GAUSSIAN PROCESSES, THE ISOMETRIC EMBEDDING PROBLEM AND TURBULENCE

GOVIND MENON

2010 *Mathematics Subject Classification.* 35F21,57Q35,60G15,82C21.

Key words and phrases. Isometric embedding, stochastic differential geometry, turbulence.

This work is supported by the National Science Foundation (DMS 1714187), the Simons Foundation and the Charles Simonyi Foundation. I am deeply grateful to Camillo De Lellis and Tom Spencer for their support. I thank Roger Brockett and Jean-Bernard Zuber for spending time mentoring my graduate students during my sabbatical year. Thanks also to Francesco Maggi, Mokshay Madiman, Bob Pego and Felix Otto for listening to these ideas at a tentative stage.

Abstract. These notes outline a new approach to the isometric embedding problem and related problems, including turbulence. We introduce a model and describe its roots in various areas of mathematics and the sciences. The goal is to document partial progress in order to begin numerics and rigorous mathematical analysis.

The main new idea is to formulate the embedding problem as a stochastic gradient descent. The probabilistic foundation of our work is the fact that spaces of Gaussian measures have a natural Riemannian geometry and that every Riemannian manifold carries an intrinsic notion of Brownian motion. The common structure of the PDE we study is that they have a natural notion of subsolution, along with a partial order on subsolutions. Informally, solutions to the PDE may be seen as the 'boundary' of the space of subsolutions. We combine these ideas to contruct probability measures on solutions by running a diffusion in the space of subsolutions that fluctuates 'outwards' to the 'boundary'. As a numerical algorithm, this corresponds to a stochastic interior-point method for optimization.

This idea takes the following form for the isometric immersion problem. Here we are given a Riemannian manifold $(\mathcal{M}^n, g_\infty)$ and our task is to find a map $u : \mathcal{M} \to \mathbb{R}^q$ such that $u^\sharp e = g_\infty$, where $u^\sharp e$ denotes the pullback of the identity metric e on $\mathbb{R}^q$. The set of subsolutions in this problem consists of (sufficiently smooth) *short* maps $v : \mathcal{M} \to \mathbb{R}^q$ such that $v^\sharp e < g_\infty$, where the inequality refers to the natural order on symmetric $(0,2)$ tensors on $\mathcal{M}$.

We reformulate this problem as a stochastic dynamical system for an infinite-dimensional Gaussian measure $\mu_t \equiv (u_t, L_t)$, $0 \le t < \infty$. Here u_t, the mean of μ_t, is a smooth short map $\mathcal{M} \to \mathbb{R}^q$. The fluctuations around u_t are described by L_t, a covariance kernel for a smooth centered Gaussian random field taking values in $\mathbb{R}^q$. Since the space of Gaussian measures has a natural Riemannian geometry, it is possible to define a diffusion on it by solving a stochastic differential equation for μ_t. In this manner, we lift the PDE $u^\sharp e = g_\infty$ to a stochastic flow of Gaussian measures μ_t.

Several natural diffusions may be constructed in this way. The typical structure we study is the stochastic flow expressed in coordinates (u_t, L_t) by

$$
\begin{aligned}
du &= \sqrt{dL}, \\
\dot{L} &= \operatorname{argmin}_{P \in T_L \mathsf{Pos}_q(\mathcal{M}, g)_+} E(u, L, P).
\end{aligned}
$$

The formal, but suggestive, notation $\sqrt{dL}$ is used to signify that $\dot{L}$ should be thought of as a 'stochastic velocity field'. Its precise meaning is that $\dot{L}$ is the covariance of a centered spatially smooth Gaussian noise. Thus, the first equation describes the stochastic kinematics of the problem. The second equation describes the energetic and gradient structure. In the simplest setting, the cost function E is analogous to a constrained Dirichlet energy. The tangent space $T_L \mathsf{Pos}_q(\mathcal{M})$ and the positive cone $T_L \mathsf{Pos}_q(\mathcal{M})_+$ are defined using the Riemannian geometry of Gaussian measures. The model provides a joint construction of a stochastically improving subsolution u_t along with an increasing family L_t of Gaussian kernels. It thus provides a probability measure supported on isometric immersions as well as information on fluctuations.

Other problems that may be reformulated in a similar way include some Hamilton-Jacobi equations, the Euler equations for incompressible fluids, and some KAM theorems. This approach also provides a bridge between the statistical theory of fields and various applications. These include the thermodynamic foundations of continuum mechanics, turbulence and Bayesian formulations of learning.

In order to complete this program, several key steps need to be established rigorously. These include the analysis of the Riemannian geometry of Gaussian measures, elliptic regularity for the minimization problem, and the development and implementation of fast numerical schemes. Several simpler problems are considered in order to develop these ideas.

Contents

1. Introduction

1.1. **The isometric embedding problem and turbulence.** The primary purpose of this work is to introduce a new constructive method in the statistical theory of fields. This work has its roots in the unexpected link between Nash's work on the isometric embedding problem and the Euler equations discovered by De Lellis and Székelyhidi [18, 21]. Their work has stimulated rapid progress on h-principles for PDE that is well documented by now [19, 20].

We approach this link in a different way, by viewing both problems through the lens of statistical mechanics. In this context, the underlying physical questions are classical: what form does the second law of thermodynamics take for a system with infinitely many degrees of freedom? How do we resolve the ultraviolet divergence present in classical field theories? If we believe that turbulence can be described by statistical mechanics, how should we formulate the process of equilibration so that it corresponds to the empirically observed universality of the cascade?

The main theme that runs through this work is an attempt to bridge PDE theory with information theory and statistical mechanics with the above examples in mind. We will view entropy as the primary concept and formulate an existence theory that replaces Nash's schemes with a stochastic flow of Gaussian measures that corresponds to a gradient descent of free energy. The construction of such stochastic flows is based on the following observations: (a) there is well-defined notion of Brownian motion, and thus stochastic differential equations, on any Riemannian manifold; (b) the space of Gaussian measures on $\mathbb{R}^n$ equipped with the 2-Wasserstein distance may itself be thought of as a (stratified) Riemannian manifold; (c) there is a well-developed theory of Gaussian measures on reproducing kernel Hilbert spaces, so that (a) and (b) can be extended to infinite dimensions.

We apply these ideas as follows. In order to solve a given PDE for a variable u, we 'lift' the PDE into a stochastic flow of infinite-dimensional Gaussian measures μ_t, $0 \leq t < \infty$ where μ_t has mean u_t and covariance kernel L_t. Here t denotes an energy scale, u_t is a smooth subsolution to the PDE and L_t serves to describe band-limited fluctuations around u_t. Heuristically, the Gaussian measure μ_t describes an approximate solution u_t along with 'error-bars' given by L_t. We then use a stochastic flow for μ_t to minimize a natural cost function, such as a relative entropy, thus driving μ_t towards a probability measure μ_∞ whose mean u_∞ is a solution to the given PDE. We design the flows so that pathwise convergence of $\mu_t \equiv (u_t, L_t)$ is an immediate consequence of the martingale convergence theorem. The main task is then to show that the limit u_∞ is a solution to the given PDE. If succesful, the method provides more than existence, since it includes a space of fluctuations parametrized by L_∞, in addition to a solution u_∞.

This approach has a thermodynamic interpretation. The stochastic evolution(s) proposed here correspond to infinite-dimensional Fokker-Planck equations that describe cascades in the isometric embedding problem and turbulence as a process of equilibration. The work of Otto and his co-workers has shown that the Fokker-Planck equations correspond to a gradient descent of free energy. This idea is implicit in our work, but we find it simpler to work with stochastic flows instead of Fokker-Planck equations on infinite-dimensional spaces of Gaussian measures. These viewpoints are loosely equivalent and the choice one makes is a matter of taste.

The main purpose of these notes is to convince the reader that this approach is natural. To this end, we focus on explaining the geometry of Gaussian measures and the stochastic flows in simpler examples. While our ideas originated in an attempt to improve Nash's work on the isometric embedding problem, the main ideas are best understood in a simpler setting. Much of this work is devoted to constructing probability measures supported on 1-Lipschitz functions on the interval $[0, 1]$. The structure of the stochastic dynamical system we propose is easiest to see in this problem. Once this example has been understood, the reader will see that these ideas extend immediately to the construction of 1-Lipschitz functions on Riemannian manifolds. This problem is of independent interest. We then return to the embedding problem and introduce a stochastic flow, which differs from the 1-Lipschitz problem mainly in the increased complexity of the associated space of Gaussian measures. Once this problem has been absorbed, it becomes clear that we have the beginnings of a general method. In order to demonstrate this point, we formulate stochastic dynamical systems for several other applications.

An assesment of the choices we make in the design of these dynamical systems, along with numerical schemes to compute the flows, reveals interesting connections with other areas. These include a probabilistic view of interior-point methods for semidefinite programming and the possible extension of universality from random matrix theory to geometry and turbulence.

1.2. **Information theory and PDE.** The embedding problem is closely tied to foundational questions in mathematics, physics and statistics. The techniques presented below do not require any familiarity with these questions, but it is helpful to discuss these briefly since these issues have guided our choice of model.

We remarked at the outset that our goal is to develop a new method in the statistical theory of fields. But how fundamental is statistical mechanics? When Shannon created information theory in the 1940s, the early speculation was that information theory should be seen as a branch of statistical mechanics, since entropy is the primary concept in both theories. However, in the 1950s, Jaynes showed that statistical mechanics may itself be derived from information theory. Jaynes' work has disturbing philosophical implications, since it causes us to question if there are *any* fundamental physical models. One of the interpretations of the entropy of a discrete random variable is that it provides the optimal length of a code for the random variable. This interpretation leads to the idea, developed mainly by computer scientists and statisticians, that the best models are those that have Minimum Description Length (MDL). This view is an extreme counterpoint to physical theories since it makes no assumptions on the structure of the source that produces the signal and states only that the best description of the data is the one that compresses it in the most efficient manner. In the MDL description, learning *is* data compression and there are no 'true' models.

It seems fair to say that such foundational concerns play no part in the analysis of PDE. Instead, analysts usually adopt the pragmatic approach that physical models, such as the equations of continuum physics, are well-founded because these have been tested by physical and numerical experiments and a rich mathematical tradition attests to their internal consistency. Such pragmatism can be justified by the large gap between the analysis of PDE and the work of statisticians, especially on the MDL principle. However, the rapid expansion of machine learning suggests that it is useful to pay close attention to the foundations and to recognize that the eventual goal of much of the analysis of the PDE, especially in continuum physics, revolves around the question of whether they constitute useful models for phenomena. Further, such introspection is valuable for an expansion of the domain of PDE theory beyond the traditional confines of continuum physics.

The embedding problem is an excellent test case for bridging this divide. The stochastic flows presented here have the following Bayesian interpretation. We view embedding as a process by which an observer makes a copy of a given Riemannian geometry by increasingly accurate estimation of a metric by measurement at finer and finer scales. Stochastic gradient descent is a particular process of measurement chosen to optimize a cost function (several choices are possible as discussed below). One does not usually think of embedding as a stochastic process, so it is worth noting that the source of noise here is not an external thermal source, but simply the errors present in any process of measurement. This viewpoint connects the embedding problem to Bayesian formulations of learning, since we see that the observer 'learns a geometry' by improving their model (the Gaussian measure μ_t) in response to signals from a Gaussian source (white noise on the given manifold $(\mathcal{M}^n, g_\infty)$). This perspective reveals close ties between the embedding theorem and Shannon's channel coding theorem by making explicit the idea that embedding is a form of information transfer from one Gaussian source to another. The embedding process is complete when signals generated by μ_∞ are indistinguishable from those generated by the source. Thus, information theory provides a microscopic view of thermodynamic equilibrium in the embedding problem. (This is a Bayesian description, not an MDL description, because we assume the signal is generated by a 'true' Gaussian source, while the observer tunes their Gaussian measure μ_t in response.)

In a similar manner, we find it useful to accept that despite their distinguished history, it is possible that the equations of continuum mechanics may 'just' be useful models of noisy signals. This means, for example, that we do not assume that the Cauchy problem for the Euler equations constitutes a fundamental description of fluid flow. Instead, we view the existence theory as complete only when all the information in the measurement of a velocity field – which we model by a given mean velocity along with a Gaussian fluctuation – is captured by the probability measure μ_∞. From this viewpoint, we do not see the constructions of De Lellis and Székelyhidi as a demonstration of 'unphysical' fluid behavior. Rather, we see it as an indication of the incompleteness of continuum mechanics as a model for physical phenomena, without the inclusion of the second law of thermodynamics. This is not a negative description, since their results also suggest that the Euler equations do constitute a complete information theoretic description of fluid flow, provided they are augmented by information on fluctuations in addition to a mean velocity (this involves interpreting the Reynolds averaged stress in their work as the fluctuation of a Gaussian field; it remains to establish this rigorously). Traditionally, entropy or viscosity conditions are chosen to augment PDE models and resolve uniqueness of weak solutions. The use of stochastic flows is consistent with these ideas, since it provides a microscopic characterization of thermodynamic equilibrium as a process of information transfer between a source and an observer, just as in the embedding problem.

