Information theory and the embedding problem for Riemannian manifolds*

Govind Menon[1]

Division of Applied Mathematics, Brown University, Providence, RI 02912, USA
govind_menon@brown.edu

Abstract. This paper provides an introduction to an information theoretic formulation of the embedding problem for Riemannian manifolds developed by the author. The main new construct is a stochastic relaxation scheme for embedding problems and hard constraint systems. This scheme is introduced with examples and context.

Keywords: Geometric information theory · Embedding theorems

1 The embedding problem for Riemannian manifolds

The purpose of this paper is to outline an information theoretic formulation of the embedding problem for Riemannian manifolds developed by the author. The main new contribution is a stochastic relaxation scheme that unifies many hard constraint problems. This paper is an informal introduction to this method.

The modern definition of a manifold was formalized by Whitney in 1936 [11]. He defined manifolds as abstract spaces covered by locally compatible charts, thus providing a rigorous description of the intuitive idea that a manifold is a topological space that 'locally looks like Euclidean space'. Whitney's definition should be contrasted with the 19th century idea of manifolds as hypersurfaces in Euclidean space. The embedding problem for a differentiable manifold checks the compatibility of these notions. Given an (abstractly defined) n-dimensional differentiable manifold $\mathcal{M}^n$, an embedding of $\mathcal{M}^n$ is a smooth map $u : \mathcal{M}^n \to \mathbb{R}^q$ that is one-to-one and whose derivative $Du(x)$ has full rank at each $x \in \mathcal{M}^n$.

An embedded manifold carries a pullback metric, denoted $u^{\#}e$, where e denotes the identity metric on $\mathbb{R}^q$. Assume that $\mathcal{M}^n$ is a Riemannian manifold equipped with a metric g. We say that an embedding $u : \mathcal{M}^n \to \mathbb{R}^q$ is *isometric* if $u^\sharp e = g$. In any local chart U, this is the nonlinear PDE

$$\sum_{\alpha=1}^{q} \frac{\partial u^\alpha}{\partial x^i}\frac{\partial u^\alpha}{\partial x^j}(x) = g_{ij}(x), \quad x \in U, \quad 1 \leq i, j \leq n. \tag{1}$$

Let us contrast equation (1) with isometric embedding of finite spaces.

* Partial support for this work was provided by the National Science Foundation (DMS 1714187), the Simons Foundation (Award 561041), the Charles Simonyi Foundation and the School of Mathematics at the Institute for Advanced Study. The author is grateful to the anonymous referees of this paper for many valuable comments.

(a) Embedding metric spaces: We are given a finite set K with a distance function $\rho(x,y)$ and we seek a map $u : K \to \mathbb{R}^q$ such that

$$|u(x) - u(y)| = \rho(x,y), \quad x, y \in K. \tag{2}$$

(b) Graph embedding: We are given a graph $G = (V, E)$ with a distance function $\rho : E \to \mathbb{R}_+$ that associates a length to each edge. The graph embedding problem is to find $u : V \to \mathbb{R}^q$ such that

$$|u(e_+) - u(e_-)| = \rho(e), \quad e \in E, \tag{3}$$

where $e_\pm$ denote the vertices at the two ends of an edge $e \in E$.

In each of these problems an abstractly defined metric space is being mapped into the reference space $\mathbb{R}^q$. The LHS is the length measured in $\mathbb{R}^q$. The RHS is the intrinsic distance on the given space. In equation (1), the equality of length is expressed infinitesimally, which is why we obtain a PDE.

Modern understanding of (1) begins with the pioneering work of Nash in the 1950s [8, 9]. His work led to the following results: for $q = n + 1$ and $g \in C^0$ there are infinitely many C^1 isometric embeddings (assuming no topological obstructions); when $q \geq n + n(n + 1)/2 + 5$ and $g \in C^\infty$ there are infinitely many C^∞ isometric embeddings. These results are improvements of Nash's original work, but follow his ideas closely. Nash's work has been systematized in two distinct ways: as Gromov's h-principle in geometry and as hard implicit function theorems in analysis. However, several fundamental questions remain unresolved [5].

