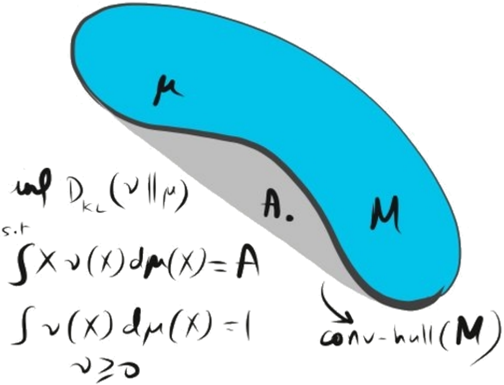
Continuous Maximum Entropy Distributions



Jonathan Leake (Weierstrass Institute / TU Berlin)

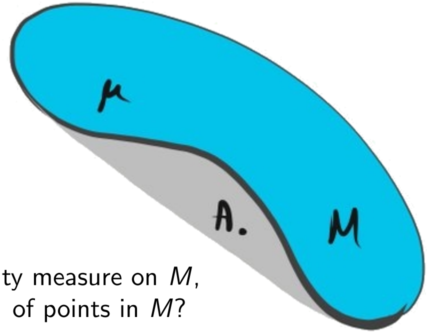
joint with

Nisheeth Vishnoi (Yale University)

A First Question

Suppose we have:

- M, a subset of $\mathbb{R}^d$,
- K, the convex hull of M, and
- A, a point in K.



Question:

Is there a canonical way to choose a probability measure on M, which expresses A as a "convex combination" of points in M?

When M is discrete:

Choose the distribution on M which **maximizes Shannon entropy** [Jaynes '57].

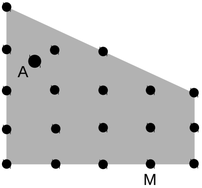
- Called **maximum entropy distributions**.
- Generalizable to minimization of **Kullback-Leibler divergence**.
- Applications to **combinatorial inequalities and discrete approximation**, algorithms for **metric TSP** approximation, **group orbit optimization and scaling algorithms**, and **statistical models**.

How do we actually compute this?

Computing Discrete Max-Entropy Distributions

Suppose we have:

- M, a subset of $\mathbb{R}^d$,
- K, the convex hull of M, and
- A, a point in K.



Primal	Dual
$\sup_{\substack{\text{supp}(\nu)\subseteq M \\ \mathbb{E}[\nu]=A}} -\sum_{x\in M} \nu(x)\log\nu(x)$	$\inf_{y\in\mathbb{R}^d} \left(\langle y, A\rangle + \log \sum_{x\in M} e^{-\langle y,x\rangle} \right)$

Dual program efficiently computable via **ellipsoid method**, given a **strong counting oracle** for the objective function [Singh-Vishnoi '15, Straszak-Vishnoi '17].

Observations:

- Dual search space dim. $\ll$ primal search space dim.
- Maximum entropy distribution $\implies \nu^\star \propto e^{-\langle y^\star, x\rangle}$.
- Dual program is **polynomial capacity** $\inf_{x>0} \frac{p(x)}{x^A}$ (see [Straszak-Vishnoi '17]).

An Aside: Polynomial Capacity

Last slide: Discrete maximum entropy $\iff$ polynomial capacity.

Polynomial capacity: $\inf_{x>0} \frac{p(x)}{x_1^{\alpha_1} \cdots x_d^{\alpha_d}}$ for $p \in \mathbb{R}[x]$, $\alpha \in \mathbb{R}^d$ due to [Gurvits '04].

