GIBBS MEASURES FOR SEMIDEFINITE PROGRAMMING

GOVIND MENON

Abstract. These notes document an ongoing study of Gibbs measures for semidefinite programming. Soft and hard constraint models are introduced and the partition function is explicitly computed in terms of Harish-Chandra (HC) integrals. The structure of this formula involves a generalized heat flow in a dual space, along with a Hamilton-Jacobi limit as $\beta \to \infty$.

Contents

1. Introduction

The purpose of these notes is to apply the formalism of Gibbs measures to semidefinite programming and to compute the related partition functions using Harish-Chandra integrals.

Key words and phrases. Semidefinite programming, Harish-Chandra integrals.
This work is supported by the National Science Foundation (DMS 1714187).

1.1. **Notation.** The abbreviations LP and SDP mean linear and semidefinite programming respectively.

We will work with the space of Hermitian and Hermitian positive definite matrices, denoted $\mathbb{H}(n)$ and $\mathbb{P}(n)$ respectively. Similar calculations hold for real symmetric matrices. We use the same terminology in both settings, writing $\mathbb{P}(n;\mathbb{C})$ and $\mathbb{H}(n;\mathbb{C})$, or $\mathbb{P}(n;\mathbb{R})$ and $\mathbb{H}(n;\mathbb{R})$, when there is a need to distinguish between the two. Standard SDP is on $\mathbb{P}(n;\mathbb{R})$ and SDP over $\mathbb{P}(n;\mathbb{C})$ can be reduced to SDP on $\mathbb{P}(n;\mathbb{R})$ [6, 4.42, pp.202]. However, the punchline here involves Harish-Chandra integrals. Since these are simpler in the Hermitian setting, we consider $\mathbb{P}(n;\mathbb{C})$ first, inverting the traditional relation. The asterisk * is used to denote the conjugate transpose, or just transpose, depending on whether one works with real or complex entries. The Frobenius inner product on $\mathbb{H}(n)$,

$$(M,N) := \operatorname{Re}\operatorname{Tr}(M^*N), \quad M,N\in\mathbb{M}(n), \tag{1.1}$$

is used throughout. The asterisk is dropped when M and N are Hermitian. The symbols $\preceq$ and $\succeq$ denote the Loewner order on Hermitian matrices, as is common in optimization [6, 13].

1.2. **Semidefinite programming (SDP).** We consider an SDP whose primal form is

$$\text{minimize} \quad c^*x \tag{1.2}$$

$$\text{subject to} \quad x_1A_1 + \dots x_mA_m + B \preceq 0. \tag{1.3}$$

Here x and c are vectors in $\mathbb{R}^m$, $\{A_j\}_{j=1}^m$ and B_j are given matrices in $\mathbb{H}(n)$. LP may be recovered from SDP by restricting attention to diagonal matrices. [1]

The dual form of the SDP (1.2)-(1.3) is

$$\text{maximize} \quad \operatorname{Tr}(BX) \tag{1.4}$$

$$\text{subject to} \quad \operatorname{Tr}(A_jX) + c_j = 0, \quad 1\le j\le m, \tag{1.5}$$

$$\text{and} \quad X \succeq 0. \tag{1.6}$$

The primal and dual formulation have the same solution when we assume strict feasibility of (1.2). This means that we assume there exists $x\in\mathbb{R}^m$ such that the following strict inequality is true:

$$\sum_{j=1}^m x_jA_j + B \prec 0. \tag{1.7}$$

The dual formulation (1.4)–(1.6) will govern most of our work. In particular, we focus on the geometry of the feasible polytope

$$\mathcal{F} := \{X\in\mathbb{P}(n)\,|\operatorname{Tr}(A_jX)+c_j = 0, \quad 1\le j\le m\}. \tag{1.8}$$

The set $\mathcal{F}$ is invariant under a $GL(n)$ action: the SDP (1.2) is equivalent to the SDP obtained through the transformations $A_j\mapsto MA_jM^*$, $B\mapsto MBM^*$ for all $M\in GL(n)$ [6, 4.39]. This invariance underlies the Riemannian geometry of $\mathbb{P}(n)$ with the trace metric that we use below.

[1] One must fix a convention of signs and whether one is working with a maximum or minimum in the primal and dual formulations. The convention adopted here is that of [6, Ex.5.11], since this allows us to check all the steps. The only differences are minor changes in notation and the fact that we work over the complex numbers and Hermitian matrices. Lovasz's notes [13] are an excellent source for the theory of SDP.

1.3. **Gibbs measures.** We will study Gibbs measures for a relaxed problem. There are two reasons for adopting this viewpoint.

(1) Our primary goal is to use SDP to construct stochastic flows on manifolds with applications to nonlinear PDE. In that setting, the cost function is constant on the boundary of the feasible set and one needs to introduce a measure on the extreme points of the feasible set to replace the usual notion of an extremum in SDP. The use of Gibbs measures with finite β is a relaxation of this problem. It is necessary to first study such Gibbs measures in finite-dimensions before extending it to PDE.

(2) Karmarkar's pioneering work on interior point methods for LP was shown to arise from a gradient flow by Bayer and Lagarias [1, 2, 3, 10]. This structure was then extended to SDP by Faybusovich [8]. We believe these structures have a probabilistic origin. Since neither LP nor SDP has a probabilistic structure at first sight, this statement requires some explanation.

   The 'cartoon' that underlies our viewpoint is this. Both LP and SDP are mathematical models for the allocation of resources in large systems. Imagine now that these large systems are equilibrium descriptions of microscopic negotiation between many parties, or a juggling of resources between different tasks, subject to the hard constraints (1.2). How could one model such underlying fluctuations? Work on interior point schemes such as Karmarkar's include a Riemannian metric on $\mathbb{P}(n)$ in addition to the constraints. But once one has a Riemannian metric, one obtains a model for fluctuations 'for free', since each choice of metric determines a Brownian motion in $\mathbb{P}(n)$. This Brownian motion may then be conditioned to respect the constraint (1.8), leading naturally to Gibbs measures for SDP.

For now, this probabilistic cartoon is implicit, not explicit. We will simply postulate a Gibbs measure and compute its partition function. Further, once one works with Gibbs measures, the cost function is less important than the interplay between the geometry of the feasible set $\mathcal{F}$ and the inverse temperature β, especially as $\beta \to \infty$. For this reason, we will first work with the case $C = 0$.

The main formulas are summarized in Section 1.5

1.4. **Outline.** The calculation consists of three steps.

(1) Choose a Riemannian metric on $\mathbb{P}(n)$. Our two favorite examples are the trace metric and the Bures-Wasserstein (BW) metric. Once the metric is chosen, we compute the associated volume form, denoted $\sqrt{\det g} DX$, on $\mathbb{P}(n)$ (this reduces to a computation of Jacobians). Thus, a metric gives us a way to integrate over $\mathbb{P}(n)$ and define probability measures on $\mathbb{P}(n)$.

