

Article

Active Inference and Functional Parametrisation: Differential Flatness and Smooth Random Realisation

Hugues Mounier ^{1,*} , Thomas Parr ²  and Karl Friston ^{3,4,*} 

¹ Laboratoire des Signaux et Systèmes, Université Paris-Saclay, Centre National de la Recherche Scientifique, CentraleSupélec, 3, rue Joliot Curie, 91192 Gif sur Yvette, France

² Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford OX3 9DU, UK; thomas.parr@ndcn.ox.ac.uk

³ Wellcome Centre for Human Neuroimaging, Queen Square Institute of Neurology, University College London, 12 Queen Square, London WC1N 3AR, UK

⁴ VERSES Research Lab, 5877 Obama Blvd, Los Angeles, CA 90016, USA

* Correspondence: hugues.mounier@universite-paris-saclay.fr (H.M.); k.friston@ucl.ac.uk (K.F.)

Abstract

This paper is a first attempt to marry constructive nonlinear control theory techniques with active inference. Specifically, we are interested in the relationship between differential flatness and the design of generative models for use in control settings. We place specific emphasis on the pathwise properties of differentially flat systems that inherit from their definition in terms of successive temporal derivatives and relate this to the use of generalised coordinates of motion in formulating continuous-time generative models in active inference. To illustrate the basic concepts, we appeal to the example of oculomotor control.

Keywords: differential flatness; active inference; periodic smooth random functions; pathwise formulations

1. Introduction

Active inference is one, if not the most, promising formal framework for computational neuroscience, with many applications in a number of areas (see, e.g., [1] and the references therein). The recent developments foregrounding pathwise formulations and Bayesian mechanics—developed in [2,3] among others—furnish a principled and natural setting to address many aspects of perception, planning, and control. The central idea that we will appeal to in this paper is that (natural or artificial) creatures make use of implicit generative world models both to draw inferences about the world and to control their sensed world. It is the latter that is more salient for our current purposes. In fact, one could regard the inferential (i.e., perceptual) role of generative models as merely a method of attaining tight bounds on marginal likelihoods—the quantity of interest when optimising sensory data through action.

We here consider some conditions whose fulfilment may help to construct generative models well suited for controlling one's environment. Crucially, conditions are in direct agreement with constructive frameworks developed in nonlinear control (see, e.g., [4–6]). We thus try to show how this framework, and in particular the differential flatness structural property, can be leveraged in applications of active inference to control problems.

Many insightful works have already been published in this general area, both on the links and differences between active inference and classical control schemes and on deploying active inference for control (see, e.g., [7–14] just to name a few).



Academic Editor: José F.F. Mendes

Received: 30 October 2025

Revised: 14 December 2025

Accepted: 6 January 2026

Published: 11 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

At first sight, differential flatness and active inference seem quite distant frameworks. The first aims to reduce a trajectory tracking error to zero, while the second minimises surprise or variational free energy; the first is inherently deterministic, and the second naturally deals with stochastic fluctuations. We shall see that the trajectory tracking error is indeed a form of surprise and that the very definition of differential flatness can be adapted to smooth random fluctuations, particularly apt for neuroscientific applications. There are two main approaches to minimise surprise or discrepancy:

- The first is to envision the problem as an optimisation procedure, as in active inference (but also in optimal control or model predictive control); in this respect, the goal is to fulfil an optimisation criterion, leading to the minimisation of target discrepancies.
- The second is to derive a deterministic control (action) scheme for the goal and estimate the fluctuations online to actively compensate for them.

The first route is that taken by active inference, while the second underwrites the differential flatness approach. We shall nevertheless see that the two frameworks are not as different as one may think. Indeed, the mode dynamics enforced by free energy minimisation—when those dynamics enjoy the differential flatness property—yield a functional parametrisation linking the two schemes intimately. Moreover, this parametrisation appears to be a most fruitful and promising tool to study nonlinear dynamics in neuroscience and biology.

More broadly, here, we try to foreground some potentially useful features of our framework, which draws from dynamical control system structural properties in a typical *pathwise* and *physically preserving* formulation. In particular, we shall see how the most striking feature of differential flatness—namely, differential parametrisation—can be of use within the active inference framework. More specifically, this differential parametrisation induces, among other relations, an invertible mapping from the sensation to action.

The problem we will be interested in in Section 5 is one in which we have a generative model that describes how (we believe) some system will evolve over time as a function of some control variables. Given a desired state or path for this system, our interest is in understanding how the control variables would have to be set in order to realise that desired configuration. More specifically, we are interested in the properties a generative model should possess in order to infer the control variables that realise some goal.

This paper is structured as follows. We first provide a brief introduction to the idea of a generative model for control and introduce some of the key definitions we will appeal to subsequently. Several of these definitions turn out to rely upon differentiable fluctuations, which leads us to a consideration of the interpretation of stochastic fluctuations that we will commit to (i.e., the use of random periodic functions). Having outlined the basic structure of the sorts of generative models we are interested in, we consider the way in which the principles that underwrite active inference constrain choices we might make when designing these models and equipping them with goals. We use the forms of the key objective functions from active inference (variational and expected free energy) to consider some of the generic properties of good generative models and consider whether these properties can be motivated by appealing to the notion of differential flatness. Following these theoretical considerations, we demonstrate the way these ideas work in practice through a worked example based upon oculomotor control. We conclude with a discussion of the relationship between some of these ideas and active inference, with a particular focus on the notion of generalised coordinates of motion, which inherit from similar ideas to that of differential flatness but turn out to play a very different role.

2. Preliminary Notions: Generative Model, Action, State, and Fluctuation Choice

2.1. Generative Models

As outlined above, we are interested in control problems that are formulated in terms of generative models. As an example to make this clear, which will be unpacked in much greater detail later on, consider how we might go about controlling the positions of our eyes such that we achieve a specific fixation point on a surface or track a moving point. To decide how to move our eyes, our brains might employ a model in which action variables (\mathbf{u}), like the contraction of extra-ocular muscles, might influence the dynamics of states (x), such as the angle of our gaze or fixation point, and in which these states return some observable (e.g., visual) data (y). We now precisely outline the notion of a generative model.

Definition 1 (Generative model). *A generative model is a set of stochastic differential equations of the following form:*

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}) + \boldsymbol{\zeta}_x \quad (1a)$$

$$\mathbf{y} = h(\mathbf{x}, \boldsymbol{\theta}) + \boldsymbol{\zeta}_y \quad (1b)$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is a vector representing the state of the system (e.g., oculomotor apparatus) to be controlled (see Definition for a precise definition), $\mathbf{u}(t) \in \mathbb{R}^m$ is the action vector (an external control input), the signal or instruction we might send to that system, $\boldsymbol{\theta}$ is a vector of the agent's model parameters, $\mathbf{y}(t) \in \mathbb{R}^p$ is the output vector, the agent's sampling of the sensor signal, and $\boldsymbol{\zeta}_x(t) \in \mathbb{R}^n$ and $\boldsymbol{\zeta}_y(t) \in \mathbb{R}^p$ are the state and output fluctuations, considered, in the Stratonovich sense, as smooth random functions (see Section 2.3 below). The functions f and h are here supposed to be meromorphic in their arguments, and we suppose that \mathbf{x}, \mathbf{y} belong to a functional space \mathcal{F} such that $\mathcal{F} \subsetneq C^0(\mathbb{R})$, with $C^0(\mathbb{R})$ being the space of continuous functions in time.

It is often convenient, in applications of active inference, to write a generative model in terms of *generalised coordinates of motion*. These coordinates are the coefficients of a Taylor series expansion of variables around the current time. In doing so, we replace a single nonlinear stochastic differential equation (for x) $\dot{x} = f(x, u, \theta) + \zeta_x$ with a set of locally linearised equations for each order of motion: $\dot{x}^{(n)} = f^{(n)}(x^{(n)}, u, \theta) + \zeta_x^{(n)}$, where the superscript in brackets indicates the order of motion (i.e., the term in the Taylor series to which the coefficient belongs or the order of the temporal derivative one must take to get to these variables from our initial variable).

The motivation for this formulation is twofold. First, it allows one to express noise processes of varying smoothness in terms of the covariance between the fluctuations assigned to each order of motion. Second, through applications of the chain rule, it allows one to determine the gradients of generalised coordinates of y with respect to u via the gradients of the generalised coordinates of x . This will be important later, when we look to the way in which descent of variational free energy gradients—by changing u —can be understood as analogous to the enaction of spinal reflex arcs that bring proprioceptive data in line with anticipated setpoints.

The need to find mappings from sensation to action is one of the key ideas common to active inference—using generalised coordinates of motion—and differential flatness as applied to control. It is interesting that both approach this problem by appealing to successive derivatives of the variables in a generative model.

Definition 2 (Mean generative model). *Given a generative model as above, the mean, or deterministic, generative model is given by the following set of differential equations:*

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}) \tag{2a}$$

$$\mathbf{y} = h(\mathbf{x}, \boldsymbol{\theta}) \tag{2b}$$

It thus corresponds to the deterministic part of the corresponding stochastic differential equation system.

2.2. Action, Output, and State

We here give precise definitions for the notions of action, or control input, output, state, and realisation, which will be useful in Section 6.2. These were given in a differential algebraic setting in [4]; see also [15] for an elementary treatment, analogous to the one below (although stated in a deterministic setting). Let us note that the subsequent definitions, especially the differential flatness definitions, are made in the presence of fluctuations. This is not to say that we lose the deterministic character of this notion, since these perturbations may well be deterministic functions. They may also be stochastic, if they are sufficiently smooth, i.e., differentiable up a certain order.

Note that the action $\mathbf{u}(t)$, or the *control input*, are functions enabling us to act on the system in order to fulfil a specified goal. The dynamics equations thus form an *undetermined* system of differential equations, since the control functions $\mathbf{u}(t)$ are not a priori determined. Once the control variables are fixed (i.e., substituted with known functions of time), the system (1) becomes determined (i.e., can be solved or integrated).

An *output* $\mathbf{y}(t)$ is generated by the model. These outputs or observations may represent signals coming from the senses, in case the agent is a living being, or from the sensors in the case of an artificial system. More precisely, we have the following definition.

Definition 3 (action (control input) and output). *Consider a model with variables $\mathbf{z}(t) = (z_1(t), \dots, z_r(t))$, fluctuations $\boldsymbol{\zeta}(t) = (\zeta_1(t), \dots, \zeta_s(t))$, and equations*

$$\Phi(\mathbf{z}, \dot{\mathbf{z}}, \dots, \mathbf{z}^{(\rho_z)}, \boldsymbol{\zeta}, \dot{\boldsymbol{\zeta}}, \dots, \boldsymbol{\zeta}^{(\rho_\zeta)}) = 0 \tag{3}$$

A control input, or action, of the model is an m -tuple (with $m < r$) of variables $\mathbf{u}(t) = (u_1(t), \dots, u_m(t))$ with the following properties.

1. *Endogenous character.*

The components of \mathbf{u} can be (locally and generically) expressed as a function of \mathbf{z} and its derivatives:

$$u_i = \psi_{ui}(\mathbf{z}, \dot{\mathbf{z}}, \dots, \mathbf{z}^{(r_{ui})}) \tag{4}$$

2. *Differential independence.*

There does not exist any nontrivial differential relation of the form

$$R_u(\mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(\beta_u)}) = 0 \tag{5}$$

3. *Any system variable is influenced by the input.*

Every z_i , $i = 1, \dots, r$ satisfies

$$E_i(z_i, \dot{z}_i, \dots, z_i^{(\delta_{zi})}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(\delta_{ui})}, \boldsymbol{\zeta}, \dot{\boldsymbol{\zeta}}, \dots, \boldsymbol{\zeta}^{(\delta_{\zeta i})}) = 0 \tag{6}$$

An output $\mathbf{y}(t) = (y_1(t), \dots, y_p(t))$ (with $p < r$) is a p -tuple, where the y_i s are functions of the model's variables:

$$y_i = h_{y_i}(\mathbf{z}, \dot{\mathbf{z}}, \dots, \mathbf{z}^{(v_z)}, \boldsymbol{\zeta}, \dot{\boldsymbol{\zeta}}, \dots, \boldsymbol{\zeta}^{(v_\zeta)}) \tag{7}$$

Remark 1 (Local genericity). Generally, in (4), we have to deal with implicit functions:

$$\Psi(\mathbf{u}, \mathbf{z}, \dot{\mathbf{z}}, \dots, \mathbf{z}^{(r_u)}) = 0 \tag{8}$$

And, locally, in the neighbourhood of generic regular points (i.e., points where the Jacobian of Ψ with respect to \mathbf{u} is regular), we can solve (8) for \mathbf{u} using the implicit function theorem. This will be termed locally and generically here and in the forthcoming definitions.

The state variables represent the instantaneous memory of the system: once the control (action) variables have been determined, the knowledge of the state variables (at time t) enables the prediction of the future state (at time $t + dt$). A complementary formulation is the following: the state of a dynamical system is a set of physical quantities, the specification of which (in the absence of external excitation) completely determines the evolution of the system. More precisely, we have the following definition.

Definition 4 (state). Consider a model with variables $\mathbf{z}(t) = (z_1(t), \dots, z_r(t))$, input $\mathbf{u}(t) = (u_1(t), \dots, u_m(t))$, fluctuations $\boldsymbol{\zeta}(t) = (\zeta_1(t), \dots, \zeta_s(t))$, and equations

$$\Phi(\mathbf{z}, \dot{\mathbf{z}}, \dots, \mathbf{z}^{(\rho_z)}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(\rho_u)}, \boldsymbol{\zeta}, \dot{\boldsymbol{\zeta}}, \dots, \boldsymbol{\zeta}^{(\rho_\zeta)}) = 0 \tag{9}$$

A state of the model is an n -tuple (with $n \leq r$) of variables $\mathbf{x}(t) = (x_1(t), \dots, x_n(t))$ with the following properties.

1. Endogenous character.

The components of \mathbf{x} can be (locally and generically) expressed as a function of \mathbf{z} and its derivatives:

$$x_i = \psi_{x_i}(\mathbf{z}, \dot{\mathbf{z}}, \dots, \mathbf{z}^{(r_{x_i})}) \tag{10}$$

2. Independence with respect to the input.

There does not exist any nontrivial differential relation of the form

$$R_x(\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(\beta_x)}) = 0 \tag{11}$$

3. Representation property.

Every variable \check{x} that is a function of \mathbf{z} and of its time derivatives can be expressed through \mathbf{x} , \mathbf{u} and its derivatives. In other words, there exists (locally and generically) a representation of the form

$$\check{x} = \phi_{\check{x}}(\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(\beta_{\check{x}})}, \boldsymbol{\zeta}, \dot{\boldsymbol{\zeta}}, \dots, \boldsymbol{\zeta}^{(\beta_{\check{\zeta}_i})}) \tag{12}$$

By virtue of the above representation property, for all $i = 1, \dots, n$, the time derivatives \dot{x}_i are expressed through \mathbf{x} , \mathbf{u} and its time derivatives. This can be otherwise stated as follows.

Definition 5 (State representation). Consider a model with variables $\mathbf{z}(t) = (z_1(t), \dots, z_r(t))$, input $\mathbf{u}(t) = (u_1(t), \dots, u_m(t))$, fluctuations $\boldsymbol{\zeta}(t) = (\zeta_1(t), \dots, \zeta_s(t))$, equations

$$\Phi(\mathbf{z}, \dot{\mathbf{z}}, \dots, \mathbf{z}^{(\rho_z)}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(\rho_u)}, \boldsymbol{\zeta}, \dot{\boldsymbol{\zeta}}, \dots, \boldsymbol{\zeta}^{(\rho_\zeta)}) = 0 \tag{13}$$

and state \mathbf{x} . Then, there exists a so-called state representation of the form

$$\dot{x}_i = \phi_{x_i}(\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(\beta_{x_i})}, \boldsymbol{\zeta}, \dot{\boldsymbol{\zeta}}, \dots, \boldsymbol{\zeta}^{(\beta_{\zeta_i})}) \tag{14}$$

When the right-hand side of (14) does not depend on the derivatives of the input \mathbf{u} (i.e., when $\beta_{x_i} = 0$ for all $i = 1, \dots, n$), the state representation is called classical.

A realisation of a model consists of a state and a state representation for this model, as the following definition states.

Definition 6 (Realisation). Consider a model with variables $\mathbf{z}(t) = (z_1(t), \dots, z_r(t))$, fluctuations $\zeta(t) = (\zeta_1(t), \dots, \zeta_s(t))$, and equations

$$\Phi(\mathbf{z}, \dot{\mathbf{z}}, \dots, \mathbf{z}^{(\rho_z)}, \zeta, \dot{\zeta}, \dots, \zeta^{(\rho_\zeta)}) = 0 \tag{15}$$

where \mathbf{z} comprises an input \mathbf{u} and output \mathbf{y} : $\{u_1, \dots, u_m, y_1, \dots, y_p\} \subset \{z_1, \dots, z_r\}$. A realisation of this model consists in the existence of a state \mathbf{x} and a state representation of the form (14) for the model.

2.3. Fluctuation Choice

An interesting point about the definitions above is that they depend upon there being differentiable (i.e., smooth or analytic) fluctuations. This means we need to think carefully about what we mean by fluctuations—an important topic generally in the study of stochastic dynamical systems. Several choices can be made for the fluctuations ζ_* , $*$ $\in \{x, y\}$ appearing in the above generative models. These include the following:

- Stochastic processes in the Itô sense.
- Stochastic processes in the Stratonovich sense.
- Nonstandard infinitesimals (see, e.g., [16–18]);
- Stochastic processes with Hölder continuous sample paths, yielding random ODEs (RODEs) (see, e.g., [19]);
- Rough paths (see, e.g., [20]);
- Random Fourier series (RFS) (see, e.g., [21,22] for metric and convergence properties, with extensions on Riemannian manifolds [23] and locally compact groups [24]; see also [25] for an engineer’s view).

Let us choose the latter because they may furnish a convenient \mathcal{C}^∞ form of fluctuations, and the solution of random ODEs can be shown to converge to solutions of a Stratonovich stochastic differential equation (see, e.g., [25], Theorem 5.1). We may thus consider so-called *periodic smooth random functions*:

$$\zeta(t) = a_0 + \sum_{j=1}^r \left[a_j \cos\left(\frac{2\pi j}{L} t\right) + b_j \sin\left(\frac{2\pi j}{L} t\right) \right], \quad r = \lfloor L/\lambda \rfloor \tag{16}$$

where $\lambda, L > 0$, each a_j and b_j is an independent sample from $\mathcal{N}(0, 1/(2r + 1))$ (where $\mathcal{N}(\mu, V)$ denotes the real normal distribution of mean μ and variance V), and $\lfloor \cdot \rfloor$ is the floor function. This type of function is L -periodic, entire, and $(2\pi/\lambda)$ -band-limited.

Then, multiplying (16) by $\sqrt{2/\lambda}$ and taking the variance to be $1/((2r + 1)\lambda)$ instead of $1/(2r + 1)$, we obtain the notion of a *big periodic smooth random function*. Note that since $r \approx L/\lambda$, we have $2/((2r + 1)\lambda) \approx 1/L$. Thus, in the big normalisation, the random coefficients of the sum ζ have variances essentially independent of λ as $\lambda \rightarrow 0$. Then, as $\lambda \rightarrow 0$, indefinite integrals of big smooth random functions converge with probability 1 to standard Brownian paths (see, e.g., Theorem 4.3 of [25] or Theorem 2, p. 236 of [21]). Regarding smooth random functions, the reader may also consult the very interesting works of R.J. Adler and J.E. Taylor, [26–28], with insightful chapters on geometry and smooth random manifolds, among others. Interestingly, this technology underwrites generative models in classical brain imaging analysis, namely, statistical parametric mapping, based upon random field theory and topological inference [29,30]. The previous convergence can also be related to the Wong–Zakai theorems (see, e.g., [31]) used in the insightful paper [32], Subsection 1.3.2.

Smooth random functions may not be an apt choice at atomic scales, where a particle’s movements are highly erratic. However, they become particularly appropriate at the cell

and mesoscopic scales and most probably appropriate at macroscopic scales at which many fluctuations are generated by dynamical systems that evolve over a timescale faster than that considered for a given control problem (see [1] and the seminal observations of Stratonovich: “a certain care must be taken in replacing an actual process by Markov process, since Markov processes have many special features, and, in particular, differ from the processes encountered in radio engineering by their lack of smoothness . . . Any random process actually encountered in radio engineering is analytic, and all its derivatives are finite with probability one” ([33], pp. 122–124)).

