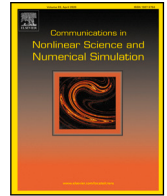




Contents lists available at ScienceDirect

# Communications in Nonlinear Science and Numerical Simulation

journal homepage: [www.elsevier.com/locate/cnsns](http://www.elsevier.com/locate/cnsns)

Research paper

## Preservation of dissipativity in dimensionality reduction

Sergey V. Stasenko<sup>a, b, \*</sup>, Alexander N. Kirdin<sup>b</sup><sup>a</sup> Artificial Intelligence Research Center, Lobachevsky University, Nizhny Novgorod, Russia<sup>b</sup> Institute of Computational Modelling, Russian Academy of Sciences, Siberian Branch, Krasnoyarsk, Russia

### ARTICLE INFO

#### Keywords:

Lyapunov function  
Dissipativity  
Dimensionality reduction  
Dynamical systems  
Adaptive control  
Optimization

### ABSTRACT

Systems with predetermined Lyapunov functions play an important role in many areas of applied mathematics, physics and engineering: dynamic optimization methods (objective functions and their modifications), machine learning (loss functions), thermodynamics and kinetics (free energy and other thermodynamic potentials), adaptive control (various objective functions, stabilization quality criteria and other Lyapunov functions). Dimensionality reduction is one of the main challenges in the modern era of big data and big models. Dimensionality reduction for systems with Lyapunov functions requires *preserving dissipativity*: the reduced system must also have a Lyapunov function, which is expected to be a restriction of the original Lyapunov function on the manifold of the reduced motion. An additional complexity of the problem is that the equations of motion themselves are often unknown in detail in advance and must be determined in the course of the study, while the Lyapunov function could be determined based on incomplete data. Therefore, the projection problem arises: for a given Lyapunov function, find a field of projectors such that the reduction of *any* dissipative system is again a dissipative system. In this paper, we present an explicit construction of such projectors and prove their uniqueness. We have also taken the first step beyond the approximation by manifolds. This is required in many applications. For this purpose, we introduce the concept of monotone trees and find a projection of dissipative systems onto monotone trees that preserves dissipativity.

### 1. Introduction

Dimensionality reduction is one of the main tasks in data analysis and mathematical modelling of complex system [1]. Many linear and non-linear methods were developed for solving this problem. Famous Principal Component Analysis (PCA) was invented by K. Pearson in 1901 [2] for data modelling and extraction essential features. Despite its very long history, this method still attracts a lot of attention from many researchers with the publication of a large number of practical recommendations and modifications [3].

Manifold learning is the first non-linear generalization of PCA [4,5]. Various methods of non-manifold low-dimensional data approximation have been also developed [6,7] and used for solving applied problems [8,9]. In addition to the injective data approximation methods, when the approximant is embedded in the dataspace, a large family of projective methods has been developed in which only the projection function onto the base space is learnt (see a brief review with a comparison of modern injective and projective methods in [10]).

The development of mathematical models of complex networks requires effective model reduction technology [11]. Many methods were created in the field of chemical dynamics [12] at the beginning of the twentieth century and then received mathematical justification. Here we should mention the 1956 Nobel Prizes awarded to Semenov and Hinshelwood, for research

\* Corresponding author.

E-mail addresses: [stasenko@neuro.nnov.ru](mailto:stasenko@neuro.nnov.ru) (S.V. Stasenko), [kirdinalexandergp@gmail.com](mailto:kirdinalexandergp@gmail.com) (A.N. Kirdin).

<https://doi.org/10.1016/j.cnsns.2024.108553>

Received 7 October 2024; Received in revised form 14 December 2024; Accepted 14 December 2024

Available online 20 December 2024

1007-5704/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

based on effective dimensionality reduction by the method of quasi-stationary concentrations (for more information, we refer to the reviews of [13,14] and the textbook [15]). The very popular machine learning algorithms for nonlinear dimensionality reduction and manifold learning, autoencoders, were invented first for chemical applications [16,17].

Two very general mathematical ideas are used in dimensionality reduction for large dynamical systems: *invariant manifold* and *Lyapunov function*. The motion reduced to an invariant manifold never leaves it. If this manifold is, at the same time, a slow manifold, then the game is over and the reduced description is ready [18,19]. Methods of invariant manifold aim to find a good approximation to the unknown slow invariant manifold.

Lyapunov functions monotonically decrease (non increase) along the systems trajectories. They appear in various forms: entropy or free energy (for physical and chemical systems [18,20–22]) and informational entropy or various divergences for machine learning [23–27], generalized fitness (for evolutionary modelling [28–30]), loss function [31] (for machine learning), or just general Lyapunov functions for many general problems of control, machine learning and adaptation in dynamical systems [32].

Information about the Lyapunov function for large dynamical systems may be much more reliable than the information about dynamic (vector field) itself. In physical and chemical applications we can operate by entropy, free energy and other thermodynamic potentials even when the detailed kinetics is not well-known. Similarly, in machine learning we can define the loss functions before detailed specification of the learning algorithm. In adaptive control and observation problem, we also define the quality functionals before construction of observers and the adaptation algorithm.

Definition of the Lyapunov function for a reduced system is very simple: we can just restrict the Lyapunov function of the initial system on the manifold of reduced description. But projection of the original dynamics is not so obvious. Of course, if the manifold of the reduced description is invariant with respect to original dynamics, then the original vector field is tangent to this manifold and belongs to its tangent space. In this case, there will be no problems with the projector. But in practice the manifolds of reduced description are not invariant beyond oversimplified toy examples. Therefore, we come to the projection problem: how should we project the original vector field onto the tangent space of an “ansatz manifold” in order to guarantee preservation of dissipativity: if the time derivative of the Lyapunov function according to the original vector field is non-positive, then the time derivative of the restriction of this Lyapunov function according to the projected vector field should be also non-positive.

This projection problem becomes interesting and non-trivial if we consider it in the universal settings:

- Let the vector field be a priori unknown; the only available a priori information is: it has a given Lyapunov function;
- Let the ansatz manifold be *a priori* unknown; the only available information is some regularity of intersections with the Lyapunov function levels (the detail follow);
- Find a projector that transforms any *a priori* unknown vector field with a given Lyapunov function into a vector field on a regular ansatz manifold equipped with a Lyapunov function, which is a simple restriction of the original Lyapunov function on this ansatz manifold.

Such a universal approach arose from the most universal physical discipline – thermodynamics. The main example was the theory of strong shock waves. The original (non-reduced) system was the general Boltzmann equation. In the middle of 20th century, Tamm and Mott-Smith proposed a very simple and useful ansatz for analysis of strong shock waves. The distribution function of particle speeds was presented as a linear combination of Maxwellian distribution at the “beginning” ( $t \rightarrow -\infty$ ) and at the “end” ( $t \rightarrow +\infty$ ). The new variable of reduced description is just the coefficient of this combination. Then the Boltzmann equation was projected onto this ansatz. The result was useful and sufficiently simple for analytic and numerical analysis (see, for example, the book [33]). Nevertheless, the projected equation may violate the Second law of thermodynamics: in some situations, entropy decreases [34]. In 1977, M. Lampis proposed a projector for the Tamm–Mott-Smith ansatz that guaranteed preservation of dissipativity [35]. Later on, Gorban and Karlin proposed a general thermodynamic theory for the Tamm–Mott-Smith ansatz and similar approximations [36]. In their work, the quasi-one dimensional nature of the Tamm–Mott-Smith ansatz was used. Some generalization were proposed in 1994 for solving the problem of instabilities of post-Navier–Stokes approximations [37]. In 2004, Gorban and Karlin published a general construction of thermodynamic projector for more general ansatz manifolds [38] and applied it to Hilbert’s 6th problem for analysis of hydrodynamic ansatz manifolds in kinetic theory [19].

There are three surprises:

1. The universal projector that preserves dissipativity does exist;
2. This projector is unique: for all other projectors there are counterexamples with non-monotone dynamic of Lyapunov function along some trajectories of reduced systems;
3. This projector preserves not only the sign of the Lyapunov function time derivative (that is the initial requirement) but also the exact value of this derivative. This is a rare case of a “free lunch”.

In our work, we analyse the mathematical problem of preservation of dissipativity in general dynamical models with a given Lyapunov function, give the explicit formula for the universal projector, and prove that it is unique. In Section 2 we give a rigorous problem statement. In Section 3 the main theorem about universal projector is proven. In Section 4 we consider the model reduction methods beyond manifolds. For this purpose, we introduce *monotone trees* and demonstrate how the results of the previous section can be used for unambiguous projection of dissipative dynamics onto such trees. In Section 5 we discuss the results and the possible directions of future work.

## 2. Notations and formal problem statement

Let  $U$  be an open convex subset of  $n$ -dimensional Euclidean space  $\mathbb{R}^n$ . The *a priori* given Lyapunov function  $H$  is given in  $U$ . Assume that:

- $H$  is twice-differentiable.
- $H$  is strongly convex in  $U$ , that is, the second differential of  $H$ ,  $D_x^2(H)$  is uniformly positive definite in  $U$ : all its eigenvalues  $\lambda$  are positive and  $\lambda > \alpha > 0$  for some  $\alpha$  for all  $x \in U$ .

In a coordinate system  $x_1, \dots, x_n$ , the second differential is presented by the Hessian matrix

$$D_x^2(H) = \frac{\partial^2 H}{\partial x_i \partial x_j}.$$

An  $m$ -dimensional ansatz manifold  $M$  is a twice-differentiable immersed  $m$ -dimensional submanifold in  $U$ .

Because most of our problems are local, it is convenient to work with a coordinate chart of  $M$ . It is an injective immersion  $F : B_m \rightarrow U$ , where  $B_m$  is an open convex subset of  $m$ -dimensional Euclidean space  $\mathbb{R}^m$ . This means that  $F$  is injective and the differential of  $F$  at each point  $p \in B_m$ ,  $D_p(F)$ , is an injective linear map from  $\mathbb{R}^m$  to  $\mathbb{R}^n$ .

