On the Computability of Continuous Maximum Entropy Distributions with Applications*

Jonathan Leake
KTH
Sweden

Nisheeth K. Vishnoi
Yale University
USA

ABSTRACT

We initiate a study of the following problem: Given a continuous domain Ω along with its convex hull $\mathcal{K}$, a point $A \in \mathcal{K}$ and a prior measure μ on Ω, find the probability density over Ω whose marginal is A and that minimizes the KL-divergence to μ. This framework gives rise to several extremal distributions that arise in mathematics, quantum mechanics, statistics, and theoretical computer science. Our technical contributions include a polynomial bound on the norm of the optimizer of the dual problem that holds in a very general setting and relies on a "balance" property of the measure μ on Ω, and exact algorithms for evaluating the dual and its gradient for several interesting settings of Ω and μ. Together, along with the ellipsoid method, these results imply polynomial-time algorithms to compute such KL-divergence minimizing distributions in several cases. Applications of our results include: 1) an optimization characterization of the Goemans-Williamson measure that is used to round a positive semidefinite matrix to a vector, 2) the computability of the entropic barrier for polytopes studied by Bubeck and Eldan, and 3) a polynomial-time algorithm to compute the barycentric quantum entropy of a density matrix that was proposed as an alternative to von Neumann entropy by Band and Park in the 1970s: this corresponds to the case when Ω is the set of rank one projection matrices and μ corresponds to the Haar measure on the unit sphere. Our techniques generalize to the setting of rank k projections using the Harish-Chandra-Itzykson-Zuber formula, and are applicable even beyond, to adjoint orbits of compact Lie groups.

CCS CONCEPTS

• **Theory of computation** → *Convex optimization*; Quantum information theory; Semidefinite programming.

KEYWORDS

Entropy, Optimization, Quantum entropy

ACM Reference Format:
Jonathan Leake and Nisheeth K. Vishnoi. 2020. On the Computability of Continuous Maximum Entropy Distributions with Applications. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC '20), June 22–26, 2020, Chicago, IL, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3357713.3384302

*Due to space restrictions, some proofs and details are omitted from this version of the paper and available in the full version of this paper: http://arxiv.org/abs/2004.07403.

STOC '20, June 22–26, 2020, Chicago, IL, USA

ACM ISBN 978-1-4503-6979-4/20/06…$15.00
https://doi.org/10.1145/3357713.3384302

1 INTRODUCTION

Entropy maximizing distributions. Let Ω be a subset of $\mathbb{R}^d$ and let $\mathcal{K} = \text{hull}(\Omega)$ denote the convex hull of Ω. Suppose one is given an $A \in \mathcal{K}$. A natural question arises: *Is there a canonical way to choose a probability measure supported on Ω that can be used to express A as a convex combination of points on Ω?* When Ω is a discrete and finite set, this problem has been extensively studied and a canonical probability distribution was proposed by Jaynes [24, 25]: among all probability distributions that can be used to express A as a convex combination of points in Ω, pick the one that maximizes the Shannon entropy. These distributions are referred to as maximum entropy (max-entropy) distributions and arise in machine learning, statistics, mathematics, and theoretical computer science (TCS). In TCS, these distributions have found many uses due to duality, connections to polynomials, and algorithms to compute them [1, 2, 11, 13, 19, 34]; see [36].

In this paper we initiate a study of the computability when Ω is a continuous (and often nonconvex) manifold. Examples of interest include

$$\mathcal{V}_1 := \{vv^\top : v \in \mathbb{R}^n\},$$
$$\mathcal{P}_1 := \{vv^* : v \in \mathbb{C}^n, \|v\|_2 = 1\},$$

the set of rank k Hermitian projection matrices

$$\mathcal{P}_k := \{Y : Y \in \mathbb{C}^{n\times n}, \text{Tr}(Y) = k, Y = Y^*, Y^2 = Y\}$$

(related to the Grassmanian), or a convex body (where $\mathcal{K} = \Omega$).

Unlike the discrete setting, in the continuous setting the notion of a max-entropy distribution is not well-defined since a canonical notion of entropy does not necessarily exist. We instead consider relative entropy, Kullback-Leibler (KL) divergence, with respect to a prior measure μ on Ω that corresponds to the density function $f(X) \equiv 1$ for all $X \in \Omega$. For all of the manifolds mentioned above, there is a canonical measure that has this property and is called the uniform measure; see Section 2. This leads us to the following infinite dimensional convex optimization problem which gives a canonical way to write A as a convex combination of points in Ω: Find a measure ν on Ω that is continuous with respect to μ and, subject to the constraint that the expected point in $\mathcal{K}$ with respect to ν is A, ν minimizes the KL divergence to μ. Note that, by choice, ν is as close to the distribution μ as possible; hence we call it a maximum entropy distribution.

The class of extremal entropy maximizing distributions that arise in this manner have several properties that have led to their appearance, implicitly or explicitly, in several different areas:

- the work of Klartag (inspired by a work of Gromov) on the isotropic constant [15, 27],

- the work of Khatri and Mardia on the Matrix Bingham distribution in statistics with applications to various scientific and engineering problems [6, 21, 26],
- as shown here, the work of Goemans and Williamson on rounding semidefinite programs [14],
- the works of Güler, Bubeck and Eldan on barrier functions for interior point methods [7, 17, 18],
- the works of Band, Park, and Slater that defined the barycentric quantum entropy and proposed it as an alternative to the von Neumann entropy in the 1970s [3, 30, 35].

Computability of entropy maximizing distributions. One of the reasons why the entropy maximizing problem defined earlier is interesting (and unifies the above problems) is *duality*: the dual optimization problem roughly has the form:

$$\inf_{Y} \langle Y, A\rangle + \log \int_{X\in\Omega} e^{-\langle Y,X\rangle} d\mu(X),$$

where $\langle\cdot,\cdot\rangle$ is an inner product and μ is the given measure. If strong duality holds, it can be shown that the optimal distribution $\nu^\star$ to the entropy maximizing problem can be described by the optimizer $Y^\star$ to the dual above: $\nu^\star(X) \propto e^{-\langle Y^\star,X\rangle}$ for $X \in \Omega$. As for computability of $\nu^\star$, $Y^\star$ lives in a small, convex, and finite dimensional (same dimension as $\mathcal{K}$) domain. Hence in principle, one could hope to represent $\nu^\star$ efficiently. However, bounding the running time of a optimization method to find $Y^\star$ reduces to 1) a bounding some norm of $Y^\star$ and, 2) coming up with efficient algorithms to compute $\int_{X\in\Omega} e^{-\langle Y,X\rangle} d\mu(X)$ for matrices Y with that norm. These are the main technical problems studied in this paper.

1.1 Our Contributions

The main contributions of this paper are to initiate a study of the computability of entropy maximizing distributions on continuous domains, to present an ellipsoid method-based framework to compute them, to derive polynomial time algorithms for computing maximum entropy distributions for specific manifolds mentioned earlier, and to present implications to some of the applications listed above.

The continuous maximum entropy framework and duality. Our general framework is presented Section 3. The focus is on the setting when the manifold Ω and the base measure μ is fixed to either the set of all rank one matrices over reals ($\mathcal{V}_1$) with the measure induced by Lebesgue measure on $\mathbb{R}^n$, or the set of all rank k projections over complexes ($\mathcal{P}_k$) for $k \geq 1$ with the appropriate Haar measure. The input consists of an element A (which is a matrix in the cases of interest) and the goal is to compute a representation for $\nu^\star$ that is the KL-divergence minimizing distribution to μ with marginal A. We start by writing down the dual of this optimization problem (see Figure 1). We show (see the full version) that strong duality holds under Slater's condition – that there is a density function that is strictly positive (and bounded) on Ω and has marginal A. This is implied by the condition that A is in the relative interior of the convex hull $\mathcal{K}$ of Ω, which is true quite generally. Strong duality implies that the optimal measure $\nu^\star$ is determined by the optimal dual solution $Y^\star$ as $\nu^\star(X) \propto e^{-\langle Y^\star,X\rangle}$; see Theorem 4.1.

Norm of the optimal dual solution. However, to solve the dual convex program one needs, at the bare minimum, that the norm of $Y^\star$ is reasonably bounded. It is not difficult to see that as A tends to the boundary of $\mathcal{K}$, the optimal measure is concentrated on a face of $\mathcal{K}$ implying that the norm of $Y^\star$ must tend to infinity. Thus, one needs some assumption on the "interiority" of A to ensure polynomial time computability. The situation is exacerbated by the fact that the $Y^\star$ appears in the exponent and, hence, to have any hope of computability of the entropy maximizing distribution, the bound on $Y^\star$ should be *polynomial* in the bit complexity of A. Unlike the case when Ω is discrete (studied in [34]), the fact that the base measure μ is continuous makes it harder. Our main contribution towards the problem of bounding the norm of $Y^\star$ involves identifying a certain "balance" property of the measure μ on the manifold Ω (Definition 4.1) and showing that, roughly, $\|Y^\star\| \leq \mathrm{poly}(d, 1/\eta)$ where η is the distance of A from the boundary of $\mathcal{K}$; see Theorem 4.2. We show that this balance property holds for a wide class of manifolds and obtain as corollaries a bound of $\mathrm{poly}(n, 1/\eta)$ for both $\Omega = \mathcal{P}_k$ (Corollary 7.4) and when Ω is an n-dimensional convex body (Corollary 4.15). This bounding box result is quite general and expected to find further applications.

Computing the integral in the dual for matrix manifolds. A bound on the norm of $Y^\star$ allows us to show that we can use the ellipsoid method to solve the dual convex program, provided the measure μ is balanced on Ω, and we can evaluate the dual and its gradient at a specified Y of norm up to that of $Y^\star$. The tasks of evaluating the dual and its gradient essentially reduce to the computation of the integral $\int_{X\in\Omega} e^{-\langle Y,X\rangle} d\mu(X)$. In the case when $\Omega = \mathcal{V}_1$ with μ being the measure induced by the Lebesgue measure, we observe that the dual optimization problem is finite only when $Y \succ 0$, and thus we need to evaluate the integral only for such a Y. The integral above then turns out to have a simple formula: roughly, $\log\det Y$.

In other interesting cases, computing such an integral turns out to be a nontrivial task. In the case when $\Omega = \mathcal{P}_1$ and μ_1 is the uniform measure induced by the Haar measure on the complex unit sphere, the entropy maximizing measure cannot be obtained by solving the problem first for $\mathcal{V}_1$ and then "projecting" it on the sphere; see Section 6.2. Further, the integral does not reduce to a product of n integrals as in the Lebesgue case, and there is no easy way around this. We need an algorithm to integrate the density e^{-v^*Yv} over the complex unit sphere where the only thing we know about Y is that it is Hermitian. Neither the density is log-concave, nor the support (unit sphere) is convex. Our main contribution here is to give an *exact* algorithm to compute this integral whose running time depends *single exponentially* on the bit complexity of the input Y to it (Theorem 4.5). As remarked earlier, because Y is being exponentiated, this is the best one can hope for and also turns out to be sufficient to obtain polynomial time algorithms for computing maximum entropy distributions on $\mathcal{P}_1$.

Interestingly, the algorithm to compute this integral and its proof relies on an connection between the manifold $\mathcal{P}_1$ and the probability simplex in n dimensions. Specifically, one can naturally push forward the entropy maximizing measure from $\mathcal{P}_1$ to a log-linear measure on the corresponding simplex. There are then algorithms to sample from such a density function on the simplex to estimate such an integral; however, to obtain an $1+\delta$ approximation to

it, the running time of these methods depends polynomially on $1/\delta$ instead $\log 1/\delta$. We give an exact algorithm to compute this integral. Our method relies on Laplace transforms, is elementary, and a significant effort is needed to deal with the case when Y has repeated eigenvalues. Importantly, this viewpoint also leads us to an exact algorithm for computing such an integral for $\mathcal{P}_k$ for $k > 1$ using the Harish-Chandra-Itzykson-Zuber formula [12, 20, 23, 37]; see Theorem 6.1.