Our interest in the area was stimulated by an unexpected link with turbulence [2]. A long-standing goal in turbulence is to construct Gibbs measures for the Euler equations of ideal incompressible fluids whose statistical behavior is in accordance with experiments. This connection suggests the application of statistical mechanics to embeddings. The construction of Gibbs measures for embeddings allows us to formalize the question 'What does a typical isometric embedding look like?'. This is in contrast with the questions 'Does an isometric embedding exist? If so, how smooth is it?' resolved by Nash and Gromov.

The Gibbs measures have a natural information theoretic construction. We model embedding as a stochastic process in which an observer in $\mathbb{R}^q$ makes a copy of a given Riemannian geometry by measurement of distances at finer and finer scales. This is a Bayesian interpretation suited to the interplay between geometry and information theory at this conference. However, prior to the author's work there was no attempt to study (1) with probabilistic methods or to treat equations (1)–(3) through a common framework. Further, the devil lies in the details, since any new attempt must be consistent with past work and must be nailed down with complete rigor. This paper discusses only the evolution equations. Analysis of these equations will be reported in forthcoming work.

2 The role of information theory

The embedding theorems are interesting both for their conceptual and technical depth. They arise in apparently unrelated fields and they have been studied by disparate techniques. Within mathematics, Nash's three papers on manifolds appear unrelated on first sight. Mathematical techniques for graph embedding, mainly stimulated by computer science, appear to have little relation to the embedding problem for manifolds [7]. The embedding problem also appears under the guise of the nonlinear sigma models in quantum field theory. In his breakthrough work, Friedan showed that the renormalization of the nonlinear sigma model is the Ricci flow [3, 4]. However, Friedan's technique, renormalization by expansion in dimension, is notoriously hard to pin down mathematically.

Such a diversity of methods and applications is bewildering until one recognizes that it offers a route to a radical conceptual simplification. In order to obtain a unified treatment, it is necessary to insist on a minimalistic formulation of embedding that does not rely in a fundamental manner on the structure of the space being embedded (e.g. whether it is a graph, manifold, or metric space). Such a formulation must be consistent with both Nash and Friedan's approach to the problem, as well as applications in computer science. This line of reasoning suggests that the appropriate foundation must be information theory – it is the only common thread in the above applications.

The underlying perspective is as follows. We view embedding as a form of information transfer between a source and an observer. The process of information transfer is complete when all measurements of distances by the observer agree with those at the source. This viewpoint shifts the emphasis from the structure of the space to an investigation of the process by which length is measured. In the Bayesian interpretation, the world is random and both the source and the observer are stochastic processes with well-defined parameters (we construct these processes on a Gaussian space to be concrete). Thus, embedding is simply 'replication' and the process of replication is complete when all measurements by the observer and the source agree on a common set of questions (here it is the question: 'what is the distance between points x and y?'). From this standpoint, there is no fundamental obstruction to embedding, except that implied by Shannon's channel coding theorem.

The challenge then is to implement this viewpoint with mathematical rigor. In order to explain our method, we must briefly review Nash's techniques. The main idea in [8] is that when $q = n + 2$ one can relax the PDE $u^\sharp g = e$ to a space of subsolutions and then introduce highly structured corrugations in the normal directions at increasingly fine scales. This iteration 'bumps up' smooth subsolutions towards a solution. In [9], Nash introduces a geometric flow, which evolves an immersion and a smoothing operator simultaneously. Unlike [8], which is brief and intuitive, the paper [9] is lengthy and technical, introducing what is now known as the Nash-Moser technique. A central insight in our approach is that one can unify these methods by introducing a reproducing kernel that evolves stochastically with the subsolution.

