Information Theory + Polyhedral Combinatorics

Sebastian Pokutta
Georgia Institute of Technology
ISyE, ARC

Joint work Gábor Braun

Information Theory in Complexity Theory and Combinatorics
Simons Institute
Berkeley, April 2015

Problems and LPs

Georgia Tech

Approximation Problems

An **approximation problem P** (max or min problem):

S: set of feasible solutions

F: set of considered objective functions (for simplicity: nonnegative)

F^*: approximation guarantees, $f^* \in \mathbb{R}$ for each $f \in F$

satisfying

$\max_{s \in S} f(s) \leq f^*$ (max problem) or $\min_{s \in S} f(s) \geq f^*$ (min problem)

Example (exact min Vertex Cover): Given a graph G

S: all vertex covers of graph G (i.e., subsets of nodes covering all edges)

F: all nonnegative weight vectors on vertices

F^*: define $f^* := \min_{s \in S} f(s)$

LPs capturing Approximation Problems

Model of [Chan, Lee, Raghavendra, Steurer 13] and [Braun, P., Zink 14]

An **LP formulation of an approximation problem $P = (S, F, F^*)$** is a linear program $Ax \leq b$ with $x \in \mathbb{R}^d$ and *realizations*:

a) *Feasible solutions:* for every $s \in S$ we have $x^s \in \mathbb{R}^d$ with

$Ax^s \leq b$ for all $s \in S$, (relaxation $conv(x^s \mid s \in S)$)

b) *Objective functions:* for every $f \in F$ we have an affine $w^f : \mathbb{R}^d \to \mathbb{R}$ with

$w^f(x^s) = f(s)$ for all $s \in S$, (linearization that is exact on S)

c) *Achieving approximation:* for every $f \in F$

$$\hat{f} = \max\{ w^f(x) \mid Ax \leq b \} \leq f^*$$

(κ, τ)-approximation: $\hat{f} \leq \kappa$ whenever $\max_{s \in S} f(s) \leq \tau$ for $f \in F$

Formulation Complexity

Approximation Problem (S, F, F^*)	→	Slack matrix of problem $M(f,s) = f^* - f(s)$	→	LP factorization $M = T \cdot U + \mu \cdot \mathbf{1}$ (restr. NMF)

Factorization theorem. Let $P = (S, F, F^*)$ be a problem and M slack matrix of P

