



Mecánica Estadística Fuera de Equilibrio a Partir del Principio de Máximo Calibre

Tesis
entregada a la
Universidad de Chile
en cumplimiento parcial de los requisitos
para optar al grado de
Doctor en Ciencias con mención en Física
Facultad de Ciencias

por

Diego Sebastián González Díaz

Noviembre, 2018

Directores de Tesis: **Dr. Sergio Davis Irarrázabal**
Dr. Gonzalo Gutiérrez Gallardo

FACULTAD DE CIENCIAS
UNIVERSIDAD DE CHILE

INFORME DE APROBACIÓN
TESIS DE DOCTORADO

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de Doctorado presentada por el candidato

Diego Sebastián González Díaz

ha sido aprobada por la Comisión de Evaluación de la Tesis como requisito para optar al grado de Doctor en Ciencias con mención en Física, en el examen de Defensa de Tesis rendido el día 29 de Noviembre de 2018.

Directores de Tesis

Dr. Sergio Davis I.

Dr. Gonzalo Gutierrez G.

Comisión de Evaluación de la Tesis

Dr. Felipe Barra

Dr. Enrique Tirapegui

Dr. Sergio Curilef

Índice

1. Introducción	1
2. Teoría de la Probabilidad	16
2.1. ¿Qué es inferencia?	17
2.2. Probabilidad Bayesiana y Frecuentista	18
2.3. Notación Bayesiana	19
2.4. Reglas de la suma y el producto	20
2.5. Teorema de Bayes	21
2.6. Expectaciones	21
2.7. Normalización y Marginalización	23
2.8. Funciones delta en probabilidad bayesiana	24
3. Principio de Máxima Entropía	26
3.1. ¿Por qué maximizar entropía?	27
3.2. Principio de Máxima Entropía	29
3.3. Ejemplos de construcción de modelos con MaxEnt	32
3.3.1. Modelo sin restricciones: entendiendo el <i>prior</i>	32
3.3.2. Modelo para un experimento con muestras aleatorias positivas	33
3.3.3. Modelo gaussiano a partir de MaxEnt	34

4. Mecánica Estadística desde el Principio de Máxima Entropía	37
4.1. Mecánica Estadística desde la probabilidad bayesiana	38
4.2. Densidades de probabilidades de diferentes <i>ensembles</i> de la ME desde MaxEnt	41
4.2.1. Ensemble Micro-Canónico	42
4.2.2. <i>Ensemble</i> Canónico	44
4.2.3. <i>Ensemble</i> Gran-Canónico	45
4.2.4. <i>Ensemble</i> Isotérmico Isobárico	46
4.2.5. Estadística de Bose-Einstein y Fermi-Dirac	48
4.3. Termodinámica desde la inferencia	53
4.3.1. Energía Libre	53
5. Principio de Máximo Calibre	56
5.1. Máximo Calibre y la Mecánica Estadística Fuera del Equilibrio	56
5.2. Notación de Funcionales	57
5.2.1. Derivada Funcional: Valores extremos de un funcional	58
5.2.2. Probabilidad Funcional	61
5.2.3. Integral Funcional: Expectaciones de Funcionales y Funciones	62
5.2.4. Delta de Dirac funcional	65
5.3. Principio de Máximo Calibre	65
5.3.1. MaxCal para restricciones de funcionales arbitrarios	66
5.3.2. MaxCal para restricciones de funciones arbitrarias	68
5.4. Implementación numérica de MaxCal	70
5.4.1. Monte Carlo Metropolis en el espacio de trayectorias	71
5.4.2. Curvas de Bézier para simular trayectorias	74

	IV
5.4.3. Aproximación de curvas como líneas rectas	75
5.5. Ejemplos	76
5.5.1. <i>Prior</i> funcional: MaxCal cuando no hay restricciones	76
5.5.2. Derivada Funcional de la acción clásica.	77
5.5.3. La trayectoria más probable: Principio de mínima acción desde MaxCal.	78
5.5.4. Dinámicas Newtonianas desde MaxCal	79
5.5.5. Relación entre n -flujos	82
5.5.6. Simulación numérica: Ecuación de movimiento de la partícula libre	84
5.5.7. Simulación numérica: Ecuación de movimiento del oscilador armónico	84
6. Mecánica Estadística Fuera del Equilibrio	90
6.1. Inferencia sobre trayectorias	92
6.1.1. Axioma fundamental de la NESM	93
6.1.2. Ecuación del Corte temporal	95
6.1.3. NESM desde inferencia de trayectorias: La hipótesis de Jaynes	97
6.1.4. Expectaciones Cortadas	99
6.1.5. Ecuación de Chapman-Kolmogorov en el espacio de trayectorias	101
6.1.6. Ecuación de continuidad	106
6.2. NESM desde MaxCal	111
6.2.1. Ecuación de Difusión	112
6.2.2. Ecuación de Liouville	115

6.2.3. Fórmula de Feynman-Kac: Ecuación tipo Schrödinger para probabilidades	120
6.3. Relaciones Fuera del Equilibrio	125
6.3.1. Primera ley de la Termodinámica microscópica	126
6.3.2. Evolución Markoviana	127
6.3.3. Teorema de Fluctuación de Crooks	130
6.3.4. Igualdad de Jarzynski	134
7. Conclusiones	136
A. Definiciones útiles en mecánica estadística	142
A.1. Teoría de distribuciones	142
A.1.1. Delta de Dirac	142
A.2. Fluctuaciones e incerteza	142
B. Ecuación de Chapman-Kolmogorov	144
B.1. Ecuación de Chapman-Kolmogorov	144
C. Teoremas para la mecánica estadística en equilibrio y fuera del equilibrio	146
C.1. Teorema de la Variable Conjugada (CVT)	146
C.2. Teorema de la Variable Conjugada para Funcionales (CVTF)	147
C.3. Teorema de Fluctuación-Disipación (FDT)	148
C.4. Teorema Fluctuación-Disipación para Funcionales (FDTF)	149
D. Publicaciones	158

RESUMEN

La Mecánica Estadística en equilibrio es una teoría probabilística bien establecida, la cual puede ser explicada por medio de un principio fundamental, el principio de máxima entropía, donde por medio de esta teoría se pueden deducir las leyes que rigen la termodinámica. Por el contrario, la Mecánica Estadística fuera del equilibrio carece de una fundamentación axiomática o incluso de una formulación unificada; más bien consta de un conjunto de teoremas y relaciones de validez particular (por ejemplo en situaciones estacionarias) sin un hilo conector claro. El único elemento común a todos estos fragmentos es el uso de probabilidades y expectativas dependientes del tiempo, típicamente relacionadas por medio de ecuaciones diferenciales parciales o relaciones de validez general.

Dado lo anterior, sería de vital interés el poder establecer un principio que nos permita deducir la estructura actual de las teorías parciales de la Mecánica Estadística fuera del equilibrio, de manera análoga a como la estructura de la Mecánica Estadística en equilibrio se deduce a partir del principio de máxima entropía.

En esta tesis se mostrará como por medio de el principio de Máximo Calibre (o Máxima Entropía de Trayectorias) y lo que llamaremos inferencia sobre trayectorias es posible derivar y relacionar por completo estos elementos que conforman la mecánica estadística fuera del equilibrio, estableciendo un método sistemático y ordenado para entender el origen de las relaciones y ecuaciones diferenciales que conforman esta teoría, Esta es una primera aproximación a una teoría que nos permita entender la termodinámica fuera del equilibrio.

Capítulo 1

Introducción

En física hay tres teorías macroscópicas clásicas bien establecidas: Mecánica Clásica, Electricidad y Magnetismo y Termodinámica [1].

Estas teorías son llamadas macroscópicas, porque describen el comportamiento de cuerpos constituidos por un gran número de partículas como por ejemplo electrones, neutrones, protones, entre otras, donde en lugar de describir el comportamiento individual de cada partícula (descripción microscópica), se describen los sistemas tomando promedios de todos sus grados de libertad microscópicos, llamando a estos promedios, propiedades macroscópicas.

El entender los sistemas de esta forma, permite tener cantidades macroscópicas que no cambian a la misma tasa que lo hacen cantidades microscópicas. Por ejemplo consideremos las variaciones que puede tener la posición de un electrón en un circuito comparado con las variaciones en la corriente del circuito: la posición tendrá muchos más cambios que la corriente, o las variaciones en las posiciones de las partículas en un vaso de agua al aire libre, con las variaciones en su temperatura: la primera tendrá más variaciones. Las propiedades macroscópicas entregan una descripción reducida de los sistemas, donde no se requiere saber con exactitud cada posición de las partículas que forman al sistema.

La distinción entre estas tres teorías macroscópicas está dada por la naturaleza de las observaciones usadas para modelarlas, es decir, cuales son los fenómenos que describen cada una de estas teorías.

En particular, en esta tesis nos centraremos como objetivo final el entender la Termodinámica [2, 3], como la teoría macroscópica que busca describir la relación entre propiedades asociadas a la transferencia de calor, como lo son la presión, la temperatura, el potencial químico, la entropía, la energía, el volumen, el número de partículas (o masa), entre otras.

La termodinámica fue desarrollada inicialmente de forma experimental por R. Boyle en 1662, quien estudió la relación entre la presión y el volumen de un gas cuando se mantenía a temperatura constante.

Con la creación de los primeros motores a vapor correspondientes a la base tecnológica de la revolución industrial, el entender su funcionamiento y como hacerlos mas eficientes motivó a distintos investigadores tales como Kelvin, Gay-Lussac, Bernoulli, Carnot, Clausius, entre otros, a estudiar las relaciones entre las propiedades macroscópicas de los sistemas, estableciendo a la Termodinámica como una de las principales ramas de la física, desarrollada teóricamente a partir de evidencias empíricas.

Uno de los resultados experimentales más importantes fue el desarrollado por Clausius [4, 5], quien estableció que para los procesos termodinámicos *es imposible efectuar una transformación cuyo único resultado final sea transferir calor de un cuerpo a una temperatura dada a un cuerpo a una temperatura mayor* [2]. Este postulado es conocido como la segunda ley de la termodinámica.

De acuerdo a los trabajos de Clausius y posteriormente Kelvin, hay una cantidad de calor que se disipa Q , la cual no puede ser utilizada como trabajo, definiendo

gracias a esta observación una cantidad llamada entropía S , donde el cambio de entropía ΔS se relaciona directamente con el calor disipado por medio de la relación $T\Delta S = \Delta Q$, donde T corresponde a la temperatura absoluta (temperatura en escala de Kelvin, definida positiva y 0 como su mínimo) en la que ocurre el proceso.

Las cantidades termodinámicas tales como la entropía S , la energía E , el volumen V y el número de partículas N , son llamadas cantidades extensivas, siendo directamente dependientes del tamaño y/o masa del sistema. Es por esto que si consideramos dos cajas, cada una con determinada entropía, energía, volumen y número de partículas (S, E, V, N) , igual para ambas, el sistema formado al unir estas dos cajas tendrá el doble de energía, de entropía, de volumen y de número de partículas $(2S, 2E, 2V, 2N)$.

Por otra parte, las cantidades tales como la temperatura T , presión p y potencial químico μ , son llamadas cantidades intensivas, y se caracterizan por ser independientes del tamaño o masa total del sistema. Estas cantidades se definen mediante el uso de cantidades extensivas, y son obtenidas por lo general mediante las relaciones entre los cambios de las cantidades extensivas, como por ejemplo la temperatura $T = \frac{\partial E}{\partial S}$. Dado esto, si tenemos dos cajas a determinada temperatura, presión y potencial químico (T, p, μ) , el sistema formado al unir estas dos cajas seguirá teniendo la misma temperatura, presión y potencial químico (T, p, μ) .

La formulación teórica de la termodinámica requiere tener en claro cuales son las cantidades que se busca relacionar en un sistema, para así definir una llamada ecuación de estado. La ecuación de estado es una ecuación que relaciona las variables que describen un sistema en equilibrio termodinámico, y que se obtiene cuando se

varía una cantidad dejando las otras constantes. Una de las ecuaciones de estado mas conocidas es la ecuación del gas ideal, $pV = Nk_B T$.

De forma axiomática, la termodinámica moderna puede ser obtenida con las llamadas leyes de la termodinámica [1]. Estas leyes son las siguientes:

- Primera Ley:

'La variación de energía de un sistema durante una transformación cualquiera es igual a la suma de la energía que el sistema recibe del medio circundante' [2].

item Segunda Ley:

'Es imposible efectuar una transformación cuyo único resultado final sea transferir calor de un cuerpo a una temperatura dada a un cuerpo a una temperatura mayor', (Postulado de Clausius). [2].

La primera ley indica que el cambio de energía de un sistema ΔE , está dado por el trabajo W realizado sobre el sistema más el calor Q disipado por el mismo. Este principio es conocido como el principio de conservación de energía para la termodinámica $\Delta E = W - Q$, visto desde un punto de vista mecánico, si realizamos trabajo sobre un sistema, esta energía se convertirá en energía interna E o en calor Q , sin perder energía, solo transformándola. Por otra parte la segunda ley hace referencia a la imposibilidad de construir un *móvil perpetuo*, en el cual todo el calor se convierta en trabajo y que luego todo el trabajo sea convertido en calor, existe amplia evidencia del cumplimiento de la segunda ley, basados principalmente en el fracaso de todos los esfuerzos realizados para construir el *móvil perpetuo* [2].

Paralelo al desarrollo de la Termodinámica, cerca del año 1860, mediante las contribuciones de Maxwell y Boltzmann, entre otros, comenzó el desarrollo de una

teoría en la cual se introdujo un análisis a nivel microscópico de los sistemas, con el cual fue posible entender la naturaleza estadística de los sistemas termodinámicos. Esta teoría es la que hoy llamamos Mecánica Estadística (ME) [6–9].

Para entender la formulación de la ME, es importante entender que todo sistema macroscópico (y en particular los sistemas termodinámicos) está formado por un conjunto de partículas, las cuales tienen una posición \vec{x}_i y momenta \vec{p}_i bien definidos para cada una de ellas, siendo posible tener una descripción microscópica completa del sistema mediante el vector $(\vec{x}_1, \dots, \vec{x}_n, \vec{p}_1, \dots, \vec{p}_n)$. A este vector con las posiciones y los momenta lo llamaremos el micro-estado del sistema. El que un sistema se encuentre en equilibrio termodinámico no quiere decir que se encuentra en un micro-estado fijo, sino por el contrario, un sistema puede tener un gran número de micro-estados compatibles con ciertas cantidades macroscópicas fijas. Un ejemplo para ver esto es pensar en un sistema donde los momenta promedio de las partículas sea \bar{p} . Pueden existir infinitas combinaciones de (p_1, \dots, p_n) tal que formen un momenta promedio $\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i$, siendo todas esas diferentes combinaciones de momenta, micro-estados compatibles con la cantidad macroscópica \bar{p} para el sistema.

La ME puede ser entendida de forma axiomática mediante un principio llamado Principio de Máxima Entropía [1, 3]. Este principio establece que la distribución de probabilidad para los diferentes micro-estados de un sistema se obtiene maximizando la entropía S , cuya forma analítica en términos de los micro-estados fue postulada en un inicio por Boltzmann, y luego generalizada por Gibbs.

En 1890 Ludwig Boltzmann definió la entropía como $S = k_B \log(\Omega)$, donde Ω corresponde al número total de micro-estados en que se puede encontrar un sistema y k_B es llamada constante de Boltzmann, utilizada para tener consistencia en las unidades al definir la entropía, teniendo unidades de energía sobre temperatura.

La entropía de Boltzmann asume que todos los micro-estados en los que se puede encontrar un sistema son equiprobables, cada uno con probabilidad $1/\Omega$.

En 1878 Josiah Gibbs [10, 11], mostró un método sistemático para construir las funciones de probabilidad de los sistemas, conocido como método de Gibbs. Este método se basa en el concepto de *ensemble*, que son copias mentales del sistema macroscópico con diferente estado microscópico, donde las cantidades de interés físico se obtienen como promedio de estas cajas, o como se acostumbra decir, promedio sobre *ensemble*. Gibbs define la entropía de la forma $S = -\sum_i P_i \log P_i$, una cantidad dependiente de la forma de las probabilidades P_i de los micro-estados que describen el sistema y que debe ser maximizada para obtener las probabilidades de sistemas en equilibrio. Este método permitió entender los fenómenos termodinámicos desde la estadística, siendo posible derivar las relaciones termodinámicas sin la necesidad de resultados experimentales, sino que mediante un análisis microscópico de los sistemas.

La ME tiene como objetivo principal encontrar la probabilidad de cada micro-estado en que se encuentra un sistema $P(x, p)$ donde $x = (x_1, \dots, x_n)$ y $p = (p_1, \dots, p_n)$. Para esto es necesario saber las condiciones que restringen al sistema. Estas condiciones pueden venir desde un análisis macroscópico, como lo es un sistema que se encuentra en contacto con un baño térmico, lo que restringe al sistema a mantener una temperatura constante, o desde un análisis microscópico, como lo es considerar un gran número de partículas interactuando por medio de un potencial.

En la ME podemos hablar de equilibrio termodinámico cuando las propiedades macroscópicas del sistema se mantienen constantes en el tiempo. Las probabilidades para cada micro-estado pueden ser obtenida por varios caminos que involucran maximizar entropía, tales como la entropía de Gibbs, la entropía de Boltzmann, entre

Condiciones del Sistema	Distribución de Probabilidad	Ensemble
Sistema Aislado (Energía Constante)	$P(x) = \frac{\delta(H(x) - E)}{Z(E)}$	Micro-Canónico
Sistema en Baño Térmico (Temperatura Constante)	$P(x) = \frac{e^{-\beta H(x)}}{Z(\beta)}$	Canónico
Sistema en Baño Térmico Y Baño de Presión (Temperatura y Presión Constante)	$P(x) = \frac{e^{-\beta H(x) - \beta p v(x)}}{Z(\beta)}$	Isotérmico-Isobárico
Sistema en Baño Térmico y Baño de Partículas (Temperatura y Potencial Químico Constante)	$P(x) = \frac{e^{-\beta H(x) - \beta \mu N(x)}}{Z(\beta)}$	Gran-Canónico

Figura 1.1: Formas de las probabilidades en Mecánica Estadística dadas condiciones externas, obtenidas maximizando entropía de Boltzmann y entropía de Gibbs.

otras y tienen una forma funcional dependiente de las condiciones macroscópicas del sistema, como se muestra en la figura 1.1.

Utilizando la función de probabilidad asociada a cada *ensemble*, se obtienen predicciones para las propiedades macroscópicas de los sistemas, emergiendo una estructura teórica para derivar la termodinámica, teoría que como se explicó anteriormente, describe cómo se relacionan dichas propiedades macroscópicas entre sí. Esta es una línea de trabajo bien definida para derivar la termodinámica, comenzando por encontrar la distribución de probabilidad para cada micro-estado dadas las condiciones

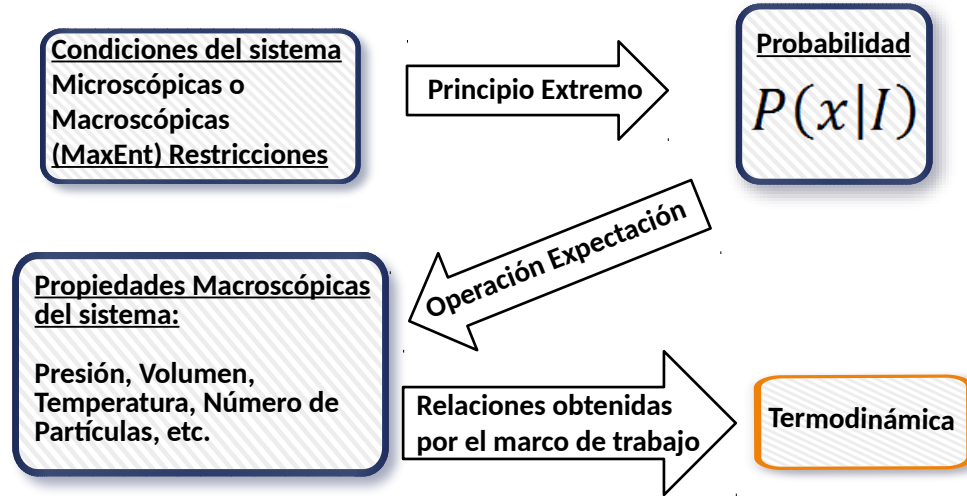


Figura 1.2: Estructura de la Mecánica Estadística para derivar Termodinámica.

macroscópicas del sistema, para luego, utilizar la probabilidad para obtener nuevas cantidades macroscópicas del sistema, buscando así, entender como se relacionan estas cantidades tal como se muestra en la figura 1.2.

En 1957 E. T. Jaynes postula el Principio de Máxima Entropía (MaxEnt) [12, 13](Cap. 3) como una manera de generalizar el método propuesto por Gibbs para entender la ME de forma axiomática a partir de un principio extremo.

En su versión moderna, MaxEnt establece que la distribución de probabilidad más objetiva (o la menos ambigua) para describir el estado de equilibrio de un sistema es aquella que maximiza la entropía de Gibbs-Shannon $S = -\sum_i P_i \log \frac{P_i}{\pi_i}$, sujeta a restricciones. Estas están dadas por las condiciones externas en que se encuentra

<u>Condiciones del Sistema y sus restricciones</u>	MaxEnt	<u>Distribución de Probabilidad</u>
Sistema Aislado $H(x) = E$		$P(x) = \frac{\delta(H(x) - E)}{Z(E)}$
Sistema en Baño Térmico $\langle H(x) \rangle = E$		$P(x) = \frac{e^{-\beta H(x)}}{Z(\beta)}$
Sistema en Baño Térmico Y Presión Constante $\langle H(x) \rangle = E \quad \langle v(x) \rangle = V$		$P(x) = \frac{e^{-\beta H(x) - \beta p v(x)}}{Z(\beta)}$
Sistema en Baño Térmico y Potencial Químico Constante $\langle H(x) \rangle = E \quad \langle n(x) \rangle = N$		$P(x) = \frac{e^{-\beta H(x) - \beta \mu N(x)}}{Z(\beta)}$

Figura 1.3: Mecánica Estadística mediante el Principio de Máxima Entropía.

el sistema, siendo P_i la probabilidad del i -ésimo micro-estado dadas las restricciones y π_i corresponde a la probabilidad del i -ésimo micro-estado cuando no se han considerado las restricciones. Este principio generaliza la idea de condiciones en las que se encuentra un sistema a restricciones, las cuales son dadas en forma de expectativa tal como se muestra en la figura 1.2.

Es así como el principio de Máxima Entropía entrega un método sistemático para asignar distribuciones de probabilidad dado un conjunto de restricciones en formas de expectativa, como se muestra en la figura 1.3.

El resultado mas importante del trabajo de Jaynes no fue derivar la Mecánica Estadística por medio de maximizar entropía, ya que esto era bien entendido an-

teriormente por Gibbs, sino que fue comprender que el maximizar entropía sujeto a restricciones, corresponde a una regla general para realizar inferencia, es decir, estimar cantidades cuando se tiene información incompleta. Esta forma de ver la mecánica estadística ha tomado gran importancia en la física contemporánea por su rol unificador de la teoría [14–16], contribuyendo en gran medida a entender de un modo diferente el significado de las diversas ecuaciones que rigen a los sistemas.

De esta forma, *ver el método de máxima entropía como una regla general de inferencia, sire para entender que esta regla puede aplicarse también al entendimiento de la Mecánica Estadística Fuera del Equilibrio (Non-equilibrium Statistical Mechanics, NESM).*

La NESM tiene como objetivo principal la descripción de la evolución temporal de propiedades macroscópicas dada una situación inicial conocida, típicamente un estado de equilibrio termodinámico. Para esta labor, se requiere una distribución de probabilidad tiempo dependiente para los micro-estados del sistema $P(x, p; t)$. Utilizando esta distribución de probabilidad es posible obtener las propiedades macroscópicas en cualquier instante.

Los principales elementos que componen la NESM tradicional [17–19] son, por un lado, ecuaciones diferenciales parciales (PDE) para la distribución de probabilidad dependiente del tiempo $P(x, p; t)$ tales como, la ecuación de Liouville [3], ecuaciones de tipo Fokker-Planck [20], ecuación maestra [19], entre otras. Por otra parte, existe relaciones e igualdades tales como la Igualdad de Jarzynski [21], la Igualdad de Crooks [22–24] y los Teoremas de Fluctuación [25–27]. Donde al igual que en la ME, por medio del cálculo de expectativas se obtienen las diferentes cantidades macroscópicas tiempo-dependientes del sistema, lo que conduce a una teoría termo-

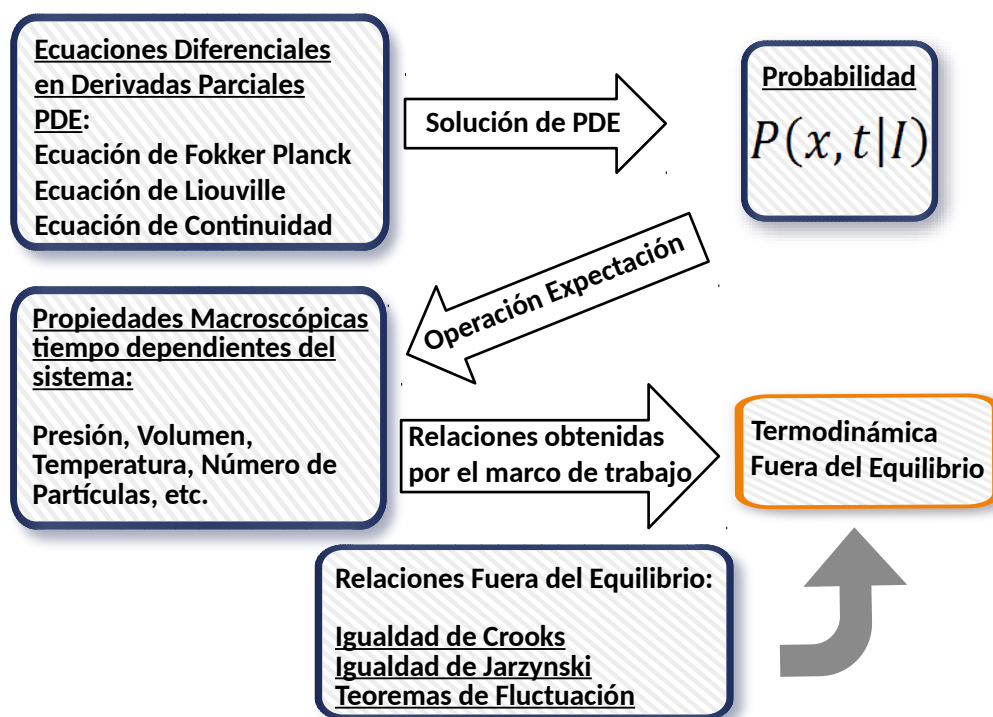


Figura 1.4: Diagrama de la estructura general NESM y sus componentes.

dinámica tiempo-dependiente o fuera del equilibrio. Estos elementos de la NESM tradicional y su estructura, se muestran en la figura 1.4.

La ME se puede derivar por completo mediante el principio de Máxima Entropía como principio extremo tal como se muestra en la figura 1.1. A partir de esto, surge la pregunta si es posible hacer lo mismo para la NESM, es decir, encontrar un método sistemático por medio del cual, una vez conocidas las condiciones del sistema (o restricciones) sea posible obtener la función de probabilidad tiempo-dependiente. La respuesta a esta pregunta es positiva, efectivamente afirmamos que mediante una generalización del principio de Máxima Entropía llamado principio de Máximo Calibre, es posible encontrar un método sistemático para derivar los principales elementos de la NESM tradicional.

Para cumplir la labor de derivar la NESM desde el principio de Máxima Entropía, E. T. Jaynes propuso el denominado Principio de Máximo Calibre (MaxCal) [28]. Este es igual a MaxEnt, solo que definido en el espacio de micro-funciones o micro-trayectorias (\tilde{x}, \tilde{p}) en lugar de micro-estados (x, p) .

De acuerdo a MaxCal, dadas restricciones de cantidades fuera del equilibrio (o restricciones tiempo-dependientes) es posible obtener la probabilidad para cada micro-función del sistema, es decir, un funcional de probabilidad dependiente de micro-funciones (\tilde{x}, \tilde{p}) . Jaynes profundizó en el entendimiento de este principio, entregando las herramientas necesarias para construir funcionales de probabilidad $P[\tilde{x}, \tilde{p}]$, pero nunca detalló cómo desde este formalismo se logra obtener la función de probabilidad tiempo-dependiente $P(x, p; t)$, que corresponde al elemento central de la NESM.

El resultado central de esta tesis es la obtención de la función de probabilidad tiempo-dependiente para los micro-estados $P(x, p; t)$ a partir del funcional de probabilidad de las micro-funciones $P[\tilde{x}, \tilde{p}]$. Esto se realiza mediante una ecuación que

llamaremos ecuación del corte temporal,

$$\int d\tilde{x} d\tilde{p} P[\tilde{x}, \tilde{p}] \delta(\tilde{x}_t - x) \delta(\tilde{p}_t - p) = P(x, p, t). \quad (1.1)$$

Esta ecuación, que será vista en detalle mas adelante, indica que para obtener $P(x, p; t)$ desde $P[\tilde{x}, \tilde{p}]$ es necesario resolver una integral en el espacio de las microfunciones o integral de caminos (*path integral*). La importancia de esta ecuación recae en entender que existe un paralelo entre resolver una integral de caminos (obtenida desde MaxCal de forma natural) y resolver una PDE, conocida como fórmula de Feynman-Kac [29, 30].

La fórmula de Feynman-Kac y la ecuación del corte temporal son las piezas faltantes para entender cómo a partir de restricciones de cantidades fuera del equilibrio se obtienen las PDE centrales que componen la NESM (como se muestra en la figura 1.5), por lo que Máximo Calibre cumple el rol de ser un principio variacional completo con el cual puede ser derivada la NESM a partir de restricciones [31, 32].

En esta tesis se mostrará además que por medio de MaxCal y el cálculo funcional [33], es posible derivar todos los elementos que componen la estructura de la NESM tradicional. De este modo, es posible sentar las bases de la termodinámica de procesos fuera del equilibrio, de la misma forma que con la Mecánica Estadística es posible derivar la Termodinámica de sistemas en equilibrio.

Esta tesis está organizada de la siguiente manera: en el segundo capítulo se explicarán las bases de la probabilidad bayesiana, pasando por los axiomas de la probabilidad de Cox y el teorema de Bayes. Las probabilidades Bayesianas, a diferencia de las probabilidades frecuentistas, se caracterizan por estar definidas bajo un contexto o estado de conocimiento, el cual corresponde a un elemento extra incluido en la notación y que como veremos más adelante permite dejar en claro qué cantidades

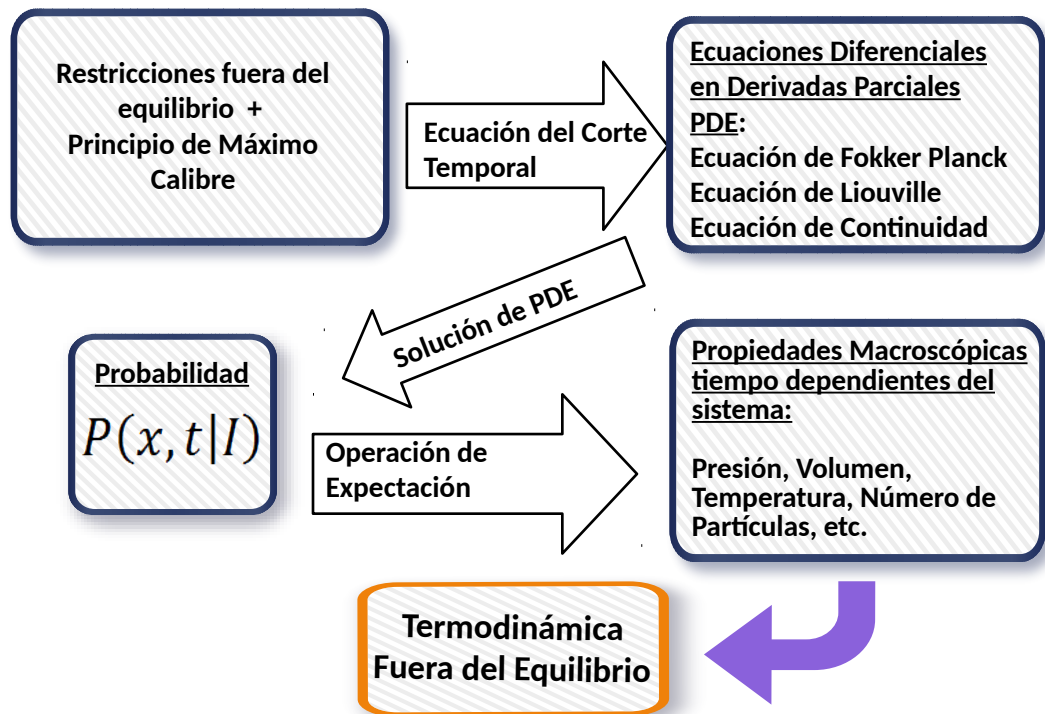


Figura 1.5: Diagrama de la estructura general NESM y sus componentes.

son conocidas para construir una probabilidad.

El tercer capítulo explica el principio de máxima entropía postulado por E.T. Jaynes, mostrando como MaxEnt es un método sistemático que permite obtener la probabilidad de los micro-estados cuando se tiene cantidades macroscópicas conocidas. En este capítulo se muestra la estructura más fundamental del formalismo y cómo operar con él.

El cuarto capítulo es una aplicación directa de MaxEnt, que muestra cómo mediante el uso de este principio es posible derivar la ME y sus componentes, permitiendo así entender los principales componentes de esta mediante la idea de restricciones.

En el quinto capítulo se mostrará la formulación del principio de máximo calibre, mostrando como este principio tiene una estructura similar a MaxEnt, solo que en este caso se trabaja en el espacio de las micro-trayectorias. Para ésto, se sentarán las bases del cálculo funcional, principal herramienta matemática a utilizar para construir el funcional de probabilidad a partir de restricciones de cantidades tiempo-dependientes (cantidades fuera del equilibrio.).

En el sexto capítulo muestra las bases de cómo MaxCal es un principio variacional fundamental para obtener la NESM, deduciendo a partir de este principio junto con la ecuación del corte temporal las principales PDEs que componen la NESM, así como también se mostrará cómo entender las relaciones fuera del equilibrio de la NESM a partir del uso de las reglas del cálculo funcional. Se demuestra por qué estas relaciones son de validez general, siendo consecuencia directa de la notación funcional y la teoría de probabilidad bayesiana.

Finalmente en el séptimo capítulo se discuten las conclusiones y se presentan las perspectivas de desarrollo de esta teoría.

Capítulo 2

Teoría de la Probabilidad

“Una paradoja es simplemente un error fuera de control; es decir, uno que ha atrapado a tantas mentes incautas que se ha hecho público, institucionalizado en nuestra literatura, y se enseña como verdad.”

— E. T. Jaynes

En este capítulo se explicarán los conceptos básicos que son necesarios para desarrollar una teoría de probabilidad bayesiana, entendiendo primero qué es la inferencia y cuándo nos enfrentamos a un contexto en el cual hay que utilizar la inferencia. Para esto utilizaremos los axiomas de la probabilidad propuestos por Richard Cox. Mostrando lo útil que es trabajar con la notación bayesiana, tanto desde un punto de vista analítico, donde se detalla el contexto bajo el cual se crea la probabilidad, como desde un punto de vista operacional, relacionando el comportamiento de las probabilidades con funciones análogas a las deltas de Dirac y Kronecker. Es por esto que desde un punto de vista fundamental, sólo la teoría de la probabilidad es necesaria para poder construir funciones de probabilidad y actualizarlas con información, mediante el uso del teorema de Bayes.

Finalmente se mostrará cómo se define la operación expectación, la cual nos

permitirá calcular cantidades macroscópicas teniendo en claro el contexto en que trabaja este operador, al igual que las probabilidades bayesianas.

2.1. ¿Qué es inferencia?

En la física y en general en ciencias, cuando queremos evaluar una hipótesis (o responder una pregunta), estamos sujetos a dos posibles ambientes: en primer lugar, uno deductivo, en el cual tenemos todos los elementos para poder dar una respuesta concreta, un verdadero o falso, posible de expresar también como 0 o 1. El otro ambiente es el de la inferencia, en el cual no tenemos todas las piezas o elementos para dar una respuesta concreta, sino que sólo podemos dar una estimación de cuán válida puede ser una respuesta, un valor intermedio entre verdadero o falso, debiendo asignar un “valor de verdad” entre 0 y 1.

Un ejemplo que ayuda a entender esto es pensar en un rompecabezas, en el cual nuestra hipótesis a evaluar es la imagen que se forma al armarlo. En el primer caso, en un ambiente deductivo, tenemos todas las piezas que conforman el rompecabezas, y si se pregunta ¿Qué imagen tiene el rompecabezas?, a pesar de que puede ser tedioso armarlo, está dentro de nuestras capacidades lograr este objetivo, por lo que si somos metódicos y rigurosos, lograremos tener la imagen completa y dar una respuesta concreta a la pregunta. En el segundo caso, en un ambiente de inferencia, es análogo a no tener todas las piezas del rompecabezas, por lo que si alguien nos pregunta ¿Qué imagen tiene el rompecabezas?, a pesar que ocupemos todos los métodos y seamos lo suficientemente rigurosos en el proceso de armado, nunca podremos completarlo, por lo que sólo podremos dar respuestas con un cierto grado de verdad, el cual será absoluto en el momento que se posea todas las piezas para armar dicho rompecabezas;

este valor intermedio entre verdadero y falso es lo que llamaremos probabilidad, un cierto grado de verdad que podemos asignar a nuestra hipótesis que estamos evaluando, dada cierta información conocida.

En esta tesis, llamaremos inferencia al uso de las leyes de la probabilidad a partir de los axiomas de Cox [34] para entender el cálculo de las probabilidades con notación bayesiana [35], junto con el uso del principio de máxima entropía propuesto por Jaynes [12, 13, 36] como un método sistemático para construir la forma que tienen las probabilidades. De esta manera, la inferencia entrega un marco de trabajo para resolver problemas mediante el uso de las probabilidades. En las siguientes secciones de este capítulo se mostrará el primer componente de la inferencia, las probabilidades bayesianas.

2.2. Probabilidad Bayesiana y Frecuentista

Cuando se habla de probabilidad puede uno referirse a dos posibles interpretaciones las cuales en ciertos límites son compatibles. Por una parte está la *probabilidad frecuentista*, donde se define la probabilidad de un suceso aleatorio A como la frecuencia con que dicho suceso ocurre, es decir, la razón

$$p(A) = \frac{\text{Número de sucesos donde ocurre } A}{\text{Número de sucesos totales}}. \quad (2.1)$$

Esta interpretación de la probabilidad es definida de forma rigurosa en base a la teoría de la medida por medio de los axiomas de Kolmogorov [37].

Por otro lado, se encuentra la *Probabilidad Bayesiana* donde se define la probabilidad de una *proposición lógica* A como un número real $P(A|\mathcal{I}_0)$ tal que $0 < P(A|\mathcal{I}_0) < 1$ y donde \mathcal{I}_0 representa un contexto o estado de conocimiento. Esta interpretación de la probabilidad se basa en los axiomas de R. Cox [34] para las

probabilidades, cuya consecuencia es la regla de la suma y la regla del producto con la cual se deduce el Teorema de Bayes [36, 38]. Estas leyes serán vistas en detalle luego de aclarar la notación bayesiana.

En una interpretación completamente frecuentista, existen variadas preguntas que no tiene mucho sentido hacer, como por ejemplo preguntarnos por un hecho histórico que pudo o no haber ocurrido: de forma frecuentista, debemos pensar que existen diferentes escenarios donde en algunos este hecho ocurrió y otros en los que no, lo que nos obliga a suponer cosas más complejas que no son necesarias en una interpretación bayesiana. Esta última responde una pregunta utilizando sólo la información que tenemos, siendo en este sentido lo más objetiva posible.

En esta tesis se utilizarán probabilidades bayesianas, debido a que buscamos entender las probabilidades desde un contexto de inferencia, es decir, dar la mejor estimación cuando tenemos información incompleta.

2.3. Notación Bayesiana

Para entender la notación bayesiana, se requiere primero comprender qué son las proposiciones lógicas y cómo se evalúan dentro de las probabilidades. Una proposición lógica es una oración a la cual se le puede asignar un valor de verdad, por ejemplo podemos tener la proposición lógica $A = \text{“Mañana es viernes”}$, la cual podemos evaluar con la información que tenemos, ya que si hoy es jueves, entonces A es verdadero, por el contrario si hoy es lunes, entonces A es falso.

Una diferencia importante entre probabilidades bayesianas y frecuentistas es que las primeras tienen que estar bajo un grado de conocimiento o contexto. Para esto, se utiliza la notación *la probabilidad de A dado B* escrito de la forma $P(A|B)$, donde A corresponde a una proposición lógica a la cual se pretende evaluar su grado de

verdad cuando sabemos que la proposición B es verdadera.

Es así como las probabilidades bayesianas nos permiten evaluar proposiciones lógicas utilizando otras proposiciones como información. Si volvemos al ejemplo anterior donde $A = \text{“Mañana es viernes”}$ y $B = \text{“Hoy es lunes”}$, entonces la probabilidad de que $\text{“Mañana es viernes”}$ dado que “Hoy es lunes” es cero, ya que tenemos la información necesaria para tener una respuesta $P(A|B) = 0$. Por otra parte, si $B = \text{“Hoy es jueves”}$ entonces $P(A|B) = 1$, ya que dada la información B podemos saber con seguridad que A es verdadero.

De esta modo, es fácil notar que la probabilidad que A sea verdadero dado que A es verdadero es una tautología, es decir, siempre es verdad

$$P(A|A) = 1. \quad (2.2)$$

Una propiedad de la notación que usaremos, es que cuando queremos evaluar más de una proposición, se entenderá el uso de la coma (,) como la conjunción lógica, la cual se lee como “y”. Es decir, si queremos preguntar por la probabilidad de que A y B sean verdaderos dado que I es verdadero, se escribe entonces $P(A, B|I)$. De forma análoga es lo mismo preguntarnos por la probabilidad de que B y A sean verdaderos dado I , $P(B, A|I)$. Entonces podemos notar que las probabilidades bayesianas siguen la igualdad

$$P(A, B|I) = P(B, A|I). \quad (2.3)$$

2.4. Reglas de la suma y el producto

Basándonos en los axiomas de la probabilidad definidos por R. Cox [34], se deducen dos reglas de la probabilidad bayesiana, la regla de la suma

$$P(A|I) = 1 - P(\neg A|I) \quad (2.4)$$

y la regla del producto

$$P(A, B|I) = P(A|I)P(B|A, I). \quad (2.5)$$

La derivación de estas reglas requiere el uso de lógica avanzada y puede ser vista en [34, 39]. Existen en la actualidad autores que siendo consistentes con las leyes de Cox, definen nuevas formas de utilizar esta notación, ampliándola a otros posibles análisis [40, 41].

2.5. Teorema de Bayes

Al utilizar la conmutatividad de la conjunción (2.3) y luego la regla del producto (2.5), se obtiene entonces el teorema de Bayes

$$P(B|A, I) = \frac{P(B|I)P(A|B, I)}{P(A|I)}. \quad (2.6)$$

Este teorema es una importante pieza en lo que respecta a la inferencia, ya que nos indica cómo cambia una probabilidad $P(B|I)$ cuando agrego información A en ella, es decir la transición de $P(B|I)$ a $P(B|A, I)$. Existen autores que utilizan este teorema como una pieza central para entender preguntas fundamentales sobre teorías físicas. Por ejemplo, la Dinámica Entrópica [42] que es utilizada para deducir Mecánica Cuántica o incluso para encontrar formas más compactas para una teoría de la probabilidad [43].

