Massachusetts Institute of Technology

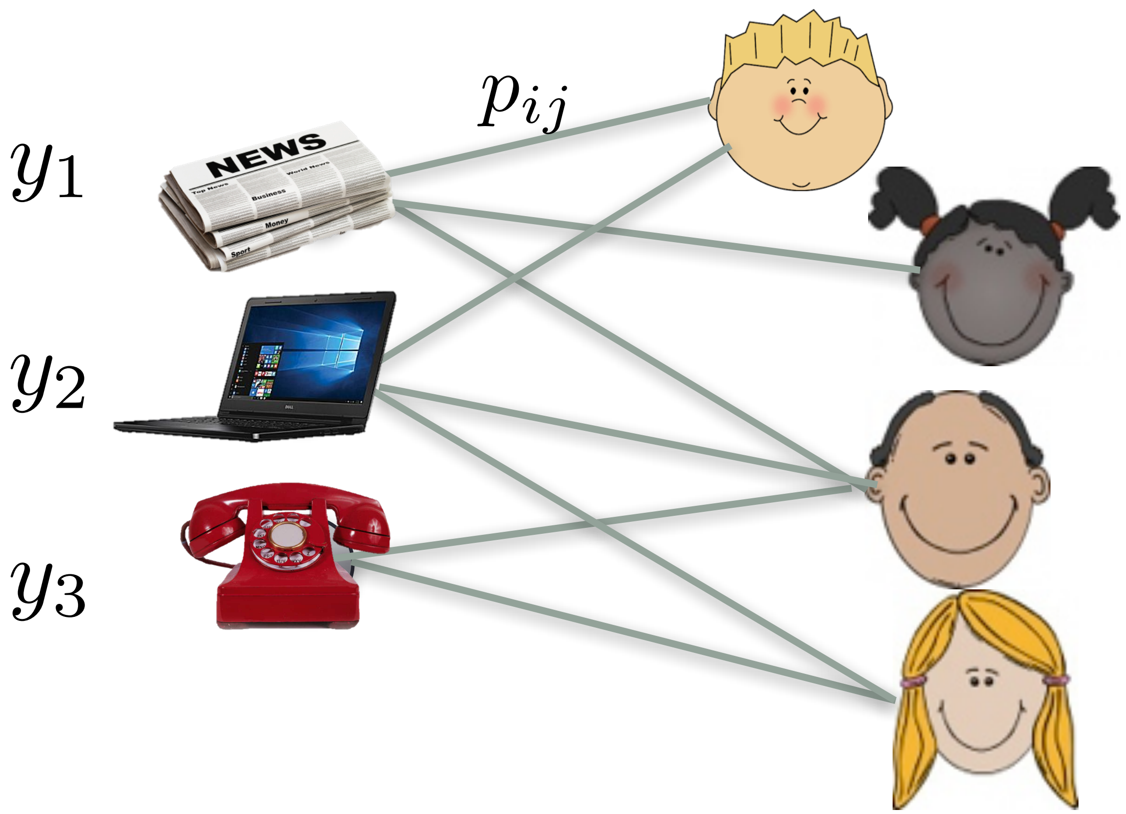
Robust Influence Maximization via Continuous Submodular Minimization

Stefanie Jegelka
joint work with Matthew Staib

Les Houches, April 13 2017

Bipartite Influence Maximization

aka "Budget Allocation"



"Independent Cascade Model":

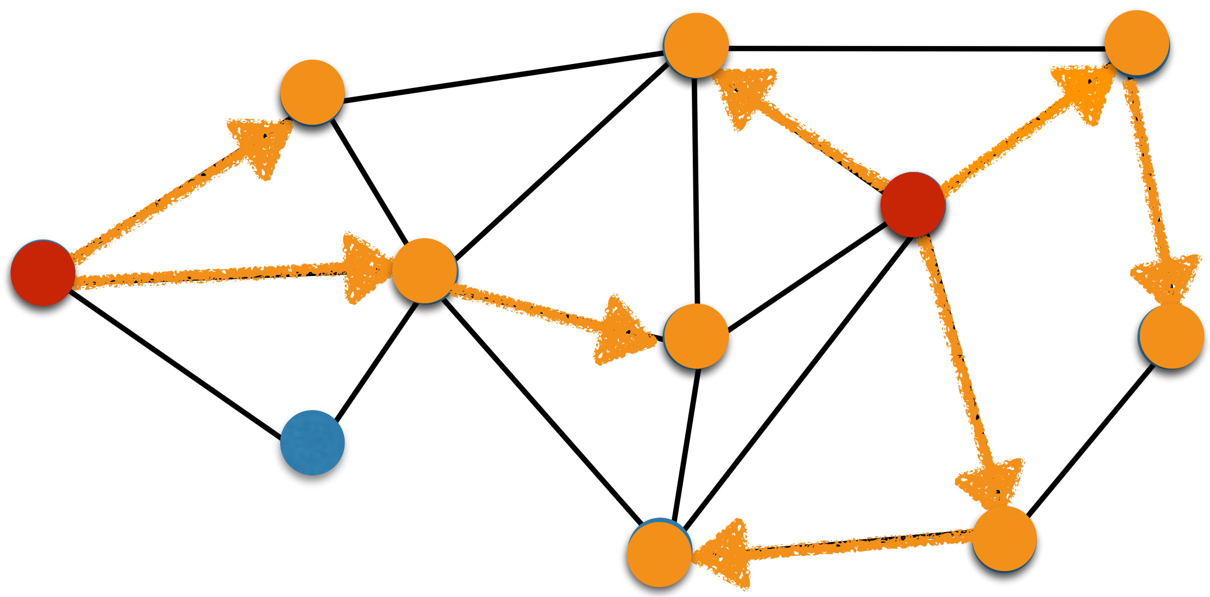
$$\mathbb{P}[j \text{ active}] = 1 - \prod_{i=1}^{n} (1 - p_{ij})^{y_i}$$

$$\mathcal{I}(y; p) = \sum_j \mathbb{P}[j \text{ active}]$$

$$\max_y \mathcal{I}(y; p) \qquad \text{s.t. } \sum_i y_i \leq k; \;\; 0 \leq y \leq u$$