For given coordinate systems in spaces  $\mathbb{R}^m (p_1, \dots, p_m)$  and  $\mathbb{R}^n (x_1, \dots, x_n)$ , the map  $F$  is given by  $n$  functions on  $B_m$ :  $x_i = f_i(p_1, \dots, p_m)$  ( $i = 1, \dots, n$ ).

The differential  $D_p(F)$  is presented by the matrix of partial derivatives  $\left. \frac{\partial f_i}{\partial p_j} \right|_p$ .

$F$  is immersion if and only if rank of  $D_p$  is constant and equal to  $m$ :  $\text{rank}(D_p(F)) = m$  for all  $p \in B_m$ .

Of course, we can always assume that  $B_m$  is an open  $m$ -dimensional ball. Nevertheless, a bit more general definition with open convex sets may be useful for some examples.

There is no simple general local criterion for the injectivity of  $F$ , but for some practical examples it can often be proved without difficulty. (For example, for the Tamm–Mott-Smith ansatz discussed above: If for two pairs of Maxwellians two linear combinations of them with non-zero coefficients coincide, then these pairs coincide too because of linear independence of different Maxwellians. This means injectivity of the Tamm–Mott-Smith ansatz.)

The tangent space to  $M$  at a point  $x = F(p)$  is the image of the differential  $D_p(F): T|_{F(p)}F(B_m) = \text{im}D_p(F)$ . In the coordinate form, it can be presented as a linear span of linear independent vectors of partial derivatives (columns of the matrix  $\left. \frac{\partial f_i}{\partial p_j} \right|_p$ ):

$$T_{F(p)} = \text{span} \left\{ \left( \frac{\partial x_1}{\partial p_i}, \frac{\partial x_2}{\partial p_i} \dots \frac{\partial x_n}{\partial p_i} \right), \quad i = 1, \dots, m \right\}.$$

The tangent space of  $U$  at any point may be considered just as  $\mathbb{R}^n$ . The first differential of  $H$  in  $U$ ,  $D_x H$ , is a linear functional: for each  $y \in \mathbb{R}^n$

$$D_x(H)(y) = \sum_i \frac{\partial H}{\partial x_i} y_i.$$

Restriction of  $H$  on  $M$  at point  $x = F(p)$  is just the value of  $H: H|_M(F(p)) = H(F(p))$ . Differential of restriction of  $H$  on  $M$  at point  $x = F(p)$  is restriction of the functional  $D_x(H)$  on  $T_{F(p)}$ . Formally, if a vector  $y \in T_{F(p)}$ , this means

$$y = \sum_{i=1}^m a_i \left( \frac{\partial x_1}{\partial p_i}, \frac{\partial x_2}{\partial p_i} \dots \frac{\partial x_n}{\partial p_i} \right)$$

for some numbers  $a_1, \dots, a_m$ . For this  $y$ ,

$$D_x(H_M)(y) = \sum_{i=1}^m a_i \sum_{j=1}^n \frac{\partial H}{\partial x_j} \frac{\partial x_j}{\partial p_i}.$$

If we transform the basis in  $T|_{F(p)}$  to  $\{e_1, \dots, e_m\}$  then for the calculation of  $D_x(H_L)(y)$  for  $y = \sum b_i e_i$  we should use representation of  $e_i$  in the old basis to find  $D_x(H)(e_i)$  and then calculate  $D_x(H_M)(y) = \sum b_i D_x(H)(e_i)$ .

The second differential of  $H$  induces a Riemannian metric on  $U$  for which the Hessian of  $H$  is the metric tensor. This is the so-called Shahshahani metric [39]. The inner product associated with this metric in the tangent space at a point  $x \in U$  is defined as

$$\langle y|z \rangle_x = (y, D_x^2(H)z),$$

or in the coordinate form

$$\langle y|z \rangle_x = \sum_{i,j=1}^n y_i \frac{\partial^2 H}{\partial x_i \partial x_j} \Big|_x z_j$$

Of course, the Shahshahani metric generated by the second differential of  $H|_M$  is also defined on the submanifold  $M$ . We use notation  $P_x^\perp$  for the orthogonal projectors of  $\mathbb{R}^n$  onto  $T_x(M)$  in the Shahshahani metrics  $\langle \cdot | \cdot \rangle_x$ . These projectors do not solve the main problem but are used in the solution.

Consider a vector field  $W(x)$  in  $U$ . If it has Lyapunov function  $H$  then the derivative of  $H$  along  $W(x)$  is non-positive. In our notations, it means that

$$D_x(H)(W(x)) \leq 0 \text{ for all } x \in U, \tag{1}$$

or in the coordinate form

$$\sum_i w_i(x) \left. \frac{\partial H}{\partial x_i} \right|_x \leq 0 \tag{2}$$

if  $w_i(x)$  are the components of the vector field  $W(x)$ .

Consider a vector field  $Q(x) \in T_{F(p)}$  defined at points  $x = F(p)$  of the submanifold  $M$ . Definition of the derivatives of  $H_M$  along this field is the same as (1), (2). Taking into account that  $x \in M$  and  $H_M = H$  on  $M$ , we write

$$D_x(H_M)(Q(x)) = D_x(H)(Q(x)).$$

The problem is: For  $x \in M$  find projectors  $P_x : \mathbb{R}^n \rightarrow T_x M$  such that for any dissipative vector field  $W(x)$  (1), (2) with Lyapunov function  $H$  the projected vector field  $Q(x)$  is also dissipative with the restriction of  $H$  on  $T_x(M)$ ,  $H_M$ .

The projectors  $P_x$  depend on point  $x \in M$ . They also depend on the tangent space  $T_x M$ . We require also that the yet unknown field of projectors should be smooth. It is usually clear, for which manifold  $M$  and tangent space  $T_x M$  a projector  $P_x$  is build, and we omit the cumbersome notation, but if it is necessary, we can use  $P_{x,M}$  for projector of  $\mathbb{R}^n$  onto  $T_x M$ .

### 3. Universal projector for submanifolds

#### 3.1. Linear submanifolds, linear vector fields, and quadratic Lyapunov functions

Linear vector fields with quadratic Lyapunov functions and linear ansatz submanifold provide us with a simplified version of the problem of dissipative projection and, at the same time, demonstrate how to use the Shahshahani metrics for the construction of the projector.

Let  $U = \mathbb{R}^n$  and  $H$  be a positive definite quadratic form  $H(x) = (x, Gx)$ , where  $G$  is a positively definite symmetric operator and  $(\cdot, \cdot)$  is the standard inner product. We select the auxiliary (Shahshahani) inner product  $\langle y|z \rangle = (y, Gz) = (Gy, z)$ .

Consider a linear vector field  $Q(x) = Qx$ , where  $Q$  is a linear operator. It is well-known that this vector field has Lyapunov functions  $H$  if and only if the time derivative of  $H$  along  $Q$  is non-positive: for all  $x \in \mathbb{R}^n$

$$\begin{aligned} \frac{d}{dt}(x, Gx) &= (\dot{x}, Gx) + (x, G\dot{x}) \\ &= (Qx, Gx) + (x, GQx) = (x, (Q^T G + GQ)x) \leq 0. \end{aligned} \tag{3}$$

This means that the symmetric operator  $Q^T G + GQ$  is negative semi-definite. The equivalent inequality is

$$\langle Qx|x \rangle \leq 0 \text{ for all } x \in \mathbb{R}^n. \tag{4}$$

A linear ansatz manifold is just a linear subspace in  $\mathbb{R}^n$ . In this case we can use for the tangent space  $T(M)$  the same notation  $M$ . A projector  $\mathcal{P} : \mathbb{R}^n \rightarrow M$  is a surjective linear operator ( $\text{im}\mathcal{P} = M$ ) with the property  $\mathcal{P}^2 = \mathcal{P}$ , or, which is the same, such a linear operator  $\mathcal{P} : \mathbb{R}^n \rightarrow M$  that its restriction on  $M$  is the unite operator:  $\mathcal{P}|_M = \text{id}_M$ .

For a given linear ansatz manifold  $M$ , projector  $\mathcal{P} : \mathbb{R}^n \rightarrow M$  and a linear vector field  $Q(x) = Qx$  on  $\mathbb{R}^n$ , the projected vector field on  $M$  is  $\mathcal{P}Qx$ ,  $x \in M$ . It is convenient to represent this projection in a more symmetric form:  $\mathcal{P}Qx = \mathcal{P}Q\mathcal{P}x$  for  $x \in M$ .

**Theorem 1.** *Projector  $\mathcal{P} : \mathbb{R}^n \rightarrow M$  projects any linear vector field  $Qx$  with quadratic Lyapunov function  $H(x) = \langle x|x \rangle$  into a linear vector field on a linear subspace  $M$  with the same quadratic Lyapunov function  $H_M(x) = \langle x|x \rangle$  (for  $x \in M$ ) if and only if  $\mathcal{P}$  is an orthogonal projector  $\mathcal{P}^\perp$  in the Shahshahani inner product  $\langle y|z \rangle$ .*

**Proof.** Assume that the projector  $\mathcal{P}$  is an orthogonal projector in the Shahshahani metric (image  $\mathcal{P}$  is orthogonal to kernel  $\mathcal{P}$ ). An equivalent condition is that projector  $\mathcal{P}$  is a self-adjoint operator in this inner product:  $\langle \mathcal{P}y|z \rangle = \langle y|\mathcal{P}z \rangle$  for all  $y, z \in \mathbb{R}^n$ . Let an operator  $Q$  satisfy the dissipativity condition (4). Then for its projection onto  $M$  we get  $\langle \mathcal{P}Q\mathcal{P}x|x \rangle = \langle Q\mathcal{P}x|\mathcal{P}x \rangle$  because  $\mathcal{P}$  is self-adjoint. This form is negative semi-definite because  $\langle Qy|y \rangle \leq 0$  for all  $y \in \mathbb{R}^n$ .