$$\mathrm{fc}(P) = rank_{LP}(M)$$

Optimal LP. $x \geq 0$ with encodings

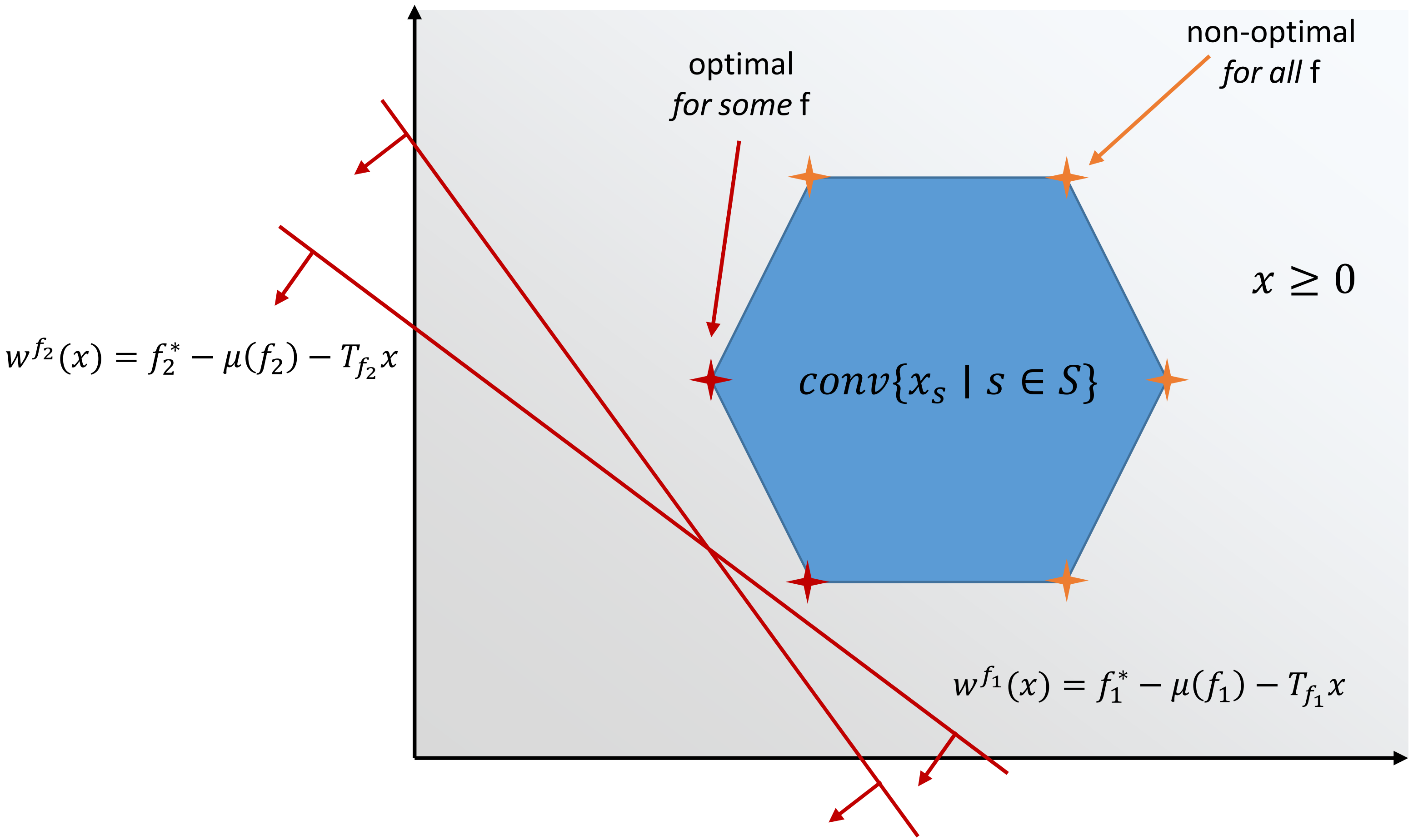
feasible solutions: $x^s := U_s$ **objective functions:** $w^f(x) := f^* - \mu(f) - T_f \cdot x$

Formulation complexity. generalization of extension complexity

- Independent of P vs. NP
- Independent of a specific polyhedral representation
- = Minimum extension complexity over all possible linear encodings
- Do not lift given representation but *construct* the optimal LP from factorization
- In fact: LP is trivial. Construct optimal encoding from factorization
- Restricted notion of nonnegative matrix factorization to support approximations

Georgia Tech

Optimal LPs



optimal
for some f
non-optimal
for all f
$x \geq 0$
$conv\{x_s \mid s \in S\}$
$w^{f_2}(x) = f_2^* - \mu(f_2) - T_{f_2}x$
$w^{f_1}(x) = f_1^* - \mu(f_1) - T_{f_1}x$

Information Theory + LPs

Information Theory: Summary

Entropy. $H[X] \coloneqq \sum_{x\in\Omega} P[X] \cdot \log\frac{1}{P[X]}$

Joint Entropy. $H[X,Y] = H[X] + H[Y|X]$

Mutual Information. $I[X;Y] \coloneqq H[X] - H[X|Y]$,

(how much information about X is leaked by observing Y)

Chain Rule. $I[(X,Y);Z] = I[X;Z] + I[Y;Z|X]$

Direct Sum Property. $Z = (Z_1, \ldots, Z_n)$ be a mutually independent

$$I[X;Z] \geq \sum_{i\in[n]} I[X;Z_i]$$

Hellinger Distance. Π_1, Π_2 distributions

$$h^2(\Pi_1, \Pi_2) = 1 - \sum_{\pi} \sqrt{P[\Pi_1 = \pi] \cdot P[\Pi_2 = \pi]}$$

Georgia Tech

NMF and Information Theory

NMF and distributions. Let $(F, S) \sim M/\,||M||_1$ and $M = \sum_\pi f_\pi s_\pi^T$ with $f_\pi, s_\pi \geq 0$.