(2) Use equation (1.8) to define either a soft constraint or hard constraint model. For soft constraints we introduce an energy $E : \mathbb{P}(n) \to \mathbb{R}$ that may be constructed from the constraints (1.8) and define the Gibbs measure

$$\mu_\beta(dX) = \frac{1}{Z_\beta} e^{-\beta E(X)} \sqrt{\det g}\, DX, \quad \beta > 0, \tag{1.9}$$

   with the partition function

$$Z_\beta = \int_{\mathbb{P}(n)} e^{-\beta E(X)} \sqrt{\det g}\, DX. \tag{1.10}$$

The soft constraints are considered first, since these immediately yield fruitful connections with random matrix theory. The disadvantage of soft constraints is that the problem is not as minimal as SDP, since additional heuristics are used to choose V given the constraints (1.8). However, hard constraints may be recovered in the limit $\beta \to \infty$.

(3) Compute the partition function Z_β. We find that Z_β may be expressed explicitly in terms of Harish-Chandra integrals for every feasible SDP; see equation (1.21) below.

Remark 1. The partition function Z_β depends implicitly on the constraints; we have not noted this explicitly in these formulas. Once Z_β is known, standard theory allows the computation of the entropy, free energy and other thermodynamic variables. The primary difficulty in all such models is the computation of Z_β. Our model is explicitly solvable. Further, the formulas seem tractable to rigorous asymptotic analysis, both as $\beta \to \infty$, as well as in the limit $n \to \infty$.

Remark 2. Karmarkar's algorithm and its generalizations are gradient flows of the cost function on $\mathcal{F}$ with respect to the trace metric. The use of the Bures-Wasserstein metric is motivated by recent work in deep learning, as well as its fundamental role in optimal transportation theory. The set of all metrics that are amenable to these methods is classified by Ando and Kubo (see Remark 5). The partition function should be computable in this setting too. There is a well-developed theory of concordant barrier functions in optimization theory. We do not use this explicitly, but for the trace metric, our computation of the volume form recovers the barrier function $\log\det X$. I don't know if the analogous Bures-Wasserstein barrier has been explored. In any case, the Ando-Kubo calculation provides a description of *all* barriers that can be constructed from an underlying Riemannian geometry.

1.5. **Harish-Chandra's integral and the partition function.** We will focus on the unitary group $U(n)$ consisting of complex matrices such that $U^*U = \mathrm{Id}$. In this setting, Harish-Chandra's integral formula, or the Harish-Chandra-Itzykzon-Zuber (HCIZ) integral is

$$\int_{U(n)} e^{\mathrm{Tr}(aUbU^*)}\, dU = \left(\prod_{p=1}^{n-1} p!\right) \frac{\det(e^{a_i b_j})_{i,j=1}^n}{V(A)V(B)} := \mathcal{H}(a,b), \tag{1.11}$$

where $a = \mathrm{diag}(a_1, \ldots, a_n)$ and $b = \mathrm{diag}(b_1, \ldots, b_n)$ are diagonal matrices, and $V(\Lambda)$ is the Vandermonde determinant associated to a diagonal matrix Λ (see equation (3.16) below). [2] Integration is with respect to normalized Haar measure dU on $U(n)$; equivalently the factor $\prod_{p=1}^{n-1} p!$ is the volume of $U(n)$ with respect to Lebesgue measure [3].

Formula (1.11) generalizes to the orthogonal group $O(n)$, but its form is more complicated since it requires the notion of a Cartan subalgebra and the form for even and odd n is different. This is why we prefer to work with Hermitian matrices, returning to real symmetric matrices once this case is understood fully.

Soft constraint models replace the constraint (1.5) with a penalty. We will consider penalty functions defined through m polynomials $p_j : \mathbb{R} \to \mathbb{R}$ each of which

[2] The switch to lower-case for matrices reflects common usage for the Cartan subalgebra.

[3] Need to check these normalizations.

has a global minimum at $x = -c_j$, with $p_j(-c_j) = 0$, and such that $p_j(x) \to +\infty$ as $|x| \to \infty$. The simplest set of such polynomials is

$$p_j(x_j) = \frac{1}{2}(x_j + c_j)^2, \quad x_j \in \mathbb{R}, \quad j = 1, \ldots, m. \tag{1.12}$$

Let us also define the polynomial and differential operator on $\mathbb{R}^m$

$$p(x) = \sum_{j=1}^{m} p_j(x_j), \quad p(\partial_x) = \sum_{j=1}^{m} p_j(\partial_{x_j}), \quad x = (x_1, \ldots, x_m) \in \mathbb{R}^m. \tag{1.13}$$

We use these polynomials to define an energy $E : \mathbb{P}(n) \to \mathbb{R}$

$$E(X) = \sum_{j=1}^{m} p_j(\mathrm{Tr}(A_j X)). \tag{1.14}$$

Clearly, $E(X) \geq 0$ and $E(X) \equiv 0$ on the feasible set $\mathcal{F}$. Several variants of this idea are possible.

The partitition function Z_β defined in equation (1.10) has the following structure

$$Z_\beta = \int_{\mathbb{P}(n)} e^{-\beta E(X)} \sqrt{\det g}\, DX, \tag{1.15}$$

where $\sqrt{\det g} DX$ is a volume form on $\mathbb{P}(n)$ that depends on the underlying metric. The volume form should be seen as the entropic part of the partition function: we compute it explicitly in Section 3 and show that it depends only on the eigenvalues of X. The energetic term, however, depends on both the eigenvalues and eigenvectors of X. Writing $X = U\Lambda U^*$ for the diagonalization and using Weyl's integration formula, Z_β may be written as

$$Z_\beta = \int_{\mathbb{R}^n} \rho_{\mathrm{ref}}(\Lambda) \left(\int_{U(n)} e^{-\beta E(U\Lambda U^*)}\, d\Lambda \right) dU := \int_{\mathbb{R}^n} \rho_{\mathrm{ref}}(\Lambda) Y_\beta(\Lambda) d\Lambda, \tag{1.16}$$

where the term $\rho_{\mathrm{ref}}(\Lambda)$ depends only on the metric, not on $\mathcal{F}$.

Such a splitting is common in random matrix theory. The inner term may be computed as a functional of Harish-Chandra's integral. We show in Section 4 that

$$Y_\beta(\Lambda) = \int_{\mathbb{R}^m} G(-x; \beta) \mathcal{H}(\sigma(A(x)), \Lambda)\, dx. \tag{1.17}$$

Here $G(x, x'; \beta)$ is the fundamental solution for the initial value problem

$$\partial_\beta G = p(\partial_x) G, \quad G(x, x'; 0) = \delta(x - x'), \quad x, x' \in \mathbb{R}^m, \tag{1.18}$$

for the differential operator $p(\partial_x)$ defined in (1.13), the Harish-Chandra integral $\mathcal{H}$ was defined in (1.11), and $\sigma(A(x))$ denotes the spectrum of the Hermitian matrix

$$A(x) = \sum_{j=1}^{m} x_j A_j, \quad x \in \mathbb{R}^m. \tag{1.19}$$

Note that x lies in the primal space $\mathbb{R}^m$, whereas Λ lies in $\mathbb{R}^n$ because X is in the dual space $\mathbb{P}(n)$. There is no a priori relation between the dimensions m and n.