A rigid adherence to an information theoretic approach to embedding contradicts Nash's results in the following sense. If all that matters is agreement in the measurement of distances between different copies of a manifold, the historical emphasis in mathematics on the role of codimension and regularity cannot be a fundamental feature of the problem. All embeddings are just copies of the same object, an abstractly defined manifold. Typical embeddings of the manifold $(\mathcal{M}^n, g)$ into $\mathbb{R}^p$ and $\mathbb{R}^q$ for $p, q \geq n$ should have different regularity: a C^∞ metric g may yield a crumpled embedding in $\mathbb{R}^p$ and a smooth embedding in $\mathbb{R}^q$ for $p < q$, but since it is only the measurement of length that matters, there is no preferred embedding. Thus, existence and regularity of solutions to (1) must be treated separately. This mathematical distinction acquires salience from its physical meaning. While Wilson renormalization is often seen as a technique for integrating out frequencies scale-by-scale, it reflects the role of gauge invariance in the construction of a physical theory. The only true measurements are those that are independent of the observer. Therefore, in order to ensure consistency between mathematical and physical approaches to the isometric embedding problem it is necessary to develop a unified theory of embeddings that does not rely substantially on codimension. Nash's methods do not meet this criterion.

Finally, embedding theorems may be used effectively in engineering only if the rigorous formulation can be supported by fast numerical methods. Here too Nash's techniques fail the test. The first numerical computations of isometric embeddings are relatively recent [1]. Despite the inspiring beauty of these images, they require a more sophisticated computational effort than is appropriate for a problem of such a fundamental nature. The numerical scheme in [1] is ultimately based on [8]. Thus, it requires the composition of functions, which is delicate to implement accurately. The models proposed below require only semidefinite programming and the use of Markov Chain Monte Carlo, both of which are standard techniques, supported by excellent software.

These remarks appear to be deeply critical of Nash's work, but the truth is more mysterious. The imposition of stringent constraints – consistency with applications, physics and numerical methods – has the opposite effect. It allows us to strip Nash's techniques down to their essence, revealing the robustness of his fundamental insights. By using information theory to mediate between these perspectives, we obtain a new method in the statistical theory of fields.

3 Renormalization group (RG) flows

3.1 General principles

The structure of our method is as follows. Many hard constraint systems and nonlinear PDE such as $u^\sharp e = g$ admit relaxations to subsolutions. We will begin with a subsolution and improve it to a solution by adding fluctuations in a bandlimited manner. These ideas originate in Nash's work [8, 9]. We sharpen his procedure as follows:

1. The space of subsolutions is augmented with a Gaussian filter. More precisely, our unknown is a subsolution u_t and a reproducing kernel L_t, $t \in [0, \infty)$. In physical terms, the unknown is a thermal system.
2. We introduce a stochastic flow for (u_t, L_t). This allows us to interpolate between the discrete time iteration in [8] and C^1 time evolution in [9], replacing Nash's feedback control method with stochastic control theory.

3. We apply the modeling principles of continuum mechanics – a separation between kinematics and energetics – to obtain a semidefinite program (SDP) for the covariance of the Itô SDE.
4. A principled resolution of the SDP is the most subtle part of the problem. We illustrate two approaches: low-rank matrix completion and Gibbs measures for the SDP. In physical terms, this is a choice of an equation of state.

The use of reproducing kernels provides other insights too. The Aronszajn-Moore theorem asserts that the reproducing kernel L_t is in one-to-one correspondence with a Hilbert space $\mathcal{H}_{L_t}$. Thus, (u_t, L_t) describes a stochastically evolving affine Hilbert space $(u_t, \mathcal{H}_{L_t})$ much like subspace tracking in machine learning.

3.2 An example: random Lipschitz functions

Let us illustrate the structure of the RG flows on a model problem. We construct random Lipschitz functions as solutions to the Hamilton-Jacobi equation

$$|\nabla u(x)|^2 = 1, \quad x \in \mathbb{T}^n, \quad u : \mathbb{T}^n \to \mathbb{R}. \tag{4}$$

