $\alpha > 0$ if

$$H(A \cup \{i\}) - H(A) \geq \alpha\Big(H(B \cup \{i\}) - H(B)\Big) \text{ for all } A \subseteq B$$

- $H$ is submodular $\Rightarrow$ $\alpha = 1$

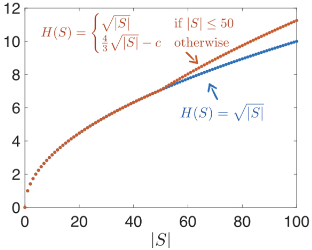


# Weakly DR-submodular functions

$H$ is $\alpha$-weakly DR-submodular [Lehmann et al., 2006], with $\alpha > 0$ if

$$H(A \cup \{i\}) - H(A) \geq \alpha\Big(H(B \cup \{i\}) - H(B)\Big) \text{ for all } A \subseteq B$$

- $H$ is submodular $\Rightarrow \alpha = 1$




- **Caveat:** $H$ should be monotone $H(A) \leq H(B) \Rightarrow \alpha \leq 1$
  $H(A) \geq H(B) \Rightarrow \alpha \geq 1$

# Problem set-up

$$\min_{S \subseteq V} H(S) := F(S) - G(S)$$

- $F$ and $G$ are both non-decreasing
- $F$ is $\alpha$-weakly DR-submodular
- $G$ is $\beta$-weakly DR-supermodular
- $F(\emptyset) = G(\emptyset) = 0$

# What set functions have this form?

$$\min_{S \subseteq V} H(S) := F(S) - G(S)$$

**Objectives in several applications:** Structured sparse learning, variance reduction in Bayesian optimization, Bayesian A-optimality in experimental design [Bian et al., 2017], column subset selection [Sviridenko et al., 2017].

# What set functions have this form?

$$\min_{S \subseteq V} H(S) := F(S) - G(S)$$

### Decomposition result

Given any set function $H$, and $\alpha, \beta \in (0, 1], \alpha\beta < 1$, we can write

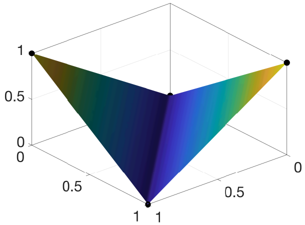
$$H(S) = F(S) - G(S)$$

- $F$ and $G$ are non-decreasing $\alpha$-weakly DR-submodular
- $G$ is $\beta$-weakly DR-supermodular

# Submodular function minimization

$$\min_{S \subseteq V} H(S) = \min_{\boldsymbol{s} \in [0,1]^d} h_L(\boldsymbol{s}) \qquad (|V| = d)$$

$h_L$ is the Lovász extension of $H$

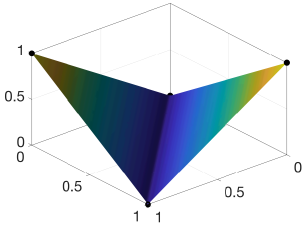


# Submodular function minimization

$$\min_{S \subseteq V} H(S) = \min_{\boldsymbol{s} \in [0,1]^d} h_L(\boldsymbol{s}) \quad (|V| = d)$$

$h_L$ is the Lovász extension of $H$

- $H$ is submodular $\Leftrightarrow$ Lovász extension is convex [Lovász, 1983]
- Easy to compute subgradients [Edmonds, 2003]: Sorting $+ \ d$ function evaluations of $H$



# Non-submodular function minimization

Can we use the same strategy?

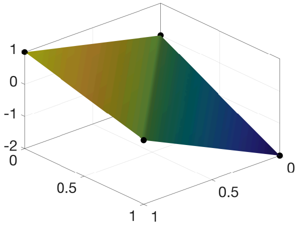
$$\min_{S \subseteq V} H(S) = \min_{\boldsymbol{s} \in [0,1]^d} h_L(\boldsymbol{s}) \quad (|V| = d)$$

# Non-submodular function minimization

Can we use the same strategy? No

$$\min_{S \subseteq V} H(S) = \min_{\boldsymbol{s} \in [0,1]^d} h_L(\boldsymbol{s}) \qquad (|V| = d)$$

- The Lovász extension $h_L$ is not convex anymore



# Non-submodular function minimization

Can we use the same strategy? Almost

$$\min_{S \subseteq V} H(S) := F(S) - G(S) = \min_{\boldsymbol{s} \in [0,1]^d} h_L(\boldsymbol{s}) := f_L(S) - g_L(S)$$