In order to get a feel for G, consider the example (1.12). Now $p(\partial_x)$ becomes the second-order elliptic operator

$$p(\partial_x) = \sum_{j=1}^{m} \left(\partial_{x_j} + c_j\right)^2 = \frac{1}{2} \triangle_x + c \cdot \nabla_x + \frac{1}{2}|c|^2, \tag{1.20}$$

and $G(x,x';\beta)$ is a Gaussian kernel which we compute in Section 4. This example is instructive because it tells us that $G(x,x';\beta)$ should always be seen as a generalized heat kernel (with a drift and killing) in the primal space $\mathbb{R}^m$ with the inverse temperature $\beta > 0$ serving as 'time'.

The expression for the partition can be simplified by interchanging limits in (1.16) and (1.17) to obtain

$$Z_\beta = \int_{\mathbb{R}^m} G(-x;\beta) W_{\mathrm{ref}}(\sigma(A(x)))\, dx, \quad W_{\mathrm{ref}}(\mu) = \int_{\mathbb{R}^n} \mathcal{H}(\mu;\lambda)\rho_{\mathrm{ref}}(\lambda)\, d\lambda, \quad \mu \in \mathbb{R}^n. \tag{1.21}$$

Equation (1.21) reflects a soft form of convex duality, since the partition function is computed by summing probabilities for the primal SDP (1.2). It should be contrasted with the expression for Z_β in (1.15), where the partition function sums probabilities on the dual space $\mathbb{P}(n)$.

This duality is reflected in the integrands too. Equation (1.15) carries the explicit Gibbs weight $e^{-\beta E(X)}\sqrt{\det g}\, DX$ for $X \in \mathbb{P}(n)$. However, in equation (1.21), the heat kernel G is dependent on the soft potential and thus the constraints, while W_{ref} splits further into a sum over the reference measure $\rho_{\mathrm{ref}}(\Lambda)$, which depends only on the choice of Riemannian metric on $\mathbb{P}(n)$, and the Harish-Chandra integral $\mathcal{H}$ which depends on the constraints only through the sum $A(x) = \sum_j x_j A_j$.

These formulas are first established formally using generating functions and then interpreted rigorously using heat kernels. In a sense, they reflect nothing more than the classical Fourier analysis trick of trading multiplication for differentiation with the Harish-Chandra integral playing the natural role of generating function. The reader with some experience with such models will recognize from (1.21) that the asymptotic analysis of Z_β in the limits $\beta \to \infty$ and $n \to \infty$ may now be approached rigorously using methods from random matrix theory. There appears to have been no prior study of this nature, despite the fundamental importance of SDP.

2. Metrics on $\mathbb{P}(n)$

We introduce the trace metric and the Bures-Wasserstein metric and discuss some of their properties.

2.1. The trace metric on $\mathbb{P}(n)$.

The material in this section originates in [16]. Good pedagogical accounts, from different mathematical perspectives, are [12, Ch XII] and [4, Ch.6].

The trace metric is the metric on $\mathbb{P}(n)$ obtained by pushing forward the Frobenius norm on $\mathbb{H}(n)$ to $\mathbb{P}(n)$ under the map

$$\mathbb{H}(n) \to \mathbb{P}(n), \quad H \mapsto \exp H = \sum_{k=0}^{\infty} \frac{1}{k!} H^k. \tag{2.1}$$

At any $P \in \mathbb{P}(n)$ the tangent space $T_P\mathbb{P}(n)$ is isomorphic to $\mathbb{H}(n)$ and for each $S \in \mathbb{H}(n)$ we define the inner-product

$$\langle S, S\rangle_P = \mathrm{Tr}\left(P^{-1}SP^{-1}S\right). \tag{2.2}$$

Note that $S = S^*$ and $P = P^*$.

This metric has several interesting properties. We list some of these.

(1) For each $M \in GL(n)$ the transformation $[M] : \mathbb{P}(n) \to \mathbb{P}(n)$ defined by $P \mapsto MPM^*$ is an isometry.

(2) The geodesic between $A, B \in \mathbb{P}(n)$ is given by [4, Thm.6.1.6]

$$P(t) = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t A^{1/2}, \quad t \in [0,1]. \tag{2.3}$$

Recall that each $P \in \mathbb{P}(n)$ has a unique square root in $\mathbb{P}(n)$.

(3) The Riemannian manifold $(\mathbb{P}(n), \langle\cdot,\cdot\rangle)$ has negative curvature.

Remark 3. This note focuses on explicit computations with this metric. However, the natural appearance of HC integrals suggests a deeper structure by analogy.

The simplest example of a space with negative curvature is the upper half plane, $\mathbb{C}_+$, equipped with the Poincaré metric. By the uniformization theorem, every compact Riemann surface is characterized by a fundamental domain in $\mathbb{C}_+$, and the automorphism group of the Riemann surface is represented as a subgroup of the conformal maps of $\mathbb{C}_+$. The manifold $(\mathbb{P}(n), \langle\cdot,\cdot\rangle)$ is a symmetric space with negative curvature that should be seen in a similar light. It was first studied by Cartan in analogy with the upper half plane [16] and the fact that the transformation $[M] : \mathbb{P}(n) \to \mathbb{P}(n)$ is an isometry is representation theoretic. It says that $M \mapsto [M]$ is a representation of $GL(n)$ in the group of isometries of $\mathbb{P}(n)$ [12, Thm 1.1].

Recall that the feasible set $\mathcal{F}$ is invariant under this $GL(n)$ action. The main structural conjecture then is that for each strictly feasible SDP, the constraints (1.8) determine a fundamental domain in $\mathbb{P}(n)$ and that the automorphism group for this domain may be explicitly determined as a subgroup of the isometries of $\mathbb{P}(n)$. In this manner, we expect that each SDP should be faithfully associated to a geometry. This is conjectural and it may be necessary to modify these statements to get things completely right. Several formulas in [1, 2, 3] suggest such a structure (see especially [3, p.298]) and the explicit construction of the automorphism group could systematize ad hoc counterexamples in LP.

A very interesting set of results that discusses the construction of self-concordant barriers as solutions to Monge-Ampere equations is discussed by Hildebrand [9]. This work seems important and provides the most direct contact between the ideas documented here and optimization theory.