(Alon-Gamzu-Tennenholtz 12; Soma et al 14; Soma & Yoshida 15; Hatano et al 15; Bian et al 17)

Influence Maximization



(Domingos & Richardson 01, Kempe et al 03, Mossel & Roch, Chen et al 09, Borgs et al 14, …)

Influence Maximization

$$\boxed{\max_{y} \; \mathcal{I}(y;p) \qquad \text{s.t.} \;\; \sum_i y_i \leq k; \;\; 0 \leq y \leq u}$$

- concave maximization over convex set
- y discrete: "monotone submodular maximization"
  $(1-\frac{1}{e})$-approximation via greedy algorithm

(Alon-Gamzu-Tennenholtz 12; Soma et al 14; Soma & Yoshida 15; Hatano et al 15; Bian et al 17)

Influence Maximization

$$\boxed{\max_y \; \mathcal{I}(y;p) \qquad \text{s.t. } \sum_i y_i \le k; \;\; 0 \le y \le u}$$

- concave maximization over convex set
- y discrete: "monotone submodular maximization"
  $(1-\frac{1}{e})$ -approximation via greedy algorithm
- *What if we don't know p?*

Robust Influence Maximization

1. small collection of functions
   (He-Kempe 16, Chen et al 16, Krause et al)

$$\max_{y \in \mathcal{B}} \ \min_{\ell \in \mathcal{L}} \ \frac{\mathcal{I}_\ell(y)}{\mathcal{I}_\ell(y_\ell^*)}$$

$$\max_{y \in \mathcal{B}} \ \min_{\ell \in \mathcal{L}} \ \mathcal{I}_\ell$$

Robust Optimization

$$\max_{y \in \mathcal{B}} \min_{p \in \mathcal{P}} \mathcal{I}(y; p)$$

- *uncertainty sets:* e.g.

$$\mathcal{P}^Q = \{p \in [0,1]^d, (p - \hat{p})^\top \Sigma^{-1} (p - \hat{p}) \leq \gamma\}$$

$$\mathcal{P}^D = \left\{p \in [0,1]^d : \sum_{(i,j) \in E} \frac{p_{ij} - \hat{p}_{ij}}{u_{ij} - p_{ij}} \leq \gamma \right\}$$

- in y: min of concave functions -> *concave*
  use e.g. **subgradient method?**
- in p: ***not*** convex or quasiconvex …

But: continuous submodular in p…

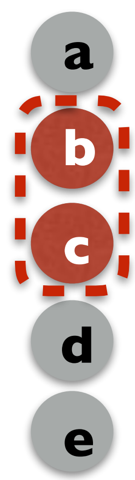
Corollary *(Staib-J17)*
Under a distinctness condition, we can solve the adversary's problem via Frank-Wolfe on a convex problem with dimension $O(n/\epsilon)$

Roadmap:

- Brief introduction to submodularity
- (Constrained) optimization and relaxations
- Implications

Submodular Set Functions

$S \subseteq V$



$$\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad x \in \{0,1\}^n$$

$F : 2^V \to \mathbb{R}$ ***submodular*** if for all *S,T*

$$F(S) + F(T) \geq F(S \cup T) + F(S \cap T)$$

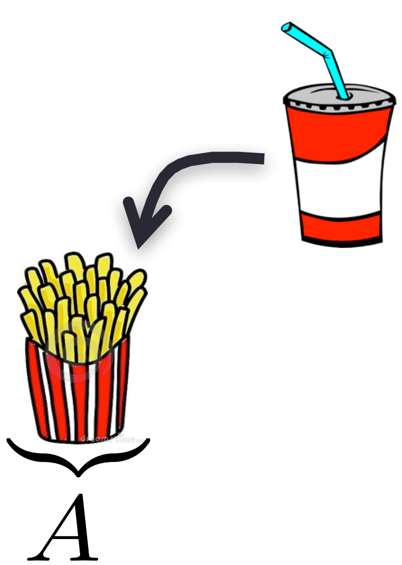
$$F(x) + F(y) \geq F(x \vee y) + F(x \wedge y)$$