- The Lovász extension $h_L$ is not convex anymore

## Main result

- Easy to compute approximate subgradient (= subgradients in the submodular case):

$$\frac{1}{\alpha} f_L(\boldsymbol{s}') - \beta g_L(\boldsymbol{s}') \geq h_L(\boldsymbol{s}) + \langle \boldsymbol{\kappa}, \boldsymbol{s}' - \boldsymbol{s} \rangle, \forall \boldsymbol{s}' \in [0,1]^d$$

- $H$ approximately submodular $\Rightarrow h_L$ is approximately convex

# Projected subgradient method (PGM)

$$s^{t+1} = \Pi_{[0,1]^d}(s^t - \eta \kappa^t) \quad \text{(PGM)}$$

$\kappa^t$ is an approximate subgradient of $h_L$ at $s^t$

# Projected subgradient method (PGM)

$$\boldsymbol{s}^{t+1} = \Pi_{[0,1]^d}(\boldsymbol{s}^t - \eta \boldsymbol{\kappa}^t) \quad \text{(PGM)}$$

$\boldsymbol{\kappa}^t$ is an approximate subgradient of $h_L$ at $\boldsymbol{s}^t$

$$\min_{S \subseteq V} H(S) := F(S) - G(S)$$

✓ PGM does not need to know $\alpha, \beta, F, G$, just $H$

# Projected subgradient method (PGM)

$$s^{t+1} = \Pi_{[0,1]^d}(s^t - \eta \kappa^t) \quad \text{(PGM)}$$

$\kappa^t$ is an approximate subgradient of $h_L$ at $s^t$

$$\min_{S \subseteq V} H(S) := F(S) - G(S)$$

✓ PGM does not need to know $\alpha, \beta, F, G$, just $H$

**Approximation guarantee**

After $T$ iterations of PGM + rounding, we obtain:

$$H(\hat{S}) \leq \frac{1}{\alpha} F(S^*) - \beta G(S^*) + O(\frac{1}{\sqrt{T}})$$

# Projected subgradient method (PGM)

$$s^{t+1} = \Pi_{[0,1]^d}(s^t - \eta \kappa^t) \quad \text{(PGM)}$$

$\kappa^t$ is an approximate subgradient of $h_L$ at $s^t$

$$\min_{S \subseteq V} H(S) := F(S) - G(S)$$

✓ PGM does not need to know $\alpha, \beta, F, G$, just $H$

## Approximation guarantee

After $T$ iterations of PGM + rounding, we obtain:

$$H(\hat{S}) \leq \frac{1}{\alpha} F(S^*) - \beta G(S^*) + O(\frac{1}{\sqrt{T}})$$

✓ Result extends to noisy oracle setting:
$P(|\hat{H}(S) - H(S)| \leq \epsilon) \geq 1 - \delta$

# Can we do better?

General set function minimization (in value oracle model):

$$\min_{S \subseteq V} H(S) := F(S) - G(S)$$

### Inapproximability result

For any $\delta > 0$, no (deterministic or randomized) algorithm achieves

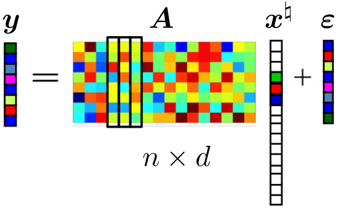
$$\mathbb{E}[H(\hat{S})] \leq \frac{1}{\alpha} F(S^*) - \beta G(S^*) - \delta$$

with less than exponentially many queries.

# Experiment: Structured sparse learning

**Problem:** Learn $\boldsymbol{x}^\natural \in \mathbb{R}^d$, whose support is an interval, from noisy linear Gaussian measurements

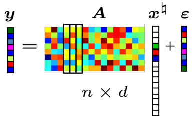
$$\min_{S \subseteq V} H(S) := \lambda F(S) - G(S)$$

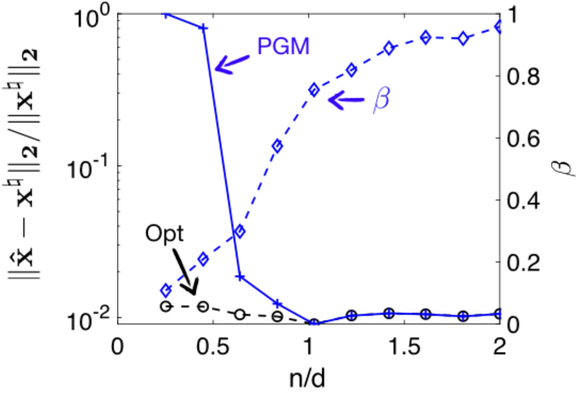



- Regularizer: $F(S) = d + \max(S) - \min(S), F(\emptyset) = 0$; $\alpha = 1$
- Loss: $G(S) = \ell(0) - \min_{\text{supp}(\boldsymbol{x}) \subseteq S} \ell(\boldsymbol{x})$, where $\ell$ is least squares loss. $G$ is $\beta$-weakly DR-supermodular; $\beta > 0$