2.2. **The Bures-Wasserstein metric on $\mathbb{P}(n)$.** The Bures-Wasserstein metric on $\mathbb{P}(n)$ has been studied more recently; a good review is [5]. This metric is obtained from the Frobenius norm on $GL(n)$ through the Riemannian submersion $\Pi : GL(n) \to \mathbb{P}(n)$, $M \mapsto MM^*$. A computation of the vertical and horizontal spaces for this submersion yields the following formula for the metric. The tangent space $T_P\mathbb{P}(n)$ is isomorphic to $\mathbb{H}(n)$ and for each $S \in T_P\mathbb{P}(n)$

$$\langle\langle S, S\rangle\rangle_P = \mathrm{Tr}(H^*H), \tag{2.4}$$

where $H = H^* \in \mathbb{H}(n)$ is the unique solution to the Lyapunov equation

$$S = PH + HP. \tag{2.5}$$

There are two useful semi-explicit formulas for the metric.

(1) Suppose we choose a basis on $\mathbb{R}^n$ in which $P = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$. In these coordinates

$$\langle\langle S, S\rangle\rangle_P = \sum_{1\le j,k\le n} \frac{|S_{jk}|^2}{\lambda_j + \lambda_k}. \tag{2.6}$$

(2) Given S and the invertible matrix P, the solution to the Lyapunov equation (2.5) is given by the formula

$$H = \int_0^\infty e^{-tP} S e^{-tP}\, dt. \tag{2.7}$$

Indeed, if we make the above ansatz for H we may compute

$$PH = \int_0^\infty P e^{-tP} S e^{-tP}\, dt = -\int_0^\infty \left(\frac{d}{dt} e^{-tP}\right) S e^{-tP}\, dt, \tag{2.8}$$

and similarly

$$HP = \int_0^\infty P e^{-tP} S e^{-tP}\, dt = -\int_0^\infty e^{-tP} S \left(\frac{d}{dt} e^{-tP}\right) dt. \tag{2.9}$$

We combine (2.8) and (2.9) to obtain (2.7).

Fix $A, B \in \mathbb{P}(n)$. The Bures-Wasserstein geodesic between A and B in $\mathbb{P}(n)$ may be parametrized by

$$M(t) = \left((1-t)\mathrm{Id} + t(A^{-1}\#B)\right) A^{1/2}, \quad P(t) = M(t)M(t)^*, \quad 0 \le t \le 1, \tag{2.10}$$

where the *geometric mean* of two matrices R and S in $\mathbb{P}(n)$ is defined by

$$R\#S = R^{-1/2} \left(R^{1/2} S R^{1/2}\right)^{1/2} R^{-1/2}. \tag{2.11}$$

The term $M(t)$ in equation (2.10) is one of many possible lifts of $P(t)$ into a straight line in $GL(n)$. For this reason, while the parametrization downstairs is satisfactory, the parametrization upstairs offers room for further exploration.

Comparing equations (2.10)–(2.11) with equation (2.3) we see that the geometric mean is the midpoint of the geodesic (in the trace metric) between R and S. In particular, $R\#S = S\#R$. Several other properties of the geometric mean are known [4, Ch.4].

Finally, the sectional curvatures of $\mathbb{P}(n)$ with the Bures-Wasserstein metric have been computed [18]. It turns out that $(\mathbb{P}(n), \langle\langle, \rangle\rangle)$ is a space with *positive* curvature.

Remark 4. The duality between the trace metric and the Bures-Wasserstein metric should be understood better, since these properties are strongly reminescent of elliptic and hyperbolic metrics in the Poincaré disk model, where a precise duality is known [15, Ch.1]. The heart of the matter is a better understanding of the geometric mean. As explained below, this concept appears in both mass transportation theory and operator theory, so there is a lot more to (2.11) than meets the eye.

Remark 5. The notion of matrix mean may be generalized, just as positive numbers admit several means – arithmetic, geometric and harmonic means being the most common choices. Several properties of such matrix means are known [4, Ch.4]. Kubo and Ando provide a complete axiomatic classification of means that reduces to Loewner's theorem on matrix monotone functions [11]. It is of interest to use this theorem to compute partition functions that generalize the two examples studied here.

Remark 6. The Bures-Wasserstein metric on $\mathbb{P}(n)$may also be obtained through mass transportation theory (in particular, the link between mass transportation and

Riemannian submersion is from [17, Sec. 4]). Every $P \in \mathbb{P}(n)$ may be identified with a Gaussian measure μ_P on $\mathbb{R}^n$ via

$$\mu_P(x)\,dx = \frac{1}{\sqrt{\det(2\pi P)}} e^{-\frac{1}{2}x^* P^{-1} x}\,dx, \quad x \in \mathbb{R}^n. \tag{2.12}$$

Equivalently, P is the covariance for a Gaussian random vector $Z \in \mathbb{R}^n$, with

$$P_{ij} = \mathbb{E}(Z_i Z_j), \quad 1 \le i,j \le n. \tag{2.13}$$

This identification is important since it says that a metric on $\mathbb{P}(n)$ is always equivalent to a metric on the space of Gaussian measures on $\mathbb{R}^n$. In particular, the Bures-Wasserstein metric on $\mathbb{P}(n)$ – which is defined here through Riemannian submersion, a geometric construct – is the restriction of the Wasserstein metric on probability measures on $\mathbb{R}^n$ to Gaussian measures that are absolutely continuous with respect to Lebesgue measure – a probabilistic construct.

Remark 7. Bures, on the other hand, was studying C^* algebras [7]. I don't understand this as well, but it seems to go roughly as follows. Each $P \in \mathbb{P}(n)$ is a generalized density matrix (generalized because a density matrix must have unit trace). The Bures metric is obtained by 'purifying' a density matrix into a pure state. In mathematical terms, this seems to involve lifting a given $P \in \mathbb{P}(n)$ into one of its square roots in $GL(n)$ which leads back to Riemannian submersion.

3. Volume forms on $\mathbb{P}(n)$

We now compute the volume forms for the trace metric and the Bures-Wasserstein metric at a point $X \in \mathbb{P}(n)$. These computations involve two steps. First, we use basic differential geometry to compute the volume form for the two metrics from Section 2. Second, we diagonalize $X = U\Lambda U^*$ and use Weyl's formula to simplify the volume form.