Assume now that a projector  $\mathcal{P} : \mathbb{R}^n \rightarrow M$  is not an orthogonal projector in the Shahshahani metric. Then the orthogonal complement to  $M = \text{im}\mathcal{P}$  does not coincide with the kernel of  $\mathcal{P}$  and there exists such a vector  $y \in \mathbb{R}^n$  that  $\langle y|z \rangle = 0$  for all  $z \in M$  but  $\mathcal{P}y \neq 0$ . Consider a one-parametric family  $A_a$  of auxiliary rank one operators: for any  $x \in \mathbb{R}^n$

$$A_a x = -(\mathcal{P}y - ay)(\mathcal{P}y - ay|x).$$

These operators are negatively semi-definite:  $\langle x|A_a x \rangle = -\langle x|\mathcal{P}y - ay \rangle^2 \leq 0$ . Calculate the projection of these operators onto  $M$ . Note that, for  $x \in M$ ,  $A_a x = -(\mathcal{P}y - ay)(\mathcal{P}y - ay|x) = -(\mathcal{P}y - ay)(\mathcal{P}y|x)$  because  $\langle y|x \rangle = 0$ . Projection of these vectors on  $M$  gives  $\mathcal{P}A_a x = -(1-a)\mathcal{P}y(\mathcal{P}y|x)$ . For  $a > 1$  these operators are not negatively semi-definite:  $\langle x|\mathcal{P}A_a x \rangle = (a-1)\langle \mathcal{P}y|x \rangle^2$ .  $\square$

Thus, for linear submanifolds (subspaces) and quadratic Lyapunov functions preservation of dissipativity in reduced systems requires orthogonal projection in the Shahshahani metrics. The next step is the analysis of the problem near the equilibrium that is the global minimizer of  $H$  in  $U$ .

### 3.2. Extension of dissipative vector fields

Transition from local analysis (in the vicinity of a point or a compact) to global statements is a very common problem. In general topology, continuation of a continuous function from a closed subset to the entire space is given by the Tietze–Urysohn–Brouwer extension theorem. For differentiable functions, the Whitney extension theorem states that it is possible to extend a differentiable function with given derivatives from a closed subset in  $\mathbb{R}^n$  to the whole space [40,41].

We have to extend a locally defined twice differentiable dissipative vector field to such a field in the entire  $U$ . The idea is very simple: we will use the classical theorems about extension of differentiable functions and “glue” the locally defined dissipative field  $Q(x)$  with the globally defined field  $\text{grad}H$ .

**Theorem 2.** *Let  $K$  be compactly contained in  $U$ :  $K \Subset U$ , and its  $\varepsilon$  neighbourhood  $K_\varepsilon = K + \varepsilon B^n$  (where  $B^n$  is the unite ball in  $\mathbb{R}^n$ ) is also compactly contained in  $U$ . If  $Q$  is a dissipative vector field in  $K_\varepsilon$  then there exist a dissipative vector field  $Q^{\text{ext}}$  in  $U$  such that  $Q(x) = Q^{\text{ext}}(x)$  for  $x \in K$ .*

**Proof.** According to the extension theorems for differentiable functions, there exists such an infinitely smooth function  $f_{K,\varepsilon}$  on  $U$  that  $f_{K,\varepsilon}(x) \geq 0$  in  $U$ ,  $f_{K,\varepsilon}(x) = 1$  for  $x \in K$  and  $f_{K,\varepsilon}(x) = 0$  for  $x \notin K_\varepsilon$ . Define  $Q^*(x) = f_{K,\varepsilon}(x)Q(x)$  for  $x \in K_\varepsilon$  and  $Q^*(x) = 0$  for  $x \notin K_\varepsilon$ . For any globally dissipative vector field  $W$  in  $U$ , the linear combination  $Q^{\text{ext}}(x) = Q^*(x) + (1 - f_{K,\varepsilon}(x))W$  is a globally dissipative vector field in  $U$ . If  $x \in K$  then  $Q^{\text{ext}}(x) = Q(x)$ , and if  $x \notin K_\varepsilon$  then  $Q^{\text{ext}}(x) = W(x)$ .  $\square$

This theorem allows us to discuss locally dissipative vector fields and then extend the results to the entire  $U$ . In particular, it gives the possibility to extend Theorem 1 from exactly linear submanifolds, linear vector fields, and quadratic Lyapunov functions to vicinities of equilibria with almost linear submanifolds, almost linear vector fields, and almost quadratic Lyapunov functions.

A special case is the extension of a dissipative vector field from a submanifold in  $U$  to the entire  $U$ . Further, we use the extension of locally defined vector fields in the vicinity of a point  $x \in M$ . Let  $Q$  be a smooth dissipative vector field on  $M$  in the  $\varepsilon$ -vicinity of  $x$ , that is,  $D_y(H_M)(Q) \leq 0$  for  $y \in M$  and  $\|x - y\| < \varepsilon$  ( $\varepsilon > 0$ ).

**Proposition 1.**

1. *If  $D_x(H_M)(Q(x)) < 0$  then there exists a dissipative vector field  $Q^{\text{ext}}$  in  $U$  such that  $Q(y) = Q^{\text{ext}}(y)$  for  $y \in M$  and  $\|x - y\| < \delta$  for some  $\delta > 0$ .*
2. *If  $Q(x) = 0$  but  $D_y(H_M)(Q(y)) < -\lambda\|x - y\|$  for  $y \in M$ ,  $\|x - y\| < \varepsilon$  and some  $\lambda > 0$ , then there exists a dissipative vector field  $Q^{\text{ext}}$  in  $U$  such that  $Q(y) = Q^{\text{ext}}(y)$  for  $y \in M$  and  $\|x - y\| < \delta$  for some  $\delta > 0$ .*

**Proof.** Locally, in a small vicinity of  $x$  in  $U$ , we can introduce such a coordinate systems  $y = (y_1, \dots, y_n)$  that  $M$  is given by the equations  $y_{m+1} = \dots = y_n = 0$ , where  $m = \dim M$  ( $\|y - x\| < \kappa$ ,  $y \in U$  for some  $\kappa > 0$ ). Locally, for  $\|y - x\| < \kappa$ ,  $(y_1, \dots, y_m)$  is a coordinate system on  $M$ . Let us use this coordinate system.

Let us prove statement 1. If  $Q(x) \neq 0$  then  $x$  is not a critical point of  $H_M$  on  $M$ . Indeed, if  $x$  is a critical point of  $H_M$  on  $M$  then, because of the strong convexity of  $H$ ,  $x$  is a point of local minimum of  $H_M$  on  $M$ , and  $D_{x+\gamma z}(H_M)(Q(x + \gamma z)) > 0$  and  $D_{x-\gamma z}(H_M)(Q(x - \gamma z)) < 0$  for some  $z = (z_1, \dots, z_m, 0, \dots, 0)$  and sufficiently small  $\gamma > 0$ . Let us use the standard theorem about rectification of vector fields (see, for example, [42,43]). In a small vicinity of  $x$  we can transform the coordinate system  $(y_1, \dots, y_m)$  on  $M$  in such a way that  $Q(y) = (1, 0, \dots, 0)$  ( $m - 1$  zeros,  $y \in M$ ,  $\|x - y\| < \nu$  for some  $\nu < 0$ ). This transformation gives, at the same time, an extension of  $Q(y)$  onto a vicinity of  $x$  in  $U$ :  $Q(y) = (1, 0, \dots, 0)$  ( $n - 1$  zeros,  $y \in U$ ,  $\|x - y\| < \zeta$  for some  $\zeta > 0$ ).

Thus, we found an extension of  $Q$  onto a vicinity of  $x$  in  $U$ . After that, we can apply Theorem 2. The first statement is proven.

Let us prove statement 2. If  $Q(x) = 0$  then the used above simple version of rectification theorem is not applicable. For any given sufficiently small  $(y_{m+1}, \dots, y_n)$ , the set  $\{(y_1, \dots, y_m, y_{m+1}, \dots, y_n)\}$  with variable  $y_1, \dots, y_m$  in a vicinity of  $x$  is an  $m$ -dimensional immersed submanifold. In the selected coordinate system, this submanifold is just a shift of a vicinity of  $x$  in  $M$  by vector  $(0, \dots, 0, y_{m+1}, \dots, y_n)$ . We use for the result of this shift the notation  $M(y_{m+1}, \dots, y_n)$ . In this notation, a sufficiently small vicinity of  $x$  in  $M$  is  $M(0, \dots, 0)$  ( $n - m$  zeros). For convenience, put  $x$  at the origin:  $x = (0, \dots, 0)$  ( $n$  zeros).

According to the assumptions,  $Q(x) = 0$ , but  $D_y(H_M)(Q(y)) < -\lambda\|x - y\|$  for  $y \in M$ ,  $\|x - y\| < \varepsilon$  and some  $\lambda > 0$ . Therefore  $x$  is a critical point of  $H_M$  on  $M$ . Due to the strong convexity of  $H$  on  $U$ , we can use the implicit function theorem and find the critical point  $x^*(y_{m+1}, \dots, y_n)$  of  $H_{M(y_{m+1}, \dots, y_n)}$  on  $M(y_{m+1}, \dots, y_n)$ , such that  $x^*(y_{m+1}, \dots, y_n)$  is a smooth function and  $x^*(0, \dots, 0) = x$ .

Let us define the extended vector field in the vicinity of  $x$  in  $U$  (extension from dimension  $m$  to dimension  $n$ ):

$$Q_n(y) = Q(y - x^*(y_{m+1}, \dots, y_n)).$$

For this vector field,  $D_y(H)(Q(y)) < -\lambda'\|x - y\|$  for  $\|x - y\| < \varepsilon'$  and some  $\lambda' > 0$ ,  $\lambda' < \lambda$ ,  $\varepsilon' < \varepsilon$ . For the further extension to the entire  $U$  we use Theorem 2. The second statement is proven.  $\square$

### 3.3. Near equilibrium

Assume that  $x^{eq} \in U$  is the global minimizer of  $H$  in  $U$ . Each dissipative vector fields is zero at  $x^{eq} \in U$ . Therefore, there exist no standard “pointwise” restriction on  $\mathcal{P}_{x^{eq}}$ : each projector projects zero into zero. Nevertheless, the dissipativity in the neighbourhood of  $x^{eq} \in U$  defines  $\mathcal{P}_{x^{eq}}$  unambiguously.