Remark 2 (Wavelet random series). *One may be tempted, in the spirit of the above, to consider wavelet random series, since wavelet expansions are better behaved than their Fourier counterparts. However, recent works by C. Esser, S. Jaffard, and B. Vede [34] (see also [35]) suggest that caution is in order; in contrast to Fourier series, the randomisation of almost every continuous function gives an almost surely nowhere locally bounded function.*

We still refer to the notion of a tube around a nominal function, i.e., for $f(t) + \zeta(t)$, where f is a deterministic function and ζ a periodic smooth random function. We shall be concerned with estimating the probability of leaving a tube of width 2λ (for $\lambda > 0$). For instance, one has the following estimate (see [36]):

$$P\left(\sup_{t \in [0, T]} \zeta(t) \geq \lambda\right) \leq \left(\frac{\sigma}{\sqrt{2\pi}\lambda} + \frac{T\sqrt{\lambda_2}}{2\pi\sigma}\right) \exp\left(\frac{-\lambda^2}{2\sigma^2}\right)$$

with $\sigma = \sqrt{E(|\zeta(t)|^2)}$, $\lambda_2 = E(|\dot{\zeta}(t)|^2)$. The reader may consult [36] for generalisations of this bound inequality. The notable paper by [37] also contains results for the volume of tubes using Gaussian–Minkowski functionals.

Example 1 (Smooth random functions). *The following two plots are examples of smooth random functions with varying wavelength λ . More precisely, the left plot is a smooth random function (a sum of sines and cosines of the form (16)) with parameters $L = 8$, $\lambda = 2$; recall that the sum has $r + 1$ terms, with $r = \lfloor L/\lambda \rfloor$. Thus, when $L = 8$, $\lambda = 2$, we obtain $r = 4$ terms in the sum. The right plot is an analogous sum with $r = \lfloor 8/0.1 \rfloor = 80$ terms (See Figure 1).*

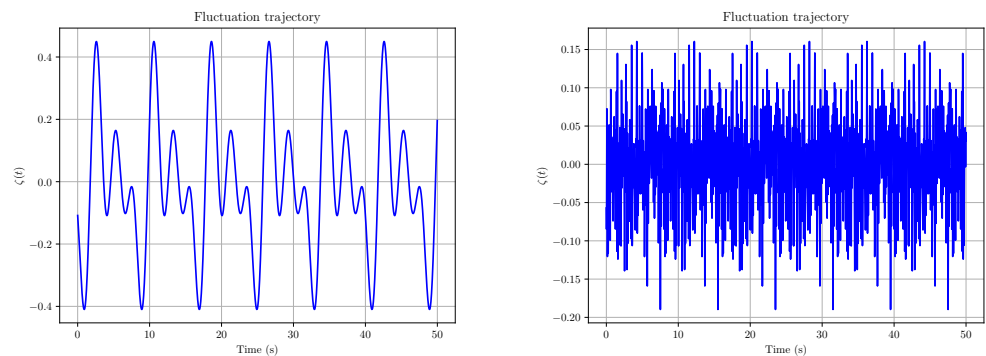


Figure 1. Examples of smooth random functions with parameters $L = 8$, $\lambda = 2$ (left plot) and $L = 8$, $\lambda = 0.1$ (right plot).

In order to offer a more concrete intuition as to the notions of state, control input, and output, let us consider a simple, although generic, example.

Example 2 (Simple generic example; model). Consider the following simple example with scalar action (control) and sensor output:

$$\dot{x}_1 = x_2 + \zeta_{x_1} \tag{17a}$$

$$\dot{x}_2 = f(x_1, x_2) + u + \zeta_{x_2} \tag{17b}$$

$$y = x_1 + \zeta_y \tag{17c}$$

where f is a smooth function: $f \in C^\infty(\mathbb{R}^2)$. The fluctuations are taken as smooth random Gaussian functions:

$$\zeta_{x_1}(t) = \sum_{k \geq 0} a_{k_1} \phi_k(t), \quad a_{k_1} = \mathcal{N}(0, \sigma_x) \tag{18a}$$

$$\zeta_{x_2}(t) = \sum_{k \geq 0} a_{k_2} \phi_k(t), \quad a_{k_2} = \mathcal{N}(0, \sigma_x) \tag{18b}$$

$$\zeta_y(t) = \sum_{k \geq 0} c_k \phi_k(t), \quad c_k = \mathcal{N}(0, \sigma_y) \tag{18c}$$

$$\phi_k(t) = \cos\left(\frac{2\pi j}{L} t\right), \quad r = \lfloor L/v \rfloor \tag{18d}$$

with $v, L > 0$. Here (17) are the equations of a generative model, in the sense of Definition 1, $\mathbf{x} = (x_1, x_2)$ is the state of the generative model, in the sense of Definition 4, u is the input, and y is an output, in the sense of Definition 3.

This type of model includes all models predicated upon Newton’s law of motion, such as

$$M\ddot{y} = F(y, \dot{y}) + v \tag{19}$$

where M is a mass, F is a model for internal forces depending on the position y and velocity \dot{y} , and v is an external force.

A more general case would be the following:

$$\dot{x}_1 = f_1(x_1, x_2) + \zeta_{x_1} \tag{20a}$$

$$\dot{x}_2 = f_2(x_1, x_2, x_3) + \zeta_{x_2} \tag{20b}$$

⋮

$$\dot{x}_n = f(x_1, \dots, x_n) + u + \zeta_{x_n} \tag{20c}$$

$$y = x_1 + \zeta_y \tag{20d}$$

with f_i s being smooth functions of their arguments, invertible with respect to the final function. All the computations made for the simple example (17) can be readily extended to the above.

3. Free Energy, Flatness, and Conceptual Similarities

3.1. Free and Expected Free Energy

One can understand both active inference- and differential flatness-informed control schemes as identifying mappings from sensation to action from models that detail the influence of action on sensation. Flatness rests upon there being an invertible mapping between action and sensation such that a desired sensory trajectory uniquely determines the actions that generate it. Active inference involves the selection of actions that bring sensations in line with the mode of a marginal density of sensory data implied by a generative model. This is mediated by reflexive actions determined by sensory data. The marginal density that identifies desired sensory trajectories is often specified in terms of an expected free energy—whose role is to determine prior plausibilities of alternative action sequences based upon their capacity to minimise the Kullback–Lieber divergence (also known as risk) between desired and anticipated sensory trajectories.

Consistent with control-theoretic formulations of the sort outlined above, active inference can be formulated as optimising a functional of a model that relates controllable variables to some observable outcomes. Specifically, it depends upon optimisation (minimisation) of a variational free energy that acts as an upper bound on the surprise or negative log marginal likelihood of those observations. Variational free energy can be formulated in several ways to quantify the performance of a system engaging in active inference in terms of energies (i.e., surprise) and divergences (i.e., relative entropies):

$$\mathcal{F}(\mathbf{u}(t), \mathbf{y}(t)) = \underbrace{- \int q(\mathbf{x}(t)|\mathbf{u}(t)) \ln p(\mathbf{x}(t), \mathbf{y}(t)|\mathbf{u}(t)) d\mathbf{x}(t)}_{\text{Energy}} + \underbrace{\int q(\mathbf{x}(t)|\mathbf{u}(t)) \ln q(\mathbf{x}(t)|\mathbf{u}(t)) d\mathbf{x}(t)}_{\text{Neg-entropy}} \quad (21a)$$

$$= \underbrace{\int q(\mathbf{x}(t)|\mathbf{u}(t)) \ln \frac{q(\mathbf{x}(t)|\mathbf{u}(t))}{p(\mathbf{x}(t)|\mathbf{y}(t), \mathbf{u}(t))} d\mathbf{x}(t)}_{\text{Divergence}} + \underbrace{(- \ln p(\mathbf{y}(t)|\mathbf{u}(t)))}_{\text{Surprise}} \quad (21b)$$

$$= \underbrace{\int q(\mathbf{x}(t)|\mathbf{u}(t)) \ln \frac{q(\mathbf{x}(t)|\mathbf{u}(t))}{p(\mathbf{x}(t)|\mathbf{u}(t))} d\mathbf{x}(t)}_{\text{Complexity}} - \underbrace{\int q(\mathbf{x}(t)|\mathbf{u}(t)) \ln p(\mathbf{y}(t)|\mathbf{x}(t), \mathbf{u}(t)) d\mathbf{x}(t)}_{\text{Accuracy}} \quad (21c)$$

The above presents variational free energy \mathcal{F} as a functional of two probability distributions (for discrete states) or densities (for continuous states). The density labelled p is that associated with our generative model, while q represents a density variously referred to as a recognition density, an approximate posterior density, or a variational density. For the purposes of this paper, the variational density is assumed to have already been optimised such that $q(\mathbf{x}(t)|\mathbf{u}(t)) = p(\mathbf{x}(t)|\mathbf{y}(t), \mathbf{u}(t))$. Each formulation of the free energy depends upon different factorisations of the generative model. When expressed as a joint density, the minimisation of free energy can be seen as a constrained maximum entropy problem. On factorising into conditionals and marginal likelihoods, the free energy is seen to be an upper bound on surprise—the negative log marginal likelihood, Bayesian model evidence, or improbability of an observation under a given generative model. Finally, factorising a generative model into priors and likelihoods gives us a balance between complexity—how far we must move from prior beliefs to explain observations—and the accuracy with which we can account for sensory inputs.

Further to this, one can formulate expected free energies that highlight other forms of discrepancy that are especially relevant for the optimisation of control. When formulated explicitly in a pathwise setting, we have

$$\mathcal{G}(\mathbf{u}(t)) = - \underbrace{\iint q(\mathbf{x}(t), \mathbf{y}(t)|\mathbf{u}(t)) \ln \frac{q(\mathbf{x}(t), \mathbf{y}(t)|\mathbf{u}(t))}{p(\mathbf{x}(t), \mathbf{y}(t)|\mathbf{u}(t))} d\mathbf{x}(t) d\mathbf{y}(t)}_{\text{Mutual information}} - \underbrace{\int q(\mathbf{y}(t)|\mathbf{u}(t)) \ln p(\mathbf{y}(t)|\mathbf{y}_r(t)) d\mathbf{y}(t)}_{\text{Expected value}} \quad (22a)$$

$$= \underbrace{\int q(\mathbf{y}(t)|\mathbf{u}(t)) \ln \frac{q(\mathbf{y}(t)|\mathbf{u}(t))}{p(\mathbf{y}(t)|\mathbf{y}_r(t))} d\mathbf{y}(t)}_{\text{Risk}} + \underbrace{\left(- \iint q(\mathbf{y}(t), \mathbf{x}(t)|\mathbf{u}(t)) \ln p(\mathbf{y}(t)|\mathbf{x}(t), \mathbf{u}(t)) d\mathbf{x}(t) d\mathbf{y}(t) \right)}_{\text{Ambiguity}} \quad (22b)$$

The expected free energy is typically used for planning, where we might integrate this quantity along future paths and assign paths of control states higher probabilities for lower expected free energies. Of particular relevance for the discussion that follows is the idea

that, by placing priors over the path of observations we wish to obtain (here indicated by conditioning upon a goal \mathbf{y}_r), optimisation of expected free energy involves determining the set of control paths that would realise these outcomes.

We can now identify at least three different aspects of optimality, namely surprise, inadequacy, and discrepancy:

1. The *discrepancy in observation* (inaccuracy) in gathering information from the world. This discrepancy is implicit in the (negative) accuracy term of the variational free energy, which under Gaussian assumptions will reduce to the square of a prediction error: $\varepsilon_{obs} = \mathbf{y} - \hat{\mathbf{y}}$, where $\hat{\mathbf{y}}$ is the predicted observation under beliefs about \mathbf{x} . This is the (sensory) prediction error that features in predictive coding or Bayesian filtering, linear quadratic control, and model predictive control.
2. The *discrepancy in action* in the way the agent acts in the world. This discrepancy is between the goal $\mathbf{y}_r(t)$, a predefined trajectory to be followed, and the actual position $\mathbf{y}(t)$: $\varepsilon_{act} = \mathbf{y} - \mathbf{y}_r$. This is implicit in the risk term of the expected free energy, which quantifies the divergence between the distribution anticipated under a set of control states and the distribution anticipated given the goal.
3. The *discrepancy in modelling* in the way the agent represents the world internally. This discrepancy would normally be quantified using the marginal likelihood of sensory observations under the model and is reflected in the surprise term of the variational free energy. When free energy is minimal with respect to q , it becomes a tight bound on this discrepancy. For this reason, variational free energy is often used as a tractable method of approximating Bayes factors to compare alternative models in statistical inference.

The free energies above (\mathcal{F} and \mathcal{G}) can be seen as so-called global Lyapunov functions (see, e.g., [38]) that capture these aspects seen in optimal control (see, e.g., [39,40]). In what follows, we shall mostly be interested in \mathcal{G} , and specifically the risk term, leaving the other aspects to future works. More precisely, we are interested in whether the notion of differential flatness coheres with the selection of generative models that optimise \mathcal{F} and \mathcal{G} . The derivation of trajectory tracking action laws on differentially flat models will be seen to minimise the above risk in \mathcal{G} (see Section 4.6).

Table 1 summarises the notations used so far, where bold symbols are vectors.

Table 1. Notation table.

Variable	Meaning
\mathbf{x}	Hidden state vector
\mathbf{y}	Output vector (e.g., sensor data)
\mathbf{u}	Action vector
ζ_x	Fluctuations in hidden state dynamics
ζ_y	Fluctuations in sensor data
\mathcal{F}	Variational free energy
\mathcal{G}	Expected free energy

3.2. Differential Flatness

3.2.1. Controllability

A ubiquitous notion—when one wishes to steer a system—is the global controllability, as stated in the following definition (see, e.g., [41]).

Definition 7. A system such as (1), but idealised such that we assume the fluctuations are infinitely precise,

$$\dot{x} = f_m(x, u) \quad (23a)$$

$$y = h_m(x) \quad (23b)$$

is said to be globally controllable if, for any time instants t_0 and t_1 and initial and final states x_0 and x_1 , there exists an action (a control law) u in a space of admissible controls steering the system from $x(t_0) = x_0$ to $x(t_1) = x_1$.

The reader may note that this definition is purely descriptive: it does not contain any constructive procedure for steering the system; it is not possible to infer, from reading the definition alone, the form of the control law to be applied to go from x_0 to x_1 . This is a definition of the existence-of-a-solution kind, not of a solution construction to the given problem.

The reader may note that this definition is purely descriptive, in that it does not contain any constructive procedure for steering the system. Moreover it is pointwise in spirit, rather than pathwise. Indeed, this definition does not say anything about the path that links the initial to the final state. This path can be implausible and still fulfil the controllability requirements. We shall see in the next subsection a stronger and, in our sense, more useful property, namely, differential flatness.

3.2.2. Motivation Through Observation and Action

The premise of active inference is that an agent seeks to minimise the surprise or divergence between its beliefs or expectations about the surrounding environment and the actual state it experiences. This minimisation can, in principle, be enforced through effective information gathering and action or be engrained in the very structure of the agent's model (such as being refined through learning and evolution).

We will examine a case where the requisite efficiency is encoded in the agent's model structure itself. More precisely the imperatives for perception and action are directly fulfilled through the following:

(Odsf) *Observation discrepancy structural fulfilment.* The state x can be recovered through what the agent is able to know directly (i.e., without any inference, reflection, or computation), that is y , u , and their time derivatives. This amounts to the system being *constructively observable*.

(Adsf) *Action discrepancy structural fulfilment.* The link between the goal y_{μ} and the action u —required to reach that goal—is direct, in that the action is given as a function of the goal and its time derivatives. This amounts to the system being *left-invertible*.

A system such as (23) is said to be *left-invertible with respect to z* if the action u is a function of z and its derivatives (see Property 5 of [42]). Thus, for a dynamical system to fulfil both (Odsf) and (Adsf), one needs to have a function ω such that both the state x and the action u can be expressed in terms of ω and its time derivatives. This corresponds to *differential flatness* [5,43,44], a property shared by a great number of practical dynamical systems (see, e.g., [45] and the references therein).

3.2.3. Motivation Through Direct and Inverse Views

The fundamental property of flat systems is that all their solutions can be functionally parameterised by a finite number of functions and their time derivatives. Although flatness is a relatively recent notion, introduced in control theory in the 1990s [5,43,44], it actually has a long history: a similar notion— of systems of undetermined differential equations

integrable without integration—dates back to Hilbert [46] and Cartan [47]. Indeed, the control system $\dot{x} = f(x, u)$ can be seen as an underdetermined differential system consisting of n equations $\dot{x}_i = f_i(x, u)$, for $i = 1, \dots, n$, and $n + m$ variables (n states and m action variables). The difference between the number of variables and the number of equations gives the degrees of freedom number of the system. It follows that m functions can be chosen freely. In the context of control systems, one usually chooses freely the input $u(t) = (u_1(t), \dots, u_m(t))$ and then integrates in order to compute the state $x(t)$. But is this the only way to achieve this?

In order to answer this question, consider the following simple single-input system:

$$\dot{x}_1 = x_2 \tag{24a}$$

$$\dot{x}_2 = u \tag{24b}$$

We can freely choose $u(t)$, then integrate it once to compute $x_2(t)$, and then integrate it a second time to compute $x_1(t)$. Let us now choose freely $x_1(t)$, differentiate it once to get $x_2(t)$, and then differentiate $x_2(t)$ to obtain $u(t)$. It follows that the system with Equation (24) admits two functional parameterisations: one via the input, for which we have to integrate twice (which is, in the general case of $\dot{x} = f(x, u)$, difficult and, sometimes, impossible analytically), and one via the state x_1 , for which we have to differentiate twice (which is always possible in a straightforward way). Hence, there are underdetermined differential systems (also known as control systems) that are solvable without integration, namely, differentially flat systems.

3.2.4. Formal Definition

Let us consider the central definition of this subsection, stated for systems with specified parameters, for ease of reading (see, e.g., [15,42,45]; see also the Supplementary Material, File AIandFP-FlatnessAndSRR-HMTPKF-2026-SupplMaterial-v1.pdf, Section D and [48] for a Python library, version 0.10.2).

Definition 8 (Differential flatness). Consider a model with variables $z(t) = (z_1(t), \dots, z_r(t)) \in \mathbb{R}^r$, fluctuations $\zeta(t) = (\zeta_1(t), \dots, \zeta_s(t)) \in \mathbb{R}^s$, and equations

$$\Phi(z, \dot{z}, \dots, z^{(\rho_z)}, \zeta, \dot{\zeta}, \dots, \zeta^{(\rho_\zeta)}) = 0 \tag{25}$$

The model is called differentially flat if there exists an m -tuple of variables, $\omega(t) = (\omega_1(t), \dots, \omega_m(t))$, named flat outputs, with the following three properties.

1. Endogenous character. We have (locally and generically)

$$\omega = h(z, \dot{z}, \dots, z^{(\eta_z)}) \quad \eta_z \in \mathbb{N} \tag{26}$$

In other words, the components of the flat output are combinations of the system’s variables.

2. Functional parameterisation.

The system’s variables z can be (locally and generically) expressed through the flat output and a finite number of its derivatives:

$$z = A_z(\omega, \dot{\omega}, \dots, \omega^{(\alpha_z)}, \zeta, \dot{\zeta}, \dots, \zeta^{(\alpha_\zeta)}) \tag{27}$$

with α_z, α_ζ integers, such that the system’s equations

$$\Phi(A_z \dot{A}_z, \dots, A_z^{(\rho_z)}, \zeta, \dot{\zeta}, \dots, \zeta^{(\rho_\zeta)}) = 0$$

are identically verified.

3. *Differential independence.* The components of the flat output are differentially independent; i.e., any differential relation

$$\Xi(\omega, \dot{\omega}, \dots, \omega^{(\beta)}) = 0$$

is necessarily trivial: $\Xi \equiv 0$.

Remark 3 (Differential flatness–state-space form case). *In case the model has the following form:*

$$\dot{x} = f(x, u) + \zeta \tag{28}$$

with state $x(t) \in \mathbb{R}^n$, action $u(t) \in \mathbb{R}^m$, and fluctuation $\zeta(t) \in \mathbb{R}^n$, the functional parametrisation (27) becomes

$$x = A(\omega, \dot{\omega}, \dots, \omega^{(\alpha_x)}, \zeta, \dot{\zeta}, \dots, \zeta^{(\alpha_x)}) \tag{29a}$$

$$u = B(\omega, \dot{\omega}, \dots, \omega^{(\alpha_u)}, \zeta, \dot{\zeta}, \dots, \zeta^{(\alpha_u)}) \tag{29b}$$

Remark 4. *Note that the conventional definition of differential flatness is made for deterministic (mean) models:*

$$\dot{x} = f(x, u) \tag{30}$$

with state $x(t) \in \mathbb{R}^n$ and action $u(t) \in \mathbb{R}^m$, in which case the functional parametrisation (27) becomes

$$x = A(\omega, \dot{\omega}, \dots, \omega^{(\alpha_x)}) \tag{31a}$$

$$u = B(\omega, \dot{\omega}, \dots, \omega^{(\alpha_u)}) \tag{31b}$$

The fluctuations considered above in (29a) and (29b) are not necessarily supposed to be known functions of time, but rather are seen as generic, sufficiently differentiable functions. The relations (29) are valid whatever the fluctuations ζ may be at this structural level. See Section 3.2.5 below for a discussion on this functional parametrisation.

Remark 5 (Local genericity). *As in Remark 1, the relations of endogenous character in (26) and of functional parameterisation in (27), naturally yield implicit functional relations:*

$$H(\omega, z, \dot{z}, \dots, z^{(\eta_z)}) = 0 \tag{32a}$$

$$\Lambda_z(z, \omega, \dot{\omega}, \dots, \omega^{(\alpha_z)}, \zeta, \dot{\zeta}, \dots, \zeta^{(\alpha_\zeta)}) = 0 \tag{32b}$$

And, locally, in the neighbourhood of generic regular points (i.e., points where the Jacobian of H with respect to ω and Λ_z with respect to z is regular), we can solve (32a) for ω and (32b) for z using the implicit function theorem.

Let us note that the differential flatness definition is made in the presence of fluctuations. The latter may be deterministic or smooth stochastic functions (i.e., differentiable up a certain order). We will consider relations furnishing, in particular, the action as a functional of the sensory and fluctuation paths. This enables one to study precisely and quantitatively the influence that each perturbation function may exert on the action. Moreover the relation yielding the action—respectively, the state—as a function of sensor and fluctuation paths is generic, in the sense that it is valid for any (sufficiently smooth) sensory and fluctuation paths. In this sense, the notion of differential flatness is *agnostic with respect to the stochasticity* of the generative model. The latter may be deterministic (as is the case for mean generative models),

subject to deterministic, but unknown, perturbations or subject to stochastic (and sufficiently smooth) fluctuations. The previous definition readily affords the following characterisation:

Proposition 1. *A differentially flat system in state-space form is a system like (28), i.e., $\dot{x} = f(x, u)$, which is observable and left invertible with respect to ω .*

To summarise what has been said so far, in order to express what we think we know about the system (our beliefs) and what we wish for the system (our expectations), we require a model structure that expresses first the agent system’s constructive observability (what we know or believe) and second the agent system’s left invertibility (what we expect and where the action u is expressed in terms of the flat output goal, ω_r). The latter amounts to differential flatness, which is no more than the *generator character of the flat output, i.e., the functional parameterisation and the differential polynomial independence of the flat output components*. The last item enables one to choose the various components $\omega_{ir}(t)$ of $\omega_r(t)$ independently of each other. Both properties are reminiscent of the basis notion in vector spaces (i.e., the minimally generating and maximally independent characters); indeed, in a differential algebraic setting, the notion of flat output corresponds to a *differential transcendence basis* (see [5]).

An interesting point of contact with active inference is that optimisation of expected free energy functionals implies a high degree of mutual information between different components of a model (specifically, between states and observations). This implies precise mappings from actions, via states, to observations. Crucially, that same mutual information means that we would expect observations to be highly informative about states and, possibly, actions. The potential to recover subsets of variables from others in such models is heuristically compatible with the differential flatness concept. Furthermore, the need for differential independence has an interesting link with the interpretation of variational free energy as an objective function for constrained maximum entropy inference—where in the absence of the ‘energy’ constraints, the best configuration is that with maximum entropy without mutual constraints between the components of a system.

Remark 6 (Tracking an arbitrary output). *Suppose that the agent’s model is differentially flat. The agent’s goal y_r may coincide with the flat output goal: $\omega_r = y_r$. If this is not the case, consider the expression of the agent’s output y as a function of the flat output ω and its derivatives:*

$$y = A_y(\omega, \dot{\omega}, \dots, \omega^{(\alpha_y)}, \zeta, \dot{\zeta}, \dots, \zeta^{(\alpha_\zeta)}) \tag{33}$$

and view this expression as a differential equation in ω :

$$\omega^{(\alpha_y)} = \tilde{A}_y(\omega, \dot{\omega}, \dots, \omega^{(\alpha_y-1)}, y, \zeta, \dot{\zeta}, \dots, \zeta^{(\alpha_\zeta)}) \tag{34}$$

Substituting y by the goal y_r yields the flat output goal ω_r as a solution of

$$\omega_r^{(\alpha_y)} = \tilde{A}_y(\omega_r, \dot{\omega}_r, \dots, \omega_r^{(\alpha_y-1)}, y_r, \zeta, \dot{\zeta}, \dots, \zeta^{(\alpha_\zeta)}) \tag{35}$$

3.2.5. Functional Parameterisation

The functional parameterisation property is an essential, if not by far the most essential feature, of differential flatness. Indeed, the original model

$$\dot{x} = f(x, u) + \zeta_x \tag{36a}$$

$$y = h(x) + \zeta_y \tag{36b}$$

is totally equivalent to its functional parametric form:

$$x = A_m(\omega, \dot{\omega}, \dots, \omega^{(\rho_x)}, \dot{\zeta}, \dots, \zeta^{(\rho_x)}) \tag{37a}$$

$$u = B_m(\omega, \dot{\omega}, \dots, \omega^{(\rho_u)}, \dot{\zeta}, \dots, \zeta^{(\rho_u)}) \tag{37b}$$

$$\zeta = (\zeta_x, \zeta_y) \tag{37c}$$

And, crucially, the functional parametric form (37) is *quasi-static*, whereas the original model (36) has a *dynamical form*. Hence, when one wishes to deal with (36), one is naturally tempted to seek the solution of the ODE system (36a), and most often, an analytical solution cannot be found. But when dealing with the functional parametric form (37), all system variables, i.e., x and u here, are parametrised by the function ω , since once we know the function ω , then we readily know the functions x and u through (37) (assuming, in a structural step such as this one, that the fluctuations ζ are known or measured).

Note that this parametrisation is a *pathwise* one and that it carries over the whole fluctuation; it can thus be envisioned either as a *tube around the mean* (see, e.g., [27,28,36]) or as a *function sheaf* (see, e.g., [49,50]; see also [51] for an elementary introduction and [52] for more complete while still accessible references).

Thus, the action, the control u , and the state x are envisioned as *functionals of the flat output and of the fluctuations*, i.e., as functions over spaces of functions:

$$x : \mathcal{F}^m \times \mathcal{F}^{n+p} \longrightarrow \mathcal{F}^n$$

$$(\omega, \zeta) \longrightarrow A_m(\omega, \dot{\omega}, \dots, \omega^{(\rho_x)}, \dot{\zeta}, \dots, \zeta^{(\rho_x)}) \tag{38a}$$

$$u : \mathcal{F}^m \times \mathcal{F}^{n+p} \longrightarrow \mathcal{F}^m$$

$$(\omega, \zeta) \longrightarrow B_m(\omega, \dot{\omega}, \dots, \omega^{(\rho_u)}, \dot{\zeta}, \dots, \zeta^{(\rho_u)}) \tag{38b}$$

$$\zeta = (\zeta_x, \zeta_y), \quad \mathcal{F} \subseteq C^\rho(\mathbb{R}), \quad \rho = \max(\rho_x, \rho_u)$$

And it is the *study of these functionals* that may be of great interest to the active inference community.

3.3. Conceptual Similarities

We can now see—with a simple concrete example—that both differential flatness and free energy minimisation enforce an inverse mapping from the sensed output to the action, namely, the control input.

Example 3 (Simple generic example; similarities). *Recall the following simple example with scalar action (control) and sensor output:*

$$\dot{x}_1 = x_2 + \zeta_{x_1} \tag{39a}$$

$$\dot{x}_2 = f(x_1, x_2) + u + \zeta_{x_2} \tag{39b}$$

$$y = x_1 + \zeta_y \tag{39c}$$

This model is differentially flat with flat output y . Indeed, the functional parameterisation is obtained as follows. Equation (39c) yields

$$x_1 = y - \zeta_y \tag{40}$$

Equation (39a) yields

$$x_2 = \dot{x}_1 - \zeta_{x_1} = \dot{y} - \dot{\zeta}_y - \zeta_{x_1} \tag{41}$$

Then, using Equation (39b),

$$u = \ddot{y} - f(y - \zeta_y, \dot{y} - \dot{\zeta}_y - \zeta_{x_1}) - \ddot{\zeta}_y - \dot{\zeta}_{x_1} - \zeta_{x_2} \tag{42}$$

Thus, the functional parametrisation associated with the differential flatness property is

$$x_1 = y - \zeta_y \tag{43a}$$

$$x_2 = \dot{y} - \dot{\zeta}_y - \zeta_{x_1} \tag{43b}$$

$$u = \ddot{y} - f(y - \zeta_y, \dot{y} - \dot{\zeta}_y - \zeta_{x_1}) - \ddot{\zeta}_y - \dot{\zeta}_{x_1} - \zeta_{x_2} \tag{43c}$$

We see in the last equation the inverse mapping from the sensory output $y(y)$ to the control input $u(t)$. Note that this relation embeds all the influences of fluctuations on the action, since it is a functional in $y, \zeta_{x_1}, \zeta_{x_2}$, and ζ_y .

In the case of a deterministic system,

$$\dot{x}_1 = x_2 \tag{44a}$$

$$\dot{x}_2 = f(x_1, x_2) + u \tag{44b}$$

$$y = x_1 \tag{44c}$$

The parametrisation becomes

$$x_1 = y \tag{45a}$$

$$x_2 = \dot{y} \tag{45b}$$

$$u = \ddot{y} - f(y, \dot{y}) \tag{45c}$$

Note that (44) is dynamically equivalent to (45).

Then, minimising the free energy \mathcal{F} enforces the dynamical mode (assuming it is possible to set all terms of this free energy to zero, at least approximately), i.e., the dynamical model obtained without fluctuations:

$$\dot{\mu}_{x_1} = \mu_{x_2} \tag{46a}$$

$$\dot{\mu}_{x_2} = f(\mu_{x_1}, \mu_{x_2}) + u \tag{46b}$$

$$y = \mu_{x_1} \tag{46c}$$

where μ_* , $* \in \{x_1, x_2, u\}$, is the mode of $*$ under the recognition density. By virtue of the previous equivalence, (46) is dynamically equivalent to

$$\mu_{x_1} = y \tag{47a}$$

$$\mu_{x_2} = \dot{\mu}_y \tag{47b}$$

$$u = \ddot{y} - f(y, \dot{y}) \tag{47c}$$

We now see that the minimisation of free energy enforces an (at least approximate) inverse mapping from the sensed output to the action. There is a further relationship that we will return to later, predicated upon generalised coordinates of motion, which we have not addressed here and which relates to the final equation above. As the action influences the rate of change of the states, which then influence the sensory data, the gradient of current sensory data with respect to the action is zero. This can be seen explicitly by noting that there is no term in the free energy in which sensory data and the mode of the action jointly appear. However, as highlighted here, there can be a non-zero

gradient associated with temporal derivatives of the sensory data (i.e., higher orders of generalised coordinates of motion). This is essential for the reflexive formulation of action under active inference.

4. References, Flatness-Based Trajectory Tracking, and Perceptual and Active Inferences

4.1. Equivalence to Linearity

4.1.1. Differential Flatness Characterisation

The class of differentially flat systems is—despite the fact that it occurs quite frequently in practice—the simplest nonlinear class with respect to the feedback equivalence classes. Indeed, we have the following.

Proposition 2. *A system is flat if, and only if, it is linearisable by endogenous feedback and a change of coordinates.*

A dynamic feedback is called endogenous if it does not include any external dynamics. More precisely, the following holds.

Definition 9. *Consider the dynamics $\dot{x} = f(x, u) + \zeta$. The feedback*

$$u = \varphi(x, \xi, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_u)}, v) \tag{48a}$$

$$\dot{\xi} = \psi(x, \xi, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_v)}, v) \tag{48b}$$

(where v is the new input) is called a dynamic endogenous feedback if the original dynamics $\dot{x} = f(x, u) + \zeta$ is equivalent to the transformed one

$$\dot{x} = f(x, \varphi(x, \xi, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_u)}, v)) \tag{49a}$$

$$\dot{\xi} = \psi(x, \xi, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_v)}, v) \tag{49b}$$

Two systems are called equivalent if there exists an invertible transformation that exchanges their trajectories.

Crucially the preceding linearisation is not local but global, and the category of flat systems is not very far from the one of linear systems, since they are equivalent through endogenous feedback and coordinate change.

4.1.2. Dynamical Extension Algorithm

This procedure enables one to determine if an m -tuple $\omega = (\omega_1, \dots, \omega_m)$ is a flat output or not and to obtain a linearizing feedback.

Phase I—Weak Brunovský Index Gathering

- (1) Differentiate ω_1 until a combination of control inputs appears; denote by κ_1 the number of successive differentiations:

$$\omega_1^{(\kappa_1)} = \gamma_1(\omega, \dot{\omega}, \dots, \omega^{(\kappa_1-1)}, u, \dot{u}, \dots, u^{(\sigma_{u1})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_{\zeta 1})})$$

with $\sigma_{u1} \leq \kappa_1 - 1$ and $\sigma_{\zeta 1} \leq \kappa_1$.

- (2) Differentiate ω_2 until a combination of control inputs (independent of the previous ones) appears; denote by κ_2 the number of successive differentiations:

$$\omega_2^{(\kappa_2)} = \gamma_2(\omega, \dot{\omega}, \dots, \omega^{(\kappa_2-1)}, u, \dot{u}, \dots, u^{(\sigma_{u2})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_{\zeta 2})})$$

with $\sigma_{u2} \leq \kappa_2 - 1$ and $\sigma_{\zeta2} \leq \kappa_2$.

⋮

- (m) Differentiate ω_m until a combination of control inputs (independent of the previous ones) appears; denote by κ_m the number of successive differentiations:

$$\omega_m^{(\kappa_m)} = \gamma_m(\omega, \dot{\omega}, \dots, \omega^{(\kappa_m-1)}, u, \dot{u}, \dots, u^{(\sigma_{um})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_{\zeta m})})$$

with $\sigma_{um} \leq \kappa_m - 1$ and $\sigma_{\zeta m} \leq \kappa_m$.

Phase II—Flatness Character Determination

Then, if $\kappa_1 + \dots + \kappa_m = n$ (n being the state dimension), the system admits ω as a flat output. If not, ω is not a flat output.

Phase III—Linearizing Feedback

In case ω is a flat output, the linearizing feedback is given by

$$\gamma_1(\omega, \dot{\omega}, \dots, \omega^{(\kappa_1-1)}, u, \dot{u}, \dots, u^{(\sigma_{u1})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_{\zeta1})}) = v_1 \tag{50a}$$

$$\gamma_2(\omega, \dot{\omega}, \dots, \omega^{(\kappa_2-1)}, u, \dot{u}, \dots, u^{(\sigma_{u2})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_{\zeta2})}) = v_2 \tag{50b}$$

⋮

$$\gamma_m(\omega, \dot{\omega}, \dots, \omega^{(\kappa_m-1)}, u, \dot{u}, \dots, u^{(\sigma_{um})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_{\zeta m})}) = v_m \tag{50c}$$

where v_1, \dots, v_m are the new action (new control input) variables. Since in Phase I, the γ_i s ($i = 1, \dots, m$) are functionally independent, the latter Equations (50) is invertible in u_1, \dots, u_m :

$$u_1 = \tilde{\gamma}_1(\omega, \dot{\omega}, \dots, \omega^{(\sigma_{\kappa_1})}, v, \dot{v}, \dots, v^{(\sigma_{v1})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\tilde{\sigma}_{\zeta1})}) \tag{51a}$$

$$u_2 = \tilde{\gamma}_2(\omega, \dot{\omega}, \dots, \omega^{(\sigma_{\kappa_2})}, v, \dot{v}, \dots, v^{(\sigma_{v2})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\tilde{\sigma}_{\zeta2})}) \tag{51b}$$

⋮

$$u_m = \tilde{\gamma}_m(\omega, \dot{\omega}, \dots, \omega^{(\sigma_{\kappa_m})}, v, \dot{v}, \dots, v^{(\sigma_{vm})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\tilde{\sigma}_{\zeta m})}) \tag{51c}$$

with $v = (v_1, \dots, v_m)$.

The original dynamics $\dot{x} = f(x, u) + \zeta$ is then transformed, via the linearising endogenous feedback (50), to a linear dynamics of the form

$$\dot{\omega}_1^{(\kappa_1)} = v_1 \tag{52a}$$

⋮

$$\dot{\omega}_m^{(\kappa_m)} = v_m \tag{52b}$$

with the new input $v = (v_1, \dots, v_m)$.

Thus, the agent’s model (36) is equivalent, through the feedback (50), to the so-called flat output dynamics:

$$\dot{\omega}_1^{(\kappa_1)} = \gamma_1(\omega, \dots, \omega^{(\kappa_1-1)}, u, \dots, u^{(\sigma_{u1})}, \zeta, \dots, \zeta^{(\sigma_{\zeta1})}) \tag{53a}$$

⋮

$$\dot{\omega}_m^{(\kappa_m)} = \gamma_m(\omega, \dots, \omega^{(\kappa_m-1)}, u, \dots, u^{(\sigma_{um})}, \zeta, \dots, \zeta^{(\sigma_{\zeta m})}) \tag{53b}$$

Thus, the original model (36) with n equations has been reduced, exactly, i.e., without any approximation, to the flat output dynamics (53) with m equations, where m is, in most practical cases, significantly smaller than n .

4.2. Differential Flatness and Controllability

A natural question a reader may ask is as follows: when is the differential flatness property verifiable? In other words, what are the checkable conditions ensuring flatness for a given system? The answer, for general nonlinear systems, is still an open problem. There are some conditions for restricted system classes or feedback equivalence classes. For instance, conditions are known for single-input systems and for static state feedback equivalence (see the Supplementary Material, File AIandFP-FlatnessAndSRR-HMTPKF-2026-SupplMaterial-v1.pdf, Section C).

There are some simple classes that are trivially flat. In the case of single-input systems, the class of systems that are, up to a feedback equivalence, in a cascade form like the following,

$$\dot{x}_1 = f_1(x_1, x_2) \tag{54a}$$

$$\dot{x}_2 = f_2(x_1, x_2, x_3) \tag{54b}$$

\vdots

$$\dot{x}_n = f_n(x_1, \dots, x_n, u) \tag{54c}$$

with the functions f_i being invertible in their last argument, are differentially flat, with x_1 as a flat output. The systems that are, up to a feedback equivalence, in a form with a finite number m of blocks like (54) are also differentially flat. We shall nevertheless see that differential flatness is a very strong property, as the following proposition suggests.

Proposition 3. *A differentially flat system is globally controllable.*

Proof. The global controllability property (see Definition 7) amounts to the following: for any initial and final states x_0 and x_1 of \mathbb{R}^n , and any initial and final times t_0 and t_1 , there exists a control law $u \in \mathcal{F}_u$, in the so-called *space of admissible controls* \mathcal{F}_u , driving the model from the initial state $x(t_0) = x_0$ to the final state $x(t_1) = x_1$. Consider now a flat system with equations

$$\dot{x} = f(x, u)$$

$$y = h(x)$$

with flat output ω . We thus have, by virtue of the flatness property, the following functional parametrisation:

$$x = A(\omega, \dot{\omega}, \dots, \omega^{(\rho_x)})$$

$$u = B(\omega, \dot{\omega}, \dots, \omega^{(\rho_u)})$$

Consider two arbitrary states x_0 and x_1 in \mathbb{R}^n , and $t_0, t_1 > 0$. Let $\omega_r(t)$ be a polynomial such that

$$x(t_0) = A(\omega_r(t_0), \dot{\omega}_r(t_0), \dots, \omega_r^{(\rho_x)}(t_0)) = x_0$$

$$x(t_1) = A(\omega_r(t_1), \dot{\omega}_r(t_1), \dots, \omega_r^{(\rho_x)}(t_1)) = x_1$$

The control steering the system from ω_0 to ω_1 is

$$u_r = B(\omega_r, \dot{\omega}_r, \dots, \omega_r^{(\rho_u)})$$

Thus the system is globally controllable. □

Example 4 (Oculomotor dynamics and flatness). As a simple example, consider a model of oculomotor control, a problem that has been addressed extensively through active inference: see, e.g., [53–57]. Here, we will consider a simplified kinematics and dynamics of eye movement. Consider a human agent whose eyes are located at a fixed distance d from a vertical visual scene. Let us concentrate of the movement of a single eye and denote the pupil’s center by E . The eye tracks a point P with coordinates x, y in the visual scene’s fixed reference frame. The agent’s head is supposed to be fixed. Let P_0 , with coordinates x_0, y_0 , be the orthogonal projection of E on the visual scene plane. We shall first describe the generative model and then demonstrate that this is differentially flat. We then consider how it might be linearised as outlined above.

According to Listing’s law [58], two angles characterise the movement of the pupil: ψ , the orientation, or yaw angle, and ϕ , the elevation, or pitch angle. Let us note that this model is extremely similar to the one of a two-degree-of-freedom gimbal system (see, e.g., [59]). We then have

$$x = x_0 + d \tan \psi + \zeta_x \tag{55a}$$

$$y = y_0 + d \tan \phi + \zeta_y \tag{55b}$$

Let us set

$$v_\psi = \dot{\psi}, \quad v_\phi = \dot{\phi}$$

The differentiation of the expressions of x and y in (55) with respect to time yields (See Figure 2)

$$\dot{x} = dv_\psi(1 + \tan^2 \psi) + \dot{\zeta}_x \tag{56a}$$

$$\dot{y} = dv_\phi(1 + \tan^2 \phi) + \dot{\zeta}_y \tag{56b}$$

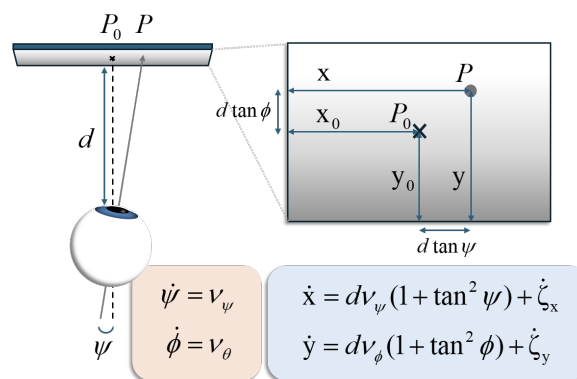


Figure 2. This graphic is intended to illustrate the set-up of our oculomotion example. On the left, we see the set-up from above, with an eye looking towards a screen at distance d . The pair of equations in the lower left panel give the (linear) flows for the yaw and pitch of the eyeball. The graphic on the right shows the screen with the fixation point P relative to the orthogonal projection P_0 . The equations in the lower right panel show the corresponding (stochastic) motion of the fixation point. The expression in terms of fixation points provides a useful nonlinear system in which we can unpack the differential flatness concept. It also provides a clear example of a control problem, as we might expect the target of an eye movement to be the fixation location that helps us resolve uncertainty about something in visual space, as opposed to a desired oculomotor angle.

Then, using the expressions of x and y in (55), we get expressions of $1 + \tan^2 \psi$ and $1 + \tan^2 \phi$ as functions of x and y :

$$d(1 + \tan^2 \psi) = \frac{d^2 + (x - x_0 - \zeta_x)^2}{d} \quad d(1 + \tan^2 \phi) = \frac{d^2 + (y - y_0 - \zeta_y)^2}{d}$$

Thus, the relations expressing \dot{x} and \dot{y} in (56) become

$$\dot{x} = v_\psi \frac{d^2 + (x - x_0 - \zeta_x)^2}{d} + \dot{\zeta}_x \tag{57a}$$

$$\dot{y} = v_\phi \frac{d^2 + (y - y_0 - \zeta_y)^2}{d} + \dot{\zeta}_y \tag{57b}$$

The relationship between v_ψ, v_ϕ and the torques exerted by the eye muscles can be written as follows:

$$2I_e \dot{v}_\psi = u_\psi \tag{58a}$$

$$I_e \dot{v}_\phi = u_\phi \tag{58b}$$

with I_e being the rotational inertia of the eye, supposed to be spherical (see the Supplementary Material, Section A, File AIandFP-FlatnessAndSRR-HMTPKF-2026-SupplMaterial-v1.pdf, for the reason why the inertia is doubled in the yaw equation above). Let us set

$$\mathbf{x} = (x_x, x_y, x_\psi, x_\phi), \quad x_x = X, \quad x_y = Y, \quad x_{v_\psi} = \psi, \quad x_{v_\phi} = \phi$$

The full generative model will then be written as follows:

$$\dot{x}_x = x_{v_\psi} \frac{d^2 + (x_x - X_0 - \zeta_x)^2}{d} + \dot{\zeta}_x \tag{59a}$$

$$\dot{x}_y = x_{v_\phi} \frac{d^2 + (x_y - Y_0 - \zeta_y)^2}{d} + \dot{\zeta}_y \tag{59b}$$

$$2I_e \dot{x}_{v_\psi} = u_\psi + \zeta_\psi \tag{59c}$$

$$I_e \dot{x}_{v_\phi} = u_\phi + \zeta_\phi \tag{59d}$$

where $\zeta_*(t), * \in \{x, y, \psi, \phi\}$, are fluctuations, which may be due to various alterations in human vision and the eye inertia $I_e = 2m_e r_e^2 / 5$, with m_e and r_e being the eye's mass and radius.

This model is differentially flat, with $\omega = (x, y)$ as a flat output. Indeed, one has the following:

1. From the first two lines of (59), we get the expressions of x_{v_ψ} and x_{v_ϕ} :

$$x_{v_\psi} = \frac{d(\dot{x}_x - \dot{\zeta}_x)}{d^2 + (x_x - X_0 - \zeta_x)^2} \tag{60a}$$

$$x_{v_\phi} = \frac{d(\dot{x}_y - \dot{\zeta}_y)}{d^2 + (x_y - Y_0 - \zeta_y)^2} \tag{60b}$$

2. From the final two lines of (59), we get the expressions of the action, the control inputs u_ψ and u_ϕ :

$$u_\psi = 2I_e \frac{d(\ddot{x}_x - \ddot{\zeta}_x)(d^2 + (x - x_0 - \zeta_x)^2) - 2d(x - x_0 - \zeta_x)(\dot{x}_x - \dot{\zeta}_x)^2}{(d^2 + (x - x_0 - \zeta_x)^2)^2} - \dot{\zeta}_\psi \tag{61a}$$

$$u_\phi = I_e \frac{d(\ddot{x}_y - \ddot{\zeta}_y)(d^2 + (y - y_0 - \zeta_y)^2) - 2d(y - y_0 - \zeta_y)(\dot{x}_y - \dot{\zeta}_y)^2}{(d^2 + (y - y_0 - \zeta_y)^2)^2} - \dot{\zeta}_\phi \tag{61b}$$

Note that the previous parametrisation is a pathwise one and that it carries over all of the fluctuations. Let us consider the reference trajectories $x_{xr}(t)$ and $x_{yr}(t)$ for x and y , depicted in Figure 3. More precisely, the plotted trajectory has the following form:

$$x_{xr}(t) = \frac{x_h - x_l}{2} (1 + \tanh(\gamma(t - t_{raise}))) + x_l, \quad t \in [0, T], \tag{62}$$

$$T = 80 \text{ s}, \quad x_l = 1, \quad x_h = 4, \quad \gamma = 4, \quad t_{raise} = 20 \text{ s}$$

where the simulation is performed on $[0, 80s]$, γ is the stiffness of the trajectory, and t_{raise} is the time where x_{x_r} has reached half of the maximum. The minimum is $\lim_{t \rightarrow -\infty} x_{x_r}(t) = x_l = 1$; the maximum is $\lim_{t \rightarrow +\infty} x_{x_r}(t) = x_h = 4$.

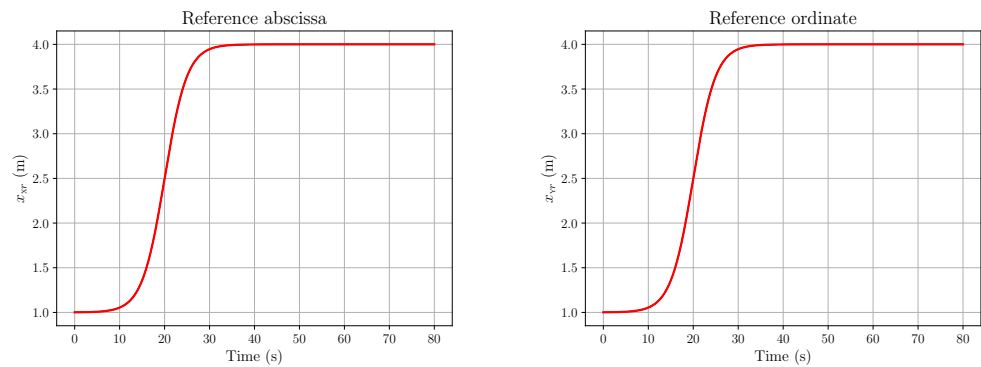


Figure 3. Simple reference trajectory examples $x_{x_r}(t), x_{y_r}(t)$.

The open-loop components of the states v_{ψ_r} and v_{ϕ_r} , corresponding to Equation (60), are shown in Figure 4, together with the influence of the fluctuations ζ_x, ζ_y and their first derivatives. These fluctuations were chosen to be

$$\zeta_* = \alpha \zeta \quad * \in \{x, y, \psi, \phi\} \tag{63}$$

with ζ being of the form (16), $L = 8, \lambda = 2$, and $\alpha = 2 \times 10^{-6}$. Recall the form (16):

$$\zeta(t) = a_0 + \sum_{j=1}^r \left[a_j \cos\left(\frac{2\pi j}{L} t\right) + b_j \sin\left(\frac{2\pi j}{L} t\right) \right], \quad r = \lfloor L/\lambda \rfloor \tag{64}$$

where $\lambda, L > 0$, each a_j and b_j is an independent sample from $\mathcal{N}(0, 1/(2r + 1))$ (where $\mathcal{N}(\mu, V)$ denotes the real normal distribution of mean μ and variance V), and $\lfloor \cdot \rfloor$ is the floor function. Since $L = 8$ and $\lambda = 2$, we have $r = 4$.

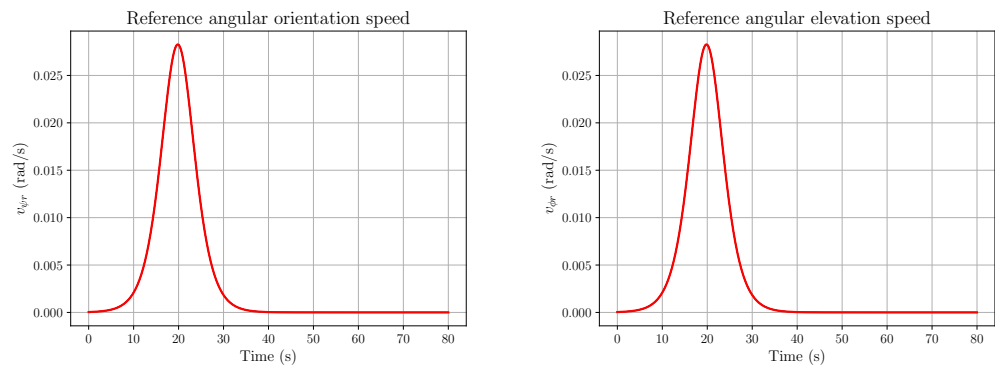


Figure 4. Orientation and elevation speeds v_{ψ_r} and v_{ϕ_r} for very small fluctuations $x_{x_r}(t), x_{y_r}(t)$.

The corresponding fluctuations ζ_x and ζ_y are depicted in Figure 5.

The open-loop action components v_{ψ_r} and v_{ϕ_r} , corresponding to Equation (61), are described in Figure 6. Although the fluctuations in Figure 5 are extremely small, their influence is already noticeable, although quite modest, in the u_{ϕ_r} plot.

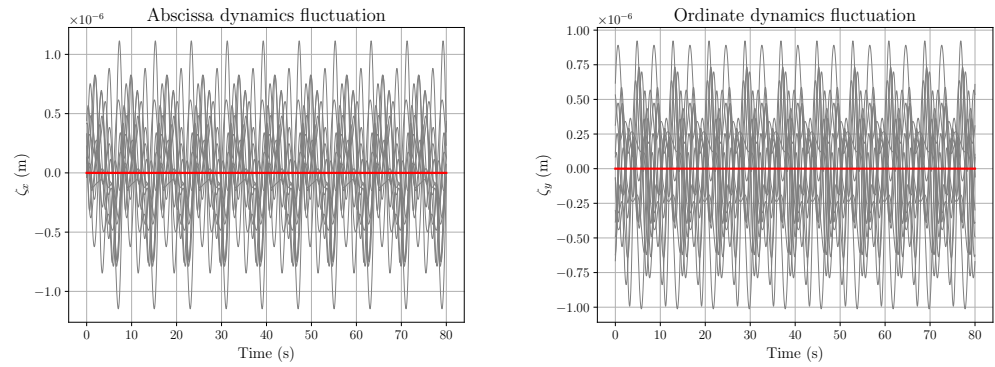


Figure 5. Fluctuations from Equation (63), with an extremely small scale ($L = 8, \lambda = 2, \alpha = 2 \cdot 10^{-6}$). Red line is the mean; grey lines are the signals with fluctuations.

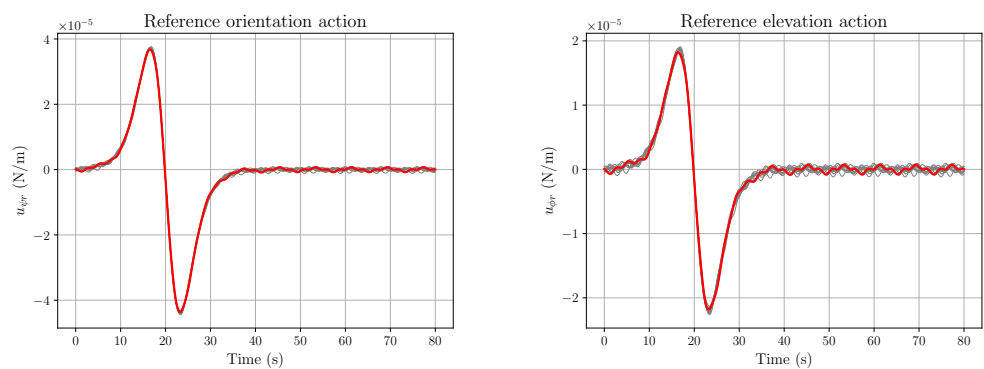


Figure 6. Orientation and elevation torques $u_{\psi r}(t), u_{\phi r}(t)$ for extremely small fluctuations ($L = 8, \lambda = 2, \alpha = 2 \times 10^{-6}$). Red lines are the means, i.e., without fluctuations; grey lines are the signals with fluctuations.

If one considers $\alpha = 10^{-4}$, which forces ζ_x and ζ_y to be still very small, the corresponding plots of $v_{\psi r}$ and $v_{\phi r}$ change radically, as depicted in Figure 7.

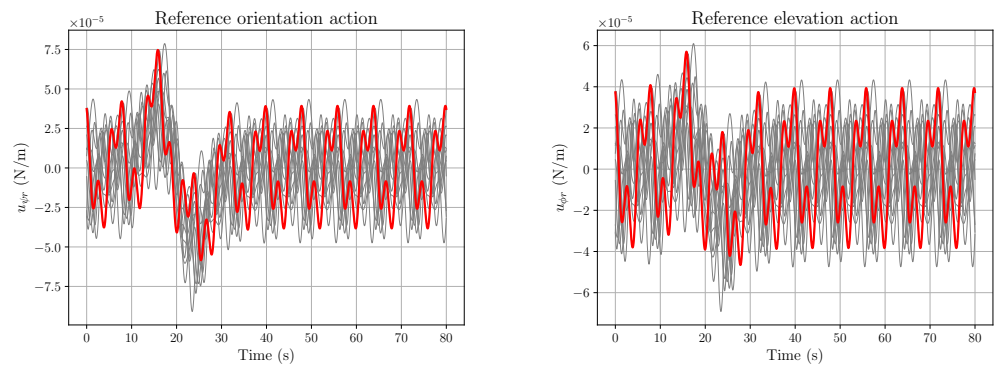


Figure 7. Orientation and elevation torques $u_{\psi r}(t), u_{\phi r}(t)$ for very small fluctuations ($L = 8, \lambda = 2, \alpha = 10^{-4}$). Red lines are the means, i.e., without fluctuations; grey lines are the signals with fluctuations.

Consider again Example 4 and its generative model:

$$\dot{x}_X = x_{v_\psi} \frac{d^2 + (x_X - x_0 - \zeta_X)^2}{d} + \dot{\zeta}_X \tag{65a}$$

$$\dot{x}_Y = x_{v_\phi} \frac{d^2 + (x_Y - Y_0 - \zeta_Y)^2}{d} + \dot{\zeta}_Y \tag{65b}$$

$$2I_e \dot{x}_{v_\psi} = u_\psi + \zeta_\psi \tag{65c}$$

$$I_e \dot{x}_{v_\phi} = u_\phi + \zeta_\phi \tag{65d}$$

To exactly linearise the model (65), one first differentiates (65c) and (65d), so that the action (control input) u_ψ, u_ϕ can appear:

$$\ddot{x}_X = \frac{1}{2dI_e} \left(d^2 + (x_X - x_0 - \zeta_X)^2 \right) (u_\psi + \zeta_\psi) + \frac{2x_{v_\psi}}{d} (\dot{x}_X - \dot{\zeta}_X) + \ddot{\zeta}_X \tag{66a}$$

$$\ddot{x}_Y = \frac{1}{dI_e} \left(d^2 + (x_Y - Y_0 - \zeta_Y)^2 \right) (u_\phi + \zeta_\phi) + \frac{2v_\phi}{d} (\dot{x}_Y - \dot{\zeta}_Y) + \ddot{\zeta}_Y \tag{66b}$$

or, in matrix form,

$$\begin{pmatrix} \ddot{x}_X \\ \ddot{x}_Y \end{pmatrix} = \frac{1}{2dI_e} \begin{pmatrix} \delta_X & 0 \\ 0 & 2\delta_Y \end{pmatrix} \begin{pmatrix} u_\psi + \zeta_\psi \\ u_\phi + \zeta_\phi \end{pmatrix} + \frac{2}{d} \begin{pmatrix} x_{v_\psi}(\dot{x}_X - \dot{\zeta}_X) \\ x_{v_\phi}(\dot{x}_Y - \dot{\zeta}_Y) \end{pmatrix} + \begin{pmatrix} \ddot{\zeta}_X \\ \ddot{\zeta}_Y \end{pmatrix} \tag{67}$$

$$\delta_X = d^2 + (x - x_0 - \zeta_X)^2 \quad \delta_Y = d^2 + (Y - Y_0 - \zeta_Y)^2 \tag{68}$$

To exactly linearise the dynamics (65), we shall equate $(\ddot{x}_X, \ddot{x}_Y)^T$ to $(v_X, v_Y)^T$, where $v_X(t), v_Y(t)$ are new action functions. Hence, we obtain

$$\frac{1}{2dI_e} \begin{pmatrix} \delta_X & 0 \\ 0 & 2\delta_Y \end{pmatrix} \begin{pmatrix} u_\psi + \zeta_\psi \\ u_\phi + \zeta_\phi \end{pmatrix} + \begin{pmatrix} \frac{2x_{v_\psi}(\dot{x}_X - \dot{\zeta}_X)}{d} + \ddot{\zeta}_X \\ \frac{2x_{v_\phi}(\dot{x}_Y - \dot{\zeta}_Y)}{d} + \ddot{\zeta}_Y \end{pmatrix} = \begin{pmatrix} v_X \\ v_Y \end{pmatrix} \tag{69}$$

Then, the nonlinear dynamics (65) is transformed into

$$\begin{aligned} \ddot{x}_X &= v_X \\ \ddot{x}_Y &= v_Y \end{aligned}$$

4.3. Trajectory Design and Planning

In order to perform tracking of a predefined trajectory, one has first to design this trajectory, the goal of the subsequent action. In some cases, this trajectory will be obvious to design, e.g., for an eye to move along a line, or in a circular motion, or for an arm to grab an object when no obstacle is present. In other circumstances, the task of planning a trajectory may be highly complex, especially in cases where both the agent and some obstacles are moving. The literature on planning is huge, and it has been studied extensively in fields like robotics, where it plays a crucial role (see, e.g., [60]).

4.4. Synthesis Law Computations: Tracking Controller

4.4.1. General Action Tracking Law

There are numerous ways to achieve trajectory tracking with stability, i.e., ensure that the discrepancy in action ϵ_{act} tends to zero asymptotically, and the corresponding literature is huge (see, e.g., [41] for a classic reference). For our purposes, i.e., the fulfilment of the inference guidelines, all of the laws described in [41] are inappropriate. This is

because they do not rely on pathwise properties, and more importantly they do not rely on a basis-like property. In contrast, the flatness-based framework is inherently pathwise while embedding the physics of the agent’s model.

Considering a differentially flat model $\dot{x} = f(x, u)$, we here wish to derive an action law (a controller) able to follow any reference trajectory $t \mapsto \omega_r(t)$. In order to compensate for model mismatch and poorly known initial conditions, one has to complement the open loop (obtained through flatness) with a closed-loop corrective term depending on the error $\omega(t) - \omega_r(t)$.

Knowing that the dynamics is flat, with flat output ω , it can be transformed via the linearising endogenous feedback (50) to a linear dynamics of the form

$$\omega_1^{(\kappa_1)} = v_1 \tag{70a}$$

⋮

$$\omega_m^{(\kappa_m)} = v_m \tag{70b}$$

with the new input $v = (v_1, \dots, v_m)$. Then, the elementary tracking feedback law

$$\begin{aligned} v_i &= \omega_{ir}^{(\kappa_i)} - \sum_{j=0}^{\kappa_i-1} \lambda_{\omega ij} (\omega_i^{(j)} - \omega_{ir}^{(j)}), & i = 1, \dots, m \\ &= \omega_{ir}^{(\kappa_i)} - \sum_{j=0}^{\kappa_i-1} \lambda_{\omega ij} e_{\omega i}^{(j)}, & \text{with } e_{\omega i} = \omega_i - \omega_{ir} \end{aligned} \tag{71}$$

with appropriately chosen λ_{zij} gains renders the error dynamics

$$e_{\omega i}^{(\kappa_i)} + \sum_{j=0}^{\kappa_i-1} \lambda_{\omega ij} e_{\omega i}^{(j)} = 0, \quad i = 1, \dots, m \tag{72}$$

asymptotically stable. In vector form, the preceding law (71) is expressed as follows:

$$v_i = \omega_r^{(\kappa_i)} - \lambda_{\omega i} \left(e_{\omega i}^{(\kappa_i-1)} \right)^T \tag{73}$$

where $e_{\omega i}^{(\kappa_i-1)} = (e_{\omega i}, \dot{e}_{\omega i}, \dots, e_{\omega i}^{(\kappa_i-1)})$ and $\lambda_{\omega i} = (\lambda_{0i}, \lambda_{1i}, \dots, \lambda_{(\kappa_i-1)i})$. Finally, the tracking action (tracking feedback control) law is expressed as follows:

$$u_1 = \tilde{\gamma}_1(\omega, v, \dot{v}, \dots, v^{(\sigma_{v1})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\tilde{\sigma}_{\zeta 1})}) \tag{74a}$$

$$u_2 = \tilde{\gamma}_2(\omega, v, \dot{v}, \dots, v^{(\sigma_{v2})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\tilde{\sigma}_{\zeta 2})}) \tag{74b}$$

⋮

$$u_m = \tilde{\gamma}_m(\omega, v, \dot{v}, \dots, v^{(\sigma_{vm})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\tilde{\sigma}_{\zeta m})}) \tag{74c}$$

$$v = (v_1, \dots, v_m) \tag{74d}$$

$$v_i = \omega_r^{(\kappa_i)} - \lambda_{zi} \left(e_{zi}^{(\kappa_i-1)} \right)^T, \quad i = 1, \dots, m \tag{74e}$$

which ensures the tracking of ω to ω_r , with stability through driving the discrepancy $\varepsilon_{act} = \omega - \omega_r$ to zero.

Remark 7 (Open-loop and model-free). *The preceding action law is named a linearising feedback controller due to the fact that the flat output dynamics is exactly linearised in a first step. Another possibility—most probably more fruitful—is to use one of the following control laws. A first possible choice is an open-loop controller, i.e., an action law obtained through the use of (29b), p. 13, in*

which ω is substituted with ω_r , a reference trajectory, with this open-loop law being supplemented by a model-free controller, in the spirit of [61]. Other possible choices include (74) supplemented by a model-free controller, an ADRC (active disturbance rejection control) one [62,63], or a sliding mode control one [64,65].

Example 5 (Simple generic example; tracking). Consider again the following simple example with scalar action (control) and sensor output:

$$\dot{x}_1 = x_2 + \zeta_{x_1} \tag{75a}$$

$$\dot{x}_2 = f(x_1, x_2) + u + \zeta_{x_2} \tag{75b}$$

$$y = x_1 + \zeta_y \tag{75c}$$

In order to elaborate on the tracking action laws, we have to obtain the sensor output y dynamics, which is obtained by differentiating (75a), thus obtaining \ddot{x}_1 :

$$\ddot{x}_1 = f(x_1, \dot{x}_1) + u + \dot{\zeta}_{x_1} + \zeta_{x_2} \tag{76}$$

and using (75c):

$$\ddot{y} = f(y - \zeta_y, \dot{y} - \dot{\zeta}_y) + u + \ddot{\zeta}_y + \dot{\zeta}_{x_1} + \zeta_{x_2} \tag{77}$$

Setting $\epsilon_y = y - y_r$, we have $y = y_r + \epsilon_y$. Then, the y -dynamics (77) becomes

$$\ddot{\epsilon}_y = -\ddot{y}_r + f(y - \zeta_y, \dot{y} - \dot{\zeta}_y) + u + \ddot{\zeta}_y + \dot{\zeta}_{x_1} + \zeta_{x_2} \tag{78}$$

The trajectory tracking goal is

$$\lim_{t \rightarrow \infty} \epsilon_y = \lim_{t \rightarrow \infty} (y - y_r) = 0 \tag{79}$$

When expressing this goal in a dynamical setting, this renders ϵ_y the solution of a target differential equation:

$$\ddot{\epsilon}_y + \lambda_{y,1}\dot{\epsilon}_y + \lambda_{y,2}\epsilon_y = 0 \tag{80}$$

with $\lambda_{y,1}, \lambda_{y,2} \in \mathbb{R}$ such that all solutions of (80) are exponentially decreasing functions of time.

In order to transform (78) into (80), we set

$$-\ddot{y}_r + f(y - \zeta_y, \dot{y} - \dot{\zeta}_y) + u + \ddot{\zeta}_y + \dot{\zeta}_{x_1} + \zeta_{x_2} = -\lambda_{y,1}\dot{\epsilon}_y - \lambda_{y,2}\epsilon_y \tag{81}$$

Hence, the control law ensuring the tracking of $y_r(t)$ with stability is given by

$$u = \ddot{y}_r - f(y - \zeta_y, \dot{y} - \dot{\zeta}_y) - \ddot{\zeta}_y - \dot{\zeta}_{x_1} - \zeta_{x_2} - \lambda_{y,1}\dot{\epsilon}_y - \lambda_{y,2}\epsilon_y \tag{82}$$

This control law assumes the full knowledge of the fluctuations ζ_{x_1} , ζ_{x_2} , and ζ_y . When this knowledge is not available, we have to derive a control law based on a deterministic generative model (which is all that the agent knows), estimate the fluctuations, and compensate for them (see, e.g., [66] for the so-called model-free control or [62,65] for other schemes).

4.4.2. Oculomotor Example

Example 6 (Oculomotor tracking). Consider again Example 4 and its generative model:

$$\dot{x}_X = x_{v_\psi} \frac{d^2 + (x_X - x_0 - \zeta_X)^2}{d} + \dot{\zeta}_X \tag{83a}$$

$$\dot{x}_Y = x_{v_\phi} \frac{d^2 + (x_Y - y_0 - \zeta_Y)^2}{d} + \dot{\zeta}_Y \tag{83b}$$

$$2I_e \dot{x}_{v_\psi} = u_\psi + \zeta_\psi \tag{83c}$$

$$I_e \dot{x}_{v_\phi} = u_\phi + \zeta_\phi \tag{83d}$$

and suppose that one wishes to track a reference trajectory $(x_{Xr}(t), x_{Yr}(t))$ for (X, Y) . We have seen in Example 4 that the following action feedback:

$$\frac{1}{2dI_e} \begin{pmatrix} \delta_X & 0 \\ 0 & 2\delta_Y \end{pmatrix} \begin{pmatrix} u_\psi + \zeta_\psi \\ u_\phi + \zeta_\phi \end{pmatrix} + \begin{pmatrix} \frac{2x_{v_\psi}(\dot{x}_X - \dot{\zeta}_X)}{d} + \ddot{\zeta}_X \\ \frac{2x_{v_\phi}(\dot{x}_Y - \dot{\zeta}_Y)}{d} + \ddot{\zeta}_Y \end{pmatrix} = \begin{pmatrix} v_X \\ v_Y \end{pmatrix} \tag{84}$$

$$\delta_X = d^2 + (x - x_0 - \zeta_X)^2 \quad \delta_Y = d^2 + (y - y_0 - \zeta_Y)^2 \tag{85}$$

transforms the nonlinear dynamics (83) into the following linear dynamics:

$$\ddot{x}_X = v_X$$

$$\ddot{x}_Y = v_Y$$

The trajectory tracking can then be enforced through the following tracking feedback law in v_X, v_Y :

$$v_X = \ddot{x}_{Xr} - \lambda_{X1}\dot{\epsilon}_X - \lambda_{X0}\epsilon_X \tag{86a}$$

$$v_Y = \ddot{x}_{Yr} - \lambda_{Y1}\dot{\epsilon}_Y - \lambda_{Y0}\epsilon_Y \tag{86b}$$

$$\epsilon_X = x_X - x_{Xr} \quad \epsilon_Y = x_Y - x_{Yr} \tag{86c}$$

where $\lambda_{X0}, \lambda_{X1}, \lambda_{Y0},$ and λ_{Y1} are suitable constants (the action or feedback control gains) ensuring the stability of the closed-loop error dynamics:

$$\ddot{\epsilon}_{act} + \lambda_1 \dot{\epsilon}_{act} + \lambda_0 \epsilon_{act} = 0$$

$$\epsilon_{act} = \begin{pmatrix} \epsilon_X \\ \epsilon_Y \end{pmatrix} \quad \lambda_0 = \begin{pmatrix} \lambda_{X0} \\ \lambda_{Y0} \end{pmatrix} \quad \lambda_1 = \begin{pmatrix} \lambda_{X1} \\ \lambda_{Y1} \end{pmatrix}$$

Here, it is sufficient to take all $\lambda_* > 0, * \in \{X0, Y0, X1, Y1\}$. Note that the preceding laws v_X and v_Y in (86) can be replaced by other ones, such as, for instance, model-free control ones (see, e.g., [66]). Finally, to express the original control law $u = (u_\psi, u_\phi)$, we must invert (84), yielding u_ψ and u_ϕ as functions of v_X and v_Y :

$$\begin{pmatrix} u_\psi \\ u_\phi \end{pmatrix} = \frac{dI_e}{\delta_X \delta_Y} \begin{pmatrix} 2\delta_Y & 0 \\ 0 & \delta_X \end{pmatrix} \left[- \begin{pmatrix} \frac{2x_{v_\psi}(\dot{x}_X - \dot{\zeta}_X)}{d} + \ddot{\zeta}_X \\ \frac{2x_{v_\phi}(\dot{x}_Y - \dot{\zeta}_Y)}{d} + \ddot{\zeta}_Y \end{pmatrix} + \begin{pmatrix} v_X \\ v_Y \end{pmatrix} \right] - \begin{pmatrix} \zeta_\psi \\ \zeta_\phi \end{pmatrix} \tag{87}$$

Then substituting (86a) into (87) yields the final action tracking controller:

$$\begin{pmatrix} u_\psi \\ u_\phi \end{pmatrix} = \frac{dI_e}{\delta_x \delta_y} \begin{pmatrix} 2\delta_y & 0 \\ 0 & \delta_x \end{pmatrix} \left[- \begin{pmatrix} \frac{2x_{v\phi}(\dot{x}_x - \dot{\zeta}_x)}{d} + \ddot{\zeta}_x \\ \frac{2x_{v\phi}(\dot{x}_y - \dot{\zeta}_y)}{d} + \ddot{\zeta}_y \end{pmatrix} + \begin{pmatrix} \ddot{x}_{xr} - \lambda_{x1}\dot{\epsilon}_x - \lambda_{x0}\epsilon_x \\ \ddot{x}_{yr} - \lambda_{y1}\dot{\epsilon}_y - \lambda_{y0}\epsilon_y \end{pmatrix} \right] - \begin{pmatrix} \zeta_\psi \\ \zeta_\phi \end{pmatrix} \quad (88)$$

which ensures the tracking of (x, y) to (x_{xr}, x_{yr}) with stability, i.e., the fulfilment of (IG-Perfo) through driving the discrepancy $\epsilon_{act} = (\epsilon_x, \epsilon_y)^T$ to zero. The quantity sensed by the agent is the position x, y of the point P in the visual scene.

Remark 8 (A more realistic example). Note that we could have considered a more realistic example than (83) by including, as in [53], a delay τ_s in sensing and an advance τ_a in acting. In order to achieve predictions, the agent uses a form of open-loop control. A possible choice could be the use of reference signals, using the functional parametrisation offered by the flatness property. In other words, using (61) after substitution of (x, y) with (x_{xr}, x_{yr}) , we get the open-loop action law:

$$u_{\psi r} = \frac{1}{I_e} \frac{d\ddot{x}_{xr}(d^2 + (x_{xr} - x_0)^2) - 2d(x_{xr} - x_0)\dot{x}_{xr}^2}{(d^2 + (x_{xr} - x_0)^2)^2} \quad (89a)$$

$$u_{\phi r} = \frac{1}{I_e} \frac{d\ddot{x}_{yr}(d^2 + (x_{yr} - y_0)^2) - 2d(x_{yr} - y_0)\dot{x}_{yr}^2}{(d^2 + (x_{yr} - y_0)^2)^2} \quad (89b)$$

Another (heavier) choice would be to simulate the generative model to get a prediction $(X(t + \tau_s + \tau_a), Y(t + \tau_s + \tau_a))$ for $(X(t), Y(t))$.

4.4.3. Simulations

The following plots refer to the simulation of the oculomotor model in (83), with the following parameters:

$$I_e = 4.032 \times 10^{-7} \text{kg.m}^2, d = 10 \text{ m}, x_0 = 0 \text{ m}, y_0 = 0 \text{ m} \quad (90)$$

The fluctuations $\zeta_x, \zeta_y, \zeta_\psi, \zeta_\phi$ are all smooth random functions (i.e., sums of cosine and sine with coefficients that are random variables with a normal distribution). More precisely, they are of the form depicted in Equation (16):

$$\zeta_*(t) = \alpha \left(a_0 + \sum_{j=1}^r \left[a_{*,j} \cos\left(\frac{2\pi j}{L} t\right) + b_{*,j} \sin\left(\frac{2\pi j}{L} t\right) \right] \right), \quad r = \lfloor L/\lambda \rfloor \quad (91)$$

$$* \in \{x, y, \psi, \phi\} \quad (92)$$

where $\lambda, L > 0$, each $a_{*,j}$ and $b_{*,j}$ is an independent sample from $\mathcal{N}(0, 1/(2r + 1))$ (where $\mathcal{N}(\mu, V)$ denotes the real normal distribution of mean μ and variance V), and $\lfloor \cdot \rfloor$ is the floor function. Here, $L = 8, \lambda = 2, \alpha = 10^{-5}$; hence $r = 4$. Let us consider a smooth eye tracking of a point along a regular curve, here a quatrefoil, shown in Figure 8 with the reference curve in red and the actual tracking (i.e., the eye movement) in blue. The trajectory type has been chosen because of its smoothness. Its precise time parametrisation is

$$x(t) = 2a \sin^2(t) \cos(t) \quad (93a)$$

$$y(t) = 2a \cos^2(t) \sin(t) \quad (93b)$$

$$a = 2 \quad (93c)$$

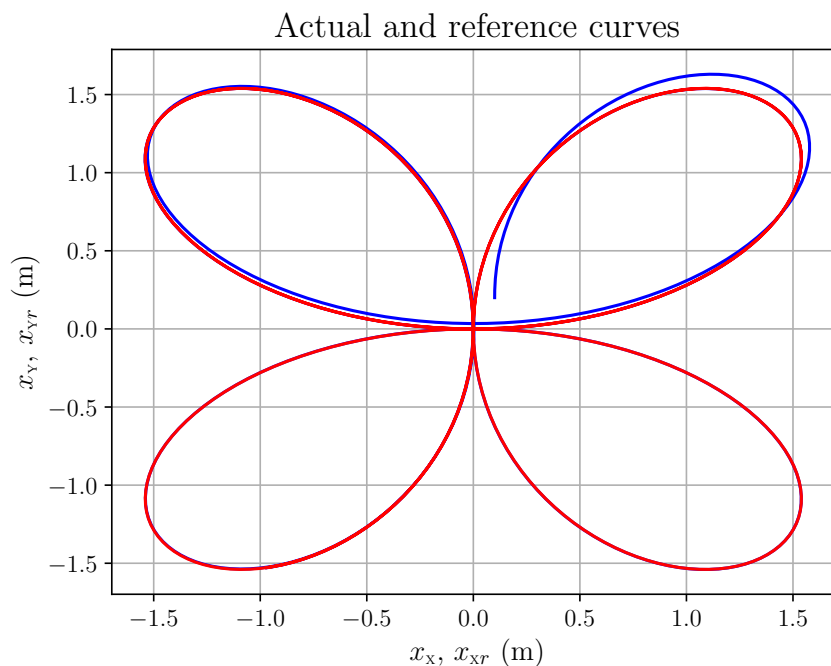


Figure 8. Reference quatrefoil trajectory in red and actual movement from the generative model in blue.

The corresponding references $x_{x_r}(t)$, $x_{y_r}(t)$ and the actual values $x_x(t)$, $x_y(t)$ are plotted in Figure 9.

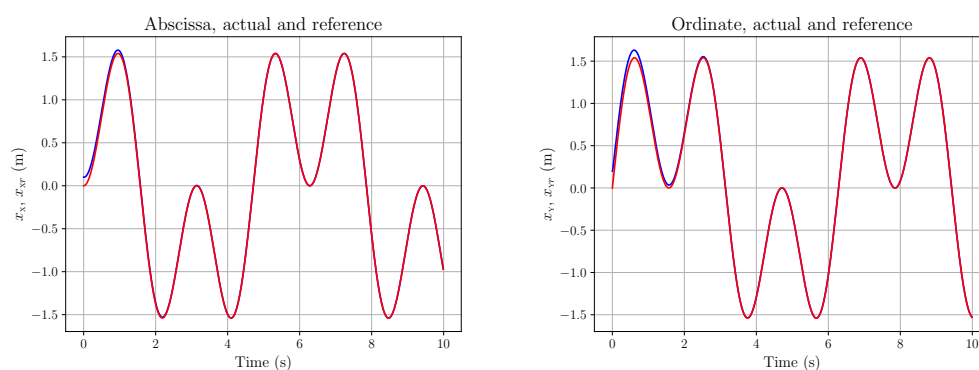


Figure 9. Plots of $x_{x_r}(t)$, $x_{y_r}(t)$ in red and of $x_x(t)$, $x_y(t)$ in blue for the quatrefoil.

The corresponding action tracking laws u_ψ and u_ϕ are shown in Figure 10. As a last example, let us consider the hypocycloid in Figure 11. We chose this type of trajectory because it may be seen as typical of saccadic eye movements. Its precise time parametrisation is

$$X(t) = (a - b) \cos(t) + b \cos\left(\frac{(a - b)t}{b}\right) \tag{94a}$$

$$Y(t) = (a - b) \sin(t) - b \sin\left(\frac{(a - b)t}{b}\right) \tag{94b}$$

$$a = 2, b = \frac{6}{5} \tag{94c}$$

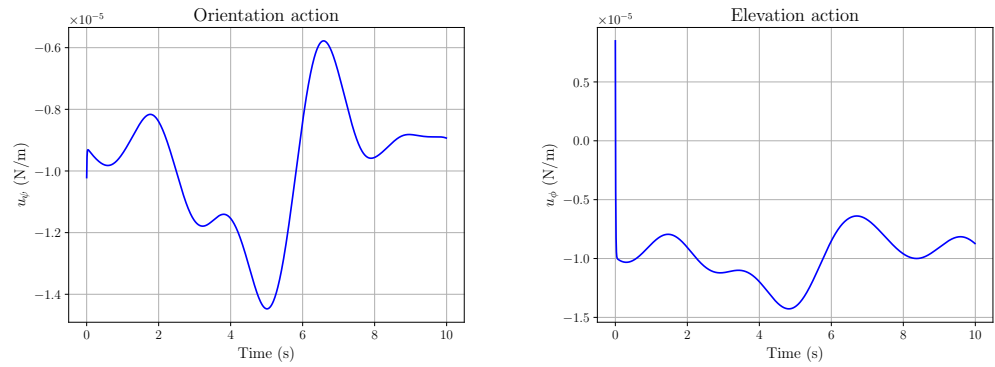


Figure 10. Plots of the actions u_ψ and u_ϕ for the quatrefoil.

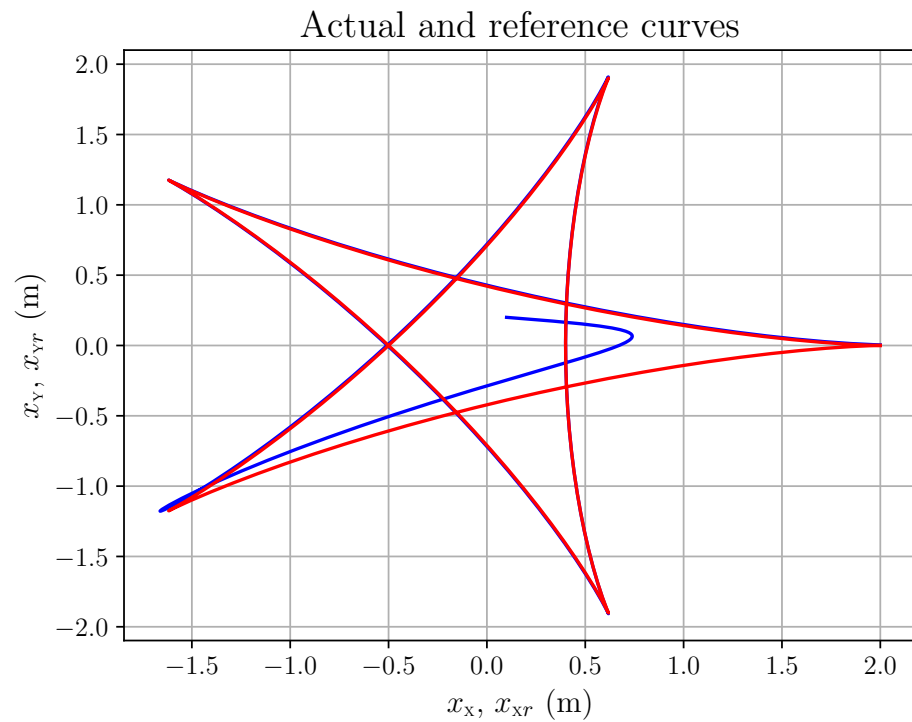


Figure 11. Reference hypocycloid trajectory in red and actual movement from the generative model in blue.

4.5. Active Inference

The principle of active inference can then be stated as extremisation of free energy via both action and perception, under a prior belief that action will extremise expected free energy (see, e.g., Figure 7 of [2]):

$$\mu_x^* = \operatorname{argmin}_{\mu_x} \mathcal{F}(\mu_x, \mu_u, \mathbf{y}(\mathbf{u})) \tag{95a}$$

$$\mathbf{u}^* = \operatorname{argmin}_{\mathbf{u}} \mathcal{F}(\mu_x^*, \mu_u, \mathbf{y}(\mathbf{u})) \tag{95b}$$

$$\mathcal{F}(\mathbf{x}, \mathbf{y}) = \mathbb{E}_{q,x} \left(\ln \frac{q(\mathbf{x}|\mathbf{u})}{p(\mathbf{x}|\mathbf{u})} \right) - \mathbb{E}_{q,x} (\ln p(\mathbf{y}|\mathbf{x}, \mathbf{u})) \tag{95c}$$

$$\mu_u = \operatorname{argmin}_u \mathcal{G}(\mathbf{u}) \tag{95d}$$

$$\mathcal{G}(\mathbf{u}) = \mathbb{E}_{q,y} \left(\ln \frac{q(\mathbf{y}|\mathbf{u})}{p(\mathbf{y}|\mathbf{y}_r)} \right) - \mathbb{E}_{q,x,y} (\ln p(\mathbf{y}|\mathbf{x}, \mathbf{u})) \tag{95e}$$

with μ_x being the agent’s estimate of the hidden state x , and \mathcal{F}, \mathcal{G} as in (21c) (complexity/accuracy) and (22b) (risk/ambiguity), written in a compact form.

$$\mathbb{E}_{q,\xi}(\phi) = \int q(\xi(t) | \mathbf{u}(t)) \phi(t) d\xi(t)$$

$$\mathbb{E}_{q,\xi,\chi}(\phi) = \iint q(\xi(t), \chi(t) | \mathbf{u}(t)) \phi(t) d\xi(t) d\chi(t)$$

The above extremisations can be formulated as the solutions to a gradient descent.

4.6. Link with Flatness-Based Tracking

The imperative for active inference is the minimisation of surprise, namely, the discrepancy between expectations, or beliefs, and actual values. In the light of the previous decomposition of variational free energy in Section 3.1, we conclude by foregrounding the links between active inference and nonlinear control. Consider the agent’s dynamics $\dot{x} = f(x, \mathbf{u}) + \zeta_x$. The dynamic feedback (see (48))

$$\mathbf{u} = \varphi(x, \xi, \zeta_x, \dot{\zeta}_x, \dots, \zeta_x^{(\sigma_u)}, \mathbf{v}) \tag{96a}$$

$$\dot{\xi} = \psi(x, \xi, \zeta_x, \dot{\zeta}_x, \dots, \zeta_x^{(\sigma_\xi)}, \mathbf{v}) \tag{96b}$$

$$\mathbf{v} = \Gamma(\boldsymbol{\varepsilon}_{act}, \dot{\boldsymbol{\varepsilon}}_{act}, \dots, \boldsymbol{\varepsilon}_{act}^{(\sigma_v)}) \tag{96c}$$

(where \mathbf{v} is the new input) transforms the agent’s dynamics to

$$\dot{x} = f(x, \varphi(x, \xi, \zeta_x, \dot{\zeta}_x, \dots, \zeta_x^{(\sigma_u)}, \mathbf{v})) \tag{97a}$$

$$\dot{\xi} = \psi(x, \xi, \zeta_x, \dot{\zeta}_x, \dots, \zeta_x^{(\sigma_\xi)}, \mathbf{v}) \tag{97b}$$

$$\mathbf{v} = \Gamma(\boldsymbol{\varepsilon}_{act}, \dot{\boldsymbol{\varepsilon}}_{act}, \dots, \boldsymbol{\varepsilon}_{act}^{(\sigma_v)}) \tag{97c}$$

The above Equation (97) then yield the following action surprise dynamics, or action discrepancy dynamics:

$$\Xi(\boldsymbol{\varepsilon}_{act}, \dot{\boldsymbol{\varepsilon}}_{act}, \dots, \boldsymbol{\varepsilon}_{act}^{(\sigma_{\boldsymbol{\varepsilon}_{act}})}) = 0 \tag{98}$$

This surprise dynamics will not, in general (i.e., for any feedback law (96)), imply the action discrepancy’s limit to be zero, i.e., a minimisation of the free energy. Let us now consider the tracking action (tracking feedback control) law stemming from (74):

$$u_1 = \tilde{\gamma}_1(\boldsymbol{\omega}, \mathbf{v}, \dot{\mathbf{v}}, \dots, \mathbf{v}^{(\sigma_{v1})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\tilde{\sigma}_{\zeta1})}) \tag{99a}$$

$$u_2 = \tilde{\gamma}_2(\boldsymbol{\omega}, \mathbf{v}, \dot{\mathbf{v}}, \dots, \mathbf{v}^{(\sigma_{v2})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\tilde{\sigma}_{\zeta2})}) \tag{99b}$$

⋮

$$u_m = \tilde{\gamma}_m(\boldsymbol{\omega}, \mathbf{v}, \dot{\mathbf{v}}, \dots, \mathbf{v}^{(\sigma_{vm})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\tilde{\sigma}_{\zeta m})}) \tag{99c}$$

$$\mathbf{v} = (v_1, \dots, v_m) \tag{99d}$$

$$v_i = \boldsymbol{\omega}_r^{(\kappa_i)} - \lambda_{zi} \left(\mathbf{e}_{zi}^{(\kappa_i-1)} \right)^T, \quad i = 1, \dots, m \tag{99e}$$

which yields a special form of (98):

$$\boldsymbol{\varepsilon}_{act}^{(n)} - \Lambda(0, \boldsymbol{\varepsilon}_{act}, \dot{\boldsymbol{\varepsilon}}_{act}, \dots, \boldsymbol{\varepsilon}_{act}^{(n-1)}) = 0 \tag{100}$$

where Λ is a linear function of its arguments such that (100) admits solely exponentially stable solutions. The tracking action (tracking feedback control) law (99) thus ensures the

tracking of ω to ω_r , with stability through driving the discrepancy $\epsilon_{act} = \omega - \omega_r$ to zero, minimising the risk in the expected free energy \mathcal{G} .

Example 7 (Oculomotor tracking control law). Consider again the oculomotor Example 4. The tracking action (tracking feedback control) law (99) takes the form of (88):

$$\begin{pmatrix} u_\psi \\ u_\phi \end{pmatrix} = \frac{dI_e}{\delta_x \delta_y} \begin{pmatrix} 2\delta_y & 0 \\ 0 & \delta_x \end{pmatrix} \left[- \begin{pmatrix} \frac{2x_{v\psi}(\dot{x}_x - \dot{\zeta}_x)}{d} + \ddot{\zeta}_x \\ \frac{2x_{v\phi}(\dot{x}_y - \dot{\zeta}_y)}{d} + \ddot{\zeta}_y \end{pmatrix} + \begin{pmatrix} \ddot{x}_{xr} - \lambda_{x1}\dot{\epsilon}_x - \lambda_{x0}\epsilon_x \\ \ddot{x}_{yr} - \lambda_{y1}\dot{\epsilon}_y - \lambda_{y0}\epsilon_y \end{pmatrix} \right] - \begin{pmatrix} \zeta_\psi \\ \zeta_\phi \end{pmatrix} \quad (101)$$

5. Prediction as a Link Between Active Inference and Differential Flatness

5.1. Delays and δ -Flatness

In real systems stemming from neuroscience or physiology, delays are present, both in sensing and acting. This has not yet been taken into account in the previous sections, although it remains of utmost importance. Indeed, there are fundamental differences between a delay-free model and one including delays; the most striking difference is the infinite-dimensional character of delay systems. To be more precise, consider a delay-differential equation of the following form:

$$x(t) = f(x(t - \tau)) \quad (102)$$

If we want to integrate such an equation, we do not need to know a pointwise initial condition such as $x(0)$ but have to know a whole function on a time interval $x : [-\tau, 0) \rightarrow \mathbb{R}$. Another view on this is that there is not just one operator acting on the variables—time differentiation—but two: time differentiation and the time delay. Therefore, the structural properties (controllability and observability) of such systems become more complex, even for linear delay systems (see, e.g., [67]). The notion of flatness has been extended to the case of delay systems (see [68,69]). We shall informally recall the definition of the so-called δ -flatness (a special case of the π -flatness where the delays do not only occur on the sensor outputs and the action variables but also on the hidden state) and then expose two different kinds of predictors, which are needed for tracking. We denote a delay operator of amplitude τ by δ_τ :

$$(\delta_\tau x)(t) = x(t - \tau) \quad (103)$$

In informal terms, a δ -flat system is a system of delay differential equations that is differentially flat when one allows for delays. Let us see how this notion unfolds in a concrete example.

Example 8. Consider Example 2 with a delay in the action:

$$\dot{x}_1(t) = x_2(t) + \zeta_{x_1}(t) \quad (104a)$$

$$\dot{x}_2(t) = f(x_1(t), x_2(t)) + u(t - \tau) + \zeta_{x_2}(t) \quad (104b)$$

$$y = x_1(t) + \zeta_y(t) \quad (104c)$$

with $\tau > 0, \tau \in \mathbb{R}$. This model is easily seen to be δ -flat, with y as a δ -flat output. Indeed, we have

$$x_1(t) = y(t) - \zeta_y(t) \tag{105a}$$

$$x_2(t) = \dot{y}(t) - \dot{\zeta}_y(t) - \zeta_{x_1}(t) \tag{105b}$$

$$u(t) = \ddot{y}(t + \tau) - f(y(t + \tau) - \zeta_y(t + \tau), \dot{y}(t + \tau) - \dot{\zeta}_y(t + \tau) - \zeta_{x_1}(t + \tau)) - \ddot{\zeta}_y(t + \tau) - \dot{\zeta}_{x_1}(t + \tau) - \zeta_{x_2}(t + \tau) \tag{105c}$$

5.2. Trajectory Tracking and Predictors

When one wishes to perform trajectory tracking, as in (99), the presence of delays in the sensor output and/or in the action control will induce the necessity to predict some or all of the hidden states or sensor output. Consider this under the previous example.

Example 9. The tracking control scheme (82) is now transformed into

$$u(t) = \ddot{y}_r(t + \tau) - f(y(t) - z_y(t), \dot{y}(t) - \dot{z}_y(t)) - \ddot{z}_y(t) - \dot{z}_{x_1}(t) - z_{x_1}(t) - \lambda_{y,1}\dot{e}_y(t) - \lambda_{y,2}e_y(t) \tag{106a}$$

$$y(t) = y(t + \tau), \quad e_y(t) = e_y(t + \tau) \tag{106b}$$

$$z_y(t) = \zeta_y(t + \tau), \quad z_{x_1}(t) = \zeta_{x_1}(t + \tau), \quad z_{x_2}(t) = \zeta_{x_2}(t + \tau) \tag{106c}$$

We now review two kinds of predictors, which may be used in tracking control schemes. Consider the following type of mean generative model with delayed action and a linear relation between the sensory output and hidden state:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t - \tau)) \tag{107a}$$

$$\mathbf{y}(t) = C\mathbf{x}(t) \tag{107b}$$

with $C \in \mathbb{R}^{n \times p}$. This model can be transformed into the advanced generative model

$$\dot{\hat{\mathbf{x}}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)) \tag{108a}$$

$$\mathbf{y}(t) = C\mathbf{x}(t) \tag{108b}$$

$$\mathbf{x}(t) = \mathbf{x}(t + \tau), \quad \mathbf{y}(t) = \mathbf{y}(t + \tau) \tag{108c}$$

The first scheme is a so-called delayed observer (see [70]):

$$\dot{\hat{\mathbf{x}}}(t) = f(\hat{\mathbf{x}}(t), \mathbf{u}(t)) + L(\mathbf{x}(t) - \hat{\mathbf{x}}(t - \tau)) \tag{109}$$

with $L \in \mathbb{R}^{n \times n}$ being the predictor gain matrix. The simulation of Equation (109) yields an estimate $\hat{\mathbf{x}}(t)$ of $\mathbf{x}(t + \tau)$. It is proven [70] that the so-called prediction observer error $\mathbf{x}(t) - \hat{\mathbf{x}}(t - \tau)$ tends to zero as t tends to infinity.

The second scheme is an integral form one (see [71–73]):

$$\mathbf{u}(t) = \kappa(\mathbf{x}(t)) \tag{110a}$$

$$\mathbf{x}(t) = \int_{t-\tau}^t f(\mathbf{x}(\zeta), \mathbf{u}(\zeta))d\zeta + \mathbf{x}(t) \tag{110b}$$

where κ is such that $\dot{\mathbf{x}} = f(\mathbf{x}, \kappa(\mathbf{x}))$ is globally asymptotically stable at $\mathbf{x} = 0$; i.e., for every trajectory $\mathbf{x}(t)$, we have $\mathbf{x}(t) \rightarrow 0$ as $t \rightarrow \infty$. Equation (110b) comes from the following simple observation:

$$\int_t^{t+\tau} \dot{z}(\zeta)d\zeta = z(t + \tau) - z(t) \tag{111}$$

The initial condition for the integral Equation (110b) for $\mathbf{x}(t)$ is defined by

$$\mathbf{x}(\sigma) = \int_{-\tau}^{\sigma} f(\mathbf{x}(\xi), \mathbf{u}(\xi)) d\xi + \mathbf{x}(0), \quad \sigma \in [-\tau, 0] \tag{112}$$

The predictor state $\mathbf{x}(t)$ is given by the implicit relation (110b), which can be solved using various approximation strategies for the integral on the right-hand side. Then, the predictive control law of the simple example shall be implemented as follows.

Example 10. *The tracking control scheme (106a) is now transformed into*

$$u(t) = \ddot{y}_r(t + \tau) - f(y(t) - z_y(t), \dot{y}(t) - \dot{z}_y(t)) - \ddot{z}_y(t) - \dot{z}_{x_1}(t) - z_{x_1}(t) - \lambda_{y,1} \dot{e}_y(t) - \lambda_{y,2} e_y(t) \tag{113a}$$

$$y(t) = \int_{t-\tau}^t f(y(\xi), u(\xi)) d\xi + y(t), \quad e_y(t) = e_y(t + \tau) \tag{113b}$$

$$z_y(t) = \zeta_y(t + \tau), \quad z_{x_1}(t) = \zeta_{x_1}(t + \tau), \quad z_{x_2}(t) = \zeta_{x_2}(t + \tau) \tag{113c}$$

5.3. Generalised Coordinates

In the active inference framework, there is a crucial distinction between what is called motion of the mean compared to mean of the motion. To unpack this distinction we first need the notion of generalised coordinates, which is obtained through a linearised differentiation. Generally a differentiation operation ∂ in algebra is defined as one such that, for any variables (here time functions) ω_1 and ω_2 , the chain rule is fulfilled:

$$\partial(\omega_1 \omega_2) = \partial \omega_1 \omega_2 + \omega_1 \partial \omega_2 \tag{114}$$

A linearised (or first-order) differentiation, d_1 , here denoted as $'$, is an operator following the above chain rule such that, for any time functions $z(t)$ and $\zeta(t)$,

$$z^i \zeta^{(j)} = 0 \tag{115}$$

whenever $j > 0, i + j > 1$ and denoting by $\zeta^{(j)}$ the j th iterated application of the $'$ operator to ζ . This type of differentiation is linked to first-order stochastic realisation problems, as briefly discussed below.

Higher-order differentiations can also be considered. A k -th-order differentiation d_k is an operator following the above chain rule such that, for any time functions $z(t)$ and $\zeta(t)$,

$$z^i d_k^j \zeta^{(j)} = 0 \tag{116}$$

whenever $j > 0, i + j > k$. This type of differentiation is linked to higher-order stochastic realisation problems, although these may induce quite involved computations.

In other words, suppose the solution $\zeta(t)$ of

$$\dot{\zeta}(t) = F(\zeta(t)) + \eta(t) \tag{117}$$

can be expressed through a series expansion:

$$\zeta(t) = \sum_{i=0}^{\infty} \frac{\zeta^{(i)}(0)}{i!} t^i \tag{118}$$

e.g., it could be an analytic function of time or a Gevrey series (i.e., belonging to a class of functions in between the analytic and the full C^∞ one—see, e.g., [74,75]; see also [76] for extensions). Then, the various derivatives at the time origin can be recovered as follows:

$$\dot{\xi}(0) = F(\xi(0)) + \eta(0) \tag{119a}$$

$$\ddot{\xi}(0) = \frac{d}{dt} F(\xi(0)) + \dot{\eta}(0) \tag{119b}$$

$$\dddot{\xi}(0) = \frac{d^2}{dt^2} F(\xi(0)) + \ddot{\eta}(0) \tag{119c}$$

⋮

assuming the fluctuations $\eta(t)$ are sufficiently smooth and

$$\frac{d}{dt} F(\xi(0)) = \nabla F(\xi(0)) \dot{\xi}(0) \tag{120a}$$

$$\frac{d^2}{dt^2} F(\xi(0)) = \nabla F(\xi(0)) \ddot{\xi}(0) + \dot{\xi}^T(0) \nabla^2 F(\xi(0)) \dot{\xi}(0) \tag{120b}$$

⋮

The so-called local linear approximation, in terms of [32], amounts to neglecting all higher-order terms in Equation (120), i.e., considering

$$\frac{d^k}{dt^k} F(\xi(0)) = \nabla F(\xi(0)) \xi^{(k)}(0) \tag{121}$$

Such an approximation, when valid, enables one to solve the so-called stochastic realisation problem quite easily (recall that a realisation is a differential equation involving a hidden state obtained from an input/action–output/sensor differential equation—see, e.g., Definition 6; see, e.g., [77,78] on the stochastic realisation problem). This approximation is also justified when studying generalised Bayesian filtering under the Laplace approximation (see, e.g., [32], Subsections 3.3.4 and 3.3.5).

The generalised coordinates may be seen as a coordinate frame that moves with the current point. In this view, it is linked with one-form transformations one can apply to nonlinear systems (see, e.g., [79,80]); the latter are more general than the endogenous dynamical feedbacks used here. It is also linked with the Cartan moving frame method (see, e.g., [81]).

6. Conclusions, Limitations, and Future Directions

We have considered the utility of differential flatness through the lens of active inference (see, e.g., [1]). This utility has been detailed in terms of control as inference. Specifically, one might conclude that if the generative models that underwrite active inference or control as inference can be limited to the class of differentially flat models, we obtain an extremely efficient control-theoretic scheme. This work therefore enables a control-theoretic perspective on active inference as control as inference by focusing on action trajectories and the minimisation of various discrepancies (afforded by variational and expected free energy). From the perspective of active inference, this work is a primer on differential flatness and its particular relevance to the kinds of generative models one might consider and be committed to. Crucially, this paper is the first (provisional) attempt to consider expected free energy in the setting of continuous state-space models.

In addition to their roles in developing systems for control, one could consider other applications of the frameworks outlined in this paper. One area is the field of computational psychiatry (see, e.g., [82–84]), where one can develop generative models of decision-making tasks—solved with active inference—and use these to understand the computational mechanisms of psychopathology. Such models are often formulated in terms of the selection among discrete alternatives. However, as noted by our reviewers, differential flatness pertains to

continuous, differentiable state spaces. While it is true that many applications of computational psychiatry focus on discrete probabilities, there are important areas of psychiatry that depend upon continuous variables of the sort addressed in differential flatness accounts. These include the altered motor dynamics associated with catatonia in psychotic disorders (see, e.g., [85]) and altered smooth pursuit eye movements in schizophrenia (see, e.g., [86,87]).

There are several related notions that could not be addressed here due to the lack of space. These include Liouvillian aspects (see [88–90] related to automatic control and [91,92] for the examination of this property in the setting of hypothalamic–pituitary–adrenal axis models and Wilson–Cowan population networks), and robust tracking with model-free control (see [66]). Liouvillian aspects, in particular, offer an opportunity to extend the notion of flatness when a model is not differentially flat. Future work will consider these and other issues related to energy transmission and controllability (see, e.g., [93,94] for a formulation of the corresponding problem) and variational function transmission. These deal with characterising a model in terms of salient features and understanding how reparameterisations might transform such features.

We discuss the following points, emphasising limitations of the current treatment and sketching some future directions.

1. Generalised coordinates, where some limitations—due to its approximate character—may be avoided through smooth random realisations.
2. Extensions of flatness: Liouvillian characters to deal with cases where the model is not differentially flat.
3. Observers and algebraic estimators to estimate the hidden state from sensor output.
4. Robust control law synthesis to cope with uncertainty under the generative model, including fluctuations.
5. Constraint fulfilment, where constraints are imposed on the hidden state, the action, and their time derivatives.
6. Feature transmission: How a feature of interest; e.g., the energy (L^2 norm), the slope or curvature, etc., is transformed through functional parametrisation.

6.1. Generalised Coordinate Limitations

The generalised coordinates are not appropriate for deriving tracking feedback action in the sense being considered here. Indeed, the dynamical extension algorithm needs to make the action appear through time differentiation, which will not be the case unless the action was already present in the state-space dynamics. More precisely, for the components of the flat output whose Brunovksy index κ_i is strictly superior to 2, the generalised coordinates are not appropriate, since successive application of d_1 will not make the action appear. For the components whose index $\kappa_i = 1$ or 2, the generalised coordinates will be sufficient.

Let us illustrate this through three simple examples.

1. First consider the longitudinal motion of a car, used, for instance, in ACCs (Automatic Cruise Controllers):

$$M\dot{V}_x = -F(V_x) + \frac{1}{r} u \quad (122)$$

with M being the car's mass, $V_x(t)$ the longitudinal speed of the vehicle's centre of gravity, $-F(V_x)$ the force due to the wind and to friction of the tyres on the ground, r the mean wheel radius, and u the traction engine propulsion torque, taken as the action. This system is trivially flat, with flat output V_x , and the preceding equation is rewritten as

$$MV'_x = -F(V_x, \dot{V}_x) + \frac{1}{r} u$$

The action is already present in the dynamics equation, $\kappa = 1$, and there is no need to differentiate more. The generalised coordinate dynamics will lead to the same action tracking controller as the original one.

2. Second, consider the oculomotor example again. Then one supplementary differentiation of x_x (resp. x_y) will be sufficient. Both of these will involve $f(x, u)$ and its partial derivatives with respect to x and u in the flat output dynamics (see (69)):

$$\begin{pmatrix} x_x \\ x_y \end{pmatrix} = \frac{1}{2dI_e} \begin{pmatrix} \delta_x & 0 \\ 0 & 2\delta_y \end{pmatrix} \begin{pmatrix} u_\psi + \zeta_\psi \\ u_\phi + \zeta_\phi \end{pmatrix} + \begin{pmatrix} \frac{2x_{v_\psi}(\dot{x}_x - \dot{\zeta}_x)}{d} + \ddot{\zeta}_x \\ \frac{2x_{v_\phi}(\dot{x}_y - \dot{\zeta}_y)}{d} + \ddot{\zeta}_y \end{pmatrix} = \begin{pmatrix} v_x \\ v_y \end{pmatrix} \tag{123a}$$

$$\delta_x = d^2 + (x - x_0 - \zeta_x)^2 \quad \delta_y = d^2 + (y - y_0 - \zeta_y)^2 \tag{123b}$$

3. Third, consider one of the most popular model for diabetes, namely the Bergman minimal model (see, e.g., [95,96]):

$$\dot{G} = -k_1(G - G_b) - XG + D = f_G(G, X, D) \tag{124a}$$

$$\dot{X} = -k_2X + k_3(I - I_b) = f_X(X, I) \tag{124b}$$

$$\dot{I} = -k_4(I - I_b) + u = f_I(I, u) \tag{124c}$$

$$D(t) = \frac{1}{2} (1 + \tanh(\gamma(t - t_{bol})))Be^{-d(t-t_{bol})} \tag{124d}$$

where $G(t)$ is the concentration of blood glucose; $X(t)$ is the concentration of insulin in the tissue fluid; $I(t)$ is the concentration of insulin in the blood; G_b and I_b are the basal concentrations of glucose and insulin; and $D(t)$ is the glycaemic influence of a meal, seen as a fluctuation. In addition, k_1, k_2 , and k_3 are positive-valued parameters that control the rates of appearance and disappearance of glucose and insulin; t_{bol} is the meal intake time (bolus intake).

We wish to control the glucose concentration $G(t)$ in order to track a desired trajectory $G_r(t)$ known in advance. To do so, we apply the dynamical extension algorithm, and we have to differentiate Equation (124a) two times in order for the action u to appear. We get

$$\ddot{G} = -(k_1 + X)\dot{G} - G\dot{X} + \dot{D}$$

and

$$\ddot{G} = -2\dot{G}\dot{X} - (k_1 + X)\ddot{G} - G\ddot{X} + \ddot{D}$$

and the control will appear in \dot{I} , present in \ddot{X} . The generalised coordinates unfold as follows. The preceding model is rewritten, within the differentiation d_1 , as follows:

$$G' = -k_1(G - G_b) - XG + D = f_G(G, X, D) \tag{125a}$$

$$X' = -k_2X + k_3(I - I_b) = f_X(X, I) \tag{125b}$$

$$I' = -k_4(I - I_b) + u = f_I(I, u) \tag{125c}$$

$$D(t) = \frac{1}{2} (1 + \tanh(\gamma(t - t_{bol})))Be^{-d(t-t_{bol})} \tag{125d}$$

The successive application of d_1 yields

$$G' = \partial_G f_G G' + \partial_X f_G X' + \partial_D f_G D' = -(k_1 + X)G' - GX' + D'$$

and

$$G'' = -(k_1 + X)G'' - GX'' + D''$$

Thus, the term $-2X'G'$ is missing when compared to its counterpart with time differentiation. And, if one attempted to use this in order to produce an action tracking feedback, the resulting error dynamics would be of the form

$$\varepsilon_G''' = -\lambda_{G2}\varepsilon_G'' - \lambda_{G1}\varepsilon_G' - \lambda_{G0}\varepsilon_G - 2X'G' \tag{126}$$

with $\varepsilon_G = G - G_r$ and $G_r(t)$ being a reference glucose trajectory. Although the λ_{Gi} are such that the solutions of

$$\varepsilon_G''' + \lambda_{G2}\varepsilon_G'' + \lambda_{G1}\varepsilon_G' + \lambda_{G0}\varepsilon_G = 0$$

are exponentially decreasing, the solutions of (126) may not tend to zero, since it is steered by $-2X'G'$.

6.2. Smooth Random Realisation

The use of this type of differentiation is also linked to the stochastic realisation problem (see, e.g., [10], 4.(c).(i), p. 15), which may be quite complex in the general case. In contrast to this, the differential flatness property yields a weak Brunovsky canonical form through the flat output (see [97], Subsection 4.1 and Definition 4.3). This canonical form yields the so-called flat output dynamics, which readily gives a smooth random realisation, as the following proposition states.

Proposition 4 (Smooth random realisation). *Consider a differentially flat model, with flat output $\omega = (\omega_1, \dots, \omega_m)$. Then, there exists integers $\kappa_1, \dots, \kappa_m$ such that $(\omega_1, \dot{\omega}_1, \dots, \omega_1^{(\kappa_1-1)}, \omega_2, \dots, \omega_m^{(\kappa_m-1)})$ is a state. This state, through the flat output dynamics, yields a smooth random realisation of the model.*

Proof. The existence of the integers $\kappa_1, \dots, \kappa_m$ such that $(\omega_1, \dot{\omega}_1, \dots, \omega_1^{(\kappa_1-1)}, \omega_2, \dots, \omega_m^{(\kappa_m-1)})$ is a state is ensured by the dynamical extension algorithm. To see this, and to obtain the associated realisation, recall the so-called flat output dynamics (Equation (53) above), which can be written as the following state-space representation:

$$\dot{x}_1 = x_2 \tag{127a}$$

⋮

$$\dot{x}_{\kappa_1-1} = x_{\kappa_1} \tag{127b}$$

$$\dot{x}_{\kappa_1} = \gamma_1(\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(\sigma_{u1})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_{\zeta 1})}) \tag{127c}$$

$$\dot{x}_{\kappa_1+1} = x_{\kappa_1+2} \tag{127d}$$

⋮

$$\dot{x}_{\kappa_1+\kappa_2-1} = x_{\kappa_1+\kappa_2} \tag{127e}$$

$$\dot{x}_{\kappa_1+\kappa_2} = \gamma_2(\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(\sigma_{u1})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_{\zeta 2})}) \tag{127f}$$

⋮

$$\dot{x}_{n-1} = x_n \tag{127g}$$

$$\dot{x}_n = \gamma_m(\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(\sigma_{um})}, \zeta, \dot{\zeta}, \dots, \zeta^{(\sigma_{\zeta m})}) \tag{127h}$$

with $\mathbf{x} = (x_1, \dots, x_n)$, since $\kappa_1 + \dots + \kappa_m = n$, and

$$\begin{aligned} x_1 &= \omega_1, & x_2 &= \dot{\omega}_1, \dots, & x_{\kappa_1} &= \omega_1^{(\kappa_1-1)} \\ x_{\kappa_1+1} &= \omega_2, & x_{\kappa_1+2} &= \dot{\omega}_2, \dots, & x_{\kappa_1+\kappa_2} &= \omega_2^{(\kappa_2-1)} \\ & & & \vdots & & \\ x_{\kappa_1+\dots+\kappa_{m-1}+1} &= \omega_m, & \dots & & \dots & x_n = \omega_m^{(\kappa_m-1)}. \end{aligned}$$

Then \mathbf{x} is a state that yields a smooth random realisation of the original system, based on the differential flatness property. \square

The present framework can be seen as an adequate proposal for the future direction in Subsection 4.1, paragraph “Stochastic control via generalized coordinates” of [32]. It also embeds, in a rather simple fashion, non-stationary smooth random signals (see Remark 4.1.1 in [32]).

6.3. Other Future Directions

We now briefly consider future directions that could not be unpacked in this paper due to lack of space.

6.3.1. Extensions of Flatness: Liouvillian Characters

Although many practical system models are differentially flat, some, among specific classes, are not; this is especially the case for many biological and neuroscience population models. Fortunately, another analogous property is shared by a much wider class, the one of so-called Liouvillian systems. Liouvillian systems can be seen as an extension of flat systems [88,89,98]. The most striking property of the latter is that all state and control variables of the system can be directly expressed—without integration of any differential equation—in terms of the flat output and a finite number of its time derivatives. So-called Liouvillian systems share a similar property, but in order to derive the trajectories of a Liouvillian system, we also need integration of a few differential equations whose solutions are known analytically. It follows that flatness-based control approaches can be extended up to solving a finite number of differential equations.

6.3.2. Observers and Algebraic Estimators

We have here dealt with control law synthesis, but we did not touch on the equally important subject of observer or estimator synthesis, i.e., the procedures aimed at estimating the hidden state from sensor measurements. Two main paths are available: The first—of so-called observers—amounts to a simulation of the mean generative model, driven by the error $\mathbf{y} - g(\mathbf{x})$ between the sensor output and its observation model, a function of the state (see, e.g., [99,100]). The second, more direct, approach is to directly make use of the so-called constructive observability (see, e.g., [4,101]), where the state is a function of sensor output, action, and their derivatives: $\mathbf{x} = \psi_{\mathbf{x}}(\mathbf{y}, \mathbf{u}, \dot{\mathbf{y}}, \dot{\mathbf{u}}, \dots, \mathbf{y}^{(\rho_{\mathbf{y}})}, \mathbf{u}^{(\rho_{\mathbf{u}})})$. This last procedure requires the estimation of sensor output derivatives (see [6,102]).

6.3.3. Robust Control Law Synthesis

When the mean generative model is a crude approximation to the real system, the cumulative effect of the fluctuations needs to be compensated for in the control scheme. So-called robust control laws aim to fulfil a control goal, such as trajectory tracking, despite the fluctuations, or perturbations, and unmodelled dynamics (i.e., model mismatch). These can be achieved via the so-called model-free control scheme, where the cumulative effects of fluctuations are estimated online and compensated for on the fly (see, e.g., [66]). The recent HEOL scheme (see [103]) achieves the same goal through slightly different tech-

niques. While model-free control is often associated with flatness-based feedback tracking control laws synthesised on a nominal, or mean generative, model, HEOL one uses an open-loop flatness-based scheme and the tangent, or variational system, associated with the simplified flat system, i.e., the linearised system around a reference trajectory of the simplified flat system. Other instances include, among others, sliding mode control and active disturbance rejection control (see, e.g., [62,64,65]).

6.3.4. Constraint Fulfilment

In real-world applications, dynamical systems are always subject to constraints: on the state (for example the configuration space of a robot is not the whole space) and on the action (e.g., muscles have finite power). These can be handled in the present framework through optimisation-based planning of the flat output trajectories (see, e.g., [104–106]). A promising framework is the one of model-free predictive control (MFPC, see [107]), mixing the popular predictive control (see, e.g., [108,109]) and model-free control, cited above.

6.3.5. Feature Transmission

Another interesting characteristic—of the functional parametrisation associated with differential flatness—is one of feature transmission. This includes the transmission of geometric features: how the curvature of the flat output trajectory is related to the curvature of the action; in other words, deducing the action’s curvature as a function of the flat output curvature and its time derivatives from the relation yielding the action as a function of the flat output and its derivatives.

Supplementary Materials: The following supporting information can be downloaded at <https://github.com/hugues-mounier/AI-FP-differential-flatness-and-smooth-random-realisation/>.

Author Contributions: Methodology, H.M., T.P. and K.F.; Software, H.M.; Writing—original draft, H.M.; Writing—review & editing, T.P. and K.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Wellcome Trust grant number 226793/Z/22/Z. The Third author (TR) is supported by an NIHR Academic Clinical Fellowship [ref: ACF-2023-13-013].

Data Availability Statement: The data and code presented in this study are openly available in <https://github.com/hugues-mounier/AI-FP-differential-flatness-and-smooth-random-realisation/> (accessed on 29 October 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Friston, K. A free energy principle for a particular physics. *arXiv* **2019**, arXiv:1906.10184. [CrossRef]
2. Friston, K.; Da Costa, L.; Sajid, N.; Heins, C.; Ueltzhöffer, K.; Pavliotis, G.A.; Parr, T. The free energy principle made simpler but not too simple. *Phys. Rep.* **2023**, *1024*, 1–29. [CrossRef]
3. Ramstead, M.J.D.; Sakthivadivel, D.A.R.; Heins, C.; Koudahl, M.; Millidge, B.; Da Costa, L.; Klein, B.; Friston, K.J. On Bayesian mechanics: A physics of and by beliefs. *Interface Focus* **2023**, *13*, 20220029. [CrossRef] [PubMed]
4. Fliess, M.; Glad, S.T. An Algebraic Approach to Linear and Nonlinear Control. In *Essays on Control*; Birkhäuser: Boston, TX, USA, 1993; pp. 223–267. [CrossRef]
5. Fliess, M.; Lévine, J.; Martin, P.; Rouchon, P. Flatness and defect of non-linear systems: Introductory theory and examples. *Int. J. Control* **1995**, *61*, 1327–1361. [CrossRef]
6. Fliess, M.; Join, C.; Ramirez, H.S. Non-linear estimation is easy. *Int. J. Model. Identif. Control* **2008**, *4*, 12. [CrossRef]
7. Baltieri, M. Active Inference: Building a New Bridge Between Control Theory and Embodied Cognitive Science. PhD thesis, University of Sussex, Brighton, UK, 2019.
8. Baltieri, M.; Buckley, C. PID Control as a Process of Active Inference with Linear Generative Models. *Entropy* **2019**, *21*, 257. [CrossRef]

9. Lanillos, P.; Meo, C.; Pezzato, C.; Meera, A.A.; Baioumy, M.; Ohata, W.; Tschantz, A.; Millidge, B.; Wisse, M.; Buckley, C.L.; et al. Active Inference in Robotics and Artificial Agents: Survey and Challenges. *arXiv* **2021**, arXiv:2112.01871. [[CrossRef](#)]
10. Da Costa, L.; Friston, K.; Heins, C.; Pavliotis, G.A. Bayesian mechanics for stationary processes. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2021**, *477*, 20210518. [[CrossRef](#)]
11. Meera, A.A.; Wisse, M. A Brain Inspired Learning Algorithm for the Perception of a Quadrotor in Wind. *arXiv* **2021**, arXiv:2109.11971. [[CrossRef](#)]
12. Tschantz, A.; Barca, L.; Maisto, D.; Buckley, C.L.; Seth, A.K.; Pezzulo, G. Simulating homeostatic, allostatic and goal-directed forms of interoceptive control using Active Inference. *Biol. Psychol.* **2022**, *169*, 108266. [[CrossRef](#)]
13. Baioumy, M.; Pezzato, C.; Ferrari, R.; Hawes, N. Unbiased Active Inference for Classical Control. *arXiv* **2022**, arXiv:2207.13409. [[CrossRef](#)]
14. Bos, F.; Meera, A.A.; Benders, D.; Wisse, M. Free Energy Principle for State and Input Estimation of a Quadcopter Flying in Wind. In Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; IEEE: Piscataway, NJ, USA, 2022. [[CrossRef](#)]
15. Rudolph, J. *Flatness-Based Control—An Introduction*, 1st ed.; Shaker GmbH: Düren, Germany, 2021. [[CrossRef](#)]
16. Francine Diener, M.D. (Ed.) *Nonstandard Analysis in Practice*; Springer: Berlin/Heidelberg, Germany, 1995. [[CrossRef](#)]
17. Fliess, M. Probabilités et fluctuations quantiques. *Comptes Rendus. Mathématique* **2007**, *344*, 663–668. [[CrossRef](#)]
18. Fliess, M.; Join, C. Towards a new viewpoint on causality for time series. *ESAIM Proc. Surv.* **2015**, *49*, 37–52. [[CrossRef](#)]
19. Han, X.; Kloeden, P.E. *Random Ordinary Differential Equations and Their Numerical Solution*; Springer: Singapore, 2017. [[CrossRef](#)]
20. Friz, P.K.; Hairer, M. *A Course on Rough Paths: With an Introduction to Regularity Structures*; Springer International Publishing: Cham, Switzerland, 2020. [[CrossRef](#)]
21. Kahane, J.P. *Some Random Series of Functions Series*, 2nd ed.; Number 5 in Cambridge Studies in Advanced Mathematics; Cambridge University Press: Cambridge, UK, 1993.
22. Hill, M. Convergence of random Fourier series. In Proceedings of the University of Chicago 2012 Summer Program for Undergraduate. 2012. Available online: <https://math.uchicago.edu/~may/REU2012/REUPapers/Hill.pdf> (accessed on 29 October 2025).
23. Tzvetkov, N. Riemannian analogue of a Paley-Zygmund theorem. *Séminaire Équations aux Dérivées Partielles (2008–2009)*. pp. 1–14. Available online: <https://eudml.org/doc/11189> (accessed on 29 October 2025).
24. Marcus, M.B.; Pisier, G. *Random Fourier Series with Applications to Harmonic Analysis (AM-101)*; Princeton University Press: Princeton, NJ, USA, 1981.
25. Filip, S.; Javed, A.; Trefethen, L.N. Smooth Random Functions, Random ODEs, and Gaussian Processes. *SIAM Rev.* **2019**, *61*, 185–205. [[CrossRef](#)]
26. Robert J. Adler, J.E.T. *Random Fields and Geometry*; Springer: New York, NY, USA, 2007. [[CrossRef](#)]
27. Adler, R.J.; Taylor, J.E. *Topological Complexity of Smooth Random Functions: École d'Été de Probabilités de Saint-Flour XXXIX-2009*; Springer: Berlin/Heidelberg, Germany, 2011. [[CrossRef](#)]
28. Adler, R.J.; Taylor, J.E.; Worsley, K.J. *Applications of Random Fields and Geometry: Foundations and Case Studies*. Book in Progress. 2016. Available online: <https://cris.technion.ac.il/en/publications/applications-of-random-fields-and-geometry-foundations-and-case-s/> (accessed on 29 October 2025).
29. Friston, K.J.; Holmes, A.P.; Worsley, K.J.; Poline, J.; Frith, C.D.; Frackowiak, R.S.J. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **1994**, *2*, 189–210. [[CrossRef](#)]
30. Worsley, K.J.; Marrett, S.; Neelin, P.; Vandal, A.C.; Friston, K.J.; Evans, A.C. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.* **1996**, *4*, 58–73. [[CrossRef](#)]
31. Gyöngy, I.; Michaletzky, G. On Wong–Zakai approximations with δ -martingales. *Proc. R. Soc. London. Ser. A Math. Phys. Eng. Sci.* **2004**, *460*, 309–324. [[CrossRef](#)]
32. Da Costa, L.; Da Costa, N.; Heins, C.; Medrano, J.; Pavliotis, G.A.; Parr, T.; Meera, A.A.; Friston, K. A Theory of Generalized Coordinates for Stochastic Differential Equations. *Stud. Appl. Math.* **2025**, *154*, e70062. [[CrossRef](#)]
33. Stratonovich, R.L. *Topics in the Theory of Random Noise*; Macmillan Education: South Yarra, Australia, 1967; Volume 2
34. Esser, C.; Jaffard, S.; Vedel, B. Regularity properties of random wavelet series. *arXiv* **2023**, arXiv:2304.00811. [[CrossRef](#)]
35. Aubry, J.M.; Jaffard, S. *Random wavelet series: Theory and applications*. In *Emergent Nature*; World Scientific: Singapore, 2002. [[CrossRef](#)]
36. Adler, R.J. On excursion sets, tube formulas and maxima of random fields. *Ann. Appl. Probab.* **2000**, *10*, 1–74. [[CrossRef](#)]
37. Krishnan, S.R.; Taylor, J.E.; Adler, R.J. The Intrinsic geometry of some random manifolds. *Electron. Commun. Probab.* **2017**, *22*, 1–12. [[CrossRef](#)]
38. Yuan, R.S.; Ma, Y.A.; Yuan, B.; Ao, P. Lyapunov function as potential function: A dynamical equivalence. *Chin. Phys. B* **2014**, *23*, 010505. [[CrossRef](#)]

39. Parr, T.; Da Costa, L.; Friston, K. Markov blankets, information geometry and stochastic thermodynamics. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2019**, *378*, 20190159. [[CrossRef](#)]
40. Wang, H.; Hu, W.; Gan, X.; Ao, P. The generalized Lyapunov function as Ao's potential function: Existence in dimensions 1 and 2. *J. Appl. Anal. Comput.* **2023**, *13*, 359–375. [[CrossRef](#)]
41. Khalil, H.K. *Nonlinear Control*; Pearson: Upper Saddle River, NJ, USA, 2014.
42. Fliess, M. A note on the invertibility of nonlinear input-output differential systems. *Syst. Control Lett.* **1986**, *8*, 147–151. [[CrossRef](#)]
43. Fliess, M.; Lévine, J.; Martin, P.; Rouchon, P. On Differentially Flat Nonlinear Systems. *IFAC Proc. Vol.* **1992**, *25*, 159–163. [[CrossRef](#)]
44. Fliess, M.; Lévine, J.; Martin, P.; Rouchon, P. A Lie-Bäcklund approach equivalence and flatness of nonlinear systems. *IEEE Trans. Automat. Control* **1999**, *44*, 922–937. [[CrossRef](#)]
45. Sira-Ramírez, H. *Differentially Flat Systems*; CRC Press: Boca Raton, FL, USA, 2004.
46. Hilbert, D. Über den Begriff der Klasse von Differentialgleichungen. *Math. Ann.* **1912**, *73*, 95–108. [[CrossRef](#)]
47. Cartan, M. Sur l'équivalence absolue de certains systèmes d'équations différentielles et sur certaines familles de courbes. *Bull. Soc. Math. Fr.* **1914**, *2*, 12–48. [[CrossRef](#)]
48. Fuller, S.; Greiner, B.; Moore, J.; Murray, R.; van Paassen, R.; Yorke, R. The Python Control Systems Library (python-control). In Proceedings of the 60th IEEE Conference on Decision and Control (CDC), Austin, TX, USA, 14–17 December 2021; IEEE: Piscataway, NJ, USA, 2021, pp. 4875–4881.
49. Perrin, D. *Algebraic Geometry*; Springer: London, UK, 2008. [[CrossRef](#)]
50. Eisenbud, D.; Harris, J. *The Geometry of Schemes*, 1st ed.; Graduate Texts in Mathematics; Springer: New York, NY, USA, 2000.
51. Agrios, M. A Very Elementary Introduction to Sheaves. *arXiv* **2022**, arXiv:2202.01379. [[CrossRef](#)]
52. Curry, J. Sheaves, Cosheaves and Applications. *arXiv* **2013**, arXiv:1303.3255.
53. Perrinet, L.U.; Adams, R.A.; Friston, K.J. Active inference, eye movements and oculomotor delays. *Biol. Cybern.* **2014**, *108*, 777–801. [[CrossRef](#)] [[PubMed](#)]
54. Adams, R.A.; Aponte, E.; Marshall, L.; Friston, K.J. Active inference and oculomotor pursuit: The dynamic causal modelling of eye movements. *J. Neurosci. Methods* **2015**, *242*, 1–14. [[CrossRef](#)] [[PubMed](#)]
55. Parr, T.; Friston, K.J. Active inference and the anatomy of oculomotion. *Neuropsychologia* **2018**, *111*, 334–343. [[CrossRef](#)]
56. Parr, T.; Friston, K.J. The computational pharmacology of oculomotion. *Psychopharmacology* **2019**, *236*, 2473–2484. [[CrossRef](#)]
57. Harris, D.; Vine, S.; Wilson, M.; Arthur, T. The Relationship Between Environmental Statistics and Predictive Gaze Behaviour During a Manual Interception Task: Eye Movements as Active Inference. *Comput. Brain Behav.* **2023**, *7*, 225–241. [[CrossRef](#)]
58. Wong, A. Listing's law: Clinical significance and implications for neural control. *Surv. Ophthalmol.* **2004**, *49*, 563–575. [[CrossRef](#)]
59. Ekstrand, B. Equations of motion for a two-axes gimbal system. *IEEE Trans. Aerosp. Electron. Syst.* **2001**, *37*, 1083–1091. [[CrossRef](#)]
60. Siciliano, B.; Khatib, O. (Eds.) *Springer Handbook of Robotics*; Springer International Publishing: Cham, Switzerland, 2016. [[CrossRef](#)]
61. Fliess, M.; Join, C.; Moussa, K.; Djouadi, S.M.; Alsager, M.W. Toward simple in silico experiments for drugs administration in some cancer treatments. *IFAC-PapersOnLine* **2021**, *54*, 245–250. [[CrossRef](#)]
62. Sira-Ramírez, H.; Luviano-Juárez, A.; Ramírez-Neria, M.; Zurita-Bustamante, E.W. *Active Disturbance Rejection Control of Dynamic Systems: A Flatness Based Approach*; Butterworth-Heinemann: Oxford, UK, 2017.
63. Sira-Ramírez, H.; Gómez-León, B.C.; Aguilar-Orduña, M.A. A two stage Active Disturbance Rejection Control design for under-actuated nonlinear systems. *IFAC-PapersOnLine* **2023**, *56*, 4539–4544. [[CrossRef](#)]
64. Utkin, V.I. *Sliding Modes in Control and Optimization*; Springer: Berlin/Heidelberg, Germany, 1992. [[CrossRef](#)]
65. Shtessel, Y.; Edwards, C.; Fridman, L.; Levant, A. *Sliding Mode Control and Observation*; Springer: New York, NY, USA, 2014. [[CrossRef](#)]
66. Fliess, M.; Join, C. Model-free control. *Int. J. Control* **2013**, *86*, 2228–2252. [[CrossRef](#)]
67. Fliess, M.; Mounier, H. Controllability and observability of linear delay systems: An algebraic approach. *ESAIM: Control. Optim. Calc. Var.* **1998**, *3*, 301–314. [[CrossRef](#)]
68. Mounier, H.; Rudolph, J. Flatness-based control of nonlinear delay systems: A chemical reactor example. *Int. J. Control* **1998**, *71*, 871–890. [[CrossRef](#)]
69. Mounier, H.; Rudolph, J. Flatness and quasi-static state feedback in non-linear delay systems. *Int. J. Control* **2008**, *81*, 445–456. [[CrossRef](#)]
70. Estrada-Sánchez, I.; Velasco-Villa, M.; Rodríguez-Cortés, H. Prediction-Based Control for Nonlinear Systems with Input Delay. *Math. Probl. Eng.* **2017**, *2017*, 7415418. [[CrossRef](#)]
71. Krstic, M. Input Delay Compensation for Forward Complete and Strict-Feedforward Nonlinear Systems. *IEEE Trans. Autom. Control* **2010**, *55*, 287–303. [[CrossRef](#)]

72. Bresch-Pietri, D.; Petit, N.; Krstic, M. Prediction-based control for nonlinear state- and input-delay systems with the aim of delay-robustness analysis. In Proceedings of the 2015 54th IEEE Conference on Decision and Control (CDC), Osaka, Japan, 15–18 December 2015; IEEE: Piscataway, NJ, USA, 2015, pp. 6403–6409. [[CrossRef](#)]
73. Karafyllis, I.; Krstic, M. *Predictor Feedback for Delay Systems: Implementations and Approximations*; Springer International Publishing: Cham, Switzerland, 2017. [[CrossRef](#)]
74. Rodino. *Linear Partial Differential Operators in Gevrey Spaces*; World Scientific: Singapore, 1993; pp. 5–59. [[CrossRef](#)]
75. Balser, W. *Formal Power Series and Linear Systems of Meromorphic Ordinary Differential Equations*; Springer: New York, NY, USA, 2000. [[CrossRef](#)]
76. Teofanov, N.; Tomić, F.; Žigić, M. An introduction to extended Gevrey regularity. *arXiv* **2024**, arXiv:2404.17366. [[CrossRef](#)]
77. Lindquist, A.; Picci, G. On the Stochastic Realization Problem. *SIAM J. Control Optim.* **1979**, *17*, 365–389. [[CrossRef](#)]
78. Veeravalli, T.; Raginsky, M. Revisiting Stochastic Realization Theory using Functional Itô Calculus. *arXiv* **2024**, arXiv:2402.10157. [[CrossRef](#)]
79. van Nieuwstadt, M.; Rathinam, M.; Murray, R. Differential flatness and absolute equivalence. In Proceedings of the 1994 33rd IEEE Conference on Decision and Control, Orlando FL, USA, 14–16 December, 1994; IEEE: Piscataway, NJ, USA, 1994; CDC-94. [[CrossRef](#)]
80. Olver, P.J. *Equivalence, Invariants and Symmetry*; Cambridge University Press: Cambridge, UK, 1995. [[CrossRef](#)]
81. Ivey, T.A.t.A. *Cartan for Beginners*; Graduate Studies in Mathematics; American Mathematical Society: Providence, RI, USA, 2003.
82. Pio-Lopez, L.; Kuchling, F.; Tung, A.; Pezzulo, G.; Levin, M. Active inference, morphogenesis, and computational psychiatry. *Front. Comput. Neurosci.* **2022**, *16*, 988977. [[CrossRef](#)]
83. Badcock, P.B.; Davey, C.G. Active Inference in Psychology and Psychiatry: Progress to Date? *Entropy* **2024**, *26*, 833. [[CrossRef](#)]
84. Harris, H.W. Active inference and psychodynamics: A novel integration with applications to depression and stress disorders. *Front. Psychiatry* **2025**, *16*, 1630858. [[CrossRef](#)] [[PubMed](#)]
85. Adams, R.A.; Stephan, K.E.; Brown, H.R.; Frith, C.D.; Friston, K.J. The Computational Anatomy of Psychosis. *Front. Psychiatry* **2013**, *4*, 47. [[CrossRef](#)]
86. Adams, R.A.; Perrinet, L.U.; Friston, K. Smooth Pursuit and Visual Occlusion: Active Inference and Oculomotor Control in Schizophrenia. *PLoS ONE* **2012**, *7*, e47502. [[CrossRef](#)]
87. Perrinet, L.U.; Adams, R.A.; Friston, K. Active inference, eye movements and oculomotor delays. *BMC Neurosci.* **2013**, *14*, P133. [[CrossRef](#)]
88. Chelouah, A. Extensions of differential flat fields and Liouvillian systems. In Proceedings of the 36th IEEE Conference on Decision and Control, San Diego, CA, USA, 12 December 1997; IEEE: Piscataway, NJ, USA, 1997; CDC-97, Volume 5, pp. 4268–4273. [[CrossRef](#)]
89. Chelouah, A. Diffieties and Liouvillian Systems. *arXiv* **2010**, arXiv:1010.3909. [[CrossRef](#)]
90. Chetverikov, V.N. Liouville systems and symmetries. *Differ. Equ.* **2012**, *48*, 1639–1651. [[CrossRef](#)]
91. Nicolau, F.; Mounier, H.; Androulakis, I.P. HPA axis differential flatness and Liouvillian study for higher resiliency investigations. *IMA J. Math. Control Inf.* **2023**, *40*, 746–788. [[CrossRef](#)]
92. Nicolau, F.; Mounier, H. Flatness of Networks of Synaptically Coupled Excitatory-Inhibitory Neural Modules. *ESAIM Control. Optim. Calc. Var.* **2023**, *29*, 89. [[CrossRef](#)]
93. Karrer, T.M.; Kim, J.Z.; Stiso, J.; Kahn, A.E.; Pasqualetti, F.; Habel, U.; Bassett, D.S. A practical guide to methodological considerations in the controllability of structural brain networks. *J. Neural Eng.* **2020**, *17*, 026031. [[CrossRef](#)]
94. Baggio, G.; Pasqualetti, F.; Zampieri, S. Energy-Aware Controllability of Complex Networks. *Annu. Rev. Control. Robot. Auton. Syst.* **2022**, *5*, 465–489. [[CrossRef](#)]
95. Nandi, S.; Singh, T. Global Sensitivity Analysis on the Bergman Minimal Model. *IFAC-PapersOnLine* **2020**, *53*, 16112–16118. [[CrossRef](#)]
96. Bergman, R.N. Origins and History of the Minimal Model of Glucose Regulation. *Front. Endocrinol.* **2021**, *11*, 583016. [[CrossRef](#)]
97. Rudolph, J. Well-formed dynamics under quasi-static state feedback. *Banach Cent. Publ.* **1995**, *32*, 349–360. [[CrossRef](#)]
98. Crespo, T.; Hajto, Z.; Mohseni, R. Real Liouvillian extensions of partial differential fields. *Symmetry Integr. Geom. Methods Appl.* **2021**, *17*, 095. [[CrossRef](#)]
99. Besançon, G. *Nonlinear Observers and Applications*; Springer: Berlin/Heidelberg, Germany, 2007. [[CrossRef](#)]
100. Isidori, A. Nonlinear Control Theory for Automation. In *Springer Handbook of Automation*; Nof, S.Y., Ed.; Springer International Publishing: Cham, Switzerland, 2023; pp. 163–187. [[CrossRef](#)]
101. Diop, S. Some Control Observation Problems and Their Differential Algebraic Partial Solutions. In *Algebraic and Symbolic Computation Methods in Dynamical Systems*; Quadrat, A., Zerz, E., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 147–160. [[CrossRef](#)]

102. Sira-Ramirez, H.J.; Garcia Rodriguez, C.; Cortes Romero, J.A.; Luviano Juarez, A. *Algebraic Identification and Estimation Methods in Feedback Control Systems*; Wiley Series in Dynamics and Control of Electromechanical Systems; John Wiley & Sons: Nashville, TN, USA, 2014.
103. Join, C.; Delaleau, E.; Fliess, M. Flatness-based control revisited: The HEOL setting. *Comptes Rendus. Mathématique* **2024**, *362*, 1693–1706. [[CrossRef](#)]
104. Greco, L.; Mounier, H.; Bekcheva, M. An approximate characterisation of the set of feasible trajectories for constrained flat systems. *Automatica* **2022**, *144*, 110484. [[CrossRef](#)]
105. Beaver, L.E.; Malikopoulos, A.A. Optimal control of differentially flat systems is surprisingly easy. *Automatica* **2024**, *159*, 111404. [[CrossRef](#)]
106. Join, C.; Delaleau, E.; Fliess, M. The Euler-Lagrange Equation and Optimal Control: Preliminary Results. In Proceedings of the 2024 12th International Conference on Systems and Control (ICSC), Batna, Algeria, 3–5 November 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 155–160. [[CrossRef](#)]
107. Join, C.; Delaleau, E.; Fliess, M. Model-Free Predictive Control: Introductory Algebraic Calculations, and a Comparison with HEOL and ANNs. *IFAC-PapersOnLine* **2025**, *59*, 255–260. . [[CrossRef](#)]
108. Richalet, J.; Rault, A.; Testud, J.; Papon, J. Model predictive heuristic control: Applications to industrial processes. *Automatica* **1978**, *14*, 413–428. [[CrossRef](#)]
109. Mayne, D.Q. Model predictive control: Recent developments and future promise. *Automatica* **2014**, *50*, 2967–2986. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.