2.6. Expectaciones

Cuando se ha definido una probabilidad, ésta puede ser utilizada para realizar estimaciones de cantidades arbitrarias. Para cumplir con esta labor, definiremos el operador expectación, descrito por paréntesis angulares $\langle \cdot \rangle$, como la suma sobre

todas las proposiciones A de una propiedad $f(A)$, pesados por las probabilidades de estas proposiciones, quedando escrito como

$$\langle f(A) \rangle_I = \sum_A P(A|I) f(A). \quad (2.7)$$

Podemos notar que el sub-índice I en la expectación, corresponde a la información dada en la probabilidad bayesiana $P(A|I)$, siendo las expectativas también bayesianas, definidas siempre bajo un contexto. Esto nos indica que, tanto cuando se escriben probabilidades bayesianas como cuando se escriben expectativas bayesianas, debemos indicar en la notación el contexto I bajo el cual se realiza la inferencia.

Por medio de la definición de expectación (2.7), es directo demostrar que este operador tiene la propiedad de ser lineal. Es decir, para α y β constantes, se tiene que

$$\begin{aligned} \langle \alpha f(A) + \beta g(A) \rangle_I &= \sum_A P(A|I) [\alpha f(A) + \beta g(A)]. \\ &= \alpha \sum_A P(A|I) f(A) + \beta \sum_A P(A|I) g(A). \\ &= \alpha \langle f(A) \rangle_I + \beta \langle g(A) \rangle_I. \end{aligned} \quad (2.8)$$

Notemos que en el contexto de Mecánica Estadística, la operación expectación nos permite obtener las llamadas cantidades macroscópicas de los sistemas. Las cantidades macroscópicas se relacionan con los promedios de acuerdo a la llamada ley de los grandes números, la cual establece que

$$\langle F \rangle_I = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_i^N F_i. \quad (2.9)$$

Es decir, la expectación de una cantidad F puede ser obtenida si se promedian $N \rightarrow \infty$ muestras de la cantidad F . Esto permite conectar con el plano experimental la teoría de probabilidad, ya que si bien teóricamente el elemento principal para

obtener los valores de expectación es la probabilidad, al realizar un experimento lo que se obtiene son un cierto número de mediciones de una cantidad, las cuales son promediadas para obtener un valor aproximado de la expectación, donde al realizar un gran número de mediciones se converge al valor exacto de la expectación que se obtendría de forma teórica.

Si bien las propiedades macroscópicas son la expresión que resume el comportamiento de un gran número de partículas en una sola cantidad, esta interpretación fuera de estar correcta o incorrecta, no es necesaria en el contexto de inferencia, ya que podemos tener sólo una partícula la cual tiene incerteza sobre su posición, por lo que este sistema de una partícula tendrá propiedades macroscópicas bien definidas y será indistinguible de un sistema con un gran número de partículas cuando se estudian sus propiedades macroscópicas.

2.7. Normalización y Marginalización

Por medio de la regla de la suma (2.4), es posible deducir una de las propiedades centrales para las probabilidades, y es que las probabilidades están normalizadas en el espacio donde se realiza inferencia [36]. En forma de expectación, esta propiedad queda descrita como

$$\langle 1 \rangle_I = \sum_A P(A|I) = 1. \quad (2.10)$$

A consecuencia de la propiedad de normalización de las probabilidades y de la regla del producto (2.5), se obtiene la regla de la marginalización [35],

$$\begin{aligned} \sum_B P(A, B|I) &= P(A|I) \sum_B P(B|A, I) \\ &= P(A|I). \end{aligned} \quad (2.11)$$

Esta propiedad nos permite reducir el número de variables de las cuales dependen

las probabilidades, por medio de marginalizar las cantidades irrelevantes y dejando sólo las de interés. En Mecánica Estadística, esto sirve para reducir una función de probabilidad dependiente de un vector de 10^{23} posiciones, a sólo una variable, marginalizando $10^{23} - 1$ posiciones. En forma de expectación, esta propiedad queda escrita así:

$$\begin{aligned}
 \langle F(A, B) \rangle_I &= \sum_A \sum_B P(A, B|I) F(A, B) \\
 &= \sum_A P(A|I) \sum_B P(B|A, I) F(A, B) \\
 &= \left\langle \left\langle F(A, B) \right\rangle_{A,I} \right\rangle_I
 \end{aligned} \tag{2.12}$$

Es bueno destacar que en la notación quedan implícitas las variables sobre las cuales se toma la expectación y cuáles corresponden a parámetros (o valores dados), como lo es en $\langle F(A, B) \rangle_{A,I}$, donde se indica que A e I son información conocida, dada en el contexto de la probabilidad, y B es la variable sobre la cual se está realizando la expectación, mediante una función de probabilidad $P(B|A, I)$.

2.8. Funciones delta en probabilidad bayesiana

Una rama de la teoría de la probabilidad, que no involucra más que las leyes de la probabilidad y sus derivados, es el llamado análisis bayesiano, teoría en que se estudian las probabilidades bayesianas y sus consecuencias para la creación de distribuciones de probabilidad [35, 38, 44]. Como se ha mostrado a lo largo de este capítulo, las probabilidades bayesianas siempre están definidas tanto en un contexto general, el cual corresponde a información conocida, pero que no entrega detalles concretos de la proposición a evaluar, como también en un contexto específico, donde la información dada corresponde a igualdades o definiciones concretas. Es así como se puede obtener la forma funcional de algunas probabilidades, entendiendo cómo

se comportan dada información específica, siendo las deltas de Dirac o Kronecker (dependiendo del espacio en que trabajemos) de gran ayuda en esta labor. Por ejemplo, si queremos saber la probabilidad de que A sea igual a B , dado el valor de B , siendo A y B números enteros, entonces la probabilidad corresponderá a una delta tipo Kronecker de la forma

$$P(A = B|B) = \delta_{AB}. \quad (2.13)$$

Análogo al caso anterior, quisiéramos saber la probabilidad de que $F(A)$ sea igual a B , dado la función F y el valor de B , con A , B y $F(A)$ valores reales, entonces tendremos una probabilidad descrita por una delta de Dirac de la forma

$$P(F(A) = B|B, F) = \delta(F(A) - B) \quad (2.14)$$

En resumen, se utiliza delta de Kronecker para variables discretas y la delta de Dirac en el caso de variables continuas, cuando en vez de probabilidad se tiene una densidad de probabilidad.

Capítulo 3

Principio de Máxima Entropía

“...solo hay una distribución que cumple las restricciones y que puede ser obtenida por medio del uso de un procedimiento que satisface axiomas de consistencia; Esta única distribución se obtiene por medio de maximizar la entropía...”

— J. E. Shore and R. W. Johnson.

La idea de maximizar la entropía comenzó con los trabajos realizados por Gibbs [11], quien postuló la forma funcional de la entropía como función de la distribución de probabilidad del sistema, la cual al ser maximizada bajo las condiciones externas que gobiernan al sistema entrega la distribución de probabilidad que describe el sistema.

A partir de esta formulación fue posible crear un método sistemático para modelar fenómenos macroscópicos, idea que fue estudiada en detalle por E.T. Jaynes [12, 13], quien basado en los axiomas de las probabilidades propuestos por R. Cox [34], postuló que la idea de maximizar entropía era no sólo un método sistemático para construir distribuciones de probabilidad, sino que correspondía a una regla general para realizar inferencia.

En este capítulo se entregarán las bases del principio de máxima entropía desde

el análisis de Jaynes¹, mostrando los fundamentos y cómo opera este principio, el cual ayudó a entender de la mecánica estadística y a la creación de modelos mediante inferencia. En la actualidad el principio de Máxima Entropía es ampliamente utilizado existiendo conferencias dedicadas a entender sus aplicaciones a problemas actuales [45, 46].

3.1. ¿Por qué maximizar entropía?

La formulación tradicional de la Mecánica Estadística (ME) [3, 7, 8, 17], basada en el concepto de *ensemble* estadístico y la cual generaliza la teoría original de Boltzmann para diversos tipos de condiciones externas, fue propuesta inicialmente por Gibbs [10, 11]. La mecánica estadística de Gibbs propone que la distribución de probabilidad p_i para los micro-estados² de un sistema, dado un conjunto de N cantidades macroscópicas fijas³ o restricciones de la forma

$$\langle F_j \rangle = \sum_i p_i F_{ij} = f_j, \quad (3.1)$$

donde f_j es un valor constante, es tal que maximiza la entropía de Gibbs

$$S = -k_B \sum_j p_j \ln p_j, \quad (3.2)$$

bajo las restricciones (3.1) y que la probabilidad es normalizable, $\sum_j p_j = 1$. Esta idea de Gibbs entregó uno de los primeros métodos sistemáticos para construir funciones de probabilidad dadas restricciones y fue utilizado con éxito para entender los fundamentos de la ME y termodinámica.

¹En castellano debería ser principio de entropía máxima, pero lo llamaremos de máxima entropía para ser coherente con las siglas en inglés: MaxEnt.

²Combinación de posiciones y momenta que describen un sistema.

³Información conocida del problema.

En 1957 Edwin T. Jaynes, basado en la interpretación de Shannon [47] de la entropía, donde en el contexto de teoría de información [48] se asocia a la información faltante de un modelo, propone que el método de Gibbs es una regla general para construir distribuciones de probabilidad de la manera menos sesgada sujeto a información incompleta, es decir, corresponde a una regla general de inferencia, formalizando lo que hoy se conoce como el Principio de Máxima Entropía abreviado como MaxEnt [12].

En el caso más general, para grados de libertad continuos, la cantidad a extremar es la llamada entropía de Shannon-Jaynes, escrita como

$$S[P(x|I), P(x|I_0)] = - \int dx P(x|I) \ln \frac{P(x|I)}{P(x|I_0)}, \quad (3.3)$$

donde $P(x|I_0)$ corresponde a la probabilidad inicial o *prior*, es decir, la probabilidad asignada previo a incorporar las restricciones y $P(x|I)$ corresponde a la probabilidad posterior, es decir la probabilidad luego de que se incorpora la información de las restricciones I en forma de expectativa.

La unicidad de la entropía de Shannon-Jaynes y su forma ha sido demostrada rigurosamente por varios autores [44, 49, 50]. Estos autores han demostrado que para ser consistentes con las leyes de la probabilidad propuestas por R. Cox [34], es decir, las leyes de la probabilidad bayesiana, la única cantidad que podemos maximizar para encontrar la distribución de probabilidad, es la entropía de Shannon-Jaynes, dejando en claro que cualquier otro tipo de entropía generalizada utilizada para obtener distribuciones de probabilidad no constituye un procedimiento de inferencia consistente.

3.2. Principio de Máxima Entropía

El principio de máxima entropía para n grados de libertad $x : (x_1, \dots, x_n)$ y funciones de probabilidad definidas como $P(x|I) : \mathbb{R}^n \rightarrow \mathbb{R}$, queda establecido como sigue. Dadas un conjunto de N restricciones de la forma

$$\langle f_i(x) \rangle_I = \int dx P(x|I) f_i(x) = F_i, \quad (3.4)$$

donde $f_i(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ y $F_i \in \mathbb{R}$, y con la probabilidad normalizable (2.10) sobre el espacio de los microestados x

$$\langle 1 \rangle_I = \int dx P(x|I) = 1, \quad (3.5)$$

la función de probabilidad menos sesgada (o mas objetiva) que cumple las restricciones (3.4), se obtendrá maximizando la entropía de Shannon-Jaynes (3.3), sujeta a N restricciones (3.4) y a la normalización (3.5).

Para maximizar hacemos uso de los multiplicadores de Lagrange [33]. Así, usando la definición $P(x|I) \equiv \rho(x)$ y $P(x|I_0) \equiv \pi(x)$ para no recargar la notación, queda

$$\frac{\delta}{\delta \rho(y)} \left[S[\rho(x), \pi(x)] - \lambda (\langle 1 \rangle - 1) - \sum_i \beta_i (\langle f_i(x) \rangle - F_i) \right] = 0, \quad (3.6)$$

donde λ y β_i corresponden a multiplicadores de Lagrange. Eliminando todo los valores independientes de $\rho(\)^4$ ya que su derivada es nula, y expresando las expectativas como integrales tal como se muestra en las ecuaciones (3.4) y (3.5),

$$\frac{\delta}{\delta \rho(y)} \left[- \int dx \rho(x) \ln \frac{\rho(x)}{\pi(x)} - \lambda \int dx \rho(x) - \sum_i \beta_i \int dx \rho(x) f_i(x) \right] = 0. \quad (3.7)$$

$$\begin{aligned} & \left[- \int dx \frac{\delta \rho(x)}{\delta \rho(y)} \ln \frac{\rho(x)}{\pi(x)} - \int dx \frac{\delta \rho(x)}{\delta \rho(y)} \right. \\ & \left. - \lambda \int dx \frac{\delta \rho(x)}{\delta \rho(y)} - \sum_i \beta_i \int dx \frac{\delta \rho(x)}{\delta \rho(y)} f_i(x) \right] = 0. \end{aligned} \quad (3.8)$$

⁴ $\rho(\)$ significa la función ρ , de la cual no importa su argumento.

Operando con las reglas usuales del cálculo funcional [51], se obtiene

$$\left[- \int dx \delta(x - y) \ln \frac{\rho(x)}{\pi(x)} - \int dx \delta(x - y) - \lambda \int dx \delta(x - y) - \sum_i \beta_i \int dx \delta(x - y) f_i(x) \right] = 0. \quad (3.9)$$

Resolviendo las integrales utilizando las propiedades de la delta de Dirac (mostradas en el apéndice A.1.1) queda

$$- \ln \frac{\rho(y)}{\pi(y)} - 1 - \lambda - \sum_i \beta_i f_i(y) = 0.$$

Aplicando la función exponencial, se obtiene finalmente la forma de la función de probabilidad,

$$\rho(y) = \pi(y) \exp(-1 - \lambda - \sum_i \beta_i f_i(y)), \quad (3.10)$$

donde sin pérdida de generalidad, la función puede ser expresada dependiente de la variable y o x . Al ser constante los dos primeros términos en la exponencial, estos se agregan como parte de la normalización, definiendo $\exp(-1 - \lambda)$ como $\frac{1}{Z(\beta)}$, la probabilidad queda:

$$\rho(x) = \frac{\pi(x)}{Z(\beta)} \exp(- \sum_i \beta_i f_i(x)), \quad (3.11)$$

donde $\beta = (\beta_1, \beta_2, \dots, \beta_n)$, con $\beta \in \mathbb{R}^n$. $Z(\beta)$ es la llamada *función partición*.

Esta función es independiente de los estados x ya que sólo depende del valor de los N multiplicadores de Lagrange β_i , es decir, depende del vector β . La *función partición* se define utilizando la condición de la normalización (2.10), reemplazando en ella la forma de la probabilidad obtenida al utilizar MaxEnt (3.11)

$$Z(\beta) = \int dx \pi(x) \exp(- \sum_i \beta_i f_i(x)), \quad (3.12)$$

quedando así explícito que $Z(\beta) = Z(\beta_1, \dots, \beta_n)$ es una función que depende sólo de los valores de los multiplicadores de Lagrange β_i . La función partición cumple un rol central relacionando las restricciones con su respectivo multiplicador de Lagrange mediante la ecuación

$$-\frac{\partial}{\partial \beta_i} \ln Z(\beta) = F_i. \quad (3.13)$$

Esta relación es directa de demostrar partiendo del lado izquierdo, reescribiendo la derivada del logaritmo como

$$-\frac{\partial}{\partial \beta_i} \ln Z(\beta) = \frac{1}{Z(\beta)} \frac{\partial}{\partial \beta_i} Z(\beta), \quad (3.14)$$

y luego reemplazando la definición de función partición (3.12) en la parte de la derivada del lado derecho de la relación, es decir,

$$\begin{aligned} -\frac{\partial}{\partial \beta_i} \ln Z(\beta) &= -\frac{1}{Z(\beta)} \frac{\partial}{\partial \beta_i} \int dx \pi(x) \exp\left(-\sum_i \beta_i f_i(x)\right). \\ &= -\frac{1}{Z(\beta)} \int dx \pi(x) \frac{\partial}{\partial \beta_i} \exp\left(-\sum_i \beta_i f_i(x)\right). \\ &= \frac{1}{Z(\beta)} \int dx \pi(x) \exp\left(-\sum_i \beta_i f_i(x)\right) f_i(x), \end{aligned}$$

donde reagrupando los términos que forman la probabilidad obtenida usando MaxEnt (3.11), es posible reescribir el lado derecho como una expectación (3.4) cuyo valor es conocido. Esto es,

$$\begin{aligned} -\frac{\partial}{\partial \beta_i} \ln Z(\beta) &= \int dx \rho(x) f_i(x), \\ &= \langle f_i(x) \rangle_I, \\ &= F_i. \end{aligned} \quad (3.15)$$

Este resultado es una relación general entre los multiplicadores de Lagrange β_i y su respectivo valor de restricción F_i , la cual nos permite resolver para los β_i dado los F_i .

El conjunto de relaciones y la forma que tendrán las distribuciones de probabilidad componen lo que llamaremos un *modelo MaxEnt*, el cual consta de un conjunto de restricciones como se muestra en (3.4), además de una distribución de probabilidad la cual es dependiente de las restricciones, un conjunto de multiplicadores de Lagrange y una función partición, como se muestra en (3.11), donde las restricciones y los multiplicadores de Lagrange quedan relacionados por (3.13). Así en un modelo de MaxEnt todos los elementos constitutivos quedan claramente identificados, lo que es consistente con las probabilidades bayesianas, que requieren conocer el contexto bajo el cual se realizará la inferencia, que en este caso son las restricciones o información dada. Es decir, para el modelo MaxEnt presentado, las probabilidades bayesianas quedan escritas como $P(x|F_1, \dots, F_n)$.

3.3. Ejemplos de construcción de modelos con MaxEnt

A continuación presentaremos tres ejemplos para mostrar cómo trabaja el formalismo de MaxEnt en su forma más general, no necesariamente ligado a problemas físicos.

3.3.1. Modelo sin restricciones: entendiendo el *prior*

Si maximizamos la entropía (3.3) sólo sujeto a la restricción de que la distribución de probabilidad sea normalizable, obtendremos que la distribución de probabilidad $p(x)$ es igual a $\pi(x)$, ya que en este caso no hemos incluido información adicional.

Desde un punto de vista fundamental, el *prior* está conectado con las simetrías del sistema y la métrica del espacio de variables [52], ya que estas propiedades son las únicas de las cuales se puede hacer uso cuando no hay información disponible para construir el modelo.

3.3.2. Modelo para un experimento con muestras aleatorias positivas

Si el promedio de N muestras de una cantidad positiva x es d , utilizando MaxEnt encontrar el modelo de probabilidad que describe las muestras.

El sistema, descrito por una distribución de probabilidad $P(x|I)$, dependiente de las muestras $x \in [0, +\infty)$, se construye utilizando la información I dada como restricción en forma de expectación,

$$\langle x \rangle_I = d, \quad (3.16)$$

con lo que de acuerdo a la forma de las probabilidades utilizando MaxEnt (3.11), obtenemos

$$P(x|I) = \frac{\pi(x)}{Z(\beta)} \exp(-\beta x). \quad (3.17)$$

Si el *prior* que corresponde a la distribución de probabilidad inicial asignada a las muestras lo consideraremos constante, es decir todas las muestras son equiprobables. Absorbiendo el *prior* en la función partición, la probabilidad queda como

$$P(x|I) = \frac{\exp(-\beta x)}{Z(\beta)}. \quad (3.18)$$

Para encontrar el multiplicador de Lagrange, utilizamos la relación entre la restricción y el multiplicador (3.13), resolviendo de forma explícita la integral definida en el dominio de las distancias x para la función partición $Z(\beta)$ (3.12)

$$\begin{aligned} -\frac{\partial}{\partial \beta} \log \left[\int_0^{\infty} \exp(-\beta x) \right] &= d. \\ -\frac{\partial}{\partial \beta} \log \left[\frac{1}{\beta} \right] &= d. \\ \beta &= \frac{1}{d}. \end{aligned} \quad (3.19)$$

Finalmente, reemplazando la forma del multiplicador de Lagrange en la distribución de probabilidad,

$$P(x|I) = \frac{\exp(-x/d)}{Z(d)} = P(x|d). \quad (3.20)$$

Podemos notar que la función de probabilidad obtenida no tiene como distancia más probable la distancia $x = d$, sino que la distancia más probable es $x = 0$, siendo el valor esperado de distancia el valor d , cumpliendo la restricción usada para crear el modelo (3.16).

Esto corresponde a la solución completa de un problema MaxEnt, que involucra tener la forma explícita de la función de probabilidad y los valores de los multiplicadores de Lagrange dadas las restricciones.

3.3.3. Modelo gaussiano a partir de MaxEnt

Una función de probabilidad gaussiana está descrita por dos propiedades: el valor de su centro μ , escrito como

$$\langle x \rangle_I = \mu, \quad (3.21)$$

y su ancho medio, escrito como σ , que corresponde al valor medio (en módulo cuadrado) de cuánto nos alejamos de μ , dado por la restricción

$$\langle (x - \mu)^2 \rangle_I = \sigma^2. \quad (3.22)$$

Dadas estas dos restricciones, la forma de las probabilidades obtenidas utilizando MaxEnt (3.11) queda como,

$$P(x|I) = \frac{\pi(x)}{Z(\beta_1, \beta_2)} \exp\left(-\beta_1 x - \beta_2 (x - \mu)^2\right). \quad (3.23)$$

Para este caso, en lugar de utilizar (3.13) para encontrar el valor de los multiplicadores de Lagrange asociados a (3.21) y (3.22), trabajaremos primero con la

segunda restricción, para esto, desarrollando el cuadrado al interior de la expectativa y utilizando la propiedad de linealidad para la expectativa

$$\langle x^2 \rangle_I - 2\mu \langle x \rangle_I + \mu^2 = \sigma^2, \quad (3.24)$$

donde reemplazando (3.21) en esta relación, se obtiene

$$\langle x^2 \rangle_I = \mu^2 + \sigma^2. \quad (3.25)$$

Entonces, dadas las restricciones (3.21) y (3.25), la forma de las probabilidades obtenidas utilizando MaxEnt (3.11) queda como,

$$P(x|I) = \frac{\pi(x)}{Z(\beta_1, \beta_2)} \exp\left(-\beta_1 x - \beta_2 x^2\right). \quad (3.26)$$

Para obtener los valores de los multiplicadores de Lagrange utilizaremos el teorema de la variable conjugada (C.2) para la distribución de probabilidad

$$\left\langle \frac{\partial v}{\partial x} \right\rangle_I = \left\langle v \frac{\partial}{\partial x} [\beta_1 x + \beta_2 x^2] \right\rangle_I. \quad (3.27)$$

Utilizando la función de prueba $v = 1$

$$0 = \beta_1 + \beta_2 \langle 2x \rangle_I. \quad (3.28)$$

Reemplazando (3.21) obtenemos que

$$\beta_1 = -2\beta_2\mu. \quad (3.29)$$

Luego utilizando la función de prueba $v = x$ en (3.27)

$$1 = \beta_1 \langle x \rangle_I + 2\beta_2 \langle x^2 \rangle_I. \quad (3.30)$$

Reemplazando (3.21) y (3.25) obtenemos que

$$1 = \beta_1\mu + 2\beta_2\mu^2 + 2\beta_2\sigma^2. \quad (3.31)$$

Entonces, reemplazando (3.29) se obtiene el valor del multiplicador β_2

$$\beta_2 = \frac{1}{2\sigma^2}, \quad (3.32)$$

por lo que el valor de β_1 queda como

$$\beta_1 = -\frac{\mu}{\sigma^2}. \quad (3.33)$$

Reemplazando los valores de los multiplicadores de Lagrange (3.32) y (3.33) en la distribución de probabilidad (3.26)

$$P(x|I) = \frac{\pi(x)}{Z(\mu, \sigma)} \exp\left(\frac{-1}{2\sigma^2}[-2\mu x + x^2]\right). \quad (3.34)$$

Completando el cuadrado en la exponencial

$$P(x|I) = \frac{\pi(x)}{Z(\mu, \sigma)} \exp\left(\frac{-1}{2\sigma^2}[x - \mu]^2\right) \exp\left(\frac{-\mu^2}{2\sigma^2}\right). \quad (3.35)$$

Finalmente incluyendo la exponencial de la derecha en la función de partición $Z(\mu, \sigma)$ y utilizando un prior constante el cual también se incluye en la normalización, se obtiene la forma usual de la gaussiana

$$P(x|I) = \frac{1}{Z(\mu, \sigma)} \exp\left(-\frac{1}{2\sigma^2}[x - \mu]^2\right). \quad (3.36)$$

También podemos utilizar (3.12) e integrar para obtener el valor de la función de partición, obteniendo

$$P(x|I) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}[x - \mu]^2\right). \quad (3.37)$$

Capítulo 4

Mecánica Estadística desde el Principio de Máxima Entropía

El uso del Principio de Máxima Entropía (MaxEnt) al ser utilizado para deducir la Mecánica Estadística (ME), entrega un método sistemático para construir la distribución de probabilidad que describe los micro-estados de un sistema, utilizando la información que se conoce de cada sistema (es decir, las condiciones en las que se encuentra) dadas en forma de expectativas.

Existen algunos autores que han explicado cómo entender la mecánica estadística desde la inferencia bayesiana [14,53–55], es decir desde la teoría de probabilidad (cap. 2) y el principio de máxima entropía (cap. 3).

En este capítulo, se deduciran los principales elementos que componen la ME, identificando las restricciones que nos permiten entender cada ensemble de la ME como se indica en la figura 1.1. Se entregará primero un entendimiento de la mecánica estadística utilizando sólo la teoría de la probabilidad (cap. 2), para así justificar el uso en todas las deducciones en este capítulo de la entropía de Shannon-Jaynes (3.3) con un *prior* constante, el cual corresponde a un caso particular de MaxEnt, descubierto por Gibbs en 1908 [1, 3, 6, 8, 11]. Luego se deduciran las densidades de probabilidad que corresponden a los principales *ensembles* que componen la ME.

Recordemos que las restricciones son la información conocida del sistema utilizada para construir el modelo, lo cual en las probabilidades bayesianas corresponde a la información dada. Esto nos permite tener un contexto sobre cuáles son las propiedades del sistema en que se realiza la inferencia, es decir, cuáles son las propiedades macroscópicas definidas en el modelo, lo que como consecuencia permite entender la termodinámica desde la inferencia. Finalmente, se mostrarán algunos componentes de la Termodinámica como la energía libre y algunas relaciones termodinámicas, entendidos desde la inferencia.

4.1. Mecánica Estadística desde la probabilidad bayesiana

La ME puede ser deducida en gran parte utilizando sólo el análisis de las probabilidades bayesianas. Para esto es necesario entender cómo se construye un modelo de probabilidad con información dada, utilizando la notación bayesiana y el teorema de Bayes, tal como se explica en sec. 2.8. A modo de ejemplo, se entrega una breve descripción de cómo entender el ensemble micro-canónico utilizando sólo el análisis bayesiano.

El *ensemble* micro-canónico describe un sistema a energía, volumen y número de partículas fijo gobernado por el hamiltoniano $H(\cdot)$, en el cual cada micro-estado x es equiprobable de acuerdo al postulado fundamental de la ME [8].

Usando la información I dada, $H(x) = E$, se construye la probabilidad de tener

un micro-estado x dado I , es decir $P(x|I)$. Utilizando el teorema de Bayes (2.6)

$$P(x|I) = P(x|H(x) = E, I_0). \quad (4.1)$$

$$= \frac{P(x|I_0) P(H(x) = E|x, I_0)}{P(H(x) = E|I_0)}. \quad (4.2)$$

Entonces, utilizando (2.14) para reescribir la probabilidad $P(H(x) = E|x, I_0)$ en el lado derecho como una delta, obtenemos

$$P(x|I) = \frac{P(x|I_0) \delta(H(x) - E)}{P(H(x) = E|I_0)}, \quad (4.3)$$

donde $P(x|I_0)$ corresponde al *prior*, es decir, la distribución de probabilidad de los micro-estados antes de incorporar la información que el sistema tiene energía constante. Para este capítulo se utilizará un *prior* constante, $p_0 \in \mathbb{R}$ con $p_0 > 0$.

Por otra parte el denominador del lado derecho $P(H(x) = E|I_0)$ cumple el rol de ser una constante de normalización, la cual se obtiene utilizando la regla de la normalización (2.10),

$$1 = \int dx \frac{p_0 \delta(H(x) - E)}{P(H(x) = E|I_0)}.$$

$$P(H(x) = E|I_0) = p_0 \int dx \delta(H(x) - E). \quad (4.4)$$

El valor de la normalización será llamado densidad de estado $\Omega(E)$ y dependerá del valor de energía E y el hamiltoniano $H(\cdot)$ del sistema, por lo que incluyendo p_0 en $\Omega(E)$ se recupera la forma tradicional de la densidad de probabilidad del *ensemble* micro-canónico,

$$P(x|I) = \frac{\delta(H(x) - E)}{\Omega(E)} = P(x|E). \quad (4.5)$$

En la derivación anterior se puede notar la utilidad de tener bien definidos todos los elementos de la inferencia, pero para no sobrecargar la notación y hacer un símil

a la ME tradicional se suele acorta la notación como

$$\rho(x) = P(x|I). \quad (4.6)$$

La notación bayesiana (lado derecho) resulta útil cuando se usan las leyes de la probabilidad (cap. 2), ayudando al entendimiento de cada una de las partes que actúan en la inferencia. Por otro lado, la notación de la mecánica estadística tradicional (lado izquierdo) resulta cómoda cuando se utiliza el calculo diferencial para trabajar con las probabilidades.

Una vez obtenido el modelo de probabilidad para un sistema que se encuentra a energía constante, el cual es conocido en ME como ensemble micro-canónico, es posible continuar derivando los demás ensembles de la ME. Para esto debemos identificar cómo cambian las probabilidades del sistema al agregar información I . Para esta labor las herramientas más utilizadas son el teorema de Bayes y la regla de la marginalización. Si el lector busca profundizar en métodos para crear distribuciones de probabilidad sólo utilizando el análisis bayesiano se recomienda la siguiente referencia [35].

Para ponernos en contexto respecto a las unidades utilizadas en los próximos desarrollos, es bueno destacar que la entropía de Shannon-Jaynes (3.3) no tiene unidades, por lo que se debe agregar una constante para poder relacionar las diferentes cantidades, conocida como la constante de Boltzmann k_B , la cual nos permite relacionar entropía con energía y cuyo valor fija la unidad de temperatura.

4.2. Densidades de probabilidades de diferentes *ensembles* de la ME desde MaxEnt

Como se mostró en la sección anterior, existe un camino para derivar las diferentes distribuciones de probabilidad dada información incompleta. El punto de tope cuando uno comienza a trabajar en inferencia sólo con las leyes de la probabilidad bayesiana, es que se requiere definir una forma funcional para las probabilidades que serán actualizadas con información utilizando el teorema de Bayes, por lo que poder formalizar la teoría mediante un método sistemático, que nos permita obtener la forma de las probabilidades dadas las condiciones del sistema, se vuelve de gran importancia en la labor de entender la ME desde la inferencia.

En la siguiente sección se mostrará cómo deducir la densidad de probabilidad de los principales ensembles de la ME utilizando MaxEnt, dejando en claro lo sistemático que se vuelve este proceso.

Para todas las próximas deducciones en este capítulo utilizaremos el *prior* con un valor constante, justificando esto con el postulado fundamental de la ME (que todos los estados en un comienzo son equiprobables). Donde luego se crean las diferentes tendencias a medida se incorpora información con el teorema de Bayes. Por otra parte, visto desde la geometría de información [56], suponer que el *prior* es constante es lo menos sesgado en términos de simetrías dado que el prior se define como la probabilidad cuando no sabemos nada del sistema por lo que suponer tendencias correspondería a suponer tendencias en la geometría del sistema.

Como hemos subrayado, a partir de MaxEnt es posible deducir la distribución de probabilidad de los micro-estados dadas las condiciones de cada sistema. Estas condiciones, cuando son entendidas desde MaxEnt, son en general¹ dadas en for-

¹MaxEnt no exige necesariamente información en forma de expectativas. Esta información

mas de expectación (3.4) y una vez que son identificadas todas las restricciones del sistema se obtiene un modelo completo para trabajar y realizar inferencia sobre el sistema, permitiendo calcular nuevas propiedades macroscópicas por medio del uso de la operación expectación (2.7).

4.2.1. Ensemble Micro-Canónico

De acuerdo con MaxEnt, para construir el modelo de probabilidad que mejor describe al sistema, se requiere identificar las restricciones del mismo. En el caso del ensemble micro-canónico de la ME, se caracteriza por describir un sistema que se encuentra completamente aislado, es decir, no intercambia energía ni partículas con el entorno y se mantiene a un volumen constante. Esto se traduce en la condición de que en el sistema cada uno de sus micro-estados permitidos x cumplen que $H(x) = E$, una cantidad constante.

La condición $H(x) = E$ no es posible de usar directamente en un marco de trabajo MaxEnt como el desarrollado en el capítulo 3, siendo necesario tener restricciones en forma de expectación (3.4). Una forma de representar la condición de tener sólo micro-estados de energía constante como expectativas, requiere entender la idea de fluctuaciones, como se muestra en el apéndice A.2. Con esto, un sistema de energía constante es equivalente a considerar un sistema con energía promedio E y sin incerteza en su energía. Es decir, el sistema sólo permite micro-estados de energía E , quedando expresadas las restricciones como

$$\langle H(x) \rangle_I = E, \quad (4.7)$$

$$\langle H(x)^2 \rangle_I = E^2, \quad (4.8)$$

pueden ser expectativas, cuocientes de expectativas o en general cualquier condición sobre $P(x)$. El uso más común es restringir expectativas pero no el único.

donde es fácil notar que, usando la definición de incerteza (A.6) y las restricciones anteriores, la incerteza de energía es nula

$$\begin{aligned} (\Delta E)^2 &= \left\langle H(x)^2 \right\rangle_I - E^2. \\ &= 0. \end{aligned} \quad (4.9)$$

Entonces, acuerdo al marco de trabajo MaxEnt, dadas las restricciones (4.7) y (4.8), la distribución de probabilidad (3.11) será de la forma

$$P(x|I) = \frac{\exp\left(-\beta_1 H(x) - \beta_2 H(x)^2\right)}{Z(\beta_1, \beta_2)}. \quad (4.10)$$

Completando el cuadrados y luego agrupando los términos se obtiene

$$P(x|I) = \frac{\exp\left(-\beta_2\left(\frac{\beta_1}{-2\beta_2} - H(x)\right)^2\right)}{Z(\beta_1, \beta_2)} \exp\left(\frac{\beta_1^2}{4\beta_2}\right), \quad (4.11)$$

siendo posible expresar el lado derecho de (4.11) como un cuadrado perfecto más una constante, formando así una la gaussiana

$$P(x|I) = \frac{\exp\left(-\frac{(H(x)-E)^2}{2(\Delta E)^2}\right)}{Z(E)\sqrt{(2\pi)(\Delta E)^2}}, \quad (4.12)$$

$$= \frac{\exp\left(-\frac{(H(x)-E)^2}{2(\Delta E)^2}\right)}{Z(E, \Delta E)}. \quad (4.13)$$

En el límite en que la incerteza de energía es cero ($\Delta E \rightarrow 0$), la probabilidad tiende a la delta de Dirac,

$$P(x|I) = \frac{\delta(H(x) - E)}{Z(E)}. \quad (4.14)$$

Donde $Z(E)$ se identifica con $\Omega(E)$, obteniendo la forma de la densidad de probabilidad del ensemble micro-canónico de la ME desde MaxEnt. De los ensembles de la ME obtenidos mediante el uso de MaxEnt, el micro-canónico es el que se obtiene

de forma “menos natural”. Esto ocurre pues en este *ensemble* la restricción no viene dada directamente en forma de expectativas, por lo que se requiere un tratamiento adicional que dificulta el uso del procedimiento. Para el caso micro-canónico, resulta entonces más natural encontrar la distribución de probabilidad sólo usando las probabilidades bayesianas como se mostró en la sección anterior.

4.2.2. *Ensemble* Canónico

El caso del *ensemble* canónico está descrito por un sistema que se encuentra en contacto con un baño térmico a temperatura T ; pudiendo intercambiar energía con su entorno, pero en promedio tiene una energía bien definida.

Desde el punto de vista de la inferencia, el modelo de probabilidad para el ensemble canónico se obtiene por medio de la restricción que el sistema tiene una energía esperada E , dado en forma de expectativa como

$$\langle H(x) \rangle_I = E. \quad (4.15)$$

De acuerdo a MaxEnt, dada la restricción (4.15), la distribución de probabilidad (3.11) será de la forma

$$P(x|I) = \frac{\exp(-\beta H(x))}{Z(\beta)}, \quad (4.16)$$

siendo posible dejar explícito en la notación la información dada, expresando ahora la probabilidad como

$$P(x|I) = P(x|\beta), \quad (4.17)$$

donde notamos que se construye de forma directa la forma de la función de probabilidad correspondiente al ensemble canónico de la ME. Debido a que es posible identificar la restricción que construye el modelo, MaxEnt nos entrega de forma clara

el contexto para realizar la inferencia, lo que en este caso tiene como información el hecho que el sistema tiene una energía promedio definida.

Para este ensemble podemos ver que existe una relación entre la restricción y el multiplicador de Lagrange dada por (3.13)

$$-\frac{\partial \log Z(\beta)}{\partial \beta} = E, \quad (4.18)$$

donde el multiplicador de Lagrange β juega el rol de ser el inverso de la temperatura [6, 53],

$$\beta = \frac{1}{k_B T}. \quad (4.19)$$

donde k_B es la constante de Boltzmann.

4.2.3. *Ensemble Gran-Canónico*

El *ensemble* Gran-Canónico está descrito por un sistema en contacto con un baño térmico a temperatura T y un baño de partículas con potencial químico μ . Estas dos condiciones caracterizan al sistema que intercambia energía y partículas con su entorno pero que en promedio tienen una energía y número de partículas bien definidos.

Desde la inferencia, un *ensemble* Gran-Canónico está dado primero por la misma restricción del ensemble canónico, es decir hay un valor esperado de energía, junto con la restricción de que existe un valor esperado para el número de partículas.

$$\langle H(x) \rangle_I = E, \quad (4.20)$$

$$\langle n \rangle_I = N, \quad (4.21)$$

por lo que de acuerdo a MaxEnt, dada las restricciones (4.20) y (4.21), la función de probabilidad (3.11) será de la forma

$$P(x, n|I) = \frac{\exp(-\beta H(x) - \beta \mu n)}{Z(\beta, \mu)}, \quad (4.22)$$

Con multiplicadores de Lagrange β y $\beta\mu$ respectivamente. Para este ensemble podemos ver que existe una relación entre las restricciones y los multiplicadores de Lagrange dadas por (3.13)

$$-\frac{\partial \log Z(\beta, \mu)}{\partial \beta} = E. \quad (4.23)$$

$$\frac{-1}{\beta} \frac{\partial \log Z(\beta, \mu)}{\partial \mu} = N. \quad (4.24)$$

Donde el multiplicador de Lagrange $\beta = 1/k_B T$ juega el rol de ser el inverso de la temperatura manteniendo la energía esperada en un valor constante y μ juega el rol del potencial químico manteniendo el valor esperado del número de partículas constante.

4.2.4. *Ensemble* Isotérmico Isobárico

El *ensemble* Isotérmico Isobárico, está descrito por un sistema que se encuentra en contacto con un baño térmico a temperatura T y un baño o reservorio de presión a presión p , teniendo promedios de energía y volumen bien definidos (como su nombre lo indica) por un sistema que se encuentra a temperatura y presión constante.

Desde la inferencia, el ensemble está restringido primero por la misma restricción del ensemble canónico, es decir hay un valor esperado de energía lo que hace que permanezca a temperatura constante como multiplicador de Lagrange, junto con la restricción de que existe un valor esperado para el volumen del sistema, lo que hace que el sistema permanezca a presión constante siendo éste su multiplicador de Lagrange. Esto es,

$$\langle H(x) \rangle_I = E. \quad (4.25)$$

$$\langle v \rangle_I = V. \quad (4.26)$$

Entonces, utilizando MaxEnt dada las restricciones (4.25) y (4.26), la función de probabilidad (3.11) será de la forma

$$P(x, v|I) = \frac{\exp(-\beta H(x) - \beta p v)}{Z(\beta, p)}, \quad (4.27)$$

con multiplicadores de Lagrange β y βp , respectivamente.

Para este ensemble podemos ver que existe una relación entre las restricciones y los multiplicadores de Lagrange dadas por (3.13)

$$-\frac{\partial \log Z(\beta, p)}{\partial \beta} = E, \quad (4.28)$$

$$\frac{-1}{\beta} \frac{\partial \log Z(\beta, p)}{\partial p} = V, \quad (4.29)$$

donde el multiplicador de Lagrange $\beta = 1/k_B T$ juega el rol de ser el inverso de la temperatura manteniendo la energía esperada en un valor constante y p juega el rol de la presión manteniendo el valor esperado del volumen constante.

Es bueno destacar que desde la inferencia no existe distinción formal entre el *ensemble* Gran-Canónico y el Ensemble Isotérmico Isobárico, siendo el tratamiento matemático idéntico para ambos, es así como la distinción aparece cuando le damos un sentido físico a las propiedades restringidas. Sin embargo hay que tener en cuenta que n y v son en estos ejemplos grados de libertad adicionales a los (x_1, \dots, x_n) , y no funciones de éstos como en la sección 3.2.

Se deja como ejercicio para el lector el construir un ensemble en el cual se tiene equilibrio térmico, químico y de presión, para lo que se obtendrá una distribución de probabilidad dada por

$$P(x, n, v|I) = \frac{\exp(-\beta H(x) - \beta \mu n - \beta p v)}{Z(\beta, \mu, p)} \quad (4.30)$$

4.2.5. Estadística de Bose-Einstein y Fermi-Dirac

Las estadísticas de Fermi-Dirac y Bose-Einstein son modelos de probabilidad para sistemas con grados de libertad discretos (típicamente relacionado con modelos de mecánica cuántica). Para estos sistemas se tiene la probabilidad de encontrar sus partículas en un cierto estado \vec{n} , donde cada una de sus componentes corresponde al número de partículas n_i contenidas en el nivel i -ésimo del sistema, por lo que cada sistema tiene un total de $\sum_i n_i$ partículas. De la misma forma, estos sistemas tienen energías discretas ϵ_i para cada uno de sus niveles i , siendo la energía total del sistema $\sum_i n_i \epsilon_i$.

Una analogía para entender esto es pensar en un edificio de L pisos (niveles), donde en cada uno de sus pisos (niveles) i , existe un cierto número de habitantes (partículas) dado por n_i , por lo que el estado del sistema (edificio) es un vector \vec{n} que contiene toda la información respecto a la ocupación del edificio, es decir $\vec{n} : (n_0, \dots, n_L)$, entonces, la cantidad total de habitantes (partículas) en el edificio es $\sum_{i=0}^L n_i$. Siguiendo con la analogía, si consideramos que cada nivel del edificio tiene un precio de ϵ_i , tendremos que el total que pagan los n_i habitantes del nivel i es de $n_i \epsilon_i$, por lo que el valor total que pagan todos los habitantes en el edificio será $\sum_i n_i \epsilon_i$. Es así como la estadística de Bose-Einstein y Fermi-Dirac, corresponde al modelo de probabilidad de la ocupación de un edificio $P(\vec{n}|I)$ en el cual conocemos que el número total de habitantes en el edificio es el valor esperado N y el monto total que pagan los habitantes es el valor esperado E .

Entonces, dada la información de que conocemos el valor esperado del número

de partículas del sistema, la primera restricción puede ser escrita como

$$N_0 = \langle N(\vec{n}) \rangle_I \quad (4.31)$$

$$= \langle \sum_i n_i \rangle_I \quad (4.32)$$

$$= \sum_i \langle n_i \rangle_I \quad (4.33)$$

$$= \sum_i \sum_{\vec{n}} P(\vec{n}|I) n_i, \quad (4.34)$$

con n_i un número entero, correspondiente al número de partículas en el nivel i -ésimo. Mientras que el valor esperado de la energía total del sistema corresponde a la segunda restricción, que puede ser escrita como

$$E_0 = \langle E(\vec{n}) \rangle_I \quad (4.35)$$

$$= \langle \sum_i n_i \epsilon_i \rangle_I \quad (4.36)$$

$$= \sum_i \epsilon_i \langle n_i \rangle_I \quad (4.37)$$

$$= \sum_i \epsilon_i \sum_{\vec{n}} P(\vec{n}|I) n_i, \quad (4.38)$$

con ϵ_i la energía asociada a cada partícula del nivel i -ésimo.

Utilizando MaxEnt, se obtiene un sistema con condiciones de equilibrio termodinámico dado por temperatura y potencial químico constantes. Esto es similar al *ensemble* Gran-Canónico, solo que para este caso el número de partículas no es un grado de libertad adicional ya que esta cantidad puede ser descrita por una función dependiente del micro-estado \vec{n} que llamaremos $N(\vec{n})$, obteniendo una probabilidad de la forma

$$P(\vec{n}|I) = \frac{\exp\left(-\beta E(\vec{n}) + \beta \mu N(\vec{n})\right)}{Z(\beta, \mu)} \quad (4.39)$$

$$= \frac{\exp\left(-\beta \sum_i \epsilon_i n_i + \beta \mu \sum_i n_i\right)}{Z(\beta, \mu)}, \quad (4.40)$$

probabilidad que reescrita de forma compacta queda como

$$P(\vec{n}|I) = \frac{1}{Z(\beta, \mu)} \prod_{i=0} \exp \left(-\beta (\epsilon_i + \mu) \right)^{n_i}. \quad (4.41)$$

Como en estos sistemas se conoce la función que nos permite obtener la energía y el número de partículas es posible obtener el valor de su función de partición², para esto se reescribe la función partición (3.12) en L funciones de partición de la forma

$$Z(\beta, \mu) = \sum_{n_i} \prod_{i=0} \exp \left(-\beta (\epsilon_i - \mu) \right)^{n_i}. \quad (4.42)$$

$$= \prod_{i=0}^L Z_i(\beta, \mu). \quad (4.43)$$

donde,

$$Z_i(\beta, \mu) = \sum_{n_i=0} \exp \left(-\beta (\epsilon_i - \mu) \right)^{n_i}. \quad (4.44)$$

Las funciones de partición obtenidas permiten obtener la relación entre las restricciones y los multiplicadores de Lagrange dadas por (3.13), de forma análoga a lo que ocurre en el ensemble Gran-Canónico. Para el caso de la restricción relacionada al número de partículas del sistema (4.31) se obtiene

$$N_0 = \frac{-1}{\beta} \frac{\partial \log Z(\beta, \mu)}{\partial \mu} \quad (4.45)$$

$$= \frac{-1}{\beta} \frac{\partial}{\partial \mu} \log \prod_{i=0}^L Z_i(\beta, \mu) \quad (4.46)$$

$$= \frac{-1}{\beta} \sum_{i=0}^L \frac{\partial}{\partial \mu} \log Z_i(\beta, \mu), \quad (4.47)$$

²Los ensembles mostrados anteriormente dependen de un Hamiltoniano $H(x)$ donde para diferentes formas funcionales del Hamiltoniano obtendremos diferentes funciones de partición, en este caso no existe tal misterio, ya que las energías son valores independientes del número de ocupación n_i . De hecho, la única información relevante acerca de H es la densidad de estados $\Omega(E)$, y ésta ya está codificada en el vector $(\epsilon_1, \dots, \epsilon_L)$

mientras que para el caso de la restricción relacionada a la energía (4.35) se obtiene

$$E = -\frac{\partial \log Z(\beta, \mu)}{\partial \beta} \quad (4.48)$$

$$= -\frac{\partial}{\partial \beta} \log \prod_{i=0}^L Z_i(\beta, \mu) \quad (4.49)$$

$$= -\sum_{i=0}^L \frac{\partial}{\partial \beta} \log Z_i(\beta, \mu) \quad (4.50)$$

$$= -\sum_{i=0}^L \frac{\partial}{\partial \beta} \log Z_i(\beta, \mu). \quad (4.51)$$

En este punto es que se hace la distinción entre la estadística de Bosé-Einstein y la estadística de Fermi-Dirac, siendo la única diferencia el considerar que las partículas del sistema son bosones, de los cuales es posible tener hasta $N \rightarrow \infty$ en cada nivel i , o fermiones, de los cuales sólo se puede tener 1 como máximo en cada nivel i , ya que no pueden existir 2 fermiones en el mismo nivel.

Bosones

Si las partículas del sistema son bosones, se tiene que $n_i \in [0, \infty]$ ya que para en cada nivel de energía no hay un número límite de ocupación, por lo que su función de partición i -ésima (4.44) será entonces

$$Z_i^B(\beta, \mu) = \sum_{n_i=0}^{\infty} \exp\left(-\beta(\epsilon_i - \mu)\right)^{n_i} \quad (4.52)$$

$$= \frac{1}{1 - \exp\left(-\beta(\epsilon_i - \mu)\right)}. \quad (4.53)$$

Entonces, para bosones la relación entre el potencial químico y el número de

partículas (4.47) queda de la forma

$$N_0 = \frac{-1}{\beta} \sum_{i=0}^L \frac{-\beta}{\exp(\beta(\epsilon_i - \mu)) - 1} \quad (4.54)$$

$$= \sum_{i=0}^L \frac{1}{\exp(\beta(\epsilon_i - \mu)) - 1}. \quad (4.55)$$

Vista de forma tradicional en ME cuando se asocia este resultado a (4.33)

$$\langle n_i \rangle_I = \frac{1}{\exp(\beta(\epsilon_i - \mu)) - 1}. \quad (4.56)$$

Por otra parte la relación entre temperatura ($\beta = \frac{1}{k_B T}$) y energía (4.51) queda de la forma

$$E = \sum_{i=0}^L \frac{\epsilon_i - \mu}{\exp(\beta(\epsilon_i - \mu)) - 1}. \quad (4.57)$$

Fermiones

Para el caso de fermiones $n_i \in [0, 1]$, ya que nunca puede haber 2 de ellos en cada nivel i , entonces la función de partición i -ésima (4.44) será entonces

$$Z_i^F(\beta, \mu) = \sum_{n_i=0}^1 \exp(-\beta(\epsilon_i - \mu))^{n_i}. \quad (4.58)$$

$$= 1 + \exp(-\beta(\epsilon_i - \mu)). \quad (4.59)$$

Entonces, para Fermiones la relación entre el potencial químico y el número de partículas (4.47) queda de la forma

$$N_0 = \frac{-1}{\beta} \sum_{i=0}^L \frac{-\beta}{\exp(\beta(\epsilon_i - \mu)) + 1}. \quad (4.60)$$

$$= \sum_{i=0}^L \frac{1}{\exp(\beta(\epsilon_i - \mu)) + 1}. \quad (4.61)$$

Vista de forma tradicional en ME cuando se asocia este resultado a (4.33)

$$\langle n_i \rangle_I = \frac{1}{\exp\left(\beta(\epsilon_i - \mu)\right) + 1}. \quad (4.62)$$

Por otra parte la relación entre temperatura ($\beta = \frac{1}{k_B T}$) y energía (4.51) queda de la forma

$$E = \sum_{i=0}^L \frac{\epsilon_i - \mu}{\exp\left(\beta(\epsilon_i - \mu)\right) + 1}. \quad (4.63)$$

4.3. Termodinámica desde la inferencia

Como se indicó en el capítulo 3, el Principio de Máxima Entropía es una regla general para realizar inferencia, la cual trabaja bajo la premisa de crear las funciones de probabilidad menos sesgadas para describir un sistema dada información incompleta. De acuerdo a esto, cuando buscamos entender termodinámica desde el contexto de MaxEnt, es importante tener claro qué elementos son conocidos y cuáles no.

A continuación se entregará una breve explicación de los elementos centrales de una teoría termodinámica, dejando en claro qué cantidades son definiciones y cuáles son obtenidas o emergen del formalismo. Por supuesto, de forma completamente análoga al tratamiento tradicional, una vez obtenida la densidad de probabilidad de cada ensemble, se puede obtener la ecuación fundamental o energía libre respectivas.

4.3.1. Energía Libre

Desde la inferencia, donde en rigor no podemos asumir la existencia de cantidades, a menos que sean definidas en el contexto I . Sabemos por ME [8] que la energía libre

D se relaciona directamente con la función partición Z (3.12) del sistema de la forma

$$\begin{aligned} Z &= \exp(-\beta D) \\ D &= -\beta^{-1} \log Z, \end{aligned} \quad (4.64)$$

con $\beta = 1/k_B T$.

Usando esto, cada *ensemble* tendrá asociado una expresión para la energía libre, que se obtiene al reemplazar la forma de la probabilidad para cada ensemble (3.11), en la entropía de Shannon-Jaynes (3.3), con un *prior* constante, como veremos a continuación.

La energía libre es, al igual que la entropía y las cantidades conjugadas (multiplicadores de Lagrange), una propiedad del modelo probabilístico, distinta de las funciones de fase $A(x, p)$.

Es bueno destacar que para el caso micro-canónico no se requiere definir una nueva energía libre D , ya que todas las cantidades en este sistema son conocidas y el realizar definiciones como la anterior mostrada llevaría a que la energía libre es igual a la entropía S .

Usando la definición de energía libre (4.64), para los diferentes ensembles, se obtienen las siguientes relaciones, las cuales son un punto de partida para entender las ecuaciones de estado para derivar la termodinámica.

Ensemble Canónico

Reemplazando (4.16) en la entropía de Shannon-Jaynes (3.3), con un *prior* constante, obtenemos

$$\frac{S}{k_B} = \beta E - \beta D, \quad (4.65)$$

donde en este caso, D corresponde a la energía libre de Helmholtz, escrita generalmente con la letra F , es decir, $D = F$. Entonces, despejando la energía libre, se llega

a

$$F = E - T S. \quad (4.66)$$

Ensemble Gran Canónico

Reemplazando (4.22) en la entropía de Shannon-Jaynes (3.3), con un *prior* constante, obtenemos

$$\frac{S}{k_B} = \beta E + \beta \mu N - \beta D, \quad (4.67)$$

donde en este caso, D corresponde a la energía libre de Gibbs, escrita generalmente con la letra G , es decir, $D = G$. Entonces, despejando la energía libre se obtiene

$$G = E + \mu N - T S. \quad (4.68)$$

Ensemble Isotérmico-Isobárico

Reemplazando (4.27) en la entropía de Shannon-Jaynes (3.3), con un *prior* constante, obtenemos

$$\frac{S}{k_B} = \beta E + \beta p V - \beta D, \quad (4.69)$$

donde en este caso, D corresponde a la energía libre de Gibbs, escrita generalmente con la letra G , es decir, $D = G$. Entonces, despejando la energía libre, se llega a

$$G = E + p V - T S. \quad (4.70)$$

Generalmente, se redefine la cantidad $H = E + p V$, conocida como entalpía, por lo que la energía libre de Gibbs es

$$G = H - T S. \quad (4.71)$$

Capítulo 5

Principio de Máximo Calibre

“...Parece intuitivamente razonable que el principio variacional de Gibbs que determina las condiciones del equilibrio heterogéneo puede generalizarse a condiciones de no equilibrio...”

The minimum entropy production principle – E. T. Jaynes.

5.1. Máximo Calibre y la Mecánica Estadística Fuera del Equilibrio

El Principio de Máximo Calibre (MaxCal) fue propuesto por E. T. Jaynes en 1980 [28], con el objetivo de crear un principio análogo al de máxima entropía pero para el estudio de sistemas fuera del equilibrio. Años antes, Jaynes demostró que el método propuesto por Gibbs [11] para construir la función de probabilidad dadas las restricciones del sistema, corresponde a una regla general para realizar inferencia, conocida como el Principio de Máxima Entropía (MaxEnt) [12, 13].

MaxEnt es un procedimiento igual al presentado por Gibbs sólo que desde el contexto de inferencia sirvió muy bien para entender la Mecánica Estadística, como se mostró en el capítulo anterior describiendo resultados ya conocidos de sistemas en equilibrio.

Entendiendo que MaxEnt corresponde a una regla general para realizar inferencia, Jaynes propuso utilizar MaxEnt en el espacio de trayectorias para entender sistemas fuera del equilibrio, cambiando los micro-estados correspondientes a descripciones microscópicas instantáneas de los sistemas por micro-trayectorias, las que representan el camino completo seguido por cada micro-estado en un intervalo de tiempo. Así emerge la idea de probabilidades para micro-trayectorias o probabilidad funcional, obtenida al maximizar una entropía definida en el espacio de las micro-trayectorias que llamó *Calibre*¹ al relacionar esta cantidad con el “diámetro” del cilindro que envuelve al conjunto de todas las micro-trayectorias de sistema, como lo es un recipiente cilíndrico lleno de spaguetti en su interior.

MaxCal es el principio de máxima entropía definido sobre el espacio de funciones \tilde{x} , las cuales serán llamadas trayectorias o caminos, en lugar de estados x . En este capítulo comenzaremos formalizando la notación de funcionales utilizada para trabajar con MaxCal, para luego mostrar cómo por medio de restringir funcionales y maximizando el calibre (o entropía de trayectorias), se obtiene que la forma funcional de las probabilidades $P[\tilde{x}|I]$ es de la familia exponencial, tal como lo son las probabilidades obtenida en los problemas de máxima entropía tradicional. Finalmente mediante ejemplos se introducirá al uso de este principio en diferentes contextos de la física para entender sistemas dependientes del tiempo.

5.2. Notación de Funcionales

Un funcional es una función que toma como argumento funciones. Estos funcionales se escriben $F[\tilde{x}]$, donde F es el funcional que toma una función \tilde{x} del espacio de funciones \mathbb{X} y entrega un valor real contenido en \mathbb{R} .

¹nombre que recibe el diámetro del cañón de un arma de fuego.

En esta tesis los principales funcionales a trabajar corresponden a cantidades reales como la acción, el trabajo, la energía acumulada y la probabilidad.

Las funciones denotadas por \tilde{x} serán llamadas trayectorias o caminos. Las trayectorias son parte del espacio de funciones $\tilde{x} \in \mathbb{X}$, el cual llamaremos de forma coloquial espacio de caminos o espacio de trayectorias.

Se utilizarán los corchetes “[]” para indicar que se está hablando de un funcional, como por ejemplo, la acción $A[\tilde{x}]$ [57], el trabajo $W[\tilde{x}]$ [57], entre otros.

Se utilizará \tilde{x} para indicar que se está hablando de trayectorias, caminos o funciones, análogo a la trayectoria seguida por una partícula \tilde{x} . Las trayectorias en general serán dependientes del tiempo, de tal forma que al evaluar una trayectoria en un tiempo determinado t , ésta toma un valor x correspondiente a un micro-estado $\tilde{x}_t = x$, donde $x \in \mathbb{R}^n$, siendo x un punto en el espacio de n dimensiones.

Se utilizará la notación \tilde{x} en lugar de $x(\)$, y \tilde{x}_t en lugar de $x(t)$, para distinguir a las funciones de los valores numéricos y reducir la notación, ya que las trayectorias son los elementos más utilizados para las derivaciones de esta tesis, y se hace necesario dejar explícitas las variables de las cuales dependen.

Es necesario destacar que el funcional F , no depende de las variables de \tilde{x} , es por esto que si \tilde{x} es una función de t , el funcional “no tiene forma de saber” cuáles son las variables implícitas de las funciones que depende, por lo que se sigue que

$$\frac{dF[\tilde{x}]}{dt} = 0. \quad (5.1)$$

5.2.1. Derivada Funcional: Valores extremos de un funcional

Un funcional puede no depender únicamente de funciones, sino que estos pueden ser dependientes además de parámetros θ , en cuyo caso escribiremos el funcional de la forma $F[\tilde{x}; \theta]$. En este caso es posible definir su derivada por medio del diferencial de

una función de varias variables. Es decir, dado $F : F[\tilde{x}, \theta]$, escribiremos su diferencial como

$$dF = \int dt \left(\frac{\delta F}{\delta \tilde{x}_t} \right) \delta \tilde{x}_t + \partial_\theta F \delta \theta. \quad (5.2)$$

Uno de los primeros elementos que no hemos introducido y que aparece en esta derivación es la *derivada funcional* $\frac{\delta}{\delta \tilde{x}_t}$, que representa la derivada de un funcional con respecto a una función de la cual depende, evaluada en un tiempo t . En esta tesis, la derivada funcional se escribirá de la forma

$$\frac{\delta A[\tilde{x}]}{\delta \tilde{x}_t}, \quad (5.3)$$

esto quiere decir que se está derivando el funcional A con respecto a su variable \tilde{x} en un tiempo t .

Es posible entender la derivada funcional a partir del cálculo infinitesimal [30, 51, 58], donde haciendo la consideración que una trayectoria puede ser entendida como un vector de infinitas componentes $\tilde{x} = (\tilde{x}_{t_0}, \tilde{x}_{t_1}, \dots, \tilde{x}_{t_N})$ el cual al tender al límite $N \rightarrow 0$ (en que $\Delta t = t_{k+1} - t_k \rightarrow 0$) nos permite definir la derivada funcional como el límite

$$\frac{1}{\Delta t} \frac{\partial}{\partial \tilde{x}_{t_k}} \rightarrow \frac{\delta}{\delta \tilde{x}_t}.$$

De esta forma, la derivada funcional vista como una generalización del gradiente de una función vectorial, tiene las mismas propiedades que la derivada usual (Reglas de Suma, Producto, Cadena, entre otras.), sólo que definidas en el contexto de funcionales [51].

En esta tesis la derivada funcional se define de acuerdo a las siguientes reglas,

$$\frac{\delta \tilde{x}_t}{\delta \tilde{x}_\tau} = \delta(t - \tau). \quad (5.4)$$

$$\frac{\delta \tilde{x}_t}{\delta \tilde{x}_t} = 1. \quad (5.5)$$

$$\frac{\delta^{(n)} \tilde{x}_t}{\delta \tilde{x}_\tau} = \delta^{(n)}(t - \tau). \quad (5.6)$$

$$\frac{\delta^{(n)} \tilde{x}_t}{\delta \tilde{x}_t} = 0. \quad (5.7)$$

donde $\tilde{x}_t^{(n)} = \frac{d^n}{dt^n}$ corresponde a la n -ésima derivada de una trayectoria respecto al tiempo (o en general con respecto a la variable de la cual depende).

La derivada funcional puede ser utilizada para encontrar los valores extremos de un funcional, buscando la trayectoria \tilde{x} para la cual el valor de un funcional $A[\tilde{x}]$ es un extremo (máximo, mínimo o punto de ensilladura). Uno de los usos más conocidos de este formalismo es en el contexto del principio de mínima acción para sistemas lagrangianos [33, 57], el cual establece que la trayectoria \tilde{x} seguida por una partícula es la que minimiza la acción A , de tal forma que \tilde{x} es un extremo de A

$$\frac{\delta A[\tilde{x}]}{\delta \tilde{x}_t} = 0. \quad (5.8)$$

donde la acción clásica A es un funcional dependiente de la trayectoria \tilde{x} , a través de una función \mathcal{L} llamada lagrangiano

$$A[\tilde{x}] = \int_{t_0}^{t_f} dt \mathcal{L}(t, x, \dot{x}, \dots, x^{(n)}) \Big|_{x = \frac{d^k}{dt^k} \tilde{x}_t}, \quad (5.9)$$

con $k \in [0, 1, \dots, n]$. Donde el lagrangiano es una función dependiente de los valores instantáneos de posición \tilde{x}_t y sus derivadas temporales $\frac{d^k}{dt^k} \tilde{x}_t$ para una trayectoria \tilde{x} en un tiempo t .

Si aplicamos la derivada funcional a la acción (5.9), tal como indica el principio de mínima acción (5.8), tendremos que entonces derivar el lagrangiano, para esto, lo derivamos como una función de varias variables, en este caso una función de $n + 1$ variables $\mathcal{L} : \mathcal{L}(\tau, x, \dot{x}, \dots, \overset{(n)}{x})$

$$\begin{aligned} 0 &= \frac{\delta A[\tilde{x}]}{\delta \tilde{x}_t} \\ &= \int_{t_0}^{t_f} d\tau \frac{\delta \mathcal{L}}{\delta \tilde{x}_t} \\ &= \int_{t_0}^{t_f} d\tau \left[\frac{\partial \mathcal{L}}{\partial \tau} \frac{\delta \tau}{\delta \tilde{x}_t} + \frac{\partial \mathcal{L}}{\partial x} \frac{\delta x}{\delta \tilde{x}_t} + \dots + \frac{\partial \mathcal{L}}{\partial \overset{(n)}{x}} \frac{\delta \overset{(n)}{x}}{\delta \tilde{x}_t} \Big|_{x=\tilde{x}_\tau} \right]. \end{aligned}$$

Evalutando la condición $\overset{(k)}{x} = \frac{d^k}{dt^k} \tilde{x}_\tau$, es posible utilizar las relaciones dadas para realizar una derivada funcional (5.4), (5.5), (5.6) y (5.7)

$$0 = \int_{t_0}^{t_f} d\tau \left[\frac{\partial \mathcal{L}}{\partial x} \delta(\tau - t) + \dots + \frac{\partial \mathcal{L}}{\partial \overset{(n)}{x}} \delta^{(n)}(\tau - t) \right].$$

Evalutando las integrales por medio de las propiedades de la delta de Dirac se obtiene

$$0 = \partial_x \mathcal{L} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{x}} + \dots + (-1)^n \frac{d^n}{dt^n} \left(\partial_{\overset{(n)}{x}} \mathcal{L} \right) \quad (5.10)$$

Este resultado es conocido como la ecuación de Euler-Lagrange, comunmente utilizado como una “receta” para obtener una derivada funcional de una acción dada por la integral de un lagrangiano (5.9).

5.2.2. Probabilidad Funcional

La notación bayesiana puede ser extendida para cuando hablamos de funcionales, escribiendo la probabilidad como $P[\tilde{x}|I]$. Esta es una probabilidad funcional, la cual representa la probabilidad de tener una trayectoria \tilde{x} , dada la información² I .

²Se entiende por información dada, un conjunto de proposiciones lógicas que se asumen como verdaderas para el cálculo de dicha probabilidad.

Esta forma de ver las probabilidades se ha introducido por autores como N. Wiener [29] y Richard Feynman [30], aunque cabe destacar que está fuera de la mecánica estadística tradicional en que tenemos probabilidades de micro-estados, es decir, la probabilidad de un vector x de dimensión n el cual representa los valores de posición y momento de una configuración para el espacio de fases, donde la probabilidad para cada micro-estado se escribe como $P(x|I)$.

Para el caso de la probabilidad funcional podemos tener la probabilidad de todas las micro-trayectorias \tilde{x} seguidas por los micro-estados en cada tiempo, de la forma $P[\tilde{x}|I]$. Estas micro-trayectorias \tilde{x} , al ser evaluadas en un tiempo t , corresponden a los micro-estados x de la ME, es decir, $\tilde{x}_t = x$ tal que x es un micro-estado. Esto puede ser visto a modo de ejemplo en la figura 5.1.

Es bueno destacar que la probabilidad funcional no depende de los argumentos de \tilde{x} , pero sí puede ser dependiente de parámetros externos, como por ejemplo condiciones de borde, multiplicadores de Lagrange, entre otros.

5.2.3. Integral Funcional: Expectaciones de Funcionales y Funciones

El Principio de Máximo Calibre, nos permite asignar probabilidades a funciones \tilde{x} , que cumplen las condiciones de borde Ω . Estas condiciones de borde están definidas en el espacio de trayectorias, y corresponden a valores conocidos de los micro-estados en tiempos dados. Por ejemplo, si conocemos que los micro-estados en $t = 0$ están en la configuración x_0 y en el tiempo $t = 1$ se encuentran en la configuración x_1 , tendremos condiciones de borde de la forma $\Omega : \tilde{x}_0 = x_0 \wedge \tilde{x}_1 = x_1$, siendo la probabilidad de las trayectorias que cumplen las condiciones de borde $P[\tilde{x}|\Omega]$. Estas condiciones pueden ser complejas o simples, dependiendo de que sabemos del

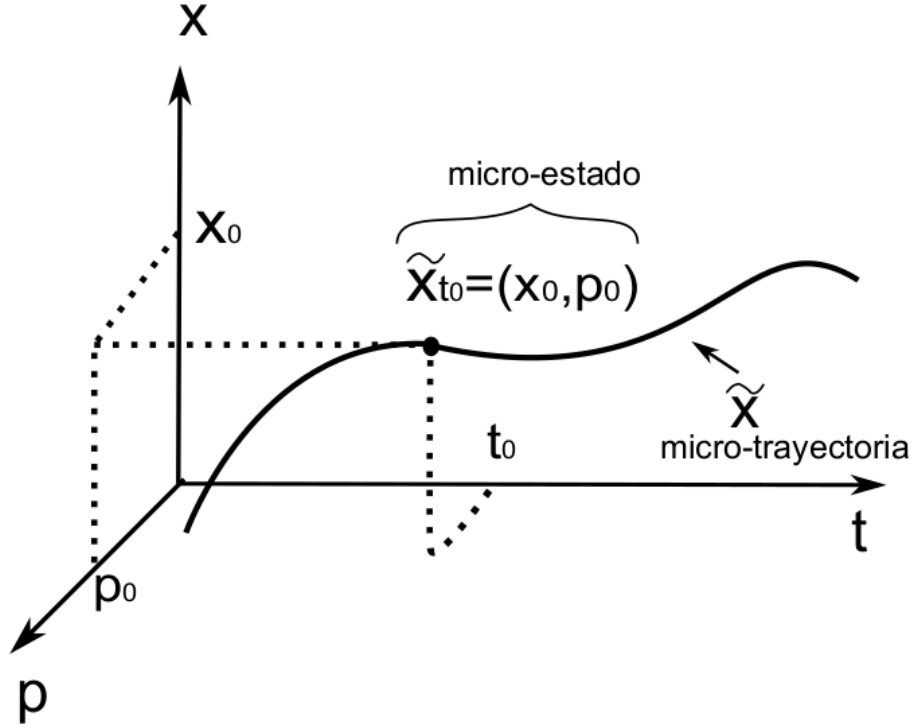


Figura 5.1: Representación esquemática de una micro-trayectoria $\tilde{x} \in \mathbb{X}$, la cual al ser evaluadas en un tiempo t_0 corresponde a un micro-estado (x_0, p_0) .

sistema, siendo posible definir sólo una condición inicial, como también dos o más condiciones de borde dadas.

De acuerdo a la regla de la normalización (2.10), la probabilidad de las trayectorias $P[\tilde{x}|I]$ tal que $\tilde{x} \in \Omega \subseteq \mathbb{X}$, está normalizada sobre el espacio de las trayectorias que cumplen las condiciones de borde Ω . En lo que sigue, utilizamos la información dada como $I : [\Omega, \text{Información Adicional}]$, debido a que las condiciones de borde son sólo una parte de toda la información que define una probabilidad. Escribimos esta normalización como

$$\langle 1 \rangle_I = \int_{\Omega} d\tilde{x} P[\tilde{x}|I] = 1. \quad (5.11)$$

La suma sobre todos los elementos del espacio de trayectorias que cumplen las

condiciones de borde es llamada una integral funcional, integral de trayectorias o integral de caminos, la cual escribiremos como $\int_{\Omega} d\tilde{x}$, y es tal que evalúa todas las funciones \tilde{x} que cumplen las condiciones de borde Ω . Es una extensión de la definición de integral desde espacio de estados, hacia el espacio de funciones \mathbb{X} .

Estas integrales son ampliamente utilizadas por ejemplo en mecánica cuántica, por medio de la formulación de integrales de caminos de R. P. Feynman [30, 59], así como en otras áreas como lo son procesos estocásticos [60, 61].

Por medio de la integral funcional es posible también definir la operación expectativa: cuando tenemos un funcional de probabilidad, ésta se escribe como

$$\langle F[\tilde{x}] \rangle = \int_{\Omega} d\tilde{x} P[\tilde{x}|I] F[\tilde{x}] = f, \quad (5.12)$$

es decir, el resultado de la operación expectativa de un funcional F , es un número real $f \in \mathbb{R}$. Por otra parte, también podemos tomar expectativa a funciones de la forma

$$\langle G(t, x, \dot{x}, \dots, \overset{(n)}{x}) \rangle_{I,t} = \int_{\Omega} d\tilde{x} P[\tilde{x}|I, t] G(t, x, \dot{x}, \dots, \overset{(n)}{x}) = g(t). \quad (5.13)$$

Como podemos ver, en este caso G es un funcional particular que representa una función dependiente de valores reales t, x, \dot{x} , etc. y no dependiente de una trayectoria completa \tilde{x} (como es en general un funcional), sino de los valores instantáneos de las trayectorias, ya que $x = \tilde{x}_t$, $\dot{x} = \frac{d}{dt} \tilde{x}_t$, ..., $\overset{(n)}{x} = \frac{d^n}{dt^n} \tilde{x}_t$. Por otra parte, el sub-índice t en la expectativa $\langle \cdot \rangle_{I,t}$ indica la dependencia temporal del resultado de la expectativa. Lo interesante de este segundo caso es que entrega una manera de obtener cantidades macroscópicas tiempo-dependientes, utilizando para ello una probabilidad de trayectorias. De esta manera se obtiene una primera aproximación para definir una probabilidad de trayectorias, permite entender las relaciones entre

las cantidades macroscópicas tiempo dependientes, esto en el espíritu de encontrar una teoría para derivar la termodinámica fuera del equilibrio.

5.2.4. Delta de Dirac funcional

Es posible definir una generalización de la delta de Dirac funcional, cuyas propiedades en una integral funcional son análogas a la delta en el espacio de estados, quedando de la siguiente forma. Si $\tilde{y} \in \Omega$, es decir \tilde{y} es una función definida bajo las condiciones de borde Ω , entonces la delta de Dirac funcional cumple

$$\int_{\Omega} d\tilde{x} \delta[\tilde{x} - \tilde{y}] = 1. \quad (5.14)$$

$$\int_{\Omega} d\tilde{x} \delta[\tilde{x} - \tilde{y}] F[\tilde{x}] = F[\tilde{y}]. \quad (5.15)$$

$$\int_{\Omega} d\tilde{x} \delta[\tilde{x} - \tilde{y}] g(t, x, \dots, \overset{(n)}{x}) \Big|_{x=\tilde{x}_t} = g(t, x, \dot{x}, \dots, \overset{(n)}{x}) \Big|_{x=\tilde{y}_t}. \quad (5.16)$$

El uso de la delta de Dirac será fundamental en los próximos capítulos para entender como pasar de una representación de funcionales para la probabilidad a una representación instantánea para la probabilidad, es decir ir de probabilidades de trayectorias a probabilidades dependientes de estados y el tiempo como las utilizadas en la mecánica estadística fuera del equilibrio.

5.3. Principio de Máximo Calibre

El formalismo de máximo calibre permite construir funcionales de probabilidad por medio de restringir funciones tiempo-dependientes, es decir, permite restringir propiedades macroscópicas fuera del equilibrio. De esta forma también nos entrega un formalismo que permite construir funcionales de probabilidad a partir de funcionales. El calibre C es un funcional dependiente de la probabilidad, con la misma forma que

la entropía pero definido en el espacio de trayectorias,

$$C[I_0 \rightarrow I] = - \int_{\Omega} d\tilde{x} P[\tilde{x}|I] \ln \frac{P[\tilde{x}|I]}{P[\tilde{x}|I_0]}. \quad (5.17)$$

Aquí podemos notar que las probabilidades corresponden a funcionales dependientes de \tilde{x} . Desde el punto de vista de la teoría de información [48], se puede entender la entropía o calibre en este caso como una distancia entre 2 probabilidades $P[\tilde{x}|I]$ y $P[\tilde{x}|I_0]$, siendo el calibre dependiente de ambas probabilidades. Para reducir la notación, se define $C[P[\tilde{x}|I], P[\tilde{x}|I_0]] := C[I_0 \rightarrow I]$, dando a entender que al maximizar el calibre C nos entregará la probabilidad luego de actualizar la información conocida de I_0 a I .

En resumen, el principio de Máximo Calibre nos permite encontrar la forma del funcional de probabilidad dadas restricciones, las cuales pueden ser de dos formas: restricciones de funciones o restricciones de funcionales, que serán vistas en detalle en la siguiente sección.

5.3.1. MaxCal para restricciones de funcionales arbitrarios

El uso de MaxCal cuando se restringen cantidades funcionales nos entregará modelos análogos a los de MaxEnt en los cuales los multiplicadores de Lagrange son valores constantes y las restricciones son independientes del tiempo.

Para un modelo de MaxCal con restricciones de funcionales, dadas N restricciones, definidas entre los tiempos t_0 y t_f , análogo a (5.12), de la forma

$$\langle F_n[\tilde{x}] \rangle = \int_{\Omega} d\tilde{x} P[\tilde{x}|I] F_n[\tilde{x}] = f_n, \quad (5.18)$$

donde $F_n : \mathbb{X} \rightarrow \mathbb{R}$ es un funcional dependiente de la función $\tilde{x} \in \mathbb{X}$, cuya expectativa tiene un valor constante $f_n \in \mathbb{R}$, si maximizamos el calibre (5.17), bajo estas

restricciones y que la probabilidad esta normalizada (5.11). Definiendo $\rho[\tilde{x}] := P[\tilde{x}|I]$ y $\pi[\tilde{x}] := P[\tilde{x}|I_0]$ para reducir la notación se obtiene

$$0 = \frac{\delta}{\delta\rho[\tilde{y}]} \left[C[\rho[\tilde{x}], \pi[\tilde{x}]] - \lambda \left(\langle 1 \rangle - 1 \right) - \sum_i \alpha_i \left(\langle F_i \rangle - f_i \right) \right].$$

Entonces, usando la misma definición para la derivada funcional, pero ahora derivando respecto a un funcional, y teniendo en cuenta que

$$\frac{\delta\rho[\tilde{x}]}{\delta\rho[\tilde{y}]} = \delta[\tilde{x} - \tilde{y}],$$

donde $\delta[\tilde{x} - \tilde{y}]$ corresponde a la Delta de Dirac funcional explicada en la sección anterior, obtenemos que

$$\begin{aligned} 0 = & - \int_{\Omega} d\tilde{x} \left(\ln \frac{\rho[\tilde{x}]}{\pi[\tilde{x}]} \delta[\tilde{x} - \tilde{y}] + \delta[\tilde{x} - \tilde{y}] \right) - \lambda \int_{\Omega} d\tilde{x} \delta[\tilde{x} - \tilde{y}] \\ & - \sum_i \alpha_i \int_{\Omega} d\tilde{x} \delta[\tilde{x} - \tilde{y}] F_i[\tilde{x}]. \end{aligned}$$

Entonces, utilizando las propiedades de la Delta de Dirac funcional (5.14) y (5.15) se obtiene

$$0 = - \ln \frac{\rho[\tilde{y}]}{\pi[\tilde{y}]} - 1 - \lambda - \sum_i \alpha_i(t) F_i[\tilde{y}].$$

Aplicando la función exponencial se obtiene la forma de ρ , es decir, la probabilidad para cada trayectoria \tilde{y} , como

$$\rho[\tilde{y}] = \pi[\tilde{y}] \exp \left(-1 - \lambda - \sum_i \alpha_i F_i[\tilde{y}] \right). \quad (5.19)$$

Entonces si se define $Z(\alpha) = \exp \left(1 + \lambda \right)$, es decir, incluimos toda la parte constante de la exponencial en la normalización, obtenemos la forma que tendrá la probabilidad funcional cuando se restringen funciones de la forma (5.18)

$$\rho[\tilde{y}] = \frac{\pi[\tilde{y}]}{Z(\alpha)} \exp \left(- \sum_i \alpha_i F_i[\tilde{y}] \right). \quad (5.20)$$

Utilizando la forma de la probabilidad encontrada y su propiedad de estar normalizada (5.11), se obtiene una relación para evaluar la función partición Z , análoga a la función partición en los modelos MaxEnt (3.12)

$$Z(\alpha) = \int_{\Omega} d\tilde{x} \pi[\tilde{y}] \exp\left(-\sum_i \alpha_i F_i[\tilde{y}]\right) \quad (5.21)$$

La relación entre las restricciones (5.18) y los multiplicadores de Lagrange $\alpha : (\alpha_1, \dots, \alpha_n)$ está dada por la función partición Z , mediante la relación análoga a la de MaxEnt (3.13),

$$-\frac{\partial}{\partial \alpha_i} \ln Z(\alpha) = f_i. \quad (5.22)$$

Este tipo de restricciones entregan modelos similares a los obtenidos con MaxEnt, donde los multiplicadores de Lagrange son valores reales constantes $\alpha_i, f_i \in \mathbb{R}$.

5.3.2. MaxCal para restricciones de funciones arbitrarias

El caso de MaxCal para restricciones de funciones arbitrarias es el más importante para desarrollar una teoría estadística fuera del equilibrio, ya que ayuda a entender cómo construir probabilidades utilizando MaxCal cuando se restringen funciones que son dependientes del tiempo, la posición y sus derivadas superiores.

Consideremos N restricciones para todo t entre los tiempos t_0 y t_f , al igual que (5.13),

$$\begin{aligned} \langle \mathcal{L}_i(t, x, \dot{x}, \dots) \Big|_{x=\tilde{x}_t} \rangle_{I,t} &= \int_{\Omega} d\tilde{x} P[\tilde{x}|I] \mathcal{L}_i(t, x, \dot{x}, \dots) \Big|_{x=\tilde{x}_t} \\ &= \tilde{L}_t^i, \end{aligned} \quad (5.23)$$

donde \mathcal{L}_i es una función dependiente del tiempo t , la posición \tilde{x}_t y derivadas temporales de \tilde{x}_t , definida su expectación para cada tiempo t entre t_0 y t_f con el valor $\tilde{L}_t^i \in \mathbb{R}$, es decir $\tilde{L}^i \in \mathbb{X}$. Si maximizamos el calibre (5.17), bajo las restricciones

(5.23) y que la probabilidad esta normalizada (5.11). Definiendo $\rho[\tilde{x}] = P[\tilde{x}|I]$ y $\pi[\tilde{x}] = P[\tilde{x}|I_0]$, se tiene

$$0 = \frac{\delta}{\delta\rho[\tilde{y}]} \left[C[\rho[\tilde{x}], \pi[\tilde{x}]] - \lambda \left(\langle 1 \rangle_I - 1 \right) - \sum_i \int_{t_0}^{t_f} dt \tilde{\beta}^i_t \left(\langle \mathcal{L}_i(t, x, \dot{x}, \dots, \overset{(n)}{x}) \Big|_{x=\tilde{x}_t} \rangle_{I,t} - \mathcal{L}_i(t) \right) \right].$$

Notemos que al estar la i -ésima restricción (5.23) definida en un intervalo de tiempo significa que para cada tiempo se tiene un valor de multiplicador de Lagrange diferente, de forma que la i -ésima restricción en un tiempo t tiene su multiplicador de Lagrange $\tilde{\beta}^i_t \in \mathbb{R}$.

Visto de una manera más general, para cada i -ésima restricción existe un multiplicador de Lagrange asociado $\tilde{\beta}^i$ (con $i \in [0, N]$), donde los multiplicadores corresponden a funciones $\tilde{\beta}^i \in \mathbb{X}$.

Usando la misma definición para la derivada funcional que empleamos en la derivación anterior, vemos que

$$0 = \int_{\Omega} d\tilde{x} \left(\ln \frac{\rho[\tilde{x}]}{\pi[\tilde{x}]} \delta[\tilde{x} - \tilde{y}] + \delta[\tilde{x} - \tilde{y}] \right) - \lambda \int_{\Omega} d\tilde{x} \delta[\tilde{x} - \tilde{y}] - \sum_i \int_{t_0}^{t_f} dt \tilde{\beta}^i_t \int_{\Omega} d\tilde{x} \delta[\tilde{x} - \tilde{y}] \mathcal{L}_i(t, x, \dot{x}, \dots, \overset{(n)}{x}) \Big|_{x=\tilde{x}_t}.$$

Si ahora utilizamos las propiedades de la Delta de Dirac funcional (5.14) y (5.16), tenemos que

$$0 = - \ln \frac{\rho[\tilde{y}]}{\pi[\tilde{y}]} - 1 - \lambda - \sum_i \int_{t_0}^{t_f} dt \tilde{\beta}^i_t \mathcal{L}_i(t, x, \dot{x}, \dots, \overset{(n)}{x}) \Big|_{x=\tilde{y}_t}. \quad (5.24)$$

Aplicando la función exponencial y despejando ρ obtenemos

$$\rho[\tilde{y}] = \pi[\tilde{y}] \exp \left(-1 - \lambda - \sum_i \int_{t_0}^{t_f} dt \tilde{\beta}^i_t \mathcal{L}_i(t, x, \dot{x}, \dots) \Big|_{x=\tilde{y}_t} \right). \quad (5.25)$$

Si se define $Z[\tilde{\beta}] = \exp(1 + \lambda)$, es decir incluimos toda la parte constante de la exponencial en la normalización, la cual será llamada funcional de partición³, obtenemos la forma que tendrá la probabilidad funcional cuando se restringen funciones de la forma (5.23)

$$\rho[\tilde{y}] = \frac{\pi[\tilde{y}]}{Z[\tilde{\beta}(\cdot)]} \exp\left(-\sum_i \int_{t_0}^{t_f} dt \tilde{\beta}_t^i \mathcal{L}_i(t, x, \dot{x}, \dots)\Big|_{x=\tilde{y}_t}\right). \quad (5.26)$$

Utilizando la forma de la probabilidad y su propiedad de estar normalizada (5.11), se muestra una forma de obtener el funcional partición Z ,

$$Z[\tilde{\beta}] = \pi[\tilde{y}] \exp\left(-\sum_i \int_{t_0}^{t_f} dt \tilde{\beta}_t^i \mathcal{L}_i(t, x, \dot{x}, \dots)\Big|_{x=\tilde{x}_t}\right). \quad (5.27)$$

La relación entre las restricciones (5.23) y los multiplicadores de Lagrange $\tilde{\beta} : (\tilde{\beta}^1, \dots, \tilde{\beta}^n)$ es dada gracias al funcional de partición Z , mediante la ecuación

$$-\frac{\delta}{\delta \tilde{\beta}_t^i} \ln Z[\tilde{\beta}] = \tilde{L}_t^i, \quad (5.28)$$

la cual es válida para todo $t \in [t_0, t_f]$ y $i \in [1, \dots, N]$.

5.4. Implementación numérica de MaxCal

Los algoritmos no deterministas para resolver problemas de optimización⁴ de forma numérica han tenido un gran aumento a medida se incrementa la capacidad de cálculo de los computadores. Uno de los algoritmos más populares es llamado algoritmo de Monte Carlo Metropolis (MCM) [62, 63], el cual corresponde a un procedimiento que muestrea los elementos de un espacio en particular. Este algoritmo

³La función partición es dependiente de cada multiplicador de Lagrange, los cuales en este caso son funciones, es decir $\tilde{\beta} : (\tilde{\beta}^1, \dots, \tilde{\beta}^n)$.

⁴Seleccionar el mejor elemento de acuerdo a un criterio dado a partir de un conjunto de elementos disponibles, comúnmente los problemas de optimización están relacionados con la búsqueda de extremos (máximo, mínimo o punto de ensilladura) de una función dada.

permite encontrar las soluciones a problemas de optimización, muestreo estadístico, cálculo de integrales multidimensionales, entre otros.

En las siguientes secciones se detallarán los elementos necesarios a considerar para la implementación del algoritmo MCM en el espacio de trayectorias.

Una ventaja de implementar el algoritmo MCM en el espacio de trayectorias es que mediante el muestreo de este espacio se obtendrá la trayectoria que minimiza un funcional dado. MCM es entonces un procedimiento numérico alternativo al utilizado en Mecánica Clásica cuando buscamos la trayectoria que minimiza la acción clásica mediante la ecuación de Euler-Lagrange.

Recordemos que la trayectoria que cumple con ser el mínimo⁵ de un funcional debe ser la solución a la ecuación de Euler-Lagrange asociada al funcional. Esto hace que MCM en el espacio de trayectorias sea una metodología válida para *resolver ecuaciones diferenciales muestreando trayectorias aleatorias* del espacio de trayectorias.

5.4.1. Monte Carlo Metropolis en el espacio de trayectorias

MCM consiste en realizar un muestreo de los elementos del espacio en que se define una probabilidad, generalmente en un espacio de estados $x \in \mathbb{R}^n$. Este muestreo entrega soluciones aproximadas a un gran número de problemas matemáticos que son complejos y costosos de evaluar con exactitud [62, 63], posibilitando así la realización de experimentos con muestreos de números aleatorios. Esta metodología es aplicable tanto a problemas de naturaleza estocástica como determinista y a diferencia de otros métodos numéricos cuyo error se basan en evaluaciones dependientes

⁵En rigor no es solo un mínimo, sino un extremo del funcional mínimo, máximo o punto de ensilladura, siendo necesario evaluar la segunda derivada para saber de qué tipo de extremo se trata. Éste problema será visto en detalle mas adelante en el capítulo.

del tamaño del espacio que se está muestreando, MCM tiene un error absoluto de la estimación que decrece como $\frac{1}{\sqrt{N}}$ [63].

Otra de las grandes ventajas de MCM, es que para realizar el muestreo requiere sólo conocer el valor mínimo entre 1 y la razón entre las probabilidades de 2 muestras x_0 y x_1 , lo que para modelos MaxEnt (3.11) queda como

$$\frac{P(x_0|I)}{P(x_1|I)} = \frac{\exp(-\beta f(x_0))}{Z(\beta)} \frac{Z(\beta)}{\exp(-\beta f(x_1))} \quad (5.29)$$

$$= \frac{\exp(-\beta f(x_0))}{\exp(-\beta f(x_1))}. \quad (5.30)$$

Este hecho permite obtener propiedades de los sistemas sin la necesidad de conocer el valor de la función de partición del modelo, muchas veces imposible de calcular en la práctica.

Para MaxCal, la probabilidad se encuentra definida en el espacio de trayectorias, por lo que para realizar el procedimiento será necesario generar trayectorias aleatorias. Los elementos involucrados y el procedimiento a realizar serán detallado a continuación. Dada una restricción de la forma

$$\langle A[\tilde{x}] \rangle_I = a, \quad (5.31)$$

donde

$$A[\tilde{x}] = \int_{t_0}^{t_f} dt l(t, \tilde{x}_t, \dot{\tilde{x}}_t) \quad (5.32)$$

es un funcional que puede ser entendido igual que la acción clásica, definido en el espacio de las trayectorias que cumplen las condiciones de borde ($\tilde{x}_{t_0} = x_0$ y $\tilde{x}_{t_f} = x_f$), podemos maximizar el calibre (5.17) con respecto a la probabilidad funcional, bajo la restricción de normalización (5.11) y la restricción anterior, y así

obtener la función de probabilidad dada por la forma de (5.20)

$$P[\tilde{x}|I] = \frac{1}{Z(\alpha)} \exp\left(-\alpha A[\tilde{x}]\right). \quad (5.33)$$

Con esta probabilidad podemos encontrar la trayectoria \tilde{x} que hace mínimo $A[\tilde{x}]$. Para esto comenzamos con una trayectoria de prueba inicial \tilde{x}^0 , que al ser evaluada en el funcional entrega $A[\tilde{x}^0] = A_0$. Entonces, realizamos de forma iterativa los siguientes pasos:

- Modificamos \tilde{x}^0 a una trayectoria cercana \tilde{x}^1 , entonces evaluamos la acción en la nueva trayectoria, $A[\tilde{x}^1] = A_1$.
- Calculamos la diferencia $\Delta A = A_1 - A_0$:
 - Si $\Delta A < 0$, entonces el cambio se acepta y repetimos el procedimiento, utilizando \tilde{x}^1 como trayectoria inicial.
 - En otro caso, es necesario generar un número aleatorio $R \in [0, 1]$, el cual se utilizará para implementar la probabilidad de aceptación de trayectorias que no son más probables que la trayectoria previa. Entonces
 - Si $R < \frac{P[\tilde{x}^1|I]}{P[\tilde{x}^0|I]} = \exp\left(-\alpha \Delta A\right)$, el cambio se acepta y se procede a repetir el procedimiento utilizando \tilde{x}^1 como trayectoria inicial.
 - En otro caso, el cambio se rechaza y se procede a repetir el procedimiento utilizando \tilde{x}^0 como trayectoria inicial.

Iterar este procedimiento nos permite generar trayectorias que se distribuyen según $P[\tilde{x}|I]$, donde los promedios calculados sobre las trayectorias muestreadas convergen al resultado de integrales de caminos gracias a la ley de los grandes números.

De la misma forma, la trayectoria que más se repite, la cual tiende a la trayectoria más probable, hace mínimo el funcional A .

Es bueno hacer notar que la trayectoria de mínima acción puede ser obtenida resolviendo la ecuación diferencial de Euler-Lagrange asociada al funcional, por lo que este algoritmo de Monte Carlo constituye un procedimiento que permite resolver ecuaciones diferenciales mediante el muestreo del espacio de trayectorias.

5.4.2. Curvas de Bézier para simular trayectorias

Para realizar el muestreo en el espacio de trayectorias, requerido por nuestra aplicación del algoritmo de Monte Carlo, es necesario representar el espacio de trayectorias utilizando una base conveniente de dimensión finita.

Dado que en este espacio las trayectorias se encuentran restringidas a condiciones de borde $\tilde{x}_{t_0} = x_0$ y $\tilde{x}_{t_f} = x_f$, es que se utilizarán curvas de Bézier para representarlas, quedando definidas como se explica a continuación.

Para una trayectoria N -dimensional \tilde{x} , con las condiciones de borde $\tilde{x}_{t_0} = P_0$ y $\tilde{x}_{t_f} = P_n$, una curva Bézier es definida como

$$\tilde{x}_t = \sum_{i=0}^n P_i b_i(t; n). \quad (5.34)$$

Donde las bases b_i son los denominados polinomios de Bernstein, dados por

$$b_i(t; n) = \binom{n}{i} \frac{(t - t_0)^i (t_f - t)^{n-i}}{(t_f - t_0)^n}. \quad (5.35)$$

Los coeficientes P_i , llamados puntos de control, son utilizados para definir cada curva, de manera que a mayor número de puntos de control utilizados, mejor es el muestreo del espacio de las trayectorias, pero más alto el costo computacional.

Dada la definición anterior, es fácil ver que la trayectoria cumple automáticamente las condiciones de borde requeridas para t_0 y t_f dadas por $\tilde{x}_{t_0} = P_0$ y $\tilde{x}_{t_f} = P_n$.

5.4.3. Aproximación de curvas como líneas rectas

Para la implementación de MCM para trayectorias es necesario evaluar un funcional asociado a cada trayectoria, las cuales son descritas como curvas Bézier. El costo computacional de cada evaluación es directamente creciente con el número de puntos de control utilizados para describir cada curva. Como se requiere utilizar un número mínimo de puntos dependiendo de la forma de la trayectoria a representar⁶, se implementó un esquema de interpolación lineal para representar la trayectoria, donde se separa el intervalo de tiempo total $\tau = t_f - t_0$ en N sub-intervalos, quedando entonces la trayectoria definida para cada k -ésimo sub-intervalo de la forma

$$\tilde{x}_t = x_k + (t - t_k) \left[\frac{x_{k+1} - x_k}{t_{k+1} - t_k} \right], \quad (5.36)$$

con $t_k \leq t < t_{k+1}$, $k \in [0, N]$ y $t_f = t_N$. Es decir, dependiente del t en que se busca evaluar la trayectoria, es necesario identificar que $t \in [t_k, t_{k+1})$ para saber los valores de x_k y x_{k+1} respectivos. Los puntos x_k serán obtenidos de acuerdo a la evaluación de las curvas de Bézier dados por,

$$x_k = \sum_{j=0}^m P'_j b_j(t_k; m), \quad (5.37)$$

por lo que para evaluar un funcional dependiente de una trayectoria, sólo es necesario evaluar los $N - 2$ valores x_k (ya que son fijos x_0 y x_N) necesarios para realizar la interpolación y así utilizar (5.36), mucho menos costosa de evaluar que la asociada al valor para cada tiempo de una curva Bézier (5.34). Esto reduce el costo computacional que se requiere para evaluar una curva Bézier. Otra característica a destacar

⁶Esto puede ser entendido de acuerdo a la siguiente regla: si se utilizan 2 puntos de control sólo es posible representar trayectorias rectas; si se utilizan 3 puntos de control sólo es posible representar curvas parabólicas y así sucesivamente. Un mayor número de puntos de control permite describir trayectorias cada vez más complejas.

de esta aproximación es que el promedio de las trayectorias calculadas usando Monte Carlo no parece ser afectado por el uso de la aproximación lineal, por lo que el algoritmo es capaz de reproducir trayectorias suaves de todas formas.

5.5. Ejemplos

5.5.1. *Prior* funcional: MaxCal cuando no hay restricciones

Recordemos que el valor absoluto de un multiplicador de Lagrange será mayor mientras más importante sea la restricción, y más cercano a cero mientras menos importante sea para el modelo a describir. Es por esto que cuando el multiplicador de Lagrange tiende a cero, el modelo tiende a la probabilidad previa o *prior*.

Como es común a todos los modelos MaxEnt, en MaxCal también existe el elemento llamado *prior*, que corresponde a la probabilidad previa a incluir información, $P[\tilde{x}|I_0]$. Una buena forma de entender el *prior* es considerar un modelo MaxCal en que no hay restricciones además de la que fija que la probabilidad sea normalizable, en cuyo caso la probabilidad corresponde al *prior*. Verificamos esto por inspección de (3.11) con $\beta_i \rightarrow 0$, o directamente maximizando el Calibre (5.17) bajo la única restricción que la probabilidad esté normalizada (5.11). Definiendo $\rho[\tilde{x}] = P[\tilde{x}|I]$ y $\pi[\tilde{x}] = P[\tilde{x}|I_0]$, escribimos

$$\frac{\delta}{\delta\rho[\tilde{y}]} \left[C[\rho[\tilde{x}], \pi[\tilde{x}]] - \lambda (\langle 1 \rangle - 1) \right] = 0,$$

donde usando la definición para la derivada funcional (5.4), vemos que

$$\int_{\Omega} d\tilde{x} \left[- \left(\ln \frac{\rho[\tilde{x}]}{\pi[\tilde{x}]} \delta[\tilde{x} - \tilde{y}] + \delta[\tilde{x} - \tilde{y}] \right) - \lambda \delta[\tilde{x} - \tilde{y}] \right] = 0.$$

Utilizando las propiedades de la Delta de Dirac funcional (5.14) y (5.15), para

resolver las integrales de camino se obtiene

$$0 = -\ln \frac{\rho[\tilde{y}]}{\pi[\tilde{y}]} - 1 - \lambda,$$

por lo tanto,

$$\rho[\tilde{y}] = \pi[\tilde{y}] \exp(-1 - \lambda). \quad (5.38)$$

Es fácil notar que usando la regla de normalización se sigue que el valor de la exponencial sólo depende del *prior* π . De esta forma, se absorbe este valor como una constante en el *prior*, y se obtiene

$$\rho[\tilde{y}] = \pi[\tilde{y}]. \quad (5.39)$$

Así vemos que un modelo MaxCal en que no hay restricciones, la probabilidad es el mismo *prior*, ya que es el *prior* la única información disponible para realizar inferencia.

5.5.2. Derivada Funcional de la acción clásica.

Si buscamos la trayectoria seguida por una partícula clásica de masa m , bajo un potencial $V(\tilde{x}_t, t)$, tendremos que dicha partícula minimizará la acción clásica A_{cl} , la cual es de la forma

$$A_{cl}[\tilde{x}] = \int_{t_0}^{t_f} dt \left[\frac{m(t)\dot{x}^2}{2} - V(x, t) \right] \Big|_{x=\tilde{x}_t}. \quad (5.40)$$

Si aplicamos la derivada funcional a la forma de la acción clásica, tal como indica

el principio de mínima acción obtenemos

$$\begin{aligned}
0 &= \frac{\delta A_{cl}[\tilde{x}]}{\delta \tilde{x}_\tau} \\
&= \frac{\delta}{\delta \tilde{x}_\tau} \int_{t_0}^{t_f} dt \left[\frac{m(t)\dot{x}^2}{2} - V(x, t) \Big|_{x=\tilde{x}_t} \right] \\
&= \int_{t_0}^{t_f} dt \left[m(t)\dot{x} \frac{\delta \dot{x}}{\delta \tilde{x}_\tau} - \partial_x V(x, t) \frac{\delta x}{\delta \tilde{x}_\tau} \Big|_{x=\tilde{x}_t} \right] \\
&= \int_{t_0}^{t_f} dt \left[m(t)\dot{x} \delta'(t - \tau) - \partial_x V(x, t) \delta(t - \tau) \Big|_{x=\tilde{x}_t} \right].
\end{aligned}$$

Evaluando las integrales mediante el uso de las propiedades de la delta de Dirac, vemos que recuperamos la ecuación de Euler-Lagrange,

$$0 = -\frac{d}{d\tau} \left(m(\tau)\dot{x} \right) - \partial_x V(x, \tau) \Big|_{x=\tilde{x}_\tau}, \quad (5.41)$$

Que corresponde entonces a la ecuación de movimiento seguida por una partícula clásica bajo un potencial $V(x, t)$. La partícula sigue una dinámica newtoniana $\frac{dp}{dt} = F$ con $p = m \dot{x}$.

5.5.3. La trayectoria más probable: Principio de mínima acción desde MaxCal.

Se procederá a demostrar cómo en un modelo MaxCal en que se restringe la acción clásica, las trayectorias más probables seguirán el principio de mínima acción. Si se restringe la acción clásica de la forma (5.13)

$$\langle A[\tilde{x}] \rangle_I = A_0, \quad (5.42)$$

el modelo de probabilidad obtenido será (5.20)

$$\rho[\tilde{x}] = \frac{\pi[\tilde{x}]}{Z(\alpha)} \exp \left(-\alpha A[\tilde{x}] \right). \quad (5.43)$$

Al ser la trayectoria más probable \tilde{x}^* automáticamente un extremo de ρ , se tiene

$$\frac{\delta \rho[\tilde{x}]}{\delta \tilde{x}_t} \Big|_{\tilde{x}=\tilde{x}^*} = 0. \quad (5.44)$$

Entonces, reemplazando la forma de la probabilidad en la ecuación anterior y utilizando un *prior* constante ($\pi[\tilde{x}] = c$) se obtiene

$$\begin{aligned} \frac{c}{Z(\alpha)} \frac{\delta}{\delta \tilde{x}_t} \exp\left(-\alpha A[\tilde{x}]\right) \Big|_{\tilde{x}=\tilde{x}^*} &= \exp\left(-\alpha A[\tilde{x}]\right) \frac{\delta}{\delta \tilde{x}_t} A[\tilde{x}] \Big|_{\tilde{x}=\tilde{x}^*} \\ &= 0. \end{aligned} \quad (5.45)$$

Para que la exponencial sea cero es necesario que $A \rightarrow \infty$ o $\alpha \rightarrow \infty$, por lo que la trayectoria más probable debe cumplir

$$\frac{\delta}{\delta \tilde{x}_t} A[\tilde{x}] = 0, \quad (5.46)$$

condición correspondiente al principio de mínima acción (5.8).

De esta forma podemos notar que un modelo MaxCal en que se restringe la acción clásica, describe trayectorias asociadas a dinámicas clásica pero con incerteza, es decir sistemas estocásticos que siguen dinámicas clásicas en promedio, no necesariamente deterministas en términos microscópicos.

5.5.4. Dinámicas Newtonianas desde MaxCal

Como se mostró en el ejemplo anterior, en un modelo MaxCal en el cual se restringe la acción clásica, las trayectorias más probables corresponden a las trayectorias que siguen el principio de mínima acción. Este es un resultado general para el cual no importa la forma que tenga la acción que fue restringida. De hecho es posible que la acción clásica sea emergente a partir de restricciones independientes.

A continuación se mostrará cuáles son las restricciones tiempo-dependientes con las cuales se obtiene un modelo que describe las dinámicas obtenidas en la mecánica clásica, es decir, se presentan las restricciones suficientes para obtener un modelo que describa este tipo de dinámicas, tal como ocurre cuando se restringe la acción clásica [64, 65].

La primera restricción que consideraremos indica el hecho que conocemos el promedio del cuadrado de la velocidad en todo tiempo, restricción que más adelante conectaremos con la energía cinética de la mecánica clásica,

$$\left\langle \frac{\dot{x}^2}{2} \Big|_{x=\tilde{x}_t} \right\rangle_{I,t} = k(t). \quad (5.47)$$

para todo $t \in [t_0, t_f]$.

La segunda restricción impone que se conoce la distribución de probabilidad para la posición en todo tiempo, restricción que más adelante conectaremos con la energía potencial de la mecánica clásica,

$$\left\langle \delta(x - y) \Big|_{x=\tilde{x}_t} \right\rangle_{I,y,t} = \rho(y, t). \quad (5.48)$$

para todo $t \in [t_0, t_f]$ y $x \in \mathbb{R}$, donde Podemos notar que esta restricción es una función dependiente de y y t , tal como lo indica el sub-índice en la expectación.

Si ahora maximizamos el calibre (5.17) con respecto a la probabilidad funcional, bajo la restricción de normalización (5.11) y las dos restricciones descritas previamente (5.47) y (5.48) correspondientes a restricciones tiempo dependientes, obtendremos la función de probabilidad dada por la forma de (5.26),

$$\rho[\tilde{x}] = \frac{\pi[\tilde{x}]}{Z[\tilde{\beta}]} \exp \left(- \int_{t_0}^{t_f} dt \frac{m(t)}{2} \dot{x}^2 + \int_{t_0}^{t_f} \int dt dy V(y, t) \delta(x - y) \Big|_{x=\tilde{x}_t} \right), \quad (5.49)$$

donde $m(t)$ es el multiplicador de Lagrange dependiente de tiempo asociado a la restricción (5.47), que en la acción clásica cumple el rol de la masa en el instante t y $V(y, t)$ es el multiplicador de Lagrange dependiente del tiempo y espacio, asociado a la restricción (5.48), que en la acción clásica cumple el rol del potencial de interacción.

Desarrollando la integral de la delta de Dirac y sin pérdida de generalidad, cambiando el signo del multiplicador de Lagrange V por $-V$, tenemos que

$$\rho[\tilde{x}] = \frac{\pi[\tilde{x}]}{Z[\tilde{\beta}]} \exp \left(- \int_{t_0}^{t_f} dt \left[\frac{m(t)}{2} \dot{x}^2 - V(x, t) \right] \Big|_{x=\tilde{x}_t} \right). \quad (5.50)$$

Podemos notar que la trayectoria más probable de este modelo cumple con seguir la segunda ley de Newton,

$$\frac{d}{dt} \left(m(t) \dot{x} \right) = - \frac{d}{dx} V(x, t).$$

Vemos que las restricciones (5.47) y (5.48), permiten construir un modelo de la forma

$$P[\tilde{x}] = \frac{1}{Z} \exp \left(- \lambda A[\tilde{x}] \right) \quad (5.51)$$

donde λ es una constante necesaria para que el argumento de la exponencial no tenga unidades, y donde A corresponde a la acción clásica,

$$A[\tilde{x}] = \int dt \left[\frac{m(t)\dot{x}^2}{2} - V(x, t) \right], \quad (5.52)$$

con $m(t)$ es la masa, la cual puede ser una función del tiempo, y $V(\tilde{x}_t, t)$ el potencial bajo el cual se encuentra la partícula, una función dependiente de la posición y el tiempo.

Si utilizamos el Teorema de la Variable Conjugada para funcionales (C.7), para el modelo correspondiente a las restricciones tiempo dependientes (5.47) y (5.48), al evaluar (5.51) en (C.7), se obtiene la identidad

$$\left\langle \frac{\delta}{\delta \tilde{x}_t} W[\tilde{x}] \right\rangle_{I,t} = \left\langle W[\tilde{x}] \frac{\delta}{\delta \tilde{x}_t} A[\tilde{x}] \right\rangle_{I,t}. \quad (5.53)$$

con W un funcional arbitrario y diferenciable.

Si utilizamos $W[\tilde{x}] = 1$, se obtiene que el sistema también sigue el principio de mínima acción en expectación

$$\begin{aligned} 0 = \left\langle \frac{\delta}{\delta \tilde{x}_t} 1 \right\rangle_{I,t} &= \left\langle 1 \frac{\delta}{\delta \tilde{x}_t} A[\tilde{x}] \right\rangle_{I,t}. \\ &= \left\langle \frac{\delta}{\delta \tilde{x}_t} A[\tilde{x}] \right\rangle_{I,t}. \end{aligned} \quad (5.54)$$

Estos resultados nos revelan que MaxCal entrega modelos de sistemas que siguen dinámicas clásicas en expectación, o también pueden ser vistos como modelos de dinámicas clásicas con incerteza (o alternativamente con fluctuaciones) en la posición.

5.5.5. Relación entre n -flujos

A continuación describiremos una aplicación para trayectorias continuas del problema de la interacción de N -flujos propuesto en el trabajo de Hazoglou *et al* (2015) [32], mostrando cómo se utiliza el formalismo presentado en esta tesis para trayectorias continuas (es decir, utilizando derivada e integral funcional) y cómo los resultados son consistentes con los formalismos reportados previamente por otros autores, donde se utilizan trayectorias como elementos discretos. Dado un sistema en el que se restringen N -flujos de la forma

$$\left\langle j_k(t, x, \dot{x}, \dots, \overset{(n)}{x}) \Big|_{x=\tilde{x}_t} \right\rangle_{I,t} = J_k(t), \quad (5.55)$$

que corresponden a la información I para crear el modelo, la forma del funcional de probabilidad obtenido al utilizar MaxCal, dado por (5.26), queda escrito como

$$\rho[\tilde{x}|I] = \frac{\exp\left(-\sum_k \int dt \tilde{\lambda}_t^k j_k(t, x, \dot{x}, \dots) \Big|_{x=\tilde{x}_t}\right)}{Z[\tilde{\lambda}]}. \quad (5.56)$$

Donde $\tilde{\lambda} = (\tilde{\lambda}^1, \tilde{\lambda}^2, \dots, \tilde{\lambda}^N)$, es un vector donde en cada componente tiene un multiplicador de Lagrange, los cuales en este contexto, son funciones, es decir, $\lambda^i \in \mathbb{X}$ con $i \in [1, N]$.

De esta forma, el Teorema de Fluctuación-Disipación para funcionales (C.16), para un funcional de probabilidad de la forma (5.56), es

$$\begin{aligned} \frac{\delta}{\delta \tilde{\lambda}_\tau^l} \langle W \rangle_I &= \left\langle \frac{\delta}{\delta \tilde{\lambda}_\tau^l} W \right\rangle_{I,\tau} - \langle W \rangle_I \frac{\delta}{\delta \tilde{\lambda}_\tau^l} \log Z[\tilde{\lambda}] \\ &+ \left\langle W \frac{\delta}{\delta \tilde{\lambda}_\tau^l} \left(\sum_k \int dt \tilde{\lambda}_t^k j_k(t, x, \dot{x}, \dots) \Big|_{x=\tilde{x}_t} \right) \right\rangle_{I,\tau}. \end{aligned} \quad (5.57)$$

La derivada de $\log Z$, puede salir de la expectación, ya que la función partición, no depende de las trayectorias \tilde{x} , que son la variable de integración al interior de la expectación. De la misma forma podemos relacionar la derivada del logaritmo de la función partición con su restricción correspondiente, como lo indica (3.13), de modo que

$$-\frac{\delta}{\delta \tilde{\lambda}^l_\tau} \log Z[\tilde{\lambda}] = \tilde{J}^l_\tau,$$

junto con utilizar (5.4), que en este caso queda de la forma

$$\frac{\delta \tilde{x}^k_t}{\delta \tilde{x}^l_\tau} = \delta(t - \tau) \delta_{kl},$$

donde δ_{kl} corresponde a la delta de Kronecker,

$$\begin{aligned} \frac{\delta}{\delta \tilde{\lambda}^l_\tau} \langle W \rangle_I &= \left\langle \frac{\delta}{\delta \tilde{\lambda}^l_\tau} W \right\rangle_{I,\tau} - \langle W \rangle_I J^l_\tau \\ &+ \left\langle W \left(\sum_k \int dt \delta(t - \tau) \delta_{kl} j_k(t, x, \dot{x}, \dots) \Big|_{x=\tilde{x}_t} \right) \right\rangle_{I,\tau}. \end{aligned} \quad (5.58)$$

Evaluando las deltas en la integral y sumatoria, vemos que el teorema de fluctuación-disipación puede ser escrito como

$$\begin{aligned} \frac{\delta}{\delta \tilde{\lambda}^l_\tau} \langle W \rangle_I &= \left\langle \frac{\delta}{\delta \tilde{\lambda}^l_\tau} W \right\rangle_{I,\tau} - \langle W \rangle_I J^l_\tau \\ &+ \left\langle W \left(j_l(\tau, x, \dot{x}, \dots) \Big|_{x=\tilde{x}_\tau} \right) \right\rangle_{I,\tau} \end{aligned} \quad (5.59)$$

Si ahora usamos la función de prueba

$$W[\tilde{x}; t] = j_k(t, y, \dot{y}, \dots) \Big|_{y=\tilde{x}_t},$$

la relación anterior queda dada por

$$\begin{aligned} \frac{\delta}{\delta \tilde{\lambda}^l_\tau} \tilde{J}^k_t &= - J^k_t \cdot J^l_\tau \\ &+ \left\langle \left(j_k(t, y, \dot{y}, \dots) \Big|_{y=\tilde{x}_t} \right) \left(j_l(\tau, x, \dot{x}, \dots) \Big|_{x=\tilde{x}_\tau} \right) \right\rangle_{I,t,\tau} \end{aligned} \quad (5.60)$$

donde la última igualdad corresponde, en términos estadísticos, a la covarianza entre j_k y j_l .

5.5.6. Simulación numérica: Ecuación de movimiento de la partícula libre

Para un sistema en el que se restringe la acción clásica de la partícula libre,

$$A[\tilde{x}] = \int_{t_0}^{t_f} dt \left(\frac{m}{2} \dot{\tilde{x}}_t^2 \right), \quad (5.61)$$

sujeta a las condiciones de borde $\tilde{x}_0 = 0$ y $\tilde{x}_1 = 1$, la solución analítica obtenida al resolver la ecuación de Euler-Lagrange, corresponde a la línea recta

$$\tilde{x}_t = \frac{t_f - t_0}{\tilde{x}_f - \tilde{x}_0} t + \tilde{x}_0 = t.$$

Realizando un muestreo en el espacio de las trayectorias que cumplen con las condiciones de borde y tomando el promedio tiempo a tiempo de las posiciones, se obtuvieron las curvas que se muestran en la figura 5.2 (Puntos Azules). En esta simulación se utilizaron sólo 5 puntos de control, sin ser necesario optimizar para tener una rápida convergencia. La solución numérica converge a la solución analítica (línea verde) luego de tan sólo 10.000 pasos de Monte Carlo.

5.5.7. Simulación numérica: Ecuación de movimiento del oscilador armónico

A continuación se muestran los resultados de una simulación numérica de MaxCal en que se restringe la acción del oscilador armónico, dado por

$$A[\tilde{x}] = \int_{t_0}^{t_f} dt \left(\frac{m}{2} \dot{\tilde{x}}_t^2 - \frac{k}{2} \tilde{x}_t^2 \right), \quad (5.62)$$

con condiciones de borde $\tilde{x}_0 = 0$ y $\tilde{x}_{t_f} = a \sin(\omega t_f)$, donde $t_f < \frac{T}{2}$ y T el período del oscilador armónico. En este primer caso, se considera el tiempo total de las trayectorias simuladas como menor a medio período, obteniéndose una convergencia

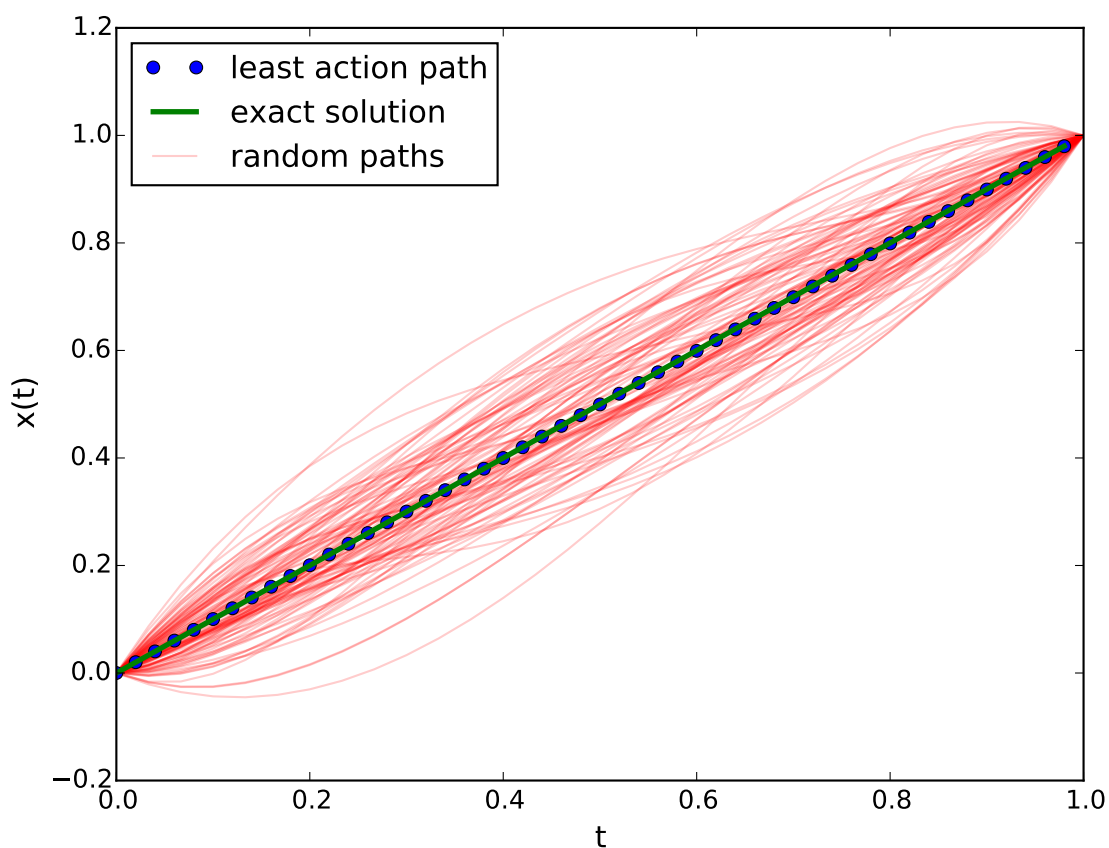


Figura 5.2: Solución numérica de la trayectoria de mínima acción para la partícula libre (Puntos azules) y trayectorias obtenidas con la probabilidad funcional que nos entrega MaxCal.

a la solución analítica en cerca de 20.000 pasos de Monte Carlo, como se muestra en la figura 5.3 (puntos azules).

Para aumentar la complejidad de las simulaciones numéricas del oscilador armónico se utilizarán las condiciones de borde $\tilde{x}_0 = 0$ y $\tilde{x}_{t_f} = a \sin(\omega t_f)$, donde $T > t_f > \frac{T}{2}$, con T el período del oscilador, es decir el tiempo total de cada trayectorias será mayor que medio período y menor que un período completo. Para este caso la solución numérica no converge a la solución analítica esperada (línea verde),

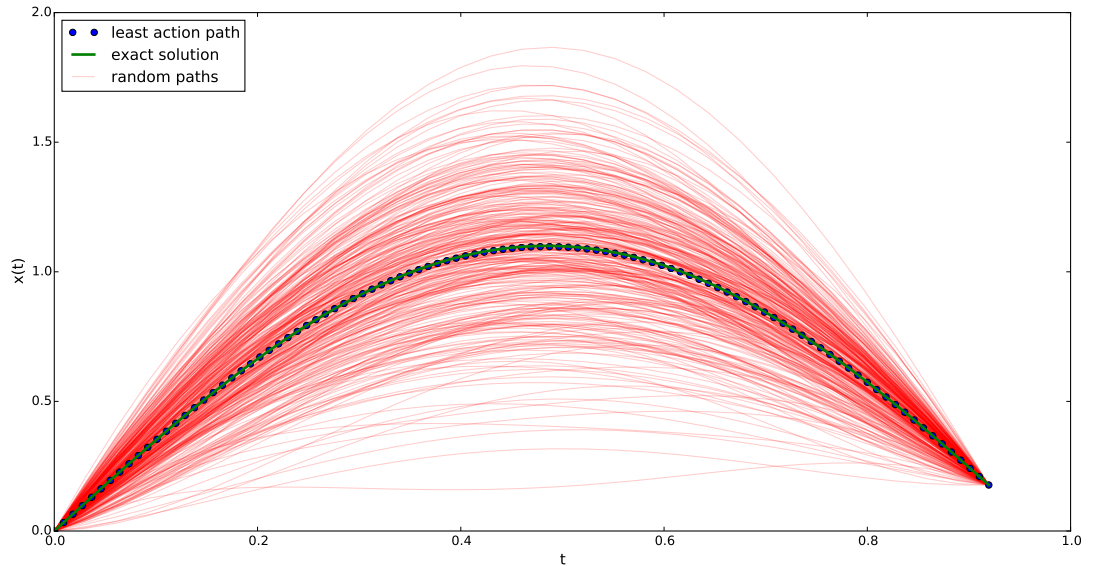


Figura 5.3: Solución numérica de la trayectoria de mínima acción para oscilador armónico, para trayectorias con un tiempo total $t < \frac{T}{2}$.

tal como se muestra en la figura 5.5.7. Al muestrear el espacio con las condiciones de borde nuevas, las trayectorias simuladas con Monte Carlo se alejan cada vez más de la solución esperada tendiendo a los bordes.

El motivo por el cual nuestro sistema no converge, es debido a que el extremo de la acción en general no siempre es un mínimo, ya que además de que la trayectoria analítica encontrada debe cumplir $\frac{\delta A[\tilde{x}]}{\delta \tilde{x}_t} = 0$, se debe cumplir la condición $\frac{\delta^2 A[\tilde{x}]}{\delta \tilde{x}_t^2} > 0$ para confirmar que la trayectoria encontrada corresponde a un mínimo. Para entender este problema con más detalle se recomienda el artículo de Gray y Taylor [66], donde se describe el comportamiento de sistemas tales que la acción no es un mínimo en las trayectorias extremas.

En este trabajo los autores estudian el principio de mínima acción mas allá de la solución analítica encontrada generalmente al utilizar la ecuación de Euler-Lagrange,

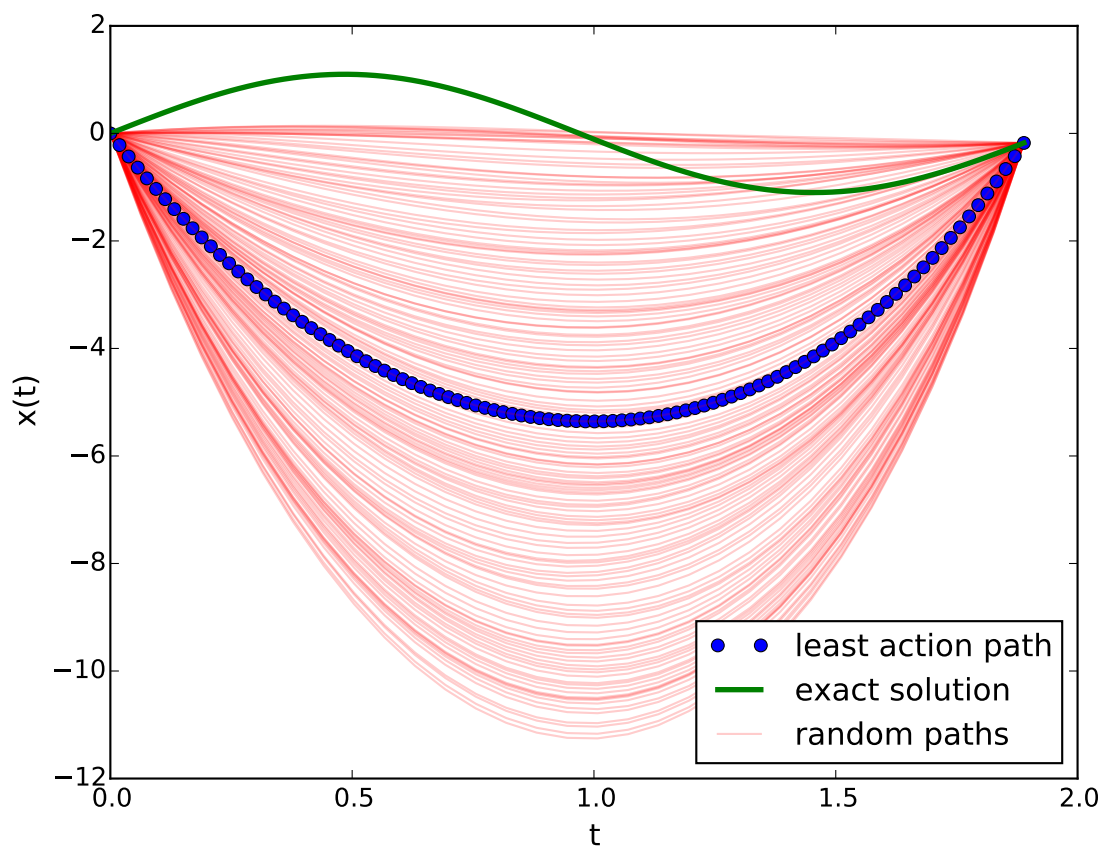


Figura 5.4: Solución numérica no convergente de la trayectoria de mínima acción para oscilador armónico, con trayectorias con un tiempo $\frac{T}{2} < t < T$.

buscando cuáles son las condiciones necesarias para que al utilizar el principio de mínima acción la solución esperada sea la misma que se obtiene de forma analítica cuando se deriva la acción clásica.

Una de los principales puntos mencionados por los autores en este trabajo, es que la acción clásica no siempre es un mínimo cuando es extrema, teniendo comportamientos de mínimo, máximo o punto de ensilladura, dependiendo de las condiciones de borde y la acción que se busca minimizar.

Los autores demuestran que para obtener la solución esperada, es decir la solución obtenida al resolver la ecuación de Euler-Lagrange, es necesario restringir el espacio de las trayectorias posibles en las que se define la acción, considerando sólo las que cumplen ciertas condiciones de borde. Las condiciones de borde necesarias son obtenidas a partir del un concepto de *foco cinético*, el *foco cinético* es un punto (x, t) que tiene la característica de ser común a todas las trayectorias del espacio, sin importar cual es la velocidad inicial o el punto final en que se definen las trayectorias. Es por esto que la forma de encontrar el foco cinético para una cierta acción A es dado por la relación

$$\frac{\partial \tilde{x}(t; v_0)}{\partial v_0} = 0, \quad (5.63)$$

donde \tilde{x} es un extremo de A . Entonces, la forma de forzar a que la acción sea siempre un mínimo, es considerar sólo las trayectorias que cumplen con pasar por el llamado foco cinético, además de las condiciones de borde que definen el problema. De manera consistente con estas condiciones, al realizar el algoritmo de Monte Carlo considerando las trayectorias que cumplen las condiciones de borde y pasan por los focos cinéticos, es que la solución numérica logra converger.

Para mostrar esto, se presenta una simulación de un oscilador armónico con condiciones de borde $\tilde{x}_0 = 0$ y $\tilde{x}_{t_f} = a \sin(\omega t_f)$, donde $\frac{3T}{2} > t_f > T$. Aquí los *focos cinéticos* son los puntos $\tilde{x}_{\frac{T}{2}} = 0$ y $\tilde{x}_T = 0$. Entonces, realizando la simulación con las condiciones de borde mencionadas y restringiendo las trayectorias a pasar por el foco cinético⁷ se logra una convergencia cercana a la solución esperada luego de 50.000 pasos de Monte Carlo (los cuales tardan más cada uno en ser evaluados), como se muestra en la figura 5.5.

⁷Obligar a las trayectorias a pasar por los focos cinéticos puede ser realizado de 2 formas: agregarlo como una condición en la simulación del MCM para trayectorias, o incluirlo como una restricción adicional en el modelo de probabilidad.

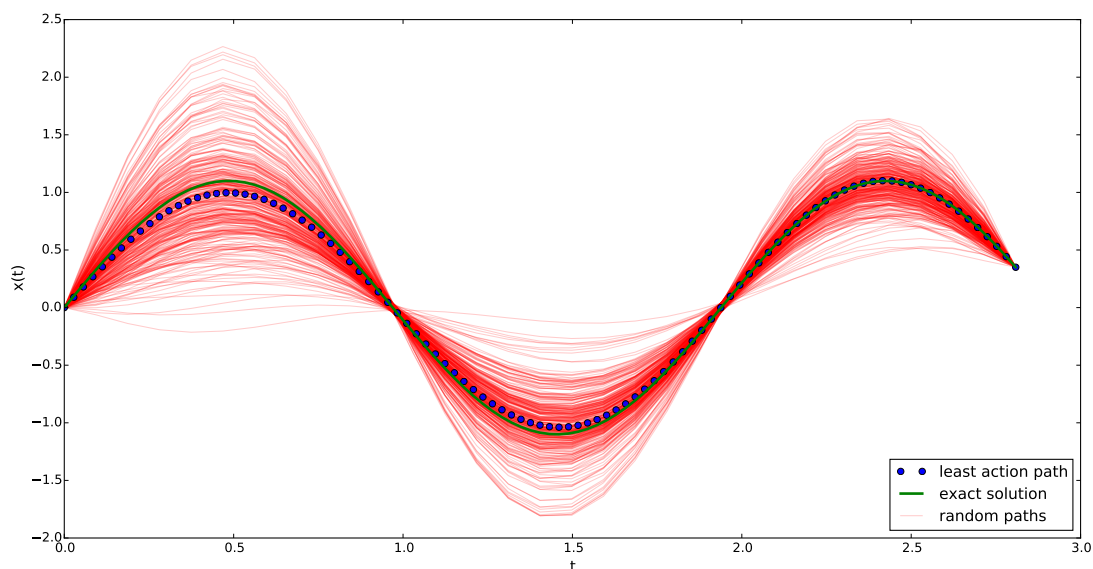


Figura 5.5: Solución numérica convergente de la trayectoria de mínima acción para oscilador armónico, con trayectorias con un tiempo total $T < t < \frac{3T}{2}$.

En conclusión, MCM para trayectorias parece ser un método viable para resolver la dinámica de sistemas físicos clásicos, esta metodología requiere resolver el problema de determinar cuáles son los focos cinéticos del sistema en cuestión, lo que en principio puede ser resuelto de forma numérica mediante (5.63). Existe un gran potencial de poder extender el método a sistemas con más grados de libertad y potenciales de interacción más complejos que el potencial armónico. Al ser MCM para trayectorias una metodología fácil de implementar, se abren grandes oportunidades de extender esta metodología a otras variantes existentes de MCM para así mejorar su convergencia.

Capítulo 6

Mecánica Estadística Fuera del Equilibrio

En este capítulo se mostrará como derivar los principales elementos que componen la mecánica estadística fuera del equilibrio (NESM¹), utilizando inferencia sobre el espacio de trayectorias y MaxCal, es decir, considerando el hecho que existe una densidad de probabilidad funcional para las trayectorias y utilizando además a MaxCal como una herramienta para entender la forma que tiene esta densidad de probabilidad funcional, con la cual se calculan las expectativas.

Es bueno destacar que, aún cuando MaxCal es una herramienta útil para entender los sistemas fuera del equilibrio mediante las restricciones tiempo dependientes, en estricto rigor no es completamente necesaria para entender la estadística fuera del equilibrio y su estructura, sino que más bien nos entrega un método sistemático para obtener la densidad de probabilidad a partir de información externa, entregando así una primera aproximación a las restricciones que gobiernan los sistemas fuera del equilibrio.

La primera parte de este capítulo trata sobre inferencia en el espacio de trayectorias, donde se postulará un axioma que permite reconstruir la base de la NESM.

¹Por su nombre en inglés *Non-Equilibrium Statistical Mechanics*.

Este axioma lo llamaremos axioma fundamental de la NESM y nos entrega una relación entre las expectativas que utilizan una densidad de probabilidad fuera del equilibrio $P(x|t, I)$, con expectativas que utilizan el funcional de probabilidad para trayectorias $P[\tilde{x}|I]$. Utilizando este axioma fundamental de la NESM se mostrará cómo derivar una expresión que llamaremos la *ecuación del corte temporal*, la cual asocia de forma directa una densidad de probabilidad funcional con una densidad de probabilidad tiempo dependiente.

Con el axioma fundamental de la NESM y la ecuación del corte temporal, es posible avanzar para completar lo que llamaremos la hipótesis de Jaynes, la cual postula que al igual que la ME puede ser derivada a partir MaxEnt utilizando restricciones en equilibrio, la NESM puede ser derivada a partir de MaxCal utilizando restricciones fuera del equilibrio.

Luego de sentar las bases teóricas mencionadas previamente, para avanzar con el entendimiento de la NESM desde la inferencia de trayectorias se procederá a derivar gran parte de los principales elementos de la NESM que ayudan a entender las probabilidades tiempo-dependientes. Entre estos se encuentran la ecuación de Fokker-Planck, la ecuación de Chapman-Kolmogorov y la ecuación de continuidad.

Continuando con los fundamentos de la NESM, donde las relaciones mencionadas previamente se derivan sólo utilizando inferencia sobre trayectorias, se procederá a mostrar cómo MaxCal permite entender restricciones que rigen a los sistemas fuera del equilibrio, sistemas descritos por ecuaciones diferenciales en derivadas parciales (PDE's) tales como la ecuación de Fokker-Planck, la ecuación de Liouville y la ecuación de difusión. Finalmente se mostrará como entender dos relaciones fuera del equilibrio muy empleadas en la práctica, la igualdad de Jarzynski [21] y el teorema de fluctuación de Crooks [23], los cuales son relaciones de validez general relacio-

nados al fenómeno de la irreversibilidad. De esta manera obtenemos una primera aproximación al entendimiento de procesos irreversibles y la termodinámica fuera del equilibrio desde el formalismo mostrado en esta tesis.

6.1. Inferencia sobre trayectorias

Definiremos la inferencia de trayectorias como la inferencia cuando se supone la existencia de un espacio de trayectorias \mathbb{X} , espacio sobre el cual es posible asignar una probabilidad para sus elementos \tilde{x} dado por $P[\tilde{x}|I] : \mathbb{X} \rightarrow \mathbb{R}$ con $0 \leq P[\tilde{x}|I]$, conocida como la densidad de probabilidad funcional.

En este nivel es necesario conocer la forma explícita de esta densidad de probabilidad funcional para entender cómo se relacionan cantidades macroscópicas en el espacio de funcionales, ya que al existir una densidad de probabilidad funcional podemos definir la operación expectación en el espacio de trayectorias (2.7). Gracias a esta operación y junto a sus propiedades es que podemos derivar el “esqueleto” que da forma a la NESM.

Como se explicó en el cap. 5, MaxCal fue postulado como un principio análogo a MaxEnt, sólo que definido para sistemas fuera del equilibrio. Por medio de MaxCal es posible obtener la densidad de probabilidad funcional, la cual asigna probabilidades a funciones del tiempo, que en esta tesis corresponden a trayectorias del sistema físico. La hipótesis de Jaynes sobre MaxCal no ha sido completada en la práctica, ya que a pesar de que el autor formalizó la estructura de la teoría, proporcionando un método sistemático para obtener la forma de la densidad de probabilidad funcional, nunca explicó cómo ir de una representación de trayectorias $P[\tilde{x}|I]$ obtenida a partir de MaxCal a una representación instantánea $P(x|t, I)$, correspondiente a la formulación tradicional de la NESM donde el elemento central es la densidad de

probabilidad dependiente del tiempo y los estados.

En el contexto de inferencia sobre trayectorias, primero se mostrará el que consideraremos el axioma fundamental para la NESM, con el cual entenderemos el origen de la forma de las ecuaciones diferenciales en derivadas parciales que aparecen en esta teoría.

6.1.1. Axioma fundamental de la NESM

Nota: Para lo que sigue en este capítulo, acortaremos la notación de modo que,

$$G(x, t) \Big|_{x=\tilde{x}_t} = G(\tilde{x}_t, t), \quad (6.1)$$

recordando que x representa la variable con la cual se opera y \tilde{x}_t es el valor de la posición para la trayectoria $\tilde{x} \in \mathbb{X}$ en un tiempo t , valor que debe ser evaluado luego de realizar cualquier operación.

Cuando realizamos inferencia sobre un espacio de trayectorias, como se ha mostrado en el capítulo 5, es posible tomar expectativas de funciones $G(x, t)$, utilizando el funcional de probabilidad de la forma

$$\begin{aligned} \langle G(\tilde{x}_t, t) \rangle_I &= \int d\tilde{x} P[\tilde{x}|I] G(\tilde{x}_t, t). \\ &= \tilde{g}_t. \\ &= g(t). \end{aligned} \quad (6.2)$$

Donde G es una función dependiente del tiempo y la posición $x = \tilde{x}_t \in \mathbb{R}^n$. Luego de evaluar la expectativa de G , se obtiene una función dependiente del tiempo \tilde{g} .

Por otra parte, en la NESM se define la operación expectación por medio de una función de probabilidad fuera del equilibrio $P(x|t, I)$, de modo que podemos definir la expectación de la misma función G anterior como,

$$\begin{aligned} \langle G(x, t) \rangle_I &= \int dx P(x|t, I) G(x, t) \\ &= \tilde{g}_t \\ &= g(t). \end{aligned} \tag{6.3}$$

De esta forma, podemos ver que existen dos caminos alternativos para obtener propiedades macroscópicas fuera del equilibrio, las cuales sólo se diferencian en el uso de un funcional de probabilidad o una densidad de probabilidad fuera del equilibrio para calcularlos. Al requerir la consistencia interna del formalismo se establece lo que llamaremos el axioma fundamental de la NESM, donde de forma consistente, se postula que el calcular expectativas de funciones tiempo-dependientes utilizando el funcional de probabilidad, (6.2) debe ser equivalente a calcular expectativas de funciones tiempo dependientes utilizando la densidad de probabilidad fuera del equilibrio (6.3). Entonces, igualando (6.2) con (6.3), se obtiene

$$\langle G(t, x) \rangle_I = \langle G(t, \tilde{x}_t) \rangle_I, \tag{6.4}$$

es decir,

$$\int dx P(x|t, I) G(x, t) = \int d\tilde{x} P[\tilde{x}|I] G(\tilde{x}_t, t), \tag{6.5}$$

para todo $G(x, t)$. Llamaremos a esta relación el axioma fundamental de la NESM, donde en el primer caso, queda escrito en forma de expectación (6.4) y en el segundo caso queda escrito en su forma integral (6.5). Este axioma impone que se debe obtener las mismas propiedades macroscópicas tiempo dependientes \tilde{g}_t , utilizando tanto la probabilidad fuera del equilibrio $P(x|t, I)$ como la probabilidad funcional $P[\tilde{x}|I]$ al realizar la operación expectación.

6.1.2. Ecuación del Corte temporal

Para lograr obtener una densidad de probabilidad tiempo-dependiente $P(x|t, I)$ mediante el uso de la densidad de probabilidad funcional $P[\tilde{x}|I]$ obtenido desde MaxCal, haremos uso de la relación que llamaremos ecuación del corte temporal [67],

$$P(y|t, I) = \int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_t - y). \quad (6.6)$$

Escrito en forma de expectativas, la ecuación del corte temporal queda como,

$$P(y|t, I) = \left\langle \delta(\tilde{x}_t - y) \right\rangle_I, \quad (6.7)$$

donde $P(y|t, I)$ es la probabilidad de tener un micro-estado y en un tiempo t , $P[\tilde{x}|I]$ es la probabilidad de una micro-trayectoria \tilde{x} y la delta de Dirac $\delta(\tilde{x}_t - y)$ es la encargada de sumar sólo las trayectorias \tilde{x} que cumplen $\tilde{x}_t = y$. Como se indica en la expectativa de la derecha, toda la dependencia con respecto a y y t , se encuentra contenida en la delta de Dirac.

El origen de la ecuación del corte temporal puede ser entendida de la siguiente forma, **la probabilidad de estar en un estado y en un tiempo t , es la suma de las probabilidades de los caminos \tilde{x} que en el tiempo t pasan por el estado y** , es decir, la suma de todas las probabilidades de los caminos que cumplen $\tilde{x}_t = y$, como se muestra en la figura 6.1

Demostración:

Utilizando el Axioma Fundamental de la NESM (6.5), con una función de prueba $G(t, x) = \delta(x - y)$ se obtiene

$$\int dx P(x|t, I) \delta(x - y) = \int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_t - y). \quad (6.8)$$

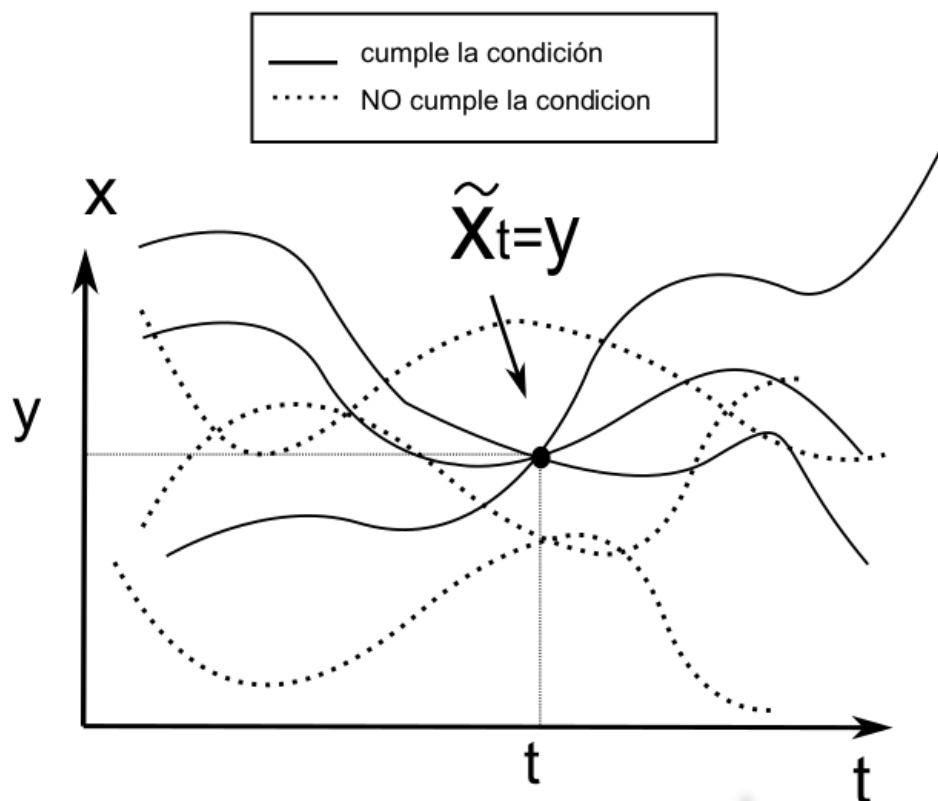


Figura 6.1: Esquema que ejemplifica la ecuación del corte temporal en el espacio de trayectorias \tilde{x} , donde la delta de Dirac es no nula cuando se cumple la condición $\tilde{x}_t = y$.

La integral del lado izquierdo puede ser resuelta utilizando la propiedad de la delta de Dirac, obteniendo la forma integral de la ecuación del corte temporal,

$$P(y|t, I) = \int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_t - y) \blacksquare \quad (6.9)$$

La ecuación del corte temporal permite relacionar la inferencia de trayectorias con el estudio de sistemas fuera del equilibrio, y su importancia se encuentra en poder relacionar la densidad de probabilidad funcional obtenida a partir de un principio

variacional como MaxCal con una densidad de probabilidad tiempo-dependiente para los estados, correspondiente al elemento central en la NESM.

Un punto a remarcar de la ecuación del corte temporal es que corresponde a una identidad matemática, la cual no requiere principios físicos para ser válida, siendo la única suposición la existencia de trayectorias $\tilde{x} \in \mathbb{X}$. Es por esto que la ecuación del corte temporal emerge como una pieza clave en el entendimiento de la NESM a partir del paradigma presentado en esta tesis, ya que es a partir de esta ecuación que se avanza en completar lo que hemos llamado la hipótesis de Jaynes.

Para profundizar en la ecuación del corte temporal con un contexto más general, se recomienda al lector la referencia [67, 68].

6.1.3. NESM desde inferencia de trayectorias: La hipótesis de Jaynes

El lograr establecer una relación entre la densidad de probabilidad funcional y la función de probabilidad tiempo-dependiente nos permite unir la idea de inferencia sobre trayectorias con la NESM, esta relación puede ser vista utilizando la llamada Ecuación del Corte temporal [67].

La Ecuación del Corte Temporal permite avanzar en gran medida la labor de demostrar la hipótesis planteada por E. T. Jaynes, ya que *para obtener la función de probabilidad de la NESM $P(y|t, I)$ por medio de restricciones tiempo-dependientes, es necesario resolver una integral de caminos en el lado derecho de (6.7), donde la forma de la densidad de probabilidad funcional $P[\tilde{x}|I]$, es obtenida mediante MaxCal (cap. 5).*

Por medio del formalismo presentado en esta tesis es posible notar que, en su for-

ma fundamental, la ME y la NESM son teorías análogas, donde quedan bien definidas cada una de sus partes, que llevan a entender teorías macroscópicas la termodinámica y la termodinámica fuera del equilibrio. Para el caso de la ME dadas las condiciones en equilibrio del sistema, utilizando MaxEnt se obtiene la forma de la densidad de probabilidad que describe el sistema, con la cual es posible calcular las propiedades macroscópicas de los sistemas y entender así como éstas se relacionan (Termodinámica). Por otra parte, en el caso de la NESM dadas restricciones tiempo-dependientes y utilizando MaxCal, se obtiene la forma de la densidad de probabilidad funcional que describe el sistema sujeto a las restricciones, y es por medio de la ecuación del corte temporal que pasamos de una representación de trayectorias a una representación instantánea, obteniendo así la forma de la densidad de probabilidad utilizada en la NESM. Con esta densidad podemos encontrar las propiedades macroscópicas fuera del equilibrio y cómo éstas se relacionan (Termodinámica fuera del equilibrio).

Es bueno destacar que al ser una integral de trayectorias y una ecuación diferencial en derivadas parciales formas análogas de resolver el mismo problema, es posible entender por qué la formulación usual de NESM involucra resolver ecuaciones diferenciales en derivadas parciales para obtener la densidad de probabilidad dependiente del tiempo, en lugar de una integral de trayectorias como se ha propuesto en esta tesis con la ecuación del corte temporal (figura 1.5.)

El entender que la probabilidad de estar en un punto del espacio de configuraciones en un tiempo es la suma de las probabilidades de los caminos que pasan por ese punto en el tiempo (ecuación del corte temporal), no es una idea nueva y ha sido entendida por autores como N. Wiener, R. Feynman [30,59] M. Kac [29] y otros [69], permitiendo relacionar integrales de caminos con ecuaciones diferenciales en derivadas parciales, procedimiento conocido como la fórmula de Feynman-Kac [29,30].

El uso más conocido de esta interpretación es el utilizado por R. Feynman la cual entrega una formulación alternativa de la mecánica cuántica basada en integrales de trayectorias [30]. Esta formulación es equivalente a resolver la ecuación de Schrödinger, que corresponde a la formulación tradicional de la mecánica cuántica [70].

6.1.4. Expectaciones Cortadas

Como se explicó en el capítulo de teoría de la probabilidad (2.8), la delta de Dirac puede ser escrita como una probabilidad y en particular como una probabilidad Bayesiana. Esta forma de ver las probabilidades puede ser entendida de mejor manera, comparando la ec. del corte temporal (6.7), con la regla de la marginalización (2.11), ya que, ambas son iguales a $P(y|t, I)$. Tomando como punto de partida la ecuación del corte temporal,

$$P(y|t, I) = \int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_t - y), \quad (6.10)$$

y empleando la regla de la marginalización (2.11),

$$P(y|t, I) = \int d\tilde{x} P[\tilde{x}, y, t|I],$$

podemos escribir

$$\int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_t - y) = \int d\tilde{x} P[\tilde{x}, y, t|I]. \quad (6.11)$$

Entonces, utilizando la regla del producto (2.5) en la probabilidad del lado derecho, se tiene que

$$\int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_t - y) = \int d\tilde{x} P[\tilde{x}|I] P(y, t|\tilde{x}, I). \quad (6.12)$$

Luego, como esta igualdad es válida para cualquier elección de $P[\tilde{x}|I]$, al comparar ambos lados podemos concluir que

$$P(y, t|\tilde{x}, I) = \delta(\tilde{x}_t - y). \quad (6.13)$$

Esta igualdad nos ayuda a entender cómo opera la delta en las expectativas de trayectorias, ya que si tomamos la expectativa de $\delta(\tilde{x}_t - y) G(\tilde{x}_t, t)$, podemos reescribir la delta como una probabilidad de la forma,

$$\left\langle \delta(\tilde{x}_t - y) G[\tilde{x}; t] \right\rangle_I = \int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_t - y) G[\tilde{x}; t]. \quad (6.14)$$

$$= \int d\tilde{x} P[\tilde{x}|I] P(y, t|\tilde{x}, I) G[\tilde{x}; t]. \quad (6.15)$$

Entonces, usando el teorema de Bayes (2.6) en el lado derecho, de modo que

$$P(y, t|\tilde{x}, I) = \frac{P(y|t, I) P[\tilde{x}|y, t, I]}{P[\tilde{x}|I]}, \quad (6.16)$$

se obtiene entonces que,

$$\left\langle \delta(\tilde{x}_t - y) G[\tilde{x}; t] \right\rangle_I = \int d\tilde{x} P[\tilde{x}|I] \frac{P(y|t, I) P[\tilde{x}|y, t, I]}{P[\tilde{x}|I]} G[\tilde{x}; t]. \quad (6.17)$$

$$= \int d\tilde{x} P(y|t, I) P[\tilde{x}|y, t, I] G[\tilde{x}; t]. \quad (6.18)$$

Sacando fuera de la integral el factor $P(y|t, I)$, que no depende de la variable de integración, se obtiene una relación que nos indica cómo opera una delta de Dirac para expectativas de trayectorias

$$\left\langle \delta(\tilde{x}_t - y) G[\tilde{x}; t] \right\rangle_I = P(y|t, I) \int d\tilde{x} P[\tilde{x}|y, t, I] G[\tilde{x}; t], \quad (6.19)$$

donde el lado derecho puede ser a su vez escrito como una expectativa, indicando en el sub-índice de la expectativa, como es usual, los parámetros de los cuales depende la probabilidad utilizada al realizar la expectativa. Es decir, finalmente obtenemos

$$\left\langle \delta(\tilde{x}_t - y) G[\tilde{x}; t] \right\rangle_I = P(y|t, I) \left\langle G[\tilde{x}; t] \right\rangle_{I, y, t}. \quad (6.20)$$

A esta forma de operar con las deltas de Dirac en expectativas de trayectorias la llamaremos *regla de las expectativas cortadas*, y será utilizada como una alternativa

reducida al uso de la regla de la marginalización y el teorema de Bayes cuando se tiene una delta de Dirac escrita como probabilidad bayesiana de acuerdo a (6.13).

Para el caso particular en que el funcional $G[\tilde{x}; t]$ es una función que sólo depende de la posición, de modo que $G[\tilde{x}; t]$ pueda ser escrito como $G(\tilde{x}_t, t)$, la regla de expectativas cortadas queda como

$$\left\langle \delta(\tilde{x}_t - y) G(\tilde{x}_t, t) \right\rangle_I = P(y|t, I) G(y, t). \quad (6.21)$$

Este resultado es directo de demostrar mediante el uso del axioma fundamental de la NESM. Utilizando una función de prueba $G(x, t) \delta(x - y)$ y desarrollando la expectativa tomada con la probabilidad instantánea (lado derecho), tenemos

$$\begin{aligned} \left\langle \delta(\tilde{x}_t - y) G(\tilde{x}_t, t) \right\rangle_I &= \left\langle \delta(\tilde{x}_t - y) G(x, t) \right\rangle_I \\ &= \int dx P(x|t, I) \delta(x - y) G(x, t). \\ &= P(y|t, I) G(y, t) \blacksquare \end{aligned} \quad (6.22)$$

Por otra parte, cuando el funcional depende de derivadas temporales de la posición, como es el caso de un lagrangiano $G[\tilde{x}; t] = \mathcal{L}(\tilde{x}_t, \dot{\tilde{x}}_t, t)$, la expectativa cortada, no puede relacionarse directamente al axioma fundamental de la NESM, siendo sólo posible expresarla como,

$$\left\langle \delta(\tilde{x}_t - y) G[\tilde{x}; t] \right\rangle_I = P(y|t, I) \left\langle \mathcal{L}(\tilde{x}_t, \dot{\tilde{x}}_t, t) \right\rangle_{I, y, t}. \quad (6.23)$$

Sin pérdida de generalidad, lo mismo ocurre para funciones dependientes de derivadas superiores de la posición.

6.1.5. Ecuación de Chapman-Kolmogorov en el espacio de trayectorias

La ecuación de Chapman-Kolmogorov es una ecuación integral utilizada para evolucionar una probabilidad en el tiempo, partiendo de $P(x_0|t_0, I)$ en un tiempo t_0

a una probabilidad posterior $P(x_1|t_1, I)$ en un tiempo t_1 , mediante el uso de una probabilidad de transición $P(x_1|x_0, t_0, t_1, I)$ quedando escrita como

$$P(x_1|t_1, I) = \int dx_0 P(x_1|x_0, t_0, t_1, I) P(x_0|t_0, I). \quad (6.24)$$

Desde las leyes de la probabilidad la ecuación de Chapman-Kolmogorov emerge como consecuencia de la regla de marginalización y la regla del producto como se muestra en apéndice B. En el contexto de inferencia de trayectorias podemos relacionar la probabilidad de transición $P(x_1|x_0, t_0, t_1, I)$ en la ecuación de Chapman-Kolmogorov (B.2), con la densidad de probabilidad funcional de la forma,

$$P(x_1|t_1, I) = \int dx_0 P(x_0|t_0, I) \int d\tilde{x} \frac{P[\tilde{x}|I] \delta(\tilde{x}_{t_0} - x_0) \delta(\tilde{x}_{t_1} - x_1)}{P(x_0|t_0, I)}, \quad (6.25)$$

donde comparando con (6.24), podemos notar que la probabilidad de transición queda escrita de la forma

$$P(x_1|x_0, t_0, t_1, I) = \int d\tilde{x} \frac{P[\tilde{x}|I] \delta(\tilde{x}_{t_0} - x_0) \delta(\tilde{x}_{t_1} - x_1)}{P(x_0|t_0, I)}. \quad (6.26)$$

En esta ecuación las deltas de Dirac se encargan de seleccionar las trayectorias \tilde{x} que cumplen, $\tilde{x}_{t_0} = x_0$ y $\tilde{x}_{t_1} = x_1$. Para el caso de la normalización (denominador del lado derecho), sólo se consideran las trayectorias que cumplen $\tilde{x}_{t_0} = x_0$,

$$P(x_0|t_0, I) = \int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_{t_0} - x_0). \quad (6.27)$$

Una forma reducida de escribir la ecuación de Chapman-Kolmogorov en el espacio de trayectorias, similar a la formulación tradicional [20] se obtiene al escribir la delta en el numerador como una probabilidad condicional (como se demostró en (6.13), donde $P(x_0|t_0, \tilde{x}, I) = \delta(\tilde{x}_{t_0} - x_0)$, obteniendo así una realización del Teorema de Bayes (2.6),

$$P[\tilde{x}|x_0, t_0, I] = \frac{P[\tilde{x}|I] P(x_0|t_0, \tilde{x}, I)}{P(x_0|t_0, I)}. \quad (6.28)$$

De esta forma, se obtiene una nueva manera de escribir la ecuación de Chapman-Kolmogorov (B.2),

$$P(x_1|t_1, I) = \int dx_0 P(x_0|t_0, I) \left[\int d\tilde{x} P[\tilde{x}|x_0, t_0, I] \delta(\tilde{x}_{t_1} - x_1) \right], \quad (6.29)$$

con una probabilidad de transición de la forma

$$P(x_1|x_0, t_0, t_1, I) = \int d\tilde{x} P[\tilde{x}|x_0, t_0, I] \delta(\tilde{x}_{t_1} - x_1). \quad (6.30)$$

Como se mencionó al comienzo de esta sección, la ecuación de Chapman-Kolmogorov puede ser utilizada para evolucionar la probabilidad en el tiempo, para lo cual es necesario plantear la forma inicial de la probabilidad $P(x_0|t_0, I)$. Entonces, resolviendo las integrales en (6.24), se obtiene la probabilidad posterior $P(x_1|t_1, I)$. Este procedimiento se puede realizar de forma iterativa, permitiéndonos obtener la dinámica de una densidad de probabilidad, es decir, utilizando $P(x_k|I)$ podemos encontrar la probabilidad posterior $P(x_{k+1}|I)$ mediante el uso de la probabilidad de transición $P(x_{k+1}|x_k, I)$, siendo posible encontrar la probabilidad n -ésima $P(x_n|I)$ partiendo desde una probabilidad inicial $P(x_0|I)$.

De igual manera se puede utilizar la ecuación de Chapman-Kolmogorov para recuperar la probabilidad previa $P(x_0|t_0, I)$ cuando se conoce la probabilidad posterior $P(x_1|t_1, I)$. La primera puede ser obtenida con una ecuación de Chapman-Kolmogorov con probabilidades de transición inversas, análoga a (B.5) pero definida en el espacio de trayectorias,

$$P(x_0|t_0, I) = \int dx_1 P(x_1|t_1, I) \int d\tilde{x} \frac{P[\tilde{x}|I] \delta(\tilde{x}_{t_0} - x_0) \delta(\tilde{x}_{t_1} - x_1)}{P(x_1|t_1, I)}, \quad (6.31)$$

la cual podemos reescribir como

$$P(x_0|t_0, I) = \int dx_1 P(x_1|t_1, I) \int d\tilde{x} P[\tilde{x}|x_1, t_1, I] \delta(\tilde{x}_{t_0} - x_0), \quad (6.32)$$

donde la probabilidad de transición está dada por

$$P(x_0|x_1, t_1, t_0, I) = \int d\tilde{x} \frac{P[\tilde{x}|I]\delta(\tilde{x}_{t_0} - x_0)\delta(\tilde{x}_{t_1} - x_1)}{P(x_1|t_1, I)}. \quad (6.33)$$

$$= \int d\tilde{x} P[\tilde{x}|x_1, t_1, I]\delta(\tilde{x}_{t_0} - x_0). \quad (6.34)$$

Vemos que es posible encontrar la $(-k)$ -ésima probabilidad previa $P(x_{-k}|I)$ dada la probabilidad posterior $P(x_0|I)$, mediante el uso de las probabilidades de transición $P(x_{-k}|x_{-k-1}, I)$. Es bueno destacar que la información dada en el funcional de probabilidad (correspondiente a la probabilidad de transición) es distinta en cada caso, ya que allí es donde se encuentra contenida la información que indica la dirección temporal en que se realiza la inferencia.

A continuación, se procederá a demostrar la ecuación de Chapman-Kolmogorov en el contexto de inferencia de trayectorias, mostrando cómo expresar dicha ecuación utilizando una densidad de probabilidad funcional, la cual puede ser obtenida utilizando MaxCal.

Demostración de la Ec. Chapman-Kolmogorov en el espacio de trayectorias:

Utilizando la ecuación del corte temporal en su forma integral (6.7) y multiplicando por $1 = \int dx_0 \delta(\tilde{x}_{t_0} - x_0)$,

$$P(x_1|t_1, I) = \int d\tilde{x} P[\tilde{x}|I]\delta(\tilde{x}_{t_1} - x_1). \quad (6.35)$$

$$= \int d\tilde{x} P[\tilde{x}|I]\delta(\tilde{x}_{t_1} - x_1) \int dx_0 \delta(\tilde{x}_{t_0} - x_0). \quad (6.36)$$

Luego, reescribiendo y separando la integral $\int dx_0$,

$$P(x_1|t_1, I) = \int dx_0 \int d\tilde{x} P[\tilde{x}|I]\delta(\tilde{x}_{t_1} - x_1)\delta(\tilde{x}_{t_0} - x_0). \quad (6.37)$$

Al agregar $1 = \frac{P(x_0|t_0, I)}{P(x_0|t_0, I)}$ en el interior de la integral y reagrupando los términos

se obtiene la forma de la ecuación de Chapman-Kolmogorov,

$$\begin{aligned}
P(x_1|t_1, I) &= \int dx_0 \frac{P(x_0|t_0, I)}{P(x_0|t_0, I)} \int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_{t_1} - x_1) \delta(\tilde{x}_{t_0} - x_0). \\
&= \int dx_0 P(x_0|t_0, I) \\
&\times \left[\int d\tilde{x} \frac{P[\tilde{x}|I] \delta(\tilde{x}_{t_1} - x_1) \delta(\tilde{x}_{t_0} - x_0)}{P(x_0|t_0, I)} \right] \blacksquare
\end{aligned} \tag{6.38}$$

Podemos notar de forma directa, que al comparar con (B.1) el término en corchetes que acompaña al operador $\int dx_0 P(x_0|t_0, I)$ corresponde a una probabilidad de transición, dado por

$$P(x_1, t_1|x_0, t_0, I) = \frac{\int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_{t_1} - x_1) \delta(\tilde{x}_{t_0} - x_0)}{\int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_{t_0} - x_0)}, \tag{6.39}$$

que corresponde a la forma correcta de la probabilidad de transición para la ecuación de Chapman-Kolmogorov en términos de la densidad de probabilidad funcional.

Es bueno recordar que la densidad de probabilidad funcional $P[\tilde{x}|I]$ apropiada para un sistema puede ser obtenida utilizando MaxCal, por lo que distintas restricciones llevarán a diferentes probabilidades de transición y por tanto a diferentes dinámicas. Tal vez pueda ser confuso para el lector el hecho que la probabilidad de transición depende de la probabilidad inicial, pero esto es una consecuencia natural de la notación, la que permite ver de forma explícita la dependencia. Esto ocurre como consecuencia directa de escribir cualquier probabilidad condicional $P(y|x, I)$ usando el teorema de Bayes (2.6) como

$$P(y|x, I) = \frac{P(x, y|I)}{P(x|I)},$$

donde ahora $P(y|x, I)$ depende de $P(x|I)$.

6.1.6. Ecuación de continuidad

La ecuación de continuidad relaciona el cambio temporal de la probabilidad con su cambio espacial y es escrita de forma tradicional [71] para un sistema de 3-dimensiones como,

$$\frac{\partial}{\partial t}P + \nabla \cdot J = 0, \quad (6.40)$$

con J conocida como la densidad de corriente de probabilidad, la cual puede ser escrita como $P v$ donde $v = (v_x, v_y, v_z)$ es un vector de 3-dimensiones asociado a una velocidad y P un escalar asociado a una probabilidad. En esta tesis se trabajará como ejemplo la ecuación de continuidad escrita en 1-dimensión para así clarificar el procedimiento que será presentado

$$\frac{\partial}{\partial t}P + \frac{\partial}{\partial x}(P v) = 0, \quad (6.41)$$

donde la densidad de corriente de probabilidad $J = P v$ será compuesta por $v : v(y, t)$ llamada velocidad de flujo de probabilidad (en analogía con fluidos) y $P : P(y|t, I)$ la densidad de probabilidad. Sin embargo, la extensión a mayor número de dimensiones es directa.

En física una ecuación de continuidad expresa una ley de conservación local de forma matemática, ya sea de forma integral como de forma diferencial. Esta ecuación puede ser vista en las principales teorías físicas expresando la conservación de propiedades tales como, conservación de la carga en la teoría electromagnética, conservación de la masa en la mecánica de fluidos, conservación de la probabilidad en mecánica cuántica y conservación de la energía y momenta en la relatividad general.

Cuando se habla de leyes de conservación se hace referencia a propiedades que se mantienen constantes² durante la evolución de los sistemas físicos, entonces la importancia de las leyes de conservación es que son una guía para entender las propiedades conservadas y las condiciones en las que ocurre su conservación.

En el contexto de la inferencia de trayectorias, se puede obtener una ecuación de continuidad a partir de la ecuación del corte temporal, de la forma

$$0 = \frac{\partial}{\partial t} P(y|t, I) + \frac{\partial}{\partial y} J(y, t). \quad (6.42)$$

$$= \frac{\partial}{\partial t} P(y|t, I) + \frac{\partial}{\partial y} \left(P(y|t, I) \int d\tilde{x} P[\tilde{x}|y, t, I] \dot{\tilde{x}} \right), \quad (6.43)$$

la cual escrita en forma de expectación queda como

$$0 = \frac{\partial}{\partial t} P(y|t, I) + \frac{\partial}{\partial y} \left(P(y|t, I) \left\langle \dot{\tilde{x}} \right\rangle_{I, y, t} \right). \quad (6.44)$$

Entonces, la velocidad de flujo de probabilidad escrita para inferencia de trayectorias puede ser vista como una integral de trayectorias, mas precisamente una expectación,

$$v(y, t) = \int d\tilde{x} P[\tilde{x}|y, t, I] \dot{\tilde{x}}. \quad (6.45)$$

$$= \left\langle \dot{\tilde{x}} \right\rangle_{I, y, t}. \quad (6.46)$$

De la misma forma, la densidad de corriente de probabilidad $J = P v$ en el contexto de inferencia sobre trayectorias puede ser escrita como,

$$P v = \int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_t - y) \tilde{x}_t. \quad (6.47)$$

$$= \left\langle \delta(\tilde{x}_t - y) \tilde{x}_t \right\rangle_I. \quad (6.48)$$

²Constantes de forma global, pero no constantes localmente: existe solo un flujo J compatible con la conservación local.

En esta tesis, la ecuación de continuidad será utilizada para construir las diferentes ecuaciones diferenciales en derivadas parciales que componen la NESM, encontrando la forma de la velocidad de flujo de probabilidad v para cada sistema físico por medio de diferentes técnicas que serán vistas mas adelante. Para ejemplificar lo dicho anteriormente, si v es de la forma

$$v = \frac{-D}{P} \frac{\partial}{\partial x} P = -D \frac{\partial}{\partial x} \log P, \quad (6.49)$$

con $D = \text{Constante}$, entonces al reemplazar v en la ec. de continuidad se obtiene la ec. de difusión,

$$\frac{\partial}{\partial t} P + D \frac{\partial^2}{\partial x^2} (P) = 0. \quad (6.50)$$

Otro ejemplo de interés es cuando la velocidad de flujo de probabilidad tiene la forma

$$v = \mu - \frac{1}{2} \frac{\partial}{\partial x} D - \frac{1}{2} D \frac{\partial}{\partial x} \log P, \quad (6.51)$$

con $\mu = \mu(x, t)$ y $D = D(x, t)$, entonces al reemplazar v en la ec. de continuidad se obtiene la ec. de Fokker-Planck,

$$\frac{\partial}{\partial t} P + \frac{\partial}{\partial x} (\mu P) - \frac{\partial^2}{\partial x^2} (D P) = 0. \quad (6.52)$$

Para ver más ejemplos en este contexto, se recomienda al lector la referencia [67]. De esta forma, la ecuación de continuidad puede ser vista como una ecuación fundamental para derivar las ecuaciones diferenciales en derivadas parciales (PDEs) que componen la NESM, ya que todas ellas son de la familia de PDEs en que se tiene primeras derivadas temporales de la probabilidad relacionada con derivadas espaciales, las cuales pueden ser de órdenes superiores.

A continuación, se mostrará una forma de derivar la ecuación de continuidad en el contexto de inferencia sobre trayectorias [67], utilizando el axioma fundamental de la NESM (6.4) y la ecuación del corte temporal (6.7). Este desarrollo, muestra como en el contexto de inferencia sobre trayectorias se puede obtener una derivación de la ecuación de continuidad que no involucra utilizar argumentos físicos como el flujo que entra o sale de una región definida en el espacio [71], ya que en este caso sólo se utilizarán las reglas del cálculo diferencial [72] y la inferencia sobre trayectorias.

Demostración

Si tomamos la derivada temporal de la ecuación del corte temporal (6.7), podemos escribir

$$\frac{\partial}{\partial t} P(y|t, I) = \int d\tilde{x} P[\tilde{x}|I] \frac{\partial}{\partial \tilde{x}_t} \delta(\tilde{x}_t - y) \dot{\tilde{x}}_t. \quad (6.53)$$

Ahora, usando la derivada de la delta de Dirac como

$$\frac{\partial}{\partial \tilde{x}_t} \delta(\tilde{x}_t - y) = -\frac{\partial}{\partial y} \delta(\tilde{x}_t - y), \quad (6.54)$$

y reemplazando en (6.53), luego de intercambiar el orden de la derivada en y con la integral de trayectorias, se obtiene entonces una igualdad similar a la ecuación de continuidad,

$$0 = \frac{\partial}{\partial t} P(y|t, I) + \frac{\partial}{\partial y} \int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_t - y) \dot{\tilde{x}}_t. \quad (6.55)$$

Utilizando la regla de las expectativas cortadas (6.20), podemos reescribir la integral de caminos como

$$0 = \frac{\partial}{\partial t} P(y|t, I) + \frac{\partial}{\partial y} \left(P(y|t, I) \int d\tilde{x} P[\tilde{x}|y, t, I] \dot{\tilde{x}}_t \right), \quad (6.56)$$

que en su forma de expectativas, queda escrita como

$$0 = \frac{\partial}{\partial t} P(y|t, I) + \frac{\partial}{\partial y} \left(P(y|t, I) \left\langle \dot{\tilde{x}}_t \right\rangle_{I, y, t} \right), \quad (6.57)$$

donde $\langle \dot{\tilde{x}}_t \rangle_{I,y,t}$, es una función dependiente de y y t , al igual que la velocidad de flujo de probabilidad en la forma tradicional de la ecuación de continuidad. Por lo tanto definiendo $v(y, t) := \langle \dot{\tilde{x}}_t \rangle_{I,y,t}$, se recupera la forma de la ecuación de continuidad (6.41),

$$0 = \frac{\partial}{\partial t} P(y|t, I) + \frac{\partial}{\partial y} \left(P(y|t, I) v(y, t) \right) \blacksquare \quad (6.58)$$

Como se mostró al inicio de esta sección, las distintas formas de v dan origen a las diferentes formas de las PDEs que componen la NESM, que en el caso de inferencia sobre trayectorias se relaciona con encontrar el valor de la expectación

$$v(y, t) = \langle \dot{\tilde{x}}_t \rangle_{I,y,t} \quad (6.59)$$

$$= \int d\tilde{x} P[\tilde{x}|y, t, I] \dot{\tilde{x}}_t, \quad (6.60)$$

donde la densidad de probabilidad funcional $P[\tilde{x}|y, t, I]$ puede ser encontrada entre otras formas utilizando MaxCal dadas restricciones tiempo-dependientes.

Es bueno destacar que desde la inferencia, la conservación global de la probabilidad en el espacio no es consecuencia de la ecuación de continuidad, sino que como se procederá a demostrar, es consecuencia directa del axioma fundamental de la NESM a través de la ec. del corte temporal. Para entender esto, si integramos la ecuación del corte temporal (6.7) en todo el espacio donde se define la probabilidad instantánea,

$$\begin{aligned} \int dy P(y|t, I) &= \int dy \int d\tilde{x} P[\tilde{x}|I] \delta(\tilde{x}_t - y). \\ &= \int d\tilde{x} P[\tilde{x}|I] \int dy \delta(\tilde{x}_t - y). \end{aligned} \quad (6.61)$$

Resolviendo la integral del lado derecho utilizando la propiedad de la delta de Dirac $\int dy \delta(\tilde{x}_t - y) = 1$, y junto con el uso de la propiedad de que el funcional

del probabilidad está normalizado en el espacio de trayectorias (5.11),

$$\int dy P(y|t, I) = \int d\tilde{x} P[\tilde{x}|I]. \quad (6.62)$$

$$= 1. \quad (6.63)$$

Se verifica entonces que para todo t , la probabilidad instantánea $P(x|t, I)$ está normalizada en todo el espacio.

6.2. NESM desde MaxCal

Como se mostró en la sección anterior es posible recuperar gran parte de la estructura matemática de la NESM utilizando lo que llamamos inferencia de trayectorias. El punto de tope en estos formalismos se encuentra en que es necesario conocer la forma de la densidad de probabilidad funcional $P[\tilde{x}|I]$. El que esto ocurra no es sorpresa, ya que de otra forma la NESM sería una teoría única y por tanto vacía, sin posibilidad de incorporar los detalles particulares de un sistema. La misma situación ocurre en la mecánica clásica o la cuántica, donde la estructura matemática está dada salvo hasta cierto punto donde debe especificarse la forma particular del Lagrangiano o Hamiltoniano para poder concluir cosas específicas de un sistema. Es por este motivo la importancia de principios como MaxCal, ya que para un sistema particular, dadas sus restricciones dependientes del tiempo, este principio nos permite encontrar la forma que tiene la densidad de probabilidad funcional.

En esta sección se mostrará cómo mediante el uso de dicho funcional obtenido a partir de MaxCal, es posible obtener la ecuación diferencial en derivadas parciales (PDE) que gobierna la evolución de la densidad de probabilidad fuera del equilibrio. Primero mostraremos una aproximación mediante el uso de la ecuación de continuidad para construir algunas de las PDEs de la NESM, y a continuación, otra

aproximación mediante el uso de la fórmula de Feynman-Kac.

6.2.1. Ecuación de Difusión

En el marco de la física estadística, en particular de la teoría cinética [8, 73], la evolución temporal de una densidad de probabilidad para un movimiento Browniano es descrita por una ecuación diferencial en derivadas parciales llamada ecuación de difusión,

$$\frac{\partial}{\partial t} P + D \frac{\partial^2}{\partial x^2} P = 0, \quad (6.64)$$

donde $P = P(x|t, I)$ y D es una constante positiva.

Para entender el origen de esta ecuación por medio de restricciones tiempo-dependientes, es necesario primero entender uno de los caminos para construir ecuaciones diferenciales, el cual involucra el uso de la ecuación de continuidad (6.44), y para lograrlo es necesario obtener la forma de la velocidad de flujo de probabilidad $\left\langle \dot{\tilde{x}}_t \right\rangle_{I,y,t}$.

Como se mencionó en un comienzo, la ecuación de difusión describe la evolución de la densidad de probabilidad para una partícula que sigue un movimiento Browniano, es decir un movimiento aleatorio donde no existe un potencial externo que genere tendencias en la dinámica. Dicho esto, si se restringe la acción de la partícula libre (es decir, se restringe la energía cinética promedio \bar{k} en un intervalo de tiempo Δt), la cual está asociada a la dinámica de una partícula sin un potencial externo

$$\left\langle \int_{t_0}^{t_f} dt m \dot{\tilde{x}}_t^2 \right\rangle_I = 2\bar{k} \Delta t, \quad (6.65)$$

se tiene que maximizando el Calibre (5.17) bajo las restricciones de normalización (5.11) y (6.65), se obtiene que la densidad de probabilidad funcional es de la forma

(5.20) (con un *prior* constante),

$$P[\tilde{x}|I] = \frac{1}{Z(\lambda)} \exp \left(-\lambda \int_{t_0}^{t_f} dt \tilde{m}_t \dot{\tilde{x}}_t^2 \right), \quad (6.66)$$

donde en este modelo las trayectorias más probables son las que minimizan la energía cinética promedio dado que $\lambda > 0$ para que $P[\tilde{x}|I]$ sea normalizable. Para encontrar la forma de la velocidad de flujo de probabilidad, se utilizará el teorema de la variable conjugada para funcionales (CVTF) (C.7). Reemplazando la forma de la probabilidad funcional (6.66) en (C.7), se obtiene que

$$\left\langle \frac{\delta}{\delta \tilde{x}_{t'}} W[\tilde{x}] \right\rangle_I = \lambda \left\langle W[\tilde{x}] \frac{\delta}{\delta \tilde{x}_{t'}} \left(\int_{t_0}^{t_f} dt \tilde{m}_t \dot{\tilde{x}}_t^2 \right) \right\rangle_I. \quad (6.67)$$

Desarrollando la derivada funcional del lado derecho se obtiene

$$\begin{aligned} \left\langle \frac{\delta}{\delta \tilde{x}_{t'}} W[\tilde{x}] \right\rangle_I &= 2\lambda \left\langle W[\tilde{x}] \left(\int_{t_0}^{t_f} dt \tilde{m}_t \dot{\tilde{x}}_t \frac{\partial}{\partial t} \delta(t-t') \right) \right\rangle_I. \\ &= 2\lambda \left\langle W[\tilde{x}] \frac{\partial}{\partial t'} \left(\tilde{m}_{t'} \dot{\tilde{x}}_{t'} \right) \right\rangle_I. \end{aligned} \quad (6.68)$$

Si ahora usamos el funcional de prueba $W[\tilde{x}; y, \tau] = \delta(\tilde{x}_\tau - y)$, vemos que

$$\left\langle \delta'(\tilde{x}_\tau - y) \delta(\tau - t') \right\rangle_I = -2\lambda \left\langle \delta(\tilde{x}_\tau - y) \frac{\partial}{\partial t'} \left(\tilde{m}_{t'} \dot{\tilde{x}}_{t'} \right) \right\rangle_I. \quad (6.69)$$

Integrando sobre la variable t' se obtiene

$$\begin{aligned} \int dt' \left\langle \delta'(\tilde{x}_\tau - y) \delta(\tau - t') \right\rangle_I &= \\ -2\lambda \left\langle \delta(\tilde{x}_\tau - y) \int dt' \frac{\partial}{\partial t'} \left(\tilde{m}_{t'} \dot{\tilde{x}}_{t'} \right) \right\rangle_I. \end{aligned} \quad (6.70)$$

Entonces,

$$\left\langle \delta'(\tilde{x}_\tau - y) \right\rangle_I = -2\lambda \left\langle \delta(\tilde{x}_\tau - y) \left(\tilde{m}_{t'} \dot{\tilde{x}}_{t'} \right) \right\rangle_I. \quad (6.71)$$

Esta es una igualdad válida para cualquier valor de τ y t' dentro del intervalo $[t_0, t_f]$. En particular con la elección $\tau = t'$ se tiene que

$$\left\langle \delta'(\tilde{x}_{t'} - y) \right\rangle_I = -2\lambda \left\langle \delta(\tilde{x}_{t'} - y) \left(\tilde{m}_{t'} \dot{\tilde{x}}_{t'} \right) \right\rangle_I. \quad (6.72)$$

Sacando fuera de la expectación lo que no depende de la trayectoria, obtenemos

$$\left\langle \delta'(\tilde{x}_{t'} - y) \right\rangle_I = -2\lambda \tilde{m}_{t'} \left\langle \delta(\tilde{x}_{t'} - y) \dot{\tilde{x}}_{t'} \right\rangle_I. \quad (6.73)$$

Reescribiendo la derivada de la delta de Dirac y utilizando la ec. de corte en el lado izquierdo, además de usar la regla de expectativas cortadas (6.20) en el lado derecho, tenemos que

$$-\frac{\partial}{\partial y} \left\langle \delta(\tilde{x}_{t'} - y) \right\rangle_I = -2\lambda \tilde{m}_{t'} \left\langle \delta(\tilde{x}_{t'} - y) \dot{\tilde{x}}_{t'} \right\rangle_I, \quad (6.74)$$

y por tanto

$$\frac{\partial}{\partial y} P(y|t', I) = 2\lambda \tilde{m}_{t'} P(y|t', I) \left\langle \dot{\tilde{x}}_{t'} \right\rangle_{I, y, t'}. \quad (6.75)$$

Se obtiene finalmente la forma de la velocidad

$$\left\langle \dot{\tilde{x}}_{t'} \right\rangle_{I, y, t'} = \frac{-1}{2\lambda \tilde{m}_{t'} P(y|t', I)} \frac{\partial}{\partial y} P(y|t', I). \quad (6.76)$$

Reemplazando esta velocidad en la ecuación de continuidad (6.44), se obtiene que la probabilidad instantánea sigue la ecuación de difusión,

$$\frac{\partial}{\partial t} P(y|t', I) - \frac{1}{2\lambda \tilde{m}_{t'}} \frac{\partial^2}{\partial x^2} P(y|t', I) = 0 \blacksquare \quad (6.77)$$

de forma consistente con los resultados obtenidos por autores como E. Nelson³ [74].

El coeficiente de difusión en nuestro caso también se comporta como el inverso de la masa, sólo que proporcional al multiplicador de Lagrange λ y no \hbar ,

$$D = \frac{1}{2\lambda \tilde{m}_{t'}},$$

valor que siempre es positivo.

³El autor examina la hipótesis de que toda partícula que recorre trayectorias continuas se mueve bajo un movimiento Browniano con coeficiente de difusión $\frac{\hbar}{2m}$ y sin rose, lo que le permite obtener una interpretación clásica de la ecuación de Schrödinger

6.2.2. Ecuación de Liouville

El teorema de Liouville es un teorema de la mecánica clásica [57], el cual describe la evolución temporal de un conjunto de partículas que componen un sistema clásico bajo una dinámica Hamiltoniana. Mediante el uso de este teorema es que, dadas las condiciones iniciales de posición y momenta de las partículas de un sistema, es posible evolucionar tiempo a tiempo las posiciones y momenta de todas las partículas de acuerdo a la función hamiltoniano que describe el sistema. Una de las principales consecuencias del teorema de Liouville es que dado un conjunto de partículas ubicadas en una región del espacio de fases, dicha región mantendrá invariante su volumen a pesar de que se deformará a medida que las partículas evolucionen.

El teorema de Liouville puede ser escrito como una ecuación dinámica en términos del corchete de Poisson. Ésta es una ecuación para la densidad de probabilidad fuera del equilibrio en el espacio de fases, de la forma

$$\frac{\partial}{\partial t}P + \{P, H\} = 0, \quad (6.78)$$

donde $P = P(x, p|t, I)$ es la densidad de probabilidad del espacio de fases, $H = H(x, p, t)$, es el hamiltoniano del sistema y donde $\{\cdot, \cdot\}$, representa los corchetes de Poisson de la mecánica clásica, definidos tal que

$$\{P, H\} = \frac{\partial P}{\partial x} \frac{\partial H}{\partial p} - \frac{\partial P}{\partial p} \frac{\partial H}{\partial x}. \quad (6.79)$$

Para entender la ecuación de Liouville en el contexto de inferencia de trayectorias, es necesario primero ver cuál es la ecuación de continuidad asociada. Como sabemos, la ecuación de Liouville emerge cuando se estudia la evolución temporal de la probabilidad para los estados en el espacio de fases, y es un caso particular de la ecuación de continuidad.

Utilizando la ecuación del corte temporal (6.6) para representar la densidad de probabilidad para el espacio de fases como una integral de trayectorias, podemos escribir

$$P(x, p|t, I) = \int d\tilde{x} d\tilde{p} P[\tilde{x}, \tilde{p}|I] \delta(\tilde{x}_t - x) \delta(\tilde{p}_t - p). \quad (6.80)$$

Tomando su derivada temporal parcial, se obtiene que la ecuación de continuidad correspondiente es

$$\frac{\partial}{\partial t} P + \frac{\partial}{\partial x} \left\langle \delta(\tilde{x}_t - x) \delta(\tilde{p}_t - p) \dot{\tilde{x}}_t \right\rangle_{I,t} + \frac{\partial}{\partial p} \left\langle \delta(\tilde{x}_t - x) \delta(\tilde{p}_t - p) \dot{\tilde{p}}_t \right\rangle_{I,t} = 0, \quad (6.81)$$

con $P : P(x, p|t, I)$ la densidad de probabilidad instantánea. Entonces, usando la propiedad de las expectativas cortadas (6.20) se reescribe el resultado anterior como una ecuación de continuidad de la forma

$$\frac{\partial}{\partial t} P + \frac{\partial}{\partial x} \left(P \left\langle \dot{\tilde{x}}_t \right\rangle_{I,x,p,t} \right) + \frac{\partial}{\partial p} \left(P \left\langle \dot{\tilde{p}}_t \right\rangle_{I,x,p,t} \right) = 0. \quad (6.82)$$

Por lo que para encontrar la PDE que siguen los sistemas definidos en un espacio de fases para las variables posición y momenta, es necesario encontrar el valor de las velocidades de flujo de probabilidad, las cuales corresponden a expectativas cortadas dadas por

$$\left\langle \dot{\tilde{x}}_t \right\rangle_{I,x,p,t} = \int d\tilde{x} d\tilde{p} P[\tilde{x}, \tilde{p}|x, p, t, I] \dot{\tilde{x}}_t. \quad (6.83)$$

$$\left\langle \dot{\tilde{p}}_t \right\rangle_{I,x,p,t} = \int d\tilde{x} d\tilde{p} P[\tilde{x}, \tilde{p}|x, p, t, I] \dot{\tilde{p}}_t. \quad (6.84)$$

Previo a realizar un procedimiento análogo al ejemplo anterior utilizando CVTF (C.7), es necesario primero profundizar en la densidad de probabilidad funcional para los caminos que siguen el sistema en el espacio de fases $P[\tilde{x}, \tilde{p}|I]$. Para esto utilizaremos la regla del producto, descomponiendo la probabilidad en dos factores,

$$P[\tilde{x}, \tilde{p}|I] = P[\tilde{x}|I] P[\tilde{p}|\tilde{x}, I]. \quad (6.85)$$

Ahora sabemos que el primer factor del lado derecho $P[\tilde{x}|I]$ puede ser construido utilizando MaxCal, de modo que, si restringimos la acción clásica $A[\tilde{x}] = \int dt \mathcal{L}(\tilde{x}_t, \dot{\tilde{x}}_t, t)$, se obtiene una densidad de probabilidad funcional de la forma

$$P[\tilde{x}|I] = \frac{\exp\left(-\lambda \int dt \mathcal{L}(\tilde{x}_t, \dot{\tilde{x}}_t, t)\right)}{Z(\lambda)}. \quad (6.86)$$

Por otra parte, para obtener el segundo factor del lado derecho, $P[\tilde{p}|\tilde{x}, I]$, es necesario ir a la definición de momento, y es que para un sistema Lagrangiano si conocemos su posición y lagrangiano, el momento queda dado por

$$\tilde{p}_t = \frac{\partial \mathcal{L}(\tilde{x}_t, \dot{\tilde{x}}_t, t)}{\partial \dot{\tilde{x}}_t}, \quad (6.87)$$

es decir, existe un único camino \tilde{p} compatible con \tilde{x} . Por lo tanto, la probabilidad de tener un camino de momento \tilde{p} dado su trayectoria de posición \tilde{x} es necesariamente una delta de Dirac funcional, la cual condiciona la trayectoria de momento a su definición

$$P[\tilde{p}|\tilde{x}, I] = \delta(\epsilon[\tilde{x}, \tilde{p}]), \quad (6.88)$$

donde hemos introducido el funcional ϵ para implementar la delta funcional. Definimos $\epsilon[\tilde{x}, \tilde{p}]$ como

$$\epsilon[\tilde{x}, \tilde{p}] = \int dt \left[\frac{\partial \mathcal{L}(\tilde{x}_t, \dot{\tilde{x}}_t, t)}{\partial \dot{\tilde{x}}_t} - \tilde{p}_t \right]^2. \quad (6.89)$$

Dado esto, es directo ver que la condición $\epsilon[\tilde{x}, \tilde{p}] = 0$ es equivalente a la igualdad de las funciones impuesta por la delta funcional, (6.87).

Entonces reemplazando (6.86) y (6.88) en (6.85), la probabilidad conjunta se escribe de la forma,

$$P[\tilde{x}, \tilde{p}|I] = \frac{\exp\left(-\lambda \int dt \mathcal{L}(\tilde{x}_t, \dot{\tilde{x}}_t, t)\right)}{Z(\lambda)} \delta(\epsilon[\tilde{x}, \tilde{p}]). \quad (6.90)$$

Como la delta restringe a las trayectorias \tilde{p} permitidas a sólo las que cumplen la condición (6.87), ahora es posible realizar la transformada de Legendre de \mathcal{L} para todo tiempo, escribiendo entonces la probabilidad como

$$P[\tilde{x}, \tilde{p}|I] = \frac{\exp\left(-\lambda \int dt \tilde{p}_t \dot{\tilde{x}}_t - H(\tilde{x}_t, \tilde{p}_t, t)\right)}{Z(\lambda)} \delta(\epsilon[\tilde{x}, \tilde{p}]). \quad (6.91)$$

Utilizando CVTF (C.7), se tienen dos relaciones tanto para la derivada funcional de \tilde{x} y otra para \tilde{p} . En el caso de la derivada funcional respecto a \tilde{x} , la relación queda de la forma

$$\begin{aligned} \left\langle \frac{\delta}{\delta \tilde{x}_{t'}} W[\tilde{x}, \tilde{p}] \right\rangle_I &= \lambda \left\langle W[\tilde{x}, \tilde{p}] \frac{\delta}{\delta \tilde{x}_{t'}} \int dt \tilde{p}_t \dot{\tilde{x}}_t - H(\tilde{x}_t, \tilde{p}_t, t) \right\rangle_I \\ &- \lambda \left\langle W[\tilde{x}, \tilde{p}] \frac{\delta}{\delta \tilde{x}_{t'}} \log \delta(\epsilon[\tilde{x}, \tilde{p}]) \right\rangle_I, \end{aligned} \quad (6.92)$$

mientras que para la derivada funcional respecto a \tilde{p} , se obtiene

$$\begin{aligned} \left\langle \frac{\delta}{\delta \tilde{p}_{t'}} W[\tilde{x}, \tilde{p}] \right\rangle_I &= \lambda \left\langle W[\tilde{x}, \tilde{p}] \frac{\delta}{\delta \tilde{p}_{t'}} \int dt \tilde{p}_t \dot{\tilde{x}}_t - H(\tilde{x}_t, \tilde{p}_t, t) \right\rangle_I, \\ &- \lambda \left\langle W[\tilde{x}, \tilde{p}] \frac{\delta}{\delta \tilde{p}_{t'}} \log \delta(\epsilon[\tilde{x}, \tilde{p}]) \right\rangle_I. \end{aligned} \quad (6.93)$$

Podemos notar que las derivadas funcionales de los logaritmos de delta de Dirac involucran derivadas de las delta de la forma

$$\frac{\delta}{\delta \tilde{x}_{t'}} \delta(\epsilon[\tilde{x}, \tilde{p}]) = 2 \left[\tilde{p}_t - \frac{\partial l}{\partial x} \right] \left[\frac{\partial^2 l}{\partial x \partial \dot{x}} - \frac{d}{dt} \frac{\partial^2 l}{\partial \dot{x}^2} \right], \quad (6.94)$$

$$\frac{\delta}{\delta \tilde{p}_{t'}} \delta(\epsilon[\tilde{x}, \tilde{p}]) = 2 \left[\tilde{p}_t - \frac{\partial l}{\partial x} \right], \quad (6.95)$$

los cuales evaluados en la condición que define la forma del momento (6.87) se anulan, por lo que eliminando los logaritmos de las deltas y desarrollando las otras derivadas funcionales se obtienen las igualdades

$$\left\langle \frac{\delta}{\delta \tilde{x}_{t'}} W[\tilde{x}, \tilde{p}] \right\rangle_I = \lambda \left\langle W[\tilde{x}, \tilde{p}] \left(-\dot{\tilde{p}}_{t'} - \frac{\partial H(\tilde{x}_{t'}, \tilde{p}_{t'}, t')}{\partial \tilde{x}_{t'}} \right) \right\rangle_I, \quad (6.96)$$

$$\left\langle \frac{\delta}{\delta \tilde{p}_{t'}} W[\tilde{x}, \tilde{p}] \right\rangle_I = \lambda \left\langle W[\tilde{x}, \tilde{p}] \left(\dot{\tilde{x}}_{t'} - \frac{\partial H(\tilde{x}_{t'}, \tilde{p}_{t'}, t')}{\partial \tilde{p}_{t'}} \right) \right\rangle_I. \quad (6.97)$$

A modo de ejemplo, podemos ver que si el funcional W es constante, entonces se obtiene que el sistema sigue dinámicas hamiltonianas en expectativa, es decir, el sistema cumple las ecuaciones de Hamilton en expectativa

$$\left\langle \dot{\tilde{p}}_{t'} \right\rangle_I = - \left\langle \frac{\partial H(\tilde{x}_{t'}, \tilde{p}_{t'}, t')}{\partial \tilde{x}_{t'}} \right\rangle_I. \quad (6.98)$$

$$\left\langle \dot{\tilde{x}}_{t'} \right\rangle_I = \left\langle \frac{\partial H(\tilde{x}_{t'}, \tilde{p}_{t'}, t')}{\partial \tilde{p}_{t'}} \right\rangle_I. \quad (6.99)$$

Retomando el objetivo de la derivación, que es encontrar el valor de las velocidades de flujo de probabilidad, si escogemos $W[\tilde{x}, \tilde{p}] = \delta(\tilde{x}_{t'} - x)\delta(\tilde{p}_{t'} - p)$ para así encontrar las expectativas faltantes en la ecuación de continuidad, entonces (6.96) queda escrita como

$$\left\langle \delta'(\tilde{x}_{t'} - x) \delta(\tilde{p}_{t'} - p) \right\rangle_I = \lambda \left\langle \delta(\tilde{x}_{t'} - x) \delta(\tilde{p}_{t'} - p) \left(-\dot{\tilde{p}}_t - \frac{\partial H(\tilde{x}_t, \tilde{p}_t, t)}{\partial \tilde{x}_t} \right) \right\rangle_I, \quad (6.100)$$

$$-\frac{\partial}{\partial x} \left\langle \delta(\tilde{x}_{t'} - x) \delta(\tilde{p}_{t'} - p) \right\rangle_I = \lambda \left\langle \delta(\tilde{x}_{t'} - x) \delta(\tilde{p}_{t'} - p) \left(-\dot{\tilde{p}}_t - \frac{\partial H(\tilde{x}_t, \tilde{p}_t, t)}{\partial \tilde{x}_t} \right) \right\rangle_I. \quad (6.101)$$

Escribiendo el lado izquierdo utilizando (6.80), y el lado derecho con la regla de expectativas cortadas (6.20) se obtiene

$$-\frac{\partial}{\partial x} P(x, p|t, I) = -\lambda P(x, p|t, I) \left[\left\langle \dot{\tilde{p}}_t \right\rangle_{I, x, p, t} + \frac{\partial H(x, p, t)}{\partial x} \right]. \quad (6.102)$$

Realizando el mismo procedimiento para (6.97), vemos que

$$\left\langle \delta(\tilde{x}_{t'} - x) \delta'(\tilde{p}_{t'} - p) \right\rangle_I = \lambda \left\langle \delta(\tilde{x}_{t'} - x) \delta(\tilde{p}_{t'} - p) \left(\dot{\tilde{x}}_t - \frac{\partial H(\tilde{x}_t, \tilde{p}_t, t)}{\partial \tilde{p}_t} \right) \right\rangle_I, \quad (6.103)$$

$$-\frac{\partial}{\partial p} \left\langle \delta(\tilde{x}_{t'} - x) \delta(\tilde{p}_{t'} - p) \right\rangle_I = \lambda \left\langle \delta(\tilde{x}_{t'} - x) \delta(\tilde{p}_{t'} - p) \left(\dot{\tilde{x}}_t - \frac{\partial H(\tilde{x}_t, \tilde{p}_t, t)}{\partial \tilde{p}_t} \right) \right\rangle_I, \quad (6.104)$$

por lo que escribiendo el lado izquierdo utilizando (6.80) y el lado derecho con la regla de expectativas cortadas (6.20) se obtiene

$$-\frac{\partial}{\partial x}P(x, p|t, I) = \lambda P(x, p|t, I) \left[\left\langle \dot{\tilde{x}}_t \right\rangle_{I, x, p, t} - \frac{\partial H(x, p, t)}{\partial p} \right]. \quad (6.105)$$

donde debemos incorporar la forma que tendrán las velocidades de flujo de probabilidad necesarias para completar la ecuación de continuidad correspondiente (6.82).

Éstas quedan dadas por

$$\left\langle \dot{\tilde{p}}_t \right\rangle_{I, x, p, t} = \frac{1}{\lambda P(x, p|t, I)} \frac{\partial}{\partial x}P(x, p|t, I) - \frac{\partial H(x, p, t)}{\partial x}, \quad (6.106)$$

$$\left\langle \dot{\tilde{x}}_t \right\rangle_{I, x, p, t} = -\frac{1}{\lambda P(x, p|t, I)} \frac{\partial}{\partial x}P(x, p|t, I) + \frac{\partial H(x, p, t)}{\partial p}, \quad (6.107)$$

y finalmente, reemplazando (6.106) y (6.107) en (6.82), se obtiene la ecuación de Liouville,

$$\frac{\partial}{\partial t}P + \frac{\partial P}{\partial x} \frac{\partial H}{\partial p} - \frac{\partial P}{\partial p} \frac{\partial H}{\partial x} = 0 \blacksquare \quad (6.108)$$

Como consecuencia de esta demostración, vemos que la ecuación de Liouville es la descripción natural de un sistema por medio de sus coordenadas del espacio de fases, donde las variables x y p , se encuentran relacionadas por medio de las ecuaciones de Hamilton. Esto es, sólo se requiere que el sistema sea descrito en términos de un modelo MaxCal con un Lagrangiano $\mathcal{L}(x, \dot{x}; t)$ en el cual sea posible realizar la transformada de Legendre.

6.2.3. Fórmula de Feynman-Kac: Ecuación tipo Schrödinger para probabilidades

La formula de Feynman-Kac establece una correspondencia entre las integrales de trayectorias y las ecuaciones diferenciales en derivadas parciales. Esta correspondencia fue descubierta inicialmente por Feynman en el contexto de su formulación

de la mecánica cuántica a partir del formalismo de integrales de trayectorias, siendo uno de los principales resultados obtenidos por Feynman el cómo expresar la ecuación de Schrödinger, un elemento central en la mecánica cuántica, como una integral de trayectorias. Luego de los resultados de Feynman, en el contexto de procesos estocásticos, fue que Kac entendió que la correspondencia encontrada por Feynman era más que un método que permitía resolver las engorrosas integrales de trayectorias sino que más bien era una regla matemática de validez general, formalizando el resultado en lo que hoy se conoce como la fórmula de Feynman-Kac.

Para entender la fórmula de Feynman-Kac en el contexto de inferencia sobre trayectorias se requiere comprender la ecuación de Chapman-Kolmogorov (6.29). Vemos que reemplazando la forma que tiene la probabilidad de transición (6.30) con la densidad de probabilidad funcional encontrada utilizando MaxCal para un sistema en el que se restringe la acción clásica, se obtiene,

$$P(x_1, t_1 | x_0, t_0, I) = \int d\tilde{x} \frac{\exp\left(-\lambda A[\tilde{x}; t_0, x_0 = \tilde{x}_{t_0}]\right)}{Z(\lambda)} \delta(\tilde{x}_{t_1} - x_1), \quad (6.109)$$

donde la acción está definida bajo la condición de borde $x_0 = \tilde{x}_{t_0}$, lo que será escrito solo como x_0 para reducir la notación. Finalmente la acción queda como

$$\begin{aligned} A[\tilde{x}; t_0, x_0] &= \int_{t_0}^{\infty} dt \mathcal{L}(\tilde{x}_t, \dot{\tilde{x}}_t, t) \\ &= \int_{t_0}^{\infty} dt \left[\frac{m \dot{\tilde{x}}_t^2}{2} - V(\tilde{x}_t, t) \right]. \end{aligned} \quad (6.110)$$

Entonces, al resolver la integral de trayectorias (6.109), podemos notar que la delta solo permitirá las trayectorias que cumplen la condición $x_1 = \tilde{x}_{t_1}$, lo que será escrito solo como x_1 para reducir la notación. De esta forma la acción queda escrita

como $A[\tilde{x}; t_0, x_0, t_1, x_1]$. Se tiene entonces una probabilidad de transición dada por

$$P(x_1, t_1 | x_0, t_0, I) = \frac{\exp\left(-\lambda A[\tilde{x}; t_0, x_0, t_1, x_1]\right)}{Z(\lambda)}. \quad (6.111)$$

$$= \frac{\exp\left(-\lambda \int_{t_0}^{t_1} dt \left[\frac{m \dot{\tilde{x}}_t^2}{2} - V(\dot{\tilde{x}}_t, t)\right]\right)}{Z(\lambda)}. \quad (6.112)$$

Este resultado es similar al propagador de Feynman utilizado para derivar la ecuación de Schrödinger, sólo que en este caso se tiene una función real correspondiente a una probabilidad de transición y en el caso de Feynman el propagador es una amplitud de probabilidad correspondiente a un número complejo.

Usando esta probabilidad de transición, la ecuación de Chapman-Kolmogorov queda como

$$P(x_1 | t_1, I) = \int_{-\infty}^{\infty} dx_0 P(x_0 | t_0, I) \frac{\exp\left(-\lambda \int_{t_0}^{t_1} dt \left[\frac{m \dot{\tilde{x}}_t^2}{2} - V(\dot{\tilde{x}}_t, t)\right]\right)}{Z(\lambda)}. \quad (6.113)$$

Los siguientes pasos son basados en el trabajo realizado por R. Feynman [30](p. 77), donde utilizando la formulación de integrales de trayectorias Feynman deriva la ecuación de Schrödinger.

Si suponemos que t_0 es cercano a t_1 , de modo que $dt = t_1 - t_0$ con $dt \ll 1$, donde $\tilde{x}_{t_0} = x_0$ y $\tilde{x}_{t_1} = x_1$, entonces, la ecuación de Chapman-Kolmogorov queda escrita de la forma

$$P(x_1 | t_0 + dt, I) = \int_{-\infty}^{\infty} dx_0 P(x_0 | t_0, I) \frac{\exp\left(-\lambda \int dt m \frac{(x_1 - x_0)^2}{2 dt^2} - \int dt \lambda V\left(\frac{x_0 + x_1}{2}, t\right)\right)}{Z(\lambda)}. \quad (6.114)$$

Si ahora hacemos el cambio de variable $x_0 = x_1 + \eta$, con η un número también

pequeño del orden de dt , tal que ηdt tiene valor finito, tenemos

$$P(x_1|t_0 + dt, I) = \int_{-\infty}^{\infty} d\eta P(x_1 + \eta|t_0, I) \frac{\exp\left(-\lambda m \frac{\eta^2}{2 dt}\right)}{Z(\lambda)} \times \exp\left(-dt \lambda V(x_1 + \frac{\eta}{2}, t_0)\right). \quad (6.115)$$

Expandiendo la probabilidad de la derecha, y factorizando lo que no depende de la variable de integración, podemos escribir

$$P(x_1|t_0 + dt, I) = \frac{1}{Z(\lambda)} \exp\left(-dt \lambda V(x_1 + \frac{\eta}{2}, t_0)\right) \times \int_{-\infty}^{\infty} d\eta \left[P(x_1|t_0, I) + \eta \frac{\partial P(x_1|t_0, I)}{\partial x_1} + \frac{\eta^2}{2} \frac{\partial^2 P(x_1|t_0, I)}{\partial x_1^2} \right] \times \exp\left(-\lambda m \frac{\eta^2}{2 dt}\right). \quad (6.116)$$

Resolviendo las integrales gaussianas,

$$\int_{-\infty}^{\infty} d\eta \exp\left(-\lambda m \frac{\eta^2}{2 dt}\right) = \sqrt{\frac{2\pi dt}{\lambda m}}, \quad (6.117)$$

$$\int_{-\infty}^{\infty} d\eta \exp\left(-\lambda m \frac{\eta^2}{2 dt}\right) \eta = 0, \quad (6.118)$$

$$\int_{-\infty}^{\infty} d\eta \exp\left(-\lambda m \frac{\eta^2}{2 dt}\right) \eta^2 = \sqrt{\frac{2\pi dt^3}{\lambda^3 m^3}}, \quad (6.119)$$

se obtiene que

$$P(x_1, t_0 + dt|I) = \frac{1}{Z(\lambda)} \exp\left(-dt \lambda V(x_1 + \frac{\eta}{2}, t_0)\right) \times \left[\sqrt{\frac{2\pi dt}{\lambda m}} P(x_1|t_0, I) + \frac{1}{2} \sqrt{\frac{2\pi dt^3}{\lambda^3 m^3}} \frac{\partial^2 P(x_1|t_0, I)}{\partial x_1^2} \right]. \quad (6.120)$$

Como la relación anterior es válida para todo valor de dt , en particular para el límite en que $dt \rightarrow 0$, se obtiene que el valor de la normalización sólo puede ser dado por

$$Z(\lambda) = \sqrt{\frac{2\pi dt}{\lambda m}}, \quad (6.121)$$

con este valor para la $Z(\lambda)$, podemos ver que si $dt \rightarrow 0$, la igualdad sigue siendo válida. Entonces, reemplazando el valor de $Z(\lambda)$, escribimos (6.120) como

$$P(x_1, t_0 + dt|I) = \exp\left(-dt \lambda V(x_1 + \frac{\eta}{2}, t_0)\right) \left[P(x_1|t_0, I) + \frac{dt}{2\lambda m} \frac{\partial^2 P(x_1|t_0, I)}{\partial x_1^2} \right]. \quad (6.122)$$

Aquí podemos expandir la exponencial del lado derecho usando series de Taylor, obteniendo

$$P(x_1, t_0 + dt|I) = \left[1 - dt \lambda V(x_1 + \frac{\eta}{2}, t_0) \right] \left[P(x_1|t_0, I) + \frac{dt}{2\lambda m} \frac{\partial^2 P(x_1|t_0, I)}{\partial x_1^2} \right]. \quad (6.123)$$

Luego expandiendo el potencial V y eliminando el término de orden ηdt como se muestra a continuación,

$$dt V(x_1 + \frac{\eta}{2}, t_0) = dt V(x_1, t_0) + dt \frac{\eta}{2} V'(x_1, t_0) \quad (6.124)$$

$$= dt V(x_1, t_0), \quad (6.125)$$

se obtiene que

$$P(x_1, t_0 + dt|I) = \left[1 - dt \lambda V(x_1, t_0) \right] \left[P(x_1|t_0, I) + \frac{dt}{2\lambda m} \frac{\partial^2 P(x_1|t_0, I)}{\partial x_1^2} \right]. \quad (6.126)$$

Expandiendo la probabilidad de la izquierda a primer orden en dt ,

$$P(x_1|t_0, I) + dt \frac{\partial P(x_1|t_0, I)}{\partial t_0} = \left[1 - dt \lambda V(x_1, t_0) \right] \times \left[P(x_1|t_0, I) + \frac{dt}{2\lambda m} \frac{\partial^2 P(x_1|t_0, I)}{\partial x_1^2} \right], \quad (6.127)$$

entonces, resolviendo el lado derecho y eliminando los términos de orden dt^2 se obtiene,

$$P(x_1|t_0, I) + dt \frac{\partial P(x_1|t_0, I)}{\partial t_0} = P(x_1|t_0, I) + \frac{dt}{2\lambda m} \frac{\partial^2 P(x_1|t_0, I)}{\partial x_1^2} - dt \lambda V(x_1, t_0) P(x_1|t_0, I). \quad (6.128)$$

Para simplificar la notación, se reemplazará $x_1 \rightarrow x$ y $t_0 \rightarrow t$. Finalmente, eliminando los términos similares a ambos lados y multiplicando por el factor $\frac{1}{dt}$, se obtiene la forma de la ecuación diferencial que seguirá la densidad de probabilidad fuera del equilibrio,

$$\frac{\partial P(x|t, I)}{\partial t} = \frac{1}{2\lambda m} \frac{\partial^2 P(x|t, I)}{\partial x^2} - \lambda V(x, t) P(x|t, I). \quad (6.129)$$

Aquí podemos notar que si no hay potencial (es decir, $V(x, t) = 0$), entonces la ecuación que seguirá la probabilidad fuera del equilibrio corresponde a la ecuación de difusión (6.77), con el mismo coeficiente de difusión obtenido previamente, $D = \frac{1}{2\lambda m}$.

6.3. Relaciones Fuera del Equilibrio

Los elementos que se encuentran fuera del esquema que describe a la NESM por medio de ecuaciones diferenciales en derivadas parciales son las llamadas relaciones fuera del equilibrio, siendo la relación central el teorema de fluctuación de Crooks [23]. Éste relaciona las probabilidades de caminos con las probabilidades de los caminos inversos, siendo este teorema uno de los pilares en el entendimiento de la irreversibilidad [75].

Algunos casos particulares del teorema de Crooks, son la igualdad de Jarzynski [21] y los teoremas de fluctuación de Evans, Hatano-Sasa, entre otros [25, 76, 77]. La igualdad de Jarzynski ha sido validada experimentalmente [24, 78], junto con ser un método utilizado en el cálculo de diferencias de energía libre en simulaciones de dinámica molecular. Los teoremas de fluctuación de Evans también han sido verificados en simulaciones de dinámica molecular [26].

En esta sección se mostrará cómo derivar el teorema de fluctuación de Crooks y

la igualdad de Jarzynski en el contexto de inferencia sobre trayectorias. Para esto, primero se mostrará cómo opera la primera ley de la termodinámica, mostrando una definición microscópica del calor; luego, se mostrará qué implica condicionar un sistema a ser Markoviano, para así comprender la relación entre la probabilidad de una trayectoria y la probabilidad de la trayectoria inversa gracias a la ayuda de la condición conocida como balance detallado.

6.3.1. Primera ley de la Termodinámica microscópica

Para entender la primera ley de la termodinámica debemos considerar que para un sistema mecánico conservativo existen dos tipos de energía, cinética y potencial, las cuales son dependientes de las velocidades y las posiciones de las partículas que componen un sistema respectivamente. La energía mecánica es la suma de ambas, y se mantiene constante en ausencia de fuerzas externas que realicen trabajo sobre el sistema. Si hay fuerzas externas, el incremento de la energía mecánica es igual al trabajo realizado sobre el sistema por estas fuerzas.

El trabajo W es una cantidad física que se expresa en unidades de energía. Desde un punto de vista microscópico de la primera ley de la termodinámica, el trabajo es una cantidad física dependiente del camino seguido por una partícula \tilde{x} , es decir, un funcional, el cual se escribe como

$$W[\tilde{x}] = \Delta H - Q[\tilde{x}], \quad (6.130)$$

donde $\Delta H = H(\tilde{x}_{t_f}, t_f) - H(\tilde{x}_{t_0}, t_0)$ corresponde a la diferencia de energía entre el micro-estado final e inicial y Q corresponde al calor que ingresa al sistema, el cual depende de la trayectoria \tilde{x} en que se realiza el trabajo. De forma funcional, el trabajo puede ser escrito como la suma del cambio de energía por el camino \tilde{x} que

une los puntos \tilde{x}_{t_0} y \tilde{x}_{t_f} , es decir,

$$W[\tilde{x}] = \int_{t_0}^{t_f} dt \left[\frac{\partial H(\tilde{x}_t, t)}{\partial t} \right], \quad (6.131)$$

donde cambiando la derivada parcial

$$\frac{\partial H(x, t)}{\partial t} = \frac{dH(x, t)}{dt} - \frac{\partial H(x, t)}{\partial x} \dot{x}_t, \quad (6.132)$$

se obtiene,

$$\begin{aligned} W[\tilde{x}] &= \int_{t_0}^{t_f} dt \left[\frac{dH(\tilde{x}_t)}{dt} - \frac{\partial H(\tilde{x}_t, t)}{\partial x} \dot{x}_t \right], \quad (6.133) \\ &= H(\tilde{x}_{t_f}, t_f) - H(\tilde{x}_{t_0}, t_0) - \int_{t_0}^{t_f} dt \left[\frac{\partial H(\tilde{x}_t, t)}{\partial x} \dot{x}_t \right], \\ &= \Delta H - \int_{t_0}^{t_f} dt \frac{\partial H(\tilde{x}_t, t)}{\partial x} \dot{x}_t, \end{aligned}$$

por lo que comparando con la primera ley de la termodinámica, se obtiene la forma funcional microscópica que tiene el calor,

$$\begin{aligned} Q[\tilde{x}] &= \int_{t_0}^{t_f} dt \left[\frac{\partial H(\tilde{x}_t, t)}{\partial x} \dot{x}_t \right] \quad (6.134) \\ &= \int_{\tilde{x}_{t_0}}^{\tilde{x}_{t_f}} dx \frac{\partial H(\tilde{x}_t, t)}{\partial x}. \end{aligned}$$

6.3.2. Evolución Markoviana

Un sistema dinámico es considerado Markoviano cuando la probabilidad de cada estado depende únicamente del estado anterior en que se encontraba el sistema. Es decir, la probabilidad de transición de un micro-estado \tilde{x}_0 a otro micro-estado \tilde{x}_1 en un instante siguiente no puede depender de los micro-estados en otro tiempo que no sea el del instante previo. Esta condición implica que, para todo estado de conocimiento I,

$$P(\tilde{x}_i | \tilde{x}_0, \dots, \tilde{x}_{i-1}, I) = P(\tilde{x}_i | \tilde{x}_{i-1}, I). \quad (6.135)$$

Esto tiene implicaciones para la forma en que se descompone una probabilidad de trayectorias en n -instantes consecutivos. Si consideramos la trayectoria \tilde{x} discretizada en el tiempo como

$$\tilde{x} = (\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_n), \quad (6.136)$$

donde $\tilde{x}_{t_i} = \tilde{x}_i$, corresponde al instante i -ésimo de la trayectoria \tilde{x} , entonces, la probabilidad funcional puede ser escrita como

$$P[\tilde{x}|I] = P(\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_n | I), \quad (6.137)$$

y descompuesta como

$$\begin{aligned} P[\tilde{x}|I] &= P(\tilde{x}_0 | I) P(\tilde{x}_1 | \tilde{x}_0, I) P(\tilde{x}_2 | \tilde{x}_1, I) \\ &\times P(\tilde{x}_3 | \tilde{x}_2, I) \dots P(\tilde{x}_{n-1} | \tilde{x}_{n-2}, I) P(\tilde{x}_n | \tilde{x}_{n-1}, I). \end{aligned} \quad (6.138)$$

De forma análoga, podemos tomar la probabilidad de la trayectoria inversa \tilde{x}^\dagger ,

$$P[\tilde{x}^\dagger|I] = P(\tilde{x}_n, \tilde{x}_{n-1}, \dots, \tilde{x}_0 | I), \quad (6.139)$$

la cual al asumir que el sistema es Markoviano, puede ser descompuesta como

$$\begin{aligned} P[\tilde{x}^\dagger|I] &= P(\tilde{x}_n | I) P(\tilde{x}_{n-1} | \tilde{x}_n, I) P(\tilde{x}_{n-2} | \tilde{x}_{n-1}, I) \\ &\times P(\tilde{x}_{n-3} | \tilde{x}_{n-2}, I) \dots P(\tilde{x}_1 | \tilde{x}_2, I) P(\tilde{x}_0 | \tilde{x}_1, I). \end{aligned} \quad (6.140)$$

El paso de (6.137) a (6.138) es como sigue. Utilizando la regla del producto (2.5)

en la probabilidad del lado derecho de la (6.137) de forma iterativa

$$\begin{aligned}
P[\tilde{x}|I] &= P(\tilde{x}_0 | I) P(\tilde{x}_1, \dots, \tilde{x}_n | \tilde{x}_0, I). \\
&= P(\tilde{x}_0 | I) P(\tilde{x}_1 | \tilde{x}_0, I) P(\tilde{x}_2, \dots, \tilde{x}_n | \tilde{x}_0, \tilde{x}_1, I). \\
&= P(\tilde{x}_0 | I) P(\tilde{x}_1 | \tilde{x}_0, I) P(\tilde{x}_2 | \tilde{x}_0, \tilde{x}_1, I) \\
&\times P(\tilde{x}_3, \dots, \tilde{x}_n | \tilde{x}_0, \tilde{x}_1, I). \\
&\vdots \\
&= P(\tilde{x}_0 | I) P(\tilde{x}_1 | \tilde{x}_0, I) P(\tilde{x}_2 | \tilde{x}_0, \tilde{x}_1, I) \\
&\times P(\tilde{x}_3 | \tilde{x}_0, \tilde{x}_1, \tilde{x}_2, I) \dots P(\tilde{x}_n | \tilde{x}_0, \dots, \tilde{x}_{n-1}, I), \quad (6.141)
\end{aligned}$$

donde las probabilidades $P(\tilde{x}_i | \tilde{x}_0, \dots, \tilde{x}_{i-1}, I)$, corresponden a probabilidades de transición, es decir, la probabilidad de estar en el punto \tilde{x}_i dado que estaba en todos los puntos $\tilde{x}_0, \dots, \tilde{x}_{n-1}$, en los tiempos anteriores. Si la evolución es Markoviana (6.135), entonces podemos escribir (6.141) como

$$\begin{aligned}
P[\tilde{x}|I] &= P(\tilde{x}_0 | I) P(\tilde{x}_1 | \tilde{x}_0, I) P(\tilde{x}_2 | \tilde{x}_1, I) \\
&\times P(\tilde{x}_3 | \tilde{x}_2, I) \dots P(\tilde{x}_{n-1} | \tilde{x}_{n-2}, I) P(\tilde{x}_n | \tilde{x}_{n-1}, I) \blacksquare \quad (6.142)
\end{aligned}$$

De forma análoga, podemos utilizar la regla del producto para descomponer la

probabilidad de la trayectoria inversa,

$$\begin{aligned}
P[\tilde{x}^\dagger|I] &= P(\tilde{x}_n | I) P(\tilde{x}_{n-1}, \dots, \tilde{x}_0 | \tilde{x}_n, I). \\
&= P(\tilde{x}_n | I) P(\tilde{x}_{n-1} | \tilde{x}_n, I) P(\tilde{x}_{n-2}, \dots, \tilde{x}_0 | \tilde{x}_n, \tilde{x}_{n-1}, I). \\
&= P(\tilde{x}_n | I) P(\tilde{x}_{n-1} | \tilde{x}_n, I) P(\tilde{x}_{n-2} | \tilde{x}_n, \tilde{x}_{n-1}, I) \times \\
&\times P(\tilde{x}_{n-3}, \dots, \tilde{x}_0 | \tilde{x}_{n-1}, \tilde{x}_{n-2}, I). \\
&\vdots \\
&= P(\tilde{x}_n | I) P(\tilde{x}_{n-1} | \tilde{x}_n, I) P(\tilde{x}_{n-2} | \tilde{x}_n, \tilde{x}_{n-1}, I) \times \\
&\times P(\tilde{x}_{n-3} | \tilde{x}_n, \tilde{x}_{n-1}, \tilde{x}_{n-2}, I) \dots P(\tilde{x}_0 | \tilde{x}_n, \dots, \tilde{x}_1, I). \quad (6.143)
\end{aligned}$$

Entonces, si la evolución es Markoviana (6.135), se tiene que

$$\begin{aligned}
P[\tilde{x}^\dagger|I] &= P(\tilde{x}_n | I) P(\tilde{x}_{n-1} | \tilde{x}_n, I) P(\tilde{x}_{n-2} | \tilde{x}_{n-1}, I) \times \\
&\times P(\tilde{x}_{n-3} | \tilde{x}_{n-2}, I) \dots P(\tilde{x}_1 | \tilde{x}_2, I) P(\tilde{x}_0 | \tilde{x}_1, I) \blacksquare \quad (6.144)
\end{aligned}$$

6.3.3. Teorema de Fluctuación de Crooks

El Teorema de Fluctuación de Crooks [23] establece que, dado un sistema inicialmente en equilibrio termodinámico A , el cual evoluciona a un estado final en equilibrio termodinámico B , realizando la evolución a temperatura constante T , la relación entre las expectativas calculadas con la densidad de probabilidad funcional para la trayectoria (F) y las expectativas calculadas con la densidad de probabilidad para las trayectorias inversas (R) es dada por

$$\left\langle A \right\rangle_F = \left\langle \hat{A} \exp \left(-\beta \Delta F + \beta W[\tilde{x}] \right) \right\rangle_R. \quad (6.145)$$

Aquí A es un funcional dependiente de la trayectoria \tilde{x} y \hat{A} es un funcional dependiente de la trayectoria inversa \tilde{x}^\dagger . Esta igualdad tiene como punto central

el describir la relación entre la densidad de probabilidad funcional de una micro-trayectoria \tilde{x} y la densidad de probabilidad funcional de la micro-trayectoria inversa \tilde{x}^\dagger , relación dada por

$$\frac{P[\tilde{x}|I]}{P[\tilde{x}^\dagger|I]} = \exp\left(-\beta\Delta F + \beta W[\tilde{x}]\right), \quad (6.146)$$

donde $\beta = \frac{1}{k_B T}$ corresponde al inverso de la temperatura del sistema, con k_B la constante de Boltzmann, $\Delta F = F_B - F_A$ es la diferencia de energía libre de Helmholtz final e inicial del sistema, y $W[\tilde{x}]$ es el trabajo (6.131).

A continuación se demostrará la relación entre las probabilidades de una trayectoria y la trayectoria inversa, dando una primera aproximación a la derivación del teorema de Crooks en el contexto de inferencia sobre trayectorias.

Demostración

Si comparamos la razón entre la probabilidad de una trayectoria y la probabilidad de la trayectoria inversa correspondiente, dado que se tiene una evolución Markoviana (6.138) y (6.140), podemos escribir

$$\begin{aligned} \frac{P[\tilde{x}|I]}{P[\tilde{x}^\dagger|I]} &= \frac{P(\tilde{x}_0|I)}{P(\tilde{x}_n|I)} \frac{P(\tilde{x}_1|\tilde{x}_0, I)}{P(\tilde{x}_0|\tilde{x}_1, I)} \frac{P(\tilde{x}_2|\tilde{x}_1, I)}{P(\tilde{x}_1|\tilde{x}_2, I)} \\ &\quad \frac{P(\tilde{x}_3|\tilde{x}_2, I)}{P(\tilde{x}_2|\tilde{x}_3, I)} \cdots \frac{P(\tilde{x}_{n-1}|\tilde{x}_{n-2}, I)}{P(\tilde{x}_{n-2}|\tilde{x}_{n-1}, I)} \frac{P(\tilde{x}_n|\tilde{x}_{n-1}, I)}{P(\tilde{x}_{n-1}|\tilde{x}_n, I)}. \end{aligned} \quad (6.147)$$

Como el estado inicial A , correspondiente al instante t_0 y el estado final B , correspondiente al instante t_n son estados de equilibrio termodinámico, entonces podemos escribir las probabilidades de dichos estados como *ensembles* canónicos,

$$P(\tilde{x}_0|I) = \frac{\exp\left(-\beta H(\tilde{x}_0, t_0)\right)}{Z_A(\beta)}, \quad (6.148)$$

$$P(\tilde{x}_n|I) = \frac{\exp\left(-\beta H(\tilde{x}_n, t_n)\right)}{Z_B(\beta)}, \quad (6.149)$$

donde $\beta H_i(\tilde{x}_i) = \beta H(\tilde{x}_i, t_i)$. Se obtiene entonces

$$\begin{aligned} \frac{P[\tilde{x}|I]}{P[\tilde{x}^\dagger|I]} &= \frac{Z_B(\beta) \exp\left(-\beta H(\tilde{x}_0, t_0)\right)}{Z_A(\beta) \exp\left(-\beta H(\tilde{x}_n, t_n)\right)} \frac{P(\tilde{x}_1 | \tilde{x}_0, I)}{P(\tilde{x}_0 | \tilde{x}_1, I)} \frac{P(\tilde{x}_2 | \tilde{x}_1, I)}{P(\tilde{x}_1 | \tilde{x}_2, I)} \\ &\quad \frac{P(\tilde{x}_3 | \tilde{x}_2, I)}{P(\tilde{x}_2 | \tilde{x}_3, I)} \cdots \frac{P(\tilde{x}_{n-1} | \tilde{x}_{n-2}, I)}{P(\tilde{x}_{n-2} | \tilde{x}_{n-1}, I)} \frac{P(\tilde{x}_n | \tilde{x}_{n-1}, I)}{P(\tilde{x}_{n-1} | \tilde{x}_n, I)}. \end{aligned} \quad (6.150)$$

En este punto, empleamos la condición de balance detallado [23], condición que establece un equilibrio en la suma de las transiciones que ocurren al evolucionar de un estado x a x' y las transiciones que ocurren al evolucionar de un estado x' a x , asociado a la condición de reversibilidad microscópica. En el contexto de esta tesis, la condición de balance detallado será vista como una relación entre las probabilidades de transición de la forma

$$\frac{P(\tilde{x}_i | \tilde{x}_{i-1}, I)}{P(\tilde{x}_{i-1} | \tilde{x}_i, I)} = \frac{P(\tilde{x}_i | t_i, I)}{P(\tilde{x}_{i-1} | t_i, I)}. \quad (6.151)$$

Esta condición puede ser vista como un balance en la razón entre las probabilidades de transición, condicionada a la razón entre las probabilidades de estar en el estado inicial y final en tiempos iguales. Sin pérdida de generalidad, las probabilidades del lado derecha pueden ser escritas en el tiempo t_{i-1} y se obtendrá la misma derivación que será vista a continuación,

$$\begin{aligned} \frac{P[\tilde{x}|I]}{P[\tilde{x}^\dagger|I]} &= \frac{Z_B(\beta) \exp\left(-\beta H(\tilde{x}_0, t_0)\right)}{Z_A(\beta) \exp\left(-\beta H(\tilde{x}_n, t_n)\right)} \frac{P(\tilde{x}_1 | t_1, I)}{P(\tilde{x}_0 | t_1, I)} \frac{P(\tilde{x}_2 | t_2, I)}{P(\tilde{x}_1 | t_2, I)} \\ &\quad \frac{P(\tilde{x}_3 | t_3, I)}{P(\tilde{x}_2 | t_3, I)} \cdots \frac{P(\tilde{x}_{n-1} | t_{n-1}, I)}{P(\tilde{x}_{n-2} | t_{n-1}, I)} \frac{P(\tilde{x}_n | t_n, I)}{P(\tilde{x}_{n-1} | t_n, I)}. \end{aligned} \quad (6.152)$$

Por lo que se tiene para cada tiempo que

$$\frac{P(\tilde{x}_i | t_i, I)}{P(\tilde{x}_{i-1} | t_i, I)} = \frac{\exp\left(-\beta H(\tilde{x}_i, t_i)\right)}{\exp\left(-\beta H(\tilde{x}_{i-1}, t_i)\right)}. \quad (6.153)$$

Si tomamos estados muy cercanos, de modo que $\tilde{x}_i = \tilde{x}_{i-1} + dx$, podemos reducir el lado derecho como

$$\frac{P(\tilde{x}_i | t_i, I)}{P(\tilde{x}_{i-1} | t_i, I)} = \frac{\exp\left(-\beta H(\tilde{x}_{i-1} + dx, t_i)\right)}{\exp\left(-\beta H(\tilde{x}_{i-1}, t_i)\right)}. \quad (6.154)$$

$$= \exp\left(-\beta \left[H(\tilde{x}_{i-1} + dx, t_i) - H(\tilde{x}_{i-1}, t_i)\right]\right). \quad (6.155)$$

Reemplazando en (6.152), llegamos a

$$\begin{aligned} \frac{P[\tilde{x}|I]}{P[\tilde{x}^\dagger|I]} &= \frac{Z_B(\beta)}{Z_A(\beta)} \frac{\exp\left(-\beta H(\tilde{x}_0, t_0)\right)}{\exp\left(-\beta H(\tilde{x}_n, t_n)\right)} \\ &\quad \prod_{i=0}^n \exp\left(-\beta \left[H(\tilde{x}_{i-1} + dx, t_i) - H(\tilde{x}_{i-1}, t_i)\right]\right), \end{aligned} \quad (6.156)$$

donde en el límite $n \rightarrow \infty$, t_i puede ser escrita como una variable continua t . Entonces, añadiendo la identidad $1 = \frac{dx}{dx}$ en la integral para definir la variable de integración se obtiene

$$\begin{aligned} \frac{P[\tilde{x}|I]}{P[\tilde{x}^\dagger|I]} &= \frac{Z_B(\beta)}{Z_A(\beta)} \frac{\exp\left(-\beta H(\tilde{x}_0, t_0)\right)}{\exp\left(-\beta H(\tilde{x}_n, t_n)\right)} \\ &\quad \exp\left(-\beta \int_{\tilde{x}_0}^{\tilde{x}_n} dx \frac{\left[H(\tilde{x}_t + dx, t) - H(\tilde{x}_t, t)\right]}{dx}\right). \end{aligned} \quad (6.157)$$

$$= \frac{Z_B(\beta)}{Z_A(\beta)} \frac{\exp\left(-\beta H(\tilde{x}_0, t_0)\right)}{\exp\left(-\beta H(\tilde{x}_n, t_n)\right)} \exp\left(-\beta \int_{\tilde{x}_0}^{\tilde{x}_n} dx \frac{\partial H(x, t)}{\partial x}\right).$$

Usando la definición microscópica del calor (6.135), tenemos que

$$\frac{P[\tilde{x}|I]}{P[\tilde{x}^\dagger|I]} = \frac{Z_B(\beta)}{Z_A(\beta)} \frac{\exp\left(-\beta H(\tilde{x}_0, t_0)\right)}{\exp\left(-\beta H(\tilde{x}_n, t_n)\right)} \exp\left(-\beta Q[\tilde{x}]\right). \quad (6.158)$$

$$= \frac{Z_B(\beta)}{Z_A(\beta)} \exp\left(\beta \Delta H - \beta Q[\tilde{x}]\right). \quad (6.159)$$

Adicionalmente, usando la definición de energía libre $F_i = \exp(-\beta Z_i(\beta))$ y luego aplicando la primera ley de la termodinámica microscópica (6.130) se obtiene finalmente que

$$\frac{P[\tilde{x}|I]}{P[\tilde{x}^\dagger|I]} = \exp\left(-\beta\Delta F + \beta\Delta H - \beta Q[\tilde{x}]\right). \quad (6.160)$$

$$= \exp\left(-\beta\Delta F + \beta W[\tilde{x}]\right). \quad (6.161)$$

Esta igualdad está directamente relacionada al entendimiento de los procesos reversibles, mostrando explícitamente que la probabilidad de que un proceso ocurra en un sentido u otro no son iguales, sino que estará relacionado con la diferencia de energía libre y el trabajo, dejando en claro que un sistema será reversible si y solo si se cumple la condición

$$\Delta F = W[\tilde{x}], \quad (6.162)$$

ya que esto implica que las densidades de probabilidad funcional serán iguales tanto para la trayectoria \tilde{x} como su inversa \tilde{x}^\dagger ,

$$P[\tilde{x}|I] = P[\tilde{x}^\dagger|I]. \quad (6.163)$$

6.3.4. Igualdad de Jarzynski

La igualdad de Jarzynski establece que dado un sistema, inicialmente en equilibrio termodinámico A , el cual se evoluciona a un sistema final, en equilibrio termodinámico B , realizando la evolución a temperatura constante T , se tiene la relación

$$\left\langle \exp(-\beta W[\tilde{x}]) \right\rangle_{\beta, I} = \exp(-\beta\Delta F), \quad (6.164)$$

Donde $\beta = \frac{1}{k_B T}$, corresponde al inverso de la temperatura del sistema, con k_B , la constante de Boltzmann. $\Delta F = F_B - F_A$ es la diferencia de energía libre final e

inicial del sistema, y el lado izquierdo es una expectación tomada sobre trayectorias, de la cantidad funcional $\exp\left(-\beta W[\tilde{x}]\right)$.

Demostración:

Una de las consecuencias del teorema de fluctuación de Crooks (6.146) visto en el ejemplo anterior puede ser reescrita como

$$P[\tilde{x}|I] \exp\left(-\beta W[\tilde{x}]\right) = P[\tilde{x}^\dagger|I] \exp\left(-\beta\Delta F\right). \quad (6.165)$$

Integrando sobre el espacio de las trayectorias \tilde{x} ,

$$\int d\tilde{x} P[\tilde{x}|I] \exp\left(-\beta W[\tilde{x}]\right) = \int d\tilde{x} P[\tilde{x}^\dagger|I] \exp\left(-\beta\Delta F\right), \quad (6.166)$$

$$= \exp\left(-\beta\Delta F\right) \int d\tilde{x} P[\tilde{x}^\dagger|I], \quad (6.167)$$

$$= \exp\left(-\beta\Delta F\right), \quad (6.168)$$

dado que $P[\tilde{x}|I]$ está correctamente normalizado, escribimos el lado izquierdo en forma de expectación y obtenemos finalmente la Igualdad de Jarzynski,

$$\left\langle \exp\left(-\beta W[\tilde{x}]\right) \right\rangle_I = \exp\left(-\beta\Delta F\right). \blacksquare \quad (6.169)$$

Capítulo 7

Conclusiones

El principio de máxima entropía (cap. 3) propuesto inicialmente por E. T. Jaynes como una extensión al principio variacional de máxima entropía de Gibbs, es un método sistemático con el cual es posible derivar por completo la estructura matemática de la mecánica estadística mediante restricciones dadas en forma de expectación (cap. 4). Con el principio de máxima entropía como herramienta para derivar la mecánica estadística se ordena la compleja estructura de esta teoría, dejando en evidencia cuales son las restricciones que gobiernan los sistemas en equilibrio y que dan origen a los llamados *ensembles*, tales como canónico, gran-canónico, isotérmico-isobárico, entre otros y a las densidades de probabilidad que emergen de la estadística de Bose-Einstein y Fermi-Dirac utilizadas para describir sistemas con propiedades discretas (cap. 4).

Para entender el principio de máximo calibre fue necesario entender la probabilidad en un contexto de inferencia estadística, donde la probabilidad siempre está definida bajo un contexto, conocido como la información dada en las probabilidades bayesianas. Esto permite generalizar el principio de máxima entropía de un método para construir funciones de probabilidad, a una regla general de inferencia, la cual

puede ser definida en el espacio donde se requiere definir una probabilidad (cap. 3), tal como lo es el espacio de funciones o trayectorias.

El potencial del principio de máximo calibre es más que una herramienta para derivar la mecánica estadística fuera del equilibrio, ya que cuando se utiliza en conjunto con el cálculo funcional, es posible entender problemas de gran importancia en física, como lo es la correlación entre n -flujos de un sistema (cap. 5.5.5) [32] y el análisis de sistemas dinámicos clásicos los cuales siguen dinámicas lagrangianas y hamiltonianas en promedio (cap. 5.5.4).

Uno de los resultados más importantes presentados en esta tesis fue demostrar que es posible realizar implementaciones numéricas de máximo calibre mediante el uso del algoritmo de Monte Carlo Metropolis en el espacio de trayectorias. El resultado mostrado en esta tesis (cap. 5.5.6) es una de las primeras implementaciones de Monte Carlo metrópolis en el espacio de trayectorias mediante el muestreo de trayectorias continuas. Este resultado tiene un gran potencial para simular sistemas dinámicos, dejando como trabajo futuro el realizar las implementaciones para sistemas con un gran número de partículas y con potenciales de interacción mas complejos.

Si bien E. T. Jaynes mostró como construir una densidad de probabilidad funcional mediante el principio de máximo calibre, la mecánica estadística fuera del equilibrio no utiliza una densidad de probabilidad funcional como elemento central, sino una densidad de probabilidad tiempo-dependiente.

Para completar la hipótesis de Jaynes, la cual plantea que el principio de máximo calibre es un principio variacional con el cual derivar la mecánica estadística fuera del equilibrio, fue necesario introducir un concepto nuevo en la inferencia, el cual llamamos inferencia sobre trayectorias. La inferencia de trayectorias es la consecuencia

directa de definir una probabilidad para trayectorias (o funciones), donde se extienden todas las reglas y consecuencias de la teoría de probabilidad bayesiana (cap. 2) al espacio de trayectorias.

El axioma fundamental de la mecánica estadística fuera del equilibrio es uno de los pilares de la inferencia de trayectorias, el cual puede ser visto como un axioma de consistencia, estableciendo que se debe obtener el mismo resultado si se calculan las expectativas utilizando la densidad de probabilidad tiempo dependiente de la mecánica estadística fuera del equilibrio y si se calculan expectativas utilizando la densidad de probabilidad funcional obtenida utilizando el principio de máximo calibre.

Una de las consecuencias más importantes de la inferencia sobre trayectorias es la llamada ecuación del corte temporal (cap. 6.1.2), ecuación que relaciona la densidad de probabilidad funcional obtenida con el principio de máximo calibre con la densidad de probabilidad tiempo dependiente de la mecánica estadística fuera del equilibrio. Es bueno destacar que la inferencia de trayectorias se encuentra por sobre el principio de máximo calibre, ya que pueden existir otros métodos para poder definir la densidad de probabilidad funcional, pero las reglas obtenidas en la inferencia de trayectorias seguirán siendo válidas sin importar el método utilizado para construir dicha probabilidad, siendo los resultados obtenidos para inferencia sobre trayectorias son sólo consecuencia del cálculo infinitesimal. Cabe destacar que la inferencia sobre trayectorias permite plantear el problema inverso para la probabilidad, es decir, encontrar la densidad de probabilidad funcional dada la función de probabilidad tiempo dependiente para todo tiempo. Es por esto que la inferencia sobre trayectorias puede ser resumida a 2 consideraciones clave, el axioma fundamental de la mecánica estadística fuera del equilibrio y la ecuación del corte temporal, junto con

el uso de las reglas de la probabilidad para una probabilidad definida en el espacio de trayectorias.

Para derivar la mecánica estadística fuera del equilibrio mediante el uso del principio de máximo calibre, se requiere además el uso de la inferencia sobre trayectorias y el cálculo funcional.

Utilizando la inferencia sobre trayectorias y el cálculo diferencial es posible derivar la ecuación de continuidad, demostrando así que esta ecuación es una consecuencia directa de realizar inferencia sobre el espacio de trayectorias [67] (cap. 6.1.6). Lograr entender la ecuación de continuidad desde la inferencia de trayectorias permite entender el origen de las diferentes ecuaciones diferenciales en derivadas parciales que componen la mecánica estadística fuera del equilibrio, dejando en claro que los diferentes sistemas que son descritos por restricciones tiempo-dependientes tendrán diferentes velocidades de flujo de probabilidad dando origen a ecuaciones centrales en la mecánica estadística fuera del equilibrio tales como la ecuación de difusión (cap. 6.2.1) que es consecuencia de restringir la acción de una partícula libre (sin potencial externo) y la ecuación de Liouville (cap. 6.2.2) que es consecuencia de restringir un sistema lagrangiano donde es posible realizar la transformada de Legendre para así poder definir un hamiltoniano [79].

Un punto clave para entender la relación del principio de máximo calibre como herramienta para derivar las ecuaciones diferenciales en derivadas parciales de la mecánica estadística fuera del equilibrio es la fórmula de Feynman-Kac, ya que esta fórmula establece una correspondencia entre una integral de trayectorias y una ecuación diferencial en derivadas parciales, como fue mostrado en la derivación de la ecuación tipo Schrödinger para las probabilidades (cap. 6.2.3). El procedimiento

mostrado tiene un gran potencial de ser una nueva metodología para construir ecuaciones diferenciales en derivadas parciales a partir de la densidad de probabilidad funcional obtenida utilizando el principio de máximo calibre.

Finalmente se demostró que las relaciones fuera del equilibrio contenidas en la mecánica estadística fuera del equilibrio son consecuencia directa de la inferencia sobre trayectorias, donde queda en evidencia el potencial de máximo calibre como un método sistemático para definir una probabilidad para trayectorias, ya que tanto la igualdad de Jarzynski [21, 80] (cap. 6.3.4) y el teorema de Crooks (cap. 6.3.3) se encuentran expresadas como expectativas en un espacio de trayectorias, demostrando que son relaciones de validez general ya que para su derivación solo es necesario el uso de la inferencia sobre trayectorias y el cálculo funcional, es decir no están restringidos a un sistema en particular.

El principal resultado de esta tesis es que el principio de máximo calibre es un principio variacional general que permite derivar por completo la mecánica estadística fuera del equilibrio [31, 32, 65]. Es gracias a este principio que por medio de restricciones tiempo dependientes, las cuales gobiernan a los sistemas fuera del equilibrio, se obtiene la densidad de probabilidad tiempo dependiente de la mecánica estadística fuera del equilibrio con la cual se pueden obtener las propiedades macroscópicas tiempo dependientes de los sistemas mediante la operación expectativa. Para esto fue necesario definir la inferencia sobre trayectorias y mostrar cómo utilizar el cálculo funcional para operar los elementos involucrados, entregando un formalismo completo para entender la mecánica estadística fuera del equilibrio por medio de restricciones tiempo dependientes, de la misma forma que es posible derivar la mecánica

estadística desde el principio de máxima entropía. Este formalismo se vuelve aún mas completo al mostrar que es posible implementar numéricamente, quedando no sólo restringido a problemas que se resuelven de forma analítica, sino a cualquier sistema en el que sea posible entender sus restricciones tiempo-dependientes.

Apéndice A

Definiciones útiles en mecánica estadística

A.1. Teoría de distribuciones

A.1.1. Delta de Dirac

La delta de Dirac es una distribución en el sentido matemático [72], se escribe como $\delta(\)$ y definida en el espacio de estados. Tiene las siguientes propiedades

$$\int dx \delta(x - y) = 1 \quad (\text{A.1})$$

$$\int dx \delta(x - y)f(x) = f(y) \quad (\text{A.2})$$

$$\int dx \delta'(x - y)f(x) = f'(y) \quad (\text{A.3})$$

A.2. Fluctuaciones e incerteza

Una cantidad macroscópica f puede ser vista como una expectación (2.7), y corresponden al valor esperado obtenido al medir una propiedad microscópica F ,

$$f = \langle F \rangle_I = \sum_j p_j F_j. \quad (\text{A.4})$$

Esta cantidad macroscópica es el resultado de una estimación, y tiene asociada una una incerteza bajo un estado de conocimiento. Se define la incerteza en la cantidad f a través de su varianza,

$$(\Delta f)^2 = \left\langle \left(F - \langle F \rangle \right)^2 \right\rangle. \quad (\text{A.5})$$

$$\begin{aligned} &= \left\langle \left(F - f \right)^2 \right\rangle. \\ &= \left\langle F^2 \right\rangle - \left\langle F \right\rangle^2. \\ &= \left\langle F^2 \right\rangle - f^2. \end{aligned} \quad (\text{A.6})$$

Esta incerteza puede ser debida a una falta de conocimiento o a la propia fluctuación de la cantidad, causada por la dinámica del sistema.

Si $F = F(\tilde{x}_t)$, entonces podemos definir su promedio temporal como

$$\bar{F} = \frac{1}{\tau} \int_0^\tau dt F(\tilde{x}_t), \quad (\text{A.7})$$

y la fluctuación

$$\left(\Delta \bar{F} \right)^2 = \frac{1}{\tau} \int_0^\tau dt \left(F(\tilde{x}_t) - \bar{F} \right)^2. \quad (\text{A.8})$$

Apéndice B

Ecuación de Chapman-Kolmogorov

B.1. Ecuación de Chapman-Kolmogorov

Con la propiedad de la marginalización y la regla del producto, es posible derivar una de las relaciones más importantes para entender la evolución de una función de probabilidad para estados x_0 que evolucionan a estados posteriores x_1 . Comenzando con la regla de Marginalización (2.11) para la probabilidad conjunta de los microestados x_0 y x_1 , escribiremos

$$P(x_1|I) = \int dx_0 P(x_0, x_1|I) \tag{B.1}$$

$$= \int dx_0 P(x_0|I) P(x_1|x_0, I). \tag{B.2}$$

Esta ecuación nos permite entender cómo se evoluciona desde una probabilidad $P(x_0|I)$ a una probabilidad $P(x_1|I)$, mediante el uso de una probabilidad de transición $P(x_1|x_0, I)$. Esta relación llamada ecuación de Chapman-Kolmogorov es, como veremos, una consecuencia directa de las leyes de la probabilidad.

La Ecuación de Chapman-Kolmogorov, puede utilizarse como una ecuación dinámica, para hacer evolucionar una probabilidad en el tiempo, ya que es posible obtener

la n -ésima probabilidad $P(x_n|I)$, dada una probabilidad inicial $P(x_0|I)$, mediante una relación de recurrencia de la forma

$$P(x_n|I) = \int dx_0 \dots dx_{n-1} P(x_0|I) P(x_0|x_1) \dots P(x_{n-1}|x_{n-2}, I). \quad (\text{B.3})$$

De forma análoga, es posible encontrar la probabilidad previa $P(x_0|I)$, dada la probabilidad posterior $P(x_1|I)$, mediante el uso de una probabilidad de transición inversa $P(x_0|x_1, I)$, de la forma

$$P(x_1|I) = \int dx_0 P(x_0, x_1|I) \quad (\text{B.4})$$

$$= \int dx_0 P(x_0|I) P(x_1|x_0, I) \quad (\text{B.5})$$

Utilizando la Ecuación de Chapman-Kolmogorov “invertida” como ecuación dinámica, para realizar la inferencia sobre cual era la probabilidad en un instante $(-k)$ -ésimo previo, se obtiene una relación de recurrencia de la forma

$$P(x_n|I) = \int dx_0 \dots dx_{-n-1} P(x_0|I) P(x_0|x_{-1}) \dots P(x_{-n-1}|x_{-n-2}, I). \quad (\text{B.6})$$

Existen distintos caminos para obtener la forma de la probabilidad de transición $P(x_1|x_0, I)$, mediante el uso de MaxEnt y también con MaxCal relacionando la probabilidad de transición con el funcional de probabilidad.

Algunos autores, entregan a la ecuación de Chapman-Kolmogorov un carácter de ecuación fundamental, derivando a partir de ella las ecuaciones diferenciales de la NESM como lo utiliza A. Caticha en Entropic Dynamics para derivar la ecuación de Fokker-Planck [14, 50, 81].

Apéndice C

Teoremas para la mecánica estadística en equilibrio y fuera del equilibrio

C.1. Teorema de la Variable Conjugada (CVT)

El teorema de la variable conjugada (CVT por su nombre en inglés *conjugate variable theorem*) [82], es una identidad matemática de las probabilidades en sistemas continuos que permite obtener relaciones entre los valores de expectación de cantidades arbitrarias. Este teorema es una consecuencia del teorema de la divergencia (o más generalmente, del teorema de Stokes) del cálculo infinitesimal en varias variables y la definición de expectación, siendo posible entender este teorema como una regla de conmutación entre el operador derivada y el operador expectación.

Dada la definición de expectación en el espacio real

$$\langle w \rangle_I = \int dx P(x|I) w(x),$$

donde $w = w(x)$, si la función a la cual se calcula su expectación puede ser escrita como una derivada, es decir $w = \frac{\partial v}{\partial x}$ existe para todo valor en que $P(x|I) > 0$, se

tiene

$$\left\langle \frac{\partial v}{\partial x} \right\rangle_I = \int dx P(x|I) \frac{\partial v}{\partial x},$$

y luego utilizando la regla de integración por partes $\int_a^b P dv = P v|_a^b - \int_a^b v dP$, para reescribir el lado derecho de la identidad, tenemos

$$\left\langle \frac{\partial v}{\partial x} \right\rangle_I = P(x|I) v \Big|_{x=-\infty}^{x=+\infty} - \int dx v \frac{\partial P(x|I)}{\partial x}.$$

reescribiendo queda entonces

$$\frac{\partial P(x|I)}{\partial x} = P(x|I) \frac{\partial \log P(x|I)}{\partial x}, \quad (\text{C.1})$$

condición válida para todo $P(x|I) > 0$, y luego rearmando la expectativa en el lado derecho,

$$\begin{aligned} \left\langle \frac{\partial v}{\partial x} \right\rangle_I &= P(x|I) v \Big|_{x=-\infty}^{x=+\infty} - \int dx P(x|I) v \frac{\partial \log P(x|I)}{\partial x}. \\ &= - \left\langle v \frac{\partial \log P(x|I)}{\partial x} \right\rangle_I. \end{aligned} \quad (\text{C.2})$$

Para el resultado final, se supone que la probabilidad es nula en los bordes $\pm\infty$. Si la probabilidad está definida entre límites finitos $x \in [a, b]$, donde la probabilidad no es necesariamente nula, entonces el teorema tendrá un término extra el cual representa condiciones de borde,

$$\left\langle \frac{\partial v}{\partial x} \right\rangle_I = P(x|I) v \Big|_{x=a}^{x=b} - \left\langle v \frac{\partial \log P(x|I)}{\partial x} \right\rangle_I. \quad (\text{C.3})$$

C.2. Teorema de la Variable Conjugada para Funcionales (CVTF)

Para derivar CVT en el contexto de funcionales, se puede realizar de forma análoga a la derivación anterior. En este caso, derivaremos la identidad a partir de la regla

del producto para la derivada funcional

$$\frac{\delta}{\delta \tilde{x}_t} \left(P[\tilde{x}|I] W[\tilde{x}] \right) = P[\tilde{x}|I] \frac{\delta}{\delta \tilde{x}_t} W[\tilde{x}] + W[\tilde{x}] \frac{\delta}{\delta \tilde{x}_t} P[\tilde{x}|I]. \quad (\text{C.4})$$

Integrando sobre el espacio de funciones, análogo al ejemplo anterior, al ser $P[\tilde{x}]$ un funcional de probabilidad nulo en la superficie del espacio, se obtiene que

$$0 = \int d\tilde{x} P[\tilde{x}|I] \frac{\delta}{\delta \tilde{x}_t} W[\tilde{x}] + \int d\tilde{x} W[\tilde{x}] \frac{\delta}{\delta \tilde{x}_t} P[\tilde{x}|I]. \quad (\text{C.5})$$

Ahora, escribiendo la derivada funcional de P de la forma

$$\frac{\delta}{\delta \tilde{x}_t} P[\tilde{x}|I] = P[\tilde{x}|I] \frac{\delta}{\delta \tilde{x}_t} \log P[\tilde{x}|I], \quad (\text{C.6})$$

combinando (C.5) con (C.6) y luego usando la definición de expectación en el espacio de trayectorias, se obtiene finalmente

$$\left\langle \frac{\delta}{\delta \tilde{x}_t} W[\tilde{x}] \right\rangle = - \left\langle W[\tilde{x}] \frac{\delta}{\delta \tilde{x}_t} \ln P[\tilde{x}] \right\rangle. \quad (\text{C.7})$$

Este teorema (C.7), tiene un funcional libre, o funcional de prueba W el cual puede ser dependiente además de n parámetros, por lo que con este teorema es posible relacionar cantidades macroscópicas fuera del equilibrio.

C.3. Teorema de Fluctuación-Disipación (FDT)

Otro teorema importante en el contexto de una teoría de probabilidad para funcionales es el teorema de fluctuación-disipación. Este teorema puede tener diferentes formas dependiendo del contexto en el cual se trabaje, pero su principal característica es entregar una relación de conmutación entre la derivada con respecto a los parámetros de una probabilidad, con el operador expectación. Esto es, si tenemos dos funciones $P := P(x, t)$ y $W := W(x, t)$, donde $x \in \mathbb{R}^n$ y $t \in \mathbb{R}$, entonces, la regla

del producto para la derivada temporal queda escrita como

$$\frac{\partial}{\partial t} (P W) = P \frac{\partial}{\partial t} W + W[\tilde{x}] \frac{\partial}{\partial t} P. \quad (\text{C.8})$$

Integrando sobre el espacio donde se define la probabilidad $\int dx$, se obtiene

$$\frac{\partial}{\partial t} \int dx (P W) = \int dx P \frac{\partial}{\partial t} W + \int dx W \frac{\partial}{\partial t} P, \quad (\text{C.9})$$

y cambiando la derivada de P de acuerdo a la identidad

$$\frac{\partial P(x, t|I)}{\partial t} = P(x, t|I) \frac{\partial \log P(x, t|I)}{\partial t}, \quad (\text{C.10})$$

se obtiene el llamado teorema de fluctuación-disipación,

$$\frac{\partial}{\partial t} \int dx (P W) = \int dx P \frac{\partial}{\partial t} W + \int dx P W \frac{\partial}{\partial t} \log P, \quad (\text{C.11})$$

el cual escrito en forma de expectativas, queda como

$$\frac{\partial}{\partial t} \langle W \rangle_{I,t} = \left\langle \frac{\partial}{\partial t} W \right\rangle_{I,t} + \langle W \frac{\partial}{\partial t} \log P \rangle_{I,t}. \quad (\text{C.12})$$

C.4. Teorema Fluctuación-Disipación para Funcionales (FDTF)

Análogo al desarrollo anterior, si tenemos dos funcionales $P = P[\tilde{x}; \tilde{\lambda}]$ y $W = W[\tilde{x}; \tilde{\lambda}]$, donde $\tilde{\lambda} \in \mathcal{X}$ es un parámetro o multiplicador de Lagrange del espacio de funciones, la regla del producto para la derivada funcional con respecto a $\tilde{\lambda}_t$ es de la forma

$$\frac{\delta}{\delta \tilde{\lambda}_t} (P W) = P \frac{\delta}{\delta \tilde{\lambda}_t} W + W[\tilde{x}] \frac{\delta}{\delta \tilde{\lambda}_t} P. \quad (\text{C.13})$$

Integrando sobre el espacio de funciones con $\int d\tilde{x}$, se obtiene

$$\frac{\delta}{\delta \tilde{\lambda}_t} \int d\tilde{x} (P W) = \int d\tilde{x} P \frac{\delta}{\delta \tilde{\lambda}_t} W + \int d\tilde{x} W \frac{\delta}{\delta \tilde{\lambda}_t} P, \quad (\text{C.14})$$

y cambiando la derivada funcional de P de la misma forma que en (C.6), tenemos que

$$\frac{\delta}{\delta \tilde{\lambda}_t} \left(\int d\tilde{x} P W \right) = \int d\tilde{x} P \frac{\delta}{\delta \tilde{\lambda}_t} W + \int d\tilde{x} P W \frac{\delta}{\delta \tilde{\lambda}_t} \log P. \quad (\text{C.15})$$

Aplicando la definición de expectación para funcionales, se obtiene finalmente

$$\frac{\delta}{\delta \tilde{\lambda}_t} \langle W \rangle_I = \left\langle \frac{\delta}{\delta \tilde{\lambda}_t} W \right\rangle_I + \left\langle W \frac{\delta}{\delta \tilde{\lambda}_t} \log P \right\rangle_I. \quad (\text{C.16})$$

Esta identidad (C.16), tiene una utilidad similar a la de CVT (C.7), relacionando cantidades macroscópicas para una teoría de probabilidad para funcionales. El uso en conjunto de ambos teoremas puede llevarnos a entender una teoría en que se estudian las relaciones entre cantidades macroscópicas fuera del equilibrio, como lo es la termodinámica.

Para el caso en que los multiplicadores de Lagrange o parámetros no son funciones, sino valores reales, es posible derivar una identidad análoga, solo que con derivadas parciales, en lugar de derivadas funcionales. Si $P = P[\tilde{x}; \lambda]$ y $W = W[\tilde{x}; \lambda]$, donde $\lambda \in \mathbb{R}^n$, también podremos escribir de manera análoga a (C.16),

$$\frac{\partial}{\partial \lambda} \langle W \rangle_I = \left\langle \frac{\partial}{\partial \lambda} W \right\rangle_I + \left\langle W \frac{\partial}{\partial \lambda} \log P \right\rangle_I. \quad (\text{C.17})$$

Referencias

- [1] H. B. Callen, *Thermodynamics and Statistical Mechanics* (1985).
- [2] E. Fermi, *Thermodynamics* (1956).
- [3] W. Greiner, L. Neise y H. Stöcker, *Thermodynamics and Statistical Mechanics* (Springer, 2001).
- [4] R. Clausius, *Annalen der Physik und Chemie* **79**, 368 (1850).
- [5] R. Clausius, *Annalen der Physik und Chemie* **125**, 351 (1865).
- [6] L. N. W. Greiner y H. Stöcker, *Thermodynamics and Statistical Mechanics* (2012).
- [7] R. Kubo, *Statistical Mechanics: An Advanced Course with Problems and Solutions* (1990).
- [8] K. Huang, *Statistical Mechanics* (1987).
- [9] R. K. Pathria, *Statistical Mechanics* (1972).
- [10] J. W. Gibbs, *Transactions of the Connecticut Academy of Arts and Sciences* págs. 108–248 and 343–524 (1874-78).
- [11] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (1902).

- [12] E. T. Jaynes, *Physical Review* **106**, 620 (1957).
- [13] E. T. Jaynes, *Physical Review* **108**, 171 (1957).
- [14] A. Caticha, *Lectures on Probability, Entropy, and Statistical Physics*. (arXiv:physics.class-ph/0710.1071v1, 2008).
- [15] P. Attard, *Thermodynamics and Statistical Mechanics: Equilibrium by Entropy Maximisation* (Academic Press, 2002).
- [16] J. G. R. Roberto Luzzi, Áurea R. Vasconcellos, *Predictive Statistical Mechanics: A Nonequilibrium Ensemble Formalism* (Springer, 2002).
- [17] L. Landau y E. Lifshitz, *Statistical Physics* (Elsevier, 1951).
- [18] S. R. de Groot y P. Mazur, *Non-equilibrium Thermodynamics* (Dover Publications Inc., 1985).
- [19] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, 2001).
- [20] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, 1996).
- [21] C. Jarzynski, *Physical Review Letters* **78**, 2690 (1996).
- [22] G. E. Crooks, *Physical Review E* **60**, 2721 (1999).
- [23] G. Crooks y E. Gavin, *Physical Review E* **61**, 2361 (2000).
- [24] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. T. Jr. y C. Bustamante, *Nature* **437**, 231 (2005).

- [25] J. Kurchan, *Journal of Physics A: Mathematical and General* **31**, 3719 (apr 1998).
- [26] D. J. Searles y D. J. Evans, *International Journal of Thermophysics* **22**, 123 (2001).
- [27] Y. Yura, H. Takayasu, D. Sornette y M. Takayasu, *Phys. Rev. Lett.* **112**, 098703 (2014).
- [28] E. T. Jaynes, *Ann. Rev. Phys. Chem.* **31**, 579 (1980).
- [29] M. Kac, *Trans. Amer. Math. Soc.* **65**, 1 (1949).
- [30] R. P. Feynman y A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 2005).
- [31] S. Pressé, K. Ghosh, J. Lee y K. A. Dill, *Reviews of Modern Physics* **85**, 1115 (2013).
- [32] M. J. Hazoglou, V. Walther, P. D. Dixit y K. A. Dill, *The Journal of Chemical Physics* **143**, 051104 (2015).
- [33] C. Lanczos, *The Variational Principles of Mechanics* (Dover Books, 1970).
- [34] R. T. Cox, *The Algebra of Probable Inference* (John Hopkins, 1961).
- [35] D. S. Sivia y J. Skilling, *Data Analysis: A Bayesian Tutoria* (OUP Oxford, 2006).
- [36] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).

- [37] A. N. Kolmogorov, *Ergebnisse der Math.* **2** (1933).
- [38] W. M. Bolstad, *Introduction to Bayesian Statistics* (J. Wiley, 2007).
- [39] R. T. Cox, *Am. J. Phys.* **14** (1946).
- [40] P. Goyal, K. H. Knuth y J. Skilling, *AIP Conference Proceedings* **1193** (2009).
- [41] S. Davis, arXiv:1607.07890v1 (2016).
- [42] A. Caticha, *Entropy* **17**, 6110 (2015).
- [43] S. Davis, arXiv:1503.03451 (2015).
- [44] J. Skilling, en *Maximum Entropy and Bayesian methods in Science and Engineering*, editado por G. J. Erickson y C. R. Smith (Kluwer Academic Publishers, 1988), págs. 173–187.
- [45] P. Goyal, A. Giffin, K. H. Knuth y E. Vrscay, *Bayesian inference and maximum entropy methods in science and engineering: proceedings of the 32th international workshop on bayesian inference and maximum entropy methods in science and engineering (maxent 2012)* (AIP Conference Proceedings, 2012).
- [46] G. Verdoolaege, *Bayesian inference and maximum entropy methods in science and engineering: proceedings of the 36th international workshop on bayesian inference and maximum entropy methods in science and engineering (maxent 2016)* (AIP Conference Proceedings, 2016).
- [47] C. Shannon, *Bell System Technical Journal* **27**, 379 (1948).
- [48] T. M. Cover y J. A. Thomas, *Elements of Information Theory* (John Wiley and Sons, 2006).

- [49] J. E. Shore y R. W. Johnson, IEEE Trans. Info. Theory **IT-26**, 26 (1980).
- [50] A. Caticha y A. Giffin, AIP Conf. Proc. **872**, 31 (2006).
- [51] L. D. Elsgolc, *Calculus of Variations* (Dover, 2007).
- [52] A. Caticha y R. Preuss, AIP Conf. Proc. **707**, 371 (2004).
- [53] L. R. S. Chibbaro y A. Vulpiani, Comm. Theor. Phys. **62**, 469 (2014).
- [54] R. C. Dewar, Entropy **11**, 931 (2009).
- [55] B. H. Lavenda, *Statistical Physics: A Probabilistic Approach* (John Wiley and Sons, INC., 1991).
- [56] S. Amari, *Information Geometry and Its Applications* (Springer Japan, 2016).
- [57] H. Goldstein, *Classical Mechanics* (1950).
- [58] I. M. Gelfand y S. V. Fomin, *Calculus of Variations* (Dover Publications, 2000).
- [59] R. P. Feynman y L. Brown, *Feynman's Thesis: A New Approach to Quantum Theory* (World Scientific Publishing Co. Pte. Ltd., 2005).
- [60] S. Machlup y L. Onsager, Physical Review **91**, 1512 (1953).
- [61] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets* (World Scientific Publishing, 2005).
- [62] J. R. Ray, Phys. Rev. A **44**, 4061 (1991).
- [63] B. D. Butler, G. Ayton, O. G. Jepps y D. J. Evans, J. Chem. Phys. **109**, 6519 (1998).

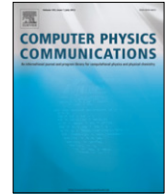
- [64] D. González, S. Davis y G. Gutiérrez, *Found. of Phys.* **44**, 923 (2014).
- [65] D. González y S. Davis, *J. Phys.: Conf. Ser.* **720**, 012006 (2016).
- [66] C. G. Gray y E. F. Taylor, *American Journal of Physics* **75** (2007).
- [67] D. González, D. Diaz y S. Davis, *The Eur. Phys. J. B* **89**, 214 (2016).
- [68] S. Davis, D. Gonzalez y G. Gutierrez, *Entropy* **20**, 696 (2018).
- [69] H. Haken, *Z. Phys. B. Cond. Matt.* **63**, 505 (1986).
- [70] C. Cohen-Tannoudji, B. Diu y F. Laloë, *Quantum mechanics* (1977).
- [71] E. M. Purcell, *Electricidad y magnetismo* (1988).
- [72] J. E. Marsden y A. J. Tromba, *Cálculo Vectorial* (Addison-Wesley Iberoamericana, 1991).
- [73] R. Soto, *Kinetic Theory and Transport Phenomena* (2016).
- [74] E. Nelson, *Phys. Rev.* **150**, 1079 (1966).
- [75] E. Roldán, *Irreversibility and Dissipation in Microscopic Systems* (Springer International Publishing, 2014).
- [76] D. J. Evans y D. J. Searles, *Advances in Physics* **51**, 1529 (2002).
- [77] D. J. Evans, D. J. Searles y S. R. Williams, *The Journal of chemical physics* **128**, 014504 (jan 2008).
- [78] C. Bustamante, J. Liphardt y F. Ritort, *Physics Today* **58**, 43 (2005).
- [79] D. Gonzalez y S. Davis, *AIP Conference Proceedings* **1757** (2016).

- [80] D. Gonzalez y S. Davis, AIP Conference Proceedings **1853** (2017).
- [81] A. Caticha y C. Cafaro, arXiv:physics.class-ph/0710.1071v1 (2007).
- [82] S. Davis y G. Gutiérrez, Phys. Rev. E **86**, 051136 (2012).

Apéndice D

Publicaciones

Durante el estudio de doctorado del autor se obtuvo como resultado las siguientes publicaciones en revistas científicas.



Fitting of interatomic potentials without forces: A parallel particle swarm optimization algorithm



Diego González*, Sergio Davis

Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile

ARTICLE INFO

Article history:

Received 29 January 2014
Received in revised form
24 July 2014
Accepted 27 July 2014
Available online 12 August 2014

Keywords:

Particle swarm optimization
Interatomic potential
Fitting

ABSTRACT

We present a methodology for fitting interatomic potentials to *ab initio* data, using the particle swarm optimization (PSO) algorithm, needing only a set of positions and energies as input. The prediction error of energies associated with the fitted parameters can be close to 1 meV/atom or lower, for reference energies having a standard deviation of about 0.5 eV/atom. We tested our method by fitting a Sutton–Chen potential for copper from *ab initio* data, which is able to recover structural and dynamical properties, and obtain a better agreement of the predicted melting point versus the experimental value, as compared to the prediction of the standard Sutton–Chen parameters.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

In Condensed Matter Physics, the task of obtaining different mechanical properties of materials, simulated atomistically with a large number of atoms under *ab initio* methods, is an almost prohibitive one, in terms of computational effort with the current computer architectures. It might even at times be impossible. Because of this, producing a “classical” interatomic potential as a substitute for the genuine quantum-mechanical interaction of the particles is highly desirable. The usual procedure is to fit some empirical interatomic potential function, depending on N parameters, requiring either agreement with certain macroscopic properties (structural, thermodynamical, etc.) or simply agreement between the predicted and observed energies and atomic forces. A standard algorithm based on force information is the force matching method [1,2].

In this work we present a methodology for fitting interatomic potentials to *ab initio* data, using the particle swarm optimization (PSO) algorithm [3]. The objective function to be minimized is the total prediction error in the energies for the configurations provided, thus the algorithm does not require any information besides the atomic positions for each configuration and their corresponding *ab initio* energies. In particular it does not require the atomic forces, as in other fitting procedures such as force matching methods.

2. Interatomic potential models

We implemented two families of interatomic potentials, pair potentials and embedded atom potentials. Among the former, we tested the well-known Lennard-Jones potential [4], given by

$$V(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \quad (1)$$

and the 6-parameters “generic” potential as implemented in Moldy [5],

$$V(r) = A \exp(-Br) + \frac{C}{r} - \frac{D^4}{r} - \frac{E^6}{r} - \frac{F^8}{r}. \quad (2)$$

From the family of embedded atom potentials [6], having the general form

$$E_i = \frac{1}{2} \sum_{j \neq i} \phi(|\mathbf{r}_i - \mathbf{r}_j|) + F \left(\sum_{j \neq i} \psi(|\mathbf{r}_i - \mathbf{r}_j|) \right) \quad (3)$$

we implemented the Sutton–Chen potential, where the pair functions and the embedding function are given by

$$F(\rho) = \epsilon C \sqrt{\rho} \quad (4)$$

$$\phi(r) = \epsilon (a/r)^n \quad (5)$$

$$\psi(r) = (a/r)^m. \quad (6)$$

* Corresponding author. Tel.: +56 229787280.

E-mail addresses: dgonzalez@gnm.cl (D. González), sdavis@gnm.cl (S. Davis).

3. Particle swarm optimization

The particle swarm optimization (PSO) algorithm is based on the idea of distributing the search procedure among a large number of “agents”, which act independently of each other. Each agent moves through the search space with a simple dynamics, reacting to fictitious forces drawing it towards its own *current best* solution and the *global best* solution for the whole swarm. In this way, when an agent finds a better solution than the current global best, it becomes the new global best and all the other agents react instantly, the swarm is directed towards the new solution.

For a set of n particles represented by their positions $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, the velocity for the i th particle and the k th step is

$$\mathbf{v}_i^k = \omega \mathbf{v}_i^{k-1} + c_1 r_1^k (\mathbf{x}_B - \mathbf{x}_i^{k-1}) + c_2 r_2^k (\mathbf{x}_G - \mathbf{x}_i^{k-1}) \quad (7)$$

and the position is given by

$$\mathbf{x}_i^k = \mathbf{x}_i^{k-1} + \mathbf{v}_i^k. \quad (8)$$

We employed the following choice of PSO parameters: $\omega = 0.7$, $c_1 = 1.4$ and $c_2 = 1.4$, after a few trial convergence runs.

4. Implementation of the fitting algorithm

For a potential function where we wish to find the parameters a_0, a_1, \dots, a_m from a set of positions \mathbf{r}_i^j and energies E_j satisfying the relation

$$V(\mathbf{r}_1^j, \mathbf{r}_2^j, \dots, \mathbf{r}_n^j; \mathbf{a}) = E_j \quad (9)$$

with

$$\mathbf{a} = (a_0, a_1, \dots, a_m),$$

we can define an objective function which is just the total prediction squared error, of the form

$$f(\mathbf{a}) = \sum_j \left(V(\mathbf{r}_1^j, \mathbf{r}_2^j, \dots, \mathbf{r}_n^j; \mathbf{a}) - E_j \right)^2, \quad (10)$$

and then for the set of parameters \mathbf{a}^* that correctly fit the potential V we have $f(\mathbf{a}^*) = 0$.

Then the problem may be solved numerically with the PSO algorithm minimizing the function $f(\mathbf{a})$.

4.1. Optimization of the algorithm

We have included some improvements on the PSO implementation, particular to our problem. For instance, we perturbed the swarm every time the procedure gets stuck in a minimum for N_S steps (N_S proportional to the number of parameters d in the potential, usually $N_S = 50d$), completely randomizing their positions.

On the other hand, we exploit the fact that for several families of potentials there is a scale parameter for the interatomic distance, let us call it σ , such that the potential depends on r only through r/σ . This is the case for the σ parameter in the Lennard-Jones potentials, for the C, D, E, F and $1/B$ parameters in the generic potential from Moldy, and also for the a parameter in the Sutton–Chen variant of the embedded atom potentials. This distance scale parameter can be constrained to be between the minimum observed distance and a multiple of this value (typically 10 times), which considerably reduces the search space.

Parallelization was achieved simply by distributing the PSO particles evenly among the different processors using the message passing interface (MPI) framework, at each step sharing the global best between all processors.

5. Results

5.1. Lennard-Jones potential

In order to test the consistency of our procedure, we randomly generate [7] a set of 20 configurations and we compute their energy according to the standard Lennard-Jones parameters for argon, $\epsilon = 0.0103048$ eV and $\sigma = 3.41$ Å.

The resulting set has a standard deviation of energy of 0.41063 eV. Then, with the information of positions and energies (in a parallel run using an AMD Opteron 6272, 2.1 GHz, shared-memory machine with 64 cores and 500 PSO particles), the time needed to find the minimum prediction error was 212.6 s. We can see that the algorithm converge quickly for each parameter, recovering their exact values at 1300 steps (the prediction error reached is below 10^{-27} meV/atom, see Figs. 1 and 2).

5.2. 6-parameter generic potential

For the 6-parameters pair potential using the same set of positions and energies obtained for the previous Lennard-Jones test, the time needed to find the minimum prediction error was 3159.9 s, using the same 64-core machine and 500 PSO particles. In this case the error for the converged set of parameters falls below 8×10^{-2} meV/atom at 9000 steps.

5.3. Embedded atom potential

We repeated the same approach for the embedded atom potential, this time using the standard Sutton–Chen parameters for copper, $\epsilon = 0.0123820$ eV, $a = 3.61$ Å, $n = 9$, $m = 6$ and $C = 39.432$. We used 4 configurations as input, and we stopped the minimization procedure after 193 015 steps (execution time was 23 h using the same 64-core machine as the section before, and using 800 PSO particles), when we reached a prediction error of about 0.8 meV/atom and the following fitted parameters: $\epsilon = 0.0145749$ eV, $a = 3.5834$ Å, $n = 8.82683$, $m = 5.67465$, and $C = 37.028$ (see Figs. 3 and 4).

As in this case the fitting error in energies is not negligible, we assessed the ability of the potential for reproducing atomic forces on each individual atom. We found an average error of 4.64%, where the average was taken over the individual relative (in percentage) errors for each atom. However, if we manually correct the magnitude of each force to match the correct one, the average error falls to 1.36%. This means about 3% of the total error in the forces is just a propagation of the error in reproducing the correct energy scale (the parameter ϵ in this case).

6. Application: an embedded atom potential for copper from *ab initio* data

In order to test our procedure on a more realist scenario and assess the quality of the fitted potentials we performed *ab initio* microcanonical molecular dynamics simulations of copper at different temperatures (covering its solid, liquid and superheated phases) but at the same room pressure (lattice constant $a = 3.61$ Å). All molecular dynamics calculations were performed using Density Functional Theory (DFT) as implemented in VASP [8]. We used Perdew–Burke–Ernzerhof (PBE) generalized gradient approximation (GGA) pseudopotentials [9] with an energy cutoff of 204.9 eV and k -point expansion around the Γ point only.

From these simulations, we generated 13 229 different atomic configurations with their respective energies, mixed from solid ($T = 738$ K), liquid ($T = 2716$ K) and superheated state ($T = 2058$ K) simulations. Among them we chose a subset of 30 with maximum standard deviation of the energy (namely 0.24

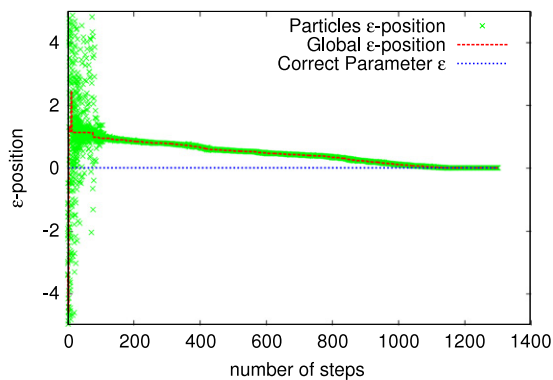


Fig. 1. (Color online) Evolution of the ϵ coordinate for the case of a Lennard-Jones potential as a function of optimization step.

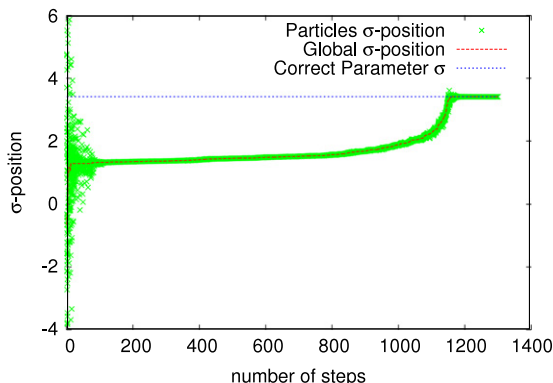


Fig. 2. (Color online) Evolution of the σ coordinate for the case of a Lennard-Jones potential as a function of optimization step.

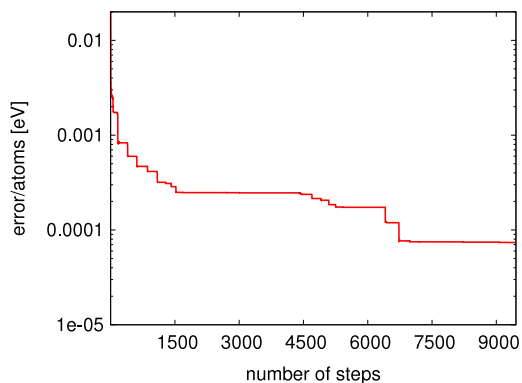


Fig. 3. Prediction error (meV) as a function of optimization steps for the case of a 6-parameter pair potential.

eV/atom), in order to increase the transferability of the fitted potential. These configurations were used as input to the fitting procedure. We found the Sutton–Chen potential parameters presented in Table 1, with a prediction error of 5.19 meV/atom (see Figs. 1–4).

As we did with the EAM fit in the previous section, we assessed the ability of the *ab initio*-fitted potential for reproducing atomic forces on each individual atom. In this case we found an average error of 22.19%, which falls to 12.53% when correcting for the magnitude of each individual force. The remaining error, of about 10%, is then due to the error in the parameter ϵ . This increase in the size of the error when attempting to reproduce individual force scales with the size of the error in energy (0.8 meV versus 5.19 meV).

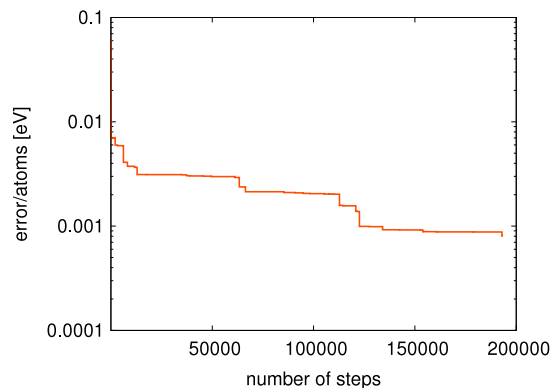


Fig. 4. Prediction error (meV) as a function of optimization steps for the case of an embedded-atom potential.

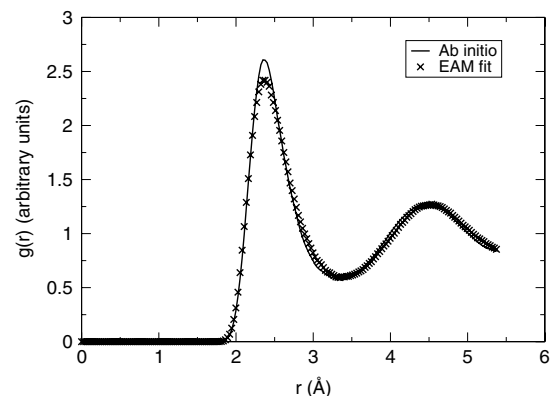


Fig. 5. Radial distribution function $g(r)$ for liquid copper at $T = 2070$ K computed from *ab initio* data and from the embedded atom potential fitted with our method.

We tested these parameters by performing classical molecular dynamics simulations using the LPMD [10] code, with a $4 \times 4 \times 4$ FCC simulation cell (256 atoms, $a = 3.61$ Å). Fig. 5 shows the radial distribution function $g(r)$ produced by our fitted Cu potential for liquid at $T = 2070$ K, together with the one computed from *ab initio* data. It reproduces the positions and widths of minima, but slightly underestimating the height of the first maximum. This $g(r)$ also agrees with the features found in a previous *ab initio* fitting study [11].

Fig. 6 shows the mean square displacement for liquid at $T = 2070$ K. From this we obtained a diffusion coefficient $D = 0.448$ Å²/ps, close to the value of $D = 0.4923$ Å²/ps obtained directly from the *ab initio* data. As a reference, the experimental value reported by Meyer [12] is $D = 0.52 \pm 0.01$ Å²/ps at $T = 1620$ K.

The quality of the potential in reproducing thermal properties was assessed by computing the melting point, using the micro-canonical Z-method [13–15]. In this method, for constant volume the $T(E)$ curve is drawn by performing different molecular dynamics simulations at different initial kinetic energies (in every simulation the system starts with the ideal crystalline configuration). The discontinuity in the isochore signals the melting point.

Fig. 7 shows the isochoric curve for different energies around the melting point, where the lowest point of the rightmost branch correspond to an upper estimate of the melting temperature T_m , in our case approx. 1700 K (the experimental value is $T_m = 1356.6$ K). The highest point is the critical superheating temperature T_{LS} , around $T = 2020$ K. For comparison we also included the isochoric curve calculated with the potential parameters by Sutton and Chen, which gives T_m around 2000 K for the same system size and number of simulation steps.

Table 1
Sutton-Chen potential parameters for Cu, fitted from *ab initio* data.

Source	a (Å)	n	m	C	ϵ (eV)
Sutton and Chen	3.61	9.0	6.0	39.432	0.012382
Belonoshko et al.	3.270	9.05	5.005	33.17	0.0225
This work	3.34385	5.93853	2.13419	32.2332	0.0846903

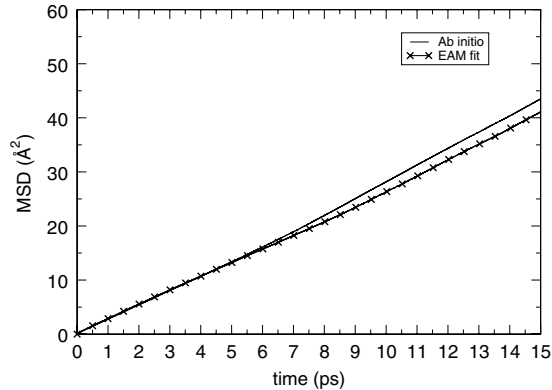


Fig. 6. Mean square displacement (MSD) for liquid copper at $T = 2070$ K computed from *ab initio* data and from the embedded atom potential fitted with our method.

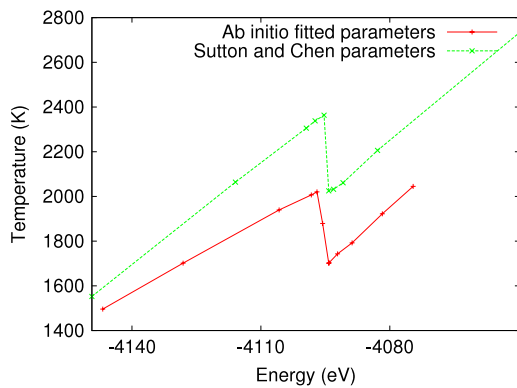


Fig. 7. (Color online) Isochoric curve (Z-curve) for copper according to our fitted potential parameters. The Z-method predicts a melting temperature $T_m = 1700$ K.

7. Concluding remarks

We have shown that it is possible to use a parallel algorithm based on particle swarm optimization to fit interatomic potentials to *ab initio* energies only.

Our procedure has been tested by fitting both pair potentials and embedded atom potentials, up to a prediction error of the order of 1 meV/atom, using between 5 and 30 different configurations. The implementation code is parallelized using message passing interface (MPI) libraries.

We demonstrated the capabilities of our method by fitting a set of Sutton-Chen parameters for copper using *ab initio* data from three thermodynamic phases. This fitted potential is able to reproduce the radial distribution function and the diffusion coefficient for liquid copper at $T = 2070$ K of the *ab initio* data, although it underestimates the diffusion coefficient with respect to experimental data. It also yields a better prediction of the melting point than the standard Sutton-Chen parameters.

Acknowledgments

SD gratefully acknowledges funding from VID Universidad de Chile and partial funding from FONDECYT 1140514. DG acknowledges the use of computer time at Grupo de Nanomateriales (GNM) and funding from CONICYT PhD fellowship 21140914.

References

- [1] F. Ercolessi, *Europhys. Lett.* 26 (1994) 583.
- [2] S. Izvekov, M. Parrinello, C.J. Burnham, G.A. Voth, *J. Chem. Phys.* 120 (2004) 10896.
- [3] J. Kennedy, R. Eberhart, *Proceedings of IEEE International Conference on Neural Networks*, Vol. IV, 1995, p. 1942.
- [4] M. Allen, D. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.
- [5] K. Refson, *Comput. Phys. Comm.* 126 (2000) 310.
- [6] M.S. Daw, S.M. Foiles, M.I. Baskes, *Mater. Sci. Rep.* 9 (1993) 251.
- [7] We generated random configurations by displacing the atoms of the ideal crystalline structure a small fraction of the lattice parameter.
- [8] G. Kresse, J. Furthmüller, *Phys. Rev. B* 54 (1996) 11169.
- [9] J.P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* 77 (1996) 3865.
- [10] S. Davis, C. Loyola, F. González, J. Peralta, *Comput. Phys. Comm.* 181 (2010) 2126.
- [11] A.B. Belonoshko, R. Ahuja, O. Eriksson, B. Johansson, *Phys. Rev. B* 61 (2000) 3838.
- [12] A. Meyer, *Phys. Rev. B* 81 (2010) 012102.
- [13] A.B. Belonoshko, N.V. Skorodumova, A. Rosengren, B. Johansson, *Phys. Rev. B* 73 (2006) 012201.
- [14] S. Davis, A.B. Belonoshko, A. Rosengren, A.C.T. van Duin, B. Johansson, *Cent. Eur. J. Phys.* 8 (2010) 789.
- [15] S. Davis, G. Gutiérrez, *Modelling Simul. Mater. Sci. Eng.* 21 (2013) 075001.

Newtonian Dynamics from the Principle of Maximum Caliber

Diego González · Sergio Davis ·
Gonzalo Gutiérrez

Received: 10 October 2013 / Accepted: 4 June 2014 / Published online: 11 July 2014
© Springer Science+Business Media New York 2014

Abstract The foundations of Statistical Mechanics can be recovered almost in their entirety from the principle of maximum entropy. In this work we show that its non-equilibrium generalization, the principle of maximum caliber (Jaynes, Phys Rev 106:620–630, 1957), when applied to the unknown trajectory followed by a particle, leads to Newton’s second law under two quite intuitive assumptions (both the expected square displacement in one step and the spatial probability distribution of the particle are known at all times). Our derivation explicitly highlights the role of mass as an emergent measure of the fluctuations in velocity (inertia) and the origin of potential energy as a manifestation of spatial correlations. According to our findings, the application of Newton’s equations is not limited to mechanical systems, and therefore could be used in modelling ecological, financial and biological systems, among others.

Keywords Newtonian mechanics · Maximum caliber

1 Introduction

Newtonian mechanics is a cornerstone of our civilization. From its simpler formulation, $F = ma$, to Lagrangian and Hamiltonian mechanics it has evolved an elegant mathematical apparatus around it. However, we still have to assume the principle of

D. González (✉) · S. Davis · G. Gutiérrez
Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias,
Universidad de Chile, Casilla 653, Santiago, Chile
e-mail: dgonzalez@gnm.cl

S. Davis
e-mail: sdavis@gnm.cl

G. Gutiérrez
e-mail: gonzalo@macul.ciencias.uchile.cl

minimum action as a postulate (that is, if we do not consider Newtonian mechanics as a limiting consequence of quantum mechanics).

There have been attempts to derive Newtonian dynamics from different principles. Particularly related to this work, it has been previously derived from information-geometric arguments leading to the idea of entropic dynamics [1]. This idea is based on the assumption of an irreducible uncertainty in the position of a particle, implying an information metric for space from which Newton's second law naturally emerges. Caticha and Cafaro's derivation is founded on Jaynes' maximum entropy principle [2] (MaxEnt for short), a generic rule for the construction of probabilistic models.

In 1957, Edwin T. Jaynes [2] postulated that Statistical Mechanics has to be understood, not as a physical theory in the same footing as, say, classical mechanics or electromagnetism, but as an application of statistical inference to a system with macroscopically large numbers of degrees of freedom. The question was reversed from "given the microscopic evolution of the system, what is the probability distribution for the macroscopic quantities?" to "given a few known macroscopic properties, what are the possible microstates compatible with said knowledge?". The answer, as initially proposed by Gibbs, was the probability distribution with maximum entropy $S = -\sum_i P_i \ln P_i$ subjected to constraints reflecting the known macroscopic properties. Jaynes, after the work of Shannon in information theory, realized that this procedure (maximization of S constrained only by the known information) is not limited to Statistical Mechanics but a valid principle in any problem of statistical inference. This is the principle of Maximum Entropy. Due to the uniqueness of Shannon's entropy in characterizing uncertainty, it is the most unbiased procedure for the construction of statistical models. Later, it has been axiomatically derived [3,4] from requirements of internal consistency, so its validity is independent of any meaning assigned to the quantity S .

A full understanding of the maximum entropy principle requires the adoption of the Bayesian viewpoint on probability: An "absolute" probability distribution $P(x)$ does not exist, only conditional probability distributions $P(x|H)$ where H represents a given state of knowledge. For a given H , MaxEnt gives us the most unbiased model we can propose with our (possibly limited) information, therefore we cannot guarantee beforehand that it is the correct model. As we acquire more and more information, $P(x|H)$ becomes increasingly more independent of H , and for all practical purposes the distribution becomes objective (independent of the observer).

The principle of Maximum Caliber [5,6] generalizes this idea to dynamical systems, including time explicitly. It has been applied recently to discrete dynamics [7] and earlier to derive the Fokker–Planck equations [8] and the Markov process formalism [9].

In this work we show that if we use the maximum caliber principle to find the unknown trajectory of a particle, there are two general conditions that automatically lead to Newton's second law, namely that (a) the expected square displacement per step is known at all times, and (b) that the time-independent probability of finding the particle at any coordinate is also known. Knowledge of both (a) and (b) leads to Newton's second law in expectation over trajectories, and what is perhaps more interesting, any dynamical system not following Newton's second law has to violate at least one of these assumptions.

2 The Maximum Entropy and Maximum Caliber Formalism

Consider a system with N degrees of freedom, whose states are denoted by vectors $\vec{x} = (x_1, \dots, x_N)$. Suppose the expectation values of M functions $f_i(\vec{x})$ are known. This represents a *Bayesian state of knowledge*, that we will denote by H . Maximization of the Shannon entropy

$$S = - \int d\vec{x} P(\vec{x}|H) \ln P(\vec{x}|H) \tag{1}$$

under the constraints $\langle f_i(\vec{x}) \rangle_H = F_i$ leads to the MaxEnt model

$$P(\vec{x}|H) = \frac{1}{Z(\vec{\lambda})} \exp\left(- \sum_{k=1}^M \lambda_k f_k(\vec{x})\right), \tag{2}$$

where the value of the Lagrange multipliers λ_k needed to impose the M constraints can be determined from

$$- \frac{\partial}{\partial \lambda_k} \ln Z(\vec{\lambda}) = F_k. \tag{3}$$

Equation 3, being a nonlinear equation is usually impractical to solve. Moreover, it needs the partition function explicitly. However, it has been recently shown [10] that for the Lagrange multipliers the equality

$$\langle \nabla \cdot \vec{v} \rangle = \sum_{k=1}^M \lambda_k \langle \vec{v} \cdot \nabla f_k \rangle \tag{4}$$

holds, with \vec{v} an arbitrary differentiable vector field, and this provides a linear system of equations for $\vec{\lambda}$. The most probable microstate \vec{x}_0 is such that

$$\nabla P(\vec{x}|H)|_{\vec{x}=\vec{x}_0} = 0,$$

therefore the function

$$G(\vec{x}) = \sum_{k=1}^M \lambda_k f_k(\vec{x})$$

which is a combination of all the constraints imposed, is also an extremum at \vec{x}_0 .

In the case of dynamical systems, we now ask for the possible microscopical trajectories compatible with some known information H (expressed as expectation values). The result is the probability distribution of trajectories $P[x(t)|H]$ which maximizes the Shannon entropy

$$\mathcal{S} = - \int \mathcal{D}x(t) P[x(t)|H] \ln P[x(t)|H] \tag{5}$$

given H . Note that S is now defined as a functional integral, as indicated by the notation $\mathcal{D}x(t)$. This conveys the idea of integration over a space of trajectories $x(t)$, as is commonly denoted in Feynman's path integral formulation of quantum mechanics [11].

The probability distribution functional for the different possible trajectories is

$$P[x(t)|H] = \frac{1}{Z[\lambda(t)]} e^{-\int dt \lambda(t) f[x(t); t]}, \quad (6)$$

where, similarly to Eq. 3, the Lagrange multiplier function can be obtained from

$$-\frac{\delta}{\delta \lambda(t)} \ln Z[\lambda(t)] = \langle f[x(t); t] \rangle. \quad (7)$$

Here the most probable trajectory extremizes a particular functional (analogous to an "action" in classical mechanics)

$$A[x(t)] = \int dt \lambda(t) f[x(t); t] \quad (8)$$

depending on the constraining function. This leads to the question: without introducing the Lagrangian of classical mechanics explicitly, could it *emerge* naturally from simpler constraints in a Maximum Caliber problem?

3 Derivation of Newton's Second Law

Consider a single particle following an unknown trajectory $x(t)$ in one spatial dimension. This can be easily generalized to many particles in arbitrary dimensions, at the cost of overcomplicated notation.

If we discretize time, such that the continuous variable t is sampled at intervals Δt and becomes $t_k = k \Delta t$ (with $k = 0, \dots, n-1$), the positions $x(t)$ have to be replaced with $x_k = x(t_k)$. Now the trajectory $x(t)$ itself becomes a vector $\vec{x} = (x_0, \dots, x_{n-1})$, and the Lagrange function $\lambda(t)$ also becomes a vector $\vec{\lambda} = (\lambda_0, \dots, \lambda_{n-1})$. This is the method of finite differences [12]. We can see that, using this discretized version, we recover Eqs. 2 and 3, therefore we can also employ the identity given by Eq. 4 for this kind of discretized Maximum Caliber problem.

We now impose the following constraints (expectations are to be interpreted over all possible trajectories)

$$\langle (x_i - x_{i-1})^2 \rangle = (\Delta t)^2 d_i^2 \quad (9)$$

$$\langle \delta(x_i - X) \rangle = P(x_i = X|H), \quad (10)$$

for all values of i and X . The first constraint recognizes the fact that the expected square displacement in one (possible infinitesimal) step is known for all times, and

is equal to an arbitrary function d_i^2 times the time step. We expressed it in this form so that d_i can remain finite when taking the limit $\Delta t \rightarrow 0$. The second constraint imposes that the static, time-independent probability distribution for the coordinate x is also known.

Here we must remark that, as we are considering the Bayesian view of probability, the last constraint can always be fulfilled: $P(x_i = X|H)$ is a model comprising our knowledge of x_i under the state H , not a physical property of the phenomenon. Different states of information H may assign different probability distributions $P(x_i = X|H)$, and the “quality” of this information will be reflected in the resulting potential energy function emerging from this formalism (as will be seen later).

The probability distribution function for \vec{x} is

$$P(\vec{x}|H) = \frac{1}{Z(\vec{\lambda})} \exp\left(-\sum_{i=0}^{n-1} \frac{\lambda_i}{(\Delta t)^2} (x_i - x_{i-1})^2 - \sum_{i=0}^{n-1} \int dX \mu(X) \delta(x_i - X)\right), \tag{11}$$

where λ_i and $\mu(X)$ are Lagrange multipliers associated to the constraints in Eqs. 9 and 10, respectively.

After integrating the Dirac delta function, Eq. 11 becomes

$$P(\vec{x}|H) = \frac{1}{Z(\vec{\lambda})} \exp\left(-\sum_{i=0}^{n-1} \frac{\lambda_i}{(\Delta t)^2} (x_i - x_{i-1})^2 - \sum_{i=0}^{n-1} \mu(x_i)\right). \tag{12}$$

This is the probability of the particle taking a well-defined discretized trajectory \vec{x} , and is precisely the solution of a Maximum Entropy problem with n degrees of freedom and n Lagrange multipliers λ_i (plus the function μ).

The most probable trajectory for the particle follows a minimum action principle. Indeed, if we define $m_k = 2\lambda_k$ and $\Phi(x) = -\mu(x)$, we recover, from Eq. 12,

$$P(\vec{x}|H) = \frac{1}{Z} \exp\left(-A(\vec{x})\right) \tag{13}$$

where a discretized version of the classical action

$$A(\vec{x}) = \sum_{i=0}^{n-1} \left[\frac{1}{2} m_i \dot{x}_i^2 - \Phi(x_i) \right] \tag{14}$$

appears in the exponential. Clearly, Eqs. 13 and 14 in the continuum limit become

$$P[x(t)|H] = \frac{1}{Z} \exp\left(-A[x(t)]\right) \tag{15}$$

$$A[x(t)] = \int dt \mathcal{L}(x(t), \dot{x}(t); t),$$

and it follows that the most probable continuous trajectory $x(t)$ is the one that extremizes the classical action $A[x(t)]$ with Lagrangian

$$\mathcal{L}(x, \dot{x}; t) = \frac{1}{2}m(t)\dot{x}^2 - \Phi(x) \quad (16)$$

and associated Hamiltonian

$$\mathcal{H}(x, p; t) = \frac{p^2}{2m(t)} + \Phi(x). \quad (17)$$

This is the formalism of Classical Mechanics for a particle with time-dependent mass $m(t)$ subjected to a potential energy $\Phi(x)$. Both m and Φ have emerged from the Lagrange multipliers associated with the constraints on the expected square of the step and the probability distribution of the coordinate, respectively. Thus we can say the following: whenever the information about the expected square of the step is important, the particle acquires mass, and whenever the information about which regions are more probable in space becomes important, the particle is subjected to a potential energy.

As Eq. 12 is a MaxEnt solution, Eq. 4 holds as

$$\left\langle \nabla \cdot \vec{v}(\vec{x}) \right\rangle = \sum_{i=0}^{n-1} \frac{\lambda_i}{(\Delta t)^2} \left\langle \vec{v}(\vec{x}) \cdot \nabla (x_i - x_{i-1})^2 \right\rangle + \sum_{i=0}^{n-1} \left\langle \vec{v}(\vec{x}) \cdot \nabla \mu(x_i) \right\rangle, \quad (18)$$

with \vec{v} an arbitrary differentiable vector field, of our choosing. If we choose \vec{v} such that it has a single component k , i.e. $v_i = \delta_{i,k} \omega(\vec{x})$ with ω an arbitrary scalar field, we obtain

$$\begin{aligned} \left\langle \frac{\partial \omega}{\partial x_k} \right\rangle &= \sum_{i=0}^{n-1} \frac{\lambda_i}{(\Delta t)^2} \left\langle \omega(\vec{x}) \cdot 2(x_i - x_{i-1})(\delta_{i,k} - \delta_{i-1,k}) \right\rangle + \sum_{i=0}^{n-1} \left\langle \omega(\vec{x}) \cdot \mu'(x_i) \delta_{i,k} \right\rangle \\ &= \frac{1}{(\Delta t)^2} \left\langle 2\omega(\vec{x}) \left[\lambda_k(x_k - x_{k-1}) - \lambda_{k+1}(x_{k+1} - x_k) \right] \right\rangle + \left\langle \omega(\vec{x}) \mu'(x_k) \right\rangle. \end{aligned} \quad (19)$$

But recalling that the discrete forward derivative is

$$\dot{a}_i \approx \frac{a_{i+1} - a_i}{\Delta t}, \quad (20)$$

we can re-write Eq. 19 as

$$\left\langle \frac{\partial \omega}{\partial x_k} \right\rangle = - \left\langle \omega \left(\dot{p}_k + \mu'(x_k) \right) \right\rangle, \quad (21)$$

where $p_k = m_k \dot{x}_k$.

Considering $\omega = 1$ we finally obtain

$$\langle \dot{p}_k \rangle = -\langle \Phi'(x_k) \rangle. \quad (22)$$

which, in the continuous limit ($\Delta t \rightarrow 0$), becomes

$$\langle \dot{p}(t) \rangle = -\langle \Phi'(x(t)) \rangle, \quad (23)$$

the expectation value of Newton's second law with momentum $p = m\dot{x}$ and potential energy $\Phi(x)$.

In appendix A we explore the validity of some aspects of the canonical formalism, namely the Poisson bracket, for the expectation over trajectories.

4 Concluding Remarks

We have found that two simple constraints are sufficient to recover Newton's second law in expectation for the probable trajectories of a particle. The first constraint, on the step size as a function of time, leads to the existence of an inertial mass $m(t)$ proportional to the Lagrange multiplier $\lambda(t)$. To understand the meaning of this, remember that for any variational problem solved using Lagrange multipliers, the larger the value of the multiplier, the more restrictive (and therefore more relevant) the constraint. An irrelevant constraint has always a vanishing multiplier. As Jaynes [13] (p. 945) clearly states, "The Lagrange multipliers λ_k in the MAXENT formalism have therefore a deep meaning: λ_k is the 'potential' of the datum R'_k , that measures how important a constraint it represents."

Now we motivate the following principle: constraints related to conserved quantities are always more relevant. For instance, this explains the fact that the canonical ensemble in equilibrium statistical mechanics is correctly derived just from a single constraint, the energy or expectation of the Hamiltonian, which is an integral of motion. Another illustration is the following: suppose we are trying to recover the trajectory of a particle from information about the distance to a particular point. If this distance is a constant, this is enough to isolate a unique trajectory, the circle. If we only know that the distance varies between r_1 and r_2 , the number of compatible trajectories will increase with $\Delta r = r_2 - r_1$, thus the strength of the constraint will correspondingly decrease with increasing Δr .

Given the earlier discussion, the closer d_i^2 is to be a conserved quantity, the more relevant the first constraint is. In this case, $\lambda(t)$ is large and therefore, $m(t)$ is also large. Conversely, if the value of m is small, this means $\lambda(t)$ is small and therefore d_i^2 has larger fluctuations. In the continuous limit it is the instantaneous speed that fluctuates (there is a non-zero acceleration). This embodies the idea of inertia, and is reminiscent of the ideas of Smolin [14] and of Nelson [15] about inertia being inversely proportional to the size of quantum fluctuations.

Acknowledgments GG and SD thank Jorge Zanelli for useful conversations at the beginning of this work. DG gratefully acknowledges the access to resources provided by Grupo de Nano Materiales (Departamento

de Física, Facultad de Ciencias, Universidad de Chile). SD acknowledges funding from FONDECYT 1140514.

Appendix 1: Canonical Coordinates and Poisson Brackets

An interesting question is how much of the formalism of classical mechanics we can recover from Eq. 22. The fact that most of the structure of classical mechanics is contained in the definition and properties of the Poisson bracket, motivates us to search for an operation analogous to this bracket under the Maximum Caliber formalism.

For arbitrary differentiable functions $f(x, p)$ and $g(x, p)$ the Poisson bracket is defined as

$$\{f, g\} = \frac{\partial f}{\partial x} \frac{\partial g}{\partial p} - \frac{\partial f}{\partial p} \frac{\partial g}{\partial x}, \quad (24)$$

and it is such that

$$\frac{df}{dt} - \frac{\partial f}{\partial t} = \{f, \mathcal{H}\} \quad (25)$$

holds. Let us compute the expectation of the left hand side,

$$\left\langle \frac{df}{dt} \right\rangle - \left\langle \frac{\partial f}{\partial t} \right\rangle = \left\langle \frac{\partial f}{\partial x_k} \dot{x}_k + \frac{\partial f}{\partial p_k} \dot{p}_k \right\rangle, \quad (26)$$

which using Eq. 21 with $\omega = \partial f / \partial p_k$ can be written as

$$\left\langle \frac{df}{dt} \right\rangle - \left\langle \frac{\partial f}{\partial t} \right\rangle = \left\langle \frac{\partial f}{\partial x_k} \dot{x}_k - \frac{\partial}{\partial x_k} \left(\frac{\partial f}{\partial p_k} \right) - \frac{\partial f}{\partial p_k} \Phi'(x_k) \right\rangle \quad (27)$$

Now using our classical Hamiltonian (Eq. 17) we recognize its derivatives

$$\dot{x}_k = \frac{\partial \mathcal{H}}{\partial p_k} \quad (28)$$

$$\Phi'(x_k) = \frac{\partial \mathcal{H}}{\partial x_k} \quad (29)$$

and, upon replacing, we have

$$\left\langle \frac{df}{dt} \right\rangle - \left\langle \frac{\partial f}{\partial t} \right\rangle = \left\langle \frac{\partial f}{\partial x_k} \frac{\partial \mathcal{H}}{\partial p_k} - \frac{\partial}{\partial x_k} \left(\frac{\partial f}{\partial p_k} \right) - \frac{\partial f}{\partial p_k} \frac{\partial \mathcal{H}}{\partial x_k} \right\rangle \quad (30)$$

leading finally to

$$\left\langle \frac{df}{dt} \right\rangle - \left\langle \frac{\partial f}{\partial t} \right\rangle = \left\langle \{f, \mathcal{H}\} \right\rangle - \left\langle \frac{\partial}{\partial x_k} \left(\frac{\partial f}{\partial p_k} \right) \right\rangle. \quad (31)$$

So, in expectation we find a Poisson bracket analog with an additional term. For the particular case $f = \mathcal{H}$, we obtain

$$\left\langle \frac{d\mathcal{H}}{dt} \right\rangle = - \left\langle \frac{\partial}{\partial x_k} \left(\frac{\partial \mathcal{H}}{\partial p_k} \right) \right\rangle, \quad (32)$$

which reduces to

$$s \left\langle \frac{d\mathcal{H}}{dt} \right\rangle = - \left\langle \frac{\partial \dot{x}_k}{\partial x_k} \right\rangle = 0, \quad (33)$$

using the centered difference,

$$\dot{a}_i \approx \frac{a_{i+1} - a_{i-1}}{2\Delta t}. \quad (34)$$

Therefore we have shown that, for a Hamiltonian with the form given in Eq. 17, the energy is conserved in expectation. Note that using one-sided differences is not appropriate here, as one obtains $\pm 1/\Delta t$ depending on forward or backward.

References

1. Caticha, A., Cafaro, C.: From information geometry to newtonian dynamics. *AIP Conf. Proc.* **954**, 165 (2007)
2. Jaynes, E.T.: Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 (1957)
3. J. E. Shore and R. W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Info. Theory, IT.* **26**, 26–37 (1980)
4. Skilling, J.: The axioms of maximum entropy. In: Erickson, G.J., Smith, C.R., (eds.) *Maximum Entropy and Bayesian Methods in Science and Engineering*, pp. 173–187. Kluwer Academic Publishers, New York (1988)
5. Jaynes, E.T.: The minimum entropy production principle. *Ann. Rev. Phys. Chem.* **31**, 579–601 (1980)
6. Press, S., Ghosh, K., Lee, J., Dill, K.A.: Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **85**, 1115 (2013)
7. Stock, G., Ghosh, K., Dill, K.A.: Maximum caliber: a variational approach applied to two-state dynamics. *J. Chem. Phys.* **128**, 194102 (2008)
8. Haken, H.: A new access to path integrals and fokker planck equations via the maximum calibre principle. *Z. Phys. B. Cond. Matt.* **63**, 505–510 (1986)
9. Ge, H., Press, S., Ghosh, K., Dill, K.: Markov processes follow from the principle of maximum caliber. *J. Chem. Phys.* **136**, 064108 (2012)
10. Davis, S., Gutiérrez, G.: Conjugate variables in continuous maximum-entropy inference. *Phys. Rev. E* **86**, 051136 (2012)
11. Feynman, R.P., Hibbs, A.R.: *Quantum Mechanics and Path Integrals*, Dover Books, (2005)
12. Gelfand, I.M., Fomin, S.V.: *Calculus of Variations*. Dover Publications, (2000)
13. Jaynes, E.T.: On the rationale of maximum-entropy methods. *Proc. IEEE* **10**, 939–952 (1982)
14. Smolin, Lee: Quantum fluctuations and inertia. *Phys. Lett.* **113A**, 408 (1986)
15. Nelson, E.: Derivation of the schrödinger equation from newtonian mechanics. *Phys. Rev.* **150**, 1079–1085 (1966)

Hamiltonian formalism and path entropy maