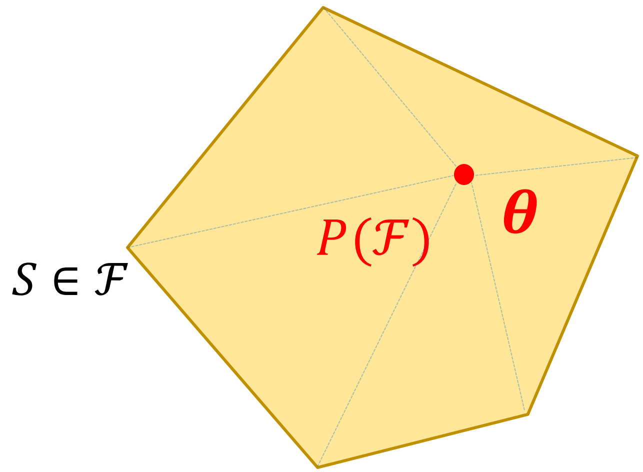
COMPUTING MAX-ENTROPY DISTRIBUTIONS



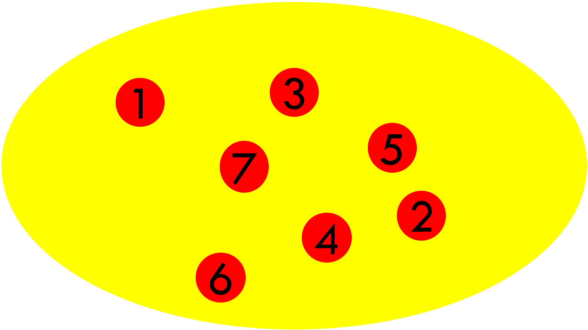
Nisheeth Vishnoi

Joint work with Mohit Singh and Damian Straszak

FOCS, October 6, 2018

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Sampling from Discrete Spaces



Universe $E = [m]$ of elements

$\mathcal{F} \subseteq 2^{[m]}$ -- ***family of allowed sets*** (possibly exponential)

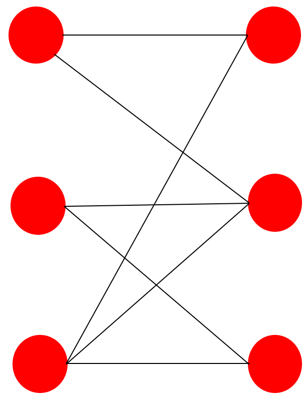
$A: E \to \mathbb{R}_{\geq 0}$ -- ***weights*** on elements

Goal: Output a set $S \in \mathcal{F}$ with probability proportional to $A_S = \prod_{e \in S} A_e$ or compute $\sum_{S \in \mathcal{F}} A_S$ (Partition function)

Applications: Physics, Mathematics, Information Theory, Statistics, Machine Learning, TCS

Perfect Matchings in Bip. Graphs

Weighted Bipartite Graph: $G = ([m] = [n] \times [n], A \in \mathbb{R}_{\geq 0}^{n \times n})$



$\mathcal{M} \subseteq 2^{[m]}$ be the ***set of perfect matchings*** in G

Fact: $P(\mathcal{M}) = \{\Theta \in \mathbb{R}_{\geq 0}^{n \times n} : \forall i \; \sum_{j=1}^{n} \Theta_{ij} = 1 \;\text{ and } \forall j \; \sum_{i=1}^{n} \Theta_{ij} = 1\}$

Sampling Problem: For $M \in \mathcal{M}$, $\mathbb{P}[M] \propto \prod_{ij \in M} A_{ij}$

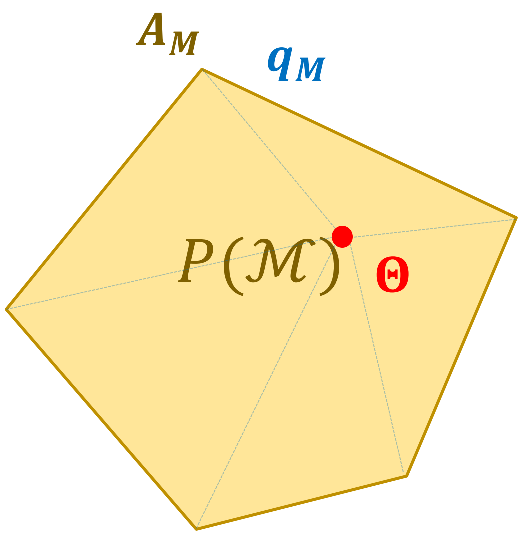
Counting Problem: $\sum_{M \in \mathcal{M}} \prod_{ij \in M} A_{ij} = \mathrm{Per}\,(A)$

