THE ISOMETRIC EMBEDDING PROBLEM AND RANDOM MATRIX THEORY

GOVIND MENON

Abstract. The purpose of these notes is to introduce a set of models that make precise some links between the isometric embedding problem and random matrix theory. The underlying formalism is a model for equlibration that separates a stochastic flow for a gauge from a (typically) deterministic evolution for an observable. Matrix models are introduced to clarify this structure. The simplest of these models is a stochastic flow in the space of Hermitian matrices $\mathbb{H}(n)$ described by the Itô equation

$$dM = [dK, M] := dKM - MdK, \quad dK_{ij} dK_{kl} = C_{ijkl}\, dt.$$

Here the noise dK takes values in the space of anti-Hermitian matrices $\mathbb{A}(n)$, so that C is a positive definite tensor on $\mathbb{A}(n)$.

Despite the fact that $M(t)$ is stochastic, it is shown that for a particular choice of covariance kernel C (see equation (4.24) below) the eigenvalues of M evolve deterministically according to Coulombic repulsion

$$\dot{\lambda}_j = \sum_{k \neq j} \frac{1}{\lambda_j - \lambda_k}, \quad t \geq 0.$$

This model is shown to be exactly solvable.

The main tasks for analysis are to show that this choice of covariance kernel may be obtained through both intrinsic and extrinsic considerations. The intrinsic problem involves showing that a particular choice of C is not just a lucky guess, but is in fact the analytic center of a polytope, so that it is itself determined as the argmin of a solution to an elliptic PDE of Monge-Ampère type. The extrinsic problem is to show that the SDE for dM is the projection of Brownian motion in the space of Hermitian matrices $\mathbb{H}(n)$ onto a coadjoint orbit. The underlying structure– deterministic evolution of an observable (the spectrum) and stochastic evolution of a gauge (the Hermitian matrix) – is representative of a larger class of models that are all motivated by a stochastic approach to the embedding theorems.

Contents

Date: December 29, 2020.
Key words and phrases. Complex Burgers equation,Stochastic gradient descent, isometric embedding.
This work is supported by the National Science Foundation (DMS 1714187).

1. Introduction

The purpose of this note is to develop some links between random matrix theory and the embedding problem for Riemannian manifolds through the analysis of new matrix models. The problems outlined here can be analyzed without any knowledge of the embedding theorems. But it is helpful at the outset to understand the underlying viewpoint and research program as motivation for this work.

1.1. **Random matrix theory and the Nash embedding theorems.**

Given a closed Riemannian manifold $(\mathcal{M}, g)$ an isometric embedding into $(\mathbb{R}^q, e)$ is a C^1 map $u : \mathcal{M} \to \mathbb{R}^q$ such that $u^\sharp e = g$. Here e denotes the Euclidean metric on $\mathbb{R}^q$ and $u^\sharp e$ is the pull-back metric on $\mathcal{M}$. Modern understanding of this problem dates to the pioneering work of Nash in the 1950s [20, 21]. All the work presented here has its origins in an attempt to find a conceptual explanation for the link between the embedding problem and turbulence discovered by De Lellis and Székelyhidi [4].

We will treat embedding as a stochastic flow of immersions $u_t : (\mathcal{M}, g) \to (\mathbb{R}^q, e)$, such that $u := \lim_{t\to\infty} u_t$ satisfies the PDE $u^\sharp e = g$. Any such stochastic flow is determined by the covariance kernel of du_t: we choose it as the solution to a semi-definite program (SDP). In general, this SDP has the character of an infinite-dimensional matrix completion problem, but simpler examples are provided below. Formally one may design the SDP so that the pullback metric g_t evolves deterministically, while u_t is stochastic. The main purpose of this note is to introduce matrix models that allows a better understanding of this structure: stochastic evolution of a 'gauge' (the immersion u_t) and deterministic evolution of an 'observable' (the metric $g_t := u_t^\sharp e$).

The simplest matrix model we study is the SDE

$$dM = [dK, M] := dKM - MdK, \quad dK_{ij} dK_{kl} = C_{ijkl}\, dt, \quad M \in \mathbb{H}(n), \tag{1.1}$$

with a specific covariance structure described in equation (4.24) below. Let $M = Q\Lambda Q^*$ denote the diagonalization of M, with the added convention that $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ with $\lambda_1 < \lambda_2 < \ldots \lambda_n$. In this setting the gauge is the matrix $M(t)$, which evolves stochastically by (1.1). The observable is the spectrum $\Lambda(t)$ which evolves deterministically by Coulombic repulsion between the eigenvalues; an associated spectral measure evolves by the complex Burgers equation.

This model is exactly solvable. This observation is not surprising to someone familiar with random matrix theory since Coulombic repulsion is simply the $\beta \to \infty$ limit of Dyson Brownian motion. What is fundamentally new is the model (1.1), the underlying viewpoint regarding equlibration, and the link between random matrix theory and the embedding theorems. Despite the extensive study of Dyson Brownian motion, it seems not to have been suspected that deterministic Coulombic repulsion of eigenvalues has an underlying stochastic structure, or that this structure may be used to understand problems in apparently unrelated areas. The main goal of this note is to present some tractable conjectures that nail down the underlying Riemannian geometry and stochastic analysis thus linking the areas of random matrix theory, the embedding theorems and turbulence. Let us briefly explain why we view this as an interesting direction.

1.2. **Gibbs measures for embedding.** Our approach to the embedding theorems involves several shifts in viewpoint from Nash's work. The question we wish to formalize and answer is 'What does a typical embedding look like?'. This is in contrast with the historical questions 'Does an embedding exist? If so, how smooth is it?' resolved by Nash and Gromov.

There are two reasons for looking to RMT for inspiration to answer these questions. First, our approach to both problems is rooted in a common technical framework. In both cases, the underlying physical process is equilibration and we model this process with stochastic flows, using an SDP to determine the covariance. We analyze this SDP with a combination of geometry and probability following recent work in the optimization community [11]. Roughly, the constraints in an SDP determine a convex polytope, as well as a canonical Riemannian metric on this polytope given by the Hessian of the unique solution, denoted F, to a critical Monge-Ampere equation (see equation (2.9) below). The function F is called the *canonical barrier* in optimization theory and it may be thought of as the optimal choice amongst the family of all *self-concordant* barriers. [1] The notion of self-concordance was introduced by Nesterov and Nemirovskii to provide a systematic understanding of Newton's method for interior-point methods for conic programs [22]. The interplay between affine and projective geometry and Newton's method dates to Karmarkar's first papers in the subject [16]; recent work sharpens this insight by showing that self-concordance is a bound on the Schwarzian derivative (or equivalently a lower bound on the Ricci curvature) for the metric induced by the barrier [5, 10, 11]. Our goals include a probabilistic understanding of this condition. This direction has not been explored to date. For these reasons, stochastic models for optimization, again focused on matrix models, are also considered in these notes. Here too the underlying structure appears to have some depth and it is necessary to first understand it precisely in simpler models.

Second, at present, RMT offers the best mathematical template for rigorous understanding of universality. This framework goes roughly as follows:

(1) Given an energy $V : \mathbb{R} \to \mathbb{R}$, consider the Gibbs measure on the space of Hermitian matrices $\mathbb{H}(n)$ defined by

$$p_\beta(M)dM = \frac{1}{Z_\beta} e^{-\beta \mathrm{Tr}(V(M))}\, dM, \quad Z_\beta = \int_{\mathbb{H}(n)} e^{-\beta \mathrm{Tr}(V(M))}\, dM,$$

and study its asymptotics as $n \to \infty$. The analog of the law of large numbers – i.e. a deterministic scaling limit in the limit $n \to \infty$ – is the equilibrium spectral measure, which is given by a fixed-point equation depending

[1]This is an unfortunate conflict of terminology with elliptic PDE theory. We will have to be careful to distinguish these notion.

on the potential V. It is possible to formally introduce such potentials for discretizations of embeddings, but there is to date no rigorous construction of Gibbs measures for embeddings.

(2) Investigate the universality of fluctuations. The fundamental laws of fluctuation in RMT are the Dyson-Gaudin-Mehta and Tracy-Widom distributions which depend on the choice of potential only through minor moment assumptions. While the universality of the Kolmogorov spectrum is an empirical fact of turbulence, there appears to have been no previous attempt to study such universality in Riemannian geometry, even numerically. It also does not appear to have been realized that there may be a relation between universality in turbulence and RMT.

Such speculation – even if it is a long way from rigorous analysis – does have a basis in both physics and machine learning. Let us discuss this in turn.

Gibbs measures for embeddings appear in the physics literature in Friedan's work on the nonlinear sigma model [6, 7]. Several of our ideas originate in this work – in particular the emphasis on equilibration, gradient flows, and a geometric view of renormalization. However, we replace Friedan's techniques – Feynman diagrams and expansion in dimension – with the use of stochastic flows and an SDP. It is tempting to draw parallels between our use of information theory and recent work by string theorists, but this is quite speculative at present [23, 28] [2]. Our purpose here is more narrow – it is to provide a dynamic construction of Gibbs measures for embedding based on the technical framework that is illustrated in Section 2. To the best of our knowledge, there appears to have been no attempt to construct Gibbs measures for embeddings rigorously prior to our work (though see [3]).

There is an extensive literature on the use of embeddings in machine learning by mathematicians and computer scientists. This includes work on manifold learning [14] as well as more recent attempts to revisit the embedding theorems to develop notions of canonical embeddings [27]. An essential idea in most of these papers is the use of heat kernel embeddings, based on the work of Bérard, Besson and Gallot [1]. Here too, while the use of heat kernels in the embedding problem is well-established, there appears to have been no prior attempt to investigate the probabilistic and thermodynamic foundations of this idea, as captured in the construction of Gibbs measures. A Bayesian approach to learning also leads directly to this construction problem. For example, Gibbs measures for embeddings are also fundamental for the construction of priors on geometric fields in the Bayesian approach to handwriting recognition. This viewpoint is described in [17, Ch.7].

In summary, while the above results, especially [1, 6, 7] were useful to us when developing our viewpoint, as a matter of technique we find it more fruitful to base our work directly on Nash's insights, combining it with an analysis of SDP. New clustering algorithms and new matrix models for optimization and deep learning that follow from this viewpoint are discussed below.

1.3. **Information theory and embedding.** At first sight, our approach is at variance with the historical development to these theorems which contain no hint of probabilistic reasoning, the importance of the embedding theorems in physics, or (until about ten years ago) their utility in applications. Instead, the main lines

[2] I am grateful to Nima Lashkari for introducing me to these papers.

of mathematical research since Nash's pioneering work are usually seen as the development of the h-principle by Gromov and the formulation of the Nash-Moser implicit function theorems [8, 9, 19, 18].

The 'philosophical core' of our approach is different: we view embedding as a form of information transfer between a source and an observer, recognizing that the process of information transfer is complete when all measurements of distances by the observer are the same as those of the source. This viewpoint is both Bayesian and information theoretic. It places the emphasis not on the structure of the manifold, but on a more primitive aspect of the problem, the measurement of length. In the Bayesian interpretation, the world is random and both the source and the observer are stochastic processes with well-defined parameters (we will construct these processes on a Gaussian space to be concrete). The process of successive approximation implict in Nash's work can now be seen as a control strategy by an observer to tune a model in response to observations of signals from the source. Thus, embedding is simply 'replication' and the process of replication is complete when all measurements by the observer and the source agree on a common set of questions (here it is the question: 'what is the distance between points x and y?').

In this interpretation, the embedding problem for Riemannian manifolds has the same character as apparently simpler problems such as the embedding problem of finite metric spaces and metric graphs into $\mathbb{R}^q$. This idea is explained at the end of this section, since it immediately reveals the mathematical structure of our stochastic flows. It also reveals how one may follow our modification of Nash's insights to create new stochastic algorithms in machine learning. (The embedding of metric spaces and graphs are of fundamental importance in applications such as geo-sensing and the design of navigation systems). Here too the 'core' mathematics is the interplay between the Riemannian geometry of SDP and stochastic flows.

Once one has adopted this intellectual position, the only fundamental bottleneck on embedding is provided by Shannon's channel coding theorem; the traditional emphasis on codimension, smoothness and the use of implicit function theorems begin to appear less fundamental. In particular, at a foundational level, the use of probabilistic or information theoretic methods in the embedding problem should be seen as natural, because it is nothing more than a modern formalism for the process of measurement of lengths. This ties the embedding theorems to much older developments in mathematics, in particular the work of Riemann.

But the truth is much more mysterious. There appear to be many subtle and unexpected ties between our approach (which is just beginning) and these landmark results. Our work began with a specific technical modification of Nash's 1954 paper and it is only through a laborious series of simplifications, guided by many analogies with the work of others, that we arrived at the viewpoint above. At the technical level, our stochastic approach to embedding interpolates between Nash's work in 1954 and 1956, replacing the discrete iteration for low-codimension embeddings in 1954 and the smooth flow for high-codimension embeddings in 1956, with a stochastic flow. The Itô correction provides the desired change of metric in this approach. The SDP arises because many choices of covariance kernel can be used to achieve the same change of metric; the analytic center of the SDP provides a principled choice amongst these. In fact, the analytic center is the argmin of the canonical barrier associated to the SDP. The primary task of this note is to construct simpler models where this SDP can be understood in detail.

By choosing an alternative viewpoint, centered mainly in the relation of mathematics to the sciences, our approach provides a systematic derivation of several new geometric stochastic flows. In each case, the key technical step is a precise understanding of an SDP which has the character of a matrix completion problem. A study of the embedding theorems from this viewpoint also seems to shed new light on the intrinsic construction of Brownian motion on Riemannian manifolds by Eells, Elworthy and Malliavin. An examination of these constructions suggests a natural 'physical' cartoon of embedding as a backward heat flow (this idea still requires a precise formulation; though a sketch is included at the end of this note).

Our main goal at present is to understand the structure of these geometric stochastic flows, both with rigorous analysis on simplified models, a probabilistic approach to SDP and optimization, as well as numerical experiments. In this way, we hope to develop new heuristics and analytic techniques, in order to attack several unresolved questions on the embedding problem.

2. Example: clustering on a finite metric space

Let us illustrate our viewpoint with an instructive example. Assume given a finite metric space (K, ρ) and consider the problem of isometrically embedding K into $\mathbb{R}^q$. A closely related problem is to find an isometric embedding of a finite graph $\mathcal{G}$ into $\mathbb{R}^q$. When $q = 1$ we are constructing a function on a metric space whose level sets may be used to partition the metric space into subsets. This idea is called clustering and it plays an important role in machine learning. These problems are well studied [12], but our take on them is new.

An essential aspect of Nash's 1954 paper is the relaxation of the PDE $u^\sharp e = g$ to a space of C^∞ subsolutions $v^\sharp e < g$ (ordering is in the sense of quadratic forms), followed by an iteration that 'pushes up lengths' by the addition of fluctuations. The purpose of this section is to ilustrate the robustness of this insight when combined with stochastic flows and SDP. For now, these ideas should be seen only as 'proof of concept'. It is of interest to improve on this basic structure to construct fast methods for embedding graphs and metric spaces.

First, consider the problem of embedding a finite metric space. Thus, assume given (K, ρ) and following Nash, let us say that a map $v : K \to \mathbb{R}^q$ is short if $|v(x) - v(y)| < \rho(x, y)$ for every pair of points $x, y \in K$. The set of short maps is clearly non-empty since we may always map all points in K to the origin in $\mathbb{R}^q$. Now let us seek an embedding via a stochastic flow $u_t : K \to \mathbb{R}^q$, $0 \le t < \infty$, insisting that u_t should be short for each t. We define such stochastic flows through Itô SDEs, writing

$$du_t(x) := du(x, t) = \sqrt{dL} \tag{2.1}$$

where the terminology $\sqrt{dL}$ means that the covariance of the mean-zero noise is given by

$$du^i(x, t)du^j(y, t) = \dot{L}^{ij}(x, y; t)\, dt. \tag{2.2}$$

The terminology for the noise is unconventional, but useful since our end result will be a closed equation for du and L.

To simplify matters further, let us assume that $q = 1$ so that we are looking at the clustering problem. It is not true in general that a finite metric space or graph can be embedded into the line; however, our method always provides a relaxation of this problem, and it is easiest to see the structure of an SDP in this setting.

When $q = 1$ equation (2.2) reduces to

$$du(x,t)du(y,t) = \dot{L}(x,y)\,dt$$

and we see that $\dot{L}$ must be a positive semi-definite matrix of size $|K| \times |K|$. For brevity, let $n = |K|$, so that $\dot{L} \in \mathbb{P}(n)$, the space of real positive semi-definite matrices of size $n \times n$. The general structure (2.2) also provides a positive semi-definite covariance tensor, but one must be more careful in analyzing positivity and our purpose here is simply to illustrate the role of SDP.

We term the use of stochastic flows in equation (2.1) as *stochastic kinematics.* The task ahead of us is to choose the matrix $\dot{L}$ using other principles, for example the inclusion of energetics, to obtain a closed evolution equation. To this end, let us again follow Nash and seek a correction of the *metric defect*, which in this discrete setting is the discrepancy between the distances provided by the short map $u_t(x)$ and the distances dictated by the metric ρ. To this end, we apply Itô's formula to the pairwise distances $|u_t(x) - u_t(y)|^2$ to obtain

$$\begin{aligned} d\,|u_t(x) - u_t(y)|^2 &= 2\,(u_t(x) - u_t(y)) \cdot (du_t(x) - du_t(y)) \\ &\quad + (du_t(x) - du_t(y)) \cdot (du_t(x) - du_t(y))\,. \end{aligned} \tag{2.3}$$

The first term on the LHS is a martingale term and its law is determined once we have prescribed the covariance $\dot{L}$. The second term is the Itô correction. It is deterministic and it can be expressed directly in terms of $\dot{L}$ as follows:

$$\left(\dot{L}(x,x) + \dot{L}(y,y) - 2\dot{L}(x,y)\right)\,dt. \tag{2.4}$$

Since u_t is short, we may now choose to 'bump it up' by choosing the above term to be $M(x,y)\,dt$ where $M = M(u) \in \mathbb{P}(n)$. This choice constitutes the *energetics* of our problem. All that matters is that M is defined by a sufficiently smooth map from $\mathbb{R}^n \to \mathbb{P}(n)$, such that $M(v) > 0$ when $v \in \mathbb{R}^n$ is short and $M \equiv 0$ when v is not short. Of course, specific choices of M could affect the rates of convergence and computational cost of the scheme, but what matters for us at this stage is simply the structure of the problem. Once $M(u)$ has been chosen, stochastic kinematics imposes a set of constraints on $\dot{L}$. In particular, if we want the expected value of each pairwise squared distance to improve by $M(x,y)\,dt$ we obtain a set of linear equations of the form

$$\dot{L}(x,x) + \dot{L}(y,y) - 2\dot{L}(x,y) = M(x,y), \quad x,y \in K. \tag{2.5}$$

This set of equations may be rewritten in the standard form of SDP. Let A_{xy} denote the matrix such that $A_{xy}(u,v) = 1$ if $u = v = x$ or $u = v = y$; $A_{xy} = -1$ when $u = x, v = y$ and $u = y, v = x$; and $A_{xy}(u,v) = 0$ otherwise. Then equation (2.5) may be rewritten as the linear constraint

$$\mathrm{Tr}(A_{xy}L) = M(x,y), \quad x,y \in K. \tag{2.6}$$

This equality constraint may also be relaxed further to the inequality constraints

$$\mathrm{Tr}(A_{xy}L) \le M(x,t), \quad x,y \in K, \tag{2.7}$$

in case equation (2.6) has no solution.

Let us take stock of what has been done so far. We have applied Nash's insight of relaxation to the problem of metric space embedding in combination with stochastic calculus and an energetic rule to choose a correction M. This has led us to another relaxation, equation (2.7). The reason we used dL in (2.1) is to allow for the fact that we would like to design our flows so that $L(t)$ is differentiable in order to ensure absolute continuity of the underlying Gaussian measures. This restriction is most interesting in infinite-dimensions where it is imposed by the Cameron-Martin theorem, but it is also interesting in finite-dimensions when constructing low-rank approximations.

This restriction is the last piece of structure we need. Differentiability of $L(t)$ imposes the condition that $\dot{L}$ must lie within the tangent space $T_L\mathbb{P}(n)$. Thus, the linear constraints in (2.7), as well as the positivity constraint

$$\dot{L} \geq 0 \tag{2.8}$$

define a convex polytope $C = C(u, L, M)$ in $T_L\mathbb{P}(n)$. This is the simplest setting in which one sees the appearance of an SDP. However, unlike the standard setting in optimization, we do *not* have a linear cost-function to minimize, what we are seeking is a way to choose a unique point within C that closes the evolution equation.

A principled answer to this question is provided by the notion of the analytic center of a polytope. Under mild non-degeneracy assumptions, a convex polytope in $\mathbb{P}(n)$ carries a natural Riemannian geometry. More precisely, there exists a unique solution to the Monge-Ampère equation [3]

$$\log\det D^2F = 2F, \quad P \in C \tag{2.9}$$

such that $\lim_{x\to\partial C} F = +\infty$. This solution provides a canonical foliation of the polytope by affine hyperspheres and the analytic center is the argmin of F. [4] These two steps in combination now yield the closed dynamical system

$$du = \sqrt{dL}, \quad \dot{L} = \mathrm{argmin}_{P\in C(u,L,M)}F(P)). \tag{2.10}$$

By construction, $\dot{L}$ is monotone increasing and u_t is a bounded martingale, and it is immediate that $\lim_{t\to\infty}(u_t(x), L_t)$ exists *if* one establishes a Lipschitz dependence of the analytic center on u, L and M. In this manner, we may construct Gibbs measures for relaxations of embedding problems. Since the above scheme maps a finite metric space into $\mathbb{R}$, what it provides is a stochastic clustering algorithm.

Once this basic structure has been identified – stochastic kinematics and a degenerate SDP that determines the noise – many variants become possible. For example, the distinction between the embedding of a finite metric space and a graph embedding, is that a graph contains additional information (the connectivity of points by edges), that may be used to 'compress' the space of embeddings. Thus, for instance when considering graph embeddings, we may use the gradients of scalar random fields to simplify the SDP and embedding. In order to avoid further distraction, these algorithms are treated in a separate section.

3. Goals and related work

3.1. **Conjectures.** The concrete tasks that we would like to establish by exploring the RMT model are as follows. Prove that:

[3]Need to transform this from the statement on $\mathbb{R}^n$ to $\mathbb{P}(n)$.

[4]While the proofs of these results use deep PDE results, it is not apparent that the full power of these results is needed in our context. What matters is simply the fact that given only the geometry of a polytope, as defined by constraints such as (2.6), there is *always* a fundamental notion of the center of the polytope. In fact, by relaxing the SDP itself, we hope to obtain a better probabilistic understanding of these results using Gibbs measures for optimization. These ideas are being explored in the thesis projects of Michael Lee and Zsolt Veraszto.

(1) The stochastic flow introduced below corresponds to motion by mean curvature of coadjoint orbits in the space $\mathbb{H}(n)$ of Hermitian operators.
(2) The motion by mean curvature is exactly solvable. This includes an explict description of the time dynamics by the complex Burgers equation, as well as an explicit description of a limiting constant mean curvature surface using Hermite polynomials.
(3) The structure of the SDP for the covariance is almost explicit for the example studied here. We conjecture that the analytic center may be determined explicitly and that it corresponds to our 'lucky guess' in equation (4.24).
(4) This approach suggests new insights into RMT too. As we show below, the scaling limit for the spectral measure of the stochastically evolving matrix $M(t) \in \mathbb{H}(n)$ is given by the complex Burgers equation. Important work of Biane and Voiculescu has shown that the complex Burgers equation should be seen as the heat equation for RMT [2, 26]. However, these results were obtained using tools from asymptotic representation theory and von Neumann algebras. There appears to be no direct construction of a diffusion in a space of bounded operators that corresponds to Voiculescu's free probability. But it seems reasonable to expect that there is a scaling limit of the construction outlined below (which is completely rigorous for any n) that provides a new description of free probability. This problem is definitely harder. It is more prudent to nail down the first three points and publish these, before addressing the question of scaling limits.

3.2. **Related work.** Matrix models of equlibration, all of which reduce to the same philosophy and tools – the construction of stochastic flows by intrinsic and extrinsic methods, SDP for covariance tensors, Gibbs measures for SDP – are being investigated in parallel work with students. In particular, Ching-Peng Huang will explore the Bures-Wasserstein analog of the construction presented here (again joint with Dominik Inauen). Michael Lee and Zsolt Veraszto are studying Riemannian analogs of the following fundamental model for equilibration with applications to optimization and deep learning respectively.

It is always helpful to think about the following classical model when one studies Gibbs measures. Given a potential $V : \mathbb{R}^n \to \mathbb{R}$ and a fixed inverse temperature $\beta > 0$ the Gibbs measure μ_β associated to the potential is the probability measure with density

$$p_\beta(x) = \frac{1}{Z_\beta} e^{-\beta V(x)}, \quad Z_\beta := \int_{\mathbb{R}^n} e^{-\beta V(x)}\, dx. \tag{3.1}$$

The main tasks in statistical mechanics usually reduce to computing the partitition function Z_β or sampling from it. The Gibbs measure may be derived in various ways. A variational approach involves minimizing free energy; the key idea being the role of entropy. This derivation has a Bayesian foundation: the Gibbs measure μ_β represents the best guess of the law of an unknown random variable, given the knowledge of the potential V. This derivation takes its sharpest form when V is either 0 or $+\infty$, so that we have a hard constraint model. The Gibbs measure is then obtained by minimizing the entropy subject to the constraints [13].

It is also helpful to think of the Gibbs measure as a dynamic quantity, so that one forms a physical feel for the process of equilibration. A commonly accepted caricature of equlibration is the Itô SDE

$$dX = -\nabla V(x)dt + \frac{1}{\sqrt{\beta}} dB, \tag{3.2}$$

which describes the motion of a particle in $\mathbb{R}^n$ that is trying to minimize the potential subject to stochastic fluctuations. In this cartoon, $B(t) = (B_1(t), \ldots, B_n(t))$ is standard Brownian motion in $\mathbb{R}^n$. The equilibrium measure for this Markov process is the Gibbs measure μ_β. The physical intuition here is that the potential drives the particle down to its minima, but then the noise kicks it out of the potential well and the equilibrium measure reflects a balance between these effects.

The work with Michael Lee and Zsolt Veraszto tackles the following problems. With Lee, our goal is to provide a stochastic formulation of interior point methods for SDP. This work is stimulated by the need to understand SDP better. Here we build on fundamental insights from the optimization community. In particular, they have shown that efficient algorithms for SDP can be understood as gradient flows on the space $\mathbb{P}(n)$ of positive definite matrices equipped with the trace (or Cartan-Hadamard) metric: at any $X \in \mathbb{P}(n)$, the length of a tangent vector $A \in T_X\mathbb{P}(n)$ is given by $\langle A, A\rangle_X = \mathrm{Tr}(X^{-1}AX^{-1}A)$. Veraszto is exploring a model of deep learning, which too reduces to a gradient flow, but this time on $GL(n)$ equipped with a metric determined by the architecture of the network.

Our approach to both these problems, reduces to the analysis of the Riemannian analog of (3.2). The setting now is as follows. We assume given a Riemannian manifold $(\mathcal{M}, g)$ and a potential $V : \mathcal{M} \to \mathbb{R}$. The Gibbs measure has density given by

$$p_\beta(x) = \frac{1}{Z_\beta} e^{-\beta V(x)}, \quad Z_\beta := \int_{\mathcal{M}} e^{-\beta V(x)} \sqrt{\det g}(dx). \tag{3.3}$$

(That is, we simply replace the volume form on $\mathbb{R}^n$ by its natural Riemannian analog). However, the SDE that replaces (3.2) is more subtle and is given by

$$dX = -\mathrm{grad}_g V(x)dt + \frac{1}{\sqrt{\beta}} dW \tag{3.4}$$

where dW is *intrinsic* Brownian motion with respect to the metric g. To the best of my knowledge, this equation has not been investigated systematically. Certainly, both the applications we consider are completely new.

It is possible to avoid SDE and work with the associated Fokker-Planck equations instead. The analysis of (3.2) then reduces to the celebrated work of Otto and his co-workers [15, 24]. The Riemannian analogs has been investigated by Sturm [25], though I have not looked at it carefully yet. But here too, the emphasis on intrinsic Brownian motion and stochastic differential geometry is new.

Finally, despite the simplicity of the results presented below, to the best of my knowledge the model and all assertions are new. It is for the reasons outlined above that I would like to begin rigorous analysis on my program with this model.

4. Stochastic gradient descent in RMT

We will construct stochastic flows of Hermitian opearators $M(t)$ such that the spectrum $\sigma(M(t))$ evolves deterministically. The main example of this construction is the following: the ordered spectrum $\lambda_1(t) < \lambda_2(t) < \ldots < \lambda_n(t)$ will evolve by

$$\dot{\lambda}_j = \sum_{k \neq j} \frac{1}{\lambda_j - \lambda_k}, \quad 1 \leq j \leq n, \tag{4.1}$$

whereas the Hermitian operators $M(t)$ will solve Itô equations of the form

$$dM = [dK, M]. \tag{4.2}$$

This Itô equation is a stochastic analogue of the deterministic equation

$$\dot{M} = [K(t), M], \tag{4.3}$$

where $K(t)$ denotes a smooth curve in the space of anti-Hermitian matrices $\mathbb{A}(n)$. Such a deterministic equation yields an isospectral flow. However, the Itô equation (4.2) is not isospectral: the Itô correction is transverse to the isospectral manifold.

An Itô equation such as (4.2) is completely prescribed by the covariance tensor $dK_{jk}dK_{lm}$. An explicit covariance tensor for which the spectrum of $M(t)$ solves equation (4.1) is given in equation (4.25) below. An equivalent description of (4.1) is provided by an evolution equation for the anti-Herglotz function

$$g(z,t) = \sum_{j=1}^{n} \frac{1}{z - \lambda_j(t)} = \mathrm{Tr}(R(z; M(t)), \quad R(z;M) := (z - M)^{-1}, \quad z \in \mathbb{C}_+. \tag{4.4}$$

Equation (4.1) is equivalent to the complex Burgers equation

$$g_t + g g_z = \frac{1}{2} g_{zz}, \quad z \in \mathbb{C}_+. \tag{4.5}$$

The specific choice of covariance such that equation (4.2) leads to equation (4.29) involves some guess work. It is of interest to systematize this procedure. We formulate a semidefinite program (SDP) to determine all tensors that 'lift' equation (4.1) to (4.2). This SDP has the structure of a matrix completion problem: we are given the diagonal entries of a positive definite matrix and our task is to make a principled guess for the matrix. We approach this question using Bayesian principles and formulate Gibbs measures for SDP that apply in particular to matrix completion.

This formalism is applied to equations (4.1) and (4.2). This construction yields the complex Burgers equation in the limit $n \to \infty$. The main rigorous challenge is to construct an associated diffusion process of operators $M(t)$ and to formulate a notion of entropy solution to (4.2) in this limit. We expect that this construction will provide a new description of free probability.

4.1. **Notation.** The space of Hermitian and positive definite matrices is denoted $\mathbb{H}(n)$ and $\mathbb{P}(n)$ respectively. By default, Hermitian means complex Hermitian, though we use the same terminology for real matrices, 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. Similarly, the unitary group is denoted $U(n)$ in both settings. However, when diagonalizing complex Hermitian matrices, we will work with the quotient group $U(n)/\mathbb{T}^n$ and $\mathbb{A}(n)$, the space of anti-Hermitian matrices with zero diagonal. For every $K \in \mathbb{A}(n)$, the curve $\exp(tK)$ lies in $U(n)/\mathbb{T}^n$, $t \in \mathbb{R}$.

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 $\mathbb{M}(n)$

$$\langle M, N \rangle := \mathrm{Re}\,\mathrm{Tr}(M^* N), \quad M, N \in \mathbb{M}(n), \tag{4.6}$$

is used throughout. The asterisk is dropped when M and N are Hermitian.

All calculations below are for real symmetric matrices. They should be extended to complex Hermitian matrices, but some care is needed when defining the possible covariances. We will come back to this. We still use the notation *, $\mathbb{A}(n)$ and $\mathbb{H}(n)$.

4.2. **Diagonalization and isospectral flows.** Let us write

$$M = Q\Lambda Q^*, \tag{4.7}$$

for the diagonalization of a matrix $M \in \mathbb{H}(n)$. We assume that the eigenvalues are distinct so that the change of variables $M \mapsto \mathbb{R}^n \times U(n)$ is locally analytic. We refer to Λ as the *observable* and Q as the *gauge*.

The tangent space $T_M\mathbb{H}(n)$ admits a direct space decomposition, which is also an orthogonal decomposition when one uses the Frobenius norm on $\mathbb{H}(n)$. Consider a smooth curve $M(t)$ and differentiate it to find

$$\dot{M} = [\dot{K}, M] + Q\dot{\Lambda}Q^*, \quad \dot{Q} = \dot{K}Q. \tag{4.8}$$

where $\dot{K} \in \mathbb{A}(n)$. Similarly, we also have

$$Q^*\dot{M}Q = [\Lambda, \dot{L}] + \dot{\Lambda}, \quad \dot{Q} = Q\dot{L}, \tag{4.9}$$

with $\dot{L} \in \mathbb{A}(n)$. The difference in these two equations is the choice of right or left-multiplication on $U(n)$. The variables $\dot{K}$ and $\dot{L}$ are related through

$$\dot{L} = Q^*\dot{K}Q. \tag{4.10}$$

Equation (4.8) is more convenient for defining SDE on $\mathbb{H}(n)$. On the other hand, equation (4.9) reveals the splitting of $T_M\mathbb{H}(n)$ more clearly. Since the matrix $[\Lambda, \dot{L}]$ vanishes on the diagonal, we see that it is orthogonal to the space of diagonal matrices with respect to $\langle \cdot, \cdot \rangle$. Thus, equation (4.9) reveals an orthogonal splitting of $T_M\mathbb{H}(n)$ with respect to the inner-product (4.6).

4.3. **The resolvent.** Define the resolvent $\mathbb{H}(n) \times \mathbb{C}_+ \to GL(n;\mathbb{C})$

$$(z, M) \mapsto R(z; M) := (z - M)^{-1}, \quad z \in \mathbb{C}_+. \tag{4.11}$$

For fixed M, the map $z \mapsto R(z)$ is analytic in the upper half plane. Similarly, holding z fixed, the map $M \mapsto R(z; M)$ is analytic in M. We compute its first and second derivatives as follows. Consider a curve $(-1, 1) \to \mathbb{H}(n)$, $t \mapsto M(t)$, with

$$M(0) = M, \quad \dot{M}(0) = S, \quad \ddot{M}(0) = 0. \tag{4.12}$$

The straight line, or geodesic in Frobenius norm, $M(t) = M + tS$ satisfies this condition. We then differentiate the identity

$$R(t)(z - M(t)) = \mathrm{Id} \tag{4.13}$$

with respect to t to obtain the idenities

$$\dot{R} = R\dot{M}R, \quad \ddot{R} = R\ddot{M}R + 2R\dot{M}R\dot{M}R. \tag{4.14}$$

We evaluate this expression at $t = 0$ to obtain the first and second derivatives of R at M in the direction S

$$DR(M)(S) = RSR, \quad D^2R(M)(S, S) = 2RSRSR. \tag{4.15}$$

We will need both these expressions when applying Itô's formula to (4.7).

4.4. **Stochastic gradient descent.** Define the projection map

$$\pi : \mathbb{H}(n) \to \mathbb{R}^n, \quad M \mapsto \Lambda. \tag{4.16}$$

We want to design flows such that M evolves stochastically, whereas $\Lambda = \pi(M)$ evolves deterministically.

The simplest situation where one sees this interplay is for the SDE

$$dM = [dK, M] = \mathrm{ad}_{dK} M. \tag{4.17}$$

An SDE of this type is completely defined by a positive definite covariance tensor

$$dK_{ij} dK_{kl} = C_{ijkl}\, dt. \tag{4.18}$$

Despite first appearances, the description of C requires no stochastic calculus. Indeed, consider any Gaussian measure on $\mathbb{A}(n)$ and let K be a random variable with respect to this measure. The covariance tensor C for real K is defined by

$$C_{ijkl} = \mathbb{E}\left(K_{ij} K_{kl}\right). \tag{4.19}$$

Such a tensor may first be prescribed in the range $1 \le i < j \le n$ and $1 \le k < l \le n$ and then extended to $\mathbb{A}(n)$ through the symmetries

$$C_{ijkl} = C_{klij} = -C_{jikl} = -C_{ijlk}. \tag{4.20}$$

These are part of the symmetries of a Riemann curvature tensor. However, the Riemann curvature tensor also includes the symmetries of the Jacobi identity, which do *not* hold in the example (4.24) that we consider below. This requires more thought. Similarly, when K is complex Hermitian, the book-keeping needs more care.

The indices are ordered so that the map $\mathbb{A}(n) \to \mathbb{A}(n)$, $A \mapsto C(A)$ defined by

$$C(A)_{ij} = C_{ijlk} A_{kl} \tag{4.21}$$

is positive definite with respect to the Frobenius norm (4.6) (note the switch in the outer two indices). Let's check this. For any $A \in \mathbb{A}(n)$ we have

$$\langle A, C(A)\rangle = \mathrm{Tr}(A^* C(A)) = A_{ji} C_{ijlk} A_{kl} \tag{4.22}$$

$$\overset{(4.19)}{=} \mathbb{E}\left(A_{ji} K_{ij} A_{kl} K_{lk}\right) = \mathbb{E}\left(\mathrm{Tr}(A^* K)\mathrm{Tr}(K^* A)\right) \tag{4.23}$$

$$= \mathbb{E}\left|\mathrm{Tr}(A^* K)\right|^2 \ge 0,$$

with equality only when A vanishes.

(It is not apparent to me that this is the best way to order the indices; the convention in Lie theory is to have a negative definite Killing form. I'd like to keep things as direct as possible in a manner that is probabilstically natural.)

In general, we will choose the covariance tensor C as a function of M. For fixed M, and thus fixed Q, it will be useful to express the covariance tensor in terms of the variable dL introduced in (4.8). We use equation (4.10) and equation (4.18) to define

$$D_{ijkl} := \frac{dL_{ij} dL_{kl}}{dt} = Q_{pi} Q_{qj} Q_{rk} Q_{sl} \frac{dK_{pq} dK_{rs}}{dt} = Q_{pi} Q_{qj} Q_{rk} Q_{sl} C_{pqrs}. \tag{4.24}$$

An important special case is the following

$$D_{ijkl} = \frac{1}{\lambda_j - \lambda_i} \frac{1}{\lambda_l - \lambda_k}, \quad i \ne j, k \ne l. \tag{4.25}$$

Let us check that D is positive definite. Let $S \in \mathbb{H}(n;\mathbb{R})$. We denote the sum over repeated indices explicitly and compute

$$D_{ijkl} = \sum_{ijkl} \frac{1}{\lambda_j - \lambda_i}\frac{1}{\lambda_l - \lambda_k} S_{ij}S_{kl} = \left(\sum_{lm} \frac{S_{lm}}{\lambda_m - \lambda_l}\right)^2 \geq 0. \tag{4.26}$$

The underlying structure has little to do with the explicit form of D_{ijkl}. What matters is that D_{ijkl} factorizes in the form $D_{ijkl} = B_{ij}B_{kl}$ for an antisymmetric matrix B. These calculations should work similarly for covariances on $\mathbb{H}(n;\mathbb{C})$, but this requires a little more care.

4.5. **Evolution of the resolvent.** Once the covariance is prescribed (and assuming existence of solutions), the evolution of the resolvent is determined by Itô's formula. We use equations (4.15) and (4.17) to obtain

$$dR = R[dK,M]R + R[dK,M]R[dK,M]R. \tag{4.27}$$

The first of these terms is a martingale term and the second is an Itô correction. Neither vanishes, but the *trace* of the martingale term vanishes. Since the trace defines a spectral measure, let us consider these terms explicitly and introduce notation for it. Define the anti-Herglotz function g and the (unnormalized) spectral measure ν_t through the formulas

$$g(z,t) := \mathrm{Tr}(R(z;M)) = \sum_{j=1}^{n} \frac{1}{z-\lambda_j(t)} := \int_{\mathbb{R}} \frac{1}{z-s}\nu_t(ds). \tag{4.28}$$

We claim that with the choice (4.25), g satisfies the complex Burgers equation

$$g_t + g g_z = \frac{1}{2} g_{zz}, \quad z \in \mathbb{C}_+. \tag{4.29}$$

The rest of this note builds up to this fact.

First, we claim that when dR solves (4.27), the evolution of g is deterministic. The proof is a computation with the martingale term

$$\mathrm{Tr}\left(R[dK,M]R\right) = \mathrm{Tr}(RdKMR) - \mathrm{Tr}(RMdKR) \tag{4.30}$$

$$= \mathrm{Tr}\left((MR^2 - R^2M)dK\right) = 0, \tag{4.31}$$

since M and R commute.

The computation of the second term in (4.28) is a little more tedious since one must compute the trace of each of the four terms

$$R[dK,M]R[dK,M]R = \tag{4.32}$$
$$R\,dKMR\,dKMR + RM\,dKRM\,dKR - RM\,dKR\,dKMR - R\,dKMRM\,dKR.$$

These calculations may be simplified by diagonalization and replacing dK with $dL = Q^* dKQ$. For brevity, let $S = Q^*RQ = (z-\Lambda)^{-1}$. Both S and Λ are diagonal matrices. Therefore, the trace of the first term in (4.32) is

$$\mathrm{Tr}(RdKMRdKMR) = \mathrm{Tr}(SdL\Lambda SdL\Lambda S) \tag{4.33}$$
$$= S_{j_1j_2}dL_{j_2j_3}(\Lambda S)_{j_3j_4}dL_{j_4j_5}(\Lambda S)_{j_5j_1} = \frac{\lambda_j}{(z-\lambda_j)^2}\frac{\lambda_k}{(z-\lambda_k)}dL_{jk}dL_{kj}$$
$$= \frac{\lambda_j}{(z-\lambda_j)^2}\frac{\lambda_k}{(z-\lambda_k)}D_{jkkj}\,dt.$$

In the second line, we first write out the trace explicitly using the summation convention. We then simplify terms using the fact that S and Λ are diagonal and relabel the two dummy indices as j and k. The same technique is used for all the terms. Thus, we compute the second term in (4.32)

$$\begin{aligned}\mathrm{Tr}(RM\,dKRM\,dK\,R) &= \mathrm{Tr}(S\Lambda\,dLS\Lambda\,dL\,S) \\ &= (S\Lambda)_{j_1j_2}dL_{j_2j_3}(S\Lambda)_{j_3j_4}dL_{j_4j_5}S_{j_5j_1} = \frac{\lambda_j}{(z-\lambda_j)^2}\frac{\lambda_k}{(z-\lambda_k)}dL_{jk}dL_{kj} \\ &= \frac{\lambda_j}{(z-\lambda_j)^2}\frac{\lambda_k}{(z-\lambda_k)}D_{jkkj}\,dt.\end{aligned} \tag{4.34}$$

Minus the third term in (4.32) is

$$\begin{aligned}\mathrm{Tr}(RM\,dK\,R\,dK\,MR) &= \mathrm{Tr}(S\Lambda\,dL\,S\,dL\,\Lambda S) \\ &= (S\Lambda)_{j_1j_2}dL_{j_2j_3}S_{j_3j_4}dL_{j_4j_5}(\Lambda S)_{j_5j_1} = \frac{\lambda_j^2}{(z-\lambda_j)^2}\frac{1}{(z-\lambda_k)}dL_{jk}dL_{kj} \\ &= \frac{\lambda_j^2}{(z-\lambda_j)^2}\frac{1}{(z-\lambda_k)}D_{jkkj}\,dt.\end{aligned} \tag{4.35}$$

Finally, minus the fourth term in (4.32) is

$$\begin{aligned}\mathrm{Tr}(R\,dK\,MRM\,dK\,R) &= \mathrm{Tr}(S\,dL\,\Lambda S\Lambda\,dL\,S) \\ &= (S)_{j_1j_2}dL_{j_2j_3}(\Lambda S\Lambda)_{j_3j_4}dL_{j_4j_5}S_{j_5j_1} = \frac{1}{(z-\lambda_j)^2}\frac{\lambda_k^2}{(z-\lambda_k)}dL_{jk}dL_{kj} \\ &= \frac{1}{(z-\lambda_j)^2}\frac{\lambda_k^2}{(z-\lambda_k)}D_{jkkj}\,dt.\end{aligned} \tag{4.36}$$

We collect all terms to obtain the evolution equation

$$dg = -\frac{(\lambda_j-\lambda_k)^2}{(z-\lambda_j)^2}\frac{1}{(z-\lambda_k)}D_{jkkj}\,dt. \tag{4.37}$$

When D is given by (4.25), and we note that $D_{jkkj}=0$ when $j=k$, this expression simplifies to

$$g_t := \partial_t g = \sum_{j\neq k}\frac{1}{(z-\lambda_j)^2}\frac{1}{(z-\lambda_k)} \tag{4.38}$$

This expression is equivalent to the complex Burgers equation (4.29), since

$$\partial_z g = -\sum_j \frac{1}{(z-\lambda_j)^2}, \quad \partial_z^2 g = \sum_l \frac{2}{(z-\lambda_l)^3}, \tag{4.39}$$

so that the sum in equation (4.38) may be expressed as

$$\sum_{j\neq k}\frac{1}{(z-\lambda_j)^2}\frac{1}{(z-\lambda_k)} = \sum_{j,k}\frac{1}{(z-\lambda_j)^2}\frac{1}{(z-\lambda_k)} - \sum_l \frac{2}{(z-\lambda_l)^3} = -gg_z + \frac{1}{2}g_{zz}. \tag{4.40}$$

Thus, we have established that the Cauchy transform of the spectral measure of M satisfies the complex Burgers equation (4.5) when $M(t)$ solves the stochastic evolution equation (4.2), with covariance kernel given by (4.25). The equivalence between (4.1) and (4.4) is seen as follows. We use the second equality in definition (4.28) along with equation (4.1) to obtain the identity

$$\partial_t g(z,t) = -\sum_j \frac{1}{(z-\lambda_j)^2}\dot{\lambda}_j = -\sum_j \frac{1}{(z-\lambda_j)^2}\sum_{k\neq j}\frac{1}{\lambda_j-\lambda_k}. \tag{4.41}$$

We then substitute the identity (for $j \neq k$)

$$\frac{1}{(z-x_j)^2}\frac{1}{x_j-x_k}+\frac{1}{(z-x_k)^2}\frac{1}{x_k-x_j} = \frac{1}{(z-x_j)^2}\frac{1}{(z-x_k)}+\frac{1}{(z-x_k)^2}\frac{1}{(z-x_j)} \tag{4.42}$$

into (4.41) to obtain equation (4.38), which implies equation (4.40) and thus (4.4). Thus, all that remains is to establish the identity (4.42). This is a partial fractions expansion. Clearly,

$$\begin{aligned} &\frac{1}{(z-x_j)^2}\frac{1}{x_j-x_k}+\frac{1}{(z-x_k)^2}\frac{1}{x_k-x_j} \\ &= \frac{1}{x_j-x_k}\left(\frac{1}{(z-x_j)^2}-\frac{1}{(z-x_k)^2}\right) \\ &= \frac{1}{x_j-x_k}\frac{1}{(z-x_j)^2}\frac{1}{(z-x_k)^2}\left((z-x_k)^2-(z-x_j)^2\right) \\ &= \frac{1}{(z-x_j)^2}\frac{1}{(z-x_k)}+\frac{1}{(z-x_k)^2}\frac{1}{(z-x_j)}. \end{aligned} \tag{4.43}$$

Equation (4.4) is *exactly solvable* by the Cole-Hopf transformation. However, this is not the usual calculation, since it involves Cole-Hopf in the complex plane. It suggests that one may approach free probability using purely analytic methods, rather than a combinatorial or representation theoretic framework.

4.6. **Gibbs measures for matrix completion.** The connection with Gibbs measures and matrix completion may now be explained. The above calculation, along with the specific choice of D in (4.25), shows that there is a covariance kernel D such that when M solves the Itô equation (4.2), its spectral measure satisfies (4.1), which is equivalent to the complex Burgers equation.

However, this choice is ad hoc. A closer examination of the calculation reveals that all that is required of D is that

(1) The tensors C, or equivalently D, must be positive definite on $\mathbb{A}(n)$.
(2) The value of D on the diagonal, i.e. pairs of indices (ij), (kl) with $i = k$,$j = l$ must be given by

$$D_{jkjk} = \frac{1}{(\lambda_j-\lambda_k)^2}. \tag{4.44}$$

The symmetry $D_{ijkl} = -D_{ijlk}$ then implies $D_{jkkj} = -1/(\lambda_j-\lambda_k)^2$, which is all that is needed for (4.38).

Equations (4.25) and (4.26) reveal that there is a choice of D such that both these conditions hold. This choice has an appealing simplicity. However, the main weakness in imposing (4.25) is that it does not provide an explanation for which, if any, of the possible solutions to conditions (1) and (2) above is canonical.

The essential structure in (1) and (2) is of a matrix completion problem. The structure of the problem is as follows. We are given the values of a positive definite tensor on its diagonal and we must reconstruct the tensor. Here too one must confront the fact that we may have many solutions to this problem. We find it most helpful to use Gibbs measures as the natural tool for resolving this indeterminancy.

Suppose one has constructed a natural Gibbs measure, μ_β supported on the set of tensors that satisfy (1) and (2) above. This allows the principled choice

$$D = \int_{\mathbb{P}(n)} \tilde{D}\, d\mu_\beta(\tilde{D}). \tag{4.45}$$

In this manner, we may see the deterministic flow 'downstairs' as being matched with a stochastic flow 'upstairs' in a manner that corresponds to a true thermodynamic evolution.

References

[1] P. Bérard, G. Besson, and S. Gallot, *Embedding Riemannian manifolds by their heat kernel*, Geometric & Functional Analysis GAFA, 4 (1994), pp. 373–398.

[2] P. Biane, *Representations of symmetric groups and free probability*, Advances in Mathematics, 138 (1998), pp. 126–181.

[3] M. Carfora, *The Wasserstein geometry of nonlinear σ models and the Hamilton-Perelman Ricci flow*, Rev. Math. Phys., 29 (2017), pp. 1750001, 71.

[4] C. De Lellis and L. Székelyhidi Jr., *Dissipative continuous Euler flows*, Invent. Math., 193 (2013), pp. 377–407.

[5] D. J. Fox, *A schwarz lemma for kähler affine metrics and the canonical potential of a proper convex cone*, Annali di Matematica Pura ed Applicata (1923-), 194 (2015), pp. 1–42.

[6] D. Friedan, *Nonlinear models in $2+\varepsilon$ dimensions*, PhD thesis, Lawrence Berkeley Laboratory, University of California, 1980.

[7] ———, *Nonlinear models in $2+\varepsilon$ dimensions*, Phys. Rev. Lett., 45 (1980), pp. 1057–1060.

[8] M. Gromov, *Partial differential relations*, vol. 9 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], Springer-Verlag, Berlin, 1986.

[9] R. S. Hamilton, *Three-manifolds with positive Ricci curvature*, J. Differential Geom., 17 (1982), pp. 255–306.

[10] R. Hildebrand, *Canonical barriers on convex cones*, Mathematics of operations research, 39 (2014), pp. 841–850.

[11] ———, *Conic optimization: affine geometry of self-concordant barriers and copositive cones*, PhD thesis, Université Grenoble Alpes, 2017.

[12] P. Indyk, J. Matoušek, and A. Sidiropoulos, *Low-distortion embeddings of finite metric spaces*, Handbook of discrete and computational geometry, 37 (2004), p. 46.

[13] E. T. Jaynes, *Information theory and statistical mechanics*, Physical review, 106 (1957), p. 620.

[14] P. W. Jones, M. Maggioni, and R. Schul, *Manifold parametrizations by eigenfunctions of the Laplacian and heat kernels*, Proceedings of the National Academy of Sciences, 105 (2008), pp. 1803–1808.

[15] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17.

[16] N. Karmarkar, *A new polynomial-time algorithm for linear programming*, Combinatorica, 4 (1984), pp. 373–395.

[17] G. Menon, *Pattern theory: old and new*, Lecture Notes, Brown University, Providence, RI 02912, 2020.

[18] J. Moser, *A rapidly convergent iteration method and non-linear differential equations. II*, Ann. Scuola Norm. Sup. Pisa (3), 20 (1966), pp. 499–535.

[19] ———, *A rapidly convergent iteration method and non-linear partial differential equations. I*, Ann. Scuola Norm. Sup. Pisa (3), 20 (1966), pp. 265–315.

[20] J. Nash, *C^1 isometric imbeddings*, Ann. of Math. (2), 60 (1954), pp. 383–396.

[21] ———, *The imbedding problem for Riemannian manifolds*, Ann. of Math. (2), 63 (1956), pp. 20–63.

[22] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*, SIAM, 1994.

[23] T. Nishioka, *Entanglement entropy: holography and renormalization group*, Rev. Modern Phys., 90 (2018), pp. 035007, 56.
[24] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.
[25] K.-T. Sturm et al., *Diffusion processes and heat kernels on metric spaces*, The Annals of Probability, 26 (1998), pp. 1–55.
[26] D. Voiculescu, *Free probability theory: random matrices and von neumann algebras*, in Proceedings of the International Congress of Mathematicians, Springer, 1995, pp. 227–242.
[27] X. Wang, K. Zhu, et al., *Isometric embeddings via heat kernel*, Journal of Differential Geometry, 99 (2015), pp. 497–538.
[28] E. Witten, *Notes on some entanglement properties of quantum field theory*, arXiv preprint arXiv:1803.04993, (2018).

Division of Applied Mathematics, Brown University, 182 George St., Providence, RI 02912.
Email address: govind_menon@brown.edu