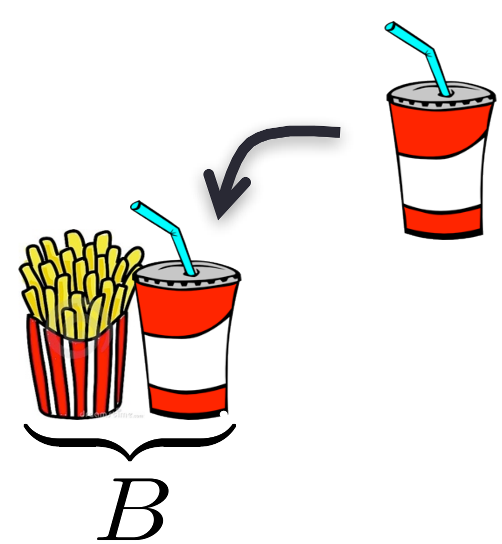
Diminishing Returns/Costs

$$F(S) \; + \; F(T) \quad \geq \quad F(S \cup T) \; + \; F(S \cap T)$$

equivalent condition: $\forall A \subseteq B \subset V, i \notin B :$

$$F(A \cup i) - F(A) \quad \geq \quad F(B \cup i) - F(B)$$



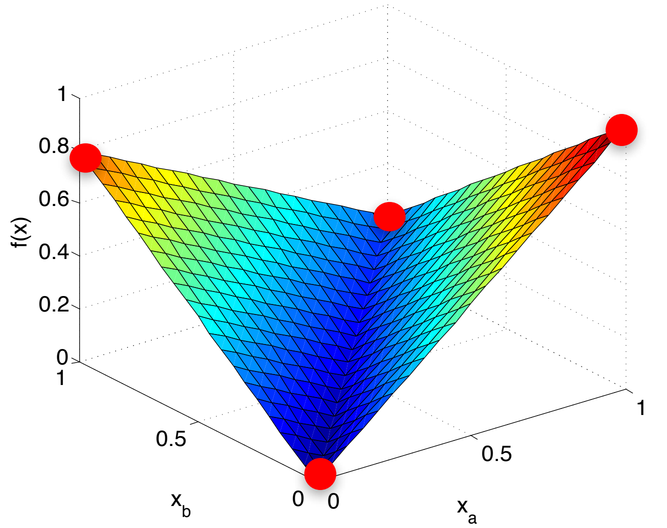


Why is submodularity useful?

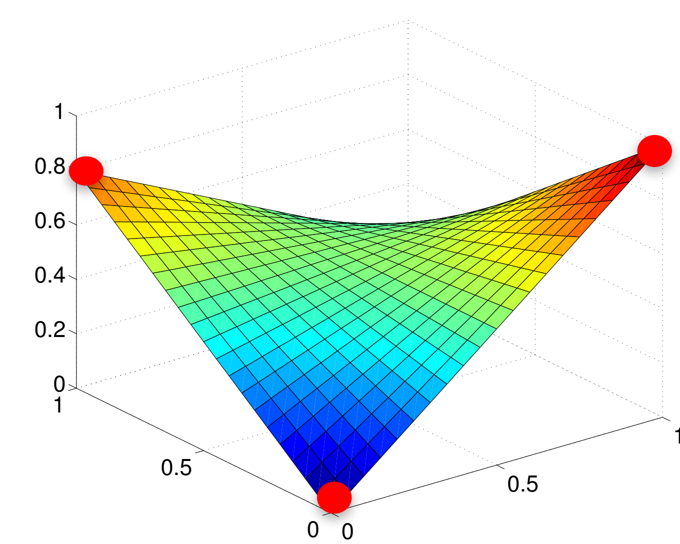
$$F(x), x \in \{0,1\}^n \text{ submodular}$$

- naturally occurs in many problems, and good optimization:

- **Minimization** over $\{0,1\}^n$ in ***polynomial time***
  "discrete convexity"
  (Groetschel-Lovasz-Schrijver 81, Schrijver 00, Iwata-Fleischer-Fujishige 01; Lee-Sidford-Wong 15; …)



- **Maximization** NP-hard, but constant-factor approximations
  "discrete concavity"
  (Nemhauser-Wolsey-Fisher 71; Calinescu-Chekuri-Pal-Vondrak 11; Vondrak ; Buchbinder-Naor-Feldman 12; …)



nonconvex optimization
lattice / continuous submodularity
many optimization & duality results generalize

probability measures
log-supermodular (→ positive assoc.)
log-submodular (← negative assoc.)
sampling, mode,
approx. partition function

submodular set functions

convexity:
minimization
max. coherence

dim. returns:
maximization
max. diversity

many examples:

- linear/modular functions
- entropy
- mutual information
- rank functions
- coverage
- diffusion in networks
- volume
- graph cut …

Submodularity beyond Sets

- straightforward generalization to (distributive) lattices:

$$F(x) \; + \; F(y) \quad \geq \quad F(x \vee y) \; + \; F(x \wedge y)$$

- for differentiable functions:

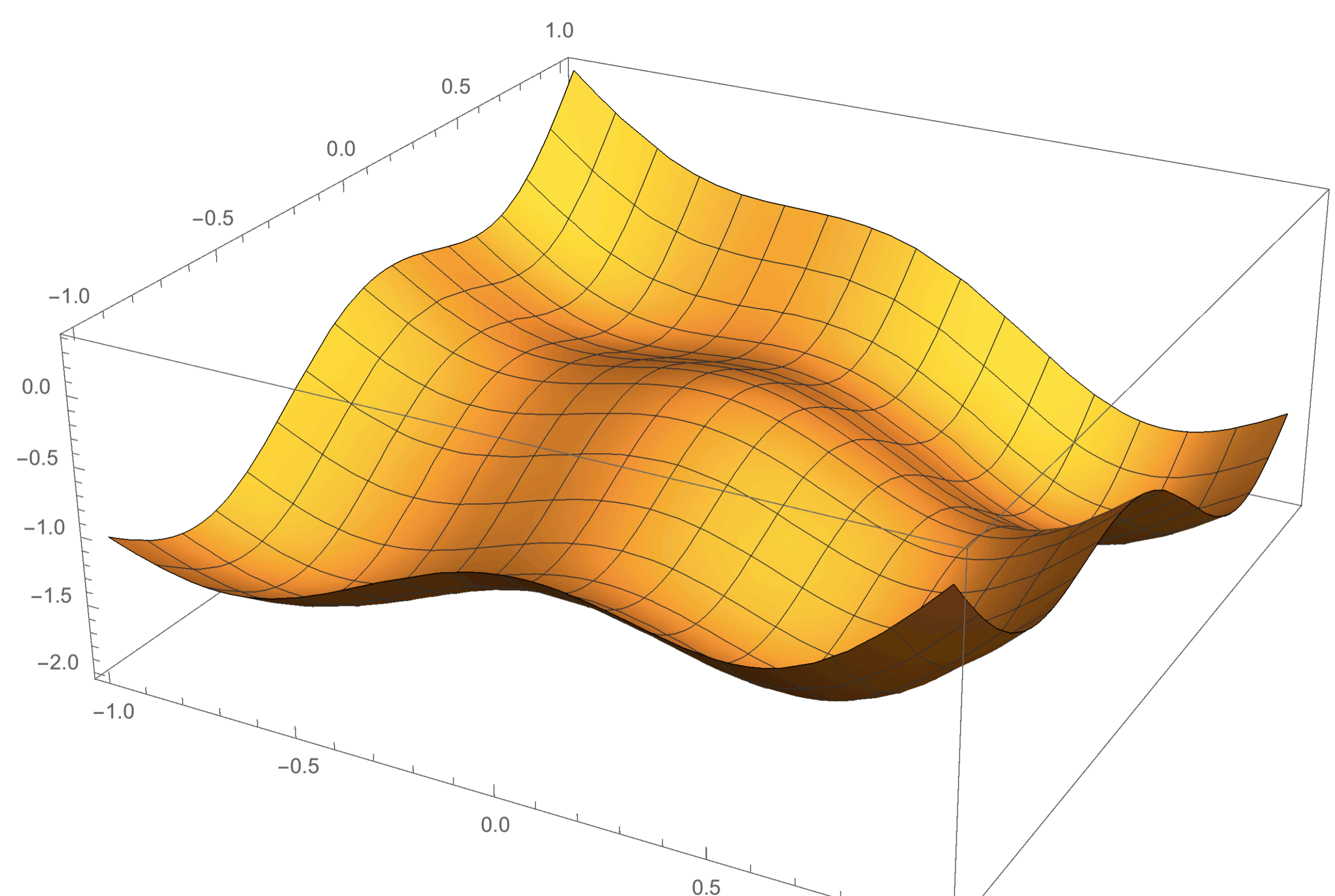
$$\frac{\partial^2 f}{\partial x_i \partial x_j} \; \leq \; 0 \qquad\qquad \forall i \neq j$$

- note: "diminishing returns" is stronger here!
  (DR-submodular)

Convex or Concave?

Examples:

- $F(x) = g(x_i - x_j)$ for convex g
- $F(x) = h(\sum_i x_i)$ for concave h
- $F(x) = x^\top Qx$ for Q with negative off-diagonals



Optimization

Minimization over Box: exact

- Reduction to sets *(Birkhoff 37 & Schrijver 00/Orlin 07)*
- Convex relaxation *(Bach 15)*

- Not all submodular functions are equal! Better for
  - *DR-submodular* functions
    (combining Ene & Nguyen 16 & e.g. Lee-Sidford-Wong 15)
  - $L^\natural$-convex functions *(Murota-Shioura)*

$$f(x) + f(y) \;\geq\; f\left(\left\lceil \frac{x+y}{2} \right\rceil\right) + f\left(\left\lfloor \frac{x+y}{2} \right\rfloor\right)$$

Maximization: greedy or non convex Frank-Wolfe

Back to Influence

$$\max_{y \in \mathcal{B}} \min_{p \in \mathcal{P}} \mathcal{I}(y; p)$$

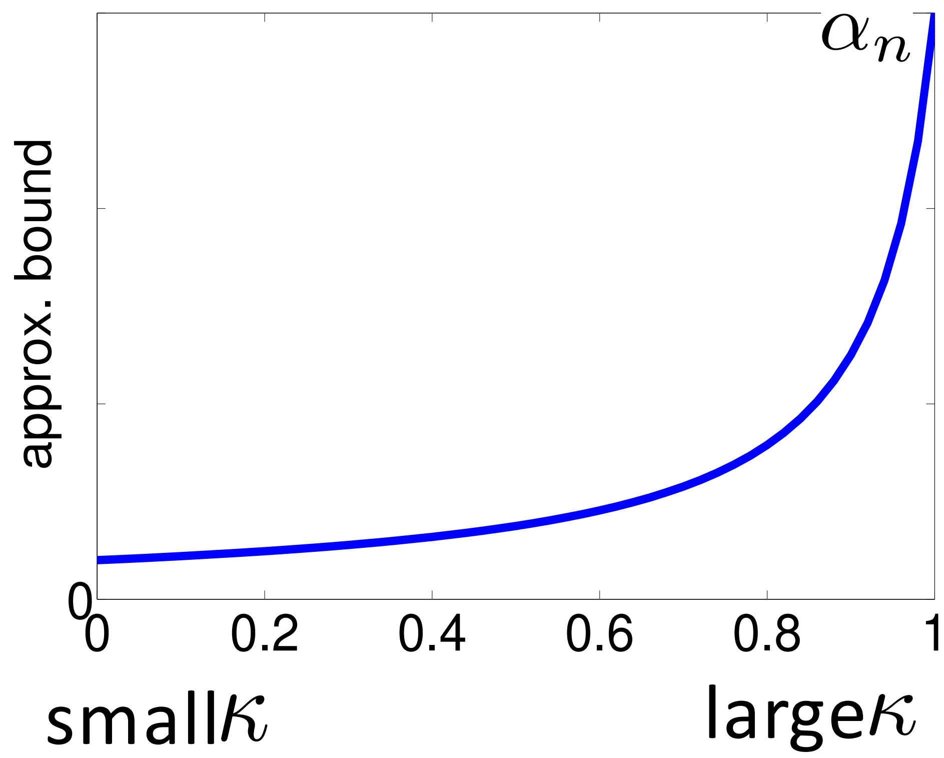
- submodular in 1-p
- DR-submodular if y integer