Efficient algorithm via the ellipsoid method. Our general ellipsoid method-based algorithm requires 1) a full dimensional embedding of hull(Ω) in a d-dimensional real Hilbert space, 2) μ is a balanced measure on Ω, 3) Ω is contained in a ball of radius r, 4) the point A is in the η-interior of hull(Ω) and, 5) that we have an exact counting/integrating oracle. It runs in time polynomial in $d, 1/\eta, \log r$ and $\log 1/\varepsilon$, to solve the dual problem to an additive ε; see Theorem 4.4. Our bound on the norm of $Y^\star$ and exact algorithms to compute the dual objective/gradient for the case of $\mathcal{P}_k$ imply a polynomial time algorithm to compute the entropy maximizing measure in this case when A is in the polynomial interior of hull($\mathcal{P}_k$); see Corollary 4.10.

1.2 Applications

SDP rounding. One approach to semi-definite programming (SDP) based approximation algorithms, starting with the work of Goemans-Williamson [14] for the maximum cut problem, is SDP rounding. Here, typically, A is a positive semi-definite (PSD) matrix, that is computed using a SDP relaxation to some non-convex problem, and one of the goals is to *round* A to a vector. This involves choosing a distribution on the set $\mathcal{V}_1$ defined above, and typical choices have been somewhat magical and lack an explanation. In the Goemans-Williamson setting, A is an $n \times n$ PSD matrix, and the density ν on Ω they choose to express A as a convex combination is as follows: pick a vector $v \in \mathbb{R}^n$ from the normal distribution with covariance matrix A. We show that this distribution is the maximum entropy distribution $\nu^\star$ (corresponding to A) on $\mathcal{V}_1$ with base measure induced by the Lebesgue measure on $\mathbb{R}^n$, thus giving an optimization characterization of this measure; see Corollary 4.14. The proof relies on strong duality and a closed form expression for the dual objective integral on $\mathcal{V}_1$; see Theorem 4.1.

Quantum entropy. In quantum mechanics, a density matrix ρ is a trace one complex $n \times n$ PSD matrix and describes the statistical state of a system. The extreme points in the set of density matrices are the pure states or $\mathcal{P}_1$. von Neumann defined a notion of entropy [38] of ρ that is computed by first writing ρ as a convex combination $\sum_{i=1}^n \lambda_i u_i u_i^*$, where $\{u_i\}_{i\in[n]}$ is an orthonormal basis for $\mathbb{C}^n$, and then computing the negative Shannon entropy of the λ_i's. While the von Neumann entropy is a mathematically elegant notion, it was vigorously argued in the 1970s that it does not capture the uncertainty in ρ [3, 30, 35]. In fact, von Neumann's way to write ρ as a convex combination of pure states can be viewed as "the most terse", or entropy minimizing one. In the same papers, an alternative way to define entropy of a density matrix was suggested – as the entropy of the entropy maximizing distribution with marginal ρ – and referred to as the barycentric quantum entropy. Unlike the von Neumann entropy, that has a simple formula ($-\operatorname{Tr}\rho\log\rho$)), the barycentric entropy did not have an efficient algorithm that could compute it. Our algorithm to compute entropy maximizing distributions for $\mathcal{P}_1$ mentioned above directly implies a polynomial time algorithm to compute the barycentric entropy of a density matrix (that is sufficiently in the interior) along with the probability density that achieves it; see Corollary 4.13.

Entropic barrier function. Bubeck and Eldan in [7] proved that the entropic barrier of a convex body $K \subseteq \mathbb{R}^d$ is a $(1 + o(1))n$-self-concordant barrier on K. Roughly speaking, this barrier function, for a point in K is defined to be the optimal value of a dual maximum entropy optimization problem when $\Omega = K$ and the measure is the Lebesgue measure on K. The computability of this barrier function for a point K is not known in general. One obstacle is to get a reasonable bound on the optimal solution. An almost direct consequence of Theorem 4.2 implies such a bound for points that are sufficiently in the interior of K; see Corollary 4.15.

2 PRELIMINARIES

Notation. Let $\mathbb{C}, \mathbb{R}, \mathbb{R}_+, \mathbb{N}$ denote the complex, real, nonnegative real, and natural numbers respectively. For $k, n \in \mathbb{N}$, let $\mathbb{C}^{k\times n}$ and $\mathbb{R}^{k\times n}$ denote the sets of $k\times n$ complex and real matrices respectively. A matrix $M \in \mathbb{C}^{n\times n}$ is said to be Hermitian if $A = A^*$ where $*$ denotes the conjugate transpose. A Hermitian matrix M is said to be PD (positive definite) and PSD (positive semidefinite) if its eigenvalues are positive and nonnegative respectively. For an $n \times n$ matrix X, we define $\operatorname{diag}(X)$ to be the length-n vector of the diagonal entries of X. If x is a vector, then we define $\operatorname{diag}(x)$ to be the diagonal matrix with entries the entries of x. For any $k, n \in \mathbb{N}$, we equip the vector space $\mathbb{C}^{k\times n}$ with the Frobenius inner product $\langle Y, Z\rangle := \operatorname{Tr}(YZ^*)$. We also denote $\|Y\| := \sqrt{\langle Y, Y\rangle}$. Note that $\langle Y, Z\rangle \in \mathbb{R}$ whenever Y, Z are Hermitian, so that the set of $n \times n$ Hermitian matrices is a real Hilbert space of dimension n^2. Also $\langle Y, Z\rangle \geq 0$ whenever Y, Z are PSD. We further let $B_\varepsilon(Y)$ denote the open ε-ball centered at Y in the space in which Y lives (e.g., the $n\times n$ Hermitian matrices). Finally, we let hull(S) denote the convex hull of a set S in some ambient vector space.

Manifolds. In general, we let Ω be any smooth manifold that is embedded in a d-dimensional real Hilbert space V with inner product $\langle\cdot,\cdot\rangle$. Let $\mathcal{L}(X) = B$ denote the affine space in which hull(Ω) is full dimensional, i.e., every element $X \in$ hull(Ω) satisfies the equation $\mathcal{L}(X) = B$. The concrete manifolds we consider are collections of matrices with some structure. In particular, for fixed $n \in \mathbb{N}$, consider the following manifold within $\mathbb{C}^{n\times n}$. An $n \times n$ *rank-k PSD projection* is a PSD matrix with k eigenvalues equal to 1 and the rest equal to 0.

$$\mathcal{P}_k = \mathcal{P}_k(n) := \{n \times n \text{ rank-}k \text{ PSD projections}\}.$$

Note that $\mathcal{P}_k$ is also a manifold within the space of $n\times n$ Hermitian matrices.[1] Other manifolds we consider are the complex unit sphere $S^n_{\mathbb{C}} \subset \mathbb{C}^n$ (which is related to $\mathcal{P}_1$), the manifold of all rank one matrices (not necessarily trace one): $\mathcal{V}_1 := \{vv^\top : v \in \mathbb{R}^n\}$, and a convex body $K \subset \mathbb{R}^n$.

[1] Note that $\mathcal{P}_k$ is homeomorphic to a Grassmannian, i.e., the manifold of k-dimensional subspaces within an n-dimensional space. The homeomorphism is explicitly given as the map which sends a rank-k PSD projection to the k-dimensional subspace given by its image.

We would also like to consider the convex hull of a given manifold Ω. To make sense of such a notion, we need to consider the manifold as being embedded in some ambient vector space. This ambient space often the space of $n \times n$ Hermitian matrices in our examples. In general, we refer to the elements of hull(Ω) as *marginals* or *marginals matrices*.

Group actions. It is useful to understand the symmetries of some of the manifolds mentioned above in terms of groups that act on them. Recall that an $n \times n$ unitary matrix is an invertible matrix U for which $U^{-1} = U^*$, and an $n \times n$ orthogonal matrix is an invertible matrix O for which $O^{-1} = O^\top$. The unitary and orthogonal groups ($U(n)$ and $O(n)$) act on the manifolds discussed above as follows:

- $U(n)$ acts on column vectors in $S^n_{\mathbb{C}}$ and on hull($S^n_{\mathbb{C}}$) by left multiplication.
- $U(n)$ acts on $\mathcal{P}_k$ and on hull($\mathcal{P}_k$) by conjugation.
- $O(n)$ acts on $\mathcal{V}_1$ and on hull($\mathcal{V}_1$) by conjugation.

Note that the actions of $U(n)$ on $S^n_{\mathbb{C}}$ and on $\mathcal{P}_1$ are compatible in the sense that for $x \in S^n_{\mathbb{C}}$ and $U \in U(n)$, we have $(Ux)(Ux)^* = U(xx^*)U^*$ where $xx^* \in \mathcal{P}_1$.

Relative interior. The convex set hull(Ω) is not necessarily full dimensional in the ambient Hilbert space. To define a notion of interior for hull(Ω), we restrict to the minimal affine subspace in which Ω lives (this is given explicitly by $\mathcal{L}(X) = B$ discussed above). More generally, we make the following definition.

Definition 2.1 (Relative interior). *Fix a convex subset S in a vector space V, and let $V_{\mathcal{L},B}$ be the minimal affine subspace in which S lives. We say that $Y \in V$ is in the η-*interior *of S (for $\eta > 0$) if*

$$B_\eta(Y) \cap V_{\mathcal{L},B} \subseteq S.$$

We say that Y is in the interior *of S if there exists $\eta > 0$ such that Y is in the η-interior of S.*

Here we usually consider $S = \mathcal{P}_k(n)$, and we will be interested in the case where $\eta \geq \frac{1}{\text{poly}(n)}$.

Measures and densities. Often, the manifolds Ω we consider have some geometric structure (e.g., it is a manifold with a group action), and we want to consider measures which interact nicely with this structure. To make sure this happens, we restrict to the class of measures which are given by continuous density functions on Ω. To make sense of this, we need a natural base measure μ on Ω which corresponds to the density function $f(X) \equiv 1$. (E.g., in the case of $\Omega = \mathbb{C}^n$ or $\Omega = \mathbb{R}^n$, the Lebesgue measure often plays this role.) In particular, the support of μ should be equal to Ω.

In the case of $\Omega = \mathcal{P}_k$, there is a canonical measure which is appropriately called the uniform measure: we define μ_k be the unique unitarily invariant measure on $\mathcal{P}_k$, where $U(n)$ acts by conjugation (as discussed above). Hence, equivalently (and more formally), we restrict to the class of measures on $\mathcal{P}_k$ which are absolutely continuous with respect to μ_k. We also consider the standard Lebesgue measure on $\mathbb{R}^n$ for convex bodies and its pushforward measure μ through the map $v \mapsto vv^\top$ on $\mathcal{V}_1$. Note that $S^n_{\mathbb{C}}$ also has a canonical unitarily invariant measure, usually called the Haar measure. The pushforward of this measure through the map $v \mapsto vv^*$ yields the unitarily invariant measure μ_1 on $\mathcal{P}_1$.

Integration/Counting oracle. We are interested in computing the following exponential integral for a given Y in our Hilbert space V.

Definition 2.2 (Exponential integrals). *Fix $n \in \mathbb{N}$ and let μ be a measure with support Ω, a manifold embedded in the real Hilbert space V. We define the following function on an input $Y \in V$:*

$$\mathcal{E}(Y) = \mathcal{E}_\mu(Y) := \log \int_\Omega e^{-\langle Y, X\rangle} d\mu(X).$$

Whenever $\mu = \mu_k$ and $\Omega = \mathcal{P}_k$, we use the following shorthand notation $\mathcal{E}_k(Y)$. We sometimes also refer to these integrals as exponential integrals.

A *strong integration/counting oracle* for Ω and μ outputs two quantities, given an element Y from the ambient Hilbert space V of Ω:

(1) $\mathcal{E}_\mu(Y)$
(2) the matrix $\nabla\mathcal{E}_\mu(Y)$, defined so that the following holds for any $Z \in V$:

$$\langle \nabla\mathcal{E}_\mu(Y), Z\rangle = \frac{d}{dt}\mathcal{E}_\mu(Y + tZ)\Big|_{t=0}.$$

In the case of $\Omega = \mathcal{P}_k$, Y and Z are Hermitian. Further, since the measure μ_k is unitarily invariant, we can assume that Y is diagonal and expect the running time of the counting oracle should depend polynomially on n and the number of bits needed to represent e^{-y_i} for any i, where $y_1, \ldots, y_n$ are the eigenvalues (diagonal elements) of Y.