To simplify the notation, let us choose the origin at the equilibrium point:  $x^{eq} = 0$ . notice that  $H(x) = \frac{1}{2}\langle x|x \rangle_0 + o(\|x\|^2)$ . Near the equilibrium, the linear functional  $D_x(H)(z)$  can be approximated as  $\langle x|z \rangle_0$ :

$$|D_x(H)(z) - \langle x|z \rangle_0| < o(\|x\|)\|z\|.$$

Following [Theorem 1](#), we can guess that  $\mathcal{P}_{x^{eq}} = \mathcal{P}_{x^{eq}}^\perp$ . To prove this guess, we have to consider projections in a vicinity of  $x^{eq}$  because, exactly at equilibrium, all dissipative vector fields vanish. Near  $x^{eq}$ , projectors that preserve dissipativity are close to  $\mathcal{P}_{x^{eq}}$ . The following theorem formalizes our guess.

**Theorem 3.** *For preservation of dissipativity by differentiable projector field  $\mathcal{P}_x$ , it is necessary that*

$$\mathcal{P}_x = \mathcal{P}_x^\perp + O(\|x - x^{eq}\|). \tag{5}$$

**Proof.** The equivalent formulation of this theorem is:  $\mathcal{P}_{x^{eq}} = \mathcal{P}_{x^{eq}}^\perp$ . At equilibrium, the projector is just an orthogonal projector in the Shahshahani metrics. Recall that the Hessian of  $H$  is strictly negative definite at any point  $x \in U$ .

Let us construct a proof by contradiction. Assume that  $M$  is a manifold of reduced description and  $0 \in M$ . Let  $\mathcal{P}_{0,M}$  be not an orthogonal projector in the Shahshahani inner product  $\mathcal{P}_{0,M} = \tilde{\mathcal{P}} \neq \mathcal{P}_0^\perp$ . The images of these projectors coincide with  $T_0(M)$ . Therefore, they should have different kernels: there exists such a vector  $y \in \mathbb{R}^n$  that  $\langle y|z \rangle_0 = 0$  for all  $z \in T_0(M)$  but  $\tilde{\mathcal{P}}y \neq 0$ , exactly as in the proof of [Theorem 1](#).

Consider a one-parametric family of auxiliary vector fields in the vicinity of the equilibrium:

$$B_a(x) = -(\mathcal{P}_x y - ay)D_x(H)(\mathcal{P}_x y - ay).$$

Here we assume that projectors  $\mathcal{P}_x$  project dissipative vector fields on  $\mathbb{R}^n$  into a dissipative vector field on  $T_x(M)$  (in our simplified notation,  $T_x(M) = M$ ). According to the problem statement, this field of projectors should be continuous and  $\|\mathcal{P}_x - \mathcal{P}_0\| = O(\|x\|)$ .

These vector fields are dissipative:

$$D_x(H)(B_a(x)) = -(D_x(H)(\mathcal{P}_x y - ay))^2 \leq 0.$$

Let us demonstrate that projection of  $B_a(x)$  on  $M$  is not dissipative. For  $x \in M$ ,  $\mathcal{P}_x B_a(x) = (a - 1)\mathcal{P}_x y D_x(H)(\mathcal{P}_x y - ay)$  and  $D_x H(\mathcal{P}_x y - ay) = \langle x|\mathcal{P}_x y \rangle_0 + o(\|x\|)$ .

Therefore, for  $x \in M$ , the time derivative of  $H$  according to the projected vector field is

$$D_x(H)(\mathcal{P}_x B_a(x)) = (a - 1)\langle x|\mathcal{P}_x y \rangle_0^2 + o(\|x\|^2).$$

This expression is positive for  $a > 1$  and some sufficiently small  $x$ , for example, for  $x = \varepsilon \mathcal{P}_x y$  and sufficiently small  $\varepsilon > 0$ . In these cases, the projected vector field is not dissipative despite the dissipativity of the original vector field in  $U$ .

According to [Theorem 2](#), vector fields dissipative in some vicinity of the equilibrium can be extended to dissipative vector fields on the entire  $U$ . Therefore, we proved that  $\mathcal{P}_{x^{eq}} = \mathcal{P}_{x^{eq}}^\perp$  and  $\mathcal{P}_x = \mathcal{P}_x^\perp + O(\|x - x^{eq}\|)$  for projectors that preserve dissipativity.  $\square$

This theorem tells us that the difference between the projectors that preserve dissipativity of all dissipative vector fields and the orthogonal projector (in the Shahshahani metrics) grows out of equilibrium not faster than  $O(\|x - x^{eq}\|)$ .

For linear submanifolds  $L \subset \mathbb{R}^n$ , [Theorem 3](#) will be applied also to dissipative vector fields on  $L \cap U$  of  $\mathbb{R}^n$  near minima of  $H$  on  $L \cap U$  without changes.

### 3.4. One-dimensional submanifolds

The orthogonal projectors in the Shahshahani metrics solve sometimes the problem of projection with preservation of dissipativity, but already simple example show that this is not the general solution of the problem. Already projections onto one-dimensional manifolds (curves) require a different approach. Reduction to essentially one-dimensional cases is not an exotic and oversimplified example. The famous Tamm – Mott-Smith approximation leads to this problem.

Return to the general problem statement (Section 2). Let the manifold of the reduced description  $M$  be one-dimensional (a curve). Its tangent space at point  $x \in M$  is a one-dimensional subspace  $T_x(M) \subset \mathbb{R}^n$ . Select a normalized vector  $e_x \in T_x(M)$ ,  $\langle e_x|e_x \rangle = 1$ . A projector  $\mathcal{P}$  can be presented by two subspace:  $\ker \mathcal{P}$  and  $\text{im} \mathcal{P}$  under conditions that

$$\ker \mathcal{P} \cap \text{im} \mathcal{P} = \{0\} \text{ and } \dim \ker \mathcal{P} + \dim \text{im} \mathcal{P} = n \tag{6}$$



The image of  $\mathcal{P}_x$  is, obviously,  $T_x(M)$ . In this subsection, it is the straight line:  $T_x(M) = \{\alpha e_x | \alpha \in \mathbb{R}\}$ , where  $e_x$  is a basis vector of  $T_x(M)$ . The following theorem gives us the kernel of  $\mathcal{P}_x$ .

**Theorem 4.** Assume that the manifold of the reduced description  $M$  is one-dimensional and  $D_x(H)(e_x) \neq 0$ . Then for the preservation of dissipativity in the projection for all dissipative vector fields it is necessary and sufficient that  $\text{im}\mathcal{P}_x = T_x M$  and  $\text{ker } \mathcal{P}_x = \text{ker } D_x(H)$ .

**Proof.** Let  $\text{im}\mathcal{P}_x = T_x(M)$  and  $\text{ker } \mathcal{P}_x = \text{ker } D_x(H)$ . Any linear operator  $\mathcal{P}$  with given one-dimensional image  $M_x$  and kernel  $\text{ker } D_x(H)$  can be presented as

$$\mathcal{P}Q = ae_x D_x(H)(Q)$$

with a real parameter  $a$ . For the projector  $\mathcal{P}_x$  this parameter can be defined using the projector property  $\mathcal{P}_x^2 = \mathcal{P}_x$ :

$$a^2 e_x (D_x(H)(e_x)) D_x(H)(Q) = ae_x D_x(H)(Q).$$

Therefore,  $a = (D_x(H)(e_x))^{-1}$  and

$$\mathcal{P}_x Q = e_x \frac{D_x(H)(Q)}{D_x(H)(e_x)}. \tag{7}$$

This is the projector with the given kernel and image.

The time derivative of  $H_M$  along the projected vector field  $\mathcal{P}_x Q$  (7) is

$$D_x(H)(\mathcal{P}_x Q) = D_x(H)(e_x) \frac{D_x(H)(Q)}{D_x(H)(e_x)} = D_x(H)(Q).$$

We can see that the time derivative of  $H$  does not change after dimensionality reduction with the projector  $\mathcal{P}_x Q$  (7). This is even more than we were looking for: Preservation of the dissipativity assumes preservation of the sign of  $dH/dt$ , and here we see preservation of the value.

All projectors onto  $T_x(M)$  have the same image,  $T_x(M)$ , but the kernels might be different under conditions (6). Let us demonstrate that all kernels of projector other than the null-space of the differential of  $H$ ,  $\text{ker } D_x(H)$ , lead to the violation of dissipativity for some vector fields  $Q$  in projection onto  $T_x(M)$ .

Consider a projector  $\tilde{\mathcal{P}}_x : \mathbb{R}^m \rightarrow T_x(M)$  that is different from  $\mathcal{P}_x$  given by formula (7). The general form of a projector on a one-dimensional subspace is  $\tilde{\mathcal{P}}_x(Q) = e_x l(Q)$ , where  $l$  is a linear functional, and  $l(e_x) = 1$  because of the projector condition  $\mathcal{P}^2 = \mathcal{P}$ . Due to the projection of a vector field  $Q$  by the projector  $\tilde{\mathcal{P}}_x$ , the time derivative of  $H_M$  is  $D_x(H)(e_x)l(Q)$ . The preservation of dissipativity requires that this time derivative is non-positive when the original dissipations,  $D_x(Q)$ , is non-positive. This means that the half-spaces in  $\mathbb{R}^n$  given by the inequalities  $\{y \in \mathbb{R}^n | l(y) \leq 0\}$  and  $\{y \in \mathbb{R}^n | D_x(H)(y) \leq 0\}$  coincide. For two linear functionals this is possible if and only if they are proportional with a positive factor. Hence, the preservation of dissipativity requires  $l = aD_x(H)$ ,  $a > 0$ , and the projector  $\tilde{\mathcal{P}}$  coincides with  $\mathcal{P}_x$  (7).