3.1. Some book-keeping.

Given a Riemannian manifold $(\mathcal{M}^d, g)$ of dimension d with metric g the associated volume form is

$$\sqrt{\det g}\,dx = \sqrt{\det\left(g_{\alpha\beta}\right)_{1\le\alpha,\beta\le d}}\,dx_1 \dots dx_d. \tag{3.1}$$

Since our manifolds are spaces of matrices, when using formulas from differential geometry such as the one above, we must replace the indices α and β with pairs of indices (ij). We will be explicit about such formulas listing the entries of a matrix $X \in \mathbb{P}(n;\mathbb{R})$ as a real vector of length $n(n+1)/2$ as follows

$$X = \begin{pmatrix} X_{11} \\ X_{22} \\ \vdots \\ X_{nn} \\ X_{12} \\ X_{13} \\ \vdots \\ X_{n-1,n} \end{pmatrix}. \tag{3.2}$$

The n diagonal entries are listed first, followed by the $n(n-1)/2$ off-diagonal entries in lexicographical order. Similarly, a matrix $X \in \mathbb{P}(n;\mathbb{C})$ defines a real vector of length n^2

$$X = \begin{pmatrix} X_{11} \\ X_{22} \\ \vdots \\ X_{nn} \\ \operatorname{Re} X_{12} \\ \operatorname{Im} X_{12} \\ \vdots \\ \operatorname{Re} X_{n-1,n} \\ \operatorname{Im} X_{n-1,n} \end{pmatrix}. \tag{3.3}$$

The heart of the matter is that for both the trace and Bures-Wasserstein metric, the metric is diagonal in these coordinates. Thus, its determinant can be calculated immediately, yielding the volume form.

3.2. **The trace metric on $\mathbb{P}(n;\mathbb{R})$.** Fix $X \in \mathbb{P}(n;\mathbb{R})$ and assume at first that $X = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. The volume form in the general case can be reduced to this situation. Assume $S \in T_X\mathbb{P}(n;\mathbb{R}) \cong \mathbb{H}(n;\mathbb{R})$. We use the definition (2.2) to obtain

$$\begin{aligned} \langle S, S\rangle_X &= \operatorname{Tr}(X^{-1}SX^{-1}S) = (X^{-1})_{ij}S_{jk}X^{-1}_{kl}S_{li} \\ &= \frac{\delta_{ij}}{\lambda_i}S_{jk}\frac{\delta_{kl}}{\lambda_k}S_{li} = \frac{1}{\lambda_i\lambda_k}S_{ik}^2 \\ &= \frac{S_{11}^2}{\lambda_1} + \ldots + \frac{S_{nn}^2}{\lambda_n} + \frac{2S_{12}^2}{\lambda_1\lambda_2} + \ldots + \frac{2S_{n-1,n}^2}{\lambda_{n-1}\lambda_n}. \end{aligned} \tag{3.4}$$

The summation convention for repeated indices has been adopted in the first two lines and the symmetry $S_{ij} = S_{ji}$ has been used. The n diagonal terms are clear; note that the $n(n-1)/2$ off-diagonal terms of S appear twice by symmetry and that we restrict the index set S_{jk} to $j < k$.

In these coordinates, the metric $\langle \cdot, \cdot \rangle_X$ is given by the $n(n+1)/2 \times n(n+1)/2$ matrix

$$g_{\alpha\beta} = \begin{pmatrix} \frac{1}{\lambda_1^2} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\lambda_n^2} & & & \\ & & & \frac{2}{\lambda_1\lambda_2} & & \\ & & & & \ddots & \\ & & & & & \frac{2}{\lambda_{n-1}\lambda_n}. \end{pmatrix} \tag{3.5}$$

Thus, the determinant of the metric is

$$\det g = 2^{n(n-1)/2}\left(\frac{1}{\lambda_1 \cdots \lambda_n}\right)^{n+1} = 2^{n(n-1)/2}\left(\det X\right)^{-(n+1)}. \tag{3.6}$$

Finally, we take the square-root of this expression to obtain the volume form for $\mathbb{P}(n;\mathbb{R})$ with the trace metric

$$\sqrt{\det g}\, DX := 2^{n(n-1)/4}\left(\det X\right)^{-(n+1)/2}\, DX, \quad X \in \mathbb{P}(n;\mathbb{R}) \tag{3.7}$$

where the notation DX means

$$DX := dX_{11} \cdots dX_{nn}\, dX_{12}\, dX_{13} \cdots dX_{n-1,n}. \tag{3.8}$$

3.3. **The trace metric on** $\mathbb{P}(n;\mathbb{C})$**.** The computation of the volume form on $\mathbb{P}(n;\mathbb{C})$ is very similar to the above calculation. The main change now is that we separate the real and imaginary components of each off-diagonal term S_{jk}, $j<k$. We find that

$$\langle S,S\rangle_X = \frac{S_{11}^2}{\lambda_1} + \ldots + \frac{S_{nn}^2}{\lambda_n} + \frac{2|S|_{12}^2}{\lambda_1\lambda_2} + \ldots + \frac{2|S|_{n-1,n}^2}{\lambda_{n-1}\lambda_n}, \tag{3.9}$$

with $|S_{jk}|^2 = (\operatorname{Re} S_{jk})^2 + (\operatorname{Im} S_{jk})^2$ for the off-diagonal terms. (The diagonal terms are always real). The matrix for the metric $\langle\cdot,\cdot\rangle_P$ is now an $n^2\times n^2$ matrix similar to that of equation (3.5), with the difference that each of the off-diagonal terms appears twice

$$g_{\alpha\beta} = \begin{pmatrix} \frac{1}{\lambda_1^2} & & & & & & \\ & \ddots & & & & & \\ & & \frac{1}{\lambda_n^2} & & & & \\ & & & \frac{2}{\lambda_1\lambda_2} & & & \\ & & & & \frac{2}{\lambda_1\lambda_2} & & \\ & & & & & \ddots & \\ & & & & & & \frac{2}{\lambda_{n-1}\lambda_n}. \end{pmatrix} \tag{3.10}$$

The determinant and volume form may be computed as before. We find

$$\sqrt{\det g}\, DX := 2^{n(n-1)/2} \left(\det X\right)^{-n}\, DX, \quad X\in\mathbb{P}(n;\mathbb{C}) \tag{3.11}$$

where the notation DX means

$$DX := dX_{11}\,\cdots dX_{nn}\, d\operatorname{Re} X_{12}\, d\operatorname{Im} X_{12}\, d\operatorname{Re} X_{13}\,\cdots d\operatorname{Im} X_{n-1,n}. \tag{3.12}$$

3.4. **The Bures-Wasserstein metric on** $\mathbb{P}(n;\mathbb{R})$ **and** $\mathbb{P}(n;\mathbb{C})$**.** We again assume that $X=\operatorname{diag}(\lambda_1,\ldots,\lambda_n)$ since the general computation may be reduced to this case. We expand (2.6) for $S\in\mathbb{H}(n;\mathbb{R})$ to obtain

$$\langle\langle S,S\rangle\rangle_X = \frac{S_{11}^2}{2\lambda_1} + \ldots + \frac{S_{nn}^2}{2\lambda_n} + \frac{2S_{12}^2}{\lambda_1+\lambda_2} + \frac{2S_{13}^2}{\lambda_1+\lambda_3} + \ldots + \frac{2S_{n-1,n}^2}{\lambda_{n-1}+\lambda_n} \tag{3.13}$$

As in equation (3.4), the off-diagonal terms appear twice. We see again that the metric is diagonal in these coordinates and we obtain the volume form

$$\sqrt{\det(g_{BW})}\, DX = 2^{n(n-1)/4}\frac{1}{\det(2X)^{1/2}}\left(\prod_{1\le j<k\le n}\frac{1}{\lambda_j+\lambda_k}\right)^{1/2} DX, \quad X\in\mathbb{P}(n;\mathbb{R}), \tag{3.14}$$

where DX is defined by equation (3.8). This formula is not as transparent as the comparable formula (3.7) because of the explicit dependence on the eigenvalues of X. However, it can be simplified using Weyl's formula as discussed below.