As we will show, in the special case when $\Omega = \mathcal{V}_1$ and μ is the pushforward of the Lebesgue measure, we can compute the integral $\mathcal{E}_\mu(Y)$ exactly in time polynomial in the bit complexity of Y due to a direct formula. This happens because the measure μ is a product measure, which is not the case for μ_k.

3 THE MAXIMUM ENTROPY FRAMEWORK

In this section we present our maximum entropy convex program. Fix a manifold Ω in a d-dimensional real Hilbert space with inner product $\langle \cdot, \cdot \rangle$, and let $\mathcal{L}(X) = B$ denote the corresponding affine space containing Ω. Let μ be the base measure on Ω and A in $\mathcal{K} := \text{hull}(\Omega)$. Our goal is to find a density function ν with marginal A that minimizes the KL-divergence with respect to μ.

We use the shorthand $\text{Prim}_\mu(A)$ (or $\text{Prim}_k(A)$ if $\mu = \mu_k$) to refer to this primal optimization program. We mainly consider the case of $\mu = \mu_k$ and $\Omega = \mathcal{P}_k$ or $\Omega = \mathcal{V}_1$ with μ the pushforward of Lebesgue measure. In these cases Y will comes from some subspace of the $n \times n$ Hermitian matrices. Drawing from the intuition that these base measures are uniform over the manifold, and hence in some sense maximize entropy, we say the KL-divergence minimizing measure is entropy maximizing. However, we note that this framework is also applicable to other base measures, in particular to the case when Ω is a convex body in $\mathbb{R}^d$ and μ is the Lebesgue measure. The fact that the entropy integral (without the minus sign) is convex as a function of the density ν follows from the fact that this integral is precisely the KL divergence between the probability distribution corresponding to ν and the distribution μ. Convexity of the KL divergence for probability distributions is then a well-known fact.

Efficiently solving this convex program directly is a priori impossible as the support of ν is infinite. To find a succinct representation

Primal

$$\sup_{\nu}\left[-\int_{\Omega}\nu(X)\log\left(\nu(X)\right)d\mu(X)\right]$$

subject to:

$$\nu:\Omega\to\mathbb{R}_{\geq 0},\ \mu\text{-measurable}$$

$$\int_{\Omega}X\nu(X)d\mu(X)=A$$

$$\int_{\Omega}\nu(X)d\mu(X)=1$$

Dual

$$\inf_{Y}F_A(Y)=\inf_{Y}\left[\langle Y,A\rangle+\log\int_{\Omega}e^{-\langle Y,X\rangle}d\mu(X)\right]$$

subject to:

$$\mathcal{L}(Y)=0$$

Figure 1: Primal and dual maximum entropy convex programs for A in the interior of hull(Ω).

for the optimal $\nu^\star$, we turn to the dual program (see Figure 1), which gives us a nice representation of the max-entropy density function $\nu^\star$. We often use the shorthand $\text{Dual}_\mu(A)$ (or $\text{Dual}_k(A)$ if $\mu=\mu_k$) to refer to this program.

In the case of $\mathcal{P}_k$ with uniform measure μ_k, the optimal solution to $\text{Dual}_k(A)$ is given by a Hermitian matrix $Y^\star$. By strong duality (see Theorem 4.1), this in turn shows that the max-entropy density function $\nu^\star$ takes on a nice form:

$$\nu^\star(X)\propto e^{-\langle Y^\star,X\rangle}.$$

As a note, in the case of $\Omega=\mathcal{P}_k$ this matrix $Y^\star$ is only unique up to a shift by a multiple of the identity matrix. Issues arising from non-uniqueness can be handled by restricting to the minimal affine subspace in which hull($\mathcal{P}_k$) lives, as referred to in the discussion surrounding Definition 2.1. However, as A tends to the boundary of hull(Ω), $Y^\star$ can be seen to tend to infinity as the support of the measure $\nu^\star$ tends to lower dimensions.

4 FORMAL STATEMENT OF OUR RESULTS

4.1 Mathematical and Computational Results

Our first result shows that strong duality holds.

Theorem 4.1 (Strong duality). *Let Ω be a manifold that is embedded in a d-dimensional real Hilbert space with an inner product $\langle\cdot,\cdot\rangle$, and let μ be a measure supported on Ω. For any A in the relative interior of the convex hull of Ω, the optimal values of the primal and dual objective functions coincide, and the corresponding max-entropy distribution has density function of the form $\nu^\star(X)\propto e^{-\langle Y^\star,X\rangle}$ for some $Y^\star$.*

The proof of this result uses standard techniques and thus is omitted from this paper (see the full version). This result applied to $\mathcal{P}_k$ and μ_k shows that optimizing $\text{Dual}_k(A)$ is in fact equivalent to optimizing $\text{Prim}_k(A)$, and therefore the max-entropy measure has the exponential form described above.

With strong duality in hand, we focus on the computability of the optimal matrix $Y^\star$ for the dual program. To do this we use a version of the ellipsoid algorithm (see Theorem 8.1), for which we need two things.

First, we need an upper bound on some norm of the dual optimal solution. If $Y^\star$ is the optimal solution, then the number of iterations of the ellipsoid algorithm depends on $\log\|Y^\star\|$. That said, it may seem that a bound depending on $e^{1/\eta}$, where η is such that $B_\eta(A)\subset$ hull(Ω), is enough to achieve polynomial dependence on $\frac{1}{\eta}$. However, this is not enough, since the integral appearing in the dual is polynomially dependent on the number of bits needs to represent e^{-y_i}, where the y_i's are the entries or eigenvalues of a given input Y. Hence, we actually need polynomial dependence on $\frac{1}{\eta}$, which is achieved in our bounding box result below. Note that this issue is not surprising, as it crops up in exactly the same way in the discrete maximum entropy case (see [34]).

We give here a bounding box result which is more general than we need for the rank-k projections case ($\Omega=\mathcal{P}_k$ and $\mu=\mu_k$). It relies on a key "balance" property of the measures. This notion extends important properties of the discrete uniform measure to continuous measures on manifolds and is one of the key notions we introduce.

Definition 4.1 (Balanced measure). *A measure μ is said to be* balanced *if for any $\delta>0$ and $X\in\Omega\subseteq\mathbb{R}^d$, we have that at least $\exp(-\text{poly}(\delta^{-1},d))$ of the mass of μ is contained in the δ-ball about X.*

We see in Definition 7.2 how this notion can be used to give a more refined notion of interior (beyond the η parameter discussed above). Conceptually, it allows us to give an measure-theoretic relaxation of the notion of a separating hyperplane.

We now state the main bounding box result. The precise bound is given in Theorem 7.2.

Theorem 4.2 (Bounding box). *Let μ be a measure supported on a manifold Ω embedded in a d-dimensional real Hilbert space. Suppose that μ is balanced, in the sense of Definition 4.1. Further, let A be an element of the η-interior of the convex hull of Ω. Then there is an optimal solution $Y^\star$ to the dual program such that: $\|Y^\star\|\leq\text{poly}(\eta^{-1},d)$.*

Corollary 7.4 and Corollary 4.15 give bounds for rank-k projections and convex bodies as corollaries.

Remark 4.3. *Our bounding box result significantly generalizes the discrete case (Theorem 2.7 in [34]). Uniform distribution in the discrete case has atoms of uniformly strictly positive (at worst singly-exponentially small) mass at all points, and this implies a bound on optimal dual solutions. In the continuous case this is no longer true, the notion of balance then fills the gap.*

Second, at each step of the ellipsoid algorithm, we need to be able to evaluate the dual objective function and its gradient at given input

Y. The hardest part of such a computation comes in evaluating $\mathcal{E}_\mu$, the exponential integral portion of the objective function. We show that if we have access to such an evaluation oracle, then under very general conditions, we can compute the maximum entropy distribution.

Theorem 4.4 (Ellipsoid method-based general algorithm). *Let μ be a balanced measure with support on a manifold Ω embedded in a d-dimensional real Hilbert space. Let the affine space in which Ω lies, $\mathcal{L}(X) = B$, be given as input $(\mathcal{L}, B)$. Assume that Ω is contained in a ball of radius r. There exists an algorithm that, given A in the η-interior of* $\text{hull}(\Omega)$*, any $\varepsilon > 0$, and a strong counting/integration oracle for the exponential integral $\mathcal{E}_\mu(Y)$, returns Y° such that*

$$F_A(Y^\circ) \leq F_A(Y^\star) + \varepsilon$$

where F_A is the objective function for the dual program $\text{Dual}_\mu(A)$*, and $Y^\star$ is an optimum of the dual program. The running time of the algorithm is polynomial in d, η^{-1},* $\log(\varepsilon^{-1})$, $\log(r)$*, and the number of bits needed to represent A, $\mathcal{L}$, and B.*

Our next result says that in fact we have an efficient strong counting oracle for $\mathcal{E}_k$ on the domain $\mathcal{P}_k$ with measure μ_k.

Theorem 4.5 (Counting oracle for $\mathcal{P}_k$). *There is an algorithm that, given $n \in \mathbb{N}$, $k \in [n]$, an $n \times n$ real diagonal matrix $Y = \text{diag}(y)$, and a $\delta > 0$, returns numbers $\bar{E}, \bar{G}$ such that*

(1) $|\bar{E} - \mathcal{E}_k(Y)| \leq \delta$
(2) $|\bar{G} - \nabla\mathcal{E}_k(Y)| \leq \delta$,

where $\mathcal{E}_k$ is the exponential integral defined above (and in Definition 2.2). The running time of the algorithm is polynomial in n, $\log(\frac{1}{\delta})$*, and the number of bits needed to represent e^{-y_i} for any $i \in [n]$.*

The proof of this theorem for $k = 1$ is elementary but relies on the interesting connection between the complex unit sphere and the probability simplex. This connection also yields an exact sampling algorithm.

Proposition 4.6 (Rank-one Sampling). *Let $Y = \text{diag}(y)$ be a real diagonal $n \times n$ matrix. The following process produces samples from the measure $e^{-\langle Y, X\rangle} d\mu_1(X)$ on $\mathcal{P}_1$.*

(1) Sample v from the measure $e^{-\langle y, v\rangle} d\mu_{\Delta_1}(v)$ on the simplex Δ_1 by iteratively sampling v_i conditioned on $v_1, \ldots, v_{i-1}$.
(2) Sample $z_1, \ldots, z_n$ independently uniformly from the complex unit circle.
(3) Construct $X := (z\sqrt{v})(z\sqrt{v})^* \in \mathcal{P}_1$ where $z\sqrt{v}$ is the column vector $(z_1\sqrt{v_1}, \ldots, z_n\sqrt{v_n})$.

For $k > 1$, the proof of Theorem 4.5 above relies on the Harish-Chandra-Itzykson-Zuber formula; see Theorem 6.1.

Remark 4.7. *In the case of $\mathcal{V}_1$ with the pushforward of Lebesgue measure, there is an exact formula to compute the corresponding dual optimum for positive definite marginals A: $Y^\star = \frac{1}{2}A^{-1}$; see Corollary 9.4. Positive-definiteness of the input Y is in fact required for the dual objective to be finite, which is in stark contrast with the $\mathcal{P}_k$ case where any Hermitian matrix is allowed. These points suggest a conceptual divide between the Lebesgue measure case and the rank-k projections case. We do not expect such a formula for $Y^\star$ in the case of $\mathcal{P}_1$ and, indeed, the lack of one has been one of the obstacles for efficient algorithms for quantum barycentric entropy and computing the normalizing constant of the matrix Bingham distribution.*

Remark 4.8. *In this paper we primarily consider the best possible setting where the running time of the counting oracle depends logarithmically on the accuracy. We refer to such counting oracles as* exact. *We note that our framework does allow for counting oracles where the dependence is polynomially in $1/\delta$.*

Remark 4.9. *Güler in [16] studies the characteristic function of a convex cone. In our language, the characteristic function of a cone is the exponential integral $\mathcal{E}_K(y)$ with respect to the Lebesgue measure on the dual cone K:*

$$\mathcal{E}_K(y) = \log \int_K e^{-\langle y, x\rangle} dx.$$

For the case of homogeneous convex cones, Güler gives a nice way to construct explicit formulas for the characteristic function. (A homogeneous cone is a cone K such that for all $u, v \in K$ theres is a linear isomorphism of K which maps u to v.) Given a fixed vector e in the interior of K, any other vector y in the interior of K, and an automorphism A_y of K mapping e to y, the dual objective for K can be written up to additive constant as:

$$F_\theta(y) = \langle y, \theta\rangle - \log \int_K e^{-\langle y, x\rangle} dx = \langle y, \theta\rangle - \frac{1}{2}\log(\det(A_y A_y^\top)).$$

Such an explicit formula gives a route to efficiently computing the dual objective function in the homogeneous case. Orthants, Lorentz cones, and semidefinite cones are all homogeneous, and in those specific cases the above formula for $F_\theta(y)$ becomes completely explicit. See Sections 3 and 7 of [16] for more details.