Thus, the requirement of preservation of dissipativity unambiguously defines projector  $\mathcal{P}_x$  (7). Moreover, the preservation of the sign of dissipation implies the preservation of the value of  $dH/dt$  in the reduces system:  $dH/dt = dH_M/dt$ .  $\square$

Informally, the proven construction of projector (7) for one-dimensional manifolds  $M$ , can be characterized by choosing  $H_M$  as the internal coordinate  $h$  on  $M$  and defining the projector  $\mathcal{P}_x$  as follows: for a vector field  $W(x)$

$$\frac{dh}{dt} = \frac{dH}{dt} = D_x(U)(W(x)).$$

Here on the left we see movement on  $M$  due to the projected vector field (a reduced system), and on the right we see the time derivative of  $H$  due to the vector field  $W$  in the original system.

Formula for one-dimensional projections (7) is important for applications, therefore we reproduce it in the coordinate form.

Let a one-dimensional manifold (a curve) be given by  $n$  functions of one variable:  $x = (f_1(p), \dots, f_n(p))$ . Here  $p$  is a real variable (parameter), and  $f_i$  are the twice differentiable functions.

- For a given  $x = (f_1(p), \dots, f_n(p))$ ,  $e_x$  is a vector of derivatives,  $e_x = (df_1(p)/dp, \dots, df_n(p)/dp)$ , and some of these derivatives are non-zero.
- The linear functional  $D_x H$  is a linear function with coefficients  $\partial H/\partial x_i$ .
- The vector field  $Q$  is  $Q = (q_1, \dots, q_n)$ .
- The value of the functional,  $D_x(H)(Q)$  is

$$D_x(H)(Q) = \sum_{i=1}^n \frac{\partial H}{\partial x_i} q_i.$$

- The value of the functional  $D_x(H)(e_x)$  is

$$D_x(H)(e_x) = \sum_{i=1}^n \frac{\partial H}{\partial x_i} \frac{df_i}{dp}.$$

- The final formula (7) for the projector  $\mathcal{P}_x$  looks in coordinates as follows:

$$(\mathcal{P}_x Q)_i = \frac{df_i}{dp} \frac{\sum_{j=1}^n \frac{\partial H}{\partial x_j} q_j}{\sum_{j=1}^n \frac{\partial H}{\partial x_j} \frac{df_j}{dp}}.$$

In particular, formula (7) and Theorem 4 explain the mathematical foundations of the results of Lampis [35] and Gorban and Karlin [36] about the projector of the Boltzmann equation on the Tamm–Mott-Smith approximation in the theory of strong shock waves.

### 3.5. General theorem about projector with preservation of dissipativity

The answer to the main question is a combination of constructions presented by Theorems 3 and 4.

Let  $M \subset U$  be an ansatz submanifold,  $x \in M$ ,  $x \neq x^{\text{eq}}$ . Assume the transversality condition: differential of  $H$  at  $x$  does not annuls the tangent space  $T_x(M)$ . This means that  $x$  is not a critical point of  $H_M$ . We are looking for projectors  $\mathcal{P}_x : \mathbb{R}^n \rightarrow T_x(M)$  that project any dissipative vector field  $Q$  ( $D_x(H)(Q) \leq 0$ ) into a dissipative vector field on  $M$  ( $D_x(H_M)(\mathcal{P}_x(Q)) \leq 0$ ). We will find the necessary conditions, prove that these conditions define unique projector, and demonstrate that this projector transforms any dissipative vector fields on  $U$  into a dissipative vector fields on  $M$ .

Consider the set of vectors  $Q_D(x) = \{Q(x)\}$ , where  $Q$  is a dissipative vector field on  $U$ . It includes the open half-space  $\{Q|D_x(H)(Q) < 0\}$  and belongs to the closed half-space  $\{Q|D_x(H)(Q) \leq 0\}$ :

$$\{Q|D_x(H)(Q) < 0\} \subset Q_D(x) \subset \{Q|D_x(H)(Q) \leq 0\}$$

The second inclusion coincides with the dissipativity condition: the derivative of  $H$  along  $H$  is non-positive. For the first inclusion we have to use Proposition 1. Use any local coordinate system in  $U$  near  $x$  and define a constant vector field  $Q(y) = Q(x)$  in a small vicinity of  $x$ . If this vicinity is sufficiently small, then there  $D_y(H)(Q(y)) < 0$ . Use the extension Theorem 2 to extend this locally defined dissipative vector field to the entire  $U$ . The first inclusion is proven.

A similar statement is valid for vector fields on  $M$ . Consider the set of vectors  $Q_{D,M}(x) = Q(x)$ , where  $Q$  is a dissipative vector field on  $M$ . It includes the open half-space  $\{Q|D_x(H_M)(Q) < 0\}$  and belongs to the closed half-space  $\{Q|D_x(H_M)(Q) \leq 0\}$ ,

$$\{Q|D_x(H_M)(Q) < 0\} \subset Q_{D,M}(x) \subset \{Q|D_x(H_M)(Q) \leq 0\},$$

where all derivatives are calculated on  $M$ .

**Lemma 1.** Projector  $\mathcal{P}_x$  projects the closed half-space  $\{Q|D_x(H)(Q) \leq 0\}$  in  $\mathbb{R}^n$  onto the closed half-space  $\{Q|D_x(H_M)(Q) \leq 0\}$  in  $T_x(M)$ :

$$\mathcal{P}_x(\{Q|D_x(H)(Q) \leq 0\}) = \{Q|D_x(H_M)(Q) \leq 0\} \tag{8}$$

**Proof.** According to Proposition 1, each vector field on  $M$  that satisfies the inequality  $D_x(H_M)(Q) < 0$  can be extended from a vicinity of  $x$  on  $M$  into a dissipative vector field in the entire  $U$ . The projection of this set of extensions is again  $Q_{D,M}(x)$ . Therefore, projection of the half-space  $\{Q|D_x(H)(Q) \leq 0\}$  by  $\mathcal{P}_x$  should coincide with the half-space  $\{Q|D_x(H_M)(Q) \leq 0\}$ . Indeed, the inclusion

$$\mathcal{P}_x(\{Q|D_x(H)(Q) \leq 0\}) \subseteq \{Q|D_x(H_M)(Q) \leq 0\}$$

holds, because for each dissipative vector field  $W(x)$ , its projection on  $T_x(M)$  should be also dissipative. The inclusion

$$\{Q|D_x(H_M)(Q) < 0\} \subseteq \mathcal{P}_x(\{Q|D_x(H)(Q) < 0\})$$

holds because each vector  $V$  of the open half-space  $\{Q|D_x(H_M)(Q) < 0\}$  can be extended to the dissipative vector field on  $U$ . The value of this field at point  $x$  is again  $V \in T_x(M)$  and its projection  $\mathcal{P}_x V = V$ . The inclusions for the closed half-spaces follow from the proven inclusions for the open half-spaces and continuity of projectors (finite-dimensional linear operators).  $\square$

This lemma also gives the following equality

**Lemma 2.**

$$\mathcal{P}_x(\{Q|D_x(H)(Q) = 0\}) = \{V|D_x(H_M)(V) = 0\} \tag{9}$$

**Proof.** Indeed, if  $D_x(H)(Q) = 0$  then  $D_x(H)(-Q) = 0$  and projections of both vectors  $Q$  and  $-Q$  belong to the closed half-space  $\{Q|D_x(H)(Q) \leq 0\}$ . Assume that  $D_x(H_M)(\mathcal{P}_x(Q)) < 0$ , then  $D_x(H_M)(\mathcal{P}_x(-Q)) > 0$  and  $\mathcal{P}_x(-Q) \notin \{Q|D_x(H_M)(Q) < 0\}$ . This contradiction with (8) proves the inclusion.

$$\mathcal{P}_x(\{Q|D_x(H)(Q) = 0\}) \subseteq \{V|D_x(H_M)(V) = 0\}.$$

Another inclusion



$$\{V|D_x(H_M)(V) = 0\} \subseteq \mathcal{P}_x(\{Q|D_x(H)(Q) = 0\})$$

is satisfied because  $\{V|D_x(H_M)(V) = 0\} \subset \{Q|D_x(H)(Q) = 0\}$ .  $\square$

On the subspace  $\ker D_x(H) = \{Q|D_x(H)(Q) = 0\}$  the projector  $\mathcal{P}_x$  acts as an orthogonal projector in the Shahshahani inner product.

**Lemma 3.** *Restriction  $\mathcal{P}_x$  onto  $\ker D_x(H)$  is an orthogonal projector in the Shahshahani inner product  $\langle \cdot | \cdot \rangle_x$ :*

$$\mathcal{P}^\perp : \ker D_x(H) \rightarrow \ker D_x(H_M)$$

**Proof.** Consider restriction of  $H$  onto the intersection of hyperplane  $x + \ker D_x(H)$  with  $U$ . On  $U \cap (x + \ker D_x(H))$  point  $x$  is a global minimizer of  $H$ . Similarly, point  $x$  is a global minimizer of  $H$  on  $U \cap (x + \ker D_x(H_M))$ . Notice that  $(U \cap (x + \ker D_x(H_M))) \subset (U \cap (x + \ker D_x(H)))$ . For projections of dissipative vector fields in  $U \cap (x + \ker D_x(H))$  into dissipative vector fields in  $U \cap (x + \ker D_x(H_M))$  near point  $x$  we can use [Theorem 3](#). According to this theorem, the restriction of  $\mathcal{P}_x$  on  $\ker D_x(H)$  is an orthogonal projector onto  $D_x(H_M)$  in the Shahshahani metrics.  $\square$

Finally, we approach a linear algebra problem: Let  $L$  be an Euclidean space with inner product  $\langle \cdot | \cdot \rangle$ . Assume that  $W$  is a subspace in  $L$ ,  $l \neq 0$  is a linear functional in  $L$ ,  $W \not\subset \ker l$ ,  $\mathcal{P} : L \rightarrow W$  is a projector,  $\text{im } \mathcal{P} = W$ ,  $\mathcal{P}(\{y \in L | l(y) \leq 0\}) = \{z \in W | l(z) \leq 0\}$ , and the restriction of  $\mathcal{P}$  on  $\ker l$ ,  $\mathcal{P}|_{\ker l} : \ker l \rightarrow (\ker l \cap W)$  is an orthogonal projector in the inner product  $\langle \cdot | \cdot \rangle$ . Demonstrate that these conditions define  $\mathcal{P}$  unambiguously and find an explicit construction of this projector.