More generally, as soon as one pays attention to a measurement process, 'thermal effects' appear in every continuum model, not just those that explicitly involve 'temperature' or 'heat flow'. We return to these questions after introducing the model, when the parallels between our approach and prior results in PDE theory can be described in greater detail.

1.3. **Outline.** The main goal of these notes is to introduce a new class of models in a way that allows easy access to numerical computations and rigorous analysis. The notes consist primarily of formal computations, heuristics and links between areas. This is unexplored territory and my purpose is to invite the reader to explore it, since a great deal of work remains. To this end, the notes are structured so that the main structure can be absorbed as easily as possible.

We first introduce the model along with a brief description of the Riemannian geometry of Gaussian measures. The emphasis in this section is to include just enough detail so that the reader understands the structure of the model, along with some numerical schemes for exploration. An important task at this point is to develop fast numerical schemes and emprirical data on the possibility of universality in these problems. This is necessary for a broader appreciation of the richness of the embedding problem, as well as for applications in turbulence and machine learning. In order to compute embeddings with convex integration one must rely on the composition of functions, which is delicate to implement numerically. Despite some beautiful new images of embeddings [2], the absence of numerical computations in more general situations has restricted our imagination. At the same time, the intricate constructions used for existence proofs of wild solutions to the Euler and Navier-Stokes equations are yet to be implemented numerically. In contrast with convex integration, the model presented here requires only semidefinite programming, Markov Chain Monte Carlo (MCMC), and numerical linear algebra. These are the engines of modern scientific computing. My hope is that this will allow the development of fast numerical schemes to compute critical exponents and fine structure in geometry, learning theory and turbulence. The interpretation of the flow as a stochastic interior-point method is of particular interest, both for the purpose of fundamental bounds on algorithms as well as practical schemes.

This is followed by a section describing the geometry of Gaussian measures in finite-dimensions in more detail. Complete proofs are included here since our interest is primarily in low-rank matrices, whereas most of the literature is dedicated to matrices of full rank. This section also allows us to explain the stratification of the space of Gaussian measures in a more convenient coordinate system than has been used in the literature. The geometric mean of two positive semi-definite matrices plays an important role here along with further connections to quantum information theory.

It is only after these sections that we turn to our main interest, the embedding problem. In contrast with Nash's work, codimension plays only a secondary role in our formulation. and we find an interesting interplay between the infinite-dimensional Riemannian geometry of spaces of Gaussian measures and the space of metrics. The Gaussian spaces have positive curvature, whereas the space of metrics has negative curvature. We then extend these ideas to stochastic flows in other gauge groups, in particular the group of diffeomorphisms. Concrete examples include diffeomorphisms of the circle and torus. This allows us to make contact with some classical KAM theorems. Finally, the formalism is applied to the Euler equations.

These later sections are formal. All that we do is to show that these models can be reformulated as stochastic gradient descent. Since the techniques for rigorous analysis require the combination of ideas from different areas of mathematics, these sections should be seen as a road map rooted in the historical development of the embedding problem along with its relation to other areas of mathematics. I have tried to include enough detail so that the plausibility of this approach is clear. The basic ideas are simple enough to admit many variations and there are many more questions than answers.

1.4. **An appreciation of Nash's work.** No one who thinks carefully about the embedding problem can fail to be inspired by Nash's work. These papers reveal entirely unexpected phenomena, while containing a wealth of powerful techniques.

All the work presented here originated in an attempt to rework Nash's papers [41, 42] from the viewpoint of statistical mechanics, in light of the link with turbulence. In our interpretation, the first of these papers [41] makes a compelling case for an entropic expansion of short metrics. Roughly, there are many different choices of iteration that lead to the existence of embeddings with low regularity and entropy is the natural tool to quantify degeneracy in statistical mechanics. The second paper [42] has a very different character, since the emphasis is on the interplay between the codimension of an embedding and its smoothness. Further, the discrete iteration of [41] is replaced by a continuous flow. However, as Nash remarks the main techniques in [42] have less to do with geometry than the use of feedback control methods to solve PDE.

The moral in our work is the same. What we have done is to replace Nash's use of deterministic control with stochastic controls and the idea of gradient descent. This allows us to interpolate between the two papers and separate the problems of existence and regularity of embeddings. Though the resulting structure emerged from a study of the embedding problem and its connections to turbulence, it is not limited to these problems.

With hindsight, it is surprising that this line of attack has not been attempted before. One of the reasons seems to be that the interplay between geometry and probability is very recent. The idea that the space of Gaussian measures has an underlying geometric structure has emerged rather slowly. In particular, the Riemannian structure we use was discovered only after the development of mass transportation theory. At the same time, despite the fundamentally geometric nature of quantum gravity and gauge theories, there were few mathematically rigorous treatments of probability measures on interesting geometries, until the introduction of Schramm-Loewner evolution.

One of the most striking aspects of Nash's first paper [41], is the generality of his results – it applies to *every* compact Riemannian manifold – along with the simplicity of his proof. Our work originated in an attempte to randomize this proof to improve the known critical exponents that separate flexibility and rigidity in this problem. It was only later that we realized that it was possible to unify the treatment of his two papers, rather than viewing them as disparate entities.

The insistence on the use of Gaussian measures follows the work Dyson, Gaudin and Mehta on random matrix theory, as well as the theory of Gaussian measures developed by Gross. A great deal of statistical mechanics is devoted to the study of lattice models. On the other hand, random matrix theory shows that restricting attention to Gaussian measures provides a powerful set of tools to understand the limit behavior for a interesting class of ensembles. Gross' work provides a completely satisfactory theory of infinite-dimensional Gaussian measures, avoiding the use of lattice models altogether. This is why we flow through Gaussian measures to construct probability measures supported on solutions to PDE.

This article concludes with a discussion of the process of discovery of the model. This section is included to reassure analysts that that one can make natural step-by-step modifications of Nash's arguments to arrive at a simple new structure. However, simplicity of formulation does not yet mean simplicity of analysis. This section also introduces the role of concentration estimates. It will be necessary to combine these tools with the older approach of several mathematicians, especially Moser and Nash, to obtain a completely rigorous understanding of the embedding problems based on the viewpoint introduced here.

2. Stochastic gradient descent

2.1. **The isometric empbedding problem.** We assume given an n-dimensional manifold $\mathcal{M}$ with C^∞ charts. The manifold is closed, i.e. it is compact and it has no boundary. In addition to these assumptions on the topological structure of the manifold, we assume that the manifold has a Riemannian metric g. A map $u : \mathcal{M} \to \mathbb{R}^q$ is an *isometric immersion* if the pullback metric equals the given metric, i.e.

$$u^\sharp e = g, \tag{2.1}$$

where e denotes the identity metric $ds^2 = dx_1^2 + \ldots + dx_q^2$ in $\mathbb{R}^q$. We may also write (2.1) in a local system of coordinates as

$$\partial_{x_i} u^\alpha \partial_{x_j} u^\alpha(x) = g_{ij}(x), \quad 1 \le i, j \le n, \quad x \in \mathcal{M} \tag{2.2}$$

We will also write (2.2) as $Du^T(x)Du(x) = g(x)$, $x \in \mathcal{M}$.

Here we adopt the convention that we sum over repeated indices, that we use Greek indices for the coordinates in $\mathbb{R}^q$ and Latin indices for coordinates in $\mathcal{M}$ (thus, u^α are the coordinates of u in $\mathbb{R}^q$ and we sum over the index α above).

The map u is an *isometric embedding* if it is an isometric immersion that is also one to one. We will focus on isometric immersions, in particular on a method to solve the PDE (2.1). The two problems are closely tied to one another.

Modern understanding of this problem begins with Nash's pioneering work in the 1950s [41, 42]. The system of equations (2.2) has $s_n = n(n+1)/2$ equations and q unknowns. Nash's results involve an interplay between the codimension $q - n$ and the smoothness of the metric g. In his 1954 paper, Nash established the existence of C^1 solutions to (2.1) when $q = n + 2$ and g is C^0 [41]. These results extend to the case $q = n + 1$ [33]. This case is overdetermined: the number of unknowns q is smaller than the number of equations s_n when $n \ge 3$. Once $q > n(n+1)/2$ the problem is underdetermined and in his 1956 paper, Nash established high codimension and high regularity solutions. For example, when $g \in C^\infty$, Nash established the existence of C^∞ solutions to (2.1) when q is large [42]. His initial estimates of q were too high. Gromov has shown that $q = n + n(n+1)/2 + 5$ is sufficient (his work is the authoritative reference in the area [28]). The factor of 5 is undoubtedly technical, however the integer $n + n(n+1)/2$ has geometric meaning: it is the maximal dimension of the osculating space at a point $u(x)$, i.e. the space spanned by the vectors $\{\partial_{x_i} u, \partial^2_{x_i x_j} u\}_{1 \le i,j \le n}$. It will play a role below.

There is an extensive PDE literature on (2.1), especially an important result of Günther [29]. We will not review these here since our goal is develop a new approach that is inspired by the connections with turbulence, and thus with statistical mechanics and constructive field theory.

2.2. **A new formalism.** We introduce a stochastic approach to solve (2.1). While this formalism was discovered through laborious calculations, it is simple with hindsight. We rely crucially on Nash's insight that 'short metrics can fluctuate upward' as well as his idea of realizing the embedding as the endpoint of a dynamic process (a discrete dynamical system in [41] and a smooth flow in [42]). However, we replace his analytic techniques, especially his use of explicit geometric constructions to vary the metric, with a stochastic dynamical system.

The "time" in our dynamical system is actually the energy scale, but we will denote it by t as usual. In order to stress the dynamic nature of embedding, let us switch notation slightly and replace g in equation (2.1) with g_∞ to denote that this is our target metric as $t \to \infty$. Our goal will be to construct a stochastic flow u_t of C^∞ maps $\mathcal{M} \to \mathbb{R}^q$ such that $g_t := u_t^\sharp e$ satisfies $g_t < g_\infty$ for $t \in [0, \infty)$ and $\lim_{t\to\infty} g_t = g_\infty$. We will generate such flows in the following way.

(a) We consider the joint evolution of (u_t, L_t) where $u_t : \mathcal{M} \to \mathbb{R}^q$ is a smooth map such that $u_t^\sharp e < g_\infty$ and $L_t \in \mathsf{Pos}_q(\mathcal{M}, g_t)$ is a covariance kernel. Here

$\mathsf{Pos}_q(\mathcal{M}, g)$ denotes the space of covariance kernels for Gaussian random fields $v : (\mathcal{M}, g) \to \mathbb{R}^q$ given in coordinates by $P^{\alpha\beta}(x, y) = \mathbb{E}(v^\alpha(x)v^\beta(y))$, $x, y \in \mathcal{M}$, $1 \le \alpha, \beta \le q$. [1]

(b) We use a gradient flow structure based on the Riemannian geometry of $\mathsf{Pos}_q(\mathcal{M}, g)$ to choose an optimal 'stochastic velocity' $\dot{L}$.

This results in the following stochastic evolution equation.

$$du = \sqrt{dL}, \tag{2.3}$$

$$\dot{L} = \mathrm{argmin}_{P \in (T_L \mathsf{Pos}_q(\mathcal{M}, u^\sharp e))_+} E(u, L, P). \tag{2.4}$$

To the best of my knowledge, the structure (2.3)–(2.4) is new [2].

The notation $\sqrt{dL}$ is suggestive shorthand for the fact that du is a stochastic differential with covariance dL. The precise structure and its consequences are explained in more detail below, as is the Riemannian geometry of $\mathsf{Pos}_q(\mathcal{M}, g)$. We will refer to equation (2.3) as stochastic kinematics and to (2.4) as the energetics. This terminology is adapted from Otto [45]. The term stochastic gradient descent is motivated by its use in the machine learning community, especially Bottou's work [13]. Several variants of stochastic gradient descent have been proposed. We will only use it to refer to equations with the above structure. This model does have an information theoretic interpretation, so this terminology is not inconsistent with its usage in learning.

An important feature of this evolution is that by construction it yields a Lipschitz martingale u_t and an increasing kernel L_t. Therefore the limit u_∞ exists by the martingale convergence theorem, and $\lim_{t\to\infty} L_t = L_\infty$ exists by monotonicity. What needs to be established is that the limit is isometric. In a similar manner, general theory ensures that each $P \in \mathsf{Pos}_q(\mathcal{M}, g)$ is equivalent to a reproducing kernel Hilbert space provided it is sufficiently smooth. Thus, if L_∞ is sufficiently regular, the method yields a limiting space of regular fluctuations in addition to an immersion. Thus, the main 'hard' task is to establish regularity of (u_t, L_t) as well as the convergence to an isometric immersion. Some speculation about this question is contained at the end of this section. For now, the primary task is to explain the structure of this model and to explain why it is natural.