The bounding box and counting oracle for μ_k and $\mathcal{P}_k$ then imply that the ellipsoid method-based algorithm from Theorem 4.4 gives a polynomial time algorithm for approximately computing $Y^\star$, the optimum of the program $\text{Dual}_k(A)$.

Corollary 4.10 (Ellipsoid method-based efficient algorithm for $\mathcal{P}_k$). *There exists an algorithm that, given $n \in \mathbb{N}$, $k \in [n]$, a trace-k PD matrix A in the η-interior of the convex hull of the set of $n \times n$ rank-k PSD projection matrices (i.e.,* $\text{hull}(\mathcal{P}_k)$*), and an $\varepsilon > 0$, returns a Hermitian matrix Y° such that*

$$F_A(Y^\circ) \leq F_A(Y^\star) + \varepsilon,$$

where F_A is the dual objective function and $Y^\star$ is an optimal solution to the dual program $\text{Dual}_k(A)$*. The running time of the algorithm is polynomial in n,* $\frac{1}{\eta}, \log(\frac{1}{\varepsilon})$*, and the number of bits needed to represent A.*

Remark 4.11. *Notice that the dependence on $\frac{1}{\eta}$ means that we do not achieve a polynomial time algorithm for A near the boundary of* $\text{hull}(\mathcal{P}_k)$*. This dependence comes from the fact that the bounding box (Theorem 4.2) is dependent on $\frac{1}{\eta}$. One may then naturally ask whether this bounding box dependence can be improved, but it turns out that it cannot in this case. Note that this differs from the discrete case, where in [36] the authors are able to remove this $\frac{1}{\eta}$ dependence under certain assumptions on the polytope.*

Finally, the closeness of the distributions associated to Y° and $Y^\star$ is then given in the following.

Proposition 4.12. *Let $Y^\star$ be the optimal to the dual objective function $F_A(Y)$, and let Y° be such that $F_A(Y^\circ) \leq F_A(Y^\star) + \varepsilon$. If*

$\mu^\star(X) \propto e^{-\langle Y^\star, X\rangle}$ and $\mu^\circ(X) \propto e^{-\langle Y^\circ, X\rangle}$ are the probability distributions associated to $Y^\star$ and Y° respectively, then

$$\|\mu^\star - \mu^\circ\|_{\mathrm{TV}} \leq \sqrt{2\varepsilon}.$$

4.2 Applications

Barycentric quantum entropy. In [35], Slater discusses the notion of barycentric quantum entropy of a density matrix, and compares it to that of von Neumann entropy. His investigation of this notion was prompted by the work of Band and Park [3, 30], who critiqued the use of von Neumann entropy as a good indicator of the uncertainty of the given density matrix. In particular, they argue that a better notion of entropy would relate to distributions on all possible pure states, whereas the von Neumann entropy is derived from the discrete distribution on the pure states corresponding to eigenvectors of the matrix. In response to this, Slater defines a notion of quantum entropy in terms of a max-entropy program on the set of all pure states. He then goes on to show how one might determine the quantum entropy in a few specific cases.

Definition 4.2 (Barycentric quantum entropy). *Let ρ be an $n \times n$ Hermitian density matrix (trace-1, positive semidefinite). Then the* barycentric quantum entropy *of ρ is defined (in our notation) as*

$$H_b(\rho) := \inf_{\nu} \int_{\mathcal{P}_1} \nu(X) \log(\nu(X)) d\mu_1(X)$$

subject to:

$$\nu(X) \geq 0 \quad \forall X \in \mathcal{P}_1$$

$$\int_{\mathcal{P}_1} X \nu(X) d\mu_1(X) = \rho$$

$$\int_{\mathcal{P}_1} \nu(X) d\mu_1(X) = 1$$

where $\mathcal{P}_1$ denotes the set of pure states and μ_1 denotes the unitarily invariant measure on $\mathcal{P}_1$.

Our results for computing max-entropy measures on $\mathcal{P}_1$ immediately imply efficient computability of the barycentric quantum entropy for density matrices that are polynomially in the interior.

Corollary 4.13 (Computability of barycentric quantum entropy). *There exists an algorithm that, given a Hermitian density matrix ρ in the η-interior of the set of Hermitian density matrices and an $\varepsilon > 0$, returns a number $\bar{H}$ such that $|\bar{H} - H_b(\rho)| < \varepsilon$. The running time of the algorithm is polynomial in n, $\frac{1}{\eta}$, $\log(\frac{1}{\varepsilon})$, and the number of bits needed to represent ρ.*

Goemans-Williamson SDP rounding. In their seminal paper, Goemans-Williamson [14] gave a rounding scheme that gives a way to round a given PD matrix A to a vector. Their method goes by drawing a vector v from a particular distribution on $\mathbb{R}^n$ based on the matrix A.

Definition 4.3 (Goemans-Williamson measure). *Given $n \in \mathbb{N}$ and a real positive definite $n \times n$ matrix A, the* Goemans-Williamson measure *μ_{GW} can be defined via a sampling process on $\mathbb{R}^n$ as follows.*

1. *Sample $g \in \mathbb{R}^n$ from the standard multivariate Gaussian distribution.*
2. *Compute $v := Vg$ where V is a square root of A, i.e., $A = VV^\top$.*
3. *v is a sample from μ_{GW}.*

It is then straightforward to compute the marginals matrix associated to this distribution as follows:

$$\mathbb{E}[vv^\top] = \int_{\mathbb{R}^n} (vv^\top) d\mu_{\mathrm{GW}}(v) = V \left[\int_{\mathbb{R}^n} gg^\top dg \right] V^\top = VV^\top = A.$$

Thus, if we map $\mathbb{R}^n$ to $\mathcal{V}_1$ via $v \mapsto vv^\top$ and also pushforward the Lebesgue measure through this map, the above is precisely the marginal constraint in our max-entropy framework. This observation implies that the pushforward of the measure μ_{GW} is a (strictly) feasible solution to the max-entropy primal program on the domain $\mathcal{V}_1$ with the pushforward of the Lebesgue measure. We show that it is also the optimal solution to the max-entropy program.

Corollary 4.14 (Goemans-Williamson measure maximizes entropy). *For any positive definite matrix A, let μ_{GW} be the measure corresponding to the Goemans-Williamson rounding scheme for A. Then the pushforward of μ_{GW} to $\mathcal{V}_1$ is the max-entropy measure with marginals A on $\mathcal{V}_1$ with respect to the pushforward of Lebesgue measure.*

Entropic barrier function. Bubeck and Eldan in [7] prove that the entropic barrier of a convex body $K \subseteq \mathbb{R}^d$ is a $(1 + o(1))n$-self-concordant barrier on K, improving a seminal result of Nesterov and Nemirovski [29]. In fact this gives the first explicit construction of a universal barrier for convex bodies with optimal self-concordance parameter.

Definition 4.4 (Entropic barrier). *Given a convex body $K \subseteq \mathbb{R}^d$, define the* entropic barrier *for K as the real-valued function on the interior of K defined as:*

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

Note that $-B_K(v)$ is precisely the maxium entropy dual program, up to negation of y in the expression.

Open questions still remain about the efficient computability of the entropic barrier. This is in particular true in the case where K is a polytope, given as a membership oracle. Towards this, the following is essentially a corollary to Theorem 4.2 (see Section 7.3 for a full proof), and can be used to efficiently compute the entropic barrier at points which are in the η-interior of K.

Corollary 4.15 (Bounding box for convex bodies). *Let μ be the uniform distribution on a d-dimensional convex body Ω contained in a ball of radius R. Given v in the η-interior of the convex hull of Ω, there is an optimal solution $y^\star$ to the dual program such that $\|y^\star\| \leq \mathrm{poly}(\eta^{-1}, d, \log(R))$.*

Details of how this implies computability of the entropic barrier are omitted from this paper.

5 ORGANIZATION OF THE PAPER

Section 6 contains a detailed technical overview of the proofs of the main results. Section 7 contains a complete proof of Theorem 4.2 (bounding box). Section 8 contains the proof of Corollary 4.10 (main ellipsoid-based algorithm). Section 9 contains the proof of Corollary 4.14 (Goemans-Williamson measure maximizes entropy). Section 10 disucusses the generalization of the maximum entropy framework to Lie groups. The rest of the proofs are omitted here

due to space restrcitions and can be found in the full version of the paper.

6 TECHNICAL OVERVIEW

In this section, we give overviews of the proofs of the main results of this paper and compare our techniques with those of previous work. We start by describing the approach of [34] in the case of discrete uniform measures μ with finite support $\Omega \subseteq \{0,1\}^d$. In this case, the marginals vector A of a measure ν on Ω is defined by setting A_k to be the expected value of the kth entry of x when x is chosen according to ν. Note that the marginals vector A is always an element of hull(Ω). The problem the authors of [34] solve is described as follows: given a finite subset Ω and a desired marginals vector A in the η-interior of hull(Ω), compute the probability measure on Ω with marginals A which maximizes entropy.

They consider the dual formulation

$$\inf_{y \in \mathbb{R}^d} \quad F_A(y) := \langle y, A \rangle + \log \sum_{x \in \Omega} e^{-\langle x, y \rangle},$$

which gives rise to measures on Ω of the following succinct form for some real vector $y^\star$:

$$\nu(x) \propto e^{-\langle x, y^\star \rangle}.$$

By strong duality $\nu = \nu^\star$ is the entropy maximizing measure, and they then use the ellipsoid method to approximate $y^\star$.

We generalize their approach to continuous measures μ on continuous domains Ω. For the most part, the ellipsoid algorithm can be applied in the same way as in the discrete case once we have the three main pieces in hand: (1) strong duality, (2) a bound on $Y^\star$, and (3) the strong counting oracle. Even in the continuous case, one can show that strong duality holds via a certain Slater-type condition. What makes the passage from the discrete case to the continuous case much more interesting and nontrivial is proving the remaining two main pieces.

6.1 Proof Overview: Bounding Box

The goal of this section is to explain the proofs of the main bounding box result and its corollaries. We first describe the approach of the discrete μ case discussed above. Note that for $B \in$ hull(Ω), there exists some $X_0 \in \Omega$ such that

$$\langle -Y^\star, X_0 - B \rangle \geq 0,$$

since every closed half-space containing B contains an extreme point $X_0 \in$ hull(Ω). If A is in the η-interior of hull(Ω), we can choose $B = A - \eta \frac{Y^\star}{\|Y^\star\|}$ to get:

$$\langle -Y^\star, X_0 - A \rangle \geq \eta \|Y^\star\|.$$

Because μ is a discrete uniform measure, we have $\mu(\{X_0\}) = |\Omega|^{-1}$. This implies a bound on $Y^\star$ as follows, via the dual objective function $F_A(Y)$:

$$\begin{aligned} 0 = F_A(0) &\geq F_A(Y^\star) \\ &= \log \int e^{\langle -Y^\star, X - A \rangle} d\mu(X) \\ &\geq \log \left(e^{\eta \|Y^\star\|} \cdot |\Omega|^{-1} \right) \\ \implies \|Y^\star\| &\leq \frac{\log |\Omega|}{\eta}. \end{aligned}$$

The lower bound on $F_A(Y^\star)$ above follows from restricting the integral (which is a sum in the discrete case) to the single point X_0. This demonstrates exactly why this argument fails in the continuous case, because in that case we have $\mu(\{X\}) = 0$ for all $X \in \Omega$.