The extension of this calculation to $\mathbb{P}(n;\mathbb{C})$ goes as follows. Each off-diagonal term in (3.13) must now be split into real and imaginary parts

$$\frac{2|S_{jk}|^2}{\lambda_j+\lambda_k} = \frac{2(\operatorname{Re} S_{jk})^2}{\lambda_j+\lambda_k} + \frac{2(\operatorname{Im} S_{jk})^2}{\lambda_j+\lambda_k}, \quad j<k,$$

and it is clear that the metric is still diagonal in these coordinates. Thus,

$$\sqrt{\det(g_{BW})}\,DX = 2^{n(n-1)/2}\frac{1}{\det(2X)^{1/2}}\left(\prod_{1\le j<k\le n}\frac{1}{\lambda_j+\lambda_k}\right)DX, \quad X\in\mathbb{P}(n;\mathbb{C}). \tag{3.15}$$

3.5. **Diagonalization and Weyl's formula.** Let $X = U\Lambda U^*$ denote the diagonalization of X (strictly speaking, we should say a diagonalization, but when the eigenvalues are distinct, the map $X\mapsto(\Lambda,U)$ is locally smooth). The Jacobian for this transformation is provided by Weyl's integration formula. To state this formula, let

$$V(\Lambda) = \begin{vmatrix} 1 & \dots & 1 \\ \lambda_1 & & \lambda_n \\ \lambda_1^2 & & \lambda_n^2 \\ \vdots & & \vdots \\ \lambda_1^{n-1} & & \lambda_n^{n-1} \end{vmatrix} = \prod_{j<k}(\lambda_k-\lambda_j) \tag{3.16}$$

denote the Vandermonde determinant associated to Λ. Then we have [14, Thm 15] the following formula for $\mathbb{P}(n;\mathbb{R})$

$$2^{n(n-1)/2}DX = |V(\Lambda)|d\Lambda dU, \tag{3.17}$$

denotes (unnormalized) Haar measure on $O(n)$. The analogous expression for $\mathbb{P}(n;\mathbb{C})$ is

$$2^{n(n-1)/2}DX = |V(\Lambda)|^2 d\Lambda dU \tag{3.18}$$

where dU now denotes (unnormalized) Haar measure on $U(n)/\mathbb{T}^n$. [4].

The expressions (3.7) and (3.12) for the volume form on $\mathbb{P}(n)$ equipped with the trace metric now take the simple form

$$\sqrt{\det g}DX = \begin{cases} \det(\Lambda)^{-\frac{n+1}{2}}|V(\Lambda)|\,d\Lambda DU, & X\in\mathbb{P}(n;\mathbb{R}) \\ \det(\Lambda)^{-n}|V(\Lambda)|^2\,d\Lambda DU, & X\in\mathbb{P}(n;\mathbb{C}) \end{cases} \tag{3.19}$$

The analogous expressions (3.14) and (3.15) for the volume form on $\mathbb{P}(n;\mathbb{R})$ equipped with the Bures-Wasserstein metric may be abbreviated by recognizing that

$$\prod_{1\le j<k\le n}\frac{1}{\lambda_j+\lambda_k} = \frac{V(\Lambda)}{V(\Lambda^2)}. \tag{3.20}$$

We then find that the volume form for the Bures-Wasserstein metric on $\mathbb{P}(n;\mathbb{R})$ is

$$\sqrt{\det g_{BW}}DX = \det(2\Lambda)^{-\frac{1}{2}}\frac{|V(\Lambda)|^{3/2}}{|V(\Lambda^2)|^{1/2}}\,d\Lambda DU, \quad X\in\mathbb{P}(n;\mathbb{R}). \tag{3.21}$$

Similarly the volume form on $\mathbb{P}(n;\mathbb{C})$ is

$$\sqrt{\det g_{BW}}DX = \det(2\Lambda)^{-\frac{1}{2}}\frac{V(\Lambda)^3}{V(\Lambda^2)}\,d\Lambda DU, \quad X\in\mathbb{P}(n;\mathbb{C}). \tag{3.22}$$

These formulas allows us to separate the entropic terms in the Gibbs measure from the energetic terms corresponding to the constraints. We now turn to these terms.

[4]Need to compute and fix normalization constants both here and [14]

4. Soft constraints

The calculation of the previous sections suffice to establish equation (1.16). Let us explain how to derive (1.17) in two steps. First, we assume that there is only one constraint, written $Ax + c = 0$, expand the integrand $e^{-\beta E(X)}$ in a Taylor series, and use the HCIZ integral to simplify this expansion. Once this case is understood, we consider m constraints and organize the terms to obtain (1.17).

4.1. **The case $m = 1$.** We find it more convenient to work with lower case and use the following notation for diagonalization:

$$X = u\lambda u^*, \quad A = vav*, \tag{4.1}$$

where $a = \operatorname{diag}(a_1, \dots, a_n)$ denote the eigenvalues of A. We further abbreviate this expression to

$$X = \mathrm{Ad}_u\lambda, \quad A = \mathrm{Ad}_v a, \tag{4.2}$$

where Ad is the adjoint action for $U(n)$. (This is the form in which these formulae generalize to other groups, so it is good to get used to this idea, even if it is just a change in notation for $U(n)$). The LHS of (1.11) may be rewritten as

$$\int_{U(n)} e^{(\mathrm{Ad}_u a, b)}\, du = \mathcal{H}(a, b). \tag{4.3}$$

We will need to evaluate sums of terms of the form

$$\int_{U(n)} (\mathrm{Ad}_u\lambda, a)^p \; du \tag{4.4}$$

for all positive integers p. The notation here is $(X, Y) = \mathrm{Tr}(XY)$ for $X, Y \in \mathbb{H}(n)$.

The main point is that the Harish-Chandra integral is a good generating function. We compute moments as derivatives of generating functions, like so

$$\int_{U(n)} (\mathrm{Ad}_u\lambda, a)^p \; du = \int_{U(n)} \partial_x^p e^{x(\mathrm{Ad}_u\lambda, a)}\Big|_{x=0} du = \partial_x^p \mathcal{H}(\lambda, ax)|_{x=0}\,. \tag{4.5}$$

(The order of operations here is first differentiate in x, then set $x = 0$.) Aside from the fact that we integrate over a group, this should be familiar.

The rest of the calculation involves an expansion in Taylor series, with a repeated use of (4.5) to swap powers $(\mathrm{Ad}_u\lambda, a)^p$ for derivatives ∂_x^p, which are then evaluated at $x = 0$. Let us begin with the orbital part of the partition function

$$J_\beta := \int_{U(n)} e^{-\beta E(X)}\, du = \sum_{k=0}^{\infty} \frac{(-\beta)^k}{k!} \int_{U(n)} E(X)^k\, du, \tag{4.6}$$

with $X = \mathrm{Ad}_u\lambda$. We will expand each integrand $E(X)^k$ once more, using the assumption that E is defined by a polynomial p through equation (1.14), which in the case $m = 1$ is simply

$$E(X) = p((\mathrm{Ad}_u\lambda, a)). \tag{4.7}$$

To this end, let us write this polynomial as $p(\theta) = \alpha_0 + \alpha_1\theta + \ldots + \alpha_s\theta^s$, where s is the degree of p. Then by the multinomial expansion

$$p(\theta)^k = \sum_{|l|\le k} \binom{k}{l_0 \dots l_s} \prod_{j=0}^{s} (\alpha_j\theta^j)^{l_j}, \quad l = (l_0, \dots, l_s). \tag{4.8}$$

Set $(\mathrm{Ad}_u\lambda, a) = \theta(X) = \theta$ for brevity, so that

$$E(X) = p((\mathrm{Ad}_u\lambda, a)) = p(\theta), \quad \text{and} \quad E(X)^k = p(\theta)^k. \tag{4.9}$$

Now integrate over $U(n)$ using (4.5), (4.8) and (4.9) to find

$$\begin{aligned}
\int_{U(n)} E(X)^k \, dU &= \sum_{|l|\le k} \binom{k}{l_0 \ldots, l_s} \left(\prod_{j=0}^{s} \alpha_j^{l_j}\right) \int_{U(n)} \theta(X)^{\sum_{r=0}^{s} r l_r} \, du \\
&= \sum_{|l|\le k} \binom{k}{l_0 \ldots l_s} \left(\prod_{j=0}^{s} \alpha_j^{l_j}\right) (\partial_x)^{\sum_{r=0}^{s} r l_r} \, \mathcal{H}(\lambda, ax)\Big|_{x=0} \\
&= \sum_{|l|\le k} \binom{k}{l_0 \ldots l_s} \left(\prod_{j=0}^{s} \alpha_j \partial_x^j\right)^{l_j} \mathcal{H}(\lambda, ax)|_{x=0} \\
&= p(\partial_x)^k \, \mathcal{H}(\lambda, ax)|_{x=0} \, .
\end{aligned} \tag{4.10}$$

With this identity in hand, we return to (4.6) and obtain

$$J_\beta = \int_{U(n)} e^{-\beta E(X)} \, du \tag{4.11}$$

$$= \sum_{k=0}^{\infty} \frac{(-\beta)^k}{k!} p(\partial_x)^k \, \mathcal{H}(\lambda, ax)|_{x=0} = \left(e^{-\beta p(\partial_x)} \mathcal{H}(\lambda, ax) \right)\Big|_{x=0} . \tag{4.12}$$

4.2. **Arbitrary** m**.** We modify the above calculations as follows. First, the notation. Replace (4.2) with

$$X = \mathrm{Ad}_u \lambda, \quad A_j = \mathrm{Ad}_{v_j} a_j, \quad 1 \le j \le m. \tag{4.13}$$

For any $x \in \mathbb{R}^m$, let

$$A(x) = \sum_{j=1}^{m} x_j A_j := \mathrm{Ad}_{v(x)} a(x), \tag{4.14}$$

where $a(x)$ and $v(x)$ denote the eigenvalues and eigenvectors of $A(x)$. The generating function for arbitrary m is the Harish-Chandra integral

$$\mathcal{H}(a(x), \lambda) = \int_{U(n)} e^{(\mathrm{Ad}_{uv(x)}\lambda, a(x))} \, du = \int_{U(n)} e^{(\mathrm{Ad}_u \lambda, a(x))} \, du. \tag{4.15}$$

By definition (4.14), we also have

$$\begin{aligned}
\mathcal{H}(a(x), \lambda) &= \int_{U(n)} e^{\sum_{j=1}^{m} x_j (\mathrm{Ad}_{uv_j}\lambda, a_j)} \, du \\
&= \int_{U(n)} \prod_{j=1}^{m} e^{x_j (\mathrm{Ad}_{uv_j}\lambda, a_j)} \, du = \int_{U(n)} \prod_{j=1}^{m} e^{x_j \mathrm{Tr}(X A_j)} \, du.
\end{aligned} \tag{4.16}$$

The last equation is added so that one sees the constraints explicitly and so that it is clear that we may differentiate the RHS with respect to each coordinate x_r and then evaluate at 0 to obtain a basic moment identity analogous to (4.5). For integers $l_1, \ldots, l_m$ we have

$$\int_{U(n)} \prod_{j=1}^{m} (\mathrm{Ad}_{uv_j}\lambda, a_j)^{l_j} du = \int_{U(n)} \left(\partial_{x_1}^{l_1} \cdots \partial_{x_m}^{l_m}\right) \prod_{j=1}^{m} e^{x_j (\mathrm{Ad}_{uv_j}\lambda, a_j)} \, du \Bigg|_{x=0} , \tag{4.17}$$

(if the typography seems confusing, let me stress again that first we differentiate, then we set $x_1 = x_2 = \ldots = x_m = 0$.)

On the other hand, we combine identities (4.15) and (4.16) to see that this moment identity must be equivalent to

$$\int_{U(n)} \prod_{j=1}^{m} (\mathrm{Ad}_{uv_j}\lambda, a_j)^{l_j}\, du = \left(\partial_{x_1}^{l_1} \cdots \partial_{x_m}^{l_m}\right) \mathcal{H}(a(x),\lambda)\big|_{x=0}. \tag{4.18}$$

This expression may be more transparent when written in the form

$$\int_{U(n)} \prod_{j=1}^{m} (\mathrm{Tr}(A_j X))^{l_j}\, du = \left(\partial_{x_1}^{l_1} \cdots \partial_{x_m}^{l_m}\right) \mathcal{H}(a(x),\lambda)\big|_{x=0}. \tag{4.19}$$

The remainder of the calculation is a somewhat more tedious version of the double expansion we used to get from (4.6) to (4.12).

First, let us revisit the energy and expressions (4.6) and (4.9). We assume that

$$E(X) = \sum_{j=1}^{m} p_j(\mathrm{Tr}(A_j X)) := \sum_{j=1}^{m} p_j(\theta_j), \quad \theta_j = \mathrm{Tr}(A_j X). \tag{4.20}$$

Now the series expansion (4.6) remains the same and we must evaluate the expression

$$\int_{U(n)} E(X)^k\, du = \int_{U(n)} \left(\sum_{j=1}^{m} p_j(\theta_j)\right)^k du. \tag{4.21}$$

This adds one layer of complexity beyond (4.10), but the basic structure is the same. First, we expand the sum within the integrand using the multinomial expansion

$$\left(\sum_{j=1}^{m} p_j(\theta_j)\right)^k = \sum_{|l|\le k} \binom{k}{l_1 \ldots l_m} p_j(\theta_j)^{l_j}. \tag{4.22}$$

Next expand each polynomial $p_j(\theta)$ in powers of θ_j

$$p_j(\theta_j) = \alpha_{j,0} + \alpha_{j,1}\theta_j + \ldots + \alpha_{j,s_j}\theta_j^{s_j}. \tag{4.23}$$

This provides a further multinomial expansion for each power $p_j(\theta_j)^{l_j}$ in (4.22)

$$p_j(\theta_j)^{l_j} = \sum_{|r|\le l_j} \binom{l_j}{r_0 \ldots r_{s_j}} \prod_{q=0}^{s_j} \left(\alpha_{j,q}\theta_j^q\right)^{r_q} \tag{4.24}$$

When equation (4.24) is combined with the expansions above, we see that the integrand in (4.20), consists of a finite sum of terms of the form

$$\int_{U(n)} \theta_1^{q_1} \cdots \theta_m^{q_m}\, du = (\partial_{x_1}^{q_1} \ldots \partial_{x_m}^{q_m}) \mathcal{H}(a(x),\lambda)\big|_{x=0}, \tag{4.25}$$

where the RHS is obtained from equation (4.17) and the definition $\theta_j = \mathrm{Tr}(A_j X)$. Thus, when integrating such monomials over $U(n)$, we may simply trade moments in θ_j for derivatives in ∂_{x_j} of the HC integral. This allows us to undo the sum as in (4.10) to obtain the identity

$$\int_{U(n)} \prod_{j=1}^{m} p_j(\theta_j)^{l_j}\, du = (p_1(\partial_{x_1})^{l_1} \ldots p_m(\partial_{x_m})^{l_m}) \mathcal{H}(a(x),\lambda)\big|_{x=0}. \tag{4.26}$$

When combined with (4.22) this yields the following identity analogous to (4.10)

$$\int_{U(n)} E(X)^k \, du = \left(\sum_{j=1}^m p_j(\partial_{x_j}) \right)^k \mathcal{H}(a(x), \lambda) \Bigg|_{x=0} . \tag{4.27}$$

Finally, we return to (4.6) to obtain the fundamental identity

$$\begin{aligned} Y_\beta &= \int_{U(n)} e^{-\beta E(X)} \, du \qquad (4.28) \\ &= \sum_{k=0}^\infty \frac{(-\beta)^k}{k!} \left(\sum_{j=1}^m p_j(\partial_{x_j}) \right)^k \mathcal{H}(a(x), \lambda)\big|_{x=0} = \left(e^{-\beta p(\partial_x)} \mathcal{H}(a(x), \lambda) \right)\Big|_{x=0} . \end{aligned}$$

where we recall the definition of the differential operator

$$p(\partial_x) = \sum_{j=1}^m p_j(\partial_{x_j}).$$

In order to complete the proof of (1.17) it is only necessary to provide a rigorous interpretation for the formal expression $e^{-\beta p(\partial_x)}$. This is why we introduce the fundamental solution in (1.18), as well as the example (1.20). A direct computation yields the heat kernel for example (1.20)

$$G(x, x'; \beta) = \left(\frac{\beta}{2\pi} \right)^{\frac{m}{2}} \prod_{j=1}^m \exp \left(-\frac{\beta}{2} (x_j - x_j')^2 - c_j (x_j - x_j') \right) . \tag{4.29}$$

It is now clear that the formal manipulations in this section may be rigorously justified. This example also reveals that we must assume the soft potential $E(X)$ and the related differential operator $p(\partial_x)$ must be chosen so that they lead to tractable heat kernels. In analogy with random matrix theory, it is prudent to first obtain a complete description for this example, before considering a general theory of τ-functions and universality for soft constraints.

References

[1] D. A. Bayer and J. C. Lagarias, *The nonlinear geometry of linear programming. I. Affine and projective scaling trajectories*, Trans. Amer. Math. Soc., 314 (1989), pp. 499–526.

[2] ———, *The nonlinear geometry of linear programming. II. Legendre transform coordinates and central trajectories*, Trans. Amer. Math. Soc., 314 (1989), pp. 527–581.

[3] ———, *Karmarkar's linear programming algorithm and Newton's method*, Math. Programming, 50 (1991), pp. 291–330.

[4] R. Bhatia, *Positive definite matrices*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2007.

[5] R. Bhatia, T. Jain, and Y. Lim, *On the Bures–Wasserstein distance between positive definite matrices*, Expositiones Mathematicae, (2018).

[6] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, Cambridge, 2004.

[7] D. Bures, *An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite W^*-algebras*, Transactions of the American Mathematical Society, 135 (1969), pp. 199–212.

[8] L. Faybusovich, *A Hamiltonian structure for generalized affine-scaling vector fields*, J. Nonlinear Sci., 5 (1995), pp. 11–28.

[9] R. Hildebrand, *Conic optimization: affine geometry of self-concordant barriers and copositive cones*, PhD thesis, (Habilitation), Université Grenoble Alpes, 2017.

[10] N. Karmarkar, *A new polynomial-time algorithm for linear programming*, Combinatorica, 4 (1984), pp. 373–395.
[11] F. Kubo and T. Ando, *Means of positive linear operators*, Math. Ann., 246 (1979/80), pp. 205–224.
[12] S. Lang, *Fundamentals of differential geometry*, vol. 191, Springer Science & Business Media, 2012.
[13] L. Lovász, *Semidefinite programs and combinatorial optimization*, in Recent advances in algorithms and combinatorics, vol. 11 of CMS Books Math./Ouvrages Math. SMC, Springer, New York, 2003, pp. 137–194.
[14] G. Menon and T. Trogdon, *Random matrix theory and numerical linear algebra*, March 2020.
[15] J. Moser and E. J. Zehnder, *Notes on dynamical systems*, vol. 12 of Courant Lecture Notes in Mathematics, New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2005.
[16] G. D. Mostow, *Some new decomposition theorems for semi-simple groups*, Mem. Amer. Math. Soc., 14 (1955), pp. 31–54.
[17] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.
[18] A. Takatsu, *Wasserstein geometry of Gaussian measures*, Osaka J. Math., 48 (2011), pp. 1005–1026.

Division of Applied Mathematics, Brown University, 182 George St., Providence, RI 02912.

Email address: govind_menon@brown.edu