The information geometry of two-field functional integrals

Eric Smith

Abstract

Two-field functional integrals (2FFI) are an important class of solution methods for generating functions of dissipative processes, including discrete-state stochastic processes, dissipative dynamical systems, and decohering quantum densities. The stationary trajectories of these integrals describe a conserved current by Liouville’s theorem, despite the absence of a conserved kinematic phase space current in the underlying stochastic process. We develop the information geometry of generating functions for discrete-state classical stochastic processes in the Doi-Peliti 2FFI form, and exhibit two quantities conserved along stationary trajectories. One is a Wigner function, familiar as a semiclassical density from quantum-mechanical time-dependent density-matrix methods. The second is an overlap function, between directions of variation in an underlying distribution and those in the directions of relative large-deviation probability that can be used to interrogate the distribution, and expressed as an inner product of vector fields in the Fisher information metric. To give an interpretation to the time invertibility implied by current conservation, we use generating functions to represent importance sampling protocols, and show that the conserved Fisher information is the differential of a sample volume under deformations of the nominal distribution and the likelihood ratio. We derive a pair of dual affine connections particular to Doi-Peliti theory for the way they separate the roles of the nominal distribution and likelihood ratio, distinguishing them from the standard dually-flat connection of Nagaoka and Amari defined on the importance distribution, and show that dual flatness in the affine coordinates of the coherent-state basis captures the special role played by coherent states in Doi-Peliti theory.
Keywords  Information geometry · Doi-Peliti theory · Liouville’s theorem · Fisher information · importance sampling · duality

1 Introduction: understanding the Liouville theorems that emerge in the probability analysis of dissipative systems

Hamiltonian state spaces, conserved phase-space densities, and two-field functional integral representations

The defining feature of dissipative systems, whether classical or quantum, is that trajectories initially distinct can merge and that distributions or densities that differ at their initial conditions become more similar over time as they are increasingly governed by local generating parameters at the expense of memory. Because such systems are intrinsically irreversible, they obey no kinematic Liouville theorem (see [35]) describing phase space densities that are conserved along flow lines. For Markov jump processes on discrete state spaces, a kinematic phase space is not defined, while for stochastically perturbed dynamical systems, its volume is not preserved under time evolution. Yet in the study of the time evolution of probability distributions for stochastic processes, a conserved volume element does arise, through a Hamilton–Jacobi theory that evolves the gradients of log-probabilities together with their values [13, 69].

In the classical dynamics of Hamiltonian systems, conserved densities in phase space are associated with reversibility and made possible by the presence of canonically conjugate pairs of momentum and position coordinates specifying states. Convergence of trajectories in the projection onto either component in such a pair is disambiguated by separation in the conjugate variable, such that time evolution defines a symplectomorphism on the state space, and the data specifying a state or distribution at one time are the image under a coordinate transformation of the data at any time.

In the Hamilton–Jacobi theory that describes distributions evolving under stochastic processes, the role of the log-probability gradient as a conjugate momentum suggests that it is the resolution in continuous-valued probability distributions that can preserve the data required for backward- as well as forward-time evolution from any instant. Two questions that remain to be clarified, then, are: (1) what is the general dynamical role of these gradients as conjugate variables in a state space, in relation to the “coordinate” variables that appear in the Hamilton–Jacobi theory; and (2) what is the conceptual interpretation of conserved densities in this phase space?

An important insight into the abstract role of dual fields was given in early work by Martin, Siggia, and Rose [52], who observed that a complete system of equations similar to the Dyson equations of quantum field theory could not be formulated for stochastically perturbed dynamical systems from functions only of the coordinate field that is dynamically evolved. Their solution—appealing explicitly to quantum field theory as a model—was to re-interpret the coordinate variables in a perturbed dynamical system as operator-valued fields and to introduce a parallel set of dual fields, non-commuting with the coordinates, with the interpretation of sources of perturbation. From the two fields together, they obtained a system of relations equivalent to a Dyson
equation between the advanced and retarded “response functions” to perturbations, and the correlation function of fluctuations of the coordinate variables.

The coordinate field along with the response field introduced in [52] can be recast from operators back to ordinary classical fields if the classical fields become variables of integration in a path integral, through the equivalence of operator formulations and path-integral formulations known in quantum mechanics since Feynman and Hibbs [28]. An especially useful observation by Kamenev [42] is that the response and correlation functions of classical stochastic processes have exactly the algebra of “observable” and “response” fields defined by Keldysh [43] as a rotation of field variables from the pairs of forward- and backward-evolving state fields in Schwinger’s time-loop formulation [63] of time-dependent quantum density matrices. Through this algebraic equivalence the role of systems of response fields introduced as duals to coordinate fields is represented generally in what we may call two-field functional integrals (2FFI). The path-integral formulation of the Schwinger time-loop was the original example, which the operator methods of [52] show may be extended to quite general stochastically perturbed dynamical systems.

For a certain class of Markov jump processes on discrete state spaces, which we may generally describe as stochastic population processes, a direct construction of the 2FFI representation for generating functions, from the transition matrix of the master equation, was worked out by Doi [21, 22] and Peliti [57, 58]. (This case was the path integral studied in [42].) The direct construction [44, 68, 71] makes clear the correspondence between the response field and the jump operator in the master equation, giving a mechanistic meaning to the operator-field interpretation of [52]. The Doi-Peliti integral representation also shows that it is the Liouville operator—the representation of the generator of time translation acting in the space of generating functions [74]—that is the Hamiltonian of the Hamilton–Jacobi theory [13, 69].

In the large-deviation limit, which will be the subject of this paper, the Liouville operator can be replaced with a function on deterministic phase-space trajectories, which becomes the generating function for the Lagrange–Hamilton duality between the tangent and cotangent bundles on the coordinate manifold of first moments of the dynamical distribution. Doi-Peliti theory is particularly expressive of the origin and nature of two-field duality because Peliti’s construction [57, 58] of the functional integral employs a representation of unity in the space of generating functions in terms of outer products of basis functions parametrized by the observable field, and projection operators parametrized by the response field introduced in [52]. The conservation of “data” entailed by a Liouville theorem is nothing other than the requirement of completeness in a representation of unity in a Hilbert space of generating functions.

The information geometry induced by symplectomorphisms acting on probability distributions evolved under stochastic processes

The 2FFI representation of Hamilton–Jacobi theory will give a direct mathematical interpretation of the conserved phase space density, as the Wigner function of the two-field integral (see Sect. 3). This function is known in quantum mechanics as a
semiclassical approximation to quantum density matrices in time-dependent thermal ensembles.

We may better understand its statistical interpretation and its relation to time reversal, however, by going beyond the simple Hamiltonian volume conservation already inherent in 2FFI theories, to derive dual parallel transport relations for vector fields induced by symplectomorphisms on the cotangent bundle, as these relate to the Riemannian geometry induced by cumulant generating functions and their Legendre duals, the large-deviation functions \[ 73 \]. The construction of those aspects of the information geometry of two-field integrals is the main work of this paper.

Information geometry derives from divergence functions between probability distributions \[ 2, 7 \], and provides coordinate-independent constructions for the metric distance between pairs of distributions, or the overlap between two directions of change in the form of the extended Pythagorean theorem \[ 3, 54 \]. In particular, these measures are invariant under the symplectomorphisms generated by time translation, so that dynamical conservation is implied by coordinate invariance.

The probability distributions corresponding to phase-space points along the large-deviation trajectories in Doi-Peliti theory may vary either through change of the underlying probability distribution (e.g., through change in initial conditions) or through change of the argument of the generating function at which that underlying distribution is being studied. The second source of change has an interpretation in terms of biased sampling from the underlying distribution, and establishes a connection between generating functions and methods of importance sampling \[ 56 \] in statistical inference.

Inference is the pertinent context within which to understand the time-reversal of Hamilton–Jacobi equations for stochastic processes, because it relates directly to the adjoint duality between the Kolmogorov forward and backward equations \[ 33 \], equivalently either evolving probability distributions forward in time or operators corresponding to observables backward in time. The representation of adjoint duality in terms of biased sampling from the underlying distribution, and establishes a connection between generating functions and methods of importance sampling \[ 56 \] in statistical inference.

Here we will study shifts in the underlying distributions evolving under some stochastic process, and shifts in the sampling bias in generating functions, as sources of vector fields which are parallel transported under the canonical transformations on phase space induced by time translation. In particular we derive affine connections under which, as differences in underlying distributions are attenuated over time due to dissipation, differences between degrees of sampling bias are amplified in the Doi-Peliti integral. The exact compensation of these two effects in Hamilton–Jacobi theory identifies conserved phase-space densities as densities of the information relevant to inference between protocols for biased sampling and features of underlying distributions.

We derive a particular form of this relation, in which the information delivered under varying degrees of sampling bias about the difference between nearby distributions is the inner product of two vector fields in the Fisher information metric. The conserved densities of sample-bias information under time translation are none other than the eigenvalues of the Fisher information. Dual parallel transport of the vector fields describing shifts in the underlying distribution and shifts in the sample-bias factor...
preserves this inner product along the one-parameter family of symplectomorphisms in the phase space generated by time translation, making the sample information about distribution change, like the phase-space density, an invariant of motion.

**Organization of the presentation**

Section 2 opens with brief reviews of the three background areas needed in this work, beginning with spaces and dualities defined at a single time and then continuing to those added with time evolution. We first consider the way general probability distributions are represented by Laplace transforms as contours in phase space, on which the fundamental Riemannian metric of information geometry, and Legendre duality between conjugate phase-space coordinates, are defined. We then review the correspondence of a generating function to an exponential family of tilted distributions with the interpretation of importance distributions produced by biased sampling of the base distribution. Third we introduce time evolution under a stochastic process generator. We review the construction of the Doi-Peliti two-field integral representation of generating functions, followed by the large-deviation limit in which the Liouville operator becomes a classical Hamiltonian and the generating function for duality between coordinates in the tangent and cotangent bundles over the configuration space.

The main results of the paper are derived in Sect. 3. The Wigner function is defined and shown to be the phase space density conserved under Liouville’s theorem. The two vector fields associated with the dual coordinates in phase space are constructed, to describe variations within exponential families through change of the sample bias, and change across families through change in the base distribution. The transport law for the Fisher metric, and dual affine connections respecting the symplectic structure, are then derived.

Section 4 contains a simple worked example illustrating all aspects of the 2FFI Liouville theorem and its associated dual geometry. Section 5 concludes, relating the familiar treatment of adjoint duality in terms of time reversal [37] to the interpretation developed here as a duality between dynamics and statistical inference.

**2 Review: Legendre duality and information geometry, importance sampling, and Lagrange/Hamilton duality**

Phase spaces with conjugate coordinates play several roles in the constructions to follow. At single times, they host the Legendre duality and Riemannian geometry of surfaces representing probability distributions through their Laplace transforms. Within the exponential family over any one distribution, conjugate fields separate the nominal distribution from other members of the family that take on the interpretation of importance distributions. When probability distributions are evolved in time, the phase spaces in which they are embedded become cotangent bundles of the dynamics, with Lagrangian–Hamiltonian dualities to the tangent bundles that depend on the generator of time translation but not on the states being evolved. In this section we establish
notations for all three roles and show how they are instantiated in Doi-Peliti functional integrals.

2.1 Phase space, foliation by exponential families, Legendre duality, and geometric relations

2.1.1 Embedding of probability distributions as surfaces in phase space, and the geometry in those surfaces

The state spaces on which the Doi-Peliti construction can be carried out to produce generating functions and functionals are non-negative integer lattices in some number of dimensions $D$, which here will be assumed finite for simplicity. Dimensions are indexed $i \in 1, \ldots, D$, and lattice points in the state space are indexed by vectors with integer components $n \equiv (n_i)$ with each $n_i \geq 0$. Examples include such phenomena as stochastic population processes, in which $i$ indexes the types of members in the population and each $n_i$ is the count of individuals of type $i$. Well-developed applications include evolutionary populations \cite{70} and chemical reaction networks \cite{44, 71}. The object of study will be continuous-valued probability distributions $\rho$ over states $n$, indexed $\rho_n$.

The Laplace transform (for a countable basis $\{n\}$, also termed the $z$-transform) of $\rho$ is the moment-generating function

$$\Psi(z) \equiv \sum_n \prod_i z_i^{n_i} \rho_n \equiv \sum_n z^n \rho_n \equiv e^{\psi(\theta)} = \sum_n \rho_n e^{\theta n}$$

If we write each component $z_i \equiv e^{\theta_i}$, $\psi(\theta) = \log \Psi(z)$ is the cumulant-generating function (CGF), which will be convex when written in coordinates $\theta$. As a shorthand we have written $z^n$ as the component-wise $n$th power of $z$. If $n$ is regarded as a column vector and $\theta$ a row vector, the Cartesian inner product is abbreviated $\theta n \equiv \theta^i n_i$, adopting the Einstein summation convention.

The gradient of $\psi(\theta)$, which we denote

$$n(\theta) \equiv \langle n \rangle_{\tilde{\rho}(\theta)} = \frac{\partial \psi}{\partial \theta},$$

is the expectation of $n$ in a distribution

$$\tilde{\rho}_n^{(\theta)} \equiv \rho_n e^{\theta n - \psi(\theta)}$$

tilted with weight $z^n$, which we will interpret later as a sampling bias or likelihood function.

The continuous coordinate vectors $n$ make up the points in a manifold $Q$ that will be the coordinate manifold with respect to which geometries and dynamics are defined. Pairs of values $(\theta, n)$ define a $2D$-dimensional phase space over the coordinate manifold $Q$. Later, when dynamics is introduced, that phase space will become the cotangent bundle $T^*Q$ over manifold $Q$, on which a Lagrangian–Hamiltonian duality
is defined. The first relation we study in this phase space, however, does not depend on its interpretation as a cotangent bundle; it is Legendre duality between values \( \theta \) and \( n \) induced by a convex CGF \( \psi \).

For convex \( \psi(\theta) \), the function \( n(\theta) \) is invertible, and the inverse function \( \theta(n) \) is a \( D \)-dimensional surface, single-valued at each \( n \), in the phase space. Each such surface has an intrinsic Riemannian geometry induced by the Hessian of \( \psi \),

\[
\frac{\partial^2 \psi(\theta)}{\partial \theta^i \partial \theta^j} = g_{ij}(\theta),
\]

known as the Fisher tensor, introduced in the interpretation as a Riemannian metric by Rao [61] (reprinted as [62]). Using the differential geometry notation in which \( \{\partial/\partial \theta^i\}_{i=1}^D \) is the set of basis elements in the tangent space to the exponential family, the Fisher metric is an inner product, which we denote

\[
g_{ij} \equiv \left( \frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \theta^j} \right) \tag{5}
\]

The manifold \( Q \) with metric (4) admits a variety of dual transport relations induced by pairs of connections compatible with that metric, as developed by Amari and Nagaoka [3, 54]. Each hypersurface \( \theta(n) \) is generated by a single underlying distribution \( \rho \) (which we will call the “base distribution”). The whole CGF \( \psi(\theta) \) is thus identified with an exponential family of distributions tilted from \( \rho \) at each value of \( z = e^{\theta} \). The dual connections of information geometry provide a way to understand the information divergences of distributions in such a family with others either inside or outside the family, in familiar geometric terms such as least-distance projections onto the family.

### 2.1.2 Foliations in phase space and the inner product between variations within and between exponential families

We will be interested here in families of base distributions, denoted \( \rho_n^{(n_0)} \), for which the exponential families defining the CGFs, over all values of the index \( n_0 \), provide a foliation of the phase space, as shown in Fig. 1. The CGFs and tilted distributions that make up each leaf are denoted

\[
e^{\psi(\theta,n_0)} \equiv \sum_n \rho_n^{(n_0)} e^{\theta_n} \quad \tilde{e}^{\psi(\theta,n_0)} \equiv \rho_n^{(n_0)} e^{\theta_n - \psi(\theta,n_0)} \tag{6}
\]

The index \( n_0 \) is the value of \( n(\theta) \) at \( \theta = 0 \) in the distribution \( \rho^{(n_0)} \).

Equation (2) implies that the Fisher metric (4) is a coordinate transformation from contravariant to covariant coordinates:

\[
g_{ij}(\theta) = \frac{\partial n_i(\theta)}{\partial \theta^j} \tag{7}
\]
If $\psi(\theta)$ is convex, the transformation (7) is invertible. The inverse of the coordinate transform, and with it the Fisher metric, is obtained from the Legendre transform of $\psi(\theta)$,

$$\psi^*(n) \equiv [\theta n - \psi(\theta)]_{\theta(n)}$$

where $\theta(n)$ is the maximizer of the argument in Eq. (8) over $\theta$ values. $\psi^*(n)$ is the Large-Deviation function (LDF). By construction its gradient gives the inverse function

$$\frac{\partial \psi^*(n)}{\partial n_i} = \theta^i(n).$$

From the definition (1) of the generating function and the extremization condition (8), $\psi(\theta)$ is seen to be, in leading exponential approximation, a kind of convex approximation to $-\log \rho_n$ for $n \approx n_0$, of which $\theta$ is the gradient. These interpretations help to give meaning to the evolution of $\psi(\theta)$ under a Hamilton–Jacobi equation derived below in Sect. 2.3.4.

From Eq. (9), the inverse coordinate transform

$$\frac{\partial \theta^i}{\partial n_j} = \frac{\partial^2 \psi^*(n)}{\partial n_i \partial n_j} \equiv g^{ij}(n)$$

is the inverse of $g(\theta)$ from Eq. (7). The distance between two distributions in the exponential family is then written in either contravariant coordinate differentials $\delta \theta$
or covariant coordinate differentials $\delta n$ as

$$\delta s^2 = \delta \theta^i \delta \theta^j g_{ij}(\theta) = \delta n_i \delta n_{j} g^{ij}(n).$$ (11)

The Fisher metric can be obtained as the projection of the Euclidean metric in $\mathbb{R}^D$ under a spherical embedding of the distribution $\rho_n$, briefly reviewed in Appendix A.1, providing a third set of coordinates for the tilted distribution $\tilde{\rho}(\theta)$. We note this embedding because for the exponential families that will play an important role in stationary-point methods later, the distance between distributions on the Fisher sphere in potentially infinite dimensions $\{n\}$ reduce to divergences of the same form on the phase space, as shown in Appendix A.2.

The distance element (11) measures the divergence between two nearby distributions in an exponential family in the geometry derived from $\psi(\theta)$. For two distributions related to a common base distribution through shifts respectively in $\theta$ or in $n_0$, we seek a measure of the degree to which the two changes overlap. The extended Pythagorean theorem [3, 54] for Kullback–Leibler (KL) divergences provides such a measure, where in the small-coordinate limit it becomes an inner product of the two coordinate displacements in the Fisher metric:

$$D\left(\tilde{\rho}_n^{(\theta, n_0+dn_0)} \bigg\| \tilde{\rho}_n^{(\theta, n_0)}\right) + D\left(\tilde{\rho}_n^{(\theta+d\theta, n_0)} \bigg\| \tilde{\rho}_n^{(\theta, n_0)}\right) - D\left(\tilde{\rho}_n^{(\theta+d\theta, n_0+d\theta, n_0)} \bigg\| \tilde{\rho}_n^{(\theta, n_0)}\right)$$

$$= \sum_n \left(\tilde{\rho}_n^{(\theta, n_0+dn_0)} - \tilde{\rho}_n^{(\theta, n_0)}\right) \log \left(\frac{\tilde{\rho}_n^{(\theta+d\theta, n_0)}}{\tilde{\rho}_n^{(\theta, n_0)}}\right)$$

$$\to \delta \theta^i \delta n_{0j} \frac{\partial^2 \psi(\theta, n_0)}{\partial \theta^i \partial n_{0j}} \equiv \delta \theta^i \delta n_{0j} \left\{ \frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial n_{0j}} \right\}$$ (12)

Recognizing that $\partial \psi / \partial \theta^i = n_i(\theta, n_0)$, the inner product in the final line of Eq. (12) is just the sensitivity

$$\left\{ \frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial n_{0j}} \right\} = \frac{\partial n_i}{\partial n_{0j}}$$ (13)

of the mean in the tilted distribution to variations in the base.

### 2.1.3 Preservation of the inner product in connection with Liouville’s theorem

A coordinate change from the mean in the base distribution to the mean in the tilted exponential distribution produces the metric in mixed coordinates that by construction is the Kronecker $\delta$,

$$\left\{ \frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial n_{0j}} \right\} = \delta^j_i,$$ (14)
and the coordinate inner product

$$\delta \theta^i \delta n_{0j} \left( \frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial n_{0j}} \right) \equiv \delta \theta^i \delta n_i.$$ \hfill (15)

We may ask, for what one-parameter families of coordinate systems \((\theta(t), n(t))\) is the coordinate inner product (14) conserved across the family? A one-parameter family of coordinates generates a one-parameter family of maps of vector fields \(\delta \theta\) and \(\delta n\) by the action

$$\left( \frac{d}{dt} \delta \theta^i \right)_i = \delta \theta^j \frac{\partial}{\partial \theta^j} \dot{\theta}^i \qquad \left( \frac{d}{dt} \delta n^i \right)_i = \delta n^j \frac{\partial}{\partial n^j} \dot{n}^i.$$ \hfill (16)

The condition

$$\frac{d}{dt} (\delta \theta^i \delta n_i) = 0$$ \hfill (17)

will be met whenever

$$\frac{\partial \dot{\theta}^i}{\partial \theta^j} = -\frac{\partial \dot{n}^j}{\partial n^i},$$ \hfill (18)

where \(\cdot\) denotes \(d/dt\). Eq. (18) is satisfied if there is a symplectic form \(\mathcal{L}(\theta, n)\) in terms of which the velocity vectors along trajectories can be written

$$\dot{\theta}^i = \frac{\partial}{\partial n^i} \mathcal{L}(\theta, n) \quad \quad \dot{n}^i = -\frac{\partial}{\partial \theta^i} \mathcal{L}(\theta, n).$$ \hfill (19)

Alternatively, if Eq. (18) holds everywhere, the form \(\mathcal{L}(\theta, n)\) can be constructed by integration.

Eq. (18) relates the dual contravariant and covariant coordinates under the Fisher metric as canonically conjugate variables in a Hamiltonian dynamical system. We return in Sect. 2.3 to derive symplectic forms \(\mathcal{L}(\theta, n)\) from the generators of stochastic processes.

### 2.2 Tilted distributions as importance distributions, and particular application to large deviations

In order to furnish an interpretation of the 2FFI symplectic transport structure, we appeal to statistical inference to assign informational meanings to the conjugate fields \(z\) that exponentially tilt the base distributions in generating functions. In importance sampling, \(z\) takes on the interpretation of a sampling bias known as a likelihood ratio, which transforms the measure for samples.

In the terminology of importance sampling [56], the base distribution \(\rho^{(n_0)}\) corresponds to the nominal distribution, and the tilted distribution \(\tilde{\rho}^{(\theta, n_0)}\) of Eq. (6) plays the role of an importance distribution. The normalized exponential tilt \(e^{\theta n - \psi(\theta, n_0)}\) is the corresponding likelihood ratio, also called the Radon–Nikodym derivative of the measure between the base and the importance distributions.
Importance distributions $\tilde{\rho}(\theta, n_0)$ can be chosen to concentrate the density of samples away from the mode of $\rho(n_0)$ to values of $n$ that are more informative about observables of interest. Tilts are typically chosen to minimize some cost function, such as the variance of samples. The large-deviation function can be derived as a leading exponential approximation to the tail weight of the base distribution, in a protocol tuned to minimize sample variance, as shown in the following construction from [66].

To illustrate with an example in one dimension, an estimate of the probability that a particle count $n$ exceeds some bound $\bar{n}$ can be obtained by sampling values of the random variable

$$h(\bar{n}) \equiv I\{n > \bar{n}\},$$

the indicator function for $n > \bar{n}$. In the base distribution $\rho(n_0)$, the probability for $n > \bar{n}$ is

$$P(n > \bar{n} | n_0) = \langle h(\bar{n}) \rangle_{(n_0)}. \quad (20)$$

An unbiased estimator for $P(n > \bar{n} | n_0)$ can be obtained by using the tilted distribution $\tilde{\rho}(\theta, n_0)$ of Eq. (6) and instead of accumulating the values $\{0, 1\}$ of the indicator $h(\bar{n})$, accumulating values of the tilted observable

$$\tilde{h}(\theta, \bar{n}) \equiv e^{\psi(\theta, n_0) - \theta \bar{n}} h_n. \quad (22)$$

The tilted estimator is unbiased because

$$\langle \tilde{h}(\theta, \bar{n}) \rangle_{(\theta, n_0)} = \langle h(\bar{n}) \rangle_{(n_0)}. \quad (23)$$

A few lines of algebra, provided in Appendix B, show that an exponential bound for the estimator at any choices of $\theta$ and $\bar{n}$ is given by

$$\langle \tilde{h}(\theta, \bar{n}) \rangle_{(\theta, n_0)} \leq e^{\psi(\theta, n_0) - \theta \bar{n}} \langle h(\bar{n}) \rangle_{(n_0)}. \quad (24)$$

The variance of the same sample estimator has a corresponding bound (see Eq. (146))

$$\left( \langle \tilde{h}(\theta, \bar{n}) \rangle_{(\theta, n_0)} \right)^2 - \langle \tilde{h}(\theta, \bar{n}) \rangle_{(\theta, n_0)}^2 \leq e^{\psi(\theta, n_0) - \theta \bar{n}} \langle h(\bar{n}) \rangle_{(n_0)}^2 - \langle h(\bar{n}) \rangle_{(n_0)}^2. \quad (25)$$

The parameter $\theta$ that minimizes the bound on sample variance (25) also gives the tightest bound (24) on the tail weight. It is the minimizing argument $\theta(\bar{n})$ of Eq. (8), so the bound is given in terms of the LDF as

$$\langle h(\bar{n}) \rangle_{(n_0)} \leq e^{-\psi^*(\bar{n})}. \quad (26)$$

Without further assumptions about $\rho(n_0)$ it is not possible to say more about the ratio $\langle h(\bar{n}) \rangle_{(n_0)}/e^{-\psi^*(\bar{n})}$. The relevant additional property, which is also associated with the use and tightness of saddle-point approximations [36] in Doi-Peliti theory, is the onset
of large-deviations scaling; that is, if \( n_0 \) and \( \bar{n} \) are increased together in proportion to some scale factor \( N \) as \( \bar{n} = N \bar{\nu}, n_0 = N \nu_0 \), the following two limits should exist:

\[
\lim_{N \to \infty} \frac{1}{N} \psi^*(N \bar{\nu}) = \bar{\psi}^*(\bar{\nu}) \quad \text{and} \quad \lim_{N \to \infty} \frac{1}{N} \psi(\theta(N \bar{\nu}); N \nu_0) = \bar{\psi}({\bar{\theta}}). \quad (27)
\]

Then the variance-minimizing tilt \( \theta \) likewise has a limit, the variance \( \partial^2 \psi / \partial \theta^i \partial \theta^j \) in Eq. (4) scales as \( N \), and the relative variance scales as \( 1/N \). Appendix B shows that in this limit the log ratio

\[
\log \left[ \langle h(\bar{n}) \rangle \langle n_0 \rangle / e^{-\psi^*(\bar{n})} \right] \leq O(N^{-1/2})
\]

compared to \( \psi^*(\bar{n}) \sim N \).

### 2.3 The Doi-Peliti integral construction of generating functions for time-dependent probability distributions

Canonical transformations on phase space [35] are introduced when distributions \( \rho \) are evolved under the generators of stochastic processes. A formally exact representation of the time-translation of distributions over finite intervals, acting directly in the representation by generating functions, is given by the Doi-Peliti construction [21, 22, 57, 58]. (See [9, 42, 68] for other self-contained introductions.) Phase-space coordinates in Doi-Peliti theory take on interpretations in terms of distributions and dual projection operators, equivalent in the path integral to the non-commuting operator fields of [52], through Peliti’s construction [57, 58] of a representation of the identity in the Hilbert space of generating functions.

In the path integral, tangent vectors \( \dot{n} \) to configurations, and phase-space dual coordinates \( \theta \), are independent dummy variables of integration. It is the large-deviation limit, realized in Doi-Peliti theory by saddle-point evaluation of the functional integral, that couples coordinate pairs \( (\dot{n}, n) \) and \( (\theta, n) \) as Lagrangian and Hamiltonian dual coordinate systems, with the Liouville operator-representation of the stochastic process generator as the Hamiltonian and the generating function of the coordinate transformation. The tangent vectors to stationary trajectories generate a family of symplectomorphisms in phase space, and the LDF evolves as a solution to the corresponding Hamilton–Jacobi equation [13, 69].

A tutorial on the defining steps in Doi-Peliti theory is given in this subsection. Supporting algebra, where needed to provide a complete self-contained definition of the method, is provided in Appendix C. Doi-Peliti theory was invented to solve Markov jump processes on integer lattices of the kind that describe population processes, and while providing a definite notation and formally-exact integral solution, neither the indexing nor the particular exponential generating functions used here are meant to be a general notation describing all 2FFIs. Appendix D, working in the simplifying limit of Gaussian fluctuations, provides derivations starting from Doi-Peliti theory of equivalent representations in terms of the Langevin stochastic differential equation and the Fokker–Planck equation [74], which manifestly have no restriction to the specific assumptions made in the Doi-Peliti construction. The derivations leading to them may be “inverted” to equivalence classes of 2FFI path integrals in the Gaussian limit, defined by a common integral kernel form explained in [42], and equivalent to the Dyson equations for general stochastic processes derived in [52]. Saddle-point
methods in Doi-Peliti theory are seen to be one instance of the eikonal approximations for large-deviation functions, developed in general form by Freidlin and Wentzell [30], and the forms derived from the action here are related to specific forms arising in that work.

2.3.1 Generator, Hilbert space, and quadrature

The starting representation for time evolution is the continuous-time master equation of the probability distribution $\rho$

$$
\frac{d\rho_n}{dt} = \sum_{n'} T_{nn'} \rho_{n'}. 
$$

(28)

$T_{nn'}$ is known as the transition matrix, and is the representation of the generator of the stochastic process acting in the space of probability distributions. It is left-stochastic, meaning $\sum_n T_{nn'} = 1, \forall n'$, ensuring conservation of probability. The matrix elements $T_{nn'}$ can depend on the time $t$, though in the example developed in Sect. 4 we will use a time-independent generator for simplicity.

The $z$-transform induces a representation of the generator of time translation acting in the space of moment generating functions in the form of a Liouville equation

$$
\frac{\partial}{\partial t} \Psi(z) = -\mathcal{L}(z, \frac{\partial}{\partial z}) \Psi(z),
$$

(29)

$\mathcal{L}(z, \partial/\partial z)$, called the Liouville operator, is conventionally defined with the minus sign of Eq. (29) because its spectrum is non-negative.

For many purposes, the properties of the generating function as an analytic function of a complex variable [29] are not needed, and the algebra of the MGF as a formal power series [77] is sufficient. An operator algebra due to Doi [21, 22] replaces the variable $z$ and derivative $\partial/\partial z$ with formal raising and lowering operators

$$
z_i \rightarrow a_i^\dagger; \quad \frac{\partial}{\partial z_i} \rightarrow a^i. 
$$

(30)

In the condensed notation (1) for vector inner products, we will regard $a$ as a column vector and $a^\dagger$ as a row vector.

Associated with operators $a_i^\dagger$ and $a^i$ are bilinear number operators $\hat{n}_i \equiv a_i^\dagger a^i$ (no Einstein sum), of which the basis monomials $z^n$ under the mapping (30) correspond to number eigenstates. Number states are denoted $|n\rangle$, and the MGF $\Psi(z)$ is represented as a state vector $|\Psi\rangle$ defined in terms of number states as

$$
\sum_n \rho_n |n\rangle \equiv |\Psi\rangle.
$$

(31)
The Liouville equation (29) becomes

$$\frac{d}{dt} |\Psi\rangle = -\mathcal{L}(a^\dagger, a) |\Psi\rangle,$$

in which the the Liouville operator $\mathcal{L}(a^\dagger, a)$ is the former function $\mathcal{L}(z, \partial/\partial z)$ under the substitution (30).

An inner product must be defined to make generating functions states in a Hilbert space. In the Doi theory this is done formally by introducing a dual state $|0\rangle$ satisfying $\langle 0 | 0 \rangle = 1$. Conjugate number states $|n\rangle$ with inner product $\langle n | n \rangle = \delta_{mn}$ are then built up using lowering operators. The correspondence of the Doi dual ground state with a projector on analytic generating functions, and the inner product known as the Glauber norm, corresponding to the evaluation of any MGF at argument $z = 1$, are given in Appendix C.1. The analytic form (1) of the MGF can be recovered using a variant of the Glauber norm, as

$$(0 | e^{za} | \Psi) = \Psi(z).$$

The objective in introducing the Doi operator formalism is to more conveniently compute the quadrature of the Liouville equation (29), formally written

$$|\Psi\rangle_T = Te^{-\int_0^T dt \mathcal{L}(a^\dagger, a)} |\Psi\rangle_0$$

$$\equiv \lim_{\delta t \to 0} T \prod_{k=1}^{T/\delta t} e^{-\delta t \mathcal{L}(a^\dagger, a)_{k\delta t}} |\Psi\rangle_0.$$  \hspace{1cm} (34)

$|\Psi\rangle_T$ is the generating function for the distribution $\rho_n$ evolved to time $t = T$ from the generating function $|\Psi\rangle_0$ for an initial distribution given at time $t = 0$. $T$ denotes time-ordering of the exponential integral, defined operationally in the second line of Eq. (34), in terms of a time-ordered product of applications of $\mathcal{L}(a^\dagger, a)$ evaluated at the sequence of times $k\delta t$.

2.3.2 The Peliti coherent-state expansion and two-field functional integral

The implications of the time-ordered operator algebra in the quadrature (34) for correlation functions may not be at all easy to derive or approximate, and the purpose of Peliti’s functional-integral construction is to supply evaluation methods, and even more importantly, approximation methods including saddle-point evaluations and a systematic approach to perturbation theory such as exists for path-integral formulations in quantum mechanics [28].

A straightforward step in the Peliti construction is to expand arbitrary generating functions in a basis of eigenstates of the lowering operator $a$. However, whereas the indices $\{n\}$ are countable, the eigenvalues of $a$ are continuous, so such a basis is overcomplete. The important and non-trivial step in the Peliti construction is to identify a representation of the identity operator in the space of generating functions, which requires defining projectors that are left-eigenstates of the raising operator $a^\dagger$. \hspace{1cm} \(\mathcal{C}\) Springer
and showing that an expansion in these two continuously-indexed sets of states is
equivalent to the identity operator on number states \( n \). It is in this second step that the
response field in phase space enters the functional-integral construction for stochastic
processes.

The eigenstates of the lowering operator will be the moment-generating functions
of products of Poisson distributions, and these distributions will play a central role in
the interpretation of stationary-point approximations in the functional integral. In the
Poisson distribution with mean \( n \),

\[
\rho_{n_i}^{(n_i)} = e^{-n_i n_i^i / n_i!},
\]

the expectations of factorial moments [10], defined (again, component-wise) as
\( n^k \equiv n!/(n-k)! \), are \( n^k \) for all \( k \). Products of Poisson marginals, or multinomial
distributions that are sections through such products at fixed \( N \equiv \sum_i n_i \),

\[
\rho_n^{(n)} = \frac{1}{N^N} \left( \frac{N!}{n_1!, \ldots, n_D!} \right) \prod_{i=1}^D n_i^{n_i},
\]

both serve as saddle-point approximations to more general distributions in the Doi-
Peliti integral, and also form an important class of exact solutions for some applications
such as chemical reaction network models [4]. Tilts of Poisson (35) or multinomial (36)
distributions remain of the same form, so that nominal and importance distributions
will have a uniform relation to phase-space points at all pairs \((\theta, n)\) in Fig. 1.

The fixed form of all moments in Poisson and multinomial distributions as functions
of the mean makes these minimum-information distributions. For these distributions
the Fisher spherical embedding of divergences of distributions \( \rho \) on the infinite index
set \( \{ n \} \) can be reduced to only \( D \) dimensions in the coordinates \( \{ n_i / N \} \), as shown
in Appendix A.2. Related simplifications, from forms of adjoint duality [38] that
generally require infinitely many parameters, to simple coordinate transforms in the
Doi-Peliti functional integral, are reviewed below in Sect. 2.3.6.

The generating function of a product distribution with a vector of Poisson param-
eters \( \phi \equiv (\phi^i) \) corresponds to the state

\[
|\phi\rangle \equiv e^{(a^\dagger - 1)\phi} |0\rangle,
\]

as may be checked by an elementary expansion of the Taylor’s series for the exponen-
tial. These states are eigenstates of the lowering operators:

\[
a^i |\phi\rangle = \phi^i |\phi\rangle.
\]
left eigenstates of the raising operator must take the form

\[ \left( \phi^\dagger \right) \equiv e^{(1-\phi^\dagger)} \phi (0) e^{\phi^\dagger a}. \]  

(39)

The normalization in Eq. (39) has been chosen for convenience, so that the inner product of two such coherent states will evaluate to

\[ \left( \phi_2^\dagger \phi_1 \right) = e^{(1-\phi_2^\dagger)(\phi_2-\phi_1)}, \]  

(40)

and in particular \( \left( \phi^\dagger \phi \right) = 1 \).

Note that the inner product (40), if expanded in the power series for the two coherent states, identifies a tilted distribution

\[ \tilde{\rho}_n = e^{-\phi_2^\dagger \phi_1} \left( \frac{\phi_2^\dagger \phi_1}{n!} \right)^n, \]  

(41)

so we recognize that the coherent-state parameters \( \phi \) index a space of base distributions, and the dual fields \( \phi^\dagger \) are the biased-sampling weights in an exponential family of generating functions in which \( \log \phi^\dagger \) takes the place of the coordinate \( \theta \) in Eq. (3). In the language of importance sampling from Sect. 2.2, Eq. (41) are importance distributions with mean parameters \( n_i = \phi_2^\dagger \phi_1 \).

The result enabling the Peliti functional integral construction is that the following integral of outer products constitutes a representation of the identity in the space of generating functions:

\[ \int \frac{dD\phi^\dagger dD\phi}{\pi D^2} \left| \phi \right> \left< \phi^\dagger \right| = I, \]  

(42)

a result demonstrated in Eq. (160) of Appendix C.2.

From the insertion of copies of the identity (42) at a sequence of times \( t = k\delta t \) for some small \( \delta t \) into the quadrature (34), some algebra gives the generating function (1) in the form of a functional integral with what Feynman and Hibbs [28] termed a “skeletonized” functional measure (defined in Eq. (162)), as

\[ e^{\psi_T(\theta)} = \int_0^T D^D\phi^\dagger D^D\phi e^{\left( z-\phi^\dagger \right) \phi_T - S + \psi_0 \left( \log \phi_0^\dagger \right)}. \]  

(43)

\( e^{\psi_0 \left( \log \phi_0^\dagger \right)} \) is the initial generating function \( \left| \Psi \right> \) in Eq. (34). \( S \) in Eq. (43) has the form of a Lagrange–Hamilton action functional,

\[ S = \int_0^T dt \left\{ -\left( d_t \phi^\dagger \right) \phi + \mathcal{L} \right\}, \]  

(44)
in which the “kinetic term” that creates a Doi-Peliti Lagrangian comes from the inner product \((40)\) between projectors and states in two expansions for the identity operator at closely-spaced times.

While the CGF \((43)\) is formally a function of the coordinate \(\theta = \log z\), the functional integral \((44)\) and action \((43)\) are at the moment expressed in terms of coherent-state fields \((\phi^\dagger, \phi)\). As noted in Eq. \((41)\), these variables are on one hand mathematically expressive, as they distinguish the mean in a base distribution from the likelihood ratio in an importance distribution. On the other hand they make calculation of the mean in the importance distribution inconvenient because it is a bilinear form in \(\phi^\dagger\) and \(\phi\), and more importantly, the Hessian of the function \(\psi_T\) is not generally positive-definite in coordinates \(z\).

Therefore we introduce the first canonical transformation from the Peliti coherent-state coordinates that will be of interest in this work, which is a simple logarithmic point transformation

\[
\phi_i^\dagger \equiv e^{\theta_i} \quad \phi_i \equiv e^{-\theta_i} n_i. \tag{45}
\]

The fields \(n\) and \(\theta\) would have the interpretation of a mean molecule number and conjugate chemical potential in a chemical-reaction model, so we refer to these as number-potential coordinates, distinguishing them from coherent-state coordinates. The action \((44)\) becomes

\[
S = \int_0^T dt \left\{ - (d_t \theta) n + L(\theta, n) \right\}, \tag{46}
\]

in which \(L(\theta, n)\) is \(L(\phi^\dagger, \phi)\) with \(\phi^\dagger\) and \(\phi\) written as functions of \(n\) and \(\theta\) by Eq. \((45)\).

### 2.3.3 The large-deviation limit and Hamiltonian stationary trajectories

The large-deviation limit of the path integral \((43)\) is a leading-exponential approximation in some value such as \(\phi\) that becomes large to reflect large population size or a similar scale variable. In that limit the integral is approximated by the value of the exponential kernel at its saddle point, which is any stationary solution of the action \(S\) and the initial and final boundary terms. The stationary trajectories for \(S\) in the form \((44)\) satisfy the pair of equations of motion

\[
\dot{\phi}_i^\dagger = \frac{\partial}{\partial \phi_i^\dagger} L(\phi^\dagger, \phi) \quad \dot{\phi}_i = -\frac{\partial}{\partial \phi_i^\dagger} L(\phi^\dagger, \phi). \tag{47}
\]

The final-time boundary value for the field \(\phi_T^\dagger\) is given by the vanishing derivative of the exponent in Eq. \((43)\) with respect to \(\phi_T^\dagger\), resulting in \(\phi_T^\dagger = z\). The initial time boundary value for \(\phi\) is given, after an integration by parts, by the vanishing derivative with respect to \(\phi_0^\dagger\), and depends on the form of the CGF \(\psi_0\) and the stationary-path value of its argument \(\phi_0^\dagger\). The two conditions are solved self-consistently through the Eq. \((47)\).
Joint stationary values for \((\phi^\dagger, \phi)\) at intermediate times \(t\), in a generating function with argument \(z\) imposed at a final time \(T\), represent the bundle of rays in the statistical model that dominate the contribution to the importance distribution at later times. For this reason the stationary trajectory in the base distribution is not generally independent of the trajectory for the tilt in systems with non-linear equations of motion.

The stationary-path equations corresponding to Eq. (47) in the logarithmic number-potential coordinates (45) are

\[
\dot{\theta}_i = \frac{\partial}{\partial n_i} \mathcal{L}(\theta, n) \quad \quad \quad \dot{n}_i = -\frac{\partial}{\partial \theta_i} \mathcal{L}(\theta, n).
\] (48)

These equations are instances of the symplectomorphisms (19). The triangle inequality of divergences (12) within and across exponential families is therefore an invariant of motion.

### 2.3.4 Duality with the Lagrangian, and the Hamilton–Jacobi equation

The equation of motion (48) for \(n\) introduces a coordinate transform from phase-space coordinates \((\theta, n)\) to coordinates \((\dot{n}, n)\) in the tangent bundle \(TQ\) over manifold \(Q\) of \(n\) values. \(-\mathcal{L}(\theta, n)\) is the generating function for the transformation and the Hamiltonian on phase space.

The Lagrangian dual to \(-\mathcal{L}\),

\[
L(\dot{n}, n) \equiv \{\theta \dot{n} + \mathcal{L}(\theta, n)\}_{\theta(\dot{n},n)},
\] (49)

is the argument of the action \(S\) up to a total time derivative in \(\theta n\). By construction, \(L\) is the generating function of the inverse coordinate transform, satisfying

\[
\frac{\partial}{\partial \dot{n}_i} L(\dot{n}, n) \bigg|_{\dot{n}} = \theta_i \quad \quad \quad \frac{\partial}{\partial n_i} L(\dot{n}, n) \bigg|_{\dot{n}} = -\frac{\partial}{\partial n_i} \mathcal{L}(\theta, n) \bigg|_{\theta}.
\] (50)

Thus Lagrange–Hamilton duality identifies the phase space with the cotangent bundle \(T^*Q\) to the coordinate manifold.

For any initial distribution \(\rho\) encoded in the initial-time generating function \(\psi_0(\theta_0)\) in Eq. (43), we may write the action (46) evaluated along a stationary trajectory as

\[
S = -\theta n|_0^T + \int_{n_0}^{n_T} \theta dn + \int_0^T \mathcal{L}(\theta, n) dt.
\] (51)

After performing the Legendre transform (8) to the LDF to cancel the surface term in Eq. (51), we find that \(\psi^*(n)\), at any time, satisfies the pair of equations

\[
\frac{\partial \psi^*}{\partial n} = \theta \quad \quad \quad \frac{\partial \psi^*}{\partial t} = \mathcal{L}\left(\frac{\partial \psi^*}{\partial n}, n\right).
\] (52)
The Hamiltonian equations (48) generate a family of maps of phase space, and under any of these maps the initial-value contour \( \theta_0(n) \) for any distribution \( \rho \) is mapped to a contour \( \theta_t(n) \), along which \( \psi^*_t(n) \) is a solution to the Hamilton–Jacobi equation (52).

2.3.5 The commutative diagram of maps from Legendre and Lagrangian–Hamiltonian dualities

The Legendre and Lagrange–Hamilton dualities of Doi-Peliti theory define a system of \( D \)-dimensional coordinate transformations that form a commutative diagram, summarized by Loek and Zhang [47]. Equations (48, 49) define a map \((\dot{n}, n) \leftrightarrow (\theta, n)\) between the tangent and cotangent bundles, \( TQ \leftrightarrow T^*Q \), describing the same trajectories in respectively Lagrangian and Hamiltonian coordinates. Integration of these equations over any finite time interval \( t \) defines a map from pairs of initial and final points \((n_t, n_0)\) in \( Q \times Q \) to points in \( TQ \) at either initial or final times, by assigning initial and final velocities \((\dot{n}_t, \dot{n}_0)\) at those points. Combined with the maps \( TQ \leftrightarrow T^*Q \) at both times, the Lagrangian map between times generates a symplectomorphism in the cotangent bundle \( T^*Q \leftrightarrow T^*Q \) equal to the direct integration along stationary trajectories (48).

The Legendre duality \( \theta(n) \) at a single time is defined separately on the exponential families over each of the base distributions \( \rho^{(m)} \) that form the leaves in the foliation of Fig. 1, but does not depend on the generator of time translation. The Lagrange–Hamilton duality depends on the generator of time translations and not on the choice of exponential families used to foliate the cotangent bundle. Because points \((\theta, n)\) on stationary trajectories in Doi-Peliti theory correspond to tilted distributions over Poisson base distributions (35), the canonical foliation of phase space in Fig. 1 is singled out by the Peliti representation of unity (42). Note that these distributions constitute a \( 2D \)-dimensional subspace of the space of all distributions on the lattice \( \{n\} \). It is shown in [69] that these special distributions, because they are minimum-information distributions, have the interpretation of macrostates for a general equilibrium or non-equilibrium thermodynamics.

In the framework of [47], the Doi-Peliti integral (43) and other two-field functional integrals are defined on the Pontryagin bundle \( TQ \oplus T^*Q \) of triples \((\theta, \dot{n}, n)\) in which all three variables are independent. It is only with the reduction to saddle-point trajectories (and only if the Doi-Peliti integral is evaluated in its continuous-time limit, and not with a finite time-step in the inner product (40)), that these triples are reduced to Lagrangian and Hamiltonian descriptions of the same trajectories with no “duality gap”.

2.3.6 Coordinate transformations in phase space expressing adjoint duality as time reversal

A large body of work on “fluctuation theorems” [64], although not required to derive the foregoing dualities or their implied conservation laws, is nonetheless helpful in interpreting the meaning of the conserved volumes in Doi-Peliti theory. Fluctuation theorems may be seen as extensions to finite time intervals of the study of instantaneous
adjoint duality between the Kolmogorov-backward and -forward evolution [33] under the transition matrix $T_{nn'}$.

Hatano and Sasa [38] recognized that application of a suitable similarity transform to the transition matrix exchanges the forms of $T_{nn'}$ and its adjoint: that is, the similarity transform exchanges the forms of the Kolmogorov-backward and -forward equations, making the evolution of distributions appear like that of observables and vice versa. Performing the similarity transform at every moment in an extended-time quadrature of the form (34) has the effect of multiplying each path of the stochastic coordinate $n$ by a weight [65], creating an extended-time generating functional for paths. The required similarity transform is generally non-locally determined and requires solution for stationary distributions on the whole state space.

An exception arises for cases in which the stationary distributions are minimum-information distributions such as the Poisson (35), for which the phase-space coordinates $(\theta, n)$ determine all fluctuation moments. In these cases the similarity transform in the potentially-infinite-dimensional state space $\{n\}$ can be projected to a coordinate transform in the phase space [69]. Moreover, the required path weights of [65] for all paths are given by a $D$-dimensional, time-local correction to the Liouville function $L$, making evolution under the adjoint generator appear like ordinary stochastic evolution in reverse time.

One such coordinate transformation, exchanging the roles of coordinate and response fields in phase space, was first used in Doi-Peliti integrals by Baish [11]. Let $n$ be the saddle-point value of the field $n$ in Eq. (45) in the steady state that would be annihilated by $T_{nn'}$ at the parameters it possesses at some time. If $T_{nn'}$ and thus $L$ is explicitly time-dependent, then the scale factor $n$ will generally be different for each time. Define coherent state fields rescaled locally by $n$ as

$$\phi_i^\dagger n_i \equiv \varphi_i^\dagger \quad \frac{\phi_i}{n_i} \equiv \varphi_i.$$  \hfill (53)

The form of the action in fields $(\varphi^\dagger, \varphi)$ remains as in Eq. (44) but the Liouville function $L$ is replaced by a possibly-shifted function

$$\tilde{L}(\varphi^\dagger, \varphi) \equiv L(\phi^\dagger, \phi) + \sum_i (\varphi_i^\dagger \varphi_i) \, dt \log n_i.$$  \hfill (54)

If we follow the Baish transformation (53) with a logarithmic transform equivalent to (45), but in the descaled fields,

$$\varphi_i^\dagger \equiv e^{-n_i} n_i \quad \varphi_i \equiv e^{n_i},$$  \hfill (55)

\footnote{It is impossible to fairly represent the motivations and scope of what has now become a significant fraction of work spanning dynamical systems and statistical mechanics. The study of generating functions for reverse-time trajectories began in dynamical systems [16, 25, 31, 32], and was later taken up in similar form for non-equilibrium stochastic processes [14, 19, 20, 24, 26, 38–41, 45, 46]. Reviews of parts of this literature from different stages in its development and from different domain perspectives include [15, 27, 37, 64].}
we coordinatize the base distributions in Fig. 1 as exponential families, and the
dependence of number $n$ on $\theta$ in mixture families. The evaluation of the triangle
inequality (12) is unchanged, but the coordinates in which vector fields are expanded
are transformed.

The action (44) in the new variables becomes

$$S = \int_0^T dt \left\{ (d_t \eta) n + \tilde{L}(n, \eta) \right\}$$  \hspace{1cm} (56)

where the modified Liouville function $\tilde{L}$ from Eq. (54) must be used, such that in the
new variables

$$\tilde{L}(n, \eta) \equiv L(\theta, n) + \sum_{i=1}^D n_i d_i \log n_i$$  \hspace{1cm} (57)

The symmetric exponential families on phase space represented by the coordinate
systems $(\theta, n)$ from Eq. (45) and $(n, \eta)$ from Eq. (55) provide natural variables in which
to expand the coefficients of affine connections constructed from the $\psi$-divergence
below in Sect. 3.3.

3 The Liouville theorem connecting dynamics to inference induced
by two-field stationary trajectories

In the review of standard results in Sect. 2, we have shown where a phase space with
conserved volume under mappings along Hamiltonian stationary trajectories originates
within Doi-Peliti theory, and identified a particular form (15) for the volume element
preserved by a mapping of coordinate differentials along these trajectories. In this sec-
tion we compute two quantities that are conserved under time translation and develop
their interpretations as information measures. The first is a scalar phase-space density,
which we identify as the Wigner function of semiclassical Liouville evolution. The
second is an inner product of vector fields, which we express in coordinate-invariant
terms by defining affine connections for dual parallel transport compatible with the
Riemannian metric (4).

The distinctive feature of the parallel transport derived here is that it is affine-flat
in the coherent-state variables of Sect. 2.3.2, rather than in the exponential coordi-
nates $\theta$ normally used to define dually-flat parallel transport, as in Ch. 6 of [3, 54].
The importance-sampling interpretation of Sect. 2.2 identifies a geometric role for
coherent-state variables that we believe has not been recognized. Dual parallel trans-
port under these affine connections expresses conservation of the Liouville volume
element as a result of compensation, in the flow of phase-space trajectories, between
the resolution of underlying base distributions and the discrimination imposed by
sampling bias. The information available from the large-deviation probability as a
sample estimator, about differences between possible initial distributions, is carried
by the eigenvalues of the Fisher information metric, which are transported as invariants
along stationary trajectories.
The canonical transformation of Sect. 2.3.6, which exchanges the roles of fields representing the base distribution and the sampling bias in the Doi-Peliti integral, is used to express the conserved information volume in symmetric form in terms of the Fisher metric. The symmetric representation of adjoint duality [37, 38] in terms of apparent time-reversal in the Doi-Peliti integral makes clear that the “reversibility” entailed by conserved phase-space volumes in the Hamilton–Jacobi equation should be understood as a reflection of duality between dynamics (the forward propagation of base distributions) and inference (the backward propagation of sample biases).

We begin with the conserved scalar density and its interpretation as a statistical model in the terms of importance sampling, and then define vector fields corresponding to the extended Pythagorean theorem (12) and derive appropriate affine connections for them.

### 3.1 The Peliti functional integral as a statistical model

The Peliti basis of coherent states (37) defines a statistical model for the stochastic process. The role of the projection operators in the representation of unity (42) in populating the model can be clarified by splitting the functional integral (44) at any intermediate time $t$, in the same fashion as the Chapman–Kolmogorov equation splits the time evolution of a probability distribution by a sum over intermediate states. This is done by integrating Eq. (43) up to time $t$, inserting an explicit representation of unity in terms of a pair of fields $(\phi^\dagger_t, \phi_t)$, and resuming the functional integral on the generating function extracted by that representation of unity:

$$e^{\psi_T(\theta)} = \int_{t+\delta t}^T D\phi^\dagger D\phi e^{\left(z-\phi^\dagger_t\right)\phi_T-S} \int D\phi e^{\phi^\dagger_{t+\delta t}-1} \phi_T \times \int \frac{d\phi^\dagger}{\pi D} e^{-\left(\phi^\dagger_{t+\delta t}-1\right)\phi_T} e^{\psi_t\left(\log \phi^\dagger_{t+\delta t}\right)}.$$ (58)

The argument $\log \phi^\dagger_{t+\delta t}$ is the CGF coordinate for an exponential family of distributions tilted from whatever base distribution the functional integral produces at time $t$. The inner product $\phi^\dagger_{t+\delta t} \phi_T$ is a dual mixture coordinate, corresponding to the mean of $n$ in the importance distribution with $\phi^\dagger_T$ as the base distribution and $\phi^\dagger_{t+\delta t}$ as the likelihood ratio. It is the mean of this importance distribution, together with the log-likelihood that is its dual coordinate in the exponential family, that must transform under symplectomorphism to satisfy the condition (19) for preservation of inner products. We show next that the stationary-path conditions provide the necessary mapping.

### 3.2 The Wigner function from the two-field identity operator plays the role of a phase-space density

The scalar density in 2FFI that fills the role of a phase space density in classical Hamiltonian mechanics is the Wigner function [76], of which versions exist for both...
classical and quantum systems.\textsuperscript{2} It is defined in terms of the representation of unity in Eq. (58), as

\[
w_t\left(\phi^+_x, \phi^+_y\right) \equiv \frac{1}{\pi^D} \int_{T + \Delta t} T D\phi^+ D\phi e^{\left(z - \phi^+_x\right)\phi_T - S \left(\phi^+_x - \phi^+_y\right)\phi^+_y} e^{\psi_1\left(\log \phi^+_1\right)}
\]

Eq. (59) implies that, for \(w_t\left(\phi^+_x, \phi^+_y\right)\) at any time,

\[
e^{\psi_T(\theta)} = \int d\phi^+ d\phi w_t\left(\phi^+_x, \phi^+_y\right)
\]

In a saddle-point approximation, one identifies arguments \(\bar{\phi}^+_x, \bar{\phi}^+_y\) for which, to leading exponential order,

\[
e^{\psi_T(\theta)} \sim w_t\left(\bar{\phi}^+_x, \bar{\phi}^+_y\right)
\]

Since Eq. (61) approximates the same function at any time \(t\), its total time derivative along the sequence of stationary points must vanish,

\[
0 = \frac{d}{dt} w_t\left(\bar{\phi}^+_x, \bar{\phi}^+_y\right) = \left(\frac{\partial}{\partial t} + \frac{d\phi^+_x}{dt} \frac{\partial}{\partial \phi^+_x} + \frac{d\phi^+_y}{dt} \frac{\partial}{\partial \phi^+_y}\right) w_t\left(\bar{\phi}^+_x, \bar{\phi}^+_y\right)
\]

Moreover, the stationary points should coincide with values along the stationary trajectories (47) of the functional integral (58), which satisfy

\[
\left(\sum_i \frac{\partial \phi^+_i}{\partial \phi^+_i} + \sum_i \frac{\partial \phi^+_i}{\partial \phi^+_i}\right)_{\bar{\phi}^+_x, \bar{\phi}^+_y} = 0
\]

Equation (62) may thus be recast as the conservation law for a 2D-dimensional current \(\left(\dot{\phi}^+_w, \dot{\phi}^+ w\right)\),

\[
0 = \frac{\partial w_t}{\partial t} + \sum_i \frac{\partial}{\partial \phi^+_i} \left(\phi^+_i w_t\right) + \sum_i \frac{\partial}{\partial \phi^+_i} \left(\phi^+_i w_t\right)
\]

which is Liouville’s theorem.

\(w_t\) is a density of rays for joint base distributions and likelihood ratios that is conserved along the Doi-Peliti stationary trajectories. \(\log w_t\) is the leading exponential approximation to the value of the CGF. It therefore integrates information along the trajectory from the final-time imposed value of \(z\) and the initial-time structure of the generating function \(\psi_0\left(\log \phi^+_0\right)\).

\textsuperscript{2} Wigner introduced this function to treat quantum density matrices, such as those arising in the Schwinger-Keldysh time-loop 2FFI [43, 63], and it is closely related to the Glauber–Sudarshan “P-representation” [17, 34, 72]. Equivalent constructions are widely applied to problems in time-frequency analysis or optics [12, 75], and an example of the map from quantum to classical 2FFI Wigner functions is given in [67].
The indirect definition (59) of the Wigner function in terms of the functional integral is convenient to manipulate but perhaps not very self-explanatory. Appendix E gives a direct construction of the stationary-point approximation in terms of a density $\rho(\theta)$ over the basis of coherent states $|\phi\rangle$ and their Laplace transforms, and verifies that the sequence of stationary points do indeed coincide with the equations of motion (47).

**Constraints and conserved current flows in reduced dimensions**

Often systems of interest will evolve under constraints arising from conservation laws, such as conserved quantities of the stoichiometry in chemical reaction networks [44, 60, 71]. Conserved quantities result in flat directions in the CGF and zero eigenvalues of the Fisher metric. Since generally the constraints will involve multiple species, and because the logarithmic canonical transform (45) is defined in the species basis, it will not be possible to factor out non-dynamical combinations. Then the transport equation (64) for the current of the Wigner density will occupy only a sub-manifold of the $2D$-dimensional Doi-Peliti coordinate space needed to define the system.

A convenient way to handle constraints is to work in the eigenbasis of the Fisher metric which we will index with subscript $\alpha$, where a number-potential counterpart to the transport equation (64) reads

$$0 = \frac{\partial w_t}{\partial t} + \sum_\alpha \frac{\partial}{\partial \theta^\alpha} (\dot{\theta}^\alpha w_t) + \sum_\alpha \frac{\partial}{\partial n_\alpha} (\dot{n}_\alpha w_t) \tag{65}$$

The picture of the Liouville equation as implying a conserved volume element

$$\frac{d}{dt} \left( \prod_\alpha \delta \theta^\alpha \delta n_\alpha \right) = 0 \tag{66}$$

with the product index $\alpha$ taken only over nonzero eigenvalues of the Fisher metric, remains nondegenerate and has a direct interpretation in terms of the product of eigenvectors of the Fisher inner product in independent dimensions.

**3.3 The Fisher metric and cubic tensor in dual canonical coordinates**

The leading-exponential equivalence of the Wigner density to the CGF from Eq. (61) suggests that the $2D$-dimensional differential of the stationary-point CGF should likewise obey a symplectic transport law, implying a transport law for the Fisher metric. To derive those results we return to the expression of the differential of the CGF in terms of the generalized Pythagorean theorem (12), and derive the Fisher metric from the $\psi$-divergence following Amari [2, Sec. 6.2].

Base distributions corresponding to points along stationary paths under the action (44) form exponential families, because they are in the class of coherent-state distributions described in Appendix A.2. Therefore label importance distributions (6) symmetrically as $\tilde{\rho}(\theta, \eta)$ with the exponential coordinates in the two logarithmic transformations (45) and (55). To study their independent variations about a reference value
(θ_R, η_R), introduce two distinct exponential families, labeled

$$\rho \equiv \tilde{\rho}(\theta, \eta)_{\eta = \eta_R} \quad \rho' \equiv \tilde{\rho}(\theta, \eta)_{\theta = \theta_R} \quad (67)$$

The ψ-divergence $D_\psi(\theta : \eta) = \psi(\theta) + \psi^*(n) - n\theta$, a Bregman divergence of the CGF, is related to the Kullback–Leibler divergence of $\rho'$ from $\rho$ as

$$D_\psi(\theta : \eta) = D_{\text{KL}}(\rho' \| \rho) = \sum_n \tilde{\rho}_n(\theta_R, \eta) \log \left( \frac{\tilde{\rho}_n(\theta_R, \eta)}{\tilde{\rho}_n(\theta, \eta_R)} \right) \quad (68)$$

The mixed second partial derivative of $-D_\psi$ gives the same variance that defines the Fisher metric. At general $\theta, \eta$, it is labeled

$$g^D_{ij} = -\frac{\partial}{\partial \theta^i} \frac{\partial}{\partial \eta^j} D_\psi(\theta : \eta)$$

$$= \sum_n \frac{\partial \tilde{\rho}_n(\theta_R, \eta)}{\partial \eta^j} \frac{\partial \log \tilde{\rho}_n(\theta, \eta_R)}{\partial \theta^i}$$

$$= \sum_n \tilde{\rho}_n(\theta_R, \eta) \left( n_j - \frac{\partial \psi}{\partial \eta^j} \right) \left( n_i - \frac{\partial \psi}{\partial \theta^i} \right) - \frac{\partial^2 \psi}{\partial \eta^k \partial \eta^i} \sum_n \tilde{\rho}_n(\theta_R, \eta) \left( n_j - \frac{\partial \psi}{\partial \eta^j} \right)^2 \left( n_i - \frac{\partial \psi}{\partial \theta^i} \right) \quad (69)$$

At $\theta = \theta_R, \eta = \eta_R$, the second line of Eq. (69) recovers exactly the differential form of the Pythagorean theorem of Eq. (12) in dual exponential coordinates.

Two third-order mixed partials define the connection coefficients for Amari’s dually-flat connections on exponential and mixture coordinates. Written in all-contravariant indices,3 these are given by

$$\Gamma^D_{kij} = \frac{\partial}{\partial \eta^k} g^D_{ij}$$

$$= -\frac{\partial^2 \psi}{\partial \theta^k \partial \theta^i} \sum_n \tilde{\rho}_n(\theta_R, \eta) \left( n_j - \frac{\partial \psi}{\partial \eta^j} \right) \left( n_i - \frac{\partial \psi}{\partial \theta^i} \right) \quad (70)$$

3 Note that it is the dual connection $\Gamma^{D*}_{kij} \equiv 0$, written in all-covariant indices, which vanishes as the affine connection on the mixture family.
Evaluated at $\theta = \theta_R$ and $\eta = \eta_R$,
\[
\begin{align*}
g_{ij}^D(\theta_R, \eta_R) &= g_{ij} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} \bigg|_{\theta_R, \eta_R} \\
\Gamma_{kij}^D(\theta_R, \eta_R) &= 0 \\
\Gamma_{kij}^*(\theta_R, \eta_R) &= T_{kij} = \frac{\partial^3 \psi}{\partial \theta^k \partial \theta^i \partial \theta^j} \bigg|_{\theta_R, \eta_R}
\end{align*}
\]

$g_{ij}$ is the Fisher metric introduced in Eq. (4), and $T_{kij}$ is the cubic tensor, also called the Amari–Chentsov tensor [2]. Below we remove the subscript $R$ and write $\theta$ and $\eta$ as the arguments of these tensors.

3.4 The dual vector fields induced by base-distribution initial conditions, and final-time tilts

From the construction of $g^D$ in Sect. 3.3, we can see how to use the stationary-path equations of motion to induce two mappings of vector fields that respect the dual arguments of the $\psi$-divergence. Variations in the likelihood act on the $\rho$ argument, while variations in the base distribution act on the $\rho'$ argument, in Eq. (67). The stationary-path equations are then used to define a 1-parameter family of maps of any basis of dual variations in initial base and final tilt parameters to intermediate times. Conservation of the Liouville volume then translates to a conserved inner product of pairs of vector fields transported respectively under the two branches of the dual mapping. Conservation of the inner product will imply a transport equation for the Fisher metric corresponding to the Eq. (65) for the Wigner function.

Introduce two vector fields corresponding to variations in $\theta$ at the final time $T$, and to variations in $\eta$ at the initial time 0. The first can be independently imposed through the arguments in $\Psi_T(\tau)$, while the second can be independently imposed in the initial data. Fields $\delta \theta_T$ and $\delta \eta_0$ are written in components as
\[
\begin{align*}
\delta \theta_T &\equiv \delta \theta_T^i \frac{\partial}{\partial \theta_T^i} \\
\delta \eta_0 &\equiv \delta \eta_0^i \frac{\partial}{\partial \eta_0^i}
\end{align*}
\]

The stationary-path conditions map the dual initial and final coordinates to pairs of coordinates at any intermediate time, which we denote $\theta(\theta_T, \eta_0, t), \eta(\theta_T, \eta_0, t)$. A one-parameter family of vector fields is defined by assigning to each such coordinate image $(\theta(\theta_T, \eta_0, t), \eta(\theta_T, \eta_0, t))$ the field values
\[
\begin{align*}
\delta \theta(\theta, \eta, t) &\equiv \delta \theta_T^i \frac{\partial}{\partial \theta_T^i} \theta(\theta_T, \eta_0, t) \\
\delta \eta(\theta, \eta, t) &\equiv \delta \eta_0^i \frac{\partial}{\partial \eta_0^i} \eta(\theta_T, \eta_0, t)
\end{align*}
\]

Under the change of coordinates from $(\theta_T, \eta_0)$ to $(\theta, \eta)$ at each time $t$, the vector fields (73) may be written in terms of the local coordinate differentials as
\[
\begin{align*}
\delta \theta(\theta, \eta, t) &\equiv \delta \theta^i(\theta, \eta, t) \frac{\partial}{\partial \theta^i} \\
\delta \eta(\theta, \eta, t) &\equiv \delta \eta^i(\theta, \eta, t) \frac{\partial}{\partial \eta^i}
\end{align*}
\]
Below we suppress the explicit \((\theta, \eta, t)\) coordinate and time arguments of \(\delta \theta^i(\theta, \eta, t)\) and \(\delta \eta(\theta, \eta, t)\), and indicate the time \(t\) in a subscript only where it is needed to avoid confusion.

The vector fields (74) have a time dependence that can be defined through the dependences of \((\theta(T, \eta_0, t), \eta(\theta(T, \eta_0, t))\) on the boundary coordinates and then transformed to the local coordinate system, becoming

\[
\left(\frac{d}{dt}\delta \theta^i\right)^j = \delta \theta^i_T \frac{\partial}{\partial \theta^i_T} \delta \theta^j_T(\theta, \eta_0, t)
\]
\[
= \delta \theta^i_T \frac{\partial}{\partial \theta^i_T} \delta \theta^j_T(\theta, \eta_0, t)
\]
\[
= \delta \theta^i_T \frac{\partial^2 L}{\partial \theta^i \partial n_j}
\]

\[
\left(\frac{d}{dt}\delta \eta^i\right)^j = \delta \eta^i_0 \frac{\partial}{\partial \eta^i_0} \delta \eta^j_0(\theta, \eta_0, t)
\]
\[
= \delta \eta^i_0 \frac{\partial}{\partial \eta^i_0} \delta \eta^j_0(\theta, \eta_0, t)
\]
\[
= -\delta \eta^i_0 \frac{\partial^2 \tilde{L}}{\partial \eta^i \partial n_j}
\]

Eq. (48) is used to arrive at the third form of each equation in terms of mixed partials of \(L\) and \(\tilde{L}\).

The coordinate transformation (7) from contravariant exponential coordinates to covariant mixture coordinates may be used in two ways to write the inner product of vector fields \(\delta \theta\) and \(\delta \eta\) in mixed form. From the definition of the inner product in terms of \(g^D\) in Eq. (69) and its equivalence to the Hessian definition of \(g\) in Eq. (71),

\[
(\delta \theta)^i g_{ij}(\delta \eta)^j \equiv (\delta \theta)^i (\delta \eta n)_j \equiv (\delta \theta n)_j (\delta \eta)^i
\]

(76)

Although the field variable \(n\) is the same in either log-transform (45) or (55), the two displacements \(\delta \eta n\) and \(\delta \theta n\) are independent vector fields.

3.4.1 The conserved inner product of dual vector fields, and directional transport of the metric

Equation (75) has a symmetric form but evolves \(\delta \theta\) and \(\delta \eta\) respectively using \(L\) and \(\tilde{L}\), making it not immediately apparent that the inner product is preserved. Writing the field \(\delta \eta\) in its dual mixture coordinate as in the first line of Eq. (76) the time derivative becomes

\[
\left(\frac{d}{dt}\delta n\right)^i = \delta n_{0j} \frac{\partial}{\partial n_{0j}} \dot{n}_i(\theta, n_0, t)
\]
\[= \delta n_{ij} \frac{\partial}{\partial n_{ij}} \dot{n}_i(\theta_T, n_0, t)\]
\[= -\delta n_{ij} \frac{\partial^2 L}{\partial n_j \partial \theta^i} \] (77)

The condition (18) is met and we have

\[
\left(\frac{d}{dt} \delta \theta\right)^j_j (\delta n)^i_i + (\delta \theta)^i_i \left(\frac{d}{dt} \delta n\right) = 0 \] (78)

Using Eq. (78) to evaluate the change in the inner product written as \((\delta \theta)^i g_{ij} (\delta \eta)^j\), substituting the derivatives (48) for \(\dot{\theta}\) and \(\dot{\eta}\), and grouping terms, we obtain the transport equation for the metric along stationary paths

\[0 = \dot{\theta}^k \frac{\partial g_{ij}}{\partial \theta^k} + \frac{\partial^2 L}{\partial \theta^i \partial n_k} g_{kj} + \frac{\partial g_{ij}}{\partial \eta^k} \dot{\eta}^k - g_{ik} \frac{\partial^2 \tilde{L}}{\partial n_k \partial \eta^j} \]
\[= \left(\frac{\partial L}{\partial n_k} \frac{\partial g_{ij}}{\partial \theta^k} + \frac{\partial^2 L}{\partial \theta^i \partial n_k} g_{kj}\right) - \left(\frac{\partial g_{ij}}{\partial \eta^k} \frac{\partial \tilde{L}}{\partial n_k} + g_{ik} \frac{\partial^2 \tilde{L}}{\partial n_k \partial \eta^j}\right) \]
\[= \frac{\partial}{\partial \theta^i} \left(\dot{\theta}^k g_{kj}\right) + \frac{\partial}{\partial \eta^j} \left(g_{ik} \dot{\eta}^k\right) \] (79)

The tensor transport equation from Eq. (79) can be compared to Eq. (65) for the transport of the Wigner density.

### 3.5 Dual connections respecting the symplectic structure of canonical transformations in the two-field system

The transport relations derived so far make use of the symplectic structure of maps generated by time-translation along Doi-Peliti stationary paths, but they are not specifically geometric. We now turn to geometric constructions that respect the symplectic structure, render its maps coordinate invariant under canonical transformations, and express the special roles of affine transport in some coordinates such as coherent states through the definition of appropriate Riemannian connections.

#### 3.5.1 Conservation of the inner product through the combined effects of two maps

The inner product (76) is preserved through the complementary action of two maps, one generated by the time-dependence of \(\theta\), and the other by the time-dependence of \(\eta\). By construction, \(\delta \theta\) depends on time only through \(\dot{\theta}\), and \(\delta \eta\) only through \(\dot{\eta}\), while the metric has no explicit time dependence but changes under both maps as the location \((\theta, \eta)\) changes. Denoting by \(d/dt|_{\dot{\theta}}\) and \(d/dt|_{\dot{\eta}}\) these separate components of change,
The time derivative of the inner product can be partitioned into two canceling terms:

\[
\left( \frac{d}{dt} \delta \theta \right)_i \delta \eta_j = \left( \frac{d}{dt} \delta \theta \right)_i g_{ij} (\delta \theta)^j + (\delta \theta)^i \dot{\theta}^k \frac{\partial g_{ij}}{\partial \theta^k} (\delta \eta)_j \equiv \left( \frac{d}{dt} \delta \eta \right)_i g_{ij} \left( \frac{d}{dt} \delta \eta \right)_j + \dot{\eta}^k \frac{\partial g_{ij}}{\partial \eta^k} (\delta \eta)^j
\]

(80)

Connection coefficients may be added within either \( \left( \frac{d}{dt} \delta \theta \right)_j \) or \( \left( \frac{d}{dt} \delta \eta \right)_i \) to make the components of change in the vector field and metric coordinate-invariant, without altering the duality between independent variations in the base distribution and in the likelihood ratio.

To introduce a connection we first replace the total derivative \( \frac{d}{dt} \) with a partial-derivative decomposition expressing the same transformation as a flow:

\[
\left( \frac{d}{dt} \delta \theta \right)_i \equiv \frac{\partial}{\partial t} (\delta \theta)_i + \dot{\theta}^k \frac{\partial}{\partial \theta^k} (\delta \theta)_i
\]

\[
\left( \frac{d}{dt} \delta \eta \right)_j \equiv \frac{\partial}{\partial t} (\delta \eta)_j + \dot{\eta}^k \frac{\partial}{\partial \eta^k} (\delta \eta)_j
\]

(81)

Connection coefficients are defined from the pullbacks \( \left( \frac{\partial}{\partial \theta^k} \right)^i \) or \( \left( \frac{\partial}{\partial \eta^k} \right)^i \) of infinitesimally transformed basis vectors in the tangent spaces to the two exponential families,

\[
\frac{\partial}{\partial \theta^k} \left( \frac{\partial}{\partial \theta^i} \right)^j \equiv \Gamma^{(\theta)}_{ki} \left( \frac{\partial}{\partial \theta^i} \right)^j \quad \frac{\partial}{\partial \eta^k} \left( \frac{\partial}{\partial \eta^i} \right)^j \equiv \Gamma^{(\eta)}_{ki} \left( \frac{\partial}{\partial \eta^i} \right)^j
\]

(82)

Superscripts \( \Gamma^{(\theta)} \) or \( \Gamma^{(\eta)} \) refer to the subspace of basis vectors \( \partial/\partial \theta \) or \( \partial/\partial \eta \) being pulled back, and the designation \( \Gamma \) or \( \Gamma^* \) distinguishes the connection associated with \( \theta \) displacement or \( \eta \) displacement, respectively. Because the \( \dot{\eta} \) component of time translation does not act in \( \delta \theta \) and vice versa, we set connection coefficients \( \Gamma^{(\eta)}_{kj} \) and \( \Gamma^{(\theta)*}_{kj} \) to zero.

Covariant derivatives of the vector fields \( \delta \theta \) and \( \delta \eta \) in the connections \( \Gamma, \Gamma^* \) of Eq. (82) are defined as

\[
\left( \nabla^{(\theta)}_k \delta \theta \right)^i = \frac{\partial}{\partial \theta^k} (\delta \theta)^i + \Gamma^{(\theta)}_{kj} (\delta \theta)^j \quad \left( \nabla^{(\eta)}_k \delta \theta \right)^i = \frac{\partial}{\partial \eta^k} (\delta \theta)^i
\]

\[
\left( \nabla^{(\theta)*}_k \delta \eta \right)_j = \frac{\partial}{\partial \theta^k} (\delta \eta)_j \quad \left( \nabla^{(\eta)*}_k \delta \eta \right)^i = \frac{\partial}{\partial \eta^k} (\delta \eta)_j + \Gamma^{(\eta)*}_{ki} (\delta \eta)^i
\]

(83)
The covariant part of the flow decomposition in Eq. (81) is defined by subtraction of the nonzero connection coefficients from the total derivatives (75), as

\[
\left(\frac{\partial}{\partial t} \delta \theta \right)^j + \dot{\theta}^k \left(\nabla_k^{(\theta)} \delta \theta \right)^j + \dot{\eta}^k \left(\nabla_k^{(\eta)} \delta \theta \right)^j = \left(\frac{d}{dt} \delta \theta \right)^j + \dot{\theta}^k \Gamma_{ki}^{(\theta)j} (\delta \theta)^i \\
\left(\frac{\partial}{\partial t} \delta \eta \right)^j + \dot{\theta}^k \left(\nabla_k^{(\theta)\ast} \delta \eta \right)^j + \dot{\eta}^k \left(\nabla_k^{(\eta)\ast} \delta \eta \right)^j = \left(\frac{d}{dt} \delta \eta \right)^j + \dot{\eta}^k \Gamma_{ki}^{(\eta)\ast j} (\delta \eta)^i
\]  

(84)

Compensating covariant derivatives of the metric are

\[
\nabla_k^{(\theta)} g_{ij} = \frac{\partial}{\partial \theta^k} g_{ij} - \Gamma_{ki}^{(\theta)l} g_{lj} \\
\nabla_k^{(\eta)\ast} g_{ij} = \frac{\partial}{\partial \eta^k} g_{ij} - \Gamma_{kj}^{(\eta)\ast l} g_{il}
\]  

(85)

Equation (84) extracts a coordinate-invariant component of the time derivative of vector fields \(\delta \theta\) and \(\delta \eta\) under canonical transformations, while Eq. (85) extracts the corresponding coordinate-invariant part of the change in the Fisher metric.

### 3.5.2 Referencing arbitrary dual connections to dually flat connections in the exponential family

The manifold for a Doi-Peliti system with \(D\) independent components has dimension \(2D\), with parallel subspaces for the base distribution and likelihood ratio. The dual connections (82) act within these two independent subspaces, in contrast to the dually-flat connections \(\Gamma_1^D\) and \(\Gamma_1^{D\ast}\) of Eq. (70), which act within the same \(D\)-dimensional exponential family. Although the Fisher metric is a function only of the overall importance distribution, which aggregates dependence from the base distribution and likelihood, the symplectic transformations from translation along stationary paths separate components of variation from within the two independent subspaces. The subspace decomposition cannot be recovered from the importance distribution alone, and thus no connection defined only from the properties of the Fisher metric is sufficient to identify the dual symplectic connections for a Doi-Peliti system.

Nonetheless, we may relate the symplectic dual connections to Amari’s dually flat connections and the Amari–Chentsov tensor through the relation (see [2, Eq. (6.27)])

\[
\partial_k \Gamma_{ij}^D = \Gamma_{kij}^D + \Gamma_{kji}^{D\ast} = 0 + T_{kji}
\]  

(86)

Substituting Eq. (86) into Eq. (85) gives expressions for the dual covariant derivatives of the Fisher metric

\[
\nabla_k^{(\theta)} g_{ij} = T_{kji} - \Gamma_{kij}^{(\theta)} \\
\nabla_k^{(\eta)\ast} g_{ij} = T_{kji} - \Gamma_{kij}^{(\eta)\ast}
\]  

(87)

### 3.5.3 Flat connections for coherent-state coordinates

Of particular interest in Doi-Peliti theory will be the canonical transformations (45) and (55) between coherent-state and number-potential coordinates. We note that the
forms of the connection coefficients for which affine transport in fields $\phi^\dagger$ is flat in the likelihood subspace, and affine transport in fields $\varphi$ is flat in the base-distribution subspace, are\(^4\):

\[
\Gamma_{kj}^{(\theta)^i} = \frac{\partial \theta^i}{\partial \varphi^j} \frac{\partial}{\partial \varphi^k} \left( \frac{\partial \varphi^l}{\partial \theta^j} \right) \quad \Gamma_{kj}^{(\theta)^s} = 0
\]

\[
\Gamma_{kj}^{(\eta)^i} = \frac{\partial \eta^i}{\partial \phi^j} \frac{\partial}{\partial \phi^k} \left( \frac{\partial \phi^l}{\partial \eta^j} \right) \quad \Gamma_{kj}^{(\eta)^s} = 0
\]

(90)

### 3.6 On the roles of coherent-state versus number-potential coordinates in the Doi-Peliti representation

The Doi-Peliti solution method is almost always introduced through the coherent-state representation [9, 42, 53], and for many applications such as chemical reaction networks [8, 10, 44, 71] or evolutionary population processes [70], coherent states are also the “native” representation in the sense that the Liouville operator is a finite-order (generally low-order) polynomial in fields. Moreover, for the importance-sampling interpretation emphasized in this paper, the coherent-state representation separates the nominal distribution and likelihood ratio.

On the other hand, Legendre duality is defined with respect to potential fields, which are the tilt coordinates $\theta$ in the exponential family of importance distributions, and it is in these coordinates, not the coherent-state coordinates, that the Fisher metric corresponds to the Hessian of the CGF. Indeed, it is not generally possible to define a dual coordinate system from the Hessian of the CGF in coherent-state fields, as we illustrate for the worked example in Sect. 4.5.

The use of Riemannian connections neatly expresses the role of each coordinate system. The elementary eigenvalues of divergence or convergence of bases and tilts, and of information susceptibilities, are often simple in coherent-state coordinates, where they are eigenvalues of coordinate divergence or convergence. In the dual connections (90), covariant derivatives retain these elementary eigenvalues, while inheriting from the exponential family the Fisher geometry that defines contravariant/covariant coordinate duality. A concrete example is given in the next section.

\(^4\) In the basis of the original species counts, by Eqs. (45) and (55), the measures between number-potential and coherent-state basis vectors are

\[
\frac{\partial}{\partial \theta^i} = \phi^{\dagger i} \frac{\partial}{\partial \phi^{\dagger i}} \quad \frac{\partial}{\partial \eta^i} = \phi^{\dagger i} \frac{\partial}{\partial \eta^i} \quad \frac{\partial}{\partial \eta^i} = e^{-\theta^i} \frac{\partial}{\partial \eta^i} \quad \frac{\partial}{\partial \phi^{\dagger i}} = e^{-\eta^i} \frac{\partial}{\partial \eta^i}
\]

(88)

Therefore the connection coefficients (90) in the species basis are

\[
\Gamma_{ki}^{(\theta)^j} = \delta_{ki} \delta_{i}^j \quad \Gamma_{ki}^{(\eta)^j} = \delta_{ki} \delta_{i}^j
\]

(89)

For various reasons it will, however, not be convenient to work in the species basis, and the resulting connection coefficients will not generally be constant.
4 A worked example: the two-state linear system

The foregoing constructions are nicely illustrated in minimal form in a simple, exactly solvable model. It is the stochastic process for \(N\) independent random walkers on a network with two states and bidirectional hopping between them. The statistical mechanics of transients, time-dependent generating functionals, and large deviations for this system has been didactically covered within the Doi-Peliti framework in [68]. Though simple, the model is nonetheless rich enough to illustrate the complementary roles of coherent states and number-potential coordinates in Doi-Peliti theory—the former as the “native” coordinates in which the system is simple, and the latter as the coordinate system carrying the Fisher geometry—and the way this relation is captured by the dual coherent-state connection (90) different from both the Levi-Civita connection and the dually-flat connections (71) of Nagaoka and Amari [3, 54].

4.1 Two-argument and one-argument generating functions on distributions with a conserved quantity

The two-state model describes (for example) a one-particle chemical reaction in a well-mixed reactor with the schema

\[
\begin{align*}
k_+ &\quad a \rightleftharpoons b \\
k_- &
\end{align*}
\]  

(91)

The probability per unit time for a reaction event is given by rate constants \(k_+\) and \(k_-\), and proportional sampling (the microphysics underlying mass-action rate laws).

A distribution initially in binomial form (36) will retain that form at all times under the master equation for the schema (91), even with time-dependent coefficients. Here for simplicity we will take \(k_+\) and \(k_-\) to be fixed. Therefore the distribution at any time is specified by descaled mean values \(\nu_a = \langle n_a \rangle / N\), \(\nu_b = \langle n_b \rangle / N\), with \(\nu_a + \nu_b = 1\).

Although the system has only one dynamical degree of freedom, it is instructive to compute both the two-argument generating function with independent weights \(z_a\) on \(n_a\) and \(z_b\) on \(n_b\), and the 1-argument generating function for the difference coordinate \(n \equiv (n_b - n_a) / 2\), to illustrate the role of conservation laws and the geometry of the coherent-state connection. The two-argument CGF (1) for the binomial distribution is

\[
e^{\psi(\log z)} = \sum_n z_a^n z_b^{n_b} \rho_{n_a, n_b} = [z_a \nu_a + z_b \nu_b]^N
\]  

(92)

Because the total number \(N = (n_b + n_a)\) is fixed, the normalized 1-variable distribution may be written
\[ \rho_n = \sqrt{v_b v_a}^N \left( \frac{N}{n} \right) \left( \frac{v_b}{v_a} \right)^n \]  
\[ \text{(93)} \]

and the terms in the generating function (92) regrouped as

\[ e^{\psi(\log z)} = \sqrt{z_b z_a}^N \sum_n \rho_n \left( \frac{z_b}{z_a} \right)^n \]  
\[ \text{(94)} \]

Introducing rotated coordinates on the exponential family of tilts

\[ h \equiv (\theta_b + \theta_a) / 2 \quad \theta \equiv (\theta_b - \theta_a) \]  
\[ \text{(95)} \]

and dividing the two-argument MGF (94) by \( \sqrt{z_a z_b}^N \), we obtain an expression for the one-argument MGF in the difference coordinate n:

\[ e^{\psi(\log z) - Nh} = \sum_n \rho_n e^{\theta n} \]  
\[ \text{(96)} \]

In what follows, \( \psi (\log z) \) will always be used to refer to the two-argument CGF (92), and the 1-argument generating function, when needed, will be written out explicitly as \( \psi (\log z) - Nh \), as in Eq. (96).

### 4.2 Generator and conserved volume element in coherent-state coordinates

The master equation for the two-state system is developed in [68], but introduces further notation, and will not be needed here. We move directly to the expression for the Liouville function of Eq. (44) after conversion to field variables, which is

\[ L = k_+ \left( \phi_a^\dagger - \phi_b^\dagger \right) \phi_a + k_- \left( \phi_b^\dagger - \phi_a^\dagger \right) \phi_b \]  
\[ \text{(97)} \]

In what follows, math boldface will be reserved for parameters in the generator such as \( k_{\pm} \) or functions of these such as the associated steady states used in Eq. (53).

Two descalings reduce the problem to parameters which are dimensionless ratios. The first defines a time coordinate \( \tau \) in units of the sum of rate parameters,

\[ \frac{d\tau}{dt} \equiv k_+ + k_- \]  
\[ \text{(98)} \]

The second expresses the equilibrium steady state under generator (97) in terms of relative hopping rates,

\[ \frac{k_+}{k_+ + k_-} = \frac{n_b}{N} \equiv v_b \quad \frac{k_-}{k_+ + k_-} = \frac{n_a}{N} \equiv v_a \]  
\[ \text{(99)} \]

As for the discrete index n, define \( v \equiv (v_b - v_a) / 2 \).
Conservation of total number $N$ results in a generator $\mathcal{L}$ that is a function only of the difference variable $(\phi_b^\dagger - \phi_a^\dagger)$. Therefore it is natural to rotate the coherent-state fields to components corresponding to conserved $N$ and dynamical $n/N$, and their dual coordinates in the generating-function argument $z$:

$$\phi^\dagger \equiv \phi_b^\dagger - \phi_a^\dagger \quad \hat{\phi} \equiv (\phi_b - \phi_a) / 2N$$

$$\Phi^\dagger \equiv \left(\phi_b^\dagger + \phi_a^\dagger\right) / 2 \quad \hat{\Phi} \equiv (\phi_b + \phi_a) / N \quad (100)$$

In rotated fields (100) the action (44) becomes

$$S = N \int d\tau \left[ -\partial_\tau \Phi^\dagger \hat{\Phi} - \partial_\tau \phi^\dagger \hat{\phi} + \phi^\dagger \left(\hat{\phi} - \nu \hat{\Phi}\right) \right]$$

$$\equiv N \int d\tau \left( -\partial_\tau \Phi^\dagger \hat{\Phi} - \partial_\tau \phi^\dagger \hat{\phi} + \hat{\mathcal{L}} \right) \quad (101)$$

A descaled Liouville function has been introduced as $N (k_+ + k_-) \hat{\mathcal{L}} \equiv \mathcal{L}$. Absence of the field $\Phi^\dagger$ from $\hat{\mathcal{L}}$ implies constancy of the expectation for $\hat{\Phi}$.\(^5\)

### 4.2.1 Splitting the symplectic structure between coherent-state conjugate field pairs

Although $\hat{\Phi}$ obeys certain time-translation invariances in correlation functions, its value even along stationary paths will not generally be 1. Therefore the coherent-state variables cannot directly be interpreted as mean values of number variables in the nominal distribution or mean weights in its dual likelihood ratio. To express the functions that are these expectation values, we recall the mean number components in the importance distribution, which are bilinear quantities in $\phi^\dagger$ and $\phi$, and then introduce a pair of dual number coordinates that, while not linear functions of the coherent-state fields, are functions respectively of $\phi^\dagger$ or of $\phi$ extracted by making use of the steady-state measure under the instantaneous value $\nu$ in the generator (97). (Along stationary paths, where some components of $\phi^\dagger$ or $\phi$ are invariant, these dual number fields will become linear functions of the remaining dynamical components of $\phi^\dagger$ or $\phi$, as we show below.)

The two components of the normalized number field in number-potential coordinates (45) are given by

$$\frac{1}{N} \phi_b^\dagger \phi_b \equiv \nu_b \equiv \left(\frac{1}{2} + \nu\right) \quad \frac{1}{N} \phi_a^\dagger \phi_a \equiv \nu_a \equiv \left(\frac{1}{2} - \nu\right) \quad (102)$$

Recall that the instantaneous steady state under the generating process is the scale variable for the dualizing canonical transform (53). To see how this reference steady

\(^5\) It implies constancy of a tower of higher-order correlation functions expressing exact conservation of the underlying variable $N$, though we do not develop the 2FFI representation of correlation functions in this paper.
state is used to separate the two conjugate variables (base and tilt) in the symplectic transformations, it is helpful to recast Eq. (102) as
\[
\frac{1}{2} \left( \phi_b^\dagger \phi_b - \phi_a^\dagger \phi_a \right) \equiv \frac{(v_b - v_a)}{2} \equiv v
\] (103)

The action of the tilt alone can be isolated, without regard to the underlying nominal distribution, by referencing the action of the \( \phi^\dagger \) fields to the steady state rather than to \( \phi \), defining an offset \( \nu \) as
\[
\frac{1}{2} \left( \phi_b^\dagger \nu_b - \phi_a^\dagger \nu_a \right) \equiv \frac{(\nu_b - \nu_a)}{2} \equiv \nu
\] (104)

Likewise, the mean value \( \bar{\nu} \) of \( n/N \) in the base (nominal) distribution is isolated by referencing the value of \( \phi \) to the uniform measure 1 instead of the dynamic measure \( \phi^\dagger \), as
\[
\frac{1}{2} \left( \phi_b - \phi_a \right) \equiv \frac{(\bar{\nu}_b - \bar{\nu}_a)}{2} \equiv \bar{\nu}
\] (105)

### 4.2.2 Stationary-path solutions and Liouville volume element

Solutions to the stationary-path equations of motion (47) for the Liouville function (97) are evaluated in Appendix F.1.

Stationary-path approximations to the time-dependent density \( \rho \) would be binomial distributions even if the exact \( \rho \) were not (the stationary point is always a pure coherent state), so the CGF at any time has the form (92), with fields \( z \) replaced by the stationary-path values of \( \phi^\dagger \) and the mean values \( \nu \) from Eq. (93) replaced by corresponding components of \( \phi \).

In particular, the initial-time generating function \( \psi_0 \left( \log \phi_{a0}^\dagger, \log \phi_{b0}^\dagger \right) \) appearing in Eq. (43) carries the mean value \( \bar{\nu}_0 \) in the starting density \( \rho_0 \), imposed as an initial-data parameter. It is through this function that the final-time tilt data in the form of the parameter \( \nu_T \), propagated backward to the stationary-path values of \( \phi_{a0}^\dagger \) and \( \phi_{b0}^\dagger \), determines the stationary path values for the fields \( \phi \) of the base distribution, establishing the potential for information coupling between initial properties of the base distribution and final-time queries in the generating function \( \psi_T \).

\( \psi_0 \) is evaluated in Eq. (190), and the value is shown to depend only on an overlap parameter between initial and final data which we denote
\[
\Lambda \equiv \frac{(\bar{\nu}_0 - \nu) (\nu_T - \nu)}{\left( \frac{1}{4} - \nu^2 \right)}
\] (106)
Fig. 2 Four trajectories (heavy blue contours), plotted in coordinates $\nu, \bar{\nu}$, that bound a region specified by $\nu_T = -1/6 \pm 0.075$, $\bar{\nu_0} = -1/4 \pm 0.05$. The steady state under the generating process sets $\nu = 1/6$. A time interval $T = 3$ between the input distribution and the final-time generating function is shown. Small rectangles (heavy dark red) show the area $\delta \nu \delta \bar{\nu}$ at five equally spaced times from start to end. The outer four trajectories (thin blue) show the possible range of joint images of $-1/2 \leq \bar{\nu_0} \leq 1/2$ and $-1/2 \leq \nu_T \leq 1/2$. Large rectangles (thin green) show the constriction of the possible range $\propto e^{-T}$. Projections of the total range and the inner trajectories are shown in thin lines on the base plane. Shading of the base plane is a grayscale plot of $\Phi_1^2$, which is constant along trajectories but variable over the $(\nu, \bar{\nu})$ coordinate range. Min and max of $\Phi_0^2$ are respectively 0.83 and 1.1

The stationary-path values of the displacement coordinates (104) and (105) are shown in Eq. (194) to follow simple exponential laws

$$\nu - \bar{\nu} = (\bar{\nu_0} - \nu) e^{-T} \quad \bar{\nu} - \nu = (\nu_T - \nu) e^{T-T}$$

Thus under independent variations of $\bar{\nu_0}$ and $\nu_T$ as described in Sect. 3.4, the trajectories of the coherent state fields $\phi^\dagger$ and $\phi$ trace out an invariant volume, illustrated in Fig. 2.

4.2.3 Invariant cumulant-generating function and the incompressible phase-space density

The stationary value of $\Phi_0$, obtained from the gradient of $\psi_0$ with respect to the components $\phi^\dagger_{0a}$ and $\phi^\dagger_{0b}$, is computed in Eq. (193). It differs from unity—the reason constructions (104) and (105) were needed—and it is equal to the stationary value of $\hat{\Phi}$ at all times as a consequence of conservation of total number $N$. The value depends only on $\Lambda$ and $T$ in the combination

$$\Phi_0 = \frac{1}{1 + \Lambda e^{-T}}$$
Moreover, as a consequence of the conserved Liouville volume element from Eq. (107), the stationary-point evaluation of the CGF at all times takes the same form as Eq. (92) and evaluates to the constant

\[ \frac{\psi}{N} = \log \left( \frac{\nu_a \ddot{v}_a + \nu_b \ddot{v}_b}{\nu_a} \right) = -\log \hat{\Phi}_0 \] (109)

\( \hat{\Phi}_0 \) in Eq. (108) is the basis for all information densities in this simple linear system. Through the stationary-point relation (61) between the Wigner function and the CGF, \(-\log \hat{\Phi}_0\) is the incompressible phase-space density convected along stationary trajectories by Eq. (64). As shown below, it is also the geometrically invariant part of sole nonzero eigenvalue of the Fisher metric.

### 4.3 Fisher metric

The Fisher metric (5) for the two-state system evaluates, along the stationary path at any time, to

\[ g = \frac{\partial \nu}{\partial \theta} \left[ \begin{array}{cc} 1 & -1 \\ -1 & 1 \end{array} \right] \] (110)

The nonzero eigenvalue comes from the single-argument generating function in Eq. (96) for the difference coordinate \( n \), and the zero eigenvalue comes from the linear CGF \( hN \) for the conserved quantity \( N \).

The term \( \partial \nu/\partial \theta \) in Eq. (110) may be converted, after some algebra, to the form

\[ \frac{\partial \nu}{\partial \theta} = \left( \frac{1}{4} - \nu^2 \right) \left( \frac{1}{4} - \ddot{\nu}^2 \right) \hat{\Phi}_0^2 \] (111)

The measure terms \( (1/4 - \nu^2) \) and \( (1/4 - \ddot{\nu}^2) \) appearing in Eq. (111) follow the divisions (104) and (105) into independent dimensions of base and tilt variation, and we will show that their effects are canceled in an appropriate covariant derivative. The remaining dependence of the eigenvalue on the initial and final data is all carried in \( \hat{\Phi}_0^2 \).

### 4.4 Dual coordinates for base and tilt, and the additive exponential family

To relate the Fisher metric in Eq. (110) to the construction of Sect. 3.3 from the \( \psi \)-divergence and to dually-symplectic parallel transport, we first express the base and tilt displacements (104) and (105) in terms of the coordinates in their respective exponential families.
Introduce reference values for the fields $\theta$ and $h$ defined in Eq. (95), corresponding to the steady-state measure under the parameters of the generating process, denoted

$$\theta \equiv \log \left( \frac{\frac{1}{2} + \nu}{\frac{1}{2} - \nu} \right)$$

$$h \equiv \frac{1}{2} \log \left( \frac{1}{4} - \nu^2 \right) = -\log \left[ \frac{2 \cosh \theta}{2} \right]$$

(112)

It is clear, in the two-argument generating function (92), that one component of variation in $z$ couples only to the conserved quantity $N$ and is not needed. It is sufficient therefore to vary along an affine coordinate in $z$ that couples to the dynamical argument $n$, and the natural choice is to fix the component of $z$ corresponding to the component of $\phi^\dagger$ that is invariant under the stationary-path equations of motion, given in Eq. (189). The resulting contour for $z$ at final time $T$ becomes

$$z_b \nu_b + z_a \nu_a = \Phi_T^\dagger + \nu \phi_T^\dagger = 1 \quad z_b - z_a = \phi_T^\dagger = \frac{\nu_T - \nu}{\frac{1}{4} - \nu^2}$$

(113)

The quantity in the first line of Eq. (113) is preserved at all times by Eq. (189), and the quantity in the second line obeys the exponential law of Eq. (194) repeated as (107).

By the definition (104), the contour (113) which is affine in coherent-state fields $\phi^\dagger$ is written in the coordinates on the exponential family of tilts as

$$\nu \equiv \frac{1}{2} \tanh \left( \frac{\theta + \eta}{2} \right)$$

$$0 \equiv h + h + \log \left[ 2 \cosh \left( \frac{\theta + \eta}{2} \right) \right]$$

$$= h - \frac{1}{2} \log \left( \frac{1}{4} - \nu^2 \right) + \frac{1}{2} \log \left( \frac{1}{4} - \nu^2 \right)$$

(114)

Likewise, in the dual exponential representation (55) of the family of base distributions, the definition (105) giving the mean number offset in the nominal distribution is expressed

$$\bar{\nu} = \frac{1}{2} \tanh \left( \frac{\eta + \theta}{2} \right)$$

(115)

in which $\eta \equiv (\eta_b - \eta_a)$ in the dual number-potential system (55), analogously to $\theta$ in Eq. (95).

Because the two exponential coordinates (base and tilt) are additive, the mean of samples in the importance distribution can likewise be written

$$\nu = \frac{1}{2} \tanh \left( \frac{\theta + \eta + \theta}{2} \right)$$

(116)
It follows that the eigenvalue (111) in the Fisher metric also has the simple expression

$$\frac{\partial \nu}{\partial \theta} = \frac{\partial \nu}{\partial \eta} = \frac{1}{4} - \nu^2 \tag{117}$$

exhibiting the equivalence of the $\psi$-divergence expression (69) and the Hessian (71) for this quantity.

### 4.5 Why coherent-state fields do not generally produce invertible coordinate transformations

The Hessian matrix is not a tensor under coordinate transform, so it is clear that the Hessian of $\psi$ with respect to the argument $z$ equivalent to the coherent-state response field $\phi^\dagger$ will not be the Fisher metric. However, since coherent states are in many ways a native basis for Doi-Peliti theory, as noted in Sect. 3.6, we may ask whether some other coordinate duality can be defined from the coherent-state Hessian of $\psi$. In fact such a duality cannot generally be defined, and it is instructive to see where it fails, to better understand why the affine connection (90) and not the Fisher geometry captures the special role of coherent states.

A divergence under the Hessian of $\psi$ in coherent-state variables, which we will denote $\Delta \delta s^2$ for reasons to become clear in a moment, if converted from the coordinates $\nu$ to coordinates $\theta$ along the $z$-affine contour (114), evaluates as

$$\frac{1}{N} \Delta ds^2 \equiv (\delta \theta)^2 \left( \frac{1}{4} - \nu^2 \right)^2 \frac{\partial^2}{\partial \psi^2} \left( \frac{\psi}{N} \right) \approx - (\delta \theta)^2 (\nu - \nu)^2 \tag{118}$$

Unlike the Fisher metric, Eq. (118) is negative-semidefinite, and degenerates if $\nu = \nu$, which is shown in Eq. (197) to hold for all $\theta$ if $\nu_0 = \nu$. At degenerate solutions, we cannot use the Hessian of $\psi$ to define a base-field variation $\delta \phi$ as a dual coordinate for a variation produced by a field $\delta \phi^\dagger$, as we could use the Hessian in the exponential family to produce a variation $\delta \nu$ as a dual coordinate to a variation $\delta \theta$.

The source of the degeneration has a nice description in terms of intrinsic and extrinsic curvatures, and advection, in the natural geometry on the exponential family. The geometric distance element (4), with $\theta$ and $h$ varied independently, is

$$\frac{1}{N} \delta s^2 = (\delta \theta)^2 \frac{\partial^2}{\partial \theta^2} \left( \frac{\psi}{N} - h \right) + (\delta h)^2 \frac{\partial^2}{\partial h^2} h$$

$$= (\delta \theta)^2 \frac{\partial \nu}{\partial \theta} + (\delta h)^2 0 \tag{119}$$

The $z$-affine contour (114) specifies a function $h(\theta)$ with extrinsic curvature in the affine coordinate manifold of the exponential family, along which the distance element is

$$\frac{1}{N} \delta s^2_{\text{CS-\extr}} = (\delta \nu)^2 \frac{d^2}{d\nu^2} (-h(\theta)) = (\delta \theta)^2 \left( \frac{1}{4} + \nu^2 \right) \tag{120}$$
The second coherent-state coordinate derivative of $\psi$ along the contour (114) can be decomposed as

$$\frac{1}{N} \Delta s^2 = (\delta \nu)^2 \frac{d^2}{d\nu^2} \left( \frac{\psi}{N} \right)$$

$$= (\delta \nu)^2 \left[ \frac{d^2}{d\nu^2} \left( \frac{\psi}{N} - h \right) + \frac{d^2}{d\nu^2} h \right]$$

$$= (\delta \theta)^2 \left[ \frac{d^2}{d\theta^2} + 2\nu \frac{d}{d\theta} \right] \left( \frac{\psi}{N} - h \right) + (\delta \nu)^2 \frac{d^2}{d\nu^2} h$$

$$= (\delta \theta)^2 \left[ \frac{\partial \nu}{\partial \theta} + 2\nu \right] + (\delta \nu)^2 \frac{d^2}{d\nu^2} h$$

$$= \frac{1}{N} \left( \delta s^2 - \delta s_{\text{CS-ext}}^2 \right) + (\delta \theta)^2 2\nu \nu$$

(121)

With some algebra, the expression (121) is shown to equal that in Eq. (118). The extrinsic curvature of the embedded contour $h(\theta)$ and the convected quantity $-2\nu \nu$ cancel against the intrinsic Fisher-Rao curvature, rendering the duality invisible to the fields $\phi^\dagger$ at degenerate points.

### 4.6 Flat transport in the coherent-state connection

The correct way to capture the simplifying role of coherent-state coordinates for simple models such as the two-state system is with the dual connections of Sect. 3.5.

We first recognize, from the forms (106) or (195) of $\Lambda$, a completely-descaled coordinate system for the dynamical parts of the coherent-state fields, defined by

$$v \equiv \frac{(\nu - \nu)}{\sqrt{\frac{1}{4} - \nu^2}} \quad u \equiv \frac{\bar{\nu} - \nu}{\sqrt{\frac{1}{4} - \nu^2}}$$

(122)

The eigenvalue of the Fisher metric in Eq. (111) then reduces to

$$\frac{\partial^2}{\partial \theta^2} \left( \frac{\psi}{N} \right) = \frac{(d\nu/d\theta)(d\nu/d\eta)}{[1 + \nu u]^2}$$

(123)

The role of the factors $\left(\frac{1}{4} - \nu^2\right) = \sqrt{\frac{1}{4} - \nu^2} (\partial \nu/\partial \theta)$ and $\left(\frac{1}{4} - \bar{\nu}^2\right) = \sqrt{\frac{1}{4} - \nu^2} (\partial \bar{\nu}/\partial \eta)$ in Eq. (111) as measure terms is now explicit, and they can be absorbed by a change of variables to $u$ and $v$. By Eq. (195) and the definitions (122) and (108),

$$[1 + \nu u] = [1 + \Lambda e^{-T}] = 1/\Phi_0$$

so the Fisher inner product (12) may be written

$$\delta \theta \delta \eta h \left[ \frac{\partial}{\partial \theta}, \frac{\partial}{\partial n} \right] = \nu \delta \bar{\nu} \left[ \frac{\partial}{\partial \nu}, \frac{\partial}{\partial \bar{\nu}} \right] = \nu \delta \bar{\nu} \Phi_0^2$$

(124)
4.6.1 Connection coefficients and absorption of measure terms

In this linear model, time evolution of $\phi^\dagger$ and $\phi$ has no cross-dependence once the initial values have been fixed through the gradients of $\psi_0$ as explained in Sect. 4.2.2. Thus $\left(\nabla_k^{(\eta)} \delta \theta\right)^j = 0$ and $\left(\nabla_k^{(\theta)} \delta \eta\right)^j = 0$.

Appendix F.4 computes connection coefficients and covariant derivatives for the vector fields corresponding to Eq. (84), and for the metric tensor corresponding to Eq. (85). Eq. (205) in the appendix gives the covariant part of the time derivatives of $\delta \theta$ and $\delta \eta$ as

$$\left(\frac{\partial}{\partial \tau} \delta \theta\right) + \theta \left(\nabla_\theta \delta \theta\right) = \delta \theta$$

$$\left(\frac{\partial}{\partial t} \delta \eta\right) + \eta \left(\nabla_\eta \delta \eta\right) = -\delta \eta$$

(125)

capturing the simple exponential scaling (107) of the coherent-state fields in the exponential-family coordinates.

The covariant part of the change in the Fisher metric, from Eq. (85) is computed in Eq. (207) to be

$$\dot{\theta} \nabla_\theta g = \left(\dot{v} \frac{\partial}{\partial v} \log \hat{\Phi}_0^2\right) g$$

$$\dot{\eta} \nabla_\eta g = \left(\dot{u} \frac{\partial}{\partial u} \log \hat{\Phi}_0^2\right) g$$

(126)

Only the dependence in the Fisher eigenvalue $\hat{\Phi}_0^2$ from Eq. (124) appears.

The two lines of Eq. (126) (which are equal and opposite) scale as $\sim e^{-T}$, and have an interpretation similar to that of a Le Chatelier principle. The term $\Lambda e^{-T}$ in Eq. (108) for $\hat{\Phi}_0$ is a susceptibility of the initial stationary value $\phi_0$ to the perturbation by the tilt variable $\phi_T^\dagger = z$, attenuated exponentially from time $T$ to time 0. The role of this attenuation, which takes $\hat{\Phi}_0 \rightarrow 1$ as $T \rightarrow \infty$, becomes clearer as a constraint on the total extractable information when we consider in Sect. 4.7 the range of all initial distributions $\rho_0$ and all tilts $z$.

4.6.2 Duality of dynamics and inference in Doi-Peliti theory

The natural separation of the coordinate transformation of the inner product of vector fields $\delta \theta$ and $\delta \eta$ generated by time translation is not between exponential and mixture coordinates, as in the dually-flat connections of Amari [2], but rather between the symplectically dual contributions from changes in $\theta$ and in $\eta$. The two contributions group as

$$0 = \frac{d}{dt} (\delta \theta \ g \ \delta \eta)$$

$$= \left(\frac{\partial}{\partial t} \delta \eta\right) \delta \eta \nu + \dot{\theta} \nabla_\theta^{(\theta)} (\delta \theta \ g) \ \delta \eta + \delta \theta \nu \left(\frac{\partial}{\partial t} \delta \eta\right) + \delta \theta \ \dot{\eta} \nabla_\eta^{(\eta)^*} (g \ \delta \eta)$$

(127)
The two rows of Eq. (127) add covariant contributions from Eq. (125) and Eq. (126) in the combinations

\[
\begin{align*}
\left( \frac{\partial}{\partial t} \delta \theta \right) \delta \eta v + \hat{\theta} \nabla_{\theta}^{(\eta)} \left( \delta \theta \right) \delta \eta &= \left( \delta \theta \delta \eta \frac{\partial}{\partial u} \log \hat{\Phi}_0 \right) \\
\delta \theta v \left( \frac{\partial}{\partial \eta} \delta \eta \right) + \delta \theta \hat{\eta} \nabla_{\eta}^{(\eta)} \left( \delta \eta \right) &= \left( \delta \theta \delta \eta \frac{\partial}{\partial u} \log \hat{\Phi}_0 \right)
\end{align*}
\]

\[128\]

Eq. (128) captures in the clearest way possible the symplectic balance of distribution dynamics (through \( \eta \)) and inference (through \( \theta \)) in Doi-Peliti theory, through both the direct effects of the exponential growth and decay eigenvalues \((\pm 1)\) and the Le Chatelier-like susceptibility of the density \( \hat{\Phi}_0 \).

### 4.7 The Fisher information density and large-deviation ratios as sample estimators

The interpretation of the vector inner product as a convected density of information can be illustrated by using ratios of large-deviation probabilities to define a sample estimator for differences in the tilt coordinate \( \eta \) between two base distributions.

Suppose that we sample from a binomial nominal distribution at a parameter \( \eta \) that is to be estimated. Recall from Eq. (24) that the probability for the value \( n \) of a sample to exceed a threshold \( n \) is given in terms of the large-deviation function by

\[
P(n \geq n | \eta) \sim e^{-\psi^*(n; \eta)}
\]

\[129\]

In a 1-dimensional system, \(^6\) for two threshold values \( n_B > n_A \), the conditional probability for \( n \) to surpass \( n_B \) given that it has surpassed \( n_A \) is the ratio

\[
P(n_B | n_A; \eta) \equiv \frac{P(n \geq n_B | \eta)}{P(n \geq n_A | \eta)} \sim e^{-[\psi^*(n_B; \eta) - \psi^*(n_A; \eta)]}
\]

\[130\]

The ratio (130) can be estimated from samples of the indicator function \( h_n \) for thresholds \( n \) as described in Sect. 2.2.

Appendix F.5 shows that if two such conditional probabilities are compared from distributions at unknown parameters \( \eta_2 \) and \( \eta_1 \), the log ratio is related to the large-deviation thresholds and the \( \eta \) values as

\[
\log \left( \frac{P(n_B | n_A; \eta_2)}{P(n_B | n_A; \eta_1)} \right) \sim \int_{n_1}^{n_2} \int_{n_A}^{n_B} d\eta_2 d\eta_1 = (n_B - n_A) (\eta_2 - \eta_1)
\]

\[131\]

\(^6\) In one dimension, the conditional probability is a ratio because the only way to escape beyond \( n_B \) is to have also exceeded \( n_A < n_B \). In higher dimensions, a similar construction of the conditional can be made, but escapes must be computed along the local least-action trajectories under the action (44), and conditions computed for thresholds that lie in sequence along those trajectories. The leading exponential approximations to such probabilities are the standard first-passage constructions of Freidlin–Wentzell theory [30].
where \( d\theta \, d\eta \) is one of the two forms of the (differential) inner product appearing in Eq (76).

Thus

\[
\log \left( \frac{P(n_B | n_A; \eta_2)}{P(n_B | n_A; \eta_1)} \right) \sim (\eta_2 - \eta_1)
\]

(132)

is a sample estimator for the difference of exponential parameters in the two underlying distributions.

The quantity (131) may be computed at any time, for instance the final time \( T \) when the thresholds \( n_B \) and \( n_A \) are imposed as experimental conditions, and \( \eta_2 \) and \( \eta_1 \) characterize evolved nominal distributions at time \( T \) from any pair of initial conditions at some earlier time \( t = 0 \). If we use the stationary-path conditions to propagate values of \( \theta \) and \( \eta \) through time, and define \( V(\tau) \) to be the area inside the image of the rectangle in Eq. (131) along these stationary trajectories, time-invariance of the inner product, and the Liouville conservation of volume elements in dual coordinates, implies that

\[
\frac{d}{d\tau} \int_V d\theta \, d\eta = 0
\]

(133)

Note that, with a coordinate transform to coherent-state variables and a corresponding redefinition of the boundary of \( V \), the relation (133) could be recast using Eq. (124) as

\[
\frac{d}{d\tau} \int_V d\nu \, d\mu \, \hat{\Phi}_0^2 = 0
\]

(134)

which is the conserved integral graphed in Fig. 2.

In Eq. (134) \( \hat{\Phi}_0^2 \), the 2-dimensional differential of the scaled CGF \( \psi / N = - \log \hat{\Phi}_0 \), appears explicitly as the density of overlap of \( d\nu \) with \( d\mu \) that, like \( \psi \) itself, is constant along stationary paths. \( \hat{\Phi}_0^2 \) is not independent of the position \((\nu, \mu)\) within the volume \( V \), but because the volume element moves along with the conserved density, the integral measures a fixed quantity of Fisher information as it is transported through different domains of base and tilt.

Although the limits of integration for \( \int d\nu \, d\mu \) in Eq. (134) are bounded, the limits on \((\eta_2 - \eta_1)\) in Eq. (131) are not, so formally the range of the sample estimator (132) remains unbounded over any duration \( T \). However, for any fixed values of \((n_B - n_A)_t = T\) and starting uncertainty \((\eta_2 - \eta_1)_{t=0}\), the total information obtainable from large-deviations sampling about differences in the initial conditions is finite and decreases as \( e^{-T} \). In Fig. 2 this limit is seen in the way any fixed ranges are squeezed exponentially at the “waist” as \( T \to \infty \). The contraction of boundaries, rather than the asymptotic behavior of the eigenvalue in the Fisher metric, measures the loss of information between initial distributions and final observations with increasing separation between the two.
5 Conclusions: the duality of dynamics and inference for irreversible and reversible processes

The three-part structure of the Fisher metric, dual Riemannian connections, and symplectic parallel transport of the Wigner density, vector fields, and the metric tensor, elegantly expresses the transport properties along 2FFI stationary paths in terms of geometric invariants. It resolves a feature of two-field constructions that at first seems paradoxical: if memory of initial conditions is continuously lost to dissipation, what concept of time-reversal is implied by invertibility of the map along stationary rays? The answer from the perspective of importance sampling is that, even if samples are finite, their expectations are computed in continuous-valued distributions, and deformations of measure through the Radon–Nikodym derivative can locally compensate for concentration of measure in the nominal distribution by expanding sensitivity of likelihood ratios. Locally in sampling space, then, time is immaterial as it is in Hamiltonian mechanics; the mappings along stationary trajectories make it possible to interpret sampling protocols from different times in an evolving distribution simply as coordinate transformations of a fixed sampling protocol on the original distribution. On the other hand, for any fixed ranges of parameter variation in the initial conditions, and fixed large-deviation thresholds compared at late time, the integrated Liouville density contracts monotonically with the separation between the two times, reflecting the absolute loss of information that can be recovered.

We have wanted to establish a concrete interpretation of time-duality in 2FFI theories as a duality of dynamics and inference, to provide an alternative to the interpretation in terms of physical reversal of paths that is the starting point in most of the literature on fluctuation theorems in stochastic thermodynamics. Microscopic reversibility can always be added later to any class of 2FFI constructions as a restriction on the scope of phenomena under study, and both stronger conclusions and additional interpretations will then follow from the added constraints. Where the existence of a duality in the mathematics itself does not depend on any such additional assumptions, taking the inference interpretation to reflect the core concepts, directly expressing Kolmogorov’s forward/backward adjoint duality, frames the special case of microscopic reversibility as one in which the system’s own dynamics contains an image of certain sampling protocols over itself.

Even if one only cares about microscopically reversible processes, making explicit the step of self-modeling, and having a concrete interpretation of conserved densities such as the Fisher information constructed here, provides a bridge between trajectory reversal in low-level mechanics and operations for sample estimation of the kind that are used by control systems. Linking limitations from path probability in a system’s autonomous dynamics to concepts of information capacity in control loops [5, 6, 18] promises a way to study the limits on spontaneous emergence of dynamical hierarchy, which has been a desired application for stochastic thermodynamics [23, 59]. These are intended topics for future work.

Acknowledgements The author thanks Supriya Krishnamurthy for ongoing collaboration and the Stockholm University Physics Department for hospitality while much of this work was done, and Nathaniel Virgo for helpful discussion. The work was supported in part by NASA Astrobiology CAN-7 award NNA17BB05A through the College of Science, Georgia Institute of Technology, and by the Japanese Ministry of Education,
A Fishersphericalembeddings

A.1 The embedding for general distributions on finite state spaces

Equation (4) in the text can be written in the form

$$\delta s^2 = 4 \delta \theta^i \delta \theta^j \sum_n \left( \sqrt{\rho_n(\theta)} \frac{\partial}{\partial \theta^i} \sqrt{\rho_n(\theta)} \frac{\partial}{\partial \theta^j} \right)$$  \hspace{1cm} (135)$$

Let $|n|$ be the cardinality of the set of states on which $\rho_n$ is defined (for example, in chemistry, only a sub-lattice of all integer-valued vectors in the positive orthant may ever be accessible as counts, given a system’s stoichiometry and conserved quantities). Suppose $|n|$ is finite in order to illustrate the Fisher embedding geometry for distributions over finite state spaces. All possible base distributions $\rho$ fall within the simplex of dimension $|n| - 1$.

Now let $\{\alpha_1, \alpha_2, \ldots \alpha_{|n|} - 1\}$ be angles associated with independent rotation axes in $\mathbb{R}^{|n|}$. Any distribution can be embedded in $\mathbb{R}^{|n|}$ by arranging the states $n$ in an (arbitrary) order $n_1, n_2, \ldots n_{|n|}$, and writing

$$\begin{align*}
p_{n_1} &\equiv \cos^2 \alpha_1 \\
p_{n_2} &\equiv \sin^2 \alpha_1 \cos^2 \alpha_2 \\
p_{n_3} &\equiv \sin^2 \alpha_1 \sin^2 \alpha_2 \cos^2 \alpha_3 \\
\vdots \n
p_{n_{|n|} - 1} &\equiv \sin^2 \alpha_1 \sin^2 \alpha_2 \cdots \sin^2 \alpha_{|n|} - 2 \cos^2 \alpha_{|n|} - 1 \\
p_{n_{|n|}} &\equiv \sin^2 \alpha_1 \sin^2 \alpha_2 \cdots \sin^2 \alpha_{|n|} - 2 \sin^2 \alpha_{|n|} - 1
\end{align*}$$  \hspace{1cm} (136)$$

A recursive calculation gives the line element (135) in terms of the angle coordinates on the radius-2 sphere as

$$\delta s^2 = 4 \left[ \delta \alpha_1^2 + \sin^2 \alpha_1 \delta \alpha_2^2 + \cdots + \left( \sin^2 \alpha_1 \cdots \sin^2 \alpha_{|n|} - 2 \right) \delta \alpha_{|n|}^2 \right]$$  \hspace{1cm} (137)$$
A.2 Embeddings in reduced dimension for exponential families on the multinomial

The Poisson (35) and multinomial (36) distributions are both in a class recognized by Anderson, Craciun, and Kurtz (ACK) [4] in connection with uniqueness of stationary solutions for chemical reaction networks. All factorial moments are powers of their first moments, causing the CGF for many particles to scale as a multiple of a single-particle CGF. It is not then surprising that the expression (4) for the Fisher metric in terms of a distribution \( \tilde{\rho}_n \) with possibly indefinitely many independent terms, for the ACK distributions projects to a function of the same form in terms of expected numbers \( n_i \) over the \( D \) independent species.

To see how this works for a distribution \( \rho_n^{(n_0)} \) with multinomial form (36), express the expected number fractions as

\[
\frac{n_{0i}}{N} = \frac{e^{\eta_i}}{\sum_{j=1}^{D} e^{\eta_j}} \quad (138)
\]

Then the mean in the distribution \( \tilde{\rho}_n^{(\theta; n_0)} \) is

\[
\frac{n_i(\theta)}{N} = \frac{e^{\eta_i+\theta_i}}{\sum_{j=1}^{D} e^{\eta_j+\theta_j}} = v_i \quad (139)
\]

and the CGF \( \psi(\theta; n_0) \) evaluates (up to a constant offset) to

\[
\psi(\theta; n_0) = N \log \left[ \sum_{j=1}^{D} e^{\eta_j+\theta_j} \right] \quad (140)
\]

The Hessian giving the Fisher metric is

\[
g_{ij}(\theta) = N \left\{ v_i \delta_{ij} - v_i v_j \right\} \quad (141)
\]

where \( v_i \) is the function of \( \eta + \theta \) in Eq. (139). Inverting Eq. (141), and projecting onto the \( \sum_i \theta_i = 0 \) plane to fix the undetermined component of \( \theta \), gives the inverse

\[
g^{-1}(n) = \frac{1}{N} \left\{ \frac{1}{v_i} \delta_{ij} - \frac{1}{D} \left[ \frac{1}{v_i} + \frac{1}{v_j} - \frac{1}{D} \sum_k \frac{1}{v_k} \right] \right\} \quad (142)
\]

One can check that both \( g(\theta) \) and \( g^{-1}(n) \) sum to zero on either index, and the product

\[
g^{ij}(n) g_{jk}(\theta) = \delta^i_k - \frac{1}{D} \quad (143)
\]

is the identity on the subspace \( \sum_{j=1}^{D} \theta^j = 0 \) or \( \sum_{j=1}^{D} n_j = N \).
If a shift of the tilted distribution in the exponential family is indexed with coordinate $\delta n$, with $\sum_{j=1}^{D} \delta n_j = 0$, the Fisher distance element from Eq. (11) becomes

$$\delta s^2 = \sum_{j=1}^{D} \frac{\delta n_j^2}{n_j}$$

which is the same function of $n$ as the function of $\tilde{\rho}$ in the third line of Eq. (4).

B Sample means and variances in the large-deviation approximation to threshold indicator expectations

The expectation of the tilted indicator function from Eq. (23) may be written in a series of inequalities culminating in the expression for the CGF, as

$$\langle \tilde{h}_{\theta} \rangle_{(\theta;n_0)} = \sum_{n>n} e^{-\theta n} e^{\theta n} \rho_n^{(n_0)}$$

$$\leq e^{-\theta \bar{n}} \sum_{n>n} e^{\theta n} \rho_n^{(n_0)}$$

$$\leq e^{-\theta \bar{n}} \sum_n e^{\theta n} \rho_n^{(n_0)}$$

$$= e^{\psi(\theta;n_0) - \theta \bar{n}}$$

providing Eq. (24) in the text.

The variance of the same sample estimator has a corresponding bound

$$\left\langle \left( \tilde{h}_{\theta} \right)^2 \right\rangle_{(\theta;n_0)} - \left\langle \tilde{h}_{\theta} \right\rangle_{(\theta;n_0)}^2 = \left( \left( \tilde{h}_{\theta} \right)^2 \right)_{(\theta;n_0)} - \langle h \rangle_{(n_0)}^2$$

$$= \left\langle e^{\psi(\theta;n_0) - \theta n \tilde{h}_{\theta}} \right\rangle_{(\theta;n_0)} - \langle h \rangle_{(n_0)}^2$$

$$\leq e^{\psi(\theta;n_0) - \theta \bar{n}} \left\langle \tilde{h}_{\theta} \right\rangle_{(\theta;n_0)} - \langle h \rangle_{(n_0)}^2$$

$$= e^{\psi(\theta;n_0) - \theta \bar{n}} \langle h \rangle_{(n_0)} - \langle h \rangle_{(n_0)}^2$$

giving Eq. (25) in the text.

To estimate the tightness of the bounds, begin by observing that in the large-deviation scaling regime (27), with all cumulants generated as derivatives of $\psi \sim N$, the expansion of central moments in terms of cumulants bounds the scaling of the $k$th central moment as

$$\left\langle (n - \langle n \rangle)^k \right\rangle / \langle n \rangle^k \leq O\left( N^{-[k/2]} \right)$$

(147)
The log ratio we wish to bound is the Bregman divergence
\[
\log \left[ \frac{\sum_{n>n} e^{-\theta(n-n)} e^{\theta n} \rho_n}{\sum_n e^{\theta n} \rho_n} \right] = \theta \bar{n} - \psi(\theta, n_0) + \log \left( h(\bar{n}) \right)_{(n_0)} \tag{148}
\]

The maximum of Eq. (148) occurs at \( \theta(\bar{n}) \) from Eq. (8), and the width of the transition for the log ratio to change by \( \mathcal{O}(N^0) \) is given by
\[
\delta \theta \approx \left( \frac{\partial^2 \psi}{\partial \theta^2} \right)_{\theta(\bar{n})}^{-1} \sim N^{-1/2} \tag{149}
\]

To estimate its maximum value we write Eq. (148) as the sum of log ratios of the two inequalities in Eq. (145), and observe that they have boundary values
\[
\log \left[ \frac{\sum_{n>n} e^{-\theta(n-n)} e^{\theta n} \rho_n}{\sum_n e^{\theta n} \rho_n} \right] \bigg|_{\theta=0} = 0 \quad \log \left[ \frac{\sum_{n>n} e^{\theta n} \rho_n}{\sum_n e^{\theta n} \rho_n} \right] \bigg|_{\theta \to \infty} = 0 \tag{150}
\]

The values of these ratios at intermediate \( \theta \) are then obtained by integrating the derivatives
\[
\frac{\partial}{\partial \theta} \log \left[ \frac{\sum_{n>n} e^{-\theta(n-n)} e^{\theta n} \rho_n}{\sum_n e^{\theta n} \rho_n} \right] = -\frac{\sum_{n>n} (n-\bar{n}) e^{\theta n} \rho_n}{\sum_{n>n} e^{\theta n} \rho_n} \quad \frac{\partial}{\partial \theta} \log \left[ \frac{\sum_{n>n} e^{\theta n} \rho_n}{\sum_n e^{\theta n} \rho_n} \right] = \frac{\sum_{n>n} (n-n(\theta)) e^{\theta n} \rho_n}{\sum_{n>n} e^{\theta n} \rho_n} \tag{151}
\]

Both log derivatives in Eq. (151) are monotone if \( \psi \) is convex, and their values sum to \( \bar{n} - n(\theta) \), the derivative of Eq (148), where \( n(\theta) \equiv \langle n \rangle_{\rho(\theta, n_0)} \).

Next, observe the leading-order scaling of the expectation \( \langle (n-\bar{n})^2 \rangle_{n>n} \), \( \langle (n-\bar{n})^2 \rangle_{n>n} \) on the half-line (as for any second moment), and likewise for \( (n-n(\theta)) \).

At \( \theta(\bar{n}) \), where \( n(\theta) = \bar{n} \), the two derivatives (151) are equal and opposite, and the boundary of the half-line \( n > \bar{n} \) is also the symmetry point of \( (n-\bar{n})^2 \). Because the skewness and higher-order central moments grow more slowly than \( \bar{n}^k \) by Eq. (147), \( \langle (n-\bar{n})^2 \rangle_{n>n} \approx \langle (n-\bar{n})^2 \rangle_{n} \times [1 + \mathcal{O}(1/N)] \), and given the large-deviations scaling of central moments (147), the derivatives in Eq. (151) scale as
\[
\frac{\sum_{n>n} (n-\bar{n}) e^{\theta n} \rho_n}{\sum_{n>n} e^{\theta n} \rho_n} \bigg|_{n(\theta) = \bar{n}} \sim \sqrt{\langle (n-\bar{n})^2 \rangle_{n}} \sim N^{1/2} \tag{152}
\]

Over the range \( \pm \delta \theta \) from Eq. (149), where the total log ratio changes by \( \mathcal{O}(N^0) \), the integral of the first derivative (152) saturates the lower limit in the first line of Eq. (150), and the upper limit in the second line of Eq. (150), to within \( \leq \mathcal{O}(N^{1/2}) \), implying that the log-ratios themselves at the midpoint scale as \( \leq \mathcal{O}(N^{1/2}) \). Hence
also the total log ratio that is their sum scales as \( \log \left[ \langle h(\bar{n}) \rangle (n_0) / e^{\psi^* (\bar{n})} \right] \leq O(N^{1/2}) \), the result used in the text.

### C Review of Doi Hilbert space and Peliti functional integral constructions

#### C.1 Doi operator algebra and inner product

The main constructs in the Doi operator formulation \([21, 22]\) of moment-generating functions as formal power series are as follows:

The identification (30) of \( z \) and \( \partial / \partial z \) with raising and lowering operators \( a^\dagger \) and \( a \) allows the commutator algebra

\[
[ a_i, a^\dagger_j ] = \delta_{ij} .
\]  

(153)

to stand for the commutator algebra between components of \( \partial / \partial z \) and factors of \( z \), applied by function composition acting to the right on MGFs.

Monomials \( z^n \) in Eq. (1) are basis elements in a linear space of MGFs, built up by multiplication on the number 1. A bracket notation for states and an inner product are introduced by the pair of denotations

\[
1 \rightarrow |0 \rangle \quad \int dDz \delta^D(z) \rightarrow (0 |
\]  

(154)

Each monomial \( z^n \) is denoted as a number state

\[
\prod_{i=1}^{D} z_i^{n_i} \times 1 \rightarrow \prod_{i=1}^{D} a^\dagger_i^{n_i} |0 \rangle \equiv |n \rangle .
\]  

(155)

The number states are eigenstates of the set of number operators \( \hat{n}_i \equiv a^\dagger_i a^i \) (no Einstein sum):

\[
\hat{n}_i |n \rangle = n_i |n \rangle .
\]  

(156)

Dual to each number state is a conjugate projection operator

\[
( m | \equiv (0 | \prod_{i=1}^{D} a_i^{m_i} ) \left\langle \int dDz \delta^D(z) \prod_{i=1}^{D} \frac{ (\partial/\partial z_i)^{m_i} }{ m_i! } \right\rangle \]

(157)

From the commutation relations of variables and their derivatives it follows that the number states and projectors have overlap

\[
(m | n) = \delta^{D}_{mn}
\]  

(158)
the Kronecker δ symbol on $D$ indices. The number states and projectors are complete, and a sum of Eq. (158) on $m$ is the *Glauber norm*

\[
(0 | \sum_i a_i | n) = 1, \quad \forall n,
\]

which defines the asymmetric inner product on the Hilbert space of generating functions.

Replacing the uniform measure $\sum_i a_i \equiv 1^T a$ in Eq. (159) with the scalar product $za$ gives the map (33) to $\Psi(z)$ in the main text. $\Psi(1) \equiv 1$; the Glauber norm of the Laplace transform of any normalized distribution is the trace of the probability distribution $\sum_n \rho_n$.

### C.2 Coherent states and Peliti functional integral

The uniform measure in the 2D-dimensional integral for the representation of unity (42) in the main text is known as the Haar measure. Using the definition (38) for coherent states and (39) for their dual projectors, and expanding the exponential functions as sums,

\[
\int \frac{d^D \phi^\dagger d^D \phi}{\pi^D} |\phi\rangle \langle \phi| = \int \frac{d^D \phi^\dagger d^D \phi}{\pi^D} \prod_{i=1}^D e^{-\phi_i^\dagger \phi_i} \sum_{n_i} \sum_{m_i} \phi_i^{n_i} \phi_i^{m_i} |n\rangle \langle m| = \sum_n |n\rangle \langle n| = I
\]

The phase component in each integral $d \phi_i^\dagger d \phi_i$ vanishes unless $n_i = m_i$, and the remaining modulus component produces a Gamma-function canceling the $n_i!$. Thus the Haar measure on coherent states is equivalent to the uniform measure on states $n$ of the classical probability distribution.\(^7\)

Mapping backward through the correspondences between analytic functions and Doi state vectors from Appendix C.1, an evaluation of the integral in terms of Dirac δ-functions\(^8\) shows that the effect of the representation of unity is the map

\[^7\] Aaronson [1, (p. 123)] has raised this equivalence as one of the reasons only the complex L2 norm of quantum mechanics results in a correspondence principle with the classical laws of probability. It is interesting that the representation of probability components $\rho_n$ as squared amplitudes (though only real-valued) also underlies the natural spherical embedding of Appendix (A) for the Fisher metric.

\[^8\] There is a notational subtlety in writing the complex area integral $\int d \phi^\dagger d \phi \equiv \int_0^\infty d |\phi| \int_0^{2\pi} |\phi| d \arg \phi$ with respect to δ-functions evaluated as complex contour integrals. For example, in $D = 1$, the integral kernel in Eq. (161) is written in the two notations as

\[
\int d \phi^\dagger d \phi e^{\zeta^\dagger \phi^\dagger (\partial/\partial z) - \phi} = \int d \phi^\dagger d \phi e^{\zeta^\dagger (\partial/\partial z) - \phi} = \int_0^\infty d |\phi|^2 e^{-|\phi|^2} \sum_{n,m=0}^{\infty} \frac{(\zeta |\phi\rangle)^n}{n!} \frac{(\langle \phi | \partial/\partial z)^m}{m!} \int_0^{2\pi} d \arg \phi e^{i(n-m) \arg \phi} = \sum_{n=0}^{\infty} \left[ \int_0^\infty d |\phi|^2 e^{-|\phi|^2} \frac{|\phi|^{2n}}{n!} \right] \frac{(\zeta \partial/\partial z)^n}{n!} = \sum_{n=0}^{\infty} \frac{(\zeta \partial/\partial z)^n}{n!}
\]
The information geometry of two-field functional integrals

\[
\int \frac{d^D \phi^\dagger d^D \phi}{\pi^D} \left| \phi \right| \left( \phi^\dagger \right) \left( \phi \right) \left( \phi^\dagger \right) \left( \phi \right) = \int \frac{d^D \phi^\dagger d^D \phi}{\pi^D} e^{z' \phi} \int d^D z \delta^D (z) e^{\phi \left( \partial / \partial z - \phi \right)}
\]

\[
= \int d^D z \delta^D (z) e^{z' \partial / \partial z} = z \mapsto z'
\]

(161)

To keep track of the scoping rules for application of complex functions and derivatives would require introducing a distinct set of variables \( z_{k \delta t} \) for each interval in the quadrature (34). In the Doi operator algebra this scoping is handled by the bracket inner product, and the map (161) becomes simply the identity map on \( a^\dagger \) and \( a \).

The integration measure that results from inserting a copy of the representation of unity (42) between each interval of time evolution in Eq. (34) is called a skeletonized measure. Its limit as the interval length \( \delta t \to 0 \)

\[
\int_T^0 D^D \phi^\dagger D^D \phi \equiv \lim_{\delta t \to 0} \prod_{k=1}^{T/\delta t} \int \frac{d^D \phi_{k \delta t}^\dagger d^D \phi_{k \delta t}}{\pi^D}
\]

(162)

defines the functional integration measure used in Eq. (43) and elsewhere.

D Gaussian-order fluctuations, diffusions, and stochastic differential equations

The Doi-Peliti construction used in this paper affords a point of departure to many other representations of perturbed dynamical systems. It is important to emphasize that the motivation and form of 2FFI representations does not depend on the integer-lattice state space and exponential generating function that enable the specific Doi-Peliti construction. To show this we reproduce from [42, 68] a variety of reductions from the full field theory to the limit of Gaussian-order fluctuations that is specified by a drift velocity and diffusion tensor that are functions only of the configuration variable \( n \). These are “classical” representations of perturbed dynamical systems ostensibly lacking two-field structure. We will show how the inverse of the steps reducing Doi-Peliti to these classical limits has been used in [30, 52] to construct two-field theories or their large-deviation asymptotics in the form produced by the Doi-Peliti 2FFI. These classical descriptions from functions on \( n \) no longer reference lattice state spaces or population processes, and the 2FFI theories to which they correspond, like the operator formulation of [30, 52] are expected to apply to much more general perturbed dynamical systems.

The measure \( \int d\phi^\dagger d\phi \) over a single complex variable must therefore be used in evaluating \( \delta \)-functions as

\[
\int \frac{d\phi^\dagger d\phi}{\pi} e^{\left( \phi^\dagger - z' \right) \left( \partial / \partial z - \phi \right)} \equiv \int d\phi^\dagger \delta \left( \phi^\dagger - z' \right)
\]

where a factor of \( 2\pi \) would be required if \( \phi^\dagger \) and \( \phi \) were distinct complex variables integrated over independent contours. This use of the measure will be needed to understand the normalization of the Wigner function in later sections.
The Gaussian-order reduction of the general action (46) is obtained by writing the pair \((\theta, n) = (\bar{\theta}, \bar{n}) + (\theta', n')\), where \((\bar{\theta}, \bar{n})\) are a stationary trajectory, and expanding \(S\) to quadratic order in primed fields as

\[
S - \bar{S} \approx \int_0^T dt \left\{ \theta' \left( d_t + \frac{\partial^2 L}{\partial \theta \partial n} \right) n' + \frac{1}{2} \theta' \frac{\partial^2 L}{\partial \theta \partial \theta} \theta'^T + \frac{1}{2} n' T \frac{\partial^2 L}{\partial n T \partial n} n' \right\}
\] (163)

Overbar in Eq. (163) indicates evaluation of \(S\) or the derivatives of \(L\) at \((\bar{\theta}, \bar{n})\).

The Hessian \(\frac{\partial^2 L}{\partial \theta \partial \theta}\) is negative-definite and time-local, so the functional determinant of its time integral may be written as a Gaussian integral over a newly-introduced auxiliary field \(\xi\) with integrand \(e^{-S_{aux}}\), where

\[
S_{aux} \approx \int_0^T dt \left\{ \frac{1}{2} \left( \xi^T - \theta' \frac{\partial^2 L}{\partial \theta \partial \theta} \right) \left( - \frac{\partial^2 L}{\partial \theta \partial \theta} \right)^{-1} \left( \xi - \frac{\partial^2 L}{\partial \theta \partial \theta} \theta'^T \right) \right\}.
\] (164)

We may insert the number 1 into the path integral (43) by inserting the explicit functional integral over \(\xi\) and dividing by the determinant that is its evaluation, producing a new functional integral with integrand \(e^{-(S + S_{aux})}\). The quadratic-order expansion of \(S\) shifted by the auxiliary-field action satisfies

\[
S - \bar{S} + S_{aux}
\approx \int_0^T dt \left\{ \theta' \left( d_t + \frac{\partial^2 L}{\partial \theta \partial n} \right) n' - \xi \right\} + \frac{1}{2} \xi \left( - \frac{\partial^2 L}{\partial \theta \partial \theta} \right)^{-1} \xi^T + \frac{1}{2} n' T \frac{\partial^2 L}{\partial n T \partial n} n'.
\] (165)

The action (165) is linear in \(\theta'\), meaning that the functional integral over \(\theta'\) along the imaginary axis at each time-index \(t\) produces a functional Dirac \(\delta\)-function for the quantity multiplying \(\theta'\). The remaining terms are quadratic forms in \(\xi\) and \(n'\). In general, the \(\delta\)-functional produced by integration over \(\theta'\) connects these two fields by a time derivative, so the implied fluctuation distribution is not simple to compute. However, about ordinary relaxation trajectories that are the classical solutions to the equations of motion, for which \(\bar{\theta} = 0\), both \(\bar{S} = 0\) and also \(\bar{\partial^2 L}/\partial n T \partial n = 0\). (See [42] for extensive explanations of why this is the case in general, and of its interpretation in terms of causality.) For this case the functional integral (43) (which we evaluate at \(z = 1\) for simplicity), becomes

\[
1 = \int_0^T D\xi Dn e^{(e^{-\theta} - 1) n T} \times \exp \left( \frac{1}{2} \int_0^T dt \xi \left( - \frac{\partial^2 L}{\partial \theta \partial \theta} \right)^{-1} \xi^T \right) \delta \left[ \left( d_t + \frac{\partial^2 L}{\partial \theta \partial n} \right) n' - \xi \right] e^{\psi_0\left(\log \phi_0^1\right)}.
\] (166)
The integral (166) is over an ensemble of trajectories each satisfying

$$\left( dt + \frac{\partial^2 L}{\partial \theta \partial n} \right) n' = \xi.$$  \hfill (167)

Eq. (167) is the approximation to linear order in the perturbing field $\xi$ of the stochastic differential equation

$$\frac{dn}{dt} + \frac{\partial L}{\partial \theta} \bigg|_{\theta=0} = \xi.$$  \hfill (168)

Because the sole remaining Gaussian kernel in Eq. (166) is time-local, $\xi$ is a Langevin white noise with correlation function

$$\left\langle \xi_t \xi_{t'}^T \right\rangle = \left( - \frac{\partial^2 L}{\partial \theta \partial \theta} \bigg|_{\theta=0} \right) \delta(t-t').$$  \hfill (169)

The noise covariance (169), although expressed in terms of the value $\bar{n}$ along a stationary trajectory at $\theta = 0$, may be understood as a function of the local coordinates $n$ in the underlying manifold, since for classical solutions stationary trajectories form a foliation of the configuration manifold.

The Gaussian-order fluctuation limit corresponding to the Langevin equation (168, 169) produces, for the continuum limit of the lattice of states $n = (n_i)$, the diffusion equation for a density indexed by continuous index $n = (n_i)$, the master equation

$$\left\{ \frac{\partial}{\partial t} - \frac{\partial}{\partial n_i} \left( \frac{\partial L}{\partial \theta_i} \bigg|_{\theta=0} \right) + \frac{1}{2} \frac{\partial^2}{\partial n_i \partial n_j} \left( \frac{\partial^2 L}{\partial \theta_j \partial \theta_i} \bigg|_{\theta=0} \right) \right\} \rho_n$$  \hfill (170)

known as the Fokker–Planck equation [74]. (This form may be obtained directly from the transition matrix corresponding to the Liouville operator—for example, as written in [44, 69, 71] for chemical reaction networks—by writing discrete shifts of the index $n_i$ as if they were exponentiated operators $e^{\partial/\partial n_i}$ acting on a continuously-indexed density $\rho_n$, and then “expanding the Taylor’s series for the exponential” to second order.)

Here we have derived the three equivalent representations of the Gaussian-fluctuation limit specified entirely by a drift velocity and diffusion kernel as functions of $n$—the quadratic action (163), the Langevin equation (168, 169), and the Fokker–Planck equation (170)—from a Doi-Peliti starting point. The reverse of the derivation from Eqs. (163–170) may be followed to arrive at either two-field forms or their large-deviation asymptotics starting from the stochastic differential equation or Fokker–Planck equation. Kamenev [42] has shown that the “tri-diagonal” arrangement of response and correlation functions in Eq. (4.1) of [52], equivalent to the quadratic action (163) here, is in fact the common motif in 2FFI for stochastic classical mechanics, and for the Schwinger-Keldysh time-loop in quantum mechanics [43, 63], allowing us to bypass the operator formalism altogether and proceed directly from
stochastic differential equations (168, 169) to the path integral up to Gaussian order. For a development of correspondence principles, where both the quantum Schwinger-Keldysh time loop and the stochastic two-field integral appear, see [67].

It is central to the aims of [52] that the tri-diagonal response and correlation functions are not limited to the “bare” constitutive parameters of the underlying stochastic differential equation, but rather are a general form for Gaussian order fluctuations in an effective theory where the drift velocity and fluctuation covariance are generally renormalized quantities, in this way subsuming higher-order fluctuation effects, which may appear explicitly in the Doi-Peliti action. Renormalization of the stochastic action as an effective action may also be done directly from the path integral with diagrammatic methods of the type appearing in [52]. For an application to evolutionary games see [70, Ch. 7].

Alternatively, the large-deviation asymptotics for excursions within, or for escapes from, a domain, may be derived directly from the stochastic differential equation, and related to boundary-value problems for solutions of diffusions such as the Fokker-Planck equation. This program, starting from representations identical to Eqs. (168–170) for Gaussian fluctuations, and covering several other classes of perturbations as well, is carried out by Freidlin and Wentzell [30], and has been applied to a wide range of first-passage and escape problems [48–51].

The connection to 2FFI methods is that an action arises as their large-deviation function from the drift and diffusion functions of the Gaussian limit, which is simply obtained from the Doi-Peliti action. By completing the square in $\theta'$ (which is $\theta$ in a neighborhood of $\bar{\theta} = 0$), the quadratic action (163) is put into the form

$$ S \rightarrow \frac{1}{2} \int_0^T dt \left\{ \left( d_t + \frac{\partial^2 L}{\partial \theta \partial n} \right) n' \right\}^T \left( -\frac{\partial^2 L}{\partial \theta \partial \theta} \right)^{-1} \left\{ \left( d_t + \frac{\partial^2 L}{\partial \theta \partial n} \right) n' \right\} + \text{quadratic in } \theta $$

(171)

The quadratic form in $\theta$ is integrated out in the functional integral to produce a functional determinant that depends on the background $\bar{n}$ only logarithmically. Again, the term in $n'$ in Eq. (171) is the leading-linear expansion of the equations of motion in small fluctuations, so at the Gaussian order to which it is defined, Eq. (171) may be approximated as

$$ S \rightarrow \approx \frac{1}{2} \int_0^T dt \left\{ \left. \frac{dn}{dt} + \frac{\partial L}{\partial \theta} \right|_{\theta=0} \right\}^T \left( -\left. \frac{\partial^2 L}{\partial \theta \partial \theta} \right|_{\theta=0} \right)^{-1} \left\{ \left. \frac{dn}{dt} + \frac{\partial L}{\partial \theta} \right|_{\theta=0} \right\} $$

(172)

a form first due to Onsager and Machlup [55], and appearing in Theorem 1.1 on p. 86 of [30].

Two-field constructions originating in classical perturbed dynamical systems (168–170) in the continuous variable $n$ may readily be extended to continuously-indexed state spaces with topologies, on which the drift velocity may include spatial gradi-
ents, as in the case of reaction-diffusion systems, without altering the program for construction of the corresponding two-field representation.

**E Stationary-point approximations to the Wigner function**

The functional integral provides the most direct route to the current conservation law (62) for the Wigner function. It is possible, with somewhat more work, to derive the same relations directly from stationary variation of the generating function, and in the process to gain some more intuition for what the Wigner function quantifies.

**The Wigner function in terms of an explicit density over coherent-state parameters**

Begin by writing any state vector as the integral of a density in the coherent-state basis:

$$|\Psi\rangle = \int d^D\phi \ |\phi\rangle \rho(\phi)$$  \hspace{1cm} (173)

The generalized Glauber norm (33) returns the analytic representation of the MFG:

$$\Psi(z) = (0| e^{za} |\Psi\rangle = \int d^D\phi \ (0| e^{za} |\phi\rangle \rho(\phi) = \int d^D\phi \ e^{(z-1)\phi} \rho(\phi)$$  \hspace{1cm} (174)

Now evaluate the integral in Eq. (59) at time $t \rightarrow T$, where the stationary value of the field $\phi^+$ will coincide with the imposed argument $z$,

$$w_T\left(\phi^\perp_1, \phi^\perp_2\right) = \frac{1}{\pi D} \int d^D\phi \ e^{\left(\phi^\perp_1-z\right)(\phi^\perp_2-\phi^\perp_1)} e^{(z-1)\phi} \rho(\phi)$$  \hspace{1cm} (175)

It follows then that

$$\int d^D\phi^\perp w_T\left(\phi^\perp_1, \phi^\perp_2\right) = e^{(z-1)\phi^\perp_1} \rho(\phi^\perp_2)$$  \hspace{1cm} (176)

A second integral over the $\phi^\perp_2$ fields yields the two equivalent expressions (60) and (174).

**Stationary-point approximations**

The stationary value $\bar{\phi}(z)$ of the tilted density $e^{(z-1)\phi} \rho(\phi)$ is given by

$$\frac{\partial \log \rho}{\partial \phi} \bigg|_{\bar{\phi}(z)} = 1 - z.$$  \hspace{1cm} (177)
The two stationarity conditions on the arguments of \( w_T \) follow from Eq. (175) as

\[
\frac{\partial}{\partial \phi^\dagger} w_T(\phi^\dagger, \phi^\ddagger) \bigg|_{\phi^\ddagger = z} = \left( z - \phi^\dagger \right) w_T(z, \phi^\ddagger)
\]

\[
\frac{\partial}{\partial \phi^\ddagger} w_T(\phi^\dagger, \phi^\ddagger) \bigg|_{\phi^\dagger = z} = \int d^D\phi \left( \phi - \phi^\ddagger \right) e^{(z-1)\phi} \rho(\phi) \sim \left( \tilde\phi - \phi^\ddagger \right) w_T(z, \phi^\ddagger)
\]

(178)

where the stationary-point approximation (177) to the mean gives the leading exponential approximation in the second expression.

The Wigner function in Eq. (175), at argument \( z \), exactly equals integral (174)

\[
w_T(z, \phi^\ddagger) = \Psi(z) \sim e^{(z-1)\phi} \rho(\phi) \bigg|_{\phi(z)}
\]

(179)

independent of the value of \( \phi^\ddagger \). While the first line of Eq. (178) shows that \( z \) is a stationary-point argument for \( \phi^\dagger \), the second line shows that only when \( \phi^\ddagger = \tilde\phi(z) \) is the other argument also a stationary value.

### Time dependence along a stationary path

Suppose now that from such a compatible pair \((z, \tilde\phi(z))\), we wish to extend \( z \) and \( \tilde\phi \) along a trajectory that preserves stationarity. The total time derivative of \( w_T \) with respect to its final-time argument is given by

\[
\frac{d}{dT} w_T(\phi^\dagger, \phi^\ddagger) = \left\{ \frac{dz}{dT} \phi^\dagger \phi^\ddagger + \frac{d\phi^\ddagger}{dT} \frac{\partial}{\partial \phi^\ddagger} + \frac{d\phi^\dagger}{dT} \frac{\partial}{\partial \phi^\dagger} \right\} w_T(\phi^\dagger, \phi^\ddagger)
\]

\[
- \frac{1}{\pi^D} \int d^D\phi e^{(\phi^\ddagger - z)(\phi - \phi^\ddagger)} \mathcal{L}(\phi^\dagger, \phi) e^{(z-1)\phi} \rho(\phi)
\]

(180)

(Note that Eq. (180) includes only contributions to the derivative of \( w_T \) from quantities defined before time \( T \); this derivative is different from the total derivative \( d/dt \) of \( w_t \) in the functional integral (59), which also includes effects of the functional integral after time \( t \).)

Ensuring that \( z + dt \left( dz/dt \right) \) is a stationary value if \( z \) is one requires that

\[
0 = \frac{\partial}{\partial \phi^\ddagger} \frac{d}{dt} w_T(\phi^\dagger, \phi^\ddagger) \bigg|_{\phi^\ddagger = z} \sim \left\{ \frac{dz}{dt} - \frac{\partial}{\partial \phi} \mathcal{L}(z, \phi) \bigg|_{\phi(z)} \right\} w_T(z, \phi^\ddagger)
\]

(181)

The term \( \partial \mathcal{L}/\partial \phi \) is obtained by an integration by parts over \( d^D\phi \), and evaluated in the stationary-point approximation. All other terms from Eq. (180) vanish at \( \phi^\dagger = z \). Thus preservation of the stationary-argument condition for \( \phi^\ddagger \) gives the stationary-path equation for \( dz/dt \).
To identify the time-dependence of the stationary argument $\phi^\ddagger$, we work directly from the stationary-point condition (177). The total time derivative of that equation is

$$\frac{\partial L}{\partial \phi} \bigg|_{\phi^\ddagger} = \frac{d}{dt} \left( \frac{\partial \log \rho}{\partial \phi} \bigg|_{\phi^\ddagger(z)} \right)$$

$$= \frac{d\phi}{dt} \left( \frac{\partial^2 \log \rho}{\partial \phi^2} \bigg|_{\phi^\ddagger} \right) + \frac{\partial}{\partial \phi} L(z, \phi^\ddagger) \bigg|_{\phi^\ddagger(z)}$$

$$= -\frac{d\phi}{dt} \left( \frac{\partial z}{\partial \phi^\ddagger} \bigg|_{t} \right) - \frac{\partial}{\partial \phi} L(z, \phi^\ddagger) \bigg|_{\phi^\ddagger(z)}$$

In passing from the second to the third line of Eq. (182), to obtain an explicit expression for $\frac{\partial \log \rho(\phi)}{\partial t} \big|_{\phi^\ddagger}$ in terms of $L(z, \phi^\ddagger)$, we evaluate $z$ as an inverse function of $\phi^\ddagger$ from Eq. (177). This functional dependence contributes the term $\frac{\partial L}{\partial z}$ in the final line, from which we obtain the stationary-path equation for the trajectory of $\phi^\ddagger(z)$ along which $\phi^\ddagger$ is to be evaluated:

$$\frac{d\phi}{dt} = -\frac{\partial}{\partial z} L(z, \phi^\ddagger) \quad (183)$$

Equations (181) and (183) imply the 2FFI counterpart to conservation of energy in Hamiltonian mechanics: $dL/dt = 0$ along the stationary path if $\partial L/\partial t = 0$.

**The stochastic effective action in stationary-path evaluations**

Note from Eq. (180) that along the contour identified to preserve stationarity of $w_T$.

$$\frac{d}{dT} \left( w_T(z, \phi^\ddagger(z)) \big|_{z(T)} \right) = \frac{dz}{dT} \phi^\ddagger - L(z, \phi^\ddagger) w_T(z, \phi^\ddagger)$$

$$= \frac{d}{dT} \int_T dt \left\{ -\frac{dz}{dt} \phi^\ddagger + L(z, \phi^\ddagger) \right\}$$

$$= \frac{d}{dT} \tilde{S}_T \quad (184)$$

Therefore the extension of the Wigner function to times $t > T$ must include the stationary-path contribution from the action, which was present for $t < T$ in the functional integral definition (59). $w_T$ thus extended satisfies

$$\frac{d}{dt} \left( e^{-\tilde{S}_T} w_T(z, \phi^\ddagger) \right) \sim 0 \quad (185)$$

recovering Eq. (62).
We have termed the stationary-path evaluation of $S$ the \textit{stochastic effective action} \cite{68}. It is the functional Legendre transform of the large-deviation functional for trajectories in Doi-Peliti integrals. The approximation (179), with $\phi(z)$ given $\rho(\phi)$, together with the contribution from $S_T$ in Eq. (185), provides the desired interpretation of the Wigner function in terms of densities in the statistical model provided by coherent states, and their exponential tilts by likelihood functions.

**F Stationary-path solutions for the two-state system**

The stationary-path equations and both initial and final values for fields are obtained from vanishing of all terms in the variational derivative of the exponential argument in Eq. (43). We begin with solutions in coherent-state variables, and then present the forms for the descaled number coordinates $\nu, \bar{\nu}$, and $\tilde{\nu}$.

**F.1 Coherent-state and number-potential solutions**

**F.1.1 Stationary-path equations and final-time conditions for response fields**

The stationary-path equations of motion for the components of the field $\phi^\dagger$ from Eq. (47), in the rotated basis (100), evaluate to

$$\partial_\tau \Phi^\dagger = \frac{\partial \hat{L}}{\partial \Phi} = -v \Phi^\dagger \quad \partial_\tau \phi^\dagger = \frac{\partial \hat{L}}{\partial \phi} = \phi^\dagger$$

(186)

The final-time values $\phi^\dagger_T$ are given by variation of $\hat{\phi}_T$, as

$$\phi^\dagger_a = z_a \quad \phi^\dagger_b = z_b$$

(187)

Fixing the magnitude of the combination $\phi^\dagger_b v_b + \phi^\dagger_a v_a$ in Eq. (104) requires varying $z$ along the contour

$$z_a \equiv \frac{v_{Ta}}{v_a} \quad z_b \equiv \frac{v_{Tb}}{v_b}$$

(188)

giving Eq. (113) in the main text. The remaining time-dependent solutions, with time argument denoted explicitly here by subscript $\tau$, are given by

$$\phi^\dagger_\tau = e^{\tau - T} \quad \Phi^\dagger_\tau = 1 - v \phi^\dagger_\tau$$

(189)

Initial data are specified in the generating function $\psi_0$, which when evaluated at the solutions for $\phi^\dagger_0$ become

$$\frac{1}{N} \psi_0 \left( \log \phi^\dagger_a, \log \phi^\dagger_b \right) = \log \left( \phi^\dagger_a v_a + \phi^\dagger_b v_b \right)$$
\[= \log \left( \Phi_0^\dagger + \bar{\nu}_0 \Phi_0^\dagger \right)\]
\[= \log \left( 1 + \Lambda e^{-T} \right)\]  \hspace{1cm} (190)

The final line of Eq. (190) is obtained by combining the two solutions (189) at \(\tau = 0\), and introduces the combination \(\Lambda\) defined in Eq. (106).

### F.1.2 Stationary-path equations and initial-time conditions for observable fields

The stationary-path equations of motion for the components of the field \(\phi\) from Eq. (47) are obtained by removing a total derivative \(dt (\phi^\dagger \phi)\) from the action (44) to shift the derivative onto \(\phi\). In the rotated basis (100), they evaluate to

\[\partial_\tau \hat{\phi} = -\frac{\partial \hat{\mathcal{L}}}{\partial \phi^\dagger} = 0 \quad \partial_\tau \phi = -\frac{\partial \hat{\mathcal{L}}}{\partial \phi^\dagger} = - \left( \hat{\phi} - \nu \hat{\Phi} \right)\]  \hspace{1cm} (191)

The total derivative cancels the final-time term \((\phi^\dagger \phi)_T\) from the exponential in Eq. (43) and introduces an initial-time term \((\phi^\dagger \phi)_0\). Variation of this term against \(\psi_0\) with respect to \(\phi_0^\dagger\) gives the initial-value conditions for the components of \(\hat{\phi}_0\), as

\[\hat{\phi}_{0a} = \frac{\partial \psi_0}{\partial \phi_{0a}} \quad \hat{\phi}_{0b} = \frac{\partial \psi_0}{\partial \phi_{0b}}\]  \hspace{1cm} (192)

Solutions to the equations of motion (191) from these initial conditions are then

\[\hat{\Phi}_t = \Phi_0 = \frac{1}{1 + \Lambda e^{-T}} \quad \hat{\phi}_t = \hat{\Phi}_0 \left[ \nu + (\bar{\nu}_0 - \nu) e^{-\tau} \right]\]  \hspace{1cm} (193)

The two displacements defined in Eqs. (104) and (105), characterizing respectively the mean in the nominal distribution and the likelihood ratio applied to the stationary measure, evaluate to

\[\bar{\nu}_\tau = \nu + (\bar{\nu}_0 - \nu) e^{-\tau} \quad \nu_\tau = \nu + (\nu_T - \nu) e^{\tau - T}\]  \hspace{1cm} (194)

These results are reproduced (dropping the explicit subscripts \(\tau\)) as Eq. (107) in the text. It follows from Eq. (194) that the combination

\[\frac{(\nu_\tau - \nu)(\bar{\nu}_\tau - \nu)}{\left(1 - \nu^2\right)} = \Lambda e^{-T}\]  \hspace{1cm} (195)

is invariant at its initial value.

The CGF for a binomial distribution at any time retains the form (92), with \(\bar{\nu}_a/\nu_a\) and \(\nu_b/\nu_b\) replacing \(z_a\) and \(z_b\), and \(\bar{\nu}_a\) and \(\bar{\nu}_b\) replacing \(\nu_a\), and \(\nu_b\). From Eq. (190)
and the invariant form (195), it follows that

$$\frac{\psi_\tau}{N} = \log \left[ 1 + \Lambda e^{-T} \right] = -\log \hat{\Phi}_0$$

(196)

giving Eq. (109) in the text.

Finally, the non-linear mean of sample values (102) in the importance distribution can be shown to evaluate to

$$\nu_\tau = \nu_\tau + \left( \frac{1}{4} - \nu_\tau^2 \right) \left[ (\bar{\nu}_\tau - \nu) \hat{\Phi}_0 \right]$$

(197)

from which the form (111) for $\partial \nu / \partial \theta$ can be derived. The ratio of measures $\left( \frac{1}{4} - \nu_\tau^2 \right) / \left( \frac{1}{4} - \nu^2 \right)$ in Eq. (197), by which the importance distribution responds to variations in the initial data, is the familiar scaling of response functions [68] in the Fluctuation-Dissipation Theorem, because expressions of the form $N \left( \frac{1}{4} - \nu^2 \right) = N \nu_u \nu_b$ are the variance of fluctuations in the binomial.

### F.2 Fisher spherical embedding

In one dimension, the Pythagorean theorem (12) for K-L divergences loses the interpretation of a direction cosine between vector fields, but still reflects scale changes between coherent-state or exponential families and the geometric coordinate.

The mean-value Fisher-sphere construction of Sect. A.2, for one variable, is the embedding on a circle:

$$\cos^2 \alpha = \frac{1}{2} + \nu \quad \sin^2 \alpha = \frac{1}{2} - \nu$$

(198)

The coordinate differential is

$$2 \frac{d\alpha}{d\theta} = \sqrt{\frac{1}{4} - \nu^2}$$

(199)

and the geometric distance element is then

$$4(\delta \alpha)^2 = \left( \frac{1}{4} - \nu^2 \right) (\delta \theta)^2 = \frac{\partial \nu}{\partial \theta} (\delta \theta)^2$$

(200)

The $\hat{\Phi}_0^2$ term in Eq. (124) becomes

$$\hat{\Phi}_0^2 = \frac{1}{[1 + uv]^2} = \frac{\left( \frac{1}{4} - \nu^2 \right)}{\left( \frac{1}{4} - \nu^2 \right) \left( \frac{1}{4} - \bar{\nu}^2 \right)}$$

$$= \frac{4 \left( \partial \alpha / \partial \theta \right) \left( \partial \alpha / \partial \eta \right)}{\left( \partial \nu / \partial \theta \right) \left( \partial u / \partial \eta \right)}$$

$$= \frac{4 \partial \alpha}{\partial \nu} \frac{\partial \alpha}{\partial u}$$

(201)
the invariant Fisher information in coherent-state coordinates. The equivalence between the two forms (69) and (71) for the Fisher metric is again recovered as

$$\frac{\partial^2}{\partial \theta^2} \left( \frac{\psi}{N} \right) = 4 \frac{\partial \alpha}{\partial \theta} \frac{\partial \alpha}{\partial \eta}$$

(202)

showing the variation of the embedding coordinate of the importance distribution with the tilt multiplied by its variation with the base.

F.3 Evaluation of the Amari–Chentsov tensor

From the form (122) of the Fisher metric in coordinates \((v, u)\), the Amari–Chentsov tensor on all contravariant indices can be computed:

$$\frac{T}{N} = \frac{\partial^3}{\partial \theta^3} \left( \frac{\psi}{N} \right)$$

$$= -2 \frac{(dv/d\theta) (du/d\eta)}{[1 + uv]^3} \left[ v (1 - uv) + \sqrt{\frac{1}{4} - \nu^2 (u + v)} \right]$$

(203)

\(T\) is symmetric under \(u \leftrightarrow v\), but its magnitude is not conserved along the stationary-path trajectories. The measure changes \((dv/d\theta)\) and \((du/d\eta)\) are the same as those in the Fisher metric, but in addition to these the term \(u + v\) is not invariant.

F.4 Connection coefficients in the coherent-state connection

Among the nonzero connection coefficients for the dual coherent-state connections (90), the only independent components are for the rotated variables \(\theta\) of Eq. (95) and \(\eta\) of Eq. (115). They are

$$\Gamma^{(\theta)}_{\theta \theta} = \frac{\partial}{\partial \theta} \log \left( \frac{\partial v}{\partial \theta} \right) \quad \Gamma^{(\eta)\eta} = \frac{\partial}{\partial \eta} \log \left( \frac{\partial \nu}{\partial \eta} \right)$$

(204)

The covariant components of the time derivative of vector fields \(\delta \theta\) and \(\delta \eta\) defined in Eq. (84), for transport respectively along \(\dot{\theta}\) and \(\dot{\eta}\), evaluate to

$$\left( \frac{\partial}{\partial \tau} \delta \theta \right) + \dot{\theta} \left( \nabla_\theta \delta \theta \right) = \left( \frac{d}{d \tau} \delta \theta \right) + \delta \theta \dot{\theta} \frac{\partial}{\partial \theta} \log \left( \frac{\partial v}{\partial \theta} \right)$$

$$= \delta \theta \left[ \frac{\partial^2 \hat{L}}{\partial \theta \partial v} + \frac{d}{d \tau} \log \left( \frac{1}{4} - \nu^2 \right) \right]$$

$$= \delta \theta \left[ \frac{\partial^2 \hat{L}}{\partial \theta \partial v} + \frac{d}{d \tau} \log \left( \frac{1}{4} - \nu^2 \right) \right]$$

$$\left( \frac{\partial}{\partial \tau} \delta \eta \right) + \dot{\eta} \left( \nabla_\eta \delta \eta \right) = \left( \frac{d}{d \tau} \delta \eta \right) + \delta \eta \dot{\eta} \frac{\partial}{\partial \eta} \log \left( \frac{\partial \nu}{\partial \eta} \right)$$
\[
= -\delta \eta \left[ \frac{\partial^2 \tilde{L}}{\partial \nu \partial \eta} - \frac{d}{d \tau} \log \left( \frac{1}{4} - \tilde{v}^2 \right) \right]
= -\delta \eta
\] (205)

The corrections from the coherent-state connection coefficients (204) may be understood immediately by using the first line of Eq. (114) and Eq. (115) together with the time dependencies (107), to write

\[
\delta \theta \tau = \delta \theta \frac{1}{4} - \bar{\nu}^2 \tau \delta \eta \frac{1}{4} - \bar{\nu}^2 \tau e^{-\tau} \quad \delta \eta \tau = \delta \eta \frac{1}{4} - \bar{\nu}^2 \tau e^{-\tau}
\] (206)

The geometric invariants (205) capture the exponential growth or decay from Eq. (206), while connection terms remove measure factors \((\frac{1}{4} - \bar{\nu}^2 \tau)\) for exponential relative to coherent-state coordinates.

The Fisher metric \(g\), unlike \(\delta \theta\) and \(\delta \eta\), has no intrinsic time dependence and changes only due to change in the net binomial parameter of the importance distribution. Its covariant derivatives (85), with connection coefficients (204), then produce the two independent components of variation from \(\dot{\theta}\) and \(\dot{\eta}\) of

\[
\dot{\theta} \nabla_{\theta} g = \dot{\theta} \left[ \frac{\partial g}{\partial \theta} - g \frac{\partial}{\partial \theta} \log \left( \frac{\tilde{v}}{\Phi_0^2} \right) \right]
= \dot{\theta} \frac{\partial g}{\partial \theta} - g \frac{d}{d \tau} \log \left( \frac{1}{4} - \bar{\nu}^2 \right)
= \left( \tilde{v} \frac{\partial}{\partial \tilde{v}} \log \Phi_0^2 \right) g
\]

\[
-\dot{\eta} \nabla^*_{\eta} g = -\dot{\eta} \left[ \frac{\partial g}{\partial \eta} - g \frac{\partial}{\partial \eta} \log \left( \frac{\tilde{v}}{\Phi_0^2} \right) \right]
= -\dot{\eta} \frac{\partial g}{\partial \eta} + g \frac{d}{d \tau} \log \left( \frac{1}{4} - \bar{\nu}^2 \right)
= -\left( \tilde{u} \frac{\partial}{\partial \tilde{u}} \log \Phi_0^2 \right) g
\] (207)

These are reproduced as Eq. (126) in the text.

**F.5 Two-dimensional divergences of the large-deviation function and their integrals**

In the exponential family of tilted distributions with tilting parameter \(\theta\), over a base distribution with exponential parameter \(\eta\), the large-deviation function of two arguments is constructed as

\[
\psi^* (n; \eta) = \theta(n; \eta) n - \psi (\theta(n; \eta); \eta)
\] (208)
Its variation with $\eta$ at fixed $n$ is given by

$$\left. \frac{\partial \psi^*}{\partial \eta} \right|_n = \left. \frac{\partial \theta}{\partial n} \right|_n \left( n - \left. \frac{\partial \psi}{\partial \theta} \right|_\eta \right) - \left. \frac{\partial \psi}{\partial \eta} \right|_\theta$$

(209)

By definition of $\theta(n; \eta)$ as the inverse function of $n(\theta; \eta) \equiv \partial \psi(\theta; \eta)/\partial \theta$, the term $(n - \partial \psi/\partial \theta) \equiv 0$ at all $n$. Therefore the second derivative

$$\frac{\partial^2 \psi^*}{\partial n \partial \eta} = - \left. \frac{\partial \theta}{\partial n} \right|_\eta \frac{\partial^2 \psi}{\partial \theta \partial \eta}$$

(210)

The third line of Eq. (210) uses additivity of $\theta$ and $\eta$ to cancel the two factors, which equal respectively $g^{-1}$ and $g$.

The log-ratio of the conditional large-deviation probabilities in Eq. (130) may be written as the integral

$$\log \left( \frac{P(n_B \mid n_A; \eta_2)}{P(n_B \mid n_A; \eta_1)} \right) = \int_{\eta_1}^{\eta_2} \frac{\partial}{\partial \eta} \int_{n_A}^{n_B} d \eta d n \frac{\partial}{\partial \eta} \left. \frac{\partial^2 \psi^*}{\partial n \partial \eta} \right|_n$$

(211)

giving Eq. (131) in the text. The conversion from the fifth to the sixth line in Eq. (211) makes use of the two alternative ways of expressing the inner product in contravariant/covariant coordinates given in Eq (76).
References

1. Aaronson, S.: Quantum Computing Since Democritus. Cambridge University Press, London (2013)
2. Amari, S.-I.: Information Geometry and its Applications. Appl. Math. Sci. vol. 194. Springer Japan (2001)
3. Amari, S.-I., Nagaoka, H.: Methods of Information Geometry. Oxford University Press, Oxford (2000)
4. Anderson, D.F., Craciun, G., Kurtz, T.G.: Product-form stationary distributions for deficiency zero chemical reaction networks. Bull. Math. Biol. 72, 1947–1970 (2010)
5. Ashby, W.R.: An Introduction to Cybernetics. Chapman and Hall, London (1956)
6. Ashby, W.R.: Requisite variety and its implications for the control of complex systems. Cybernetica 1, 83–99 (1958)
7. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information Geometry. Schwinger International, Cham (2017)
8. Baez, J.C.: Quantum techniques for reaction networks. arXiv:1306.3451
9. Baez, J.C., Biamonte, J.D.: Quantum techniques for stochastic mechanics. https://math.ucr.edu/home/baez/stoch_stable.pdf
10. Baez, J.C., Fong, B.: Quantum techniques for studying equilibrium in reaction networks. J. Complex Netw. 3, 22–34 (2014). https://academic.oup.com/comnet/article-abstract/3/1/22/490572/Quantum-techniques-for-studying-equilibrium-in?redirectedFrom=fulltext
11. Baish, A.J.: Deriving the Jarzynski relation from Doi-Peliti field theory. Bucknell University Honors Thesis (2015)
12. Bazarov, I.V.: Synchrotron radiation representation in phase space. Phys. Rev. ST Accel. Beams 15, 050703 (2012)
13. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., Landim, C.: Towards a nonequilibrium thermodynamics: a self-contained macroscopic description of driven diffusive systems. J. Stat. Phys. 135, 857–872 (2009)
14. Chernyak, V., Chertkov, M., Jarzynski, C.: Path-integral analysis of fluctuation theorems for general Langevin processes. J. Stat. Mech., P08001 (2006). https://doi.org/10.1088/1742-5468/2006/08/P08001
15. Chetrite, R., Gawedzki, K.: Fluctuation relations for diffusion processes. Commun. Math. Phys. 282, 469–518 (2008)
16. Cohen, E.D.G., Gallavotti, G.: Note on two theorems in nonequilibrium statistical mechanics. J. Stat. Phys. 96, 1343–1349 (1999)
17. Cohen, L.: Generalized phase-space distribution functions. J. Math. Phys. 7, 781–786 (1966)
18. Conant, R.C., Ashby, W.R.: Every good regulator of a system must be a model of that system. Int. J. Syst. Sci. 1, 89–97 (1970)
19. Crooks, G.E.: Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys. Rev. E 6, 2721–2726 (1999)
20. Crooks, G.E.: Path-ensemble averages in systems driven far from equilibrium. Phys. Rev. E 61, 2361–2366 (2000)
21. Doi, M.: Second quantization representation for classical many-particle system. J. Phys. A 9, 1465–1478 (1976)
22. Doi, M.: Stochastic theory of diffusion-controlled reaction. J. Phys. A 9, 1479 (1976)
23. England, J.L.: Statistical physics of self-replication. J. Chem. Phys. 139, 121923 (2013)
24. Esposito, M., Van den Broeck, C.: Three detailed fluctuation theorems. Phys. Rev. Lett. 104, 090601 (2010)
25. Evans, D.J., Cohen, E.G.D., Morriss, G.P.: Probability of second law violations in shearing steady states. Phys. Rev. Lett. 71, 2401–2404 (1993)
26. Evans, D.J., Seares, D.J.: Fluctuation theorem for stochastic systems. Phys. Rev. E 60, 159–164 (1999)
27. Evans, D.J., Seares, D.J.: The fluctuation theorem. Adv. Phys. 51, 1529–1585 (2002)
28. Feynman, R.P., Hibbs, A.R.: Quantum Mechanics and Path Integrals. McGraw-Hill, New York (1965)
29. Flajolet, P., Sedgewick, R.: Analytic Combinatorics. Cambridge University Press, London (2009)
30. Freidlin, M.I., Wentzell, A.D.: Random Perturbations in Dynamical Systems, 2nd edn. Springer, New York (1998)
31. Gallavotti, G., Cohen, E.D.G.: Dynamical ensembles in non-equilibrium statistical mechanics. Phys. Rev. Lett. 74, 2694–2697 (1995)
32. Gallavotti, G., Cohen, E.D.G.: Dynamical ensembles in stationary states. J. Stat. Phys. 80, 931–970 (1995)
33. Gardiner, C.: Stochastic Methods: A Handbook for the Natural and Social Sciences. Springer, Heidelberg (1996)
34. Glauber, R.J.: Coherent and incoherent states of the radiation field. Phys. Rev. 131, 2766–2788 (1963)
35. Goldstein, H., Poole, C.P., Safko, J.L.: Classical Mechanics, 3rd edn. Addison Wesley, New York (2001)
36. Goutis, C., Casella, G.: Explaining the saddlepoint approximation. Am. Stat. 53, 216–224 (1999)
37. Harris, R.J., Schütz, G.M.: Fluctuation theorems for stochastic dynamics. J. Stat. Mech., P07020 (2007). https://doi.org/10.1088/1742-5468/2007/07/P07020
38. Hatano, T., Sasa, S.-I.: Steady state thermodynamics of Langevin systems. Phys. Rev. Lett. 86, 3463–3466 (2001)
39. Jarzynski, C.: Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. Phys. Rev. E 56, 5018–5035 (1997)
40. Jarzynski, C.: Nonequilibrium equality for free energy differences. Phys. Rev. Lett. 78, 2690–2693 (1997)
41. Jarzynski, C.: Nonequilibrium work relations: foundations and applications. Eur. Phys. J. B 64, 331–340 (2008)
42. Kamenev, A.: Keldysh and Doi-Peliti techniques for out-of-equilibrium systems. In: Lerner, I.V., Althsuler, B.L., Fal’ko, V.I., Giamarchi, T. (eds.) Strongly Correlated Fermions and Bosons in Low-Dimensional Disordered Systems. Springer, Heidelberg, pp. 313–340 (2002)
43. Keldysh, L.V.: Diagram technique for nonequilibrium processes. Sov. Phys. JETP 20, 1018 (1965)
44. Krishnamurthy, S., Smith, E.: Solving moment hierarchies for chemical reaction networks. J. Phys. A Math. Theor. 50, 425002 (2017)
45. Kurchan, J.: Fluctuation theorem for stochastic dynamics. J. Phys. A 31, 3719 (1998)
46. Kurchan, J.: Non-equilibrium work relations. J. Stat. Mech. 2007, P07005 (2007)
47. Leok, M., Zhang, J.: Connecting information geometry and geometric mechanics. Entropy 19, 518 (2017)
48. Maier, R.S., Stein, D.L.: Escape problem for irreversible systems. Phys. Rev. E 48, 931–938 (1993)
49. Maier, R.S., Stein, D.L.: Oscillatory behavior of the rate of escape through an unstable limit cycle. Phys. Rev. Lett. 77, 4860–4863 (1996)
50. Maier, R.S., Stein, D.L.: A scaling theory of bifurcations in the symmetric weak-noise escape problem. J. Stat. Phys. 83, 291 (1996)
51. Maier, R.S., Stein, D.L.: Asymptotic exit location distributions in the stochastic exit problem. SIAM J. Appl. Math. 57, 752 (1997)
52. Martin, P.C., Siggia, E.D., Rose, H.A.: Statistical dynamics of classical systems. Phys. Rev. A 8, 423–437 (1973)
53. Mattis, D.C., Glasser, M.L.: The uses of quantum field theory in diffusion-limited reactions. Rev. Mod. Phys. 70, 979–1001 (1998)
54. Nagaoka, H., Amari, S.: Differential geometry of smooth families of probability distributions. Technical Report METR 82-7, U. Tokyo (1982)
55. Onsager, L., Machlup, S.: Fluctuations and irreversible processes. Phys. Rev. 91, 1505 (1953)
56. Owen, A.B.: Monte Carlo theory, methods and examples. http://statweb.stanford.edu/~owen/mc/ (2013)
57. Peliti, L.: Path-integral approach to birth-death processes on a lattice. J. Phys. 46, 1469 (1985)
58. Peliti, L.: Renormalization of fluctuation effects in a + a → a reaction. J. Phys. A 19, L365 (1986)
59. Perunov, N., Marsland, R., England, J.: Statistical physics of adaptation. Phys. Rev. X 6, 021036 (2015). arXiv:1412.1875v1 [physics.bio-ph]
60. Poletti, M., Esposito, M.: Irreversible thermodynamics of open chemical networks. I. Emergent cycles and broken conservation laws. J. Chem. Phys. 141, 024117 (2014)
61. Rao, C.R.: Information and accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc. 37, 81–91 (1945)
62. Rao, C.R.: Information and accuracy attainable in the estimation of statistical parameters. In: Kotz, S., Johnson, N.L. (eds.) Breakthroughs in Statistics: Springer Series in Statistics, pp. 235–247. Springer, New York (1992)
63. Schwinger, J.: Brownian motion of a quantum oscillator. J. Math. Phys. 2, 407–32 (1961)
64. Seifert, U.: Stochastic thermodynamics, fluctuation theorems, and molecular machines. Rep. Prog. Phys. 75, 126001 (2012). arXiv:1205.4176v1
65. Seifert, U., Speck, T.: Fluctuation-dissipation theorem++ in nonequilibrium steady states. Europhys. Lett. 89, 10007 (2010)
66. Siegmund, D.: Importance sampling in the Monte Carlo study of sequential tests. Ann. Stat. 4, 673–684 (1976)
67. Smith, E.: Quantum-classical correspondence principles for locally non-equilibrium driven systems. Phys. Rev. E 77, 021109 (2008). Originally as SFI preprint # 06-11-040
68. Smith, E.: Large-deviation principles, stochastic effective actions, path entropies, and the structure and meaning of thermodynamic descriptions. Rep. Prog. Phys. 74, 046601 (2011). arXiv:1102.3938 [cond-mat.stat-mech]
69. Smith, E.: Intrinsic and extrinsic thermodynamics for stochastic population processes with multi-level large-deviation structure. Entropy 22, 1137 (2020)
70. Smith, E., Krishnamurthy, S.: Symmetry and Collective Fluctuations in Evolutionary Games. IOP Press, Bristol (2015)
71. Smith, E., Krishnamurthy, S.: Flows, scaling, and the control of moment hierarchies for stochastic chemical reaction networks. Phys. Rev. E 96, 062102 (2017)
72. Sudarshan, E.C.G.: Equivalence of semiclassical and quantum mechanical descriptions of statistical light beams. Phys. Rev. Lett. 10, 277–279 (1963)
73. Touchette, H.: The large deviation approach to statistical mechanics. Phys. Rep. 478, 1–69 (2009). arxiv:0804.0327
74. van Kampen, N.G.: Stochastic Processes in Physics and Chemistry, 3rd edn. Elsevier, Amsterdam (2007)
75. Wahlberg, P.: The random Wigner distribution of gaussian stochastic processes with covariance in $s_0 \left( -2d \right)$. J. Funct. Spaces Appl. 3, 163–181 (2005)
76. Wigner, E.: On the quantum correction for thermodynamic equilibrium. Phys. Rev. 40, 749–759 (1935)
77. Wilf, H.S.: Generatingfunctionology, 3rd edn. A K Peters, Wellesley (2006)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.