This is the first difficulty we must overcome. We need a way to restrict the dual objective integral to a region of Ω which has positive mass, emulating the role of atoms in the discrete case.

We introduce a two-parameter interior for the measure μ. We say that A is in the (η, δ)-interior of μ if every half-space intersecting the η-ball about A contains at least δ mass of μ (Definition 7.2). Instead of restricting the dual integral to a single point of Ω, we restrict it to the appropriate δ-mass to obtain a bound on $\|Y^\star\|$:

$$\begin{aligned} 0 &\geq \log \int e^{\langle -Y^\star, X - A \rangle} d\mu(X) \geq \log \left(e^{\eta \|Y^\star\|} \cdot \delta \right) \\ &\implies \|Y^\star\| \leq \frac{1}{\eta} \log \frac{1}{\delta}. \end{aligned}$$

We explain this formally in Lemma 7.1.

This leads to the second difficulty. Our bounding box theorem only refers to the η parameter, and so we need a way to handle or control δ in terms of η and d.

Here is where the key *balance* property comes into play. We say that a measure μ is balanced if for all $\varepsilon > 0$ and $X \in \Omega$, the ε-ball about X contains $\exp(-\text{poly}(\varepsilon^{-1}, d))$ of the mass of μ (Definition 4.1). This links the two interiority parameters: from any point of the ε-interior of hull(Ω), there will be at least $\exp(-\text{poly}(\varepsilon^{-1}, d))$ mass in the direction of any $X \in \Omega$ on the boundary.

The crucial feature of the balance property is then how this linking of the parameters allows one to transfer between them. Specifically for a balanced measure, the η-interior of hull(Ω) is contained in the $(\frac{\eta}{2}, \exp(-\text{poly}(\frac{\eta}{2}, d)))$-interior of μ. To see this, let A be in the η-interior of hull(Ω). Hence, any half space which intersects the $\frac{\eta}{2}$-ball about A contains another $\frac{\eta}{2}$-ball in hull(Ω). By translating this ball toward a point of Ω, we can assume that the half-space contains an $\frac{\eta}{2}$-ball about a point of Ω. Since μ is balanced, this implies A is in the $(\frac{\eta}{2}, \exp(-\text{poly}(\frac{\eta}{2}, d)))$-interior of μ.

At this point, the rest of the proof of Theorem 4.2 is straightforward. For balanced μ and A in the η-interior of hull(Ω), we actually have that A is in the $(\frac{\eta}{2}, \exp(-\text{poly}(\frac{\eta}{2}, d)))$-interior of μ. The two parameter bound described above then implies $\|Y^\star\| \leq \text{poly}(\frac{1}{\eta}, d)$.

To obtain bounding boxes for μ_k on $\mathcal{P}_k$, $n \times n$ rank k projections, (Corollary 7.4) and to uniform measures on convex bodies (Corollary 4.15), we then demonstrate balance properties. In the case of μ_k, $\mathcal{P}_k \subset B_{\sqrt{k}}(0)$ can be covered by at most $\exp(\text{poly}(\log \delta^{-1}, n))$ balls of radius δ for any $\delta > 0$, morally because:

$$\frac{\text{vol}(B_{\sqrt{k}})}{\text{vol}(B_\delta)} = \frac{(\pi\sqrt{k})^n / n!}{(\pi\delta)^n / n!} = \left(\frac{\sqrt{k}}{\delta} \right)^n = \exp(\text{poly}(\log \delta^{-1}, n)).$$

Therefore a δ-ball about *some* point of $\mathcal{P}_k$ must contain at least $\exp(-\text{poly}(\log \delta^{-1}, n))$ of the mass of μ_k, and unitary invariance then implies that this is actually true for *all* points of $\mathcal{P}_k$.

For uniform measures μ on convex bodies K contained in a ball of radius R, we prove the bounding box using similar arguments

as follows. By the volume ratio computation above, every δ-ball contained in K contains at least $(\frac{\delta}{R})^d$ of the mass of μ. Therefore every A in the η-interior of hull(Ω) is also in the $(\frac{\eta}{2}, (\frac{\eta}{2R})^d)$-interior of μ, since every half-space intersecting the $\frac{\eta}{2}$-ball about A contains another $\frac{\eta}{2}$-ball in K. The bounding box then follows from the two-parameter bound discussed above (Lemma 7.1).

6.2 Proof Overview: Counting Oracle for $\mathcal{P}_1$ and $\mathcal{V}_1$

The goal of this section is to explain why we can efficiently evaluate and compute the gradient of

$$\mathcal{E}_\mu(Y) = \log \int_\Omega e^{-\langle Y,X\rangle} d\mu(X)$$

in the case of $\Omega = \mathcal{P}_1$ and $\Omega = \mathcal{V}_1$.

First consider the case of $\Omega = \mathcal{V}_1$, where μ is the pushforward of the Lebesgue measure through $x \mapsto xx^\top$. In this case we have a very explicit formula whenever Y is positive definite:

$$\mathcal{E}_\mu(Y) = \log \int e^{-\langle Y,X\rangle} d\mu(X) = \frac{n}{2}\log(\pi) - \frac{1}{2}\log\det(Y).$$

Since μ is the pushforward of the Lebesgue measure through $x \mapsto xx^\top$, this expression follows from the following classical Gaussian integral formula:

$$\int_{\mathcal{V}_1} e^{-\langle Y,X\rangle} d\mu(X) = \int_{\mathbb{R}^n} e^{-x^\top Y x} dx = \sqrt{\det(\pi Y)}. \tag{1}$$

We show how leads to our optimality characterization of the Goemans-Williamson measure at the end of this section.

The above Gaussian formula for $\mathcal{V}_1$ suggests a natural approach for computing $\mathcal{E}_1$ on $\mathcal{P}_1$. Allowing complex Hermitian matrices, note that $\mathcal{P}_1$ is the set of norm-1 elements of $\mathcal{V}_1$. Hence, we "integrate out" the norm of the elements of $\mathcal{V}_1$, in an attempt to obtain a similar formula for $\mathcal{P}_1$. We do this via a standard change of variables (equalities are up to scalar):

$$\begin{aligned}\int_{\mathcal{V}_1} e^{-\langle Y,X\rangle} d\mu(X) &= \int_{\mathcal{P}_1}\int_0^\infty e^{-\langle Y, r^2X\rangle} r^{2n-1} dr d\mu_1(X)\\ &= \int_{\mathcal{P}_1} \langle Y,X\rangle^{-n} d\mu_1(X) \neq \mathcal{E}_1(Y).\end{aligned}$$

This shows that this approach fails: that is, integrating out the norm does not provide us a formula for $\mathcal{E}_1(Y)$.

This demonstrates the first difficulty for constructing a counting oracle for $\mathcal{P}_1$. Normalizing the max-entropy measure on $\mathcal{V}_1$ as above yields a measure on $\mathcal{P}_1$ which is *not* a max-entropy measure. Max-entropy measures on $\mathcal{P}_1$ an $\mathcal{V}_1$ are therefore fundamentally different objects, and thus constructing the associated counting oracles requires different techniques. In particular the well-known Gaussian integral formulas cannot help us in the case of $\mathcal{P}_1$.

The remarkable fact is then that max-entropy measures on $\mathcal{P}_1$ *can* be translated into max-entropy measures on a very simple polytope: the standard simplex in $\mathbb{R}^n$. We have the following equality for real $Y = \mathrm{diag}(y)$, where m is the Lebesgue measure on the simplex $\Delta_1 := \{p \in \mathbb{R}^n_+ : \sum_{i=1}^n p_i = 1\}$:

$$\int_{\mathcal{P}_1} e^{-\langle Y,X\rangle} d\mu_1(X) = \int_{\Delta_1} e^{-\langle y,x\rangle} dm(x).$$

Put another way, max-entropy measures on $\mathcal{P}_1$, a nonconvex manifold, correspond to max-entropy measures on Δ_1, a convex polytope. To see this, first note the following for any $m_1, \ldots, m_n$. The first equality is the Bombieri inner product formula (see e.g. Lemma 3.2 of [31]), and the second inequality is a basic induction after a change of variables:

$$\begin{aligned}\int_{\mathcal{P}_1} X_{11}^{m_1}\cdots X_{nn}^{m_n} d\mu_1(X) &= \frac{m_1!\cdots m_n!(n-1)!}{(m_1+\cdots+m_n+n-1)!}\\ &= \int_{\Delta_1} x_1^{m_1}\cdots x_n^{m_n} dm(x).\end{aligned}$$

The exponential equality then follows from limiting, since $\mathcal{P}_1$ and Δ_1 are compact and since $e^{-\langle Y,X\rangle}$ and $e^{-\langle y,x\rangle}$ are limits of polynomials.

This argument also implies the more general fact: that m is the pushforward of μ_1 through the map $\phi : X \mapsto \mathrm{diag}(X)$:

$$\int_{\mathcal{P}_1} f(\phi(X)) d\mu_1(X) = \int_{\Delta_1} f(x) dm(x).$$

This transfer to the simplex now leads to an explicit computation for $\mathcal{E}_1(Y)$ when $Y = \mathrm{diag}(y)$. (Considering diagonal Y is actually without loss of generality.) By making a change of variables, the simplex integral is an iterated convolution:

$$\begin{aligned}&\frac{1}{(n-1)!}\int_{\mathcal{P}_1} e^{-\langle Y,X\rangle} d\mu_1(X)\\ &\quad= \int_0^1\int_0^{1-x_1}\cdots\int_0^{1-x_1-\cdots-x_{n-2}} e^{-\langle y,x\rangle} dx\\ &\quad= (e^{-y_1 t} * \cdots * e^{-y_n t})\big|_{t=1}.\end{aligned}$$

Applying the Laplace transform $\mathcal{L}$ converts this convolution into a partial fraction decomposition problem for distinct values of y_i:

$$\begin{aligned}(e^{-y_1 t} * \cdots * e^{-y_n t})\big|_{t=1} &= \mathcal{L}^{-1}\left[\frac{1}{\prod_i (s+y_i)}\right](1)\\ &= \mathcal{L}^{-1}\left[\sum_i \frac{c_i}{s+y_i}\right](1)\\ &= \sum_i c_i e^{-y_i}.\end{aligned}$$

Computing the values of c_i via a standard partial fractions formula gives:

$$\begin{aligned}\frac{1}{(n-1)!}\int_{\mathcal{P}_1} e^{-\langle Y,X\rangle} d\mu_1(X) &= \sum_{i=1}^n \frac{e^{-y_i}}{\prod_{j\neq i}(y_j - y_i)}\\ &= \frac{\det(M(-y))}{\prod_{i<j}(y_j - y_i)}.\end{aligned}$$

Here $M(-y)$ is a Vandermonde-like matrix which arises when forming the common denominator of the last expression, defined as follows.

Definition 6.1 (Matrix for dual integral, $k = 1$). *Given $y_1, \ldots, y_n \in \mathbb{R}$, let $\lambda_1 < \cdots < \lambda_k$ denote the distinct values of y_i*

with multiplicities $m_1, \dots, m_k$. We define an $n \times n$ matrix $M(y)$ as:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots \\ \lambda_1 & 1 & \cdots & 0 & \lambda_2 & 1 & \cdots \\ \lambda_1^2 & \binom{2}{1}\lambda_1 & \cdots & 0 & \lambda_2^2 & \binom{2}{1}\lambda_2 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \cdots \\ \lambda_1^{n-2} & \binom{n-2}{1}\lambda_1^{n-3} & \cdots & \binom{n-2}{m_1-1}\lambda_1^{n-m_1-1} & \lambda_2^{n-2} & \binom{n-2}{1}\lambda_2^{n-3} & \cdots \\ \frac{e^{\lambda_1}}{0!} & \frac{e^{\lambda_1}}{1!} & \cdots & \frac{e^{\lambda_1}}{(m_1-1)!} & \frac{e^{\lambda_2}}{0!} & \frac{e^{\lambda_2}}{1!} & \cdots \end{bmatrix}.$$

This brings us to the second difficulty for constructing a counting oracle for $\mathcal{P}_1$. When the values of y_i are not distinct, then the denominator vanishes and this formula cannot be used. Even though $\mathcal{E}_1$ is continuous, this could still be a major problem: if for example the gradient of $\mathcal{E}_1(Y)$ becomes large as y_i approaches y_j, then computing $\mathcal{E}_1(Y)$ could become computationally infeasible.