Let  $L_0 = \ker l$  and  $W_0 = W \cap \ker l$ . Find a unit normal vector  $v$  to  $W_0$ . Space  $L$  can be presented as an orthogonal sum of three subspaces:

$$L = W_0 \oplus \{y \in L_0 | y \perp W_0\} \oplus \{\alpha v | \alpha \in R\}.$$

Subspace  $W$  can be presented as an orthogonal sum of two subspaces:

$$W = W_0 \oplus \{\alpha v_W | \alpha \in R\},$$

where  $v_W \in W$  is the unit normal vector to  $W_0$  in  $W$ . Note, that in a general situation  $v_W \neq v$ . For convenience, we choose such orientation of normal vectors that  $l(v) < 0$  and  $l(v_W) < 0$ .

If a vector  $z \in L$  is a orthogonal sum

$$z = z_0 \oplus z_0^\perp \oplus \zeta v \tag{10}$$

( $z_0 \in W_0$ ,  $z_0^\perp \in \{y \in L_0 | y \perp W_0\}$  and  $\zeta \in R$ ), then its projection onto  $W$  should have a form:

$$\mathcal{P}z = z_0 \oplus \beta \zeta v_W,$$

where  $\beta$  is a real number. This coefficient is defined by the condition  $\mathcal{P}(v_W) = v_W$  because projector  $\mathcal{P}$  is the identity operator on its image.

Vector  $v_W$  can be represented in  $L$  as an orthogonal sum:  $v_W = v_W^\parallel \oplus v_W^\perp$ , where  $v_W^\perp = \langle v_W | v \rangle v$  and  $v_W^\parallel = v_W - v_W^\perp$ .  $v_W^\parallel \in W_0$  because  $\langle v_W^\parallel | v \rangle = 0$ . Therefore,  $\mathcal{P}v_W = \beta \langle v_W | v \rangle v_W$ .

The condition  $\mathcal{P}v_W = v_W$  gives:  $\beta \langle v_W | v \rangle = 1$  and

$$\beta = \frac{1}{\langle v_W | v \rangle}.$$

Finally, the projector  $\mathcal{P}$  is unambiguously defined and for the projection of the orthogonal sum (10) we obtain:

$$\mathcal{P}(z_0 \oplus z_0^\perp \oplus \zeta v) = z_0 \oplus \frac{\zeta v_W}{\langle v_W | v \rangle} \tag{11}$$

Note that this projector preserves the value of  $l(z)$ . First of all,

$$l(v_W) = l(v_W^\perp) = \langle v_W | v \rangle l(v)$$

because  $v_W^\parallel \in \ker l$ . Note that  $l(z_0 \oplus z_0^\perp \oplus \zeta v) = \zeta l(v)$ . Therefore

$$l(\mathcal{P}(z_0 \oplus z_0^\perp \oplus \zeta v)) = l\left(z_0 \oplus \frac{\zeta v_W}{\langle v_W | v \rangle}\right) = \zeta l(v) = l(z_0 \oplus z_0^\perp \oplus \zeta v).$$

Thus, we required the preservation of the sign of  $l(z)$  and obtained the preservation of its value.

[Lemmas 1, 2 and 3](#) together with formula (11) prove the following theorem. Let  $x \in U$  be a non-critical point of  $H$ . Assume that  $M \subset U$  is a smooth immersed ansatz manifold,  $x \in M$ , and the transversality condition holds:  $T_x(M) \not\subset \ker D_x H$  (that is,  $M$  is not tangent to the level set of  $H$  at point  $x$ ). Let us use the notations  $L_0 = \ker D_x H$  and  $W_0 = T_x(M) \cap L_0$ . We use the Shahshahani inner product  $\langle \cdot | \cdot \rangle_x$ . Find unit normal vectors  $v$  for  $L_0$  and  $v_W \in T_x(M)$  for  $W_0$ . For most applications, it is convenient to select the ‘‘antigradient’’ orientations of these normal vectors:  $D_x(H)(v) < 0$  and  $D_x(H_M)(v_W) < 0$ . Represent  $L_0$  in the form of the orthogonal sum:  $L_0 = W_0 \oplus \{y \in L_0 | y \perp W_0\}$ . Each vector  $Q \in \mathbb{R}^n$  can be represented as an orthogonal sum:  $Q = Q_0 \oplus Q_0^\perp \oplus \zeta v$ .

**Theorem 5.** *Projector field  $P_x$ , that solves the problem of dimensionality reduction with preservation of dissipativity acts as follows:*

$$P_x(Q) = Q_0 \oplus \frac{\zeta v_W}{\langle v_W | v \rangle} \quad \square \tag{12}$$

The applications of Theorem 5 are the same for all problems of dimension reduction:

- Find Hessian of  $H$  and define the Shahshahani inner product  $\langle \cdot | \cdot \rangle_x$ .
- Define the unit normal vector  $v = v(x)$  to  $L_0 = \ker D_x(H(x))$ . For each  $x \in U$  this vector is the normalized antigradient  $H$  in the inner product  $\langle \cdot | \cdot \rangle_x$ :  $v(x) = -\frac{\text{grad}_x H(x)}{\|\text{grad}_x H(x)\|}$ .
- For a given ansatz manifold  $M$  and any  $x \in M$  find the unit normal vector  $v_W = v_W(x) \in T_x(M)$  to the subspace  $W_0 = T_x(M) \cap \ker D_x(H_M)$ . For each  $x \in M$  this vector is the normalized antigradient  $H_M$  in the restriction of the inner product  $\langle \cdot | \cdot \rangle_x$  onto  $T_x(M)$ :  $v_W(x) = -\frac{\text{grad}_x H_M(x)}{\|\text{grad}_x H_M(x)\|}$ .
- Find the orthogonal decomposition  $L_0 = W_0 \oplus \{y \in L_0 | y \perp W_0\}$ .
- Use the projector (12) for any dissipative vector field.

The only computationally expensive task in high dimensionality may be the orthogonal decomposition for a given ansatz manifold. Calculation of gradients requires several comments. Gradient of a function in a Euclidean space is the Riesz representation of its differential and depends on the inner product. By definition, for the Shahshahani inner product  $\langle \cdot | \cdot \rangle_x$ ,  $D_x H(x)(Q) = \langle \text{grad}_x H(x), Q \rangle_x$ . Similarly,  $D_x H_M(x)(Q) = \langle \text{grad}_x H_M(x), Q \rangle_x$ , but for  $\text{grad}_x H_M(x)$  all calculations are performed in the space  $T_x(M)$  and the vector of gradient belongs to this space. In general,  $\text{grad}_x H_M(x) \neq \text{grad}_x H(x)$ . The standard coordinate representation of the gradient as the vector of partial derivatives is valid only in the orthonormal basis in the Shahshahani inner product. It may be worth mentioning that the antigradient in the Shahshahani inner product is the direction of descent in Newton’s method for minimizing a function. Indeed, let  $e_x$  be a gradient  $H(x)$  at point  $x \in U$  with respect to the inner product  $\langle \cdot | \cdot \rangle_x$ . Let us use the initial (standard) coordinates in  $\mathbb{R}^n$ . Recall that  $\langle y | z \rangle_x = (y, \text{Hes}(H)z)$ , where  $\text{Hes}(H)$  is the Hessian matrix of second-order partial derivatives of  $H$  and  $(y, z) = \sum y_i z_i$  is the standard inner product in  $\mathbb{R}^n$ . The gradient vector  $e_x$  is defined by equations:

$$\sum_{i=1}^n \frac{\partial H}{\partial x_i} Q_i = \sum_{i=1}^n Q_i \left( \sum_{j=1}^n (\text{Hes}_x(H))_{ij} (e_x)_j \right) \text{ for all } Q.$$

Vectors  $Q$  can be excluded from these equations and we obtain

$$\nabla_x H = \text{Hes}_x(H)e_x; \quad e_x = (\text{Hes}_x(H))^{-1} \nabla_x H,$$

where  $\nabla_x H$  is the vector of partial derivatives of  $H$  in the initial coordinate system. Finally,  $v$  in the definition of projector (12) is:

$$v = -\frac{e_x}{\sqrt{\langle e_x | e_x \rangle_x}}.$$

The similar calculations in an internal coordinate system on an ansatz manifold  $M$  give us the unit normal vector to  $\ker D_x H_M(x)$ ,  $v_W \in T_x(M)$ .

For convenience, in the next subsection we collect the most popular functions  $H(x)$  and their Hessian matrices.

### 3.6. Shahshahani inner products and gradients for several popular divergences

Many of the applied Lyapunov function are particular cases of the following family:

$$H(x) = \sum_i x_i^{\text{eq}} f \left( \frac{x_i}{x_i^{\text{eq}}} \right), \tag{13}$$

where  $f$  is a convex function on the positive semi-axis.

These functions were introduced by Rényi [44] in the same work, where he introduced the Rényi entropy. He proved that they are Lyapunov functions for Markov kinetics. It is noteworthy that the function  $H(x)$  (13) depends only on the current value of  $x$  and on the equilibrium value  $x^{\text{eq}}$ , and does not depend on the kinetic coefficients, which are usually not as well known as the equilibrium. There exist no other Lyapunov functions for the Markov kinetics with this property (independence of kinetic constants). This characterization of  $H(x)$  was proven by P. Gorban in 2003 [45] and then independently by S.-I. Amari in 2009 [46]. These functions were also rediscovered by Morimoto [47] and studied further by Csiszar [48]. They are known as  $f$ -divergences or the Csiszar-Morimoto relative entropies.