Polynomial capacity / discrete max-entropy applications:

- Approximate/bound **permanent** of non-negative matrices, **mixed discriminant** of PSD matrices, **mixed volume** of convex bodies [Gurvits '00s].
- Approximate/bound **integer points/volume** of polytopes [Barvinok '00s, Barvinok-Hartigan '09, Gurvits '15, Gurvits-L '20, Csikvári-Schweitzer '20. Brändén-L-Pak '21].
- **Matroid** counting/optimization [Straszak-Vishnoi '17, Anari-Oveis Gharan '17, Anari-Liu-Oveis Gharan-Vinzant '19].
- **Metric TSP** approximation [Karlin-Klein-Oveis Gharan '21]

Capacity, beyond polynomials:

- Matrix, operator, tensor **scaling** [Garg-Gurvits-Oliveira-Wigderson '15, Franks '18, Bürgisser-Franks-Garg-Oliveira-Walter-Wigderson '18, Allen-Zhu-Garg-Li-Oliveira-Wigderson '18, van Apeldoorn-Gribling-Li-Nieuwboer-Walter-de Wolf '20, Bürgisser-Li-Nieuwboer-Walter '20].
- **Non-commutative optimization** and **orbit intersection** [Bürgisser-Garg-Oliveira-Walter-Wigderson '17, Allen-Zhu-Garg-Li-Oliveira-Wigderson '18, Bürgisser-Franks-Garg-Oliveira-Walter-Wigderson '19, Franks-Walter '20, Franks-Reichenbach '21].
- **Stat models** [Améndola-Kohn-Reichenbach-Seigal '21, Franks-Oliveira-Ramachandran-Walter '21]

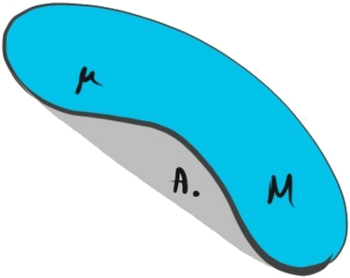
Back to Maximum Entropy Distributions

Discrete case: Efficiently computable, with many applications / generalizations.

What about non-discrete support?

Suppose we have:

- M, a **manifold** embedded in $\mathbb{R}^d$,
- K, the convex hull of M, and
- A, a point in K.



Maximum entropy program:

$$\sup_{\substack{\text{supp}(\nu)\subseteq M \\ \mathbb{E}[\nu]=A}} -\int_M \nu(x)\log\nu(x)dx$$

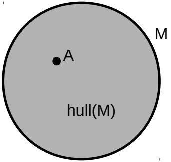
New problems:

- What is dx here? **No canonical notion of entropy.**
- Search space now **infinite-dimensional**.
- M can only be **implicitly expressed** (e.g., "all unit vectors").

Computing Continuous Max-Entropy Distributions

Suppose we have:

- M, a **manifold** embedded in $\mathbb{R}^d$,
- μ a **probability measure** on M,
- K, the convex hull of M, and
- A, a point in K.



Primal	Dual
$\sup_{\substack{\text{supp}(\nu)\subseteq M \\ \mathbb{E}[\nu]=A}} -\int_M \nu(x)\log\nu(x)d\mu(x)$	$\inf_{y\in\mathbb{R}^d}\left(\langle y, A\rangle + \log\int_M e^{-\langle y,x\rangle}d\mu(x)\right)$

Dual search space is **finite-dimensional** and by strong duality [L-Vishnoi '20], the maximum entropy distribution is of the form $d\nu^\star(x) \propto e^{-\langle y^\star, x\rangle} d\mu(x)$.

Questions remain:

- Why do we care about this?
- Efficient computability? (Now relies on an **integration oracle**.)
- Efficient sampling?

Motivating Example: Quantum Entropy

Von Neumann entropy: Derived from **discrete** distribution on eigenvector pure states:

$$\mathcal{H}_{\mathrm{vN}}(A) = -\operatorname{tr}(A \log A) = -\sum_{i=1}^{m} \lambda_i \log \lambda_i,$$

where λ_i are the eigenvalues of the **density matrix** A (PSD, $\operatorname{tr}(A) = 1$).

Issues:

- Critiqued by [Band-Park '76] as an indicator of uncertainty of density matrix A.
- Can be viewed as the entropy of a **minimum** entropy distribution ("most terse" description of A).