To handle this difficulty, we take limits by successively applying L'Hopital's rule. One iteration for $y_1 = y_2$ goes as follows:

$$\lim_{y_2 \to y_1} \frac{\det(M(-y))}{\prod_{i<j}(y_j - y_i)} = \left.\frac{\partial_{y_2} \det(M(-y))}{\partial_{y_2} \prod_{i<j}(y_i - y_j)}\right|_{y_2 = y_1} = \frac{\det(M(-y))}{\prod_{2<i}(y_i - y_1)^2 \prod_{2<i<j}(y_i - y_j)}.$$

The key observation here is the fact that the numerator is still a determinant, due to the fact that only one column of $M(-y)$ depends on y_i for all i. Applying L'Hopital's rule as many times as is necessary leads to the following, where λ_i represent the distinct values of y with multiplicities m_i:

$$\exp(\mathcal{E}_1(Y)) = \int_{\mathcal{P}_1} e^{-\langle Y, X\rangle} d\mu_1(X) = (n-1)! \frac{\det(M(-y))}{\prod_{i<j}(\lambda_i - \lambda_j)^{m_i m_j}}.$$

Note that the definition of $M(y)$ (Definition 6.1 above) already handles the non-distinctness of the eigenvalues. A similar expression for the gradient is achieved using the same techniques, and so we state it here without further detail:

$$(\nabla \mathcal{E}_1(Y))_l = -\sum_{i \neq p} \frac{m_i}{\lambda_p - \lambda_i} - \frac{\det(M_p(-y))}{\det(M(-y))}.$$

Here $M_p(y)$ is obtained by applying the operator $\frac{\partial_{\lambda_p}}{m_p}$ to the rightmost column of $M(y)$ that depends on λ_p.

Since the entries of $M(-y)$ and $M_p(-y)$ have bit complexity polynomial in n and the bit complexity of e^{-y_i}, their determinants have the same bit complexity. Therefore these formulas, for $\mathcal{E}_1(Y)$ and its gradient, lead to an efficient counting oracle for $\mathcal{P}_1$.

Proof overview: sampling for $\mathcal{P}_1$. We now discuss how to sample from max-entropy distributions on $\mathcal{P}_1$. Our main algorithm (Theorem 4.4) gives an efficient oracle for approximating the max-entropy density function:

$$\nu(X) \propto e^{-\langle Y^\star, X\rangle}.$$

The main problem is that it is not at all clear how to use such a density function to sample from a manifold.

We avoid this difficulty by transferring the problem of sampling to the simplex Δ_1 for $Y^\star = \mathrm{diag}(y^\star)$, using the following fact discussed in the previous section:

$$\frac{1}{(n-1)!} \int_{\mathcal{P}_1} e^{-\langle Y^\star, X\rangle} d\mu_1(X) = \int_0^1 \int_0^{1-x_1} \cdots \int_0^{1-x_1-\cdots-x_{n-2}} e^{-\langle y^\star, x\rangle} dx.$$

The sampling process for $\mathcal{P}_1$ then occurs in two parts.

First, we sample from the max-entropy distribution on the simplex, one coordinate at a time. We use the right-hand side of the above expression to compute the cumulative density function (CDF) for each coordinate, conditioned on the previously sampled coordinates. Formulas and computations for these conditioned CDFs are very similar to that of the counting oracle, and hence we omit them here.

Once we have a sample x on the simplex, we need to convert it into a sample on $\mathcal{P}_1$ by considering its inverse image under the map $\phi : X \mapsto \mathrm{diag}(X)$. The difficulty that now arises is the fact that there are many elements of $\mathcal{P}_1$ which map to the same simplex element under ϕ.

Fortunately, there is a principled way to select from these possibilities. The fiber $\phi^{-1}(x)$ is an orbit of the action of diagonal unitary matrices on $\mathcal{P}_1$ by conjugation. Since $Y^\star$ is diagonal, this implies the max-entropy measure $\nu(X)$ is uniform when restricted to $\phi^{-1}(x)$. Given x, we then sample X from $\phi^{-1}(x)$ by picking an arbitrary $X_0 \in \phi^{-1}(x)$ and conjugating by a uniformly random diagonal unitary matrix.

Hence, to sample X from $\mathcal{P}_1$ we (1) sample x from the simplex, and then (2) sample X uniformly from $\phi^{-1}(x)$. This samples X from the correct measure due to the disintegration theorem (see [9]), which says the following for any f:

$$\int_{\mathcal{P}_1} f(X) d\mu_1(X) = \int_{\Delta_1} \int_{\phi^{-1}(x)} f(X) d\mu_{\phi^{-1}(x)}(X) dm(x).$$

That is, the measure μ_1 can be split into measures on Δ_1 and on the fibers $\phi^{-1}(x)$.

Therefore, the above sampling process efficiently samples the max-entropy measure on $\mathcal{P}_1$ with density $\nu(X)$.

6.3 Proof Overview: Extending the Counting Oracle for $\mathcal{P}_1$ to $\mathcal{P}_k$

For the case of $\mathcal{P}_k$ and μ_k, we want to generalize the formulas of the $k = 1$ case. To do this, we make use of the famous Harish-Chandra-Itzykson-Zuber formula for integrals over the Haar measure of the unitary group $U(n)$.

Theorem 6.1 (HCIZ formula). *For $n \times n$ Hermitian matrices Y and B with distinct eigenvalues $y_1 < \cdots < y_n$ and $\beta_1 < \cdots < \beta_n$ respectively, we have the following where μ is the Haar measure on the unitary group $U(n)$:*

$$\int_{U(n)} e^{\langle Y, UBU^*\rangle} d\mu(U) = \left(\prod_{p=1}^{n-1} p!\right) \frac{\det([e^{y_i \beta_j}]_{1 \le i,j \le n})}{\prod_{i<j}(y_j - y_i)(\beta_j - \beta_i)}.$$

For $B = \mathrm{diag}(1, \dots, 1, 0, \dots, 0)$ with k 1s and $n-k$ 0s, notice that $\mathcal{P}_k = \{UBU^* : U \in U(n)\}$. This leads to the following:

$$\exp(\mathcal{E}_k(Y)) = \int_{\mathcal{P}_k} e^{-\langle Y, X\rangle} d\mu_k(X) = \int_{U(n)} e^{-\langle Y, UBU^*\rangle} dU.$$

To handle the issue of the denominator vanishing, and to compute the gradient, one can apply all the same techniques which were required for the $k = 1$ case. These formulas end up having the right bit complexity, and so they immediately imply an efficient strong counting oracle for $\mathcal{P}_k$.

Unlike in the case of $k = 1$, the problem of sampling in the case of $k > 1$ is more difficult as the image of $\mathcal{P}_k$ under the map $\phi : X \mapsto \text{diag}(X)$ is much more complicated. Thus we leave as an open problem the question of sampling from the associated maximum entropy distributions in the case of $\mathcal{P}_k$ for $k > 1$.

7 BOUNDING BOX

In this section, we prove the general bounding box result (Theorem 4.2). With this, we then specialize to the cases of rank-k projections and convex bodies.

7.1 General Bounding Box

In what follows we will discuss "interiors" of a probability distribution μ given by two parameters, (η, δ). The η parameter will control how far we are from the boundary, and the δ parameter will control how well-distributed μ is on its support. At the end of the day, we will prove that for nice situations one only needs to consider the η parameter (as in the bounding box result of [34]).

We now define the two-parameter interior. In what follows, we will let $V_{\mathcal{L}}$ be the vector subspace given by $\mathcal{L}(X) = 0$, where $\mathcal{L}(X) = B$ is the maximal set of linearly independent equality constraints for Ω. More informally, $V_{\mathcal{L}}$ is the vector space corresponding to the minimal affine space in which $\mathcal{K} = \text{hull}(\Omega)$ lives (i.e., translate the affine space so that $0 \in V_{\mathcal{L}}$). The fact that $\mathcal{L}(X) = B$ is a maximal linearly independent set means that the optimal solution to the dual program is unique when restricted to $V_{\mathcal{L}}$.

Definition 7.1. *We define the $(0, \delta)$-interior of μ to be the set of all $A \in \mathcal{K}$ such that for all $Y \in V_{\mathcal{L}}$ we have:*

$$\mu(\{X \in \Omega \mid \langle X - A, Y\rangle \geq 0\}) > \delta.$$

Morally, this says that every closed half-space containing A contains more than δ of the mass of μ. Note that this is not always an open set (which is perhaps a bit odd for something called the "interior", but this will be our convention).

Definition 7.2 (Two-parameter interior). *We define the (η, δ)-interior of μ to be the set of all $A \in \mathcal{K}$ such that the ball of radius η about A is contained in the $(0, \delta)$-interior of μ. Note that this is not necessarily an open set.*

The next lemma is then precisely how to combine the two parameters to get a bounding box for the optimal solution to the dual program.

Lemma 7.1 (Two-parameter bounding box). *Given $A \in \mathcal{K}$, let $Y^\star \in V_{\mathcal{L}}$ be the optimal solution to the dual program. Recall the dual objective:*

$$\inf_Y F_A(Y) = \inf_Y \log \int_\Omega e^{-\langle Y, X-A\rangle} d\mu(X).$$

If A is in the (η, δ)-interior of μ, then $\|Y^\star\| \leq \frac{1}{\eta} \log\left(\frac{1}{\delta}\right)$.

Proof. By definition, we have that $A - \eta \cdot \frac{Y^\star}{\|Y^\star\|}$ is in the $(0, \delta)$-interior of $\mathcal{K}$. Therefore:

$$\begin{aligned}\delta &\leq \mu(\{X \in \Omega \mid \langle X - (A - \eta \cdot Y^\star/\|Y^\star\|), -Y^\star\rangle \geq 0\}) \\ &= \mu(\{X \in \Omega \mid \langle X - A, -Y^\star\rangle \geq \eta \cdot \|Y^\star\|\}).\end{aligned}$$

This gives the bound:

$$\log \int e^{\langle -Y^\star, X-A\rangle} d\mu(X) \geq \log\left(\delta \cdot e^{\eta \cdot \|Y^\star\|}\right) = \log(\delta) + \eta \cdot \|Y^\star\|.$$

On the other hand, plugging in $Y = 0$ gives an upper bound on the optimal value of the above dual program:

$$0 \geq \log \int e^{\langle -Y^\star, X-A\rangle} d\mu(X) \geq \log(\delta) + \eta \cdot \|Y^\star\|.$$

Rearranging this gives the result. ∎

This gives us a good way of bounding solutions corresponding to interior points of $\mathcal{K}$. In general however, trying to get a bound on the δ parameter of the interior is much more difficult than that of the η parameter. To deal with this we define a property of μ which allows us to only have to consider the η parameter.

Definition 7.3 (δ-balanced measure). *We say that μ is δ-*balanced *if for any $X \in \Omega$, we have that at least $\exp(-\text{poly}(\delta^{-1}, d))$ of the mass of μ is contained in the δ-ball about X (where d is the dimension of $\mathcal{K}$). If f is the polynomial in the exponent (i.e., $\exp(-f(\delta^{-1}, d))$), then we say that μ is δ-*balanced with bound f.

We now prove the main bounding box theorem for such balanced measures. We then use this to obtain a bounding box for rank-k projections and for convex bodies in the following sections.