Equation (4) is simpler than equation (1) because the unknown u is a scalar. We say that $v : \mathbb{T}^n \to \mathbb{R}$ is a smooth subsolution if $v \in C^\infty$ and $|\nabla v(x)| < 1$ for $x \in \mathbb{T}^n$. Define the residual $r(x; v)$, trace l, and the density matrix P by

$$r(x;v) = (1 - |\nabla v(x)|^2)_+^{1/2}, \quad l = \int_{\mathbb{T}^n} L(x,x)\,dx, \quad P(x,y) = \frac{1}{l} L(x,y), \; x, y \in \mathbb{T}^n. \tag{5}$$

The simplest RG flow associated to equation (4) is the stochastic evolution

$$du_t(x) du_t(y) = (PSP)(x,y)\,dt, \quad \dot{P} = PSP - \mathrm{Tr}(PSP)P, \quad \frac{\dot{l}}{l} = \mathrm{Tr}(PSP), \tag{6}$$

where S is a covariance kernel constructed from $r(x; u_t)$ as follows

$$S(x,y) = \nabla r(x) \cdot \nabla r(y) = \sum_{i=1}^n \partial_{x_i} r(x;u_t) \partial_{y_i} r(y;u_t), \quad x, y \in \mathbb{T}^n. \tag{7}$$

Both S and P are integral operators on $L^2(\mathbb{T}^n)$ and PSP denotes the natural composition of such operators.

The first equation in (6) reflects *stochastic kinematics*. As in Nash's work, we are bumping up a subsolution, but now by stochastic fluctuations with covariance tensor PSP. The density matrix P smoothes the correction S, so that fluctuations are band-limited. The last equation shows that l_t is slaved to (u_t, P_t). The study of (u_t, P_t) and (u_t, L_t) is equivalent for this reason and both choices offer different insights. The equation for (u_t, L_t) (after a change of time-scale) is

$$du_t(x) du_t(y) = \dot{L}(x,y)\,dt, \quad \dot{L} = LSL. \tag{8}$$

Observe that this equation is invariant under reparametrization of time.

The specific relation between S and r in equation (7) emerges from an explicit rank-one solution to a matrix completion problem. More generally, all RG flows for equation (4) require the resolution of an SDP that provides a covariance kernel S, given a residual r. An alternate resolution of this question, involving Gibbs measures for the SDP, is described in the next section.

3.3 Interpretation

The intuition here is as follows. Assume given initial conditions (u_0, L_0) where u_0 is a smooth subsolution and L_0 is a band-limited reproducing kernel. Standard SDE theory implies the existence of strong solutions to equations (6) and (8). Equation (8) tells us that L_t is increasing in the Loewner order, so that the Hilbert spaces $\mathcal{H}_{L_t}$ are ordered by inclusion. This corresponds to the subsolutions getting rougher and rougher, while staying band-limited. On the other hand, equation (6) tells us that u_t and P_t are bounded martingales, so that $(u_\infty, P_\infty) := \lim_{t\to\infty}(u_t, P_t)$ exists by the martingale convergence theorem. The limit u_∞ is always a random subsolution to (4). Our task is to find the smallest space L_0 so that u_∞ is a solution to (4), thus providing random Lipschitz functions.

The analogous evolution for the isometric embedding problem (1) is obtained from similar reasoning. A subsolution is a map $v : \mathcal{M}^n \to \mathbb{R}^q$ such that $v^\sharp e$ satisfies the matrix inequality $v^\sharp e(x) < g(x)$ at each $x \in \mathcal{M}^n$. The residual $r(x; v)$ is the matrix square-root of the metric defect $g - v^\sharp e(x)$. The covariance kernel L_t is also now a matrix valued kernel. Thus, the generalization reflects the tensorial nature of (1) and does not change the essence of equation (6). The associated flow makes precise the idea that embedding is a process of estimation of the metric g by estimators $u_t^\sharp e$. At each scale t, we choose the best correction to u_t given the Gaussian prior P_t and a principled resolution of an SDP.

