Guaranteed Non-convex Optimization: Submodular Maximization over Continuous Domains

ETH zürich

A. An Bian, Baharan Mirzasoleiman, Joachim M. Buhmann, Andreas Krause

Characterizations of Submodular Continuous Functions:

Properties	Submodular **cont.** function $f(\cdot)$	Convex function $g(\cdot), \lambda \in [0,1]$
0th order	$f(\boldsymbol{x}) + f(\boldsymbol{y}) \geq f(\boldsymbol{x} \vee \boldsymbol{y}) + f(\boldsymbol{x} \wedge \boldsymbol{y})$	$\lambda g(\boldsymbol{x}) + (1-\lambda) g(\boldsymbol{y}) \geq g(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y})$
1st order	weak DR (this work)	$g(\boldsymbol{y}) - g(\boldsymbol{x}) \geq \langle \nabla g(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle$
2nd order	$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} \leq 0, \forall i \neq j$ (off-diagonal Hessian)	$\nabla^2 g(\boldsymbol{x}) \succcurlyeq 0$

$\wedge$: coordinate-wise min., $\vee$: coordinate-wise max.

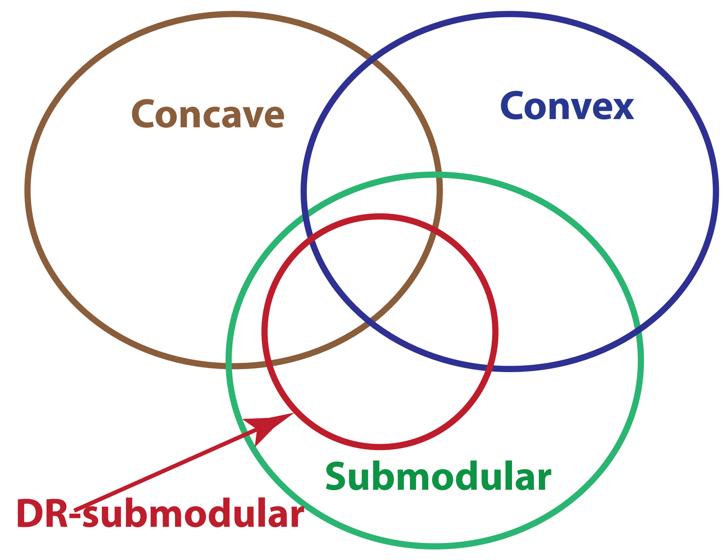
Definition (weak DR):
$\forall \boldsymbol{a} \leq \boldsymbol{b} \in \mathrm{dom} f, \forall i \in E \; s.t. \; a_i = b_i, \forall k \geq 0,$
$f(k\chi_i + \boldsymbol{a}) - f(\boldsymbol{a}) \geq f(k\chi_i + \boldsymbol{b}) - f(\boldsymbol{b}).$

Submodular $\Leftrightarrow$ weak DR

$E := \{e_1, e_2, \dots, e_n\}$: ground set of n elements,
χ_i: characteristic vector for element e_i

weak DR: unified characterization for the submodularity of all set, integer-lattice & continuous functions

weak DR as 1st order property:
$\nabla_{\chi_i} f(\boldsymbol{a}) \geq \nabla_{\chi_i} f(\boldsymbol{b}), \forall i \in E \; s.t. \; a_i = b_i$



Toy example (quadratic):
$f(\boldsymbol{x}) = 0.5 * \boldsymbol{x}^{\mathrm{T}} \mathbf{H} \boldsymbol{x} + \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x}$
Hessian is $\mathbf{H}$

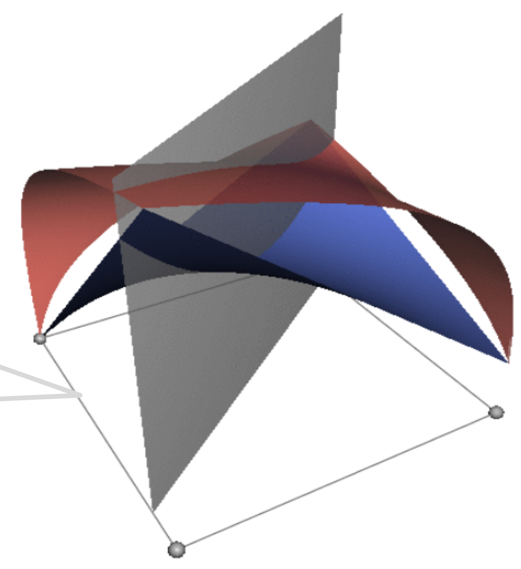
$\mathbf{H} = \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix}$, eigenvalues: $\binom{1}{-3}$

Non-convex/non-concave, but submodular & DR-submodular

Definition (DR-submodular): All entries of Hessian non-positive

Submodular: only **off-diagonal** entries of Hessian non-positive

Submodular + coordinate-wise concave $\Leftrightarrow$ DR-submodular



DR-submodular examples:
Softmax (red) & multilinear (blue) extensions of submodular set fns
[Fig. from Gillenwater et al '12]

Problem Setting (I)

$\max_{\boldsymbol{x} \in \mathcal{P}} f(\boldsymbol{x})$
$f(\boldsymbol{x})$: monotone, **DR-submodular**
$\mathcal{P}$: down-closed convex set

Hardness: (I) is NP-hard. When $\mathcal{P}$ is a polytope, the **optimal** approximation ratio is $(\mathbf{1 - 1/e})$, unless RP = NP.