Theorem 7.2 (Bounding box for balanced measures). *Suppose μ is $\frac{\eta}{2}$-balanced with bound f. If A is in the $(\eta, 0)$-interior of μ and $Y^\star \in V_{\mathcal{L}}$ is the optimal solution to the corresponding dual program, then $\|Y^\star\| \leq 2\eta^{-1} \cdot f(2\eta^{-1}, d) = \text{poly}(\eta^{-1}, d)$.*

Proof. We first show that the $(\frac{\eta}{2}, 0)$-interior of μ is contained in the $(0, \exp(-f(\frac{2}{\eta}, d)))$-interior of μ. To see this, let A_0 be some element of the $(\frac{\eta}{2}, 0)$-interior of μ. Then any closed half-space containing A_0 also contains an $\frac{\eta}{2}$-ball about some $X \in \Omega$. That is, for every $Y \in V_{\mathcal{L}}$ there exists X such that:

$$B_{\eta/2}(X) \subseteq \{Z \in \Omega \mid \langle Z - A_0, Y\rangle \geq 0\}.$$

Since μ is $\frac{\eta}{2}$-balanced, we have that $\exp(-f(\frac{2}{\eta}, d))$ of the mass of μ is contained in the $\frac{\eta}{2}$-ball about X. This implies:

$$\exp(-f(2/\eta, d)) \leq \mu(B_{\eta/2}(X)) \leq \mu(\{Z \in \Omega \mid \langle Z - A_0, Y\rangle \geq 0\}).$$

That is, A_0 is in the $(0, \exp(-f(\frac{2}{\eta}, d)))$-interior of μ.

Now for A in the $(\eta, 0)$-interior of μ, we have that the $\frac{\eta}{2}$-ball about A is contained in the $(\frac{\eta}{2}, 0)$-interior of μ. Therefore A is in the $(\frac{\eta}{2}, \exp(-f(\frac{2}{\eta}, d)))$-interior of μ. By Lemma 7.1, this implies $\|Y^\star\| \leq 2\eta^{-1} \cdot f(2\eta^{-1}, d)$. ∎

Remark 7.3. *Note that Theorem 7.2 is immediately applicable to uniform discrete measures on (singly) exponentially sized sets S. In particular, such a measure is automatically balanced with constant bound $f = \log|S|$.*

7.2 Rank-k Projections

We now prove bounding box result for $\mathcal{P}_k$, by showing that μ_k is balanced and applying the previous theorem. Note that in this case $\mathcal{L}(X) = B$ reduces to $\mathrm{Tr}(X) = k$, and so $V_{\mathcal{L}}$ is the set of traceless Hermitian matrices in this case.

Corollary 7.4 (Bounding box for $\mathcal{P}_k$). *Let μ_k be the uniform distribution on $\mathcal{P}_k$. Then given A in the $(\eta, 0)$-interior of μ_k, the optimal traceless solution $Y^\star$ of the corresponding dual program is such that $\|Y^\star\| \leq \frac{2n^2}{\eta} \log\left(\frac{8n\sqrt{k}}{\eta}\right)$.*

Proof. We prove that μ_k is balanced and then apply the previous proposition. The number of balls of size δ required to cover the unit ball in $\mathbb{R}^{n^2}$ (with Euclidean/Frobenius norm) is at most $(2n/\delta)^{n^2}$. Since the set of projections of rank k is contained in the sphere of radius $\sqrt{k}$, we have that it requires at most $(2n\sqrt{k}/\delta)^{n^2}$ δ-balls to cover all such projections. With this, there exists some δ-ball (call it B_δ) in this cover which contains at least $(2n\sqrt{k}/\delta)^{-n^2}$ of the mass of μ_k. Pick some $X \in \mathcal{P}_k \cap B_\delta$, and let $B_{2\delta}(X)$ be the ball of radius 2δ which is centered at X. Thus, in fact $B_{2\delta}(X)$ contains at least $(2n\sqrt{k}/\delta)^{-n^2}$ of the mass of μ_k. By unitary invariance of μ_k, we have that the ball of radius 2δ about any point of $\mathcal{P}_k$ contains at least $(2n\sqrt{k}/\delta)^{-n^2}$ of the mass of μ_k. That is, μ_k is δ-balanced with bound $f(\delta^{-1}, n) = n^2 \log(4n\sqrt{k} \cdot \delta^{-1})$ for all $\delta > 0$. Applying the previous proposition then gives the result. ∎

7.3 Convex Bodies

We now prove bounding box result for convex bodies. Instead of applying the previous theorem directly, we make a simpler computation in the same spirit.

Proof of Corollary 4.15. Note that v in the η-interior of μ is automatically in the $\left(\frac{\eta}{2}, \left(\frac{\eta}{4R}\right)^d\right)$-interior of μ, since:

$$\mu(B_{\eta/4}) \geq \frac{\mathrm{vol}(B_{\eta/4})}{\mathrm{vol}(B_R)} = \left(\frac{\eta}{4R}\right)^d.$$

By Lemma 7.1, this implies $\|y^\star\| \leq \frac{2d}{\eta} \log(\frac{4R}{\eta})$. ∎

8 COMPUTING MAXIMUM ENTROPY MEASURES

In this section we describe the entire algorithm for computing the optimum $Y^\star$ for the dual program $\mathrm{Dual}_\mu(A)$, which is essentially an application of the ellipsoid algorithm. First though, we discuss how the linear equality constraints $\mathcal{L}(X) = B$ come into play here. We want to restrict our search space to the vector space $V_{\mathcal{L}}$ defined as the set of all X such that $\mathcal{L}(X) = 0$. The main reason for this is, since the constraints given by $\mathcal{L}(X) = B$ pick out an affine space in which $\mathcal{K}$ is full dimensional, restricting the search space to $V_{\mathcal{L}}$ causes the optimum $Y^\star$ to be unique. Further, the bounding box results above apply specifically to this particular $Y^\star$.

Since we are given $\mathcal{L}$ effectively and explicitly, we assume for the ellipsoid algorithm that we can project the gradient (given by the strong counting oracle) onto $V_{\mathcal{L}}$. That said, we will from now on assume $V_{\mathcal{L}}$ to be the domain in which we are optimizing.

8.1 The Ellipsoid Framework

Using a standard argument via Hölder's inequality, we have that the dual objective function is convex:

$$F_A(Y) := \langle Y, A\rangle + \mathcal{E}_\mu(Y) = \langle Y, A\rangle + \log\left(\int_\Omega e^{-\langle Y, X\rangle} d\mu(X)\right).$$

With this, the main optimization tool we use to approximate the the dual optimum $Y^\star$ is the ellipsoid algorithm. Recall the following from [34] Theorem 2.13, which was essentially taken from [5].

Theorem 8.1 (Ellipsoid algorithm). *Given any $\beta > 0$ and $R > 0$, there is an algorithm which, given a strong first-order oracle for F_A, returns a $Y^\circ \in V_{\mathcal{L}}$ such that:*

$$F_A(Y^\circ) \leq \inf_{Y \in V_{\mathcal{L}}, \|Y\|_\infty \leq R} F_A(Y) + \beta\left(\sup_{Y \in V_{\mathcal{L}}, \|Y\|_\infty \leq R} F_A(Y) - \inf_{Y \in V_{\mathcal{L}}, \|Y\|_\infty \leq R} F_A(Y)\right).$$

The number of calls to the strong first-order oracle for F_A is bounded by a polynomial in d, $\log R$, and $\log(1/\beta)$. Here, d is the dimension of the ambient Hilbert space in which Ω lies.

We now prove the main theorem (Theorem 4.4) regarding the existence of an algorithm for approximating the optimum to the dual objective.

Proof of Theorem 4.4. To apply the ellipsoid algorithm, we need to choose the two parameters, β and R. Since μ is balanced with some polynomial bound f, we choose for R the bounding box given for balanced measures in Theorem 4.2:

$$R := 2\eta^{-1} \cdot f(2\eta^{-1}, d).$$

So, the set $\{Y \in V_{\mathcal{L}} : \|Y\| \leq R\} \subset \{Y \in V_{\mathcal{L}} : \|Y\|_\infty \leq R\}$ contains the optimal $Y^\star$ for the dual program. Next, we need to choose β. Note that for $\|Y\|_\infty \leq R$ we have:

$$\begin{aligned} |F_A(Y)| &\leq |\langle Y, A\rangle| + \left|\log \int e^{-\langle Y, X\rangle} d\mu(X)\right| \\ &\leq r\|Y\|_\infty + r\|Y\|_\infty \\ &\leq 2r\sqrt{d}\|Y\| \leq 2rR\sqrt{d}. \end{aligned}$$

Therefore, choosing $\beta := \frac{\varepsilon}{4rR\sqrt{d}}$ implies:

$$\beta = \frac{\varepsilon}{4rR\sqrt{d}} \leq \frac{\varepsilon}{\sup_{Y \in V_{\mathcal{L}}, \|Y\|_\infty \leq R} F_A(Y) - \inf_{Y \in V_{\mathcal{L}}, \|Y\|_\infty \leq R} F_A(Y)}.$$

The ellipsoid algorithm then guarantees a Y° such that:

$$F_A(Y^\circ) \leq \inf_{Y \in V_{\mathcal{L}}, \|Y\|_\infty \leq R} F_A(Y) + \varepsilon = F_A(Y^\star) + \varepsilon.$$

The number of calls to the strong counting oracle is polynomial in d, $\log(R) = \log(2\eta^{-1} \cdot f(2\eta^{-1}))$ and $\log(1/\beta) = \log(4rR\sqrt{d}\varepsilon^{-1})$. Given the bounding box, each oracle call (now including computing $\langle Y, A\rangle$) can be implemented in time polynomial in d, η^{-1}, and the number of bits needed to represent A. This completes the proof. ∎

8.2 Rank-k Projections

Next we apply the above result to the case of $\Omega = \mathcal{P}_k$ and $\mu = \mu_k$, i.e., the case of rank-k projections. To do so we make a few tweaks to the proof of the theorem for the general algorithm given in the previous section. In particular, even though our domain $\mathcal{P}_k$ lies in the space of Hermitian matrices, our strong counting oracle for $\mathcal{E}_k$ only applies to real diagonal matrices Y. That said, we now prove the result for rank-k projections (Corollary 4.10) and discuss such issues in the proof.

Proof of Corollary 4.10. The result essentially follows from the general case, with a few details that need to be dealt with. First, the maximal linear equalities for $\mathcal{P}_k$ boils down to something very simple within the space of Hermitian matrices. It is simply given by $\mathrm{Tr}(X) = k$. Thus, our search space $V_{\mathcal{L}}$ then becomes the set of traceless Hermitian matrices.

Next, by unitary invariance of μ_k we can assume A is diagonal by unitary conjugation. Once we obtain an approximate optimum Y° for the diagonalized A, we can obtain an approximate optimum for the original A via conjugation by this unitary. Next, by the Schur-Horn theorem [22, 33] we can further assume that $Y^\star$ is diagonal. That is, we can assume A is real diagonal and restrict the domain of $F_A(Y)$ to real diagonal matrices Y.

Once we make this simplifying assumption, we have access to a strong counting/integration oracle for $\mathcal{E}_k(Y)$ by Theorem 4.5. The proof for the general case then goes through (using this strong counting oracle and the bounding box result for rank-k projections), giving the desired result. ∎

9 THE GOEMANS-WILLIAMSON MEASURE

In this section, we demonstrate how the measure associated to the Goemans-Williamson SDP rounding scheme can be interpreted as a max-entropy measure. We abuse notation in this section by letting μ_{GW} refer to the pushforward measure on $\mathcal{V}_1$. Omitted proofs are simple and appear in the full version of the paper.

We describe the Goemans-Williamson SDP rounding scheme formally as follows.

Definition 9.1 (Goemans-Williamson rounding scheme). *Given an $n \times n$ real symmetric positive definite matrix A, let V be a real $n \times n$ matrix such that $VV^\top = A$. The Goemans-Williamson rounding scheme proceeds as follows:*

(1) Sample a random standard Gaussian vector g from $\mathbb{R}^n$.
(2) Return the rank-1 PSD matrix $(Vg)(Vg)^\top$.

The measure associated to this sampling process we refer to as the Goemans-Williamson measure *and denote it μ_{GW}. This measure is supported on the rank-1 real symmetric PSD matrices, which is the set of extreme points of the real symmetric PSD cone.*

Now let m be the Lebesgue measure on $\mathbb{R}^n$, and let μ be the measure on the real symmetric PSD cone which is the pushforward of m through the map $\Phi : x \mapsto xx^\top$. With this we can also give an explicit description of the Goemans-Williamson measure.