Valiant '79: #P-Hard to compute $\mathrm{Per}(A)$

Jerrum-Sinclair-Vigoda '04: FPRAS for $\mathrm{Per}(A)$ using MCMC

Linial-Samorodnitsky-Wigderson '00: e^n-approximation for $\mathrm{Per}(A)$ using SCALING

Entropy and Counting



$\text{PRIMAL}(\Theta) =$

$$\sup_q \sum_{M \in \mathcal{M}} q_M \log \frac{A_M}{q_M}$$

q – prob. distribution over $\mathcal{M}$

The expectation of q is Θ

Fact 1: $\text{PRIMAL}(\Theta) \leq \log \text{Per}(A) \quad \forall\, \Theta \in P(\mathcal{M})$

Fact 2: $\text{PRIMAL}(\tilde{\Theta}) = \log \text{Per}(A)$ for $\tilde{\Theta} = \frac{1}{\sum A_M} \sum_M 1_M A_M$

Corollary: $\sup_{\Theta \in P(\mathcal{M})} \text{PRIMAL}(\Theta) = \log \text{Per}(A)$

Nothing special about perfect matchings -- holds for any $\mathcal{F}$!

Duality

$\mathrm{PRIMAL}(\Theta) =$

$$\sup_q \; \sum_{M\in\mathcal{M}} q_M \log \frac{A_M}{q_M}$$

$$\sum_{M\in\mathcal{M}} q_M = 1\,;\, q_M \geq 0$$

$$\forall \boldsymbol{ij},\; \sum_{\boldsymbol{M\in\mathcal{M};ij\in M}} \boldsymbol{q_M} = \boldsymbol{\Theta_{ij}}$$

$\mathrm{DUAL}(\Theta) =$

$$\inf_{\lambda\in\mathbb{R}^{n\times n}} \log \sum_M A_M e^{\langle\lambda, 1_M\rangle} - \langle\lambda, \Theta\rangle$$

Substitute $\mathbf{e}^{\lambda_{ij}} = \boldsymbol{z_{ij}}$ to obtain $\mathrm{DUAL}(\Theta) = \inf_{z\in\mathbb{R}^{n\times n}_{>0}} \log \sum_M A_M \prod_{ij\in M} z_{ij} - \sum_{i,j} \Theta_{ij} \log z_{ij}$

Exponentiate to obtain $\mathrm{DUAL}(\Theta) = \inf_{z\in\mathbb{R}^{n\times n}_{>0}} \frac{\sum_M A_M \prod_{ij\in M} z_{ij}}{\prod_{i,j} {z_{ij}}^{\Theta_{ij}}}$

Corollary [Folklore]: $\mathbf{sup}_{\boldsymbol{\Theta\in P(\mathcal{M})}} \, \mathbf{inf}_{\boldsymbol{z\in\mathbb{R}^{n\times n}_{>0}}} \frac{\sum_M A_M \prod_{ij\in M} z_{ij}}{\prod_{i,j} {z_{ij}}^{\Theta_{ij}}} = \mathrm{Per}(A)$

A Relaxation

$$\sup_{\Theta \in P(\mathcal{M})} \inf_{z \in \mathbb{R}^{n\times n}_{>0}} \frac{\boldsymbol{p}(\boldsymbol{z})}{\prod_{i,j} {z_{ij}}^{\Theta_{ij}}} = \mathrm{Per}(A)$$

Hard to evaluate!

Easy to evaluate!

Observation: $\boldsymbol{P(\mathcal{M}) = P(\mathcal{M}_1)} \cap P(\mathcal{M}_2)$ (intersection of 2 matroids)

$P(\mathcal{M}_1) = \{\Theta \in \mathbb{R}^{n\times n}_{\geq 0} : \forall j \ \sum_{i=1}^{n} \Theta_{ij} = 1 \ \}$ and $P(\mathcal{M}_2) = \{\Theta \in \mathbb{R}^{n\times n}_{\geq 0} : \forall i \ \sum_{j=1}^{n} \Theta_{ij} = 1 \ \}$

Let $\tilde{p}(z) = \prod_i \sum_j A_{ij} z_{ij}$ -- ***Newton polytope*** of $\tilde{p}$ is $P(\mathcal{M}_2)$ (If $A > 0$)