Algorithm: Frank-Wolfe variant with adaptive step sizes
Input: Prespecified step size $\gamma \in (0,1]$
1) $\boldsymbol{x}^0 \leftarrow 0, t \leftarrow 0, k \leftarrow 0$ // k : iteration idx
2) **While** $t < 1$ **do**
3) $\boldsymbol{v}^k \leftarrow \mathrm{argmax}_{\boldsymbol{v} \in \mathcal{P}} \langle \boldsymbol{v}, \nabla f(\boldsymbol{x}^k) \rangle$ //allow multi. & additive errors
4) find step size $\gamma_k \in (0,1]$, set $\gamma_k \leftarrow \min\{\gamma_k, 1-t\}$
5) $\boldsymbol{x}^{k+1} \leftarrow \boldsymbol{x}^k + \gamma_k \boldsymbol{v}^k, t \leftarrow t + \gamma_k, k \leftarrow k+1$
Return $\boldsymbol{x}^K$ // K inerations in total

Frank-Wolfe has $(1 - 1/e)$-approximation & sub-linear rate

Proof idea: $f(\cdot)$ is DR-submodular $\Rightarrow$ **concave** along non-negative directions

Assuming $g_{\boldsymbol{x},\boldsymbol{v}}(\xi)$ has L-Lipschitz continuous derivative in $[0,1]$

Auxiliary fn. $g_{\boldsymbol{x},\boldsymbol{v}}(\xi) := f(\boldsymbol{x} + \xi \boldsymbol{v})$, $\xi \geq 0, \boldsymbol{v} \in \mathbb{R}^n_+$ is concave

$f(\boldsymbol{x}^{k+1}) - f(\boldsymbol{x}^k) \geq \gamma_k \langle \boldsymbol{v}^k, \nabla f(\boldsymbol{x}^k) \rangle - L\gamma_k^2/2$

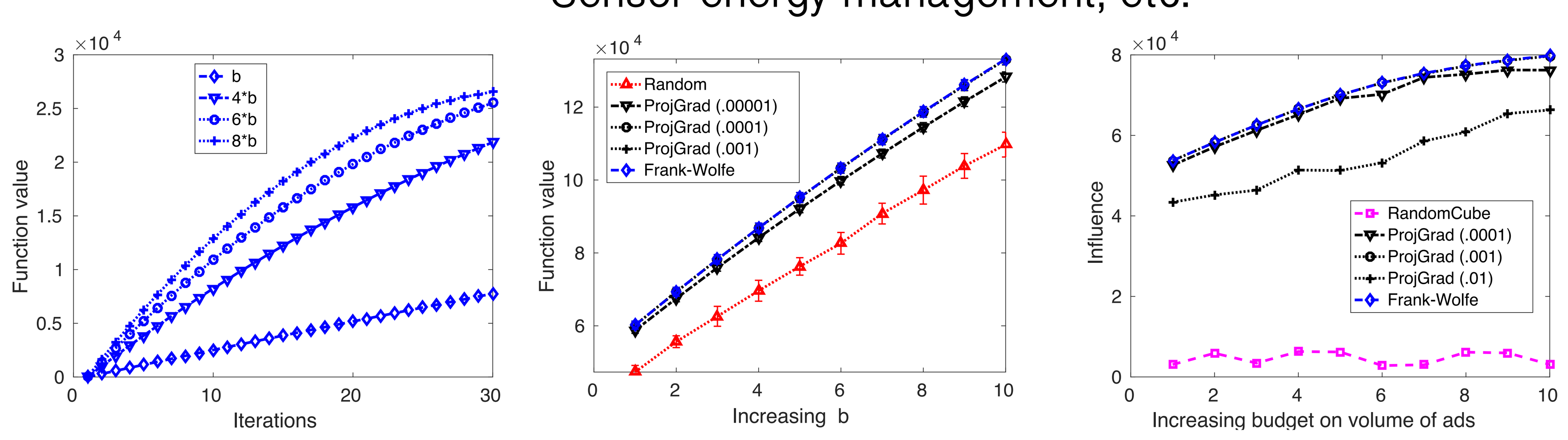
$\langle \boldsymbol{v}^k, \nabla f(\boldsymbol{x}^k) \rangle \geq f(\boldsymbol{x}^*) - f(\boldsymbol{x}^k)$

$$f(\boldsymbol{x}^K) \geq (1 - e^{-1}) f(\boldsymbol{x}^*) - \frac{L}{2} \sum_k \gamma_k^2$$

constant step size $\gamma_k = K^{-1}$

$$f(\boldsymbol{x}^K) \geq (1 - e^{-1}) f(\boldsymbol{x}^*) - \frac{L}{2K}$$

Applications:
- Influence max. with continuous assignments
- Sensor energy management, etc.