# Experiment: Structured sparse learning

$$\min_{S \subseteq V} H(S) := \lambda F(S) - G(S)$$

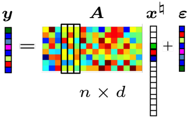


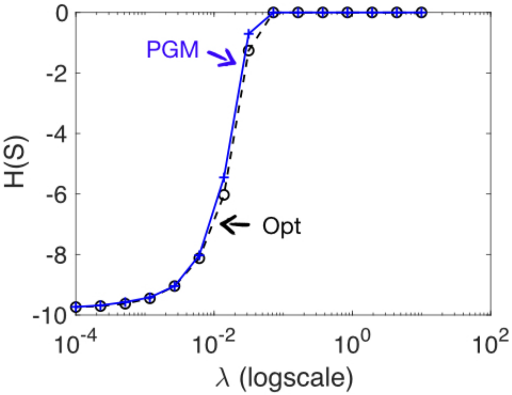




$$d = 250, k = 20, \sigma = 0.01$$

# Experiment: Structured sparse learning

$$\min_{S \subseteq V} H(S) := \lambda F(S) - G(S)$$




$$d = 250, k = 20, \sigma = 0.01, n = 306$$

# Take home message

Approximate submodularity $\Rightarrow$ guaranteed tight approximate solutions using efficient convex methods

# References I


- Bai, W., Iyer, R., Wei, K., and Bilmes, J. (2016).
  Algorithms for optimizing the ratio of submodular functions.
  In *International Conference on Machine Learning*, pages 2751–2759.

- Bian, A. A., Buhmann, J. M., Krause, A., and Tschiatschek, S. (2017).
  Guarantees for greedy maximization of non-submodular functions with applications.
  In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 498–507. JMLR. org.

- Das, A. and Kempe, D. (2011).
  Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection.
  *arXiv preprint arXiv:1102.3975*.

- Edmonds, J. (2003).
  Submodular functions, matroids, and certain polyhedra.
  In *Combinatorial Optimization—Eureka, You Shrink!*, pages 11–26. Springer.

# References II


- Hassidim, A. and Singer, Y. (2018).
  Optimization for approximate submodularity.
  In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 394–405. Curran Associates Inc.

- Horel, T. and Singer, Y. (2016).
  Maximization of approximately submodular functions.
  In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 3045–3053. Curran Associates, Inc.

- Krause, A., Singh, A., and Guestrin, C. (2008).
  Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies.
  *Journal of Machine Learning Research*, 9(Feb):235–284.

# References III


- Kuhnle, A., Smith, J. D., Crawford, V. G., and Thai, M. T. (2018).
  Fast maximization of non-submodular, monotonic functions on the integer lattice.
  *arXiv preprint arXiv:1805.06990.*
- Lehmann, B., Lehmann, D., and Nisan, N. (2006).
  Combinatorial auctions with decreasing marginal utilities.
  *Games and Economic Behavior*, 55(2):270–296.
- Lovász, L. (1983).
  Submodular functions and convexity.
  In *Mathematical Programming The State of the Art*, pages 235–257. Springer.
- Mairal, J., Jenatton, R., Bach, F. R., and Obozinski, G. R. (2010).
  Network flow algorithms for structured sparsity.
  In *Advances in Neural Information Processing Systems*, pages 1558–1566.

# References IV


- Qian, C., Shi, J.-C., Yu, Y., Tang, K., and Zhou, Z.-H. (2017).
  Optimizing ratio of monotone set functions.
  In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, IJCAI'17, pages 2606–2612. AAAI Press.
- Sviridenko, M., Vondrák, J., and Ward, J. (2017).
  Optimal approximation for submodular and supermodular optimization with bounded curvature.
  *Mathematics of Operations Research*, 42(4):1197–1218.
- Wang, Y.-J., Xu, D.-C., Jiang, Y.-J., and Zhang, D.-M. (2019).
  Minimizing ratio of monotone non-submodular functions.
  *Journal of the Operations Research Society of China*.