New Relaxation: $\sup_{\Theta \in P(\mathcal{M}_1)} \inf_{z \in \mathbb{R}^{n\times n}_{>0}} \frac{\tilde{p}(z)}{\prod_{i,j} {z_{ij}}^{\Theta_{ij}}}$

Substitute: $x_i = z_{ij}$ for all j

Recovering Gurvits' capacity-based relaxation [Straszak-V. 17a]:

$$\sup_{\Theta \in P(\mathcal{M}_1)} \inf_{z \in \mathbb{R}^{n\times n}_{>0}} \frac{\tilde{p}(z)}{\prod_{i,j} {z_{ij}}^{\Theta_{ij}}} = \inf_{x \in \mathbb{R}^{n}_{>0} 0} \frac{\prod_{i \in [n]} \sum_{j \in [n]} A_{ij} x_j}{\prod_i x_i}$$

Beyond Permanent

Input:

Evaluation oracle to polynomial $p \in \mathbb{R}_{\geq 0}[x_1, x_2, \ldots, x_m]$

$$p(x) = \sum_{S \subseteq [m]} p_S x^S$$

Separation oracle to the polytope of $\mathcal{F} \subseteq 2^{[m]}$

Problem

$\sum_{S \in \mathcal{F}} p_S$?

[**Applications:** Fair Sampling from determinantal measures [Celis et al. 16, 17, 18]]

Relaxation: $\mathrm{Cap}_{\mathcal{F}}(p) = \sup_{\Theta \in P(\mathcal{F})} \inf_{z \in \mathbb{R}^m_{>0}} \frac{p(z)}{\prod_i {z_i}^{\Theta_i}}$

Theorem [Straszak-V. STOC '17]: Assuming p is homogeneous and real-stable

$$\frac{\mathrm{Cap}_{\mathcal{F}}(p)}{M} \leq \sum_{S \in \mathcal{F}} p_S \leq \mathrm{Cap}_{\mathcal{F}}(p)$$

$\boldsymbol{M < \infty}$ whenever $\mathcal{F}$ ***supports a real-stable polynomial; depends only on*** $\mathcal{F}$

Similar and independent result by [Anari-OveisGharan'17]

Application: Rank-1 Brascamp-Lieb

Given m vectors $v_1, v_2, \dots, v_m \in \mathbb{R}^n$ and a $\theta \in \mathbb{R}^m_+$

Brascamp-Lieb Constant

$$\inf_{z \in \mathbb{R}^m_{>0}} \frac{\det \sum_j \theta_j z_j v_j v_j^\top}{\prod_i z_i^{\theta_i}} = \inf_{z \in \mathbb{R}^m_{>0}} \frac{\sum_{S \subseteq [m], |S|=n} \theta^S z^S \det V_S V_S^\top}{\prod_i z_i^{\theta_i}}$$

Complexity studied by [GargGurvitsOlivieraWigderson '16] – pseudo-polynomial time scaling-based algorithm in the bit complexity of θ (for the general BL case)

Can we compute $\inf_{z \in \mathbb{R}^m_{>0}} \frac{p(z)}{\prod_i z_i^{\theta_i}} = \inf_{y \in \mathbb{R}^m} \log p(e^y) - \sum \theta_i y_i$ ***for all*** $\theta \in P$**?**

Unlike permanent, not clear how to scale efficiently …

Ellipsoid Method

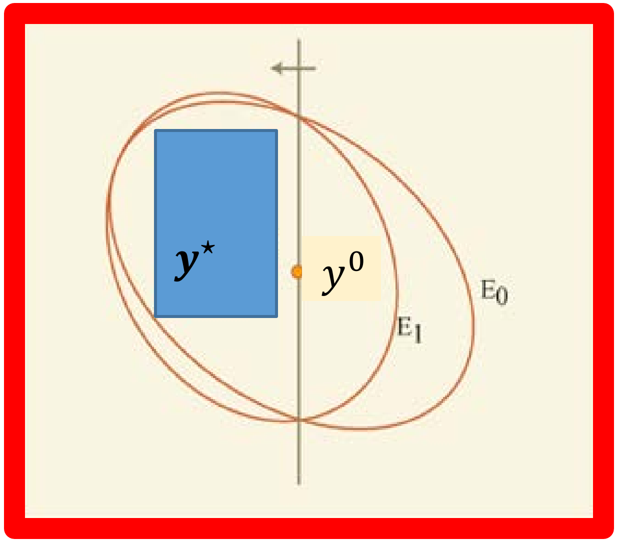
$$\textbf{OPT} = \inf_{y \in \mathbb{R}^m} \log p(e^y) - \sum \theta_i y_i = \inf_{y \in \mathbb{R}^m} f(y)$$