Objectives by Frank-Wolfe; DR-submodular quadratic function; Influence max. on real data

Baselines: Projected Gradient, Random

Problem Setting (II)

$\max_{\boldsymbol{x} \in [\underline{\boldsymbol{u}}, \overline{\boldsymbol{u}}]} f(\boldsymbol{x})$
$f(\boldsymbol{x})$: **non-monotone**, submodular
$[\underline{\boldsymbol{u}}, \overline{\boldsymbol{u}}]$: "box" constraint

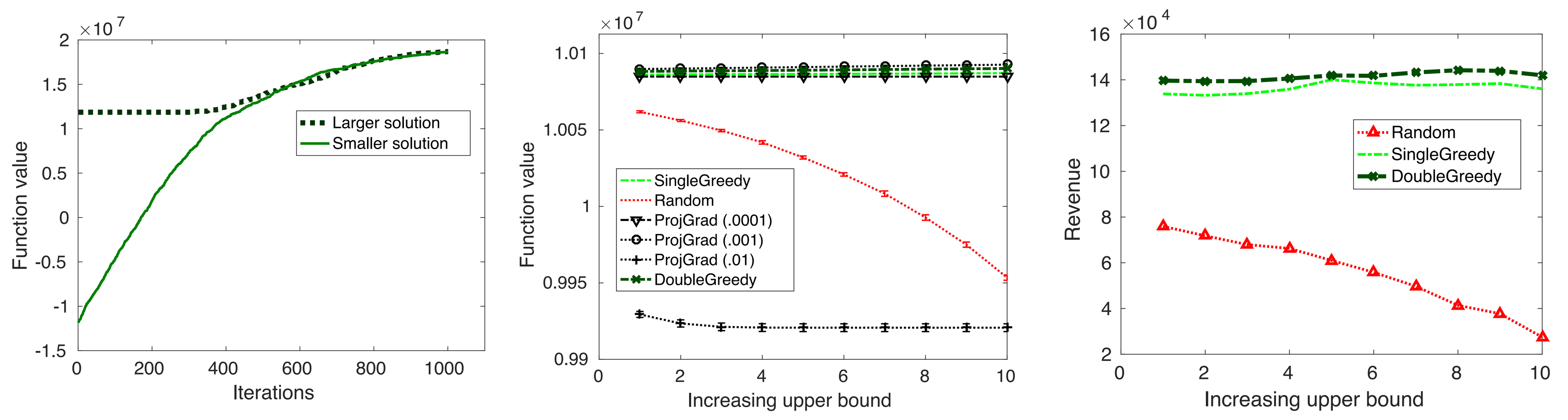
Hardness: (II) is NP-hard. There is no $\left(\frac{1}{2} + \boldsymbol{\varepsilon}\right)$-approximation $\forall \boldsymbol{\varepsilon} > \mathbf{0}$, unless RP = NP.

Algorithm: DoubleGreedy (coordinate ascent with 2 solutions)
Input: $f(\underline{\boldsymbol{u}}) + f(\overline{\boldsymbol{u}}) \geq 0$
1) $\boldsymbol{x} \leftarrow \underline{\boldsymbol{u}}, \boldsymbol{y} \leftarrow \overline{\boldsymbol{u}}$
2) **For** each coordinate $i \in [n]$ **do**
3) find max. gain δ_a when increasing x_i // allow additive error
4) find max. gain δ_b when decreasing y_i // in Steps 3), 4)
5) change x_i and y_i according to whether $\delta_a \geq \delta_b$ or not
Return $\boldsymbol{x}^n$ $(= \boldsymbol{y}^n)$

DoubleGreedy has 1/3-approximation, i.e., $f(\boldsymbol{x}^n) \geq 1/3 \, f(\boldsymbol{x}^*)$

Proof idea: Bound the loss in objective value from the assumed optimal value between every two consecutive steps

Applications:
- Revenue max. with continuous assignments
- Multi-resolution summarization, etc.



Objectives by DoubleGreedy; Submodular quadratic function; Revenue max. on real data

Baselines: SingleGreedy, Projected Gradient, Random

More applications:

- Many classical discrete submodular problems → naturally generalize to continuous settings
- Facility location considering scale of each facility
- Maximum coverage with confidence level, etc

Open problems:

- Projected gradient → approximation guarantee?
- For Problem (II), ½-approximation reachable?
- Non-monotone DR-submodular max. s.t. convex constraints?