Massachusetts Institute of Technology

Between Discrete and Continuous Optimization: Submodularity & Optimization

Stefanie Jegelka, MIT

Simons Bootcamp Aug 2017

Submodularity

set function: $F(S)$

- submodularity = *“diminishing returns”* $\forall S \subseteq T,\ a \notin T$

$$F(S \cup \{a\}) \;-\; F(S) \;\geq\; F(T \cup \{a\}) \;-\; F(T)$$

Submodularity

set function: $F(S)$

- diminishing returns: $\forall S \subseteq T,\ a \notin T$

$$F(S \cup \{a\}) - F(S) \geq F(T \cup \{a\}) - F(T)$$

- equivalent general definition: $\forall\ A, B \subseteq V$

$$F(A) + F(B) \geq F(A \cup B) + F(A \cap B)$$

Why is this interesting?

Importance of convex functions *(Lovász, 1983):*

- *"occur in many models in economy, engineering and other sciences", "often the only nontrivial property that can be stated in general"*
- *preserved under many operations and transformations: larger effective range of results*
- *sufficient structure for a "mathematically beautiful and practically useful theory"*
- *efficient minimization*

"It is less apparent, but we claim and hope to prove to a certain extent, that a similar role is played in discrete optimization by submodular set-functions" […]

Examples of submodular set functions

- linear functions
- discrete entropy
- discrete mutual information
- matrix rank functions
- matroid rank functions (“combinatorial rank”)
- coverage
- diffusion in networks
- volume (by log determinant)
- graph cuts
- …

Roadmap

- Optimizing submodular set functions:
  discrete optimization via continuous optimization

- Submodularity more generally:
  continuous optimization via discrete optimization

- Further connections

Roadmap

- Optimizing submodular set functions via continuous optimization

Key Question:

Submodularity = Discrete Convexity or Discrete Concavity?

(Lovász, Fujishige, Murota, …)

Continuous extensions

$$\min_{S \subseteq V} F(S) \qquad \Leftrightarrow \qquad \min_{x \in \{0,1\}^n} F(x)$$

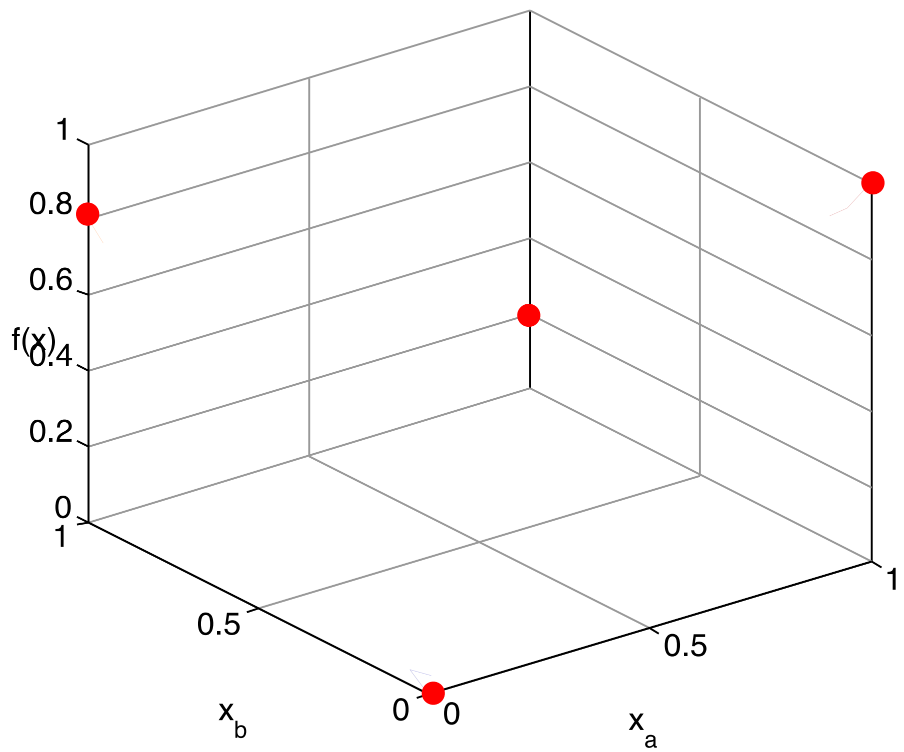
- LP relaxation?
  nonlinear cost function: exponentially many variables…

$$F : \{0,1\}^n \to \mathbb{R} \qquad \longrightarrow \qquad f : [0,1]^n \to \mathbb{R}$$