Finally, equation (8) has a simple physical interpretation. The RG flows model quasistatic equilibration of the thermal system (u_t, L_t). This is perhaps the most traditional thermodynamic picture of the flow of heat, dating back to Clausius, Gibbs and Maxwell. What is new is the mathematical structure. The mean and covariance evolve on different time-scales, so that the system is always in local equilibrium. This insight originates in Nelson's derivation of the heat equation [10]. Like Nelson, we stress the foundational role of stochastic kinematics and time-reversibility. However, unlike Nelson, we rely on information theory as the foundation for heat flow, not *a priori* assumptions about a background field. The flows are designed so that $\dot{L}_t$ is always the 'most symmetric' fluctuation field with respect to the prior. This offers a rigorous route to the construction of Gibbs measures by renormalization, using different techniques from Friedan's work. This is why we term our model an RG flow.

4 Isometric embedding of finite metric spaces into $\mathbb{R}^q$

In this section we show that RG flows for equations (1)–(3) may be derived from common principles. This goes roughly as follows: the discrete embeddings (2)–(3)

have subsolutions and we use a stochastic flow analogous to (8) to push these up to solutions. The main insights in this section are the role of an underlying SDP and the use of low-rank kernels L_t as finite-dimensional analogs of smoothing operators. Formally, we expect embeddings of the manifold to be the continuum limit of discrete embeddings of geodesic triangulations of the manifold. However, this has not yet been established rigorously.

Assume given a finite metric space (K,ρ). Equation (2) is a hard constraint system that may not have a solution. For example, an equilateral triangle cannot be isometrically embedded into $\mathbb{R}$. It is necessary to relax the problem. Following Nash [8], let us say that a map $v : K \to \mathbb{R}^q$ is *short* if $|v(x)-v(y)| < \rho(x,y)$ for each pair of distinct points $x, y \in K$. These are our subsolutions.

Let $\mathbb{P}(n,q)$ denote the space of covariance tensors for $\mathbb{R}^q$-valued centered Gaussian processes on K. Our state space is $\mathcal{S}_q = \{(u,L) \in \mathbb{R}^{nq} \times \mathbb{P}(n,q)\}$. The RG flow analogous to equation (8) is the Itô SDE

$$du_t^i(x)du_t^j(y) = \dot{L}(x,y)^{ij}\,dt, \quad \dot{L}_t = C(u_t,L_t), \quad x,y \in K, \quad 1 \le i,j \le q. \tag{9}$$

The rest of this section describes the use of SDP to determine C. First, we set $C(u,L) \equiv 0$ when u is not short. When u is short its *metric defect* is

$$r^2(x,y;u) = \left(\rho^2(x,y) - |u(x)-u(y)|^2\right)_+, \quad x,y \in K. \tag{10}$$

We'd like to choose $C(u,L)$ to correct a solution by $r^2(x,y;u)$ on average. To this end, assume $du^i(x)du^j(y) = Q^{ij}(x,y)\,dt$ and use Itô's formula to compute

$$d\,|u(x)-u(y)|^2 = 2\,(u(x)-u(y))\cdot(du(x)-du(y)) + \Diamond Q\,dt. \tag{11}$$

The Itô correction, captured by the $\Diamond$ operator defined below, provides the expected bump up in lengths

$$(\Diamond Q)\,(x,y) := \sum_{j=1}^{q} \left(Q^{jj}(x,x) + Q^{jj}(y,y) - 2Q^{jj}(x,y)\right). \tag{12}$$

In order to correct by the metric defect, Q must satisfy the linear constraints

$$(\Diamond Q)\,(x,y) = r^2(x,y;u), \quad x,y \in K. \tag{13}$$

A second set of constraints is imposed by the Cameron-Martin theorem: the Gaussian measure associated to $\dot{L}$ must be absolutely continuous with respect to that of L. Explicitly, this means that $Q = ALA^T$ where A is a linear transformation, given in coordinates by $A^{ij}(x,y)$. This restriction is trivial when L has full-rank; but when L is rank-deficient, it provides $\mathbb{P}(n,q)$ with a sub-Riemannian geometry. We use the notation $Q \in T_L\mathbb{P}(n,q)$ to recognize this constraint.