2.3. **Some heuristics.**

2.3.1. *Cellina's differential inclusion.* Equation (2.3)–(2.4) admits a natural extension to other problems. What changes is the geometry of the space of covariance kernels, the cost function E and the nature of subsolutions. One of these problems, which is the first test case of the method, is the following question motivated by Cellina's work on differential inclusions.

How does one construct natural probability measures that are supported on the space of solutions to the differential equation

$$|u'(x)|^2 = 1, \quad x \in (0, 1), \quad u(0) = u(1) = 0\,? \tag{2.5}$$

---

[1] It is possible that the intrinsic Gaussian geometry of $\mathsf{Pos}_q(\mathcal{M})$ is all we need; but right now I also use the metric g to define its geometry, which is why I use $\mathsf{Pos}_q(\mathcal{M}, g)$.

[2] Its closest relative in the literature could be Nelson's stochastic mechanics [43]. However, despite several attempts to link these ideas, I still don't see how to do it. The noise in our model arises from errors in measurement, not from a background field hypothesis. Nevertheless, there are several common themes and it is hard to escape the feeling that the models must be related.

Equation (2.5) has a viscosity solution. This means that uniqueness is determined by a pointwise maximum principle. What we are asking is whether the viscosity solution has a stochastic origin, i.e. whether the probability measures supported on solutions to (2.5) converge to the viscosity solution in a suitable small noise limit. This is what we mean by constructing a 'natural' probability measure on the space of solutions to (2.5).

The importance of (2.5) as a test case is that it begins to reveal a structure that has more to do with general PDE theory than the specific geometric perturbation devices used by Nash. The stochastic flow associated to (2.5) is now a diffusion in the space of subsolutions. These are smooth maps $\{v : (0,1) \to \mathbb{R}\}$ that satisfy the boundary condition $v(0) = v(1) = 0$ as well as the strict inequality $|v'(x)| < 1$ at all $x \in (0,1)$. Given a smooth subsolution and Gaussian kernel (u_0, L_0) our task is to find a stochastic flow (u_t, L_t) in the space of smooth subsolutions that solves

$$du = \sqrt{dL}, \tag{2.6}$$

$$\dot{L} = \mathrm{argmin}_{P \in T_L \mathsf{Pos}(0,1)} E(u, L, P), \tag{2.7}$$

and satisfies $\lim_{t\to\infty} |u_t'(x)| = 1$ for every $x \in (0,1)$. The space $\mathsf{Pos}(0,1)$ is defined in Section (5) below. We see that equations (2.6)–(2.7) have the same character as (2.3)–(2.4), though of course the underlying energy and space of covariance kernels is different.

Thus, our constructive approach to the isometric embedding problem also sheds light on the structure of typical Lipschitz functions on a Riemannian manifold. It is easy to generalize (2.5) and the associated stochastic flow (2.6)–(2.7) to other Hamilton-Jacobi problems with convex Hamiltonians.

The probabilistic structure of the problem is simplest to understand when we consider (2.5). In this case, the set of subsolutions is convex and we may think of it as an infinite-dimensional polytope. This allows us to visualize the stochastic evolution as a growth process that introduces fluctuations in a manner that pushes subsolutions 'outwards' towards the boundary of the polytope. This is a useful heuristic, provided one keeps in mind that it is atypical; in higher-dimensional problems such as (2.1) the set of subsolutions is not (naively) convex. Nevertheless, it illustrates the idea that it is useful to think of PDEs such as (2.1) and (2.5) as an infinite systems of constraints that formally define a polytope. What we are trying to do is to create a diffusion in the space of subsolutions that continues to push outwards until all the constraints are saturated on the boundary of the polytope. This approaches stresses the role of (statistical mechanical) entropy in PDE theory, unlike the conventional emphasis on energy methods. It also provides a useful connection with work in the 1980s that revealed unexpected dynamical structure in fundamental numerical algorithms [6, 7, 15, 22]. As an MCMC scheme, it is useful to think of (2.3)–(2.4) as a stochastic interior point algorithm.

2.3.2. *Controllability and Hypoellipticity.* The main point to note about (2.3)–(2.4) is that the gradient structure determines the covariance of the fluctuations. The problem does not separate into the sum of an applied gradient and stochastic forcing. What we have is a system that evolves only through fluctuations. As we show below, the fact that $\dot{L}$ lies in $T_L \mathsf{Pos}_q(\mathcal{M}, g)$ necessarily means that $\dot{L}$ inherits smoothing properties from L. For this reason, we will say that $\dot{L}$ is bandlimited. One should think of a map u_t that gets rougher with time, but in a manner that is controlled by the balance between the descent of the energy and the expansion of $\dot{L}$. Thus, (2.3)–(2.4) has the character of a Carnot-Caratheodory diffusion with an expanding range of fluctuations. In order to prove that $u_t^\sharp e \to g_\infty$ it will be necessary to prove that for L_0 sufficiently large the flow is hypoelliptic. Similar results – though not exactly what we need– play an important role in control theory, probability and turbulence [14, 39, 30].

2.3.3. *Singular behavior at the boundary.* This model suggests that the simplest physical cartoon for equilibration in turbulence is not the familiar picture of a particle in a potential well subject to thermal noise. This caricature is best illustrated with the Ornstein-Uhlenbeck process

$$dX = -X\,dt + dB = -\nabla V(X)\,dt + dB, \quad X \in \mathbb{R}, \tag{2.8}$$

where we have chosen $V(x) = x^2/2$ for simplicity. Instead, our work suggests that a better paradigm is the Feller diffusion

$$dX = \sqrt{X}\,dB = -\sqrt{\nabla V(X)}\,dB, \quad X > 0 \tag{2.9}$$

Loosely speaking, this reflects the fact that entropic fluctuations give rise to a one-sided transport in the embedding problem and turbulence with singular behavior at an exit boundary. [3]

3. Stochastic kinematics for the embedding problem

We will always assume that $u(x,t) := u_t(x)$ and $L(x,y,t) := L_t(x,y)$ are C^∞ in the spatial variables x,$y \in \mathcal{M}$. Equation (2.3) shows that u_t evolves stochastically in time. However, L_t evolves differentiably in time and $dL = \dot{L}\,dt$ in equation (3.2) is a classical differential. [4]

The formal, but suggestive, notation $\sqrt{dL}$ is used to signify that $\dot{L}$ should be thought of as a *stochastic velocity field*. This notation was introduced by Lévy to describe stochastic differentials [38]. While Lévy's notation is no longer standard, it is very useful to resurrect it in our setting since it allows us to collapse the stochastic kinematics to a single suggestive equation. Its precise meaning is that $u(x,t)$ satisfies the Stratonovich stochastic differential equation

$$du(x,t) = X(x,t,\circ dB), \quad x \in \mathcal{M}, t > 0, \tag{3.1}$$

where X is a Gaussian noise on $C^\infty(\mathcal{M}, \mathbb{R}^q)$ with reproducing kernel $\dot{L}$ [4, 34]. Here $B(t) = (B_1(t), B_2(t), \ldots)$ denotes a sequence of independent standard Brownian motions and dB and $\circ dB$ denote its Itô and Stratonovich differentials respectively.

The Stratonovich formulation is necessary to ensure invariance under coordinate transformations on a manifold. However, to fix ideas it is simpler to think in

---

[3]These diffusions are related in an interesting way [46]. For example, the Feller diffusion (with a drift) may be obtained by projecting Brownian motion in $\mathbb{R}^N$ onto the half-line via $X_t = |B_t|^2$; in a similar way, the map $Y_t = |B_t|$ gives the Bessel process. What is entropic about these processes is the fact that B_t has N degrees of freedom while X_t and Y_t have only one degree of freedom. Entropy is simply the volume factor we obtain when we integrate out the angular degrees of freedom of B_t to determine the evolution of X_t and Y_t. The reason this cartoon is relevant in the embedding problem is that the stochastic process u_t, like B_t, has many more degrees of freedom than $g_t = u_t^\sharp e$ because of gauge transformations. Thus, it is necessary to integrate these out, scale by scale, and that is what we are attempting with a stochastic flow.

[4]Of course, it is necessary to establish this rigorously for given initial data (u_0, L_0). The key step is to establish elliptic regularity for (2.4), since the standard theory of stochastic flows may be used after that.

terms of coordinates and Itô differential equations since this allows us to 'see the fluctuations'. (At the first pass it is best to suppose that $\mathcal{M}$ is the torus $\mathbb{T}^n$ so that we have a global coordinate system. In this setting, one may simply view the formulation (2.3)–(2.4) as a technique for solving the PDE (2.1) using stochastic gradient descent with an Itô SDE).

The relationship between u and $\dot{L}$ is transparent in this setting: [5]

$$\mathbb{E}\left(X^\alpha(x,t,dB)X^\beta(y,t,dB)\right) = dL^{\alpha\beta}(x,y,t), \quad x,y\in\mathcal{M}, 1\le\alpha,\beta\le q. \tag{3.2}$$

As described in Section 8, our use of stochastic flows was at first motivated by Nash's paper on C^1 embeddings [41]. In this paper, Nash constructs immersions by adding fluctuations in physical space. The use of Gaussian fluctuations is a natural extension of this idea. What is more surprising is that the same approach provides a unified treatment of embedding in any codimension and leads naturally into Nash's work on smooth embeddings. Let us explain this idea with the following comparison.

In order to construct smooth embeddings, Nash used a geometric flow of the form [42]

$$\partial_t u(x,t) = v(x,t), \tag{3.3}$$

where $v(x,t)$ is a C^∞ vector field that is normal to an immersion u. That is, we assume (recall that $1\le\alpha\le q$ and that we sum over repeated indices)

$$v^\alpha \partial_{x^i} u^\alpha = 0, \quad x\in\mathcal{M}, \quad 1\le i\le n. \tag{3.4}$$

It immediately follows from (3.3), (3.4) and the definition $g_t = u_t^\sharp e$ that

$$\partial_t g_{ij} = -2\left(v^\alpha \partial^2_{x^i x^j} u^\alpha\right). \tag{3.5}$$

This equation may be viewed as a linear system relating $\partial_t g$ and v. Nash prescribes $\partial_t g$ and chooses the least squares solution for v, assuming q is large enough. A necessary condition for existence of a least squares solution is that v should *not* be normal to the osculating plane (if so, $v^\alpha \partial^2_{x^i x^j} u^\alpha = 0$ for all i,j, and it would be impossible to change the metric). The fact that the change of metric is driven by an interaction between v^α and the second derivatives $\partial^2_{x^i x^j} u^\alpha$ gives his proof a very delicate character (this effect is in addition to his use of a smoothing operator that evolves with u_t). Thus, the situation when v is normal to the osculating plane is the worst case scenario for the deterministic flow (3.3).

By contrast, let us consider the analogous computation for a stochastic flow [6]

$$du = X(dB) = \sqrt{dL}, \tag{3.6}$$

assuming L is supported on spatially smooth vector fields normal to the immersion. The change in metric now includes an Itô correction because the metric is quadratic in Du. We apply Itô's formula to find

$$dg_{ij} = -2\left(X^\alpha(dB)\partial^2_{x^i x^j} u^\alpha\right) + \partial_{x^i} X^\alpha(dB)\partial_{x^j} X^\alpha(dB). \tag{3.7}$$

The Itô correction is a function of $\dot{L}$ alone. Indeed, differentiating the covariance kernel using the definition

$$X^\alpha(x,dB)X^\beta(y,dB) = dL^{\alpha\beta}(x,y), \tag{3.8}$$

[5]The $\mathbb{E}$ is actually redundant here, but is included for clarity since the formal multiplication rule $dB_k dB_l = \delta_{kl}\, dt$ implicitly involves an expectation.

[6]We suppress the arguments (x,t) when possible. So $X(dB) = X(x,t,dB)$.

we find that

$$\partial_{x^i} X^\alpha(dB) \partial_{x^j} X^\alpha(dB) = \partial_{x^i}\partial_{y^j} dL^{\alpha\alpha}(x,y)\big|_{x=y}. \tag{3.9}$$

This linear operator on covariance kernels appears so often that it is useful to introduce notation for it. We define the operator $\lozenge$ that maps symmetric kernels to symmetric $(0,2)$ tensors by [7]

$$(\lozenge P)_{ij}(x) := \partial_{x^i}\partial_{y^j} P^{\alpha\alpha}(x,y)\big|_{x=y}. \tag{3.10}$$

The linear operator $\lozenge$ is analogous to a divergence operator because it involves a first derivative in each component and a trace on the diagonal. I am not sure if it has appeared in the literature (if so, I will correct the terminology).

With these definitions in place, let us summarize our calculations. We have found that a stochastic flow normal to u given by (3.6) yields

$$dg = dm + (\lozenge \dot{L})dt, \tag{3.11}$$

where we have defined the martingale term dm by

$$dm_{ij} := -2\left(X^\alpha(dB)\partial^2_{x^i x^j} u^\alpha\right). \tag{3.12}$$

Equation 3.11 is the stochastic counterpart of equation (3.7). We see immediately that the martingale term dm is the analog of the right hand side of (3.7).

The dimension q has played no role so far in the analysis. Let us now assume that $q > n + n(n+1)/2$ and that it is possible to choose spatially smooth $X(dB)$ that is normal to the osculating plane. This is the worst case scenario for Nash's deterministic flow. However, it is the best case scenario for a stochastic flow, because now $dm \equiv 0$ and equation (3.11) becomes the *deterministic* evolution

$$dg = \lozenge dL, \quad \text{i.e} \quad \partial_t g = \lozenge \dot{L}. \tag{3.13}$$

Thus, we observe a stark change in the evolution of the metric when q crosses the threshold $q = n + n(n+1)/2$. Although u evolves stochastically, if the fluctuations are normal to the osculating plane the metric evolves deterministically. In analogy with problems in phase transitions it is tempting to speculate that this value of q is a threshold that separates exponential decay of $g_\infty - g_t$ from algebraic decay. Since these decay rates are in the energy scale t, they should (I hope) translate to a sharp gap in regularity of immersions constructed by this method. This is one of our reasons for treating the problem of existence for all q in a unified manner.

We will now try to exploit this idea in a weak formulation by choosing an energy minimization principle to determine $\dot{L}$. By adding a term that penalizes gradients it is possible to add smoothing in a relatively conventional manner. More interesting is the fact that equation (2.4) includes a subtle smoothing mechanism that is related to the geometry of $\mathsf{Pos}_q(\mathcal{M}, g)$. Let us first explain this geometry and then describe the energetics.

4. The geometry of $\mathsf{Pos}_q(\mathcal{M}, g)$

The space of covariance kernels $\mathsf{Pos}_q(\mathcal{M}, g)$ has a natural (infinite-dimensional) Riemannian geometry that is used to define the tangent space $T_L\mathsf{Pos}_q(\mathcal{M}, g)$ and the positive cone $T_L\mathsf{Pos}_q(\mathcal{M}, g)_+$. This geometry has not been studied in the generality we need, but it is well known in simpler settings, which allows us to work by analogy. A useful expository account is [10].

[7] This needs to be done properly in an invariant fashion with the Stratonovich formulation.

4.1. **The geometry of positive definite matrices.** The simplest example is the following. Let $\mathsf{Symm}(N)$ and $\mathsf{Pos(N)}$ denote the space of real symmetric and symmetric positive semi-definite (psd) of size N. The natural inner product on $\mathsf{Symm}(N)$ is the Frobenius (or Hilbert-Schmidt) inner product $\langle A, B\rangle = \mathrm{Tr}(A^T B)$. When this inner product is pushed forward to $\mathsf{Pos(N)}$ by the map $X \mapsto \exp X$ it yields the Riemannian structure of $\mathsf{Pos(N)}$ as a symmetric space [9, 35]. At any $L \in \mathsf{Pos(N)}$ we find that

$$T_L\mathsf{Pos(N)} = \{M \,|M = LA + AL, \quad A \in \mathsf{Symm}(N)\,\}. \tag{4.1}$$

Further, the inner product on $T_L\mathsf{Pos(N)}$ is given by

$$\langle M_A, M_B\rangle_L = \mathrm{Tr}(ALB), \quad M_A = LA + AL, \quad M_B = LB + BL. \tag{4.2}$$

In a basis in which L is diagonal with positive eigenvalues $\{l_i\}_{i=1}^N$, this may also be written as

$$\langle M_A, M_B\rangle_L = \sum_{i,j} \frac{l_i}{(l_i + l_j)^2}(M_A)_{ij}(M_B)_{ij}, \quad M_A, M_B \in T_L\mathsf{Pos(N)}. \tag{4.3}$$

This metric on $\mathsf{Pos(N)}$ is canonical because it may also be obtained in several other ways. Algebraically, $\mathsf{Pos(N)}$ is the quotient space $GL(N)/O(N)$, and the above inner product is inherited by Riemannian submersion. Thus, it is the only metric that respects (discrete) gauge invariance.

More relevant for our needs is the following probabilistic interpretation. Consider a centered Gaussian vector $v = (v_1, \dots, v_N) \in \mathbb{R}^N$. The covariance kernel for v is the matrix $L_{ij} = \mathbb{E}(v_i v_j) \in \mathsf{Pos(N)}$. This matrix determines (and is equivalent to) a Gaussian measure on $\mathbb{R}^N$. A natural metric on probability measures on $\mathbb{R}^N$ is the Wasserstein-2 transport metric. It turns out that the inner-product (4.2) may also be obtained by restricting the Wasserstein-2 metric to Gaussian measures [10].

The above formulas continue to hold when L is positive semi-definite. The following analogy is helpful to understand the smoothing that is implicit in (2.4). We have seen that the tangent space $T_L\mathsf{Pos(N)}$ is the image of the linear map

$$\mathsf{Symm}(N) \to T_L\mathsf{Pos(N)}, \quad A \mapsto AL + LA. \tag{4.4}$$

Now suppose that L has finite rank. Choosing a basis in which L is diagonal, say $L_{ii} = l_i$, $i = 1, \dots, r$ and $l_i = 0$, $r < i \le N$, we find that each matrix $M \in T_L\mathsf{Pos(N)}$ has coordinates

$$M_{ij} = (l_i + l_j)\, A_{ij},$$

so that $M_{ij} = 0$ when i and j are greater than r. Thus, a rank restriction on L imposes restrictions on the tangent space $T_L\mathsf{Pos(N)}$.

While there are subtleties in extending these ideas to the operator theoretic setting of (2.4), the above connection with the Wasserstein-2 metric provides a natural way to understand the Riemannian geometry of $\mathsf{Pos}_q(\mathcal{M}, g)$ in the setting of (2.4). While the precise results we need do not exist in the literature, there has been some related work, though mainly in finite-dimensions [47].

4.2. **The geometry of bandlimited Gaussian kernels.** Let us briefly explain how the finite-dimensional calculations extend to infinite dimensions when $q = 1$, since this captures the essence of the situation. These ideas may then be extended to $\mathbb{R}^q$ valued Gaussian random fields for arbitrary q. Assume given $(\mathcal{M}, g)$, consider the fixed reference space $L^2(\mathcal{M}, g)$ and let $\mathsf{Symm}(\mathcal{M}, g)$ denote the space of Hilbert-Schmidt operators on $L^2(\mathcal{M}, g)$. Now consider a centered scalar Gaussian field

$v : (\mathcal{M}, g) \to \mathbb{R}$ with covariance $L(x, y) = \mathbb{E}(v(x)v(y))$. In this case, the formal tangent space $T_L\mathsf{Pos}(\mathcal{M}, g)$ has the form

$$T_L\mathsf{Pos}(\mathcal{M}, g) = \{M \,|M = LA + AL, \quad A \in \mathsf{Symm}(\mathcal{M}, g)\,\}, \tag{4.5}$$

where AL and LA denote composition of operators in the obvious way

$$(LA)(x, y) = \int_{\mathcal{M}} L(x, s)A(s, y)\,\sqrt{G}ds, \tag{4.6}$$

where $G = \det(g)$ so that $\sqrt{G}ds$ is the volume form defined by the metric g.

It is at this point that the inner-product defined by L begins to play a role. If $L(x, y)$ is sufficiently smooth in x and y, it defines a trace-class operator on $L^2(\mathcal{M}, g)$. By Mercer's theorem, it has a complete eigenfunction expansion and by Lidskii's theorem $\mathrm{Tr}L = \sum_i l_i < \infty$, where $l_i \geq 0$ denote the eigenvalues of L. We now see that in infinite-dimensions the inner-product (4.3) becomes a weighted norm defined by the spectrum of L. More precisely, the tangent space $T_L\mathsf{Pos}_q(\mathcal{M}, g)$ is the Hilbert space obtained by closing the linear space (4.5) in the norm

$$\langle M, M\rangle_L := \sum_{i,j} \frac{l_i}{(l_i + l_j)^2} M_{ij}^2. \tag{4.7}$$

Here M_{ij} denote the coordinates of M in the basis of eigenfunctions of L.

The case $q > 1$ is slightly more complicated because the composition rule (4.6) must account for the fact that L and A act on vector valued functions in $L^2(\mathcal{M}, g; \mathbb{R}^q)$. However, this does not affect the basic observation that $T_L\mathsf{Pos}_q(\mathcal{M}, g)$ carries a weighted norm that is determined by L.

5. Stochastic gradient descent for Cellina's problem

We first explain the stochastic kinematics and energetic for the simplified problem (2.5) and the associated stochastic flows (2.6)–(2.7).

The stochastic velocity fields in this problem are covariance kernels for scalar Gaussian fields on $(0, 1)$ and at any time t we have

$$dL_t(x, y) = X(x, t, dB)X(y, t, dB), \quad x, y \in (0, 1). \tag{5.1}$$

Let $\mathsf{Pos}(0, 1)$ denote the space of such kernels. More generally, let $\mathsf{Symm}(0, 1)$ denote the space of symmetric Hilbert-Schmidt kernels on $(0, 1)$ with Lebesgue measure. Given a smooth positive kernel L, we let $\{l_i\}_{i=1}^\infty$ denote its eigenvalues and $\{\varphi_i\}_{i=1}^\infty$ its eigenfunctions. Then every kernel in $T_L\mathsf{Pos}$ is of the form

$$M = AL + LA, \quad A \in \mathsf{Symm}(0, 1) \tag{5.2}$$

subject to the condition that the squared norm

$$\langle M, M\rangle_L = \sum_i \frac{l_i}{(l_i + l_j)^2} M_{ij}^2 < \infty, \tag{5.3}$$

where the coefficients M_{ij} are obtained using the eigenfunctions of L

$$M_{ij} = \int_0^1 \int_0^1 M(x, y)\varphi_i(x)\varphi_j(y)\,dx\,dy. \tag{5.4}$$

Now let us turn to the energetics. In what follows we will assume that we choose a kernel $P \in T_L\mathsf{Pos}(0, 1)$ as a candidate for $\dot{L}$. We finally choose $\dot{L}$ by optimizing over all such P. For brevity, let $\rho_t(x) := u'(x,t)^2$. An application of Itô's formula analogous to equations (3.7)–(3.10) yields

$$d\rho = 2u'du' + du'du' := dm + \Diamond P\, dt. \tag{5.5}$$

In this setting, the $\Diamond$ operator is a linear map from $\mathsf{Symm}(0,1)$ to functions on the interval $(0,1)$ given by

$$(\Diamond P)(x) := \left. \partial_x \partial_y P(x,y)\right|_{x=y}, \tag{5.6}$$

The $\Diamond$ operator is positivity preserving in the sense that if P is the smooth covariance kernel of a Gaussian process, $\Diamond P$ is a non-negative density on $(0,1)$. In a similar manner, the quadratic variation of the martingale dm is given in terms of $\Diamond P$ as follows

$$[dm, dm] := (2u'du')(2u'du') = 4\rho \Diamond P\, dt. \tag{5.7}$$

Our task is to use the stochastic flow to drive $\rho_t(x)$ to 1 in a manner that is constrained by the geometry of $\mathsf{Pos}(0,1)$. If we had no restrictions, a natural way to force $\rho_t(x)$ to approach 1 would be to impose the condition

$$\partial_t \rho_t(x) = 1 - \rho_t(x), \quad \text{or} \quad d\rho_t(x) = (1-\rho_t(x))\, dt. \tag{5.8}$$

This is our 'ideal' evolution. On the other hand, equation (5.5) denotes the evolution that is accessible with stochastic kinematics. So we now try to do the best we can subject to the kinematic constraints. This means that we'd like to match the means and minimize the fluctuations, i.e.

$$(1-\rho_t(x)) - \Diamond P) \approx 0, \quad [dm, dm] \approx 0. \tag{5.9}$$

For these reasons, we choose the cost function

$$E(\rho, P) = \int_0^1 \left((1-\rho(x) - \Diamond P\right)^2 dx + 4\int_0^1 \rho\, \Diamond P\, dx \tag{5.10}$$

We have penalized the terms in (5.9) equally. Clearly, there are many other possibilities here – we may choose different weights, different ideal evolutions that replace (5.8), etc. It is also possible to incorporate a smoothing term by minimizing the quadratic variation of du''. Let us define

$$(\Diamond^2 P)(x) := \left. \partial_x^2 \partial_y^2 P(x,y)\right|_{x=y}. \tag{5.11}$$

Then noting that

$$du''du'' = \Diamond^2 P, \tag{5.12}$$

we fix a parameter $\varepsilon > 0$ and consider the energy

$$E_\varepsilon(\rho, P) = \int_0^1 \left((1-\rho(x) - \Diamond P\right)^2 dx + 4\int_0^1 \rho \Diamond P\, dx + \varepsilon \int_0^1 (\Diamond^2 P)\, dx. \tag{5.13}$$

Finally, the evolution is prescribed by setting

$$\dot{L} = \mathrm{argmin}_{P \in T_L \mathsf{Pos}(0,1)_+} E_\varepsilon(\rho, P). \tag{5.14}$$

The cone $T_L\mathsf{Pos}(0,1)_+$ is convex and the energy E_ε is strictly convex on this set. Thus, there is a unique minimizer $\dot{L}$.