Proposition 9.1 (Goemans-Williamson density function). *The Goemans-Williamson measure on the set of rank-1 real symmetric PSD matrices is given by*

$$d\mu_{\mathrm{GW}}(X) \propto e^{-\langle \frac{1}{2}A^{-1}, X\rangle} d\mu(X),$$

where μ is the pushforward of Lebesgue measure through $x \mapsto xx^\top$.

Proof. Let $A = VV^\top$ as in the definition of μ_{GW}. Since a standard Gaussian g is distributed according to $e^{-\frac{1}{2}\|g\|^2} dm(g)$, we can apply the change of variables formula to determine how $x := Vg$ is distributed. We have:

$$x \sim e^{-\frac{1}{2}\|V^{-1}x\|^2} \cdot \det(V^{-1}) dm(x) = e^{-\langle \frac{1}{2}A^{-1}, xx^\top\rangle} \cdot \sqrt{\det(A^{-1})} dm(x).$$

Considering the pushforward of this measure through $x \mapsto xx^\top$ gives the desired result. ∎

Note that strong duality then immediately implies μ_{GW} is a max-entropy measure with respect to μ, since its density function is of the correct form. To demonstrate this more concretely, we prove this explicitly below via an explicit formula $\mathcal{E}_\mu(Y)$. First, the following observation tells us that it is sufficient to restrict $\mathcal{E}_\mu(Y)$ to positive definite Y.

Lemma 9.2. *If Y is not PD, then $\int_{\mathcal{V}_1} e^{-\langle Y, X\rangle} d\mu(X) = +\infty$.*

We now give an explicit formula for $\mathcal{E}_\mu(Y)$ on positive definite Y.

Proposition 9.3 (Lebesgue evaluation formula). *We have the following explicit expression for $\mathcal{E}_\mu(Y)$ for $n \times n$ real symmetric positive definite Y:*

$$\mathcal{E}_\mu(Y) := \log \int_{\mathcal{V}_1} e^{-\langle Y, X\rangle} d\mu(X) = \frac{n}{2}\log(\pi) - \frac{1}{2}\log\det(Y).$$

This then leads to the main result of this section.

Corollary 9.4 (Max-entropy, SDP rounding). *Given an $n \times n$ real symmetric positive definite marginals matrix A, the Goemans-Williamson measure μ_{GW} is the max-entropy measure with respect to μ, the pushforward through $x \mapsto xx^\top$ of the Lebesgue measure on $\mathbb{R}^n$. That is, μ_{GW} is the optimal measure for* $\mathrm{Prim}_\mu(A)$.

Proof. Using Proposition 9.3, a standard computation gives:

$$\nabla \mathcal{E}_\mu(Y) = -\frac{1}{2}\nabla \log\det(Y) = -\frac{1}{2}Y^{-1}.$$

This implies the following regarding the gradient of the dual program objective $\mathrm{Dual}_\mu(A)$ for positive definite A:

$$0 = \nabla F_A(Y) = \nabla(\langle Y, A\rangle + \mathcal{E}_\mu(Y)) = A - \frac{1}{2}Y^{-1} \iff Y = \frac{1}{2}A^{-1}.$$

That is, $Y^\star = \frac{1}{2}A^{-1}$ is the optimum for the dual program. By strong duality for μ and the density function for μ_{GW} given in Proposition 9.1 above, this implies the result. ∎

10 GENERALIZATION OF THE MAXIMUM ENTROPY FRAMEWORK TO LIE GROUPS

Recent work (e.g., [8, 10]) has demonstrated interesting connections between Lie theory and TCS, and the max-entropy framework fits into this context as well. In what follows we will briefly discuss the case of $\Omega = \mathcal{P}_k$ and $\mu = \mu_k$, as well as how this can be generalized. However, a more detailed investigation of the computational aspects of the max-entropy framework in this context is outside the scope of this paper.

We first describe the case of $\Omega = \mathcal{P}_k$ and $\mu = \mu_k$ in a more general way. The unitary group $U(n)$ acts on the real vector space of $n \times n$ Hermitian matrices by conjugation. This group action

partitions the vector space into orbits, with X and Y being in the same orbit if and only if they have the eigenvalues. Given any Hermitian matrix F, we denote the orbit of F by $\mathcal{O}(F)$.

Consider now the matrix $P_k := \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$ where k denotes the number of 1s that appear in the matrix. Then the orbit $\mathcal{O}(P_k)$ is precisely the set of rank-k projections. That is, $\mathcal{O}(P_k) = \mathcal{P}_k$, and so the unitarily invariant measure μ_k on $\mathcal{P}_k$ induces such a measure on $\mathcal{O}(P_k)$. In fact such a unitarily invariant measure μ_F exists for any orbit $\mathcal{O}(F)$ allowing us to extend our maximum entropy framework to such orbits of $U(n)$.

This can be generalized beyond the group $U(n)$, to the general setting of a Lie group G and its corresponding Lie algebra $\mathfrak{g}$ upon which G naturally acts. The primal and dual programs for this generalized setting are the same as in the general case, with one exception. The element $F \in \mathfrak{g}$ is now an input, and any algorithm for approximating an optimum for $\mathrm{Dual}_{\mu_F}(A)$ will necessarily depend on the complexity of F. That said, strong duality holds in this case whenever A is in the interior of $\mathcal{K} = \mathrm{hull}(\mathcal{O}(F)) \subset \mathfrak{g}$, and so the bounding box and the strong counting oracle are the two main results needed to obtain the polynomial-time ellipsoid-based algorithm described in this paper. As an aside, in this case $\mathcal{K} = \mathrm{hull}(\mathcal{O}(F))$ is called an *orbitope* (e.g., see [4, 32]).

Thus, the following optimization problem is a natural generalization of the (dual) maximum entropy problem considered in this paper. The G-invariant inner product used in the exponent here can be derived from the so-called Killing form of $\mathfrak{g}$ when G is compact (e.g., see [28], Corollary 4.26).

$$\inf_{Y \in \mathfrak{g}} \left[\langle Y, A \rangle + \log \int_{\mathcal{O}(F)} e^{-\langle Y, X \rangle} d\mu_F(X) \right]$$

Computability of this problem will be a subject of future work.

ACKNOWLEDGMENTS

The authors would like to thank Sushant Sachdeva, Sebastien Bubeck, and Umesh Vazirani for useful discussions. They would also like to thank Simons Institute for the Theory of Computing where this work was initiated. This research was partially supported by NSF CCF-1908347 grant and by Vetenskapsrådet.

REFERENCES

[1] Zeyuan Allen Zhu, Yuanzhi Li, Rafael Oliveira, and Avi Wigderson. Much faster algorithms for matrix scaling. In *FOCS'17: Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science*, 2017.

[2] Arash Asadpour, Michel X Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. An $O(\log n / \log \log n)$-approximation algorithm for the asymmetric traveling salesman problem. *Operations Research*, 65(4):1043–1061, 2017.

[3] William Band and James L. Park. New information-theoretic foundations for quantum statistics. *Foundations of Physics*, 6(3):249–262, Jun 1976.

[4] Alexander Barvinok and Isabella Novik. A centrally symmetric version of the cyclic polytope. *Discrete & Computational Geometry*, 39(1-3):76–99, 2008.

[5] Aharon Ben-Tal and Arkadi Nemirovski. Optimization III: Convex analysis, nonlinear programming theory, nonlinear programming algorithms. Lecture Notes, 2012.

[6] Christopher Bingham. An antipodally symmetric distribution on the sphere. *Ann. Statist.*, 2(6):1201–1225, 11 1974.

[7] Sébastien Bubeck and Ronen Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. In *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 279–279, Paris, France, 03–06 Jul 2015. PMLR.

[8] Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Mendes de Oliveira, Michael Walter, and Avi Wigderson. Efficient algorithms for tensor scaling, quantum marginals, and moment polytopes. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 883–897, 2018.

[9] Joseph T Chang and David Pollard. Conditioning as disintegration. *Statistica Neerlandica*, 51(3):287–317, 1997.

[10] Matthias Christandl, Brent Doran, Stavros Kousidis, and Michael Walter. Eigenvalue distributions of reduced density matrices. *Communications in mathematical physics*, 332(1):1–52, 2014.

[11] Michael B. Cohen, Aleksander Madry, Dimitris Tsipras, and Adrian Vladu. Matrix scaling and balancing via box constrained Newton's method and interior point methods. In *FOCS'17: Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science*, 2017.

[12] J. J. Duistermaat and G. J. Heckman. On the variation in the cohomology of the symplectic form of the reduced phase space. *Inventiones mathematicae*, 69(2):259–268, Jun 1982.

[13] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. Operator scaling: theory and applications. *Foundations of Computational Mathematics*, pages 1–68, 2015.

[14] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, November 1995.

[15] M. Gromov. *Convex sets and Kahler manifolds*, pages 1–38. 1990.

[16] Osman Güler. Barrier functions in interior point methods. *Mathematics of Operations Research*, 21(4):860–885, 1996.

[17] Osman. Güler. On the self-concordance of the universal barrier function. *SIAM Journal on Optimization*, 7(2):295–303, 1997.

[18] Osman Güler and Levent Tunçel. Characterization of the barrier parameter of homogeneous convex cones. *Mathematical Programming*, 81(1):55–76, Mar 1998.

[19] Leonid Gurvits and Alex Samorodnitsky. A deterministic polynomial-time algorithm for approximating mixed discriminant and mixed volume, and a combinatorial corollary. *Discrete & Computational Geometry*, 27:531–550, 2002.

[20] Harish-Chandra. Differential operators on a semisimple lie algebra. *American Journal of Mathematics*, 79(1):87–120, 1957.

[21] Peter D. Hoff. Simulation of the matrix bingham-von mises-fisher distribution, with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics*, 18(2):438–456, 2009.

[22] Alfred Horn. Doubly stochastic matrices and the diagonal of a rotation matrix. *American Journal of Mathematics*, 76(3):620–630, 1954.

[23] C. Itzykson and J. Zuber. The planar approximation. ii. *Journal of Mathematical Physics*, 21(3):411–421, 1980.

[24] Edwin T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106:620–630, May 1957.

[25] Edwin T. Jaynes. Information theory and statistical mechanics. II. *Physical Review*, 108:171–190, October 1957.

[26] C. G. Khatri and K. V. Mardia. The von mises-fisher matrix distribution in orientation statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):95–106, 1977.

[27] B. Klartag. On convex perturbations with a bounded isotropic constant. *Geometric & Functional Analysis GAFA*, 16(6):1274–1290, Dec 2006.

[28] Anthony W Knapp. *Lie groups beyond an introduction*, volume 140. Springer Science & Business Media, 2013.

[29] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13. Siam, 1994.

[30] James L. Park and William Band. Rigorous information-theoretic derivation of quantum-statistical thermodynamics. i. *Foundations of Physics*, 7(3):233–244, Apr 1977.

[31] Damián Pinasco. Lower bounds for norms of products of polynomials via bombieri inequality. *Transactions of the American Mathematical Society*, 364(8):3993–4010, 2012.

[32] Raman Sanyal, Frank Sottile, and Bernd Sturmfels. Orbitopes. *Mathematika*, 57(2):275–314, 2011.

[33] Issai Schur. Uber eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie. *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 22(9-20):51, 1923.

[34] Mohit Singh and Nisheeth K Vishnoi. Entropy, optimization and counting. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 50–59. ACM, 2014.

[35] Paul B. Slater. Relations between the barycentric and von neumann entropies of a density matrix. *Physics Letters A*, 159(8):411 – 414, 1991.

[36] Damian Straszak and Nisheeth K. Vishnoi. Maximum entropy distributions: Bit complexity and stability. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 2861–2891, Phoenix, USA, 25–28 Jun 2019. PMLR.

[37] Michèle Vergne. Convex polytopes and quantization of symplectic manifolds. *Proceedings of the National Academy of Sciences*, 93(25):14238–14242, 1996.

[38] J. von Neumann and R.T. Beyer. *Mathematical Foundations of Quantum Mechanics*. Goldstine Printed Materials. Princeton University Press, 1955.