Quantum entropy: Derived from **continuous** distribution on **all** pure states:

$$\mathcal{H}_q(A) = \sup_{\mathbb{E}[\nu]=A} -\int_{\mathcal{P}_1} \nu(X) \log \nu(X) d\mu_1(X),$$

where $\mathcal{P}_1$ = all Hermitian **rank-one PSD projections**, $\operatorname{hull}(\mathcal{P}_1)$ = all **density matrices**, and μ_1 = the **Haar measure** on $\mathcal{P}_1$ (via action of unitary group).

Efficient computability: Low-dim. cases [Slater '91], general case* [L-Vishnoi '20]

Motivating Example: Interior Point Methods

Entropic barrier function for convex body $K \subset \mathbb{R}^d$:

$$B_K(v) = \sup_{y \in \mathbb{R}^d} \left[\langle y, v \rangle - \log \int_K e^{\langle y, x \rangle} dx \right].$$

- Barrier for K with **optimal self-concordance parameter** [Bubeck-Eldan '15].
- $-B_K(v)$ is the optimal value of **maximum entropy dual program**.
- Open manifold $K \subset \mathbb{R}^d$ with $\mu =$ **Lebesgue measure** restricted to K.

Open problem: Is the entropic barrier efficiently computable for K, given as a membership oracle?

Partial answer: **Yes**, given an efficient **integration oracle** for K.

Other connections beyond quantum entropy and interior point methods:

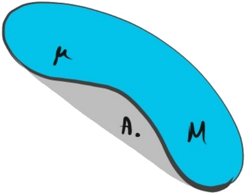
- Isotropic constant [Klartag '06], see also [Gromov '90]
- Langevin/Bingham distributions on matrix spaces [Khatri-Mardia '77, Chikuse '03]
- SDP rounding, from matrix to vector [Goemans-Williamson '95]

Question: What about efficient **computability** (and **sampling**)?

Computability of Continuous Max-Entropy Distributions

Suppose we have:

- M, a **manifold** embedded in $\mathbb{R}^d$,
- μ a **probability measure** on M,
- K, the convex hull of M, and
- A, a point in K.



Ellipsoid method-based algorithm [L-Vishnoi '20] to efficiently approximate the corresponding maximum entropy distribution whenever:

- μ is "balanced" - intuitively, μ is close to **uniform/symmetric**,
- A is not too near the boundary of K, and
- we have a **strong integration oracle** for μ; i.e. oracles for

$$F(Y) = \log \int_M e^{-\langle Y, X \rangle} d\mu(X) \qquad \text{and} \qquad \nabla F(Y).$$

Question: When can we obtain these **strong integration oracles**?

Integration Oracles via Symmetry

Focus: $M =$ **orbit of the unitary group** in the space of Hermitian matrices,

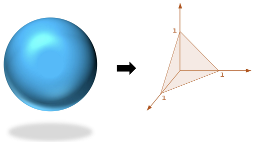
$$\mathcal{O}_\lambda := \{U \operatorname{diag}(\lambda) U^* : U \in \mathrm{U}(n)\} = \{\text{Hermitian matrices with e-vals } \lambda\}.$$

- **Invariant measure** μ_λ on $\mathcal{O}_\lambda$ derived from **Haar measure** on $\mathrm{U}(n)$.
- **Question:** What about the oracle for $\log \int_{\mathcal{O}_\lambda} e^{-\langle Y,X\rangle} d\mu_\lambda(X)$?

Quantum entropy case: $\lambda = e_1 = (1, 0, 0, \ldots, 0) \implies \mathcal{O}_\lambda =$ rank-1 PSDs.