NMF => writing complicated distribution as mix of product distributions

Common information. M nonnegative matrix, Z conditional [Wyner, 1975]

$$C[M \mid Z] \coloneqq \inf_{\substack{\Pi:\text{NMF of M} \\ \Pi \perp Z | F,S}} I[F,S;\Pi \mid Z].$$

Lower bounding $\mathrm{rk}_+(M)$. M nonnegative matrix

$$C[M \mid Z] \leq \inf_{\substack{\Pi:\text{NMF of } M \\ \Pi \perp Z | F,S}} H[\Pi \mid Z] \leq \log rk_+(M)$$

Georgia Tech

NMF and Information Theory

Cut-and-Paste (for NMF). M nonnegative matrix, $\Pi_{a,b} \coloneqq \Pi | A = a, B = b$

$$\sqrt{M(f_1, s_1) \cdot M(f_2, s_2)} \left(1 - h^2\left(\Pi_{f_1,s_1}; \Pi_{f_2,s_2}\right)\right)$$
$$= \sqrt{M(f_1, s_2) \cdot M(f_2, s_1)} \left(1 - h^2\left(\Pi_{f_1,s_2}; \Pi_{f_2,s_1}\right)\right)$$

1	1
1	0

=> Information-theoretic fooling set method

$$0 = 1 \cdot \left(1 - h^2\left(\Pi_{f_1,s_2}; \Pi_{f_2,s_1}\right)\right) \Leftrightarrow h^2\left(\Pi_{f_1,s_2}; \Pi_{f_2,s_1}\right) = 1$$

(f_1, s_2) and (f_2, s_1) cannot be in the same rank-1 factor.

Georgia Tech

General strategy. Let M be a slack matrix. Bound $I[F,S;\Pi \mid Z]$ for all possible Π:

1. Identify a conditional Z decomposing $I[F,S;\Pi \mid Z]$ via direct sum theorem:

$$I[F,S;\Pi \mid Z] \geq \sum_{i=1,\ldots,l} I[F_i,S_i;\Pi \mid Z] \geq l \cdot \min_i I[F_i,S_i;\Pi \mid Z]$$

   where for each i we have a smaller sub-problem.

2. Lower bound $I[F_i,S_i;\Pi \mid Z]$ via polyhedral/inf-theoretic argument:

$$I[F_i,S_i;\Pi \mid Z] \geq C$$

This then suffices:

$$\Rightarrow \mathrm{fc(P)} = \mathrm{rk}_+(M) \geq 2^{l \cdot C}$$

Nice side effect. We automatically get inapproximability results (due to continuity).

Today: Only indication of these steps.

Correlation Polytope

The correlation polytope

Functions. For any $b \in \{0,1\}^n$

$$f_b(x) := (1 - x^T b)^2$$

Solutions. For any $x \in \{0,1\}^n$

$$s_x := x$$

Associated **Slack Matrix.** $M_n(x, b) := (1 - x^T b)^2$

=> Contains UDISJ matrix as submatrix

Polyhedral equivalent is **correlation polytope**

$$\mathrm{COR}(n) := \mathrm{conv}\{xx^T \mid x \in \{0,1\}^n\}$$

Georgia Tech

The correlation polytope

UDISJ submatrix as **probability distribution.** For some $c > 0$

$$P[A = a, B = b] = \begin{cases} c & \text{if } a \cap b = \emptyset \\ c(1-\varepsilon) & \text{if } |a \cap b| = 1 \end{cases}$$

Decomposing **conditional.**

1. $C = (C_1, \dots, C_n)$ independent fair coins
2. New RVs $D = (D_1, \dots, D_n)$ with $D_i = \begin{cases} A_i & \text{if } C_i = 0 \\ B_i & \text{if } C_i = 1 \end{cases}$

=> Conditioning on $D = 0, C$ ensures $\{(A_i, B_i) : i \in [n]\}$ are independent