constrained submodular minimization?

Constrained Minimization?

- can be ***hard.***
  (Goel-Karande-Tripathi-Wang 09; Iwata-Nagano 09; Svitkina-Fleischer 11)

- ***But***, there is hope:
  for increasing functions,
  tight bounds via curvature
  (Iyer-Jegelka-Bilmes 13)



approximation

optimal

Constrained Minimization?

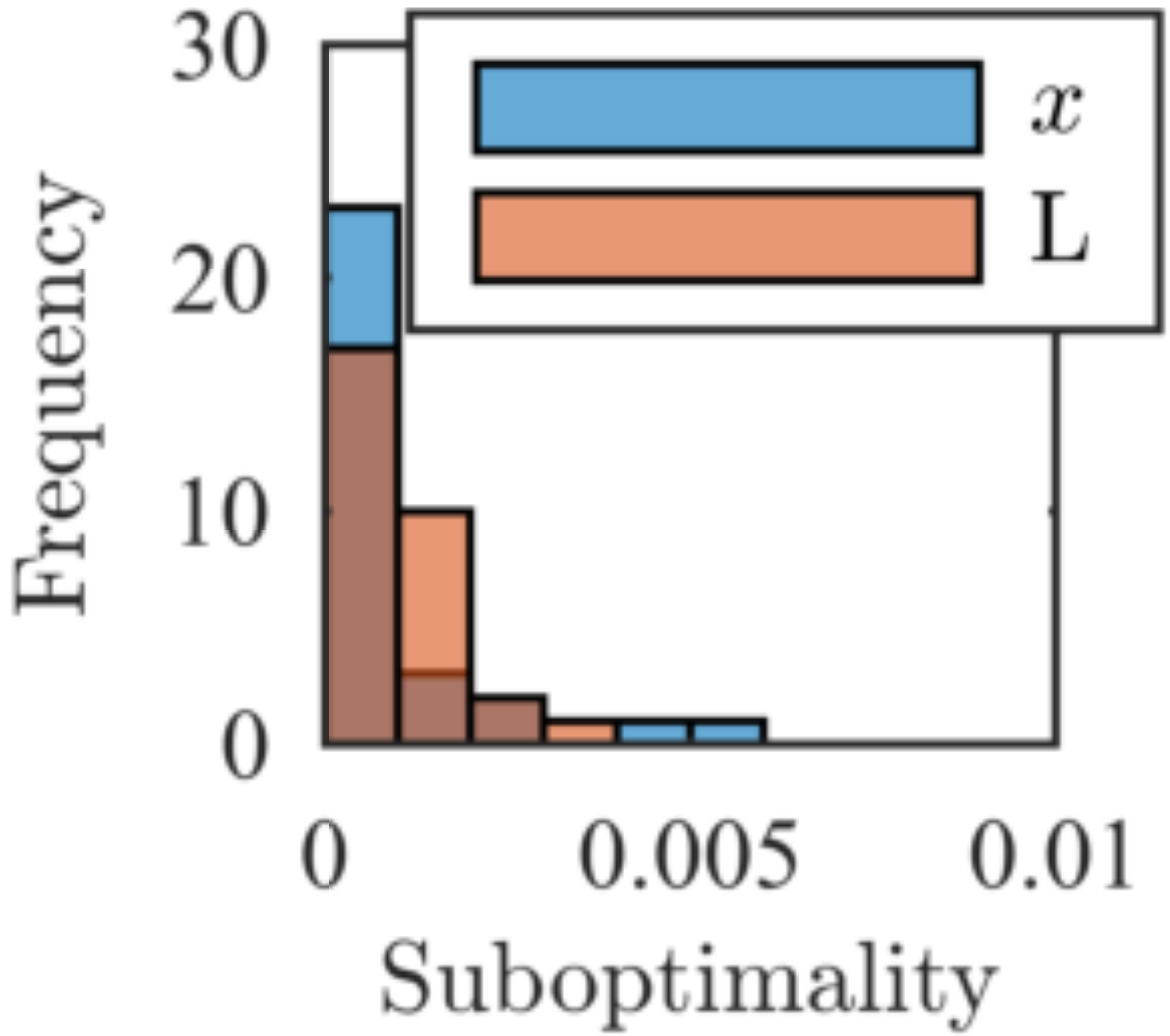
$$\begin{aligned} \min \quad & F(x) \\ \text{s.t.} \quad & R(x) \leq B \\ & \ell \leq x \leq u \end{aligned}$$

Theorem *(Staib-J 17)*

Under distinctness condition, for monotone non-increasing submodular F and separable, increasing R, we can this solve to accuracy ϵ via Frank-Wolfe for smooth convex problem with dimension $O(n/\epsilon)$.