Polytope connection: "Moment map" on $\mathcal{O}_{e_1}$:

$\pi : \mathcal{O}_{e_1} \to \mathbb{R}^n$ via $\pi : X \mapsto$ diag. entries of X



- Image of π acting on $\mathcal{O}_{e_1}$ is **the standard simplex** $\Delta_n = \mathrm{hull}(e_1, \ldots, e_n)$
- That is, $\pi(\mathcal{O}_{e_1}) = \Delta_n$, but **even better** $\pi(d\mu_{e_1}) = dx|_{\Delta_n}$:

$$\int_{\mathcal{O}_{e_1}} e^{-\langle Y,X\rangle} d\mu_{e_1}(X) = \int_{\Delta_n} e^{-\langle y,x\rangle} dx \qquad \text{for} \quad Y = \operatorname{diag}(y).$$

Bonus: Bijection between max-entropy distributions on $\mathcal{O}_{e_1}$ and Δ_n.

Integration Oracles via Symmetry

Focus: $M =$ **orbit of the unitary group** in the space of Hermitian matrices,

$$\mathcal{O}_\lambda := \{U \operatorname{diag}(\lambda) U^* : U \in \mathrm{U}(n)\} = \{\text{Hermitian matrices with e-vals } \lambda\}.$$

Last slide: $\mathcal{O}_{e_1} \iff \Delta_n$ connection: measures, max-entropy, integration.

What about other λ? Want to compute $\log \int_{\mathcal{O}_\lambda} e^{-\langle Y, X\rangle} d\mu_\lambda(X)$.

- Apply "moment map" π to $\mathcal{O}_\lambda$ to get a **polytope**:

$$\pi(\mathcal{O}_\lambda) = \operatorname{hull}\{\sigma \cdot \lambda : \sigma \in S_n\}.$$

- **Problem:** $\pi(d\mu_\lambda) =$ **piecewise polynomial density** [Duistermaat-Heckman '82].
- **Fix** is possible via different map/polytope (see **Colin's talk** next).
- **Solution:** The **Harish-Chandra-Itzykson-Zuber formula**:

$$\int_{\mathrm{U}(n)} e^{\operatorname{tr}(AUBU^*)} dU = \left(\prod_{p=1}^{n-1} p!\right) \frac{\det([e^{\alpha_i \beta_j}]_{i,j=1}^n)}{\prod_{i<j}(\beta_j - \beta_i)(\alpha_j - \alpha_i)}.$$

Question: Why do we care about these orbits? Other manifolds?

Integration Oracles via Symmetry

Previous slides: Integration oracles for Hermitian matrix unitary orbits:

$$\mathcal{O}_\lambda := \{U \operatorname{diag}(\lambda) U^* : U \in \mathrm{U}(n)\} = \{\text{Hermitian matrices with e-vals } \lambda\}.$$

Implies efficient algorithm for **computing max-entropy distributions** on $\mathcal{O}_\lambda$.

Question: Where do these distributions appear?

- The $\mathcal{O}_{e_1}$ case is equivalent to computing **quantum entropy**.
- The $\mathcal{O}_{e_1+e_2+\cdots+e_k}$ case is yields max-entropy distributions on the **complex Grassmanian**: utilized in **private low-rank approx.** [L-McSwiggen-Vishnoi '21]
- Complex Langevin/Bingham matrix distributions [Khatri-Mardia '77, Chikuse '03]
- **HCIZ densities** in physics, random matrix theory, ...

Question: What about other manifolds?

- Need a **measure** (via symmetry) and an **integration oracle**.
- **E.g.:** HCIZ formula extends to other adjoint orbits of **compact Lie groups**, with measure via the **Haar measure**. (see [L-Vishnoi '20])

Conclusion

This talk:

- **Discrete** maximum entropy distributions and applications: combinatorics, discrete approximation, orbit problems, norm minimization, etc.
- **Continuous** maximum entropy distributions and (**different**) applications: quantum entropy, interior point methods, low-rank approximation, etc.
- Specific examples related to **conjugation orbits of unitary groups**.

Upshot: In the continuous case, **manifold symmetries** give rise to measures and integration oracles which **make efficient computability possible**.

Question: But what about **sampling**?

- **Answer:** See next talk by Colin.

Open problems:

- Many **combinatorial** applications of **discrete** maximum entropy distributions. **Anything like this for the continuous case?**
- **Entropy interpretation** of the norm minimization generalization?

Thanks!