Reduce to Feasibility: Given A, check if **OPT** is $\leq A + \varepsilon$ or $> A$

Assume $\|y^\star\| \leq R, f \in [-F, F]$

Ellipsoid Algorithm:

- **Start** with an ellipsoid E_0 that contains $y^\star$
- At kth step, let E_k be the ellipsoid centered at y^k
  - **IF** $f(y^k) \leq A$, **DONE**
  - **ELSE**
    - use evaluation oracle for p to get $\nabla f(y^k)$
    - $E_{k+1} \supseteq E_k \cap \{y: \langle y - y^k, \nabla f(y^k) \rangle \leq 0\}$
- **Stop** when the radius of the ellipsoid becomes $\leq \varepsilon R/M$



Invariant: If $f(y^\star) \leq A$ then $y^\star \in E_k$ for all k

Proof: Convexity of f implies $\langle y^\star - y^k, \nabla f(y^k) \rangle + f(y^k) \leq f(y^\star) \leq A$

Since $f(y^k) > A$, $\langle y^\star - y^k, \nabla f(y^k) \rangle < 0$

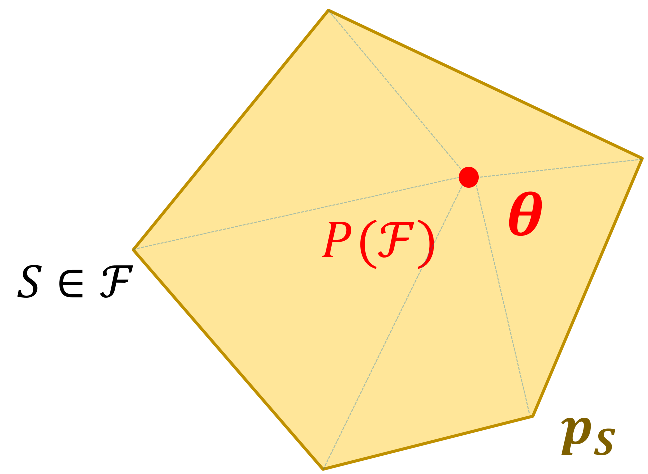
Running Time: $\mathrm{poly}(m, t_f, t_{\nabla f}, \log \frac{RF}{\varepsilon})$

Bounding R and M?

$$\inf_{y\in\mathbb{R}^m} \log p(e^y) - \sum \theta_i y_i$$

$$\sup_{q} \sum_{S\in\mathcal{F}} q_S \log \frac{p_S}{q_S}$$

$\Rightarrow F \leq m$



- q – prob. distribution over $\mathcal{F}$
- The expectation of q is θ

Bounding R?: As θ comes close to the boundary, $y^\star$ must blow up. By how much?

Special case of spanning tree polytope: [Asadpour et al. '10, Oveis-Gharan et al '11]

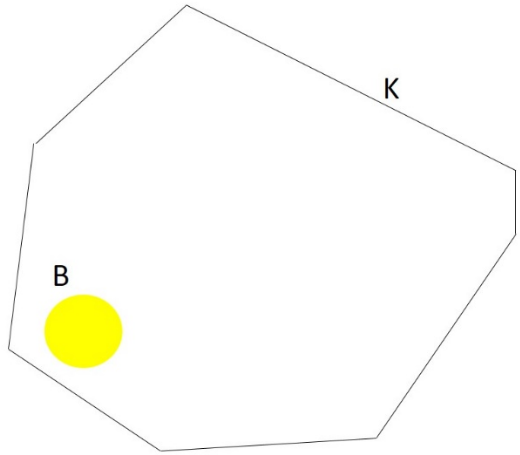
Theorem [SinghV. '14]: If $\Theta \in \mathrm{Int}_\eta(P)$ $R \leq \mathrm{poly}(1/\eta)$ – all combinatorial polytopes

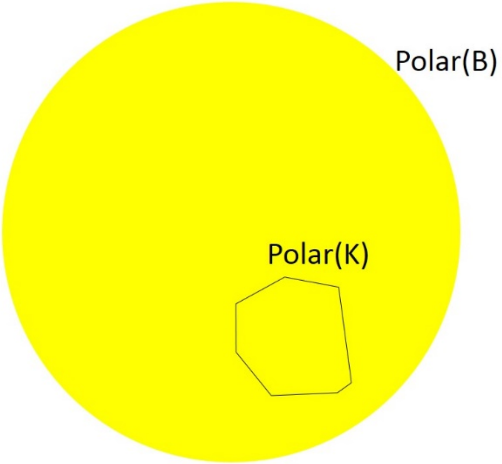
Theorem [StraszakV.]: If the unary complexity of all facets is polynomial in m then, $R \leq \mathrm{poly}(m)$ – includes almost all combinatorial polytopes ***(along with tightness result)***