The derivatives of the  $f$ -divergences (13) are

$$\frac{\partial H}{\partial x_i} = f' \left( \frac{x_i}{x_i^{\text{eq}}} \right),$$

where  $f'$  is the derivative of  $f$ .

The Hessian matrix  $\text{Hes}_x(H)$  is

$$\frac{\partial^2 H}{\partial x_i \partial x_j} = \delta_{ij} \frac{1}{x_i^{\text{eq}}} f'' \left( \frac{x_i}{x_i^{\text{eq}}} \right), \tag{14}$$

where  $f''$  is the second derivative of  $f$ . This matrix is diagonal, therefore, the corresponding Shahshahani inner product is very simple:

$$\langle z|y \rangle_x = \sum_i f'' \left( \frac{x_i}{x_i^{\text{eq}}} \right) \frac{z_i y_i}{x_i^{\text{eq}}}.$$

The gradient of  $H(x)$  with respect to this inner product is a vector  $e_x$  with components

$$e_{xi} = x_i^{\text{eq}} \frac{f' \left( \frac{x_i}{x_i^{\text{eq}}} \right)}{f'' \left( \frac{x_i}{x_i^{\text{eq}}} \right)}.$$

If  $f(z) = z \ln z$  then the  $f$ -divergence (13) is the Kullback–Leibler divergence very popular both in kinetics and machine learning:

$$H(x) = \sum_i x_i \ln \left( \frac{x_i}{x_i^{\text{eq}}} \right).$$

For this function,

$$\frac{\partial H}{\partial x_i} = 1 + \ln \left( \frac{x_i}{x_i^{\text{eq}}} \right), \quad \frac{\partial^2 H}{\partial x_i \partial x_j} = \frac{\delta_{ij}}{x_i}, \quad \langle z|y \rangle_x = \sum_i \frac{z_i y_i}{x_i}, \quad e_{xi} = x_i \left( 1 + \ln \left( \frac{x_i}{x_i^{\text{eq}}} \right) \right).$$

If  $f(z) = z(\ln z - 1)$  then the formulas become even simpler:

$$H(x) = \sum_i x_i \left( \ln \left( \frac{x_i}{x_i^{\text{eq}}} \right) - 1 \right).$$

$$\frac{\partial H}{\partial x_i} = \ln \left( \frac{x_i}{x_i^{\text{eq}}} \right), \quad \frac{\partial^2 H}{\partial x_i \partial x_j} = \frac{\delta_{ij}}{x_i}, \quad \langle z|y \rangle_x = \sum_i \frac{z_i y_i}{x_i}, \quad e_{xi} = x_i \ln \left( \frac{x_i}{x_i^{\text{eq}}} \right).$$

If  $f(z) = -\ln z$  then  $H(x)$  is the relative Burg entropy:

$$H(x) = - \sum_i x_i^{\text{eq}} \ln \left( \frac{x_i}{x_i^{\text{eq}}} \right).$$

$$\frac{\partial H}{\partial x_i} = - \frac{x_i^{\text{eq}}}{x_i}, \quad \frac{\partial^2 H}{\partial x_i \partial x_j} = \delta_{ij} \frac{x_i^{\text{eq}}}{x_i^2}, \quad \langle z|y \rangle_x = \sum_i \frac{z_i y_i x_i^{\text{eq}}}{x_i^2}, \quad e_{xi} = -x_i.$$

For all other examples of  $f$ -divergences calculations are essentially the same. Calculations on an arbitrary subspace requires more efforts with orthogonalization, at least, partial.

#### 4. Beyond manifolds: monotone trees

Approximating data sets or dynamical systems by projection onto a manifold often seems suboptimal, since a relatively high-dimensional approximation may contain too much “air” (empty or nearly empty space irrelevant to the applied problem), while a manifold of lower dimension and limited curvature may lose the quality of the approximation.

The idea of data approximation by “principal graphs and trees” is well-elaborated and there exists an open source software for use in many areas [49], from single cell omics in bioinformatics to dynamic phenotyping of diseases in clinical medicine and even in astronomy for analysis of distributions of galaxies. Nevertheless, projection of dynamics onto principal graphs and trees remains an open problem [8,50].

In this section, we present a solution of the problem of projection dynamics onto trees with preservation of dissipativity. Trees in the graph-theoretical sense may be also viewed as topological spaces by representation of each edge as an arc. A tree is smoothly immersed in  $U$  if each arc is a smooth curve. In a tree, each two nodes are connected by a unique path that consists of arcs (edges). A tree is monotone with respect to function  $H$  if  $H$  is a strongly monotone functions along each of these paths and its derivative along edges does not vanish.

The problem of preservation of dissipativity for projection of any dissipative vector field on a monotone tree is de-facto solved in Section 3.4. We should just apply Theorem 4 to the paths on the tree. Technically, the result is very simple. Let  $T \subset U$  be a monotone tree. Find the node  $x^* \in T$  with minimal value of  $H$ . This node is unique because each two nodes are connected by the unique path and  $H$  is a strongly monotone function along these paths. Let  $x$  be a point on the tree and the line  $L$  be a path in  $T$  between  $x$  and  $x^*$ . Select  $H$  as an internal coordinate on the  $L$  ( $H_L$ ). For every dissipative vector field  $Q$  define motion on this path by the condition

$$\frac{dH_L(x)}{dt} = D_x(H(x))(Q).$$

According to Theorem 4, this definition gives a unique projector of any dissipative vector fields into dissipative motion along the path  $L$ . Note that the path  $L$  is piecewise smooth, therefore we have to apply Theorem 4 to arcs and then glue motions in nodes.

For non-monotonic trees, the problem of flow distribution at bifurcation points arises. This problem is stated as important for applied work [8,50]. Here we cannot even hypothesize about a deterministic solution. Perhaps, to reduce dissipative systems to non-monotonic graphs, it is necessary to introduce stochastic models with randomized flow distributions at branching points.

## 5. Conclusion and outlook

We have presented a detailed solution to the projection problem for the reduction of a dissipative system with incomplete information about vector fields. For this purpose, we have found projectors that project any dissipative vector field into a dissipative vector field on ansatz manifolds. This work is a further mathematical development of the previous work on physical and chemical kinetics. The general form of projectors is universal and can be used in any field to reduce dimensionality when there is no detailed knowledge of the vector field. Further refinement of the model can be performed after reducing the dimension. There are many possible applications, from control and nonequilibrium thermodynamics to machine learning.

The answer to the problem is presented in [Theorem 5](#). It is simple and transparent. Calculations of the gradient of  $H$  with respect to the Shahshahani inner product  $\langle \cdot | \cdot \rangle_x$  in  $\mathbb{R}^n$  are straightforward for many popular Lyapunov functions that are useful both in machine learning and in nonequilibrium thermodynamics (Section 3.6). For an ansatz manifold of small dimension or codimension, calculations are also simple (see, for example, Section 3.4). Nevertheless, calculations of projections for an ansatz manifold of intermediate dimension (neither small dimension, nor small codimension) require additional effort for the development of optimized algorithms.

We have to stress that for some applied problems simpler projectors are used with success, for example, just an orthogonal projector in the Shahshahani inner product. In particular, if the ansatz manifold is invariant then any projector on its tangent space will preserve dissipativity. If the ansatz manifold is “almost invariant” with small normal component of the vector field? then the choice of projector may be also non-unique.

An important observation was made by Roberts [51]: He studied deformation of almost invariant systems and proved that approximate invariant manifolds of a system are exact invariant manifolds of a specially constructed approximate system. This theorem extends the area of application of exactly invariant manifolds. Nevertheless, beyond invariant and almost invariant manifolds, the choice of projector is unique and other projectors may disrupt the dissipation.

Reduces dynamics beyond manifolds is very attractive in many applications. The algorithms for data approximation by graphs and, in particular, by trees are well developed and the open source software is available (for its description and further references we refer to [49]). Despite of these successes, the problem of projection of dynamics onto ansatzes that are not manifolds has been repeatedly described as important in applications and unsolved [8,50]. Here we have taken the first step towards solving this problem: we have introduced monotone trees and found a projection operator that projects any dissipative dynamics in space into the dissipative dynamics on the tree. The problem of projection onto non-monotonic trees and graphs requires a change in the framework of consideration: it seems necessary to consider not deterministic dynamics, but dynamics with stochastic redistribution of the phase flow at bifurcation points. Stochastic modelling is beyond the scope of this article, but is certainly the first task on the “to do” list.

Finally, let us repeat the main message of the paper: the projection of dissipative dynamics with preservation of dissipativity requires a special design of the projector. Surprisingly, a universal projector exists, and it is unique.

## CRedit authorship contribution statement

**Sergey V. Stasenko:** Writing – review & editing, Investigation, Formal analysis. **Alexander N. Kiridin:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by the Analytical Center for the Government of the Russian Federation (agreement identifier 000000D730324P540002, grant No 70-2023-001320 dated 27.12.2023).

## Data availability

No data was used for the research described in the article.