Always: bounds on optimality

Optimality in Practice?



Yahoo Webscope data (>50k edges), 30 rep.

Relaxation

- Lagrangian relaxation

$$\begin{aligned} \min \quad & F(x) \\ \text{s.t.} \quad & R(x) \leq B \\ & \ell \leq x \leq u \end{aligned} \longrightarrow \begin{aligned} \min \quad & F(x) + \lambda R(x) \\ & \ell \leq x \leq u \end{aligned}$$

- Solve parametric problem (discretize)

$$\begin{aligned} \min \quad & F(z) + \lambda R(z) \\ & z_i \in [k_i] \end{aligned}$$

- Pick best feasible $x(\lambda)$

entire solution path
via a single optimization

Parametric Problem

$$\min \;\; F(z) + \lambda R(z)$$
$$z_i \in [k_i]$$

scalar smooth convex function

"decoding":

$$z(\lambda)_i = \rho_i^{-1}(\lambda)$$

$$\min_{\rho} \;\; f(\rho) + \sum_i \sum_{j=1}^{k_i - 1} a_{ij}(\rho_i(j))$$
$$\text{s.t.} \;\; \rho_i \in \mathbb{R}_{\downarrow}^{k_i - 1}$$

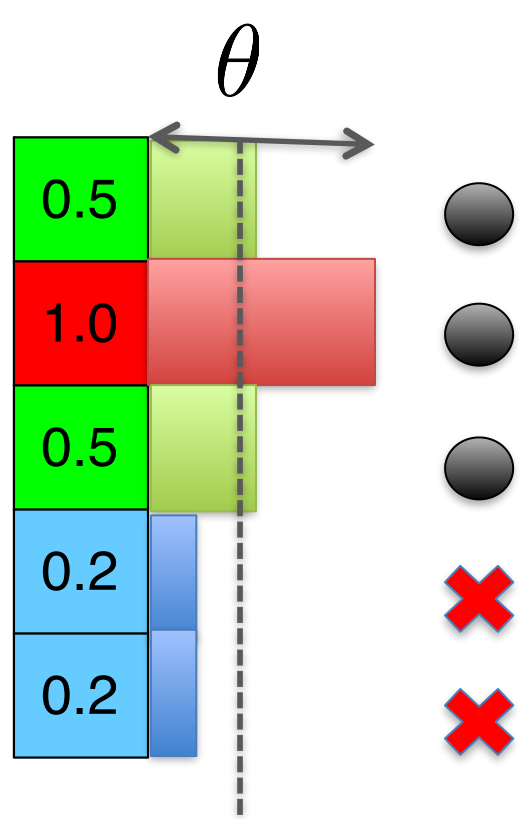
reverse CDF
(decreasing vector)

Set functions: Lovasz extension

$$F(S) \longrightarrow f(u), u \in [0,1]^n$$

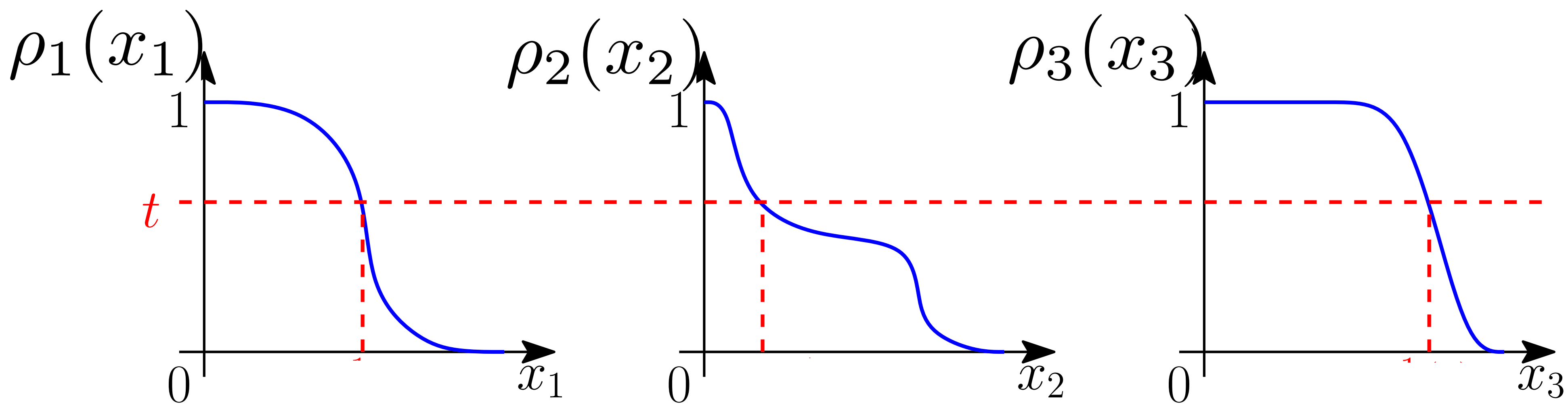
$$f(u) = \mathbb{E}[F(S_\theta)]$$

e.g. $\theta = 0.4$



- "couples/aligns" dimensions
- u_i "distribution" of coordinate i

Generalized extension



- distribution over values 1, … k
- reverse CDF:

$$\rho_i(x_i) = \mu_i\{y \in [k_i], y \geq x_i\}$$

- extension:

$$f(\mu_1, \ldots, \mu_n) = \int_0^1 F(\rho_1^{-1}(t), \ldots, \rho_n^{-1}(t)dt$$