With this conditional (for minimal Π):

$$\log rk_+(M) \geq I[A, B; \Pi \mid D = 0, C] \geq \sum_{i \in [n]} I[A_i, B_i; \Pi \mid D = 0, C] \geq \varepsilon/8 \cdot n$$

Georgia Tech

The case $n = 1$

Consider the term:

$$I[A_1, B_1; \Pi \mid D = 0, C] = \frac{I[A_1, B_1; \Pi \mid A_1 = 0] + I[A_1, B_1; \Pi \mid B_1 = 0]}{2}$$

With $\Pi_{a,b} \coloneqq \Pi \mid A = a, B = b$ we have (Lemma by Bar-Yossef et al.)

$$I[A_1, B_1; \Pi \mid A_1 = 0] \geq \boldsymbol{h^2(\Pi_{00}; \Pi_{01})}$$
$$I[A_1, B_1; \Pi \mid B_1 = 0] \geq \boldsymbol{h^2(\Pi_{00}; \Pi_{10})}$$

Not a smart idea though: $h^2(\Pi_{00}; \Pi_{01}) = 0$ possible as $00, 01$ can be in the same rank-1 factor. (Similarly for $h^2(\Pi_{00}; \Pi_{10}) = 0$)

Georgia Tech

Simultaneous estimation via CS and Δ-inequality

$$\frac{I[A_1, B_1; \Pi \mid A_1 = 0] + I[A_1, B_1; \Pi \mid B_1 = 0]}{2} \geq \frac{h^2(\Pi_{00}; \Pi_{01}) + h^2(\Pi_{00}; \Pi_{10})}{2}$$

$$\geq \frac{(h(\Pi_{00}; \Pi_{01}) + h(\Pi_{00}; \Pi_{10}))^2}{4}$$ **(Cauchy—Schwarz Inequality)**

$$\geq \frac{h^2(\Pi_{10}; \Pi_{01})}{4}$$ **(*Δ*—Inequality)**

Apply **Cut-and-Paste.**

$$h^2(\Pi_{10}; \Pi_{01}) \geq 1 - \sqrt{\frac{M(0,0) \cdot M(1,1)}{M(0,1) \cdot M(1,0)}} \geq 1 - \sqrt{1 - \varepsilon} \geq \frac{\varepsilon}{2}$$

1	1
1	$1 - \varepsilon$

Average case hardness for COR(n)

Theorem. Any LP approximating $\mathrm{COR}(n)$ within a factor $n^{1-\varepsilon}$ is of size $2^{\frac{\varepsilon}{8}n}$.

Note: same result was obtained earlier by [Braverman, Moitra 13] (see talk)

However, common information captures all types of average case hardness:

Perturbation	$\mathbf{Log\ rk_+ \geq}$	Remarks
UDISJ	$\frac{6-3\log 3}{4}\ n$	(optimal estimation)
Shifts of UDISJ	$\frac{1}{8\rho}\ n$	$(\rho-1)$-shift
Sets of fixed size $\frac{n}{4} + O(n^{1-\varepsilon})$	$\frac{1}{8\rho}\ n - O(n^{1-\varepsilon)})$	
Random $2^{(1-\alpha)\ n} \times 2^{(1-\beta)\ n}$	$(\frac{1}{8\rho} - \alpha - \beta)\ n$	In expectation
Adversarial $(1-\alpha)\ 2^n \times (1-\beta)\ 2^n$	$\left(\frac{1}{8\rho} - \alpha - \beta\right) n - \log 3$	Removal of fraction per size

Georgia Tech

The matching problem – a much more complicated case

Via a generalization of Razborov's technique:

Theorem. [Rothvoss 14] Any LP formulation of the matching polytope is of exponential size.

This is very special and important:

1. Matching can be solved in polynomial time
2. Yet any LP capturing it is of exponential size

=> Separates the power of P from polynomial size LPs

With common information: ruling out the existence of FPTAS-type LP formulations

Theorem. [Braun, P. 14] For some $\varepsilon > 0$ any LP approximating the Matching Polytope within a factor $1 + \frac{\varepsilon}{n}$ is of exponential size.

Georgia Tech

Thank you!