## References

- [1] Gorban AN, Yablonsky GS. Grasping complexity. *Comput Math Appl* 2013;10(65):1421–6. <http://dx.doi.org/10.1016/j.camwa.2013.04.023>.
- [2] Pearson K. On lines and planes of closest fit to systems of points in space. *Lond Edinb Dublin Philos Mag J Sci* 1901;2(11):559–72. <http://dx.doi.org/10.1080/14786440109462720>.
- [3] Greenacre M, Groenen PJ, Hastie T, d'Enza AI, Markos A, Tuzhilina E. Principal component analysis. *Nat Rev Methods Primers* 2022;2(1):100. <http://dx.doi.org/10.1038/s43586-022-00184-w>.
- [4] Lee JA, Verleysen M. *Nonlinear dimensionality reduction*. New York: Springer; 2007.
- [5] Gorban AN, Kégl B, Wunsch DC, Zinovyev AY. Principal manifolds for data visualization and dimension reduction. In: *Lecture notes in computational science and engineering (LNCSE)*, volume 58, Berlin: Springer; 2008.
- [6] Gorban AN, Zinovyev A. Principal manifolds and graphs in practice: from molecular biology to dynamical systems. *Int J Neural Syst* 2010;20(03):219–32. <http://dx.doi.org/10.1142/S0129065710002383>.
- [7] Zhang H, Pedrycz W, Miao D, Zhong C. A global structure-based algorithm for detecting the principal graph from complex data. *Pattern Recognit* 2013;46(6):1638–47. <http://dx.doi.org/10.1016/j.patcog.2012.11.015>.
- [8] Golovenkin SE, Bac J, Chervov A, Mirkes A, Orlova YV, Barillot E, et al. Trajectories, bifurcations, and pseudo-time in large clinical datasets: applications to myocardial infarction and diabetes data. *GigaScience* 2020;9(11):giaa128. <http://dx.doi.org/10.1093/gigascience/giaa128>.
- [9] Chen H, Albergante L, Hsu JY, Lareau CA, Lo Bosco G, Guan J, et al. Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat Commun* 2019;10(1):1903. <http://dx.doi.org/10.1038/s41467-019-09670-4>.
- [10] Bac J, Zinovyev A. Lizard brain: Tackling locally low-dimensional yet globally complex organization of multi-dimensional datasets. *Front Neurobot* 2020;13:110. <http://dx.doi.org/10.3389/fnbot.2019.00110>.
- [11] Gorban AN, Kazantzis NK, Kevrekidis IG, Öttinger HC, Theodoropoulos C, editors. *Model reduction and coarse-graining approaches for multiscale phenomena*. Springer Berlin – Heidelberg; 2006.
- [12] Angeli D. A tutorial on chemical network dynamics. *Eur J Control* 2009;15:398–406. <http://dx.doi.org/10.23919/ECC.2009.7074477>.
- [13] Gorban AN, Yablonsky GS. Three waves of chemical dynamics. *Math Model Nat Phenomena* 2015;10(5):1–5. <http://dx.doi.org/10.1051/mmnp/201510501>.
- [14] Gorban AN. Model reduction in chemical dynamics: slow invariant manifolds, singular perturbations, thermodynamic estimates, and analysis of reaction graph. *Curr Opin Chem Eng* 2018;21:48–59. <http://dx.doi.org/10.1016/j.coche.2018.02.009>.
- [15] Marin GB, Yablonsky GS, Constaes D. *Kinetics of chemical reactions*. John Wiley & Sons: Decoding Complexity; 2019.
- [16] Kramer MA. Nonlinear principal component analysis using autoassociative neural networks. *AIChE J* 1991;37(2):233–43. <http://dx.doi.org/10.1002/aic.690370209>.
- [17] Kramer MA. Autoassociative neural networks. *Comput Chem Eng. Neural Netw Appl Chem Eng* 1992;16(4):313–28. [http://dx.doi.org/10.1016/0098-1354\(92\)80051-A](http://dx.doi.org/10.1016/0098-1354(92)80051-A).
- [18] Gorban AN, Karlin IV. *Invariant manifolds for physical and chemical kinetics*. Berlin, Heidelberg: Springer; 2005.
- [19] Gorban A, Karlin I. Hilbert's 6th problem: exact and approximate hydrodynamic manifolds for kinetic equations. *Bull Amer Math Soc* 2014;51(2):187–246. <http://dx.doi.org/10.1090/S0273-0979-2013-01439-3>.
- [20] Jaynes ET. Information theory and statistical mechanics, I; II. *Phys Rev* 1957;106:620–63, 108:171–190.
- [21] Grmela M. Role of thermodynamics in multiscale physics. *Comput Math Appl* 2013;65(10):1457–70. <http://dx.doi.org/10.1016/j.camwa.2012.11.019>.
- [22] Pavelka M, Klika V, Grmela M. *Multiscale thermo-dynamics: introduction to GENERIC*. Berlin, Boston: Walter de Gruyter; 2018.
- [23] Principe JC. *Information theoretic learning: renyi's entropy and kernel perspectives*. New York: Springer; 2010.
- [24] Filippi S, Cappé O, Garivier A. Optimism in reinforcement learning and Kullback–Leibler divergence. In: 2010 48th annual allerton conference on communication, control, and computing (allerton) (115–122). IEEE; 2010. <http://dx.doi.org/10.1109/ALLERTON.2010.5706896>.
- [25] Nowozin S, Cseke B, Tomioka R. F-gan: Training generative neural samplers using variational divergence minimization. In: *Proceedings of 30th conference on neural information processing systems (NIPS 2016)*, Barcelona, Spain. *Advances in Neural Information Processing Systems*; 2016, p. 271–9.
- [26] Semenov A, Boginski V, Pasiliao EL. Neural networks with multidimensional cross-entropy loss functions. In: Tagarelli A, Tong H, editors. *Computational data and social networks. CSonNet 2019*. Lecture notes in computer science, Vol. 11917, Springer; 2019. [http://dx.doi.org/10.1007/978-3-030-34980-6\\_5](http://dx.doi.org/10.1007/978-3-030-34980-6_5).
- [27] Ji S, Zhang Z, Ying S, Wang L, Zhao X, Gao Y. Kullback–Leibler divergence metric learning. *IEEE Trans Cybern* 2020;52(4):2047–58. <http://dx.doi.org/10.1109/TCYB.2020.3008248>.
- [28] Gorban AN. Selection theorem for systems with inheritance. *Math Model Nat Phenomena* 2007;2(4):1–45. <http://dx.doi.org/10.1051/mmnp:2008024>.
- [29] Kareva I, Karev G. *Modeling evolution of heterogeneous populations: theory and applications*. San Diego: Academic Press; 2019.
- [30] Morozov A. Modelling biological evolution: Developing novel approaches. *Bull Math Biol* 2019;81:4620–4. <http://dx.doi.org/10.1007/s11538-019-00670-5>.
- [31] Li H, Xu Z, Taylor G, Studer C, Goldstein T. Visualizing the loss landscape of neural nets. *Adv Neural Inf Process Syst* 2018;31.
- [32] Tyukin I. *Adaptation in dynamical systems*. Cambridge University Press; 2011.
- [33] Cercignani C. *The Boltzmann equation and its applications*. New York: Springer; 1988.
- [34] Hosokawa I, Inage S. Local entropy balance through the shock wave. *J Phys Soc Japan* 1986;55(10):3402–9. <http://dx.doi.org/10.1143/JPSJ.55.3402>.
- [35] Lampis M. New approach to the Mott-Smith method for shock waves. *Mechanica* 1977;12:171–3. <http://dx.doi.org/10.1007/BF02179931>.
- [36] Gorban AN, Karlin IV. Thermodynamic parameterization. *Phys A* 1992;190:393–404. [http://dx.doi.org/10.1016/0378-4371\(92\)90044-Q](http://dx.doi.org/10.1016/0378-4371(92)90044-Q).
- [37] Gorban AN, Karlin IV. Method of invariant manifolds and regularization of acoustic spectra. *Transp Theory Stat Phys* 1994;23(5):559–632. <http://dx.doi.org/10.1080/00411459408204345>.
- [38] Gorban AN, Karlin IV. Uniqueness of thermodynamic projector and kinetic basis of molecular individualism. *Phys A* 2004;336(3–4):391–432. <http://dx.doi.org/10.1016/j.physa.2004.01.039>.
- [39] Shahshahani S. Dissipative systems on manifolds. *Inventiones Mathematicae* 1972;16:177–90. <http://dx.doi.org/10.1007/BF01391216>.
- [40] Whitney H. On the extension of differentiable functions. *Bull Amer Math Soc* 1944;50(12):76–81.
- [41] Malgrange B. Ideals of differentiable functions. In: *Tata institute of fundamental research studies in mathematics. Vol. 3*, Oxford: Oxford University Press; 1967.
- [42] Arnold VI. *Ordinary differential equations*. Berlin, Heidelberg: Springer; 1992.
- [43] Arnold VI. *Geometrical methods in the theory of ordinary differential equations*. Berlin, Heidelberg: Springer; 2012.
- [44] Rényi A. On measures of entropy and information. In: *Proceedings of the 4th Berkeley symposium on mathematics, statistics and probability 1960. Vol. 1*, Berkeley, CA, USA: University of California Press; 1961, p. 547–61.
- [45] Gorban PA. Monotonically equivalent entropies and solution of additivity equation. *Phys A* 2003;328(3–4):380–90. [http://dx.doi.org/10.1016/S0378-4371\(03\)00578-8](http://dx.doi.org/10.1016/S0378-4371(03)00578-8).
- [46] Amari S-I. Divergence, optimization, geometry. In: Leung CS, Lee M, Chan JH, editors. *The 16th international conference on neural information processing (ICONIP 2009)*, Bangkok, Thailand, 1–5 2009. *Lecture notes in computer science, vol 5863*, Heidelberg: Springer Berlin; 2009, p. 185–93. [http://dx.doi.org/10.1007/978-3-642-10677-4\\_21](http://dx.doi.org/10.1007/978-3-642-10677-4_21).
- [47] Morimoto T. Markov processes and the  $H$ -theorem. *J Phys Soc Jap* 1963;12:328–31.

- [48] Csiszár I. Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von markoffschen ketten. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* 1963;8:85–108.
- [49] Albergante L, Mirkes E, Bac J, Chen H, Martin A, Faure L, et al. Robust and scalable learning of complex intrinsic dataset geometry via EIPiGraph. *Entropy* 2020;22(3):296. <http://dx.doi.org/10.3390/e22030296>.
- [50] Bell AJ, Ram S, Labaki WW, Murray S, Kazerooni EA, Galban S, et al. Temporal exploration of COPD phenotypes: Insights from the COPDGene and SPIROMICS cohorts. *Am J Respir Crit Care Med* 2024. <http://dx.doi.org/10.1164/rccm.202401-0127OC>, (ja).
- [51] Roberts AJ. Backwards theory supports modelling via invariant manifolds for non-autonomous dynamical systems. 2018, <http://dx.doi.org/10.48550/arXiv.1804.06998>, arXiv preprint [arXiv:1804.06998](https://arxiv.org/abs/1804.06998).