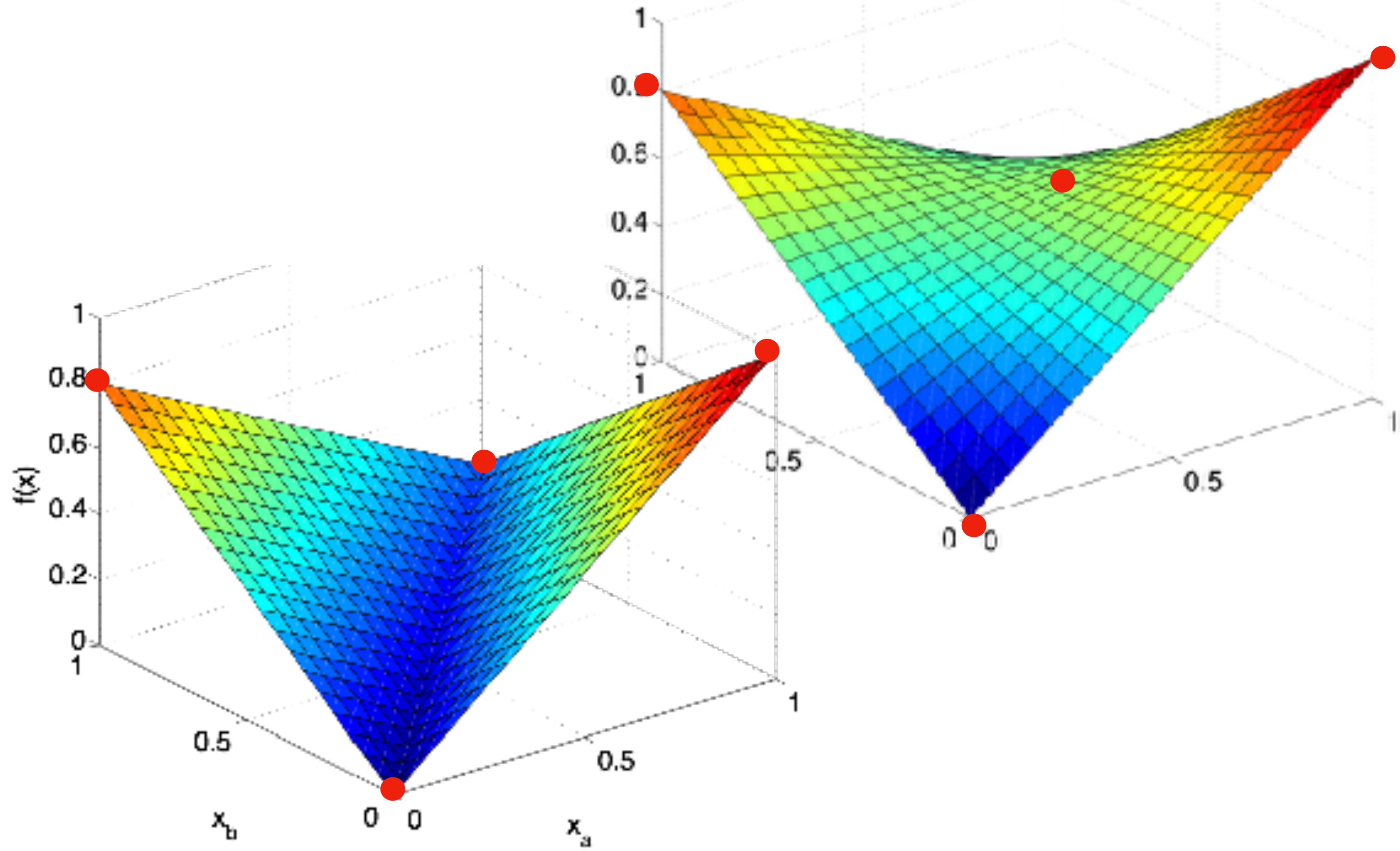
Nonlinear extensions & optimization

nonlinear extension/optimization

$$F : \{0,1\}^n \to \mathbb{R}$$



$$f : [0,1]^n \to \mathbb{R}$$



$$\min_{x \in \mathcal{C} \subseteq \{0,1\}^n} F(x) \qquad \min_{z \in \mathrm{conv}(\mathcal{C}) \subseteq [0,1]^n} f(z)$$

Generic construction

$F : \{0,1\}^n \to \mathbb{R}$ ⟶ $f : [0,1]^n \to \mathbb{R}$

- Define probability measure over *subsets* (joint over coordinates) such that marginals agree with z:

$$\mathbb{P}(i \in S) = z_i$$

- Extension: $$f(z) = \mathbb{E}[F(S)]$$

- for discrete z: $$f(z) = F(z)$$

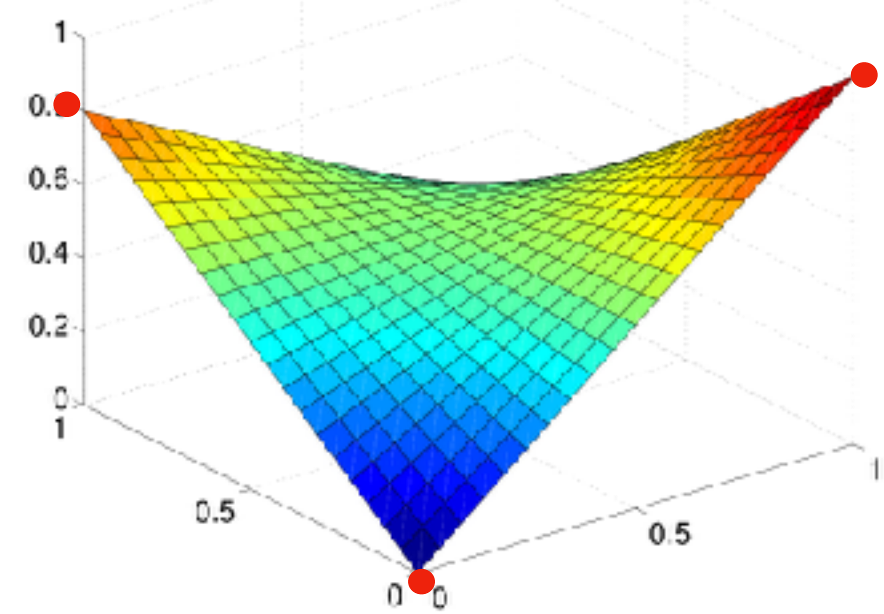
Independent coordinates

$$f(z) = \mathbb{E}[F(S)]$$

a	.5
b	.5
c	0
d	.8

$$P(S) = \prod_{i \in S} z_i \cdot \prod_{j \notin S} (1 - z_j)$$

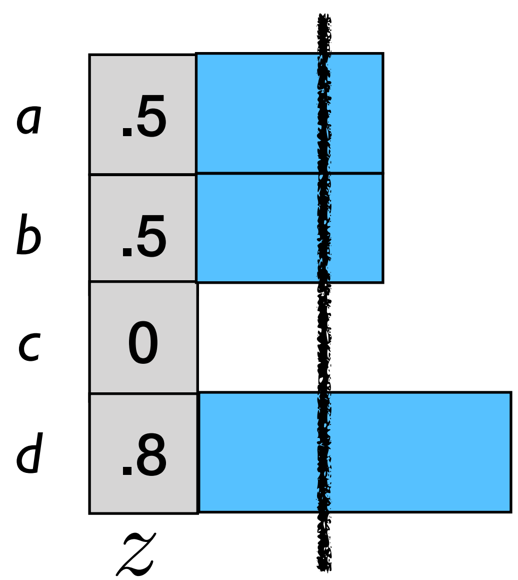
- $f(z)$ is a multilinear polynomial: ***multilinear extension***
- neither convex nor concave…



Lovász extension

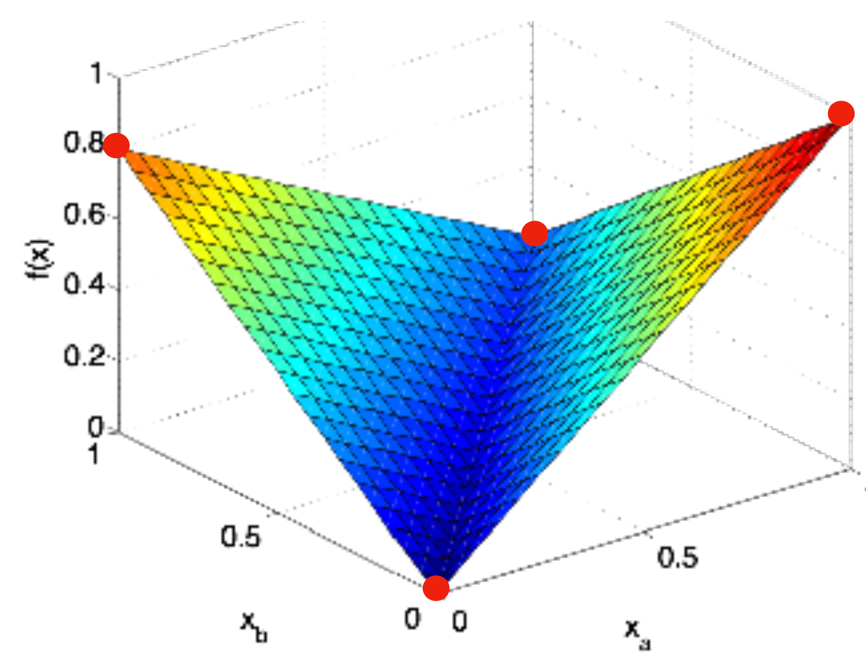
$$f(z) = \mathbb{E}[F(S)] \qquad \mathbb{P}(i \in S) = z_i$$

- "coupled" distribution defined by level sets



$$S_0 = \{\},\ S_1 = \{d\},\ S_2 = \{a, b, d\},$$
$$S_3 = \{a, b, c, d\}$$

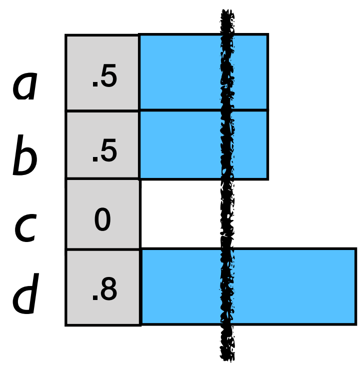
$\mathbb{E}[F(S)]$ = *Choquet integral of F*



Theorem (Lovász 1983)
$f(z)$ is convex iff $F(S)$ is submodular.

Convexity and subgradients

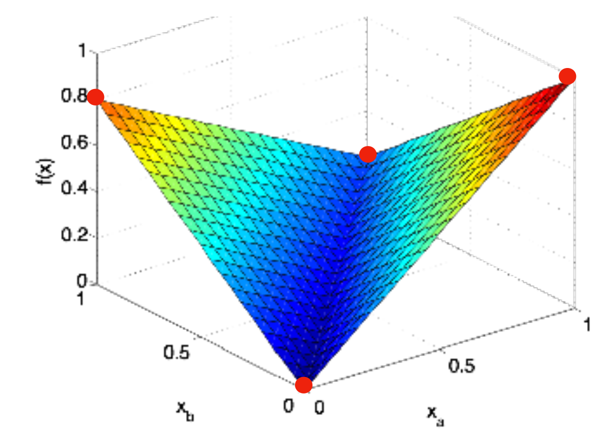
if *F* is submodular *(Edmonds 1971, Lovász 1983)*:



$$f(z) = \mathbb{E}[F(S)] = \max_{s \in \mathcal{B}_F} \langle s, z \rangle$$

Base Polytope of F

- can compute subgradient of *f(z) in O(n log n)*
- rounding: use one of the level sets of *z**



exact convex relaxation!