Properties of Extensions I

2 ways to create extension:

- **"coupling/alignment":**
  computable but not always convex

- **convex envelope:**
  convex but not always computable

for submodular functions, these are the same!
(Lovasz 1982, Bach 15, Carlier 03)
via multi-marginal optimal transport

Properties of Extensions II

For

$$\min_S \quad F(S) + \lambda |S|$$

solve

$$\min_u \quad f(u) + \tfrac{1}{2}\|u\|^2$$

u^* gives optimal solutions for all λ

works also for integer functions

Insight: ***we can encode R(x) this way!***

Parametric Problem

$$\min \; F(z) + \lambda R(z)$$
$$z_i \in [k_i]$$

"decoding":

$$z(\lambda)_i = \rho_i^{-1}(\lambda)$$

scalar smooth
convex function

$$\min_{\rho} \; f(\rho) + \sum_i \sum_{j=1}^{k_i - 1} a_{ij}(\rho_i(j))$$

$$\text{s.t. } \; \rho_i \in \mathbb{R}_{\downarrow}^{k_i - 1}$$

reverse CDF
(decreasing vector)

solving relaxation:
pairwise Frank-Wolfe
(Lacoste-Julien & Jaggi 15)
on dual

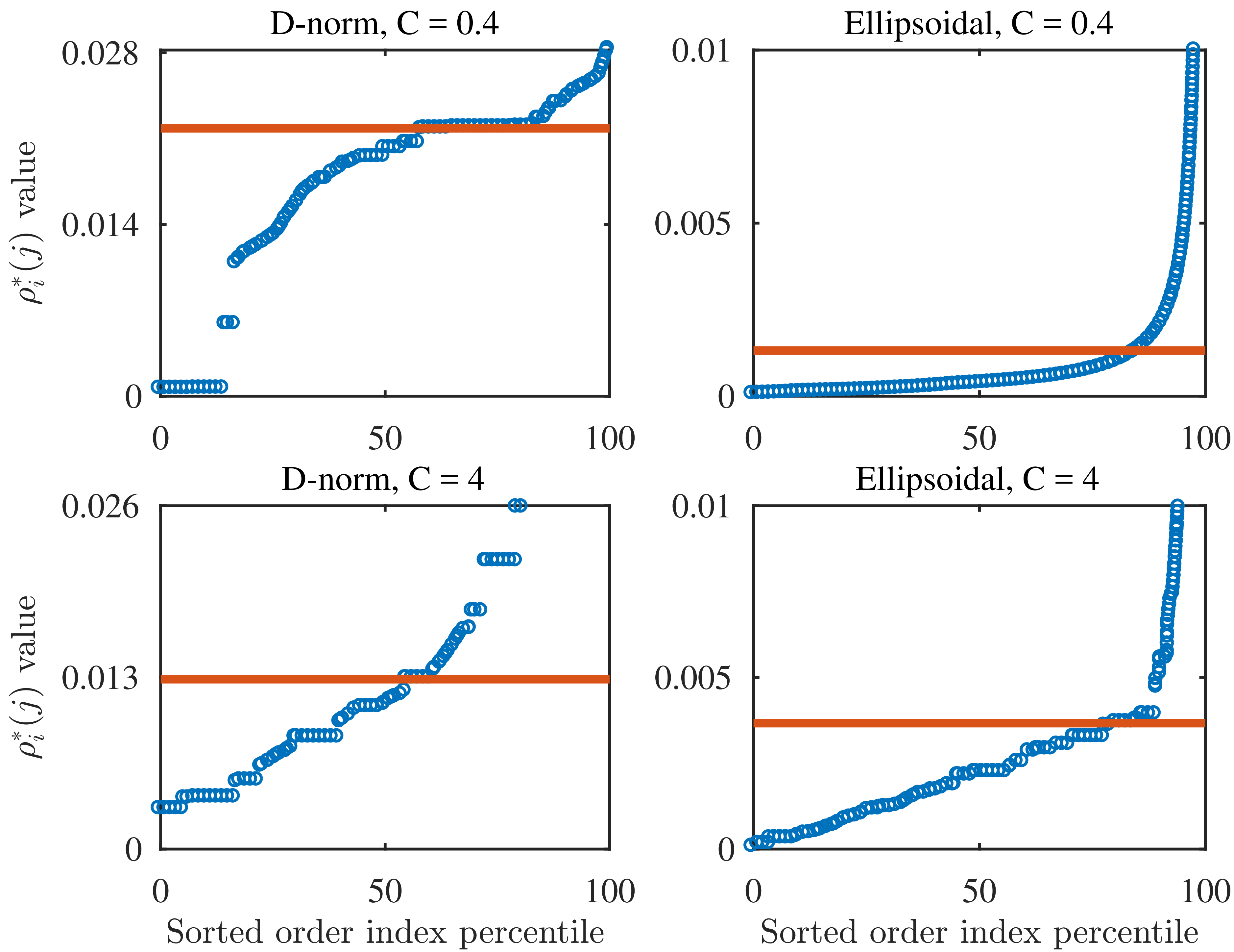
Optimality

"decoding":

$$z(\lambda)_i = \rho_i^{-1}(\lambda)$$

- **Distinctness condition:** all entries of ρ_i are distinct

Distinctness condition in practice



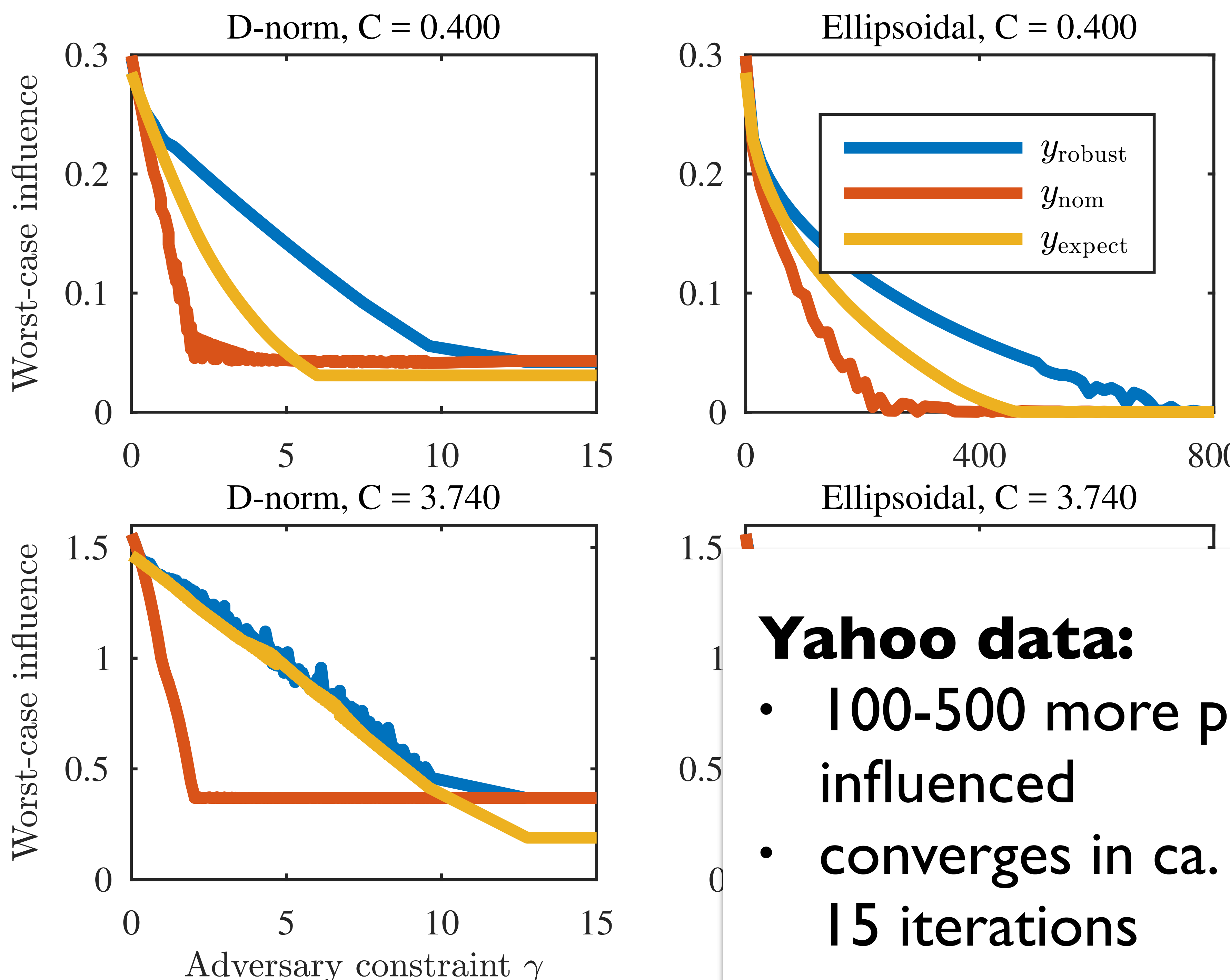
Putting things together

$$\max_{y \in \mathcal{B}} \min_{p \in \mathcal{P}} \mathcal{I}(y; p)$$

$$
\begin{aligned}
&\textbf{loop} \\
&\quad \boxed{p^{(k)} \leftarrow \operatorname{argmin}_{p \in \mathcal{P}} \mathcal{I}(y^{(k)}; p)} \\
&\quad g^{(k)} \leftarrow \nabla_y \mathcal{I}(y^{(k)}; p^{(k)}) \\
&\quad y^{(k+1)} \leftarrow \operatorname{proj}_{\mathcal{Y}}(y^{(k)} + \gamma^{(k)} g^{(k)}) \\
&\quad k \leftarrow k + 1 \\
&\textbf{end loop}
\end{aligned}
$$

continuous submodular minimization

Empirical Results



Yahoo data:

- 100-500 more people influenced
- converges in ca. 15 iterations

Summary

- Optimization for continuous constrained submodular minimization via parametric relaxation
- Optimality bounds
- Applications to robust influence maximization/budget allocation

M. Staib & S. Jegelka. Robust Budget Allocation via Continuous Submodular Functions. arXiv, 2017.