Bounding Box 1

- **Theorem [SinghV. '14]:** If $\boldsymbol{\Theta} \in \mathbf{Int}_{\boldsymbol{\eta}}(\boldsymbol{P})$ then $\|\boldsymbol{\lambda}^{\star}\| = R \leq \mathrm{m}/\eta$
- ***Proof:*** Since entropy over a discrete set of size at most 2^m is at most $\boldsymbol{m}$
  - $\langle \lambda^*, \theta \rangle + \log \sum_S e^{-\langle \lambda^*, 1_S \rangle} = \sum_S \ln e^{\langle \lambda^*, \theta \rangle - \langle \lambda^*, 1_S \rangle} \leq m \Rightarrow \langle \lambda^*, \theta \rangle - \langle \lambda^*, 1_S \rangle \leq m \quad \forall S$
  - $\left\langle -\frac{\lambda^*}{m}, v - \theta \right\rangle \leq 1 \; \forall \, v \in \mathrm{convhull}(\{1_S\})$
    - (up to a centering and assuming convhull full dimensional)
  - $-\frac{\lambda^*}{m} \in \mathrm{polar} - \mathrm{convhull}$
  - θ is in η interior of convhull implies $\mathrm{polar} -$ convhull contained in ball of radius $\frac{1}{\eta}$

$\Rightarrow \|\lambda^*\| \leq m/\eta$





Bounding Box 2

- $g(\theta) = \inf_{y \in \mathbb{R}^m} h(\theta, y) := \inf_{y \in \mathbb{R}^m} \log \sum_{\alpha \in \mathcal{F}} p_\alpha e^{\langle \alpha - \theta, y \rangle}$

- **Theorem [Straszak-V.]**: Consider a polytope

$$P = \{x \in \mathbb{R}^m : \langle a_i, x \rangle \leq b_i \ \forall i \in I\} \cap H \text{ with } a_i \in \mathbb{Z}^m \text{ and } \|a_i\| \leq M$$

Then $\forall\, \theta \in P \ \ \exists y \in \mathbf{B}\left(\mathbf{0}, \mathbf{poly}\left(\boldsymbol{m}, \boldsymbol{M}, \max_{\alpha \in \mathcal{F}} \mathbf{log}\, |\mathbf{p}_\alpha|, \mathbf{log}\left(\frac{\boldsymbol{m}}{\boldsymbol{\varepsilon}}\right)\right)\right)$ *s.t.* $h(\theta, y) \leq g(\theta) + \varepsilon$

Proof: Let $y^\star$ be an optimal solution

Step1: Write $y^\star = \sum \beta_i a_i$ -- as a non-negative combination of normal vectors of tight constraints at a vertex $\alpha^\star$. Select an α that maximizes $\langle \alpha, y^\star \rangle$ and then use Farkas Lemma

Step 2: For $\boldsymbol{\Delta} = \mathbf{m} + \mathbf{M} + \max_{\alpha \in \mathcal{F}} \mathbf{log}\, |\mathbf{p}_\alpha| + \mathbf{log}\left(\frac{\boldsymbol{m}}{\boldsymbol{\varepsilon}}\right)$, let $y^\circ = \sum \min(\beta_i, \Delta) \quad a_i$. Then

$$h(\theta, y^\circ) \leq h(\theta, y^\star) + \varepsilon$$

Relies on the fact that the coefficients of the inequalities defining P are integral – thus for any α that does not lie on a facet $\langle a_i, x \rangle = b_i$, $\langle \alpha^\star - \alpha, a_i \rangle \geq 1$

Step 3: Thus $\|y^\circ\| \leq m\Delta \left\| a_i \right\| \leq poly(m)\Delta\mathrm{M}$.

Entropy interpretation seems important to obtain the bit complexity bounds

Summary and Challenges

- Resolved the bit complexity of max-entropy for a large class of polytopes
- Applications to constrained sampling/optimization, rank 1 Brascamp-Lieb, matrix scaling

- Tight bit complexity examples for 0/1 polytopes?
- Faster (scaling/interior point) algorithms for max-entropy?
- Polynomial time algorithm for Brascamp-Lieb constant for rank 2?

Thanks! Questions?