These constraints describe a *matrix completion problem*: choose $Q \in T_L\mathbb{P}(n,q)$ that satisfies (13). This may not have a solution, so we introduce the convex set

$$\mathcal{P} = \{Q \in T_L\mathbb{P}(n,q) \,\big|\, \Diamond Q(x,y) \le r^2(x,y),\ x,y \in K\}. \tag{14}$$

Our model design task is to make a principled choice of a point $C(u,L)$ in $\mathcal{P}$.

Equation (7) is obtained by choosing a rank-one solution to the analogous operator completion problem for (4). But one may also use the theory of SDP to provide other resolutions of the above matrix completion problem. Interior point methods for SDP associate a barrier $F_{\mathcal{P}}$ to $\mathcal{P}$ and it is natural to choose

$$C(u,L) = \mathrm{argmin}_{Q\in\mathcal{P}} F_{\mathcal{P}}(Q). \tag{15}$$

When $F_{\mathcal{P}}$ is the *canonical barrier* associated to $\mathcal{P}$ we find that C is the *analytic center* of $\mathcal{P}$. This choice is similar in its minimalism to (7). The barrier $F_{\mathcal{P}}$ is a convex function on $\mathcal{P}$ whose Hessian $D^2F_{\mathcal{P}}$ provides a fundamental Riemannian metric on $\mathcal{P}$ [6]. It provides a natural microcanonical ensemble for embedding.

We may also introduce Gibbs measures on $T_L\mathbb{P}(n,q)$ that have the density

$$p_\beta(Q) = \frac{1}{Z_\beta} e^{-\beta E_r(Q)},\ Z_\beta = \int_{T_L\mathbb{P}(n,q)} e^{-\beta E_r(Q)}\, dQ,\ C = \int_{T_L\mathbb{P}(n,q)} Q\, p_\beta(Q)\, dQ, \tag{16}$$

where the energies E_r replace the constraints $\Diamond Q \leq r^2$ with suitable penalties. In these models, the covariance $C(u,L)$ is the most symmetric choice at scale t, with respect to the Gibbs measure p_β. As noted in Section 3.3, this is why these models may be termed renormalization group flows.

References

1. Borrelli, V., Jabrane, S., Lazarus, F., Thibert, B.: Flat tori in three-dimensional space and convex integration. Proc. Natl. Acad. Sci. USA **109**(19), 7218–7223 (2012). https://doi.org/10.1073/pnas.1118478109
2. De Lellis, C., Székelyhidi Jr., L.: Dissipative continuous Euler flows. Invent. Math. **193**(2), 377–407 (2013). https://doi.org/10.1007/s00222-012-0429-9
3. Friedan, D.: Nonlinear models in $2+\varepsilon$ dimensions. Ph.D. thesis, Lawrence Berkeley Laboratory, University of California (1980)
4. Friedan, D.: Nonlinear models in $2+\varepsilon$ dimensions. Phys. Rev. Lett. **45**(13), 1057–1060 (1980). https://doi.org/10.1103/PhysRevLett.45.1057
5. Gromov, M.: Geometric, algebraic, and analytic descendants of Nash isometric embedding theorems. Bull. Amer. Math. Soc. (N.S.) **54**(2), 173–245 (2017). https://doi.org/10.1090/bull/1551, https://doi.org/10.1090/bull/1551
6. Hildebrand, R.: Canonical barriers on convex cones. Mathematics of Operations Research **39**(3), 841–850 (2014)
7. Indyk, P., Matoušek, J., Sidiropoulos, A.: Low-distortion embeddings of finite metric spaces. Handbook of discrete and computational geometry **37**, 46 (2004)
8. Nash, J.: C^1 isometric imbeddings. Ann. of Math. (2) **60**, 383–396 (1954). https://doi.org/10.2307/1969840
9. Nash, J.: The imbedding problem for Riemannian manifolds. Ann. of Math. (2) **63**, 20–63 (1956). https://doi.org/10.2307/1969989
10. Nelson, E.: Dynamical theories of Brownian motion. Princeton University Press, Princeton, N.J. (1967)
11. Whitney, H.: Differentiable manifolds. Ann. of Math. (2) **37**(3), 645–680 (1936). https://doi.org/10.2307/1968482