$$\min_{S \subseteq V} F(S) = \min_{z \in [0,1]^n} f(z)$$

Submodular minimization: a brief overview

$$\min_{z \in [0,1]^n} f(z)$$

convex optimization

- ellipsoid method *(Grötschel-Lovász-Schrijver 81)*
- subgradient method *(improved: Chakrabarty-Lee-Sidford-Wong 16)*

combinatorial optimization

- network flow based *(Schrijver 00, Iwata-Fleischer-Fujishige-01)*
  $O(n^4 T + n^5 \log M)$ *(Iwata 03)*, $O(n^6 + n^5 T)$ *(Orlin 09)*

convex + combinatorial

- cutting planes *(Lee-Sidford-Wong 15)*

$$O(n^2 T \log nM + n^3 \log^c nM) \qquad O(n^3 T \log^2 n + n^4 \log^c n)$$

How far does relaxation go?

- **strongly convex** version:

$$\min_{z \in [0,1]^n} f(z) \quad \dashrightarrow \quad \min_{z \in \mathbb{R}^n} f(z) + \frac{1}{2}\|z\|^2$$

$$\updownarrow$$

dual: $$\min_{s \in \mathcal{B}_F} \frac{1}{2}\|s\|^2$$

- Fujishige-Wolfe / minimum-norm point algorithm
- actually solves *parametric submodular minimization*

- **But:** no relaxation is tight for constrained minimization
  typically hard to approximate

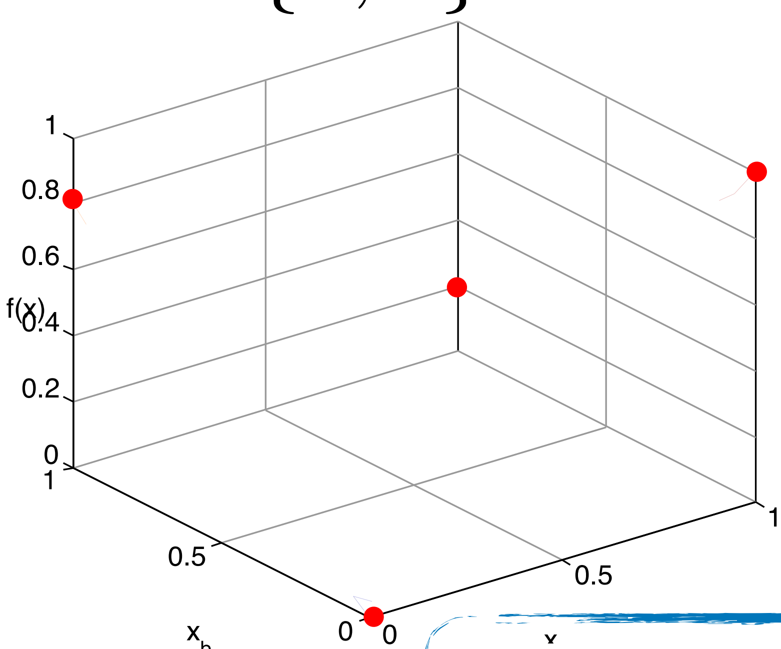
Submodular maximization

$$\max_{S \subseteq V} F(S) \qquad\qquad \max_{|S| \leq k} F(S) \quad *$$

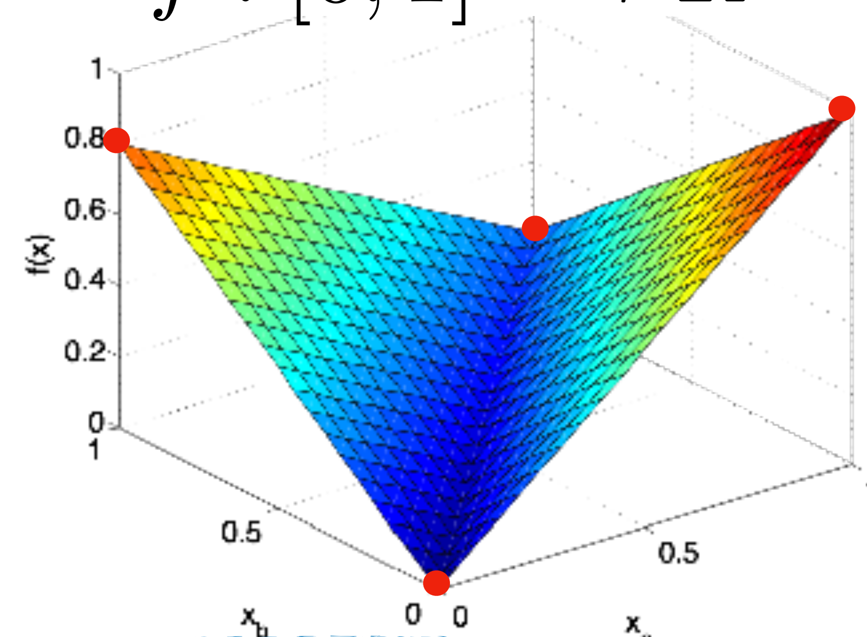
NP-hard

- simple cases (*, *monotone*):
  discrete greedy algorithm is optimal *(Nemhauser-Wolsey-Fisher 1972)*
- more complex cases *(complicated constraints, non-monotone)*:
  continuous extension + rounding

$$F : \{0,1\}^n \to \mathbb{R}$$



$$f : [0,1]^n \to \mathbb{R}$$

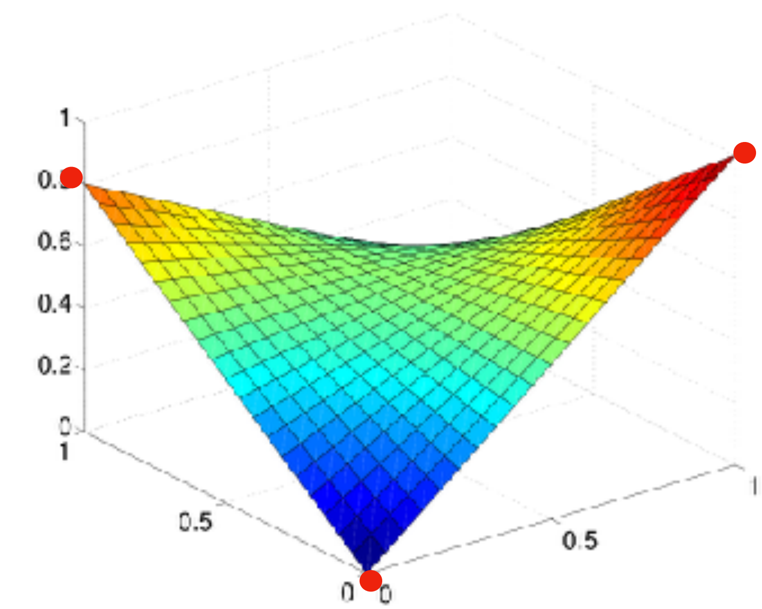


concave envelope is intractable, but …

Independent coordinates

$$f(z) = \mathbb{E}[F(S)]$$

$$P(S) = \prod_{i \in S} z_i \cdot \prod_{j \notin S} (1 - z_j)$$



- $\frac{\partial^2 f}{\partial x_i \partial x_j} \leq 0$ for all *i,j*
- $f(z)$ concave in increasing directions
  (diminishing returns)
- $f(z)$ convex in "swap" directions

- continuous maximization (monotone): despite nonconvexity!
  (Calinescu-Chekuri-Pal-Vondrak 2007, Feldman-Naor-Schwartz 2011,…, Hassani-Soltanolkotabi-Karbasi 2017, …)
- similar approach for non-monotone functions
  (Buchbinder-Naor-Feldman 2012,…)

“Continuous greedy” as Frank-Wolfe

Initialize: $z^0 = 0$

for t=1, . . . T:

$$s^t \in \arg\max_{s \in P} \langle s, \nabla f(z^t) \rangle$$

$$z^{t+1} = z^t + \alpha_t s^t$$

- concavity in positive directions:
  for all $z \in [0,1]^n$ there is a $v \in P$:

$$\langle v, \nabla f(z) \rangle \geq \mathrm{OPT} - f(z)$$

- Analysis:

$$\begin{aligned} f(z^{t+1}) \;&\geq\; f(z^t) + \alpha \langle s^t, \nabla f(z^t) \rangle - \tfrac{C}{2}\alpha^2 \\ &\geq\; f(z^t) + \alpha[\mathrm{OPT} - f(z^t)] - \tfrac{C}{2}\alpha^2 \end{aligned}$$

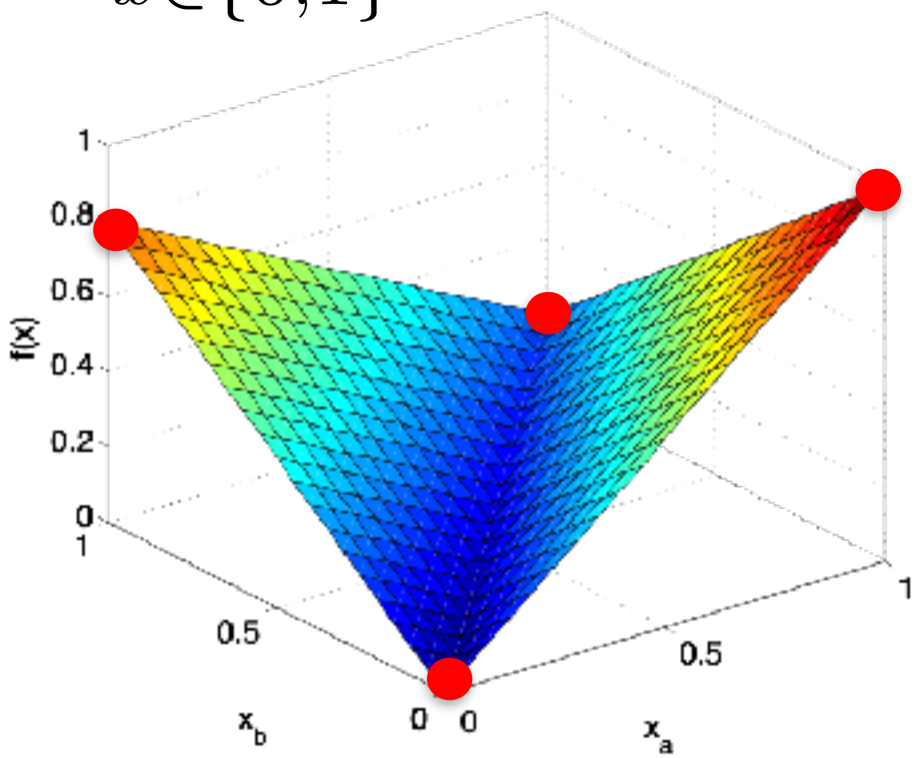
$$\Rightarrow \;\mathrm{OPT} - f(z^{t+1}) \;\leq\; (1-\alpha)[\mathrm{OPT} - f(z^t)] + \tfrac{C}{2}\alpha^2$$

- with $\alpha = 1/T$

$$f(z^T) \;\geq\; (1 - (1 - \tfrac{1}{T})^T)\mathrm{OPT} - \tfrac{C}{2T}$$

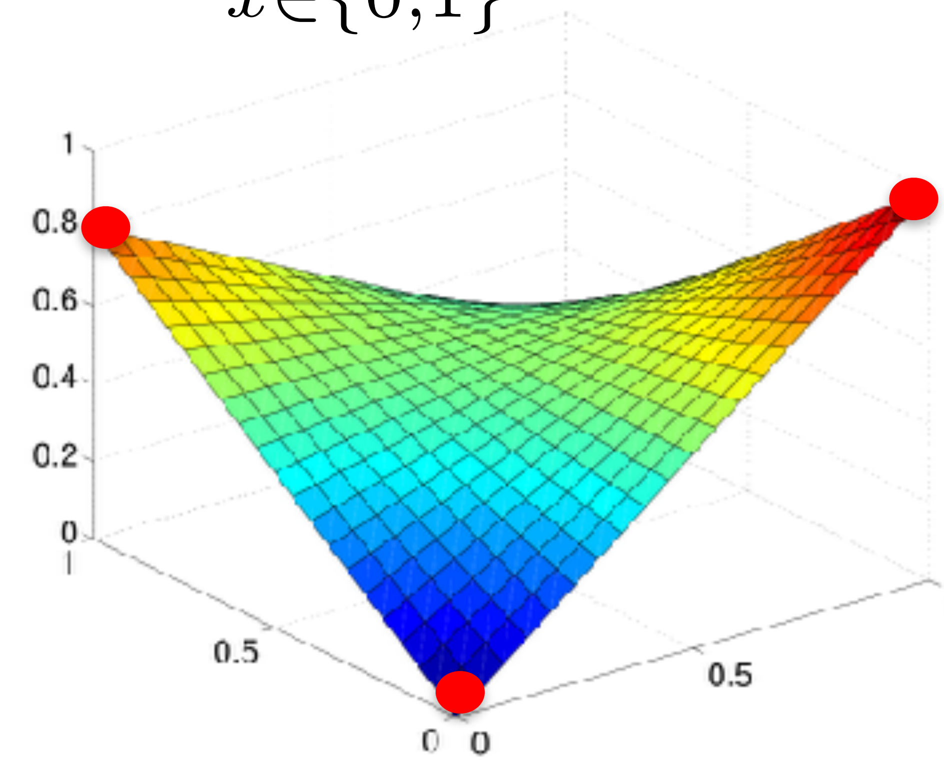
Binary / Set function optimization

$$\min_{x \in \{0,1\}^n} f(x)$$



- exact convex relaxation
- Lovász extension
- But: constrained is hard
- convexity

$$\max_{x \in \{0,1\}^n} f(x)$$



- NP-hard
- But: constant-factor approximations for constraints
- multilinear extension
- diminishing returns

Roadmap

- Optimizing submodular set functions:
  discrete optimization via continuous optimization

- Submodularity more generally:
  continuous optimization via discrete optimization

- Further connections

Submodularity beyond sets

- sets: for all subsets $A, B \subseteq V$

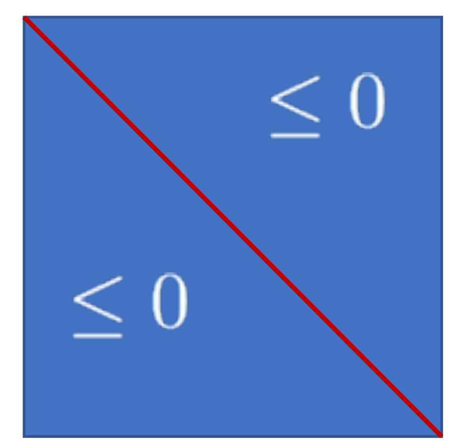
$$F(A) + F(B) \geq F(A \cup B) + F(A \cap B)$$

- replace sets by vectors:

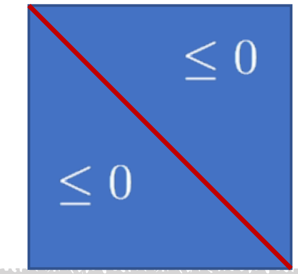
$$F(x) + F(y) \geq F(x \vee y) + F(x \wedge y)$$

- or: Hessian has all off-diagonals <= 0. *(Topkis 1978)*

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

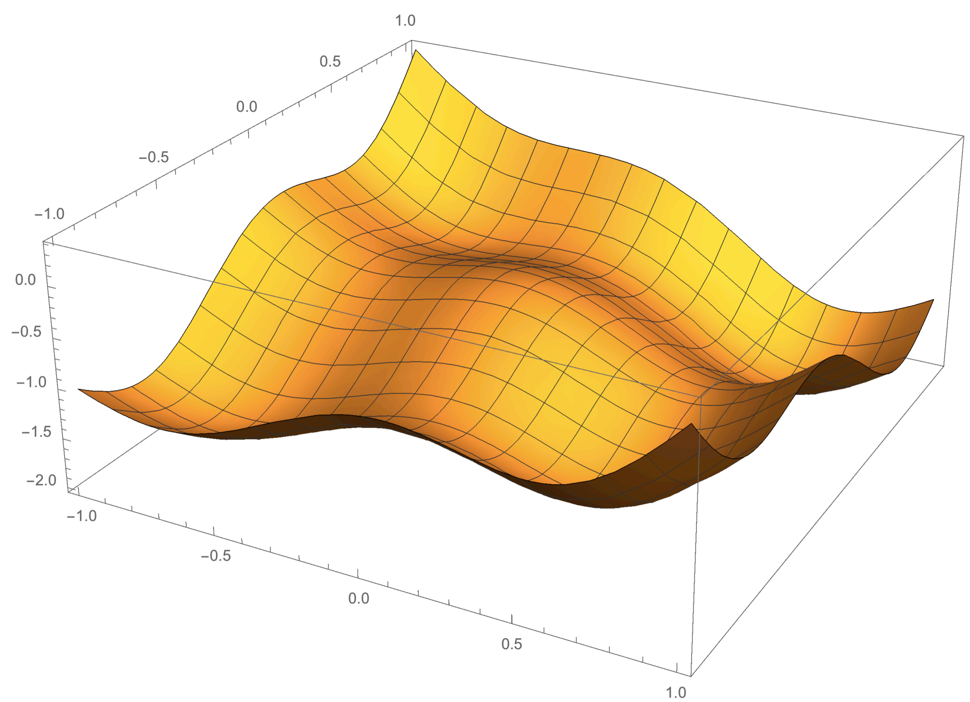


Examples



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

$$\frac{\partial^2 F}{\partial x_i \partial x_j} \leq 0 \quad \forall i \neq j$$



submodular function can be convex, concave or neither!

- any separable function $F(x) = \sum_{i=1}^{n} F_i(x_i)$
- $F(x) = g(x_i - x_j)$ for concave g
- $F(x) = h(\sum_i x_i)$ for convex h

Maximization

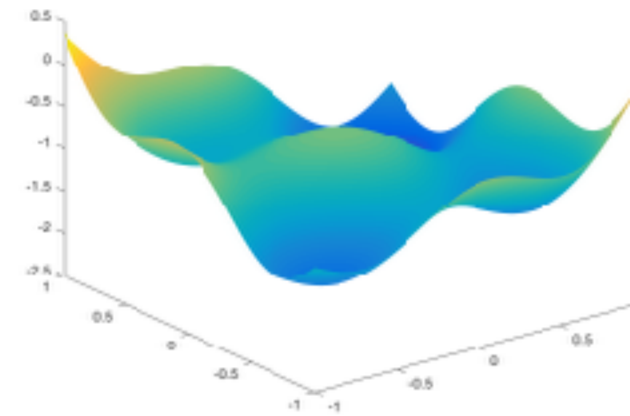
- General case:
  diminishing returns stronger than submodularity
- DR-submodular function:

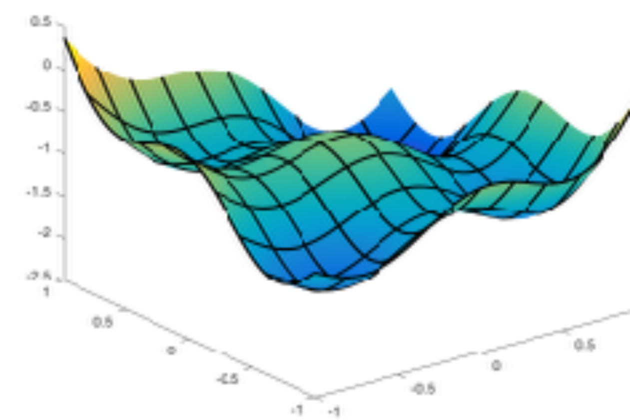
$$\partial^2 F / \partial x_i \partial x_j \leq 0 \quad \text{for } \textbf{\textit{all}} \quad i, j$$

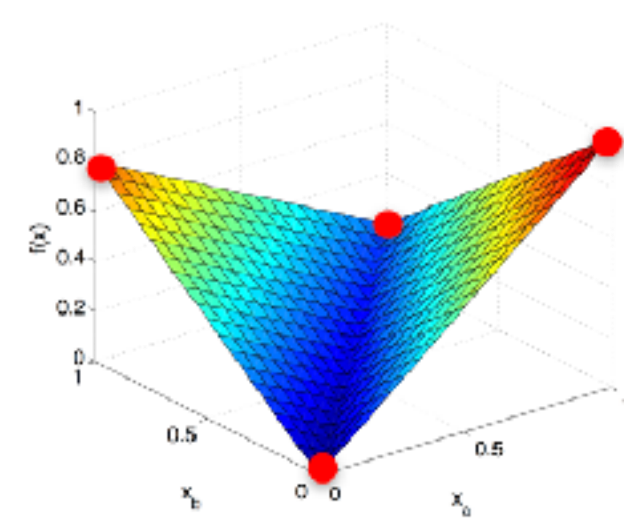
- with DR, many results generalize
  (including "continuous greedy")
  (Kapralov-Post-Vondrák 2010, Soma et al 2014-15, Ene & Nguyen 2016, Bian et al 2016, Gottschalk & Peis 2016)

Minimization

- discretize continuous functions: factor $O(1/\epsilon)$
- Option I:
  transform into set function optimization
  (Birkhoff 1937, Schrijver 2000, Orlin 2007)
  better for DR-submodular
  (Ene & Nguyen 2016)
- Option II:
  convex extension for integer submodular function *(Bach 2015)*



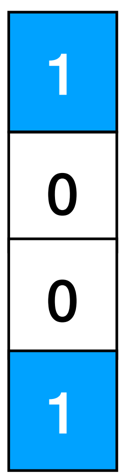




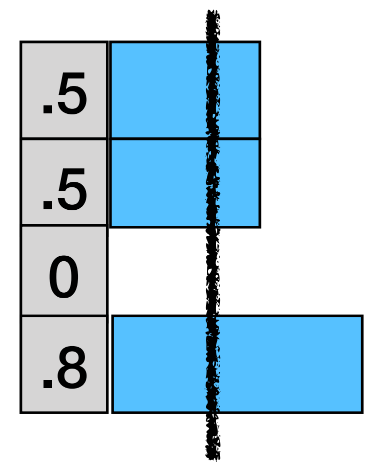
Convex extension

- **Set functions:** efficient minimization via convex extension

$F : \{0,1\}^n \to \mathbb{R}$

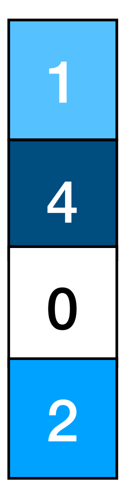


$f : [0,1]^n \to \mathbb{R}$

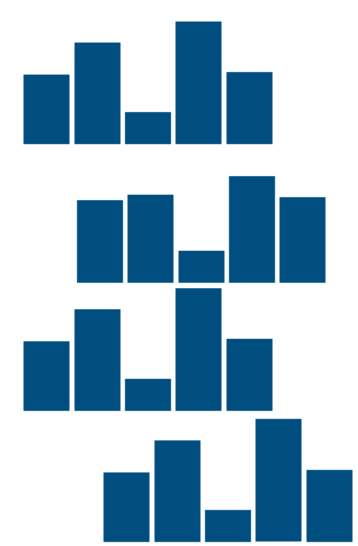


$f(z) = \mathbb{E}[F(S)]$

- **Integer vectors:** distribution over {*0*,…*k*} for each coordinate



$F : \{0, \ldots k\}^n \to \mathbb{R}$

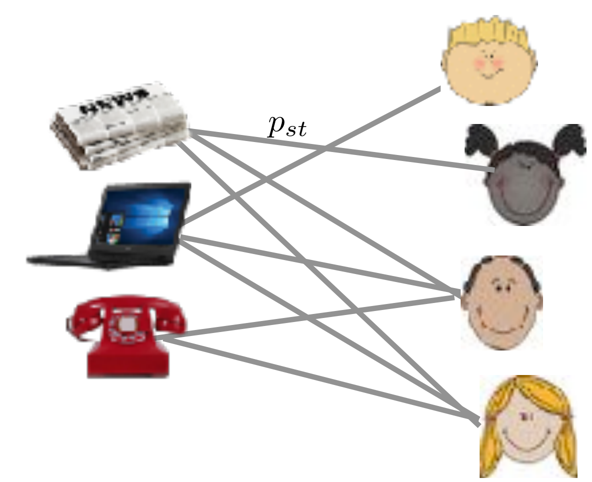


$f(z) = \mathbb{E}[F(x)]$

Applications

- robust optimization of bipartite influences *(Staib-Jegelka 2017)*

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



- non-convex isotonic regression *(Bach 2017)*

$$\min_{x \in [0,1]^n} \sum_{i=1}^{n} G(x_i - z_i) \quad \text{s.t. } x_i \geq x_i \, \forall (i,j) \in E$$

Roadmap

- Optimizing submodular set functions:
  discrete optimization via continuous optimization

- Submodularity more generally:
  continuous optimization via discrete optimization

- Further connections

Log-sub/supermodular distributions

$$P(S) \propto \exp(F(S)) \qquad\qquad P(x) \propto \exp(F(x))$$

- *-F(S)* submodular: multivariate totally positive, FKG lattice condition

$$P(S)\,P(T) \;\leq\; P(S \cup T)\,P(S \cap T)$$

  - implies *positive association*:
    for all monotonically increasing *G,H*:

$$\mathbb{E}[G(S)H(S)] \;\geq\; \mathbb{E}G(S)\mathbb{E}H(S)$$

- F(S) submodular?

$$P(S)\,P(T) \;\geq\; P(S \cup T)\,P(S \cap T)$$

Negative association and stable polynomials

- sub-class satisfies *negative association*:
  for all monotonically increasing *G,H with disjoint support:*

$$\mathbb{E}[G(S)H(S)] \leq \mathbb{E}G(S)\mathbb{E}H(S)$$

- Condition implies conditionally negative association:

$$q(z) = \sum_{S \subseteq V} P(S) \prod_{i \in S} z_i, \quad z \in \mathbb{C}^n$$

should be *real stable.* *Strongly Rayleigh measures*

(Borcea, Bränden, Liggett 2009)

Implications

- Concentration of measure *(Pemantle-Peres 2011)*
- *P(|S|)* log-concave
- Fast-mixing Markov Chains
  (Feder-Mihail 1982, …, Anari-Oveis-Gharan-Rezaei 2016, Li-Sra-Jegelka 2016)
- Approximate partition functions / counting and optimization
  (Gurvits 2006, Nikolov-Singh 2016, Straszak-Vishnoi 2016, …)
- …

Summary

Optimizing submodular set functions:
discrete optimization via continuous optimization

- extensions via expectations
- convex and partially concave

Further connections:

- Submodularity more generally:
  continuous optimization via discrete optimization
- Negative dependence and stable polynomials