6. Energetics for the embedding problem

6.1. **Scaling the metric defect.** Let us now extend the ideas of Section 5 to the embedding problem. While the main ideas are roughly the same, we will explore two different 'ideal' evolutions. The first is the analog of (5.8)

$$\partial_t g_t = g_\infty - g_t, \quad \text{or} \quad dg_t = (g_\infty - g_t)\, dt. \tag{6.1}$$

If this equation held, we would have exponential decay of the metric defect $g_\infty - g_t = e^{-t}(g_\infty - g_0)$.

The second ideal evolution is a gradient flow on the space of metrics $\mathsf{Met}(\mathcal{M})$ on $\mathcal{M}$. In order to describe this gradient flow we must introduce the Riemannian structure of $\mathsf{Met}(\mathcal{M})$. The formal tangent space $T_g\mathsf{Met}(\mathcal{M})$ at a metric g consists of symmetric $(0,2)$ tensors. If $h, h' \in T_g\mathsf{Met}(\mathcal{M})$ the natural inner-product on $T_g\mathsf{Met}(\mathcal{M})$ is

$$\langle h, h' \rangle_g = \int_{\mathcal{M}} g^{ij} g^{kl} h_{ik} h'_{jl} \sqrt{G}, \tag{6.2}$$

where $G = \det(g)$. [8] We need this inner-product to form the Dirichlet energy that is the analog of (5.10).

We have already seen in equation (3.11) that if $\dot{L} = P \in T_L\mathsf{Pos}_q(\mathcal{M}, g)$, stochastic kinematics restrict us to

$$dg = dm(P) + \lozenge P\, dt \tag{6.3}$$

where the components of the martingale term are[9]

$$dm_{ij} = \partial_{x^i} X^\alpha(dB) \partial_{x^j} u^\alpha + \partial_{x^j} X^\alpha(dB) \partial_{x^i} u^\alpha. \tag{6.4}$$

Therefore, its quadratic variation is

$$\begin{aligned} E_0(P) &:= \langle dm(P), dm(P) \rangle_g = \int_{\mathcal{M}} g^{ij} g^{kl} dm_{ik} dm_{jl} \sqrt{G} \\ &= 2 \int_{\mathcal{M}} g^{ij} g^{kl} \partial_{x^j} u^\alpha \partial_{x^l} u^\beta \partial_{x^i} \partial_{y^k}\, P^{\alpha\beta}(x,y)\big|_{x=y} \sqrt{G}. \end{aligned} \tag{6.5}$$

In a similar manner, the natural penalty for enforcing $\lozenge P \approx g_\infty - g$ is

$$E_1(P) = \langle g_\infty - g - \lozenge P, g_\infty - g - \lozenge P \rangle_g \tag{6.6}$$

Thus, the Dirichlet energy that is the analog of (5.10) is $E(P) = E_0(P) + E_1(P)$. In a similar manner, we may also incorporate a penalty for smoothing. Again, these give strictly convex energies on the cone $T_L(\mathsf{Pos}_q(\mathcal{M}, g))$.

6.2. **An energy function and ideal gradient flow for metrics.** The evolution in equation (6.1) is not entirely satisfactory since it is not apparent that it corresponds to a gradient flow. Interestingly, there is a natural 'ideal' gradient flow but

[8] I am following the physicists here [26]; this must be completely standard in mathematics too.

[9] We do not want to assume that X^α is normal to the map u a priori; this should emerge from energy minimization. This is why we do not write this term in the form of equation (3.12).

it does not yield (6.1) [10]. Let us define the following energy functionals on $\mathsf{Met}(\mathcal{M})$:

$$\mathrm{Vol}(g) = \int_{\mathcal{M}} \sqrt{G} \tag{6.7}$$

$$A(g) = -\int_{\mathcal{M}} g^{ab}(g_\infty)_{ab}\sqrt{G_\infty}, \tag{6.8}$$

$$F(g) = A(g) - 2\mathrm{Vol}(g). \tag{6.9}$$

where $G = \det g$ and $G_\infty = \det(g_\infty)$.

The terms in this energy have an intuitive meaning: clearly the volume is monotonic under the order on metrics. The energy $A(g)$ may be thought of as an applied field that drive the metric g_t to g_∞. So it is a bit like the quadratic energy for the Ornstein-Uhlenbeck process, except that it is more subtle because of the nonlinearity in the metric, as well as the one-sided condition $g < g_\infty$.

Computing the gradient of F with respect to the metric $\langle\cdot,\cdot\rangle_g$ on $\mathsf{Met}(\mathcal{M})$ we find the gradient flow

$$\partial_t g = \mathrm{grad}_g A(g) = (g_\infty - g) + \left(\sqrt{\frac{G_\infty}{G}} - 1\right) g_\infty. \tag{6.10}$$

This flow has the desired positivity property $\mathrm{grad}_g A(g) > 0$ in the cone $g_\infty > g$. Further, $\mathrm{grad}_g A(g) > (g_\infty - g)$ in the region $0 < g < g_\infty$. Thus, we may use it as our 'ideal evolutions' instead of the flow chosen in (6.1) by choosing the associated energy penalty and Dirichlet energy

$$E_2(P) = \langle \mathrm{grad}_g A(g) - \Diamond P, \mathrm{grad}_g A(g) - \Diamond P\rangle_g, \quad \tilde{E}(P) = E_0(P) + E_2(P). \tag{6.11}$$

The computation of (6.10) goes as follows. Consider a curve $g(\tau)$ with $g(0) = g$ and $\dot{g}(0) = h$. Then we find that

$$\dot{g}^{ij}\big|_{\tau=0} = -g^{ip}h_{pq}g^{qj}, \quad \frac{d}{d\tau}\sqrt{G}\bigg|_{\tau=0} = \frac{1}{2}g^{ab}h_{ab}\sqrt{G}. \tag{6.12}$$

Therefore, we find that

$$\langle \mathrm{grad}_g A(g), h\rangle_g = \frac{d}{d\tau}A(g(\tau))\bigg|_{\tau=0} = \int_{\mathcal{M}} g^{ak}g^{bl}h_{kl}(g_\infty)_{ab}\sqrt{G_\infty} \tag{6.13}$$

$$\langle \mathrm{grad}_g \mathrm{Vol}(g), h\rangle_g = \frac{d}{d\tau}\mathrm{Vol}(g(\tau))\bigg|_{\tau=0} = \frac{1}{2}\int_{\mathcal{M}} g^{ab}h_{ab}\sqrt{G}. \tag{6.14}$$

We now use the definition of the inner-product in (6.2) and the fact that h is an arbitrary smooth, symmetric $(0,2)$ tensor to obtain the pointwise relations

$$g^{\alpha\beta}g^{\gamma\delta}\left(\mathrm{grad}_g A(g)\right)_{\alpha\gamma} h_{\beta\delta} = g^{ak}g^{bl}h_{kl}(g_\infty)_{ab}\sqrt{G_\infty}, \tag{6.15}$$

$$g^{\alpha\beta}g^{\gamma\delta}\left(\mathrm{grad}_g \mathrm{Vol}(g)\right)_{\alpha\gamma} h_{\beta\delta} = \frac{1}{2}g^{ab}h_{ab}\sqrt{G}. \tag{6.16}$$

Contracting indices and using the fact that h is arbitrary, we obtain

$$\mathrm{grad}_g A(g) = \sqrt{\frac{G_\infty}{G}}g_\infty, \quad \mathrm{grad}_g \mathrm{Vol}(g) = \frac{1}{2}g. \tag{6.17}$$

[10] This may just be because I'm not clever enough to find a primitive that yields (6.1).

7. Intrinsic fluctuations and turbulence

In this section, we formulate a model that leads us closer to a conceptual link between the embedding problem and turbulence. As we have seen in Section 3, the use of stochastic flows allows us to treat embeddings into all $\mathbb{R}^q$ in a unified manner. We now take this idea to its natural extreme and seek a completely intrinsic problem that shares the character of the embedding problem. There are several other reasons – pure and applied – for considering such problems. These are discussed at the end of this section once the model has been introduced.

7.1. **The model.** Given a Riemannian manifold $(\mathcal{M}, g_\infty)$, and a metric g_0 such that $g_0 < g_\infty$ we seek a stochastic flow $\{\varphi_t\}_{t\geq 0}$, $\varphi_0 = I$ of diffeomorphisms such that the pushforwards $g_t := \varphi_t^* g_0$ converge to g_∞. Such diffeomorphisms do not constitute 'true' changes of the metric, since they are simply gauge transformations. However, a systematic study of gauge transformations is interesting because it leads us towards the study of Brownian motions in the diffeomorphism group. It also leads us to pay attention to the entropy implicit in the choice of a gauge. The assumption that $g_0 < g_\infty$ is of course inspired by Nash's observation that short metrics can fluctuate 'upward'. Our purpose is to show that it is a meaningful starting point even in a purely intrinsic setting.

Our task as in Section 2.2 is to derive a stochastic gradient descent for this problem. In equations (2.3)–(2.4) our stochastic vector fields took values in $\mathbb{R}^q$ and were completely prescribed by covariance kernels for maps $\mathcal{M} \to \mathbb{R}^q$. Now they must take values in the tangent bundle $T\mathcal{M}$. While it is a somewhat subtle problem to define Gaussian measures on vector bundles in a purely intrinsic way, a fully satisfactory theory exists [3]. (For example, it is not obvious what the covariance kernel must be, since we need to compare vectors $v(x)$ and $v(y)$ in the tangent spaces $T_x\mathcal{M}$ and $T_y\mathcal{M}$ respectively.)

In fact, we will make life easier for ourselves and restrict attention to a simpler class of vector fields – gradients of scalars. A smooth scalar Gaussian scalar field $p : \mathcal{M} \to \mathbb{R}$ is completely prescribed by its correlation kernel

$$L(x,y) = \mathbb{E}\left(p(x)p(y)\right). \tag{7.1}$$

Since $(\mathcal{M}, g)$ is a Riemannian manifold, if L is sufficiently smooth, so is every realization p and we obtain a well-defined vector field $\mathrm{grad}_g p(x) \in T_x\mathcal{M}$. We use the letter p to signify that this is a pressure field and the subscript g to denote the metric being used for the gradient.

Now let us extend this notion to an evolving family of covariance kernels and the associated stochastic flows. As in the previous sections, in order to state the minimization for stochastic gradient descent, we will first let $P \in T_L\mathsf{Pos}$ denote the covariance kernel for the pressure field. We study the stochastic kinematics and then choose the stochastic velocity $\dot{L}$ by optimizing over P (we abuse notation and continue to call $\dot{L}_t$ the stochastic velocity field, though it is actually the covariance kernel for the pressure field). Thus, consider a spatially smooth Gaussian noise

$$p(x,t,dB)p(y,t,dB) = P(x,y)\,dt, \quad x,y \in \mathcal{M}, \tag{7.2}$$

and the associated stochastic vector field

$$X(x,t,dB) = \mathrm{grad}_g p(x,t,dB) = g^{ij}\partial_{x^j} p(x,t,dB)\partial_{x^i}, \tag{7.3}$$

where the last formula expresses X in coordinates.

Given a spatially smooth stochastic flow, the evolution of the metric is given in the Itô form by [34]

$$dg = \mathcal{L}_{X(dB)}g + \frac{1}{2}\mathcal{L}^2_{X(dB)}g, \tag{7.4}$$

where $\mathcal{L}$ denotes the Lie derivative. The martingale term analogous to (3.12) is given by

$$dm(P) := \mathcal{L}_{X(dB)}g = \mathcal{L}_{\mathrm{grad}_g p(dB)}g = 2\mathrm{Hess}_g\left(p(dB)\right). \tag{7.5}$$

Here we use an identity relating the Lie derivative of the metric to the covariant derivative (Y, Z are arbitrary smooth vector fields in TM) [27]

$$\mathcal{L}_X g(Y,Z) = g(D_Y X, Z) + g(Y, D_X Z). \tag{7.6}$$

Then we further simplify this identity using the fact that $X = \mathrm{grad}_g p$. (The fact that X is a stochastic differential is irrelevant for this calculation; all that matters is that it is smooth in space).

A tedious, but similar calculation, provides the Itô correction to the metric

$$\frac{1}{2}\mathcal{L}^2_{X(dB)}g = 4\left(\mathrm{Hess}_g(p(dB))\right)^2. \tag{7.7}$$

This term is deterministic and its positivity may be seen as follows. Diagonalize the kernel P and write $p(x,dB) = \sum_k p_k(x)dB_k$. Since Hess_g is a linear operation, $\mathrm{Hess}_g(p(x,dB)) = \sum_k \mathrm{Hess}_g(p_k(x))dB_k$, and by the Itô multiplication rule $dB_j dB_k = \delta_{jk}\,dt$, we see that

$$\left(\mathrm{Hess}_g(p(x,dB))\right)^2 = \sum_j \left(\mathrm{Hess}_g(p_j(x)\right)^2\,dt. \tag{7.8}$$

In particular, the Itô correction for the stochastic flows generated by pressure gradients pushes the metric 'upwards' as desired. The above formula again provides a positivity preserving map from the space of covariance kernels to metrics analogous to (3.10). Since it involves second derivatives, we define

$$\lozenge^2 P(x) := 4\left(\mathrm{Hess}_g(p(x,dB))\right)^2, \quad p(x,t,dB)p(y,t,dB) = P(x,y)\,dt. \tag{7.9}$$

(We define the operation in this form, since it is clearer than when written in coordinates).

We now observe that the energies defined in Section 6 may continue to be used here. More precisely, as in equation (6.5) we consider the quadratic variation of the martingale term

$$E_0(g,P) = \langle dm(P), dm(P)\rangle_g = \int_{\mathcal{M}} g^{ij}g^{kl}dm_{ik}dm_{jl}\sqrt{G} \tag{7.10}$$

where the martingale term is now given by (7.5). In a similar manner, we may define the energy

$$E_1(g,P) = \langle g_\infty - g - \lozenge^2 P, g_\infty - g - \lozenge^2 P\rangle_g \tag{7.11}$$

so that the mean change in metric stays close to $\partial_t g = g_\infty - g$. In a completely analogous manner to (6.11), we may also consider

$$E_2(g,P) = \langle \mathrm{grad}_g A(g) - \lozenge^2 P, \mathrm{grad}_g A(g) - \lozenge^2 P\rangle_g \tag{7.12}$$

Thus, choosing $E = E_0 + E_1$ or $E_0 + E_2$ we find our desired stochastic gradient descent:

$$d\varphi = \sqrt{dL} \tag{7.13}$$

$$\dot{L} = \mathrm{argmin}_{P \in T\mathsf{Pos}(\mathcal{M},g)} E(g,P) \tag{7.14}$$

Observe that once L is chosen by the minimization rule (7.14), the evolution of the metric is completely determined by the kinematic relation (7.4).

7.2. **Some remarks on the model.**

7.2.1. *From embedding theorems to KAM theory.* This model is introduced as a 'proof of concept' that the use of Gaussian processes to introduce fluctuations may be used to formulate *intrinsic* problems of equilibration that share a conceptual and technical similarity with Nash's work. Of course, an important such example is turbulence and our primary goal in introducing fluctuations in the above model is to make it possible to formulate turbulence (as modeled by the Euler equations) from a similar point of view.

More generally, our purpose here is to show how one can 'move' in a targeted way within the diffeomorphism group with stochastic flows. This question is of interest even in the simplest setting of the diffeomorphism groups of the circle and tori. Despite the historical ties between the embedding problems and KAM theory, I am unaware of constructions of circle maps that truly rely on Nash's work. This is another 1-D model problem which is an important test of our technique.

7.2.2. *Intrinsic constructions of Brownian motion.* The background to this model is the intrinsic construction of Brownian motion on $(\mathcal{M}, g)$ by Eells, Elworthy and Malliavin [24, 39, 31]. Baxendale used these ideas to define a notion of Brownian motion in the diffeomorphism group. However, his approach requires restrictive smoothness assumptions. It is natural to ask if it is possible to layer stochastic flows using stochastic gradient descent to yield natural notions of Brownian motion in the diffeomorphism group with critical regularity. While Nash's embedding theorems have been used for (extrinsic) constructions of Brownian motion on manifolds, it appears not to have been recognized that the intrinsic constructions of Brownian motion may themselves be used to shed light on the embedding theorems.

7.2.3. *Geodesic flows in infinite-dimensional Lie groups.* Geodesic flows on gauge groups arise in several applications [1, 40]). These gauge groups are typically treated from the point of view of their tangent spaces – for example, the (formal) tangent space to the group of diffeomorphisms is the space of divergence-free vector fields. The theory of stochastic flows [5, 34] provides a good description of Hilbert subgroups of these gauge groups. (A typical theorem is of the following kind: if one considers an RKHS that is supported on C^∞ divergence-free vector fields on $\mathbb{T}^n$, then the stochastic flow it generates fills out the Hilbert Lie group obtained by taking the iterated commutators of all vector fields in the RKHS). Thus, it should be possible to formulate stochastic gradient descents within these gauge groups that 'layer' C^∞ stochastic flows in a manner that is completely analogous to (2.3)–(2.4).

While we have focused on the diffeomorphism group for now, it is natural to ask if the general technique introduced here cannot be used in other contexts. The approach is again the same: we avoid the use of all explicit geometric constructions and simply use general control theoretic principles to drive us towards a desired geometry. Good test cases to consider are conformal maps and symplectic transformations. I have not yet studied this problem carefully and a sticking point may be the analog of (7.8).

7.2.4. *Learning a metric.* The above problem is also motivated by Bayesian models of learning. The gauge transformations φ_t should be interpreted as a process of calibrating a metric by an intrinsic observer, i.e. an observer restricted to the space $\mathcal{M}$. Roughly, g_0 corresponds to an initial guess about the 'true' metric g_∞ and the initial kernel L_0 corresponds to the possible modes of exploration by the observer. The pushforwards $g_t = \varphi_t^* g_0$ correspond to improved estimates of the 'true' metric g_∞ in response to measurements at finer and finer scales. In the same vein, all our models for the embedding problem of $(\mathcal{M}, g)$ into $\mathbb{R}^q$ have a Bayesian interpretation: these correspond to learning a metric by an *external* observer in $\mathbb{R}^q$. It seems natural to first understand the intrinsic problem in the above setting, since this involves minimal complexity (no codimension, scalar fluctuation fields).

Fast numerical algorithms are of great importance in machine learning. In principle, the flows presented here may be numerically implemented by discretizing Riemannian manifolds by metric graphs. There are no analytical subtleties in this setting and all our choices of gradient structure and energy give rise to stochastic gradient descent algorithms. The real question though is which of these correspond to fast and reliable algorithms.

8. Discovering a model

8.1. **Outline.** The tone of this part of the notes is conversational, since it is primarily a discussion of wandering that finally led to something interesting. The primary audience here are analysts familiar with the search for critical exponents in the Euler equations [21]. However, I also include a schematic description of Nash's 1954 proof of low-regularity embeddings for those unfamiliar with it because this proof provides valuable insight into the nature of fluctuations. Further, it shows immediately the need for concentration estimates (though these appear in a cumbersome way in the iterative scheme and are simpler when one has a stochastic flow).

The route to (2.3)–(2.4) involved the following 'locally natural' steps:

(1) Randomizing Nash 54 in discrete time along with concentration estimates.
(2) The use of heat kernels to remove the local assumption in Nash 54.
(3) The switch from an iterative scheme to stochastic flows.
(4) The treatment of smoothing as part of the unknowns.
(5) Search for a gradient flow structure.

The overall effect, however, was that my initial idea of using randomness to improve the critical exponent morphed into a completely different class of problems.

Steps (4) and (5) took me a rather long time because I explored the use of a gradient flow through operators (i.e. flowing Green's functions and/or Laplacians). This is natural from the perspective of random matrix theory, but seems complicated with hindsight. Similarly, the calculations of Section 3 reveal the kinematic advantages of using stochastic flows, so there is no reason to review it again.

What I will focus on in detail is the first two steps, since these stay closest to h-principles for the Euler equation and are likely to be of most relevance to others. Of particular importance is the use of an 'auxiliary Gaussian space' to introduce fluctuations. This idea stems from an explicitly probabilistic interpretation of the

Bérard-Besson-Gallot embedding and it reveals quite clearly how one can use stochastic methods to construct fluctuation fields with good error estimates. This step was important to me because it was the first point at which probabilistic methods felt more natural than Nash's constructions.

8.2. **Randomizing Nash's scheme: low codimension.** My initial goal was to improve the exponent α for $C^{1,\alpha}$ embeddings of a closed n dimensional manifold $\mathcal{M}$ into $\mathbb{R}^q$ when $q = n+1$ using a randomization scheme. My interest in this question was to extend the results in [17], which contain the best general bound on the local and global exponent α for a $C^{1,\alpha}$ isometric embedding of an n-dimensional manifold $\mathcal{M}$ into $\mathbb{R}^{n+1}$, $q = n+1$. We say the bound is local when the manifold has a single chart. In this case [17, Thm.1], the result is

$$\alpha = \frac{1}{1+2s_n}, \quad s_n = \frac{n(n+1)}{2}. \tag{8.1}$$

A clever simplicial decomposition used by Nash may be combined with the local exponent to yield $\alpha = 1/(1+2ns_n)$ [17, Thm.2] for a closed manifold. This exponent is determined by a balance between the iterative addition of oscillations via Nash's scheme and the requirement that the iterates $u_n : \mathcal{M} \to \mathbb{R}^q$ converge in $C^1(\mathcal{M})$.

Let us briefly review Nash's proof to explain why randomization is a reasonable approach to improve the exponent. We will only consider the local bound, since the main difficulties are contained here. Further, we assume that $g \in C^\infty$ and $q = n+2$ instead of $q = n+1$ since the main issues are present in this case and the calculations are easier.

The intuitive picture behind Nash's proof is simple: imagine that we wanted to embed the two-dimensional torus $\mathbb{T}^2$ into $\mathbb{R}^3$. We know that we can always immerse $\mathbb{T}^2$ in $\mathbb{R}^3$ (this is a topological fact, which is nontrivial for a general manifold). By rescaling $\mathbb{R}^3$ we can ensure that any immersion $u : \mathbb{T}^2 \to \mathbb{R}^3$ has the property that the length between any two points $u(x)$ and $u(y)$ (measured along the immersion in $\mathbb{R}^3$) is strictly less than the distance between x and y (measured in $\mathbb{T}^2$). What the Nash-Kuiper scheme does is to 'inflate' the immersion u by adding fine-scale oscillations in a highly structured manner. Images of this process may be found in [2, 12].

Let us make this more precise. We say that a metric g' is short with respect to a metric g if $g'(x) < g(x)$ in the sense of quadratic forms; i.e. $g'(x)(v,v) < g(x)(v,v)$ for every $x \in \mathcal{M}$ and every $v \in T_x\mathcal{M}$. This is equivalent to the fact that the geodesic distance between any two points x and y in these two metrics is strictly ordered, i.e. $d_{g'}(x,y) < d_g(x,y)$ for every pair of points $x, y \in \mathcal{M}$. In a similar manner, we may order immersions and embeddings in $\mathbb{R}^q$. An immersion $u : \mathcal{M} \to \mathbb{R}^q$ induces the metric $u^\sharp e$ on $\mathcal{M}$, where e denotes the standard metric on $\mathbb{R}^q$. The starting point of Nash's proof is a *short* immersion $u : \mathcal{M} \to \mathbb{R}^q$ such that $u^\sharp e < g$. As we have seen in the case of $\mathbb{T}^2$, if there are no topological obstructions, we always have an initial condition for the iteration.

The iteration scheme consists of an outer loop (termed stages, indexed by m) and an inner loop (termed steps, with s_n steps, indexed by k). At the beginning of the m-th stage we are given a short immersion u_m such that the metric defect $g - u^\sharp e$ is small. At the end of the s_n steps in the m-th stage, we will obtain a new immersion u_{m+1} that is still short with respect to g, but has expanded u_m in the sense that $u_m^\sharp e < u_{m+1}^\sharp e$. In order to obtain convergence in C^1, we need error estimates that ensure that u_{m+1} is short and quantify just how much smaller than $g - u_m^\sharp e$ the new metric defect $g - u_{m+1}^\sharp e$ is.

The steps in Nash's iteration involve a very clever idea. The above assumptions allow a decomposition of the metric defect into rank-one tensors termed primitive metrics:

$$(g - u_m^\sharp e)(x) = \sum_{k=1}^{s_n} a_k^2(x)\, \nu_k \otimes \nu_k, \quad x \in Q. \tag{8.2}$$

Here the ν_k are fixed unit vectors independent of x and m and Q is a fixed domain in space (we are only solving the local problem; it is also necessary to assume that $g - u_m^\sharp e$ is small to ensure that the decomposition (8.2) is possible; these are minor assumptions).

The k-th step in the m-th stage approximately corrects the metric defect by the primitive metric $\theta a_k^2\, \nu_k \otimes \nu_k$, where $\theta \in (0,1)$ may be chosen later. Such a metric correction is obtained by choosing $v_0 = u_m$ and $v_k = v_{k-1} + w_k$, where w_k is a 'Nash twist' [11]

(8.3)

$$w_k(x) = \frac{\sqrt{\theta}\, a_k(x)}{\lambda_k} \left(\cos(\lambda_k x \cdot \nu_k)\, \mathbf{n}_{k-1}(x) + \sin(\lambda_k x \cdot \nu_k)\, \mathbf{b}_{k-1}(x)\right), \quad k = 1, 2, \ldots, s_n.$$

Here λ_k is a large parameter and $\mathbf{n}_{k-1}$ and $\mathbf{b}_{k-1}$ denote the normal and binormal vector fields to the short immersion v_{k-1} (we need two normal directions to introduce oscillations with sines and cosines, which is why we chose $q = n + 2$; Kuiper introduced a more delicate oscillation which works in codimension one [33]).

The derivative Dw_k may be written as the sum $A_k + B_k + C_k$ where

$$\begin{aligned} A_k &= \sqrt{\theta} a_k(x) \left(-\sin(\lambda_k x \cdot \nu_k)\, \mathbf{n}_{k-1}(x) + \cos(\lambda_k x \cdot \nu_k)\, \mathbf{b}_{k-1}(x)\right) \otimes \nu_k, \\ B_k, C_k &= O(\frac{1}{\lambda_k}). \end{aligned} \tag{8.4}$$

An essential aspect of the Nash twist is that $A_k^T D v_{k-1} = 0$. Therefore,

$$(u_m + w_1 + \ldots + w_{s_n})^\sharp\, e = u_m^\sharp e + \sum_{k=1}^{s_n} w_k^\sharp e + O(\frac{1}{\lambda}). \tag{8.5}$$

Let us now briefly explain the error estimates and the appearance of the exponent α. Each step introduces higher order derivatives, since the $O(1/\lambda_k)$ term includes the derivatives $D\mathbf{n}_{k-1}$ and $D\mathbf{b}_{k-1}$ which depend on $\|u_{k-1}\|_{C^2}$. Hölder continuity is obtained by combining Nash's iteration with a smoothing step to control the growth of the C^2 norm [17]. The Hölder exponent is determined by the balance between the convergence of u_m in C^1 and divergence of u_m in C^2. The exponent $\alpha = 1/(1 + 2s_n)$ arises from the fact that the steps are performed serially. If we could apply all the Nash twists simultaneously, we would obtain $\alpha = 1/3$.

We can now finally explain the role of randomization. We will try to eliminate the steps and consider a stage-by-stage iteration. There is a clear complication – the steps are performed serially to preserve orthogonality and ensure (8.5). However, by adding random coefficients we hope that the cross terms vanish *on average* and that this cancellation will still allow convergence.

---

[11] Of course, w_k, $\mathbf{n}_k$ and $\mathbf{b}_k$ depend on the stage m too, but we drop m from the notation of the steps to avoid clutter.

To this end, we modify the scheme as follows. Index the stages by m as earlier, assume u_m is short and C^∞ and let $\mathbf{n}_m$ and $\mathbf{b}_m$ be the normal and binormal fields to u_m. We choose a sequence of iid, standard normal random variables $X_{m,1}, \ldots, X_{m,s_n}$, a scaling parameter $\theta_m \in (0,1)$, and consider a random superposition of Nash twists

(8.6)
$$w_m = \frac{\sqrt{\theta_m}}{\lambda_m} \sum_{k=1}^{s_n} \sum_{k=1}^{s_n} X_{m,k} a_k(x) \left(\cos\left(\lambda_m x \cdot \nu_k\right) \mathbf{n}_m(x) + \sin\left(\lambda_m x \cdot \nu_k\right) \mathbf{b}_m(x) \right),$$

and set $u_{m+1} = u_m + w_m$ as before.

Note that equation (8.5) no longer holds. However, $Dw_m^T Du_m = O(1/\lambda_m)$ and $\mathbb{E}\left(w_m^\sharp e\right) = \theta_m(g - u_m^\sharp e) + O(1/\lambda_m)$ by construction, so that if we denote $F_m = \mathbb{E}\left(w_m^\sharp e\right) - w_m^\sharp e$ we obtain

$$g - u_{m+1}^\sharp e = (1-\theta_m)(g - u_m^\sharp e) + F_m + O(1/\lambda_m). \tag{8.7}$$

The fluctuation field F_m does not vanish, but it is a homogeneous Gaussian chaos of degree 2 with mean zero (so estimates exist to control it uniformly).

The catch is that it is no longer true that u_{m+1} is short with probability one, even if u_m is. This is a crucial aspect of Nash's argument (if not, one cannot continue the iteration!). Thus, in order to iterate the scheme, we need concentration estimates that state that u_{m+1} is short with high probability if λ_m is chosen sufficiently large. I spent a few months trying to make this work using two fundamental inequalities for Gaussian processes:

(1) Dudley's inequality for the supremum of Gaussian processes [23].
(2) The Borell-Ibragimov-Tsirelson-Sudakov concentration inequality [11].

These two estimates are standard tools in the theory of Gaussian processes (though they were new to me). The basic reference is the book by Ledoux and Talagrand [37], though there are also more sophisticated estimates (bounds on quadratic Gaussian chaos) that are directly relevant for this scheme, which are laid out in Ledoux's lecture notes [36, Sec.5].

In order to parallel Nash's approach we need to iterate these estimates scale-by-scale. For example, Dudley's inequality is applied as follows. The covariance kernel for each of the processes w_m, Dw_m, A_m, and B_m may be computed in terms of u_m and its derivatives. What one then has to do is to choose parameters λ_m and θ_m and optimize the bound in a way that is suited to the Nash iteration. For instance, a calculation shows that (see Appendix **??**)

$$\mathbb{E}\left(\|F_m\|_\infty\right) \leq C\theta_m \, \|g - u_m^\sharp e\|_\infty \, \log\left(\frac{\lambda_m \|D^2 u_m\|_\infty}{\|Du_m\|_\infty} \right). \tag{8.8}$$

This bound is essentially sharp. However, when we iterate these estimates and include the error terms, the calculations get very messy. After trying for a few months to make this work, I ran out of steam and put the problem aside.

However, I did learn something important in this approach: Dudley's inequality is an *entropic bound*. This comes about as follows. A mean zero Gaussian process on an arbitrary set induces a metric on the set. More precisely, if X is a set and $\varphi : X \to \mathbb{R}$ is a Gaussian process such that $\mathbb{E}(\varphi(x)) = 0$, $x \in X$ and $K(x,y) =$

$\mathbb{E}(\varphi(x)\varphi(y))$, then

$$d_K^2(x,y) := \det \begin{vmatrix} K(x,x) & K(x,y) \\ K(y,x) & K(y,y) \end{vmatrix} = \mathbb{E}\left(\varphi(x)-\varphi(y)\right)^2, \tag{8.9}$$

defines a metric on X. Gaussian bounds, beginning with Dudley, control the supremum of φ in terms of the *metric entropy* of the metric d_K. Roughly, we cover the metric space (X, d_K) with increasingly fine collections of balls and we control the supremum of φ by comparison with its average on these balls.

I found Dudley's argument, as well as the BITS inequality, fascinating. Physicists who study soft matter think about crumpling in entropic terms, but it was never clear to me how to formulate their ideas mathematically. Thus, despite being overwhelmed by calculations, the fact that the idea of entropy made its appearance in this problem – in a precise mathematical sense – was very encouraging. Nash's proof didn't seem quite as invulnerable any more. The fluctuations of Gaussian processes are completely prescribed by a covariance kernel. A good kernel – such as that constructed from the Nash twists – makes for a tight proof. But, clever as the Nash twist may be, it is *ad hoc*. Entropy, on the other hand, is as fundamental as it gets.

8.3. **The BBG embedding and heat kernel asymptotics.** An important result of Bérard, Besson and Gallot [8] provides a precise understanding of the embedding problem in infinite dimensions that sheds light on the embedding problem in finite dimensions. This work is also of importance in machine learning.

The BBG embedding uses the heat kernel associated to the metric g. Let $\triangle_g$ be the Laplacian associated to g, whose action on a test function ψ is given in local coordinates by

$$\triangle_g \psi(x) = \frac{1}{\sqrt{G}} \frac{\partial}{\partial x^i}\left(\sqrt{G} g^{ij} \frac{\partial \psi}{\partial x^j}\right), \quad G = \det g. \tag{8.10}$$

Let $k_g(t;x,y)$ denote the resolvent of the heat kernel $e^{-t\triangle_g}$. [12] Since $\mathcal{M}$ is a closed manifold, it has an eigenfunction expansion

$$k_g(t;x,y) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \varphi_i(x)\varphi_i(y), \tag{8.11}$$

where $\{\varphi_i\}_{i=0}^{\infty}$ is a complete basis for $L^2(\mathcal{M}, g)$. [13] The operator k_g is trace-class and

$$Z_g(t) := \mathrm{Tr}(e^{-t\triangle_g}) = \int_{\mathcal{M}} k_g(t;x,x)\, d\mathrm{vol}_g(x) = \sum_{i=0}^{\infty} e^{-\lambda_i t}. \tag{8.12}$$

The BBG embedding is the map $u_t : \mathcal{M} \to l^2$ given by

$$u_t(x) = \sqrt{2}(4\pi)^{n/4} t^{(n+2)/4} \{e^{-\lambda_j t/2} \varphi_j(x)\}_{j=1}^{\infty}. \tag{8.13}$$

The zero-th eigenfunction is constant, which is why we only consider $j \geq 1$.

Let e_∞ denote the canonical inner product on l^2: for $\psi, \psi' \in l^2$, $e_\infty(\psi,\psi') = \sum_{j=1}^{\infty} \psi_j \psi'_j$. The basic result of Bérard, Besson and Gallot [8, Thm.5] is the following description of the pullback metric $u_t^\sharp e_\infty$.

---

[12] Here we use the convention of [8] so that the formulas are the same as in [8]. In most other places, we use the probabilists convention $e^{-t\triangle_g/2}$ which differs by a factor of 2.

[13] This notation means we consider the volume form dx induced by the metric g.

(1) For each $t > 0$ the map $u_t : \mathcal{M} \to l^2$ is an embedding.
(2) The mapping u_t is *asymptotically isometric* in the sense that

$$u_t^\sharp e_\infty = g + \frac{t}{3}\left(\frac{1}{2}\mathrm{Scal}(g)g - \mathrm{Ric}(g)\right) + O(t^2), \quad t \to 0. \tag{8.14}$$

Here $\mathrm{Scal}(g)$ and $\mathrm{Ric}(g)$ denote the scalar and Ricci curvature of g respectively.

In order to facilitate direct comparison with [8], let us write the formulas in terms of the map $\Phi_t : \mathcal{M} \to l^2$ defined by

$$\Phi_t(x) = \{e^{-\lambda_j t/2}\varphi_j(x)\}_{j=1}^\infty. \tag{8.15}$$

The proof of part (1) is soft: it uses only the fact that $\{\varphi_j(x)\}_{j=1}^\infty$ is a basis for l^2 (modulo constants). Thus, it separates points and $\Phi_t(x) = \Phi_t(y)$ if and only if $x = y$. The proof of part (2) relies on classical heat kernel asymptotics. The key calculation is the following: if $v \in T_xM$, then by (8.15)

$$\Phi_t^\sharp e_\infty(x)(v,v) = \|d\Phi_t(x)v\|_{l^2}^2 = \sum_{j\geq 1} e^{-\lambda_j t}\left|d\varphi_j(x)v\right|^2. \tag{8.16}$$

We use (8.11) to rewrite the last term above in the form

$$\Phi_t^\sharp e_\infty(x)(v,v) = (\Diamond k_g(t))(x)v^iv^j. \tag{8.17}$$

Here is where I saw the linear operator $\Diamond$ for the first time (compare equations (3.10) and (8.17)). Bérard, Besson and Gallot use different notation; the use of $\Diamond$ prevents confusion with the use of d for stochastic differentials, while sticking with standard LaTeX symbols.

The formula (8.14) is obtained by combining the classical Minakshisundaram-Pleijel asymptotics for the heat kernel k_g with (8.17). These calculations are straightforward, but they are quite interesting since we begin to see the interplay between geodesics, Ricci curvature, and (implicitly) Brownian motion.

While the BBG embedding is infinite-dimensional and it is only asymptotically isometric, it has the appealing feature of being more directly connected to the geometry of the manifold than Nash's scheme. In particular, the appearance of the scalar and Ricci curvatures in (8.17) gives us hope that there may be 'typical' embeddings of manifolds even in finite dimension. I did not find anything truly probabilistic in the literature, but there are many papers – pure and applied – that show quite clearly that the BBG embedding is of fundamental importance. For example, an attempt to construct canonical C^∞ embeddings beginning with the BBG embedding is considered in [48]. There is also a large literature on the use of heat kernel embeddings in machine learning. For example, it is the BBG embedding that underlies so-called diffusion geometry for various pattern recognition appliations. A mathematical view of this world may be found in [32]; there are many similar papers in applied and computational harmonic analysis. While none of the above literature is directly relevant to our goals, it does suggest again the importance of diffusion and entropy in the embedding problem.

Let us now return to embeddings in finite dimensions and Nash 54. For someone with my tastes, it is more natural to think of Brownian motion rather than heat kernels; that is, I'd like to switch from an implicitly probabilistic viewpoint to an explicitly probabilistic viewpoint. So what I attempted to do next is to stick with the general framework of Nash 54, but to use the BBG embedding instead of the Nash twists and simplicial decomposition to build fluctuation fields with more transparent estimates.

Thus, let us put ourselves back in the Nash-Kuiper setup, further reducing to codimension one. We assume that we have a short immersion $u_m : \mathcal{M} \to \mathbb{R}^{n+1}$, with $u_m^\sharp e < g$. Let $g_m = g - u_m^\sharp e$ denote the metric defect and let $\triangle_{g_m}$ and k_{g_m} denote the Laplacian and heat-kernel associated to the metric defect g_m. Let us also use the somewhat cumbersome notation $\{\lambda_j^m\}_{j\geq 0}$ and $\{\varphi_j^m\}_{j\geq 0}$ to denote the eigenvalues and eigenfunctions of $\triangle_{g_m}$.

We use the BBG embedding to improve our earlier guess for the fluctuation field (8.6) as follows. Fix a smoothing parameter t_m and define the scalar random field $\psi_m : \mathcal{M} \to \mathbb{R}$

$$\psi_m(x) = \sqrt{2}(4\pi)^{n/4} t_m^{(n+2)/4} \sum_{j=1}^{\infty} X_{j,m} e^{-\lambda_j^m t_m/2} \varphi_j^m(x), \tag{8.18}$$

where $X_{j,m}$ are iid standard Gaussian random variables. The scaling factors in (8.18) are precisely those of the BBG embedding (see (8.13)).

We use the scalar random field ψ_m and the normal vector field $\mathbf{n}_m$ to u_m to define a vector-valued random field $w_m : \mathcal{M} \to \mathbb{R}^{n+1}$ in the natural way:

$$w_m(x) = \psi_m(x)\mathbf{n}_m(x), \quad x \in \mathcal{M}. \tag{8.19}$$

This fluctuation field should be compared with (8.6). What's going on here is that by using the iid random variables, we have found an 'auxiliary' l^2 Gaussian space of fluctuations (i.e. an approximate infinite dimensional embedding) and we have combined it with the normal to construct a 'physical' space correction akin to the Nash-Kuiper construction.

Let us check this. The spatial derivatives of w_m are given by

$$Dw_m = \mathbf{n}_m \otimes D\psi_m + \psi_m D\mathbf{n}_m. \tag{8.20}$$

By construction, the first term on the right hand side of (8.20) is orthogonal to Du_m. The second term should be thought of as a small scalar multiplied by an extrinsic curvature (ψ_m is small if t_m is small, because of the prefactor $t_m^{(n+1)/4}$ in (8.18) and Dudley's inequality). Therefore, with e denoting the standard inner product on $\mathbb{R}^{n+1}$, and using the fact that $\mathbf{n}_m$ is a unit vector, we find that $w_m^\sharp e$ is given in coordinates by

$$\left(w_m^\sharp e\right)_{ij} := \frac{\partial w_m^\alpha}{\partial x^i}\frac{\partial w_m^\alpha}{\partial x^j} = \frac{\partial \psi_m}{\partial x^i}\frac{\partial \psi_m}{\partial x^j} + \psi_m^2 \frac{\partial n_m^\alpha}{\partial x^i}\frac{\partial n_m^\alpha}{\partial x^j}. \tag{8.21}$$

Similarly, computing $u_{m+1}^\sharp e$ in coordinates we find

$$\left(u_{m+1}^\sharp e\right)_{ij} = \left(u_m^\sharp e\right)_{ij} + \left(w_m^\sharp e\right)_{ij} + \psi_m(x)\left(\frac{\partial n_m^\alpha}{\partial x^i}\frac{\partial u_m^\alpha}{\partial x^j} + \frac{\partial n_m^\alpha}{\partial x^j}\frac{\partial u_m^\alpha}{\partial x^i}\right). \tag{8.22}$$

These computations should be compared with (8.5) and (8.6). The (small) smoothing parameter t_m plays roughly the same role as the (large) oscillation parameter λ_m. In particular, the definition (8.18) shows that ψ_m has mean zero, so that conditional on u_m

$$\mathbb{E}\left(u_{m+1}^\sharp e\right)_{ij} = \left(u_m^\sharp e\right)_{ij} + \mathbb{E}\left(w_m^\sharp e\right)_{ij}. \tag{8.23}$$

Now using (8.18), (8.19) and (8.22) and choosing t_m small we find that

$$\mathbb{E}\left(w_m^\sharp e\right)_{ij} = (g_m)_{ij} + \frac{t_m}{3}\left(\frac{1}{2}\mathrm{Scal}(g_m)g_m - \mathrm{Ric}(g_m)\right) + ct_m^{(n+2)/2}k_{g_m}(t_m;x,x)\frac{\partial n_m^\alpha}{\partial x^i}\frac{\partial n_m^\alpha}{\partial x^j} + O(t_m^2), \tag{8.24}$$

where c is a universal constant.

In summary, we see that while the BBG embedding is infinite-dimensional, viewing it as an embedding into Gaussian space allows us to develop a random Nash-Kuiper scheme in codimension one. This scheme removes the *ad hoc* nature of the Nash twist or Kuiper's correction and provides a clearer geometric understanding of the error terms in terms of intrinsic and extrinsic curvature.

There is still a price to pay: we need concentration estimates to show that u_{m+1} is short. Nevertheless, at this point it is quite clear that it is possible with some technical effort to improve the exponents in [17].

What I found more tantalizing is the possibility that there is a deeper structure in the problem that was intrinsically probabilistic. Since its much simpler to deal with Brownian motion rather than random walks, I decided to switch to stochastic flows. As soon as one looks at [42] through this lens, the kinematic advantages of stochastic flows becomes immediately apparent. However, it is still subtle to understand how to evolve the smoothing operator. Nash includes an evolution equation, but it struck me as very cumbersome, and certainly not as natural as a gradient flow. Given Otto's work, it now becomes a matter of finding an infinite-dimensional Riemannian geometry that is the analog of the space of probability measures in [45]. A natural analog is the space of metrics $\mathsf{Met}(\mathcal{M})$ used by the physicists. However, this is not quite right in an extrinsic setting, because of the codimension. After some exploration, the geometry of the space of Gaussian kernels began to feel like the natural candidate.

References

[1] V. Arnoldʹ and B. Khesin, *Topological methods in hydrodynamics*, Springer Verlag, 1998.

[2] E. Bartzos, V. Borrelli, R. Denis, F. Lazarus, D. Rohmer, and B. Thibert, *An explicit isometric reduction of the unit sphere into an arbitrarily small ball*, Found. Comput. Math., 18 (2018), pp. 1015–1042.

[3] P. Baxendale, *Gaussian measures on function spaces*, Amer. J. Math., 98 (1976), pp. 891–952.

[4] ———, *Stochastic flows and Malliavin calculus*, in 1981 20th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes, IEEE, 1981, pp. 595–599.

[5] ———, *Brownian motions in the diffeomorphism group. I*, Compositio Math., 53 (1984), pp. 19–50.

[6] 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.

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

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

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

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

[11] C. Borell, *On polynomial chaos and integrability*, Probab. Math. Statist., 3 (1984), pp. 191–203.

[12] V. Borrelli, S. Jabrane, F. Lazarus, and B. Thibert, *Flat tori in three-dimensional space and convex integration*, Proc. Natl. Acad. Sci. USA, 109 (2012), pp. 7218–7223.

[13] L. Bottou, *Large-scale machine learning with stochastic gradient descent*, in Proceedings of the 19th International Conference on Computational Statistics (COMPSTAT'2010), Y. Lechevallier and G. Saporta, eds., Paris, France, August 2010, Springer, pp. 177–187.

[14] R. W. Brockett, *Lie algebras and lie groups in control theory*, in Geometric methods in system theory, Springer, 1973, pp. 43–82.

[15] R. W. Brockett, *Differential geometry and the design of gradient algorithms*, in Differential geometry: partial differential equations on manifolds (Los Angeles, CA, 1990), vol. 54 of Proc. Sympos. Pure Math., Amer. Math. Soc., Providence, RI, 1993, pp. 69–92.

[16] P. Constantin, W. E, and E. S. Titi, *Onsager's conjecture on the energy conservation for solutions of Euler's equation*, Communications in Mathematical Physics, 165 (1994), pp. 207–209.

[17] S. Conti, C. De Lellis, and L. Székelyhidi, Jr., *h-principle and rigidity for $C^{1,\alpha}$ isometric embeddings*, in Nonlinear partial differential equations, vol. 7 of Abel Symp., Springer, Heidelberg, 2012, pp. 83–116.

[18] C. De Lellis and L. Székelyhidi, Jr., *The Euler equations as a differential inclusion*, Ann. of Math. (2), 170 (2009), pp. 1417–1436.

[19] C. De Lellis and L. Székelyhidi, Jr., *High dimensionality and h-principle in PDE*, Bull. Amer. Math. Soc. (N.S.), 54 (2017), pp. 247–282.

[20] C. De Lellis and L. Székelyhidi Jr., *The h-principle and the equations of fluid dynamics*, Bulletin of the American Mathematical Society, 49 (2012), pp. 347–375.

[21] ———, *Dissipative continuous Euler flows*, Invent. Math., 193 (2013), pp. 377–407.

[22] P. Deift, T. Nanda, and C. Tomei, *Ordinary Differential Equations and the Symmetric Eigenvalue Problem*, SIAM Journal on Numerical Analysis, 20 (1983), pp. 1–22.

[23] R. M. Dudley, *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes*, J. Functional Analysis, 1 (1967), pp. 290–330.

[24] J. Eells and K. Elworthy, *Wiener integration on certain manifolds*, in Problems in nonlinear analysis, Springer, 2010, pp. 69–94.

[25] G. L. Eyink, *Energy dissipation without viscosity in ideal hydrodynamics I. Fourier analysis and local energy transfer*, Physica D: Nonlinear Phenomena, 78 (1994), pp. 222–240.

[26] D. Friedan, *Introduction to Polyakov's string theory*, in Recent advances in field theory and statistical mechanics (Les Houches, 1982), North-Holland, Amsterdam, 1984, pp. 839–867.

[27] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian geometry*, vol. 3, Springer, 1990.

[28] 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.

[29] M. Günther, *Isometric embeddings of Riemannian manifolds*, in Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990), Math. Soc. Japan, Tokyo, 1991, pp. 1137–1143.

[30] M. Hairer and J. C. Mattingly, *Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing*, Ann. of Math. (2), 164 (2006), pp. 993–1032.

[31] E. P. Hsu, *Stochastic analysis on manifolds*, vol. 38, American Mathematical Soc., 2002.

[32] 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.

[33] N. Kuiper, *On C^1-isometric imbeddings. I, II*, in Proc. Kon. Acad. Wet. Amsterdam A, vol. 58, 1955, p. 683689.

[34] H. Kunita, *Stochastic flows and stochastic differential equations*, vol. 24 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1997. Reprint of the 1990 original.

[35] S. Lang, *Fundamentals of differential geometry*, vol. 191, Springer Science & Business Media, 2012.

[36] M. Ledoux, *Isoperimetry and Gaussian analysis*, in Lectures on probability theory and statistics (Saint-Flour, 1994), vol. 1648 of Lecture Notes in Math., Springer, Berlin, 1996, pp. 165–294.

[37] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Classics in Mathematics, Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.

[38] P. Lévy, *Processus stochastiques et mouvement brownien*, Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics], Éditions Jacques Gabay, Sceaux, 1992. Followed by a note by M. Loève, Reprint of the second (1965) edition.

[39] P. Malliavin, *Stochastic analysis*, vol. 313, Springer, 2015.

[40] D. Mumford and A. Desolneux, *Pattern theory: the stochastic analysis of real-world signals*, AK Peters/CRC Press, 2010.

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

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

[43] E. Nelson, *Quantum fluctuations*, Princeton Series in Physics, Princeton University Press, Princeton, NJ, 1985.

[44] T. Nishioka, *Entanglement entropy: holography and renormalization group*, Rev. Modern Phys., 90 (2018), pp. 035007, 56.

[45] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.

[46] T. Shiga and S. Watanabe, *Bessel diffusions as a one-parameter family of diffusion processes*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 27 (1973), pp. 37–46.

[47] A. Takatsu, *Wasserstein geometry of Gaussian measures*, Osaka J. Math., 48 (2011), pp. 1005–1026.

[48] X. Wang, K. Zhu, et al., *Isometric embeddings via heat kernel*, Journal of Differential Geometry, 99 (2015), pp. 497–538.

[49] E. Witten, *Notes on some entanglement properties of quantum field theory*, arXiv preprint arXiv:1803.04993, (2018).

Current: School of Mathematics, Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540.

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

Email address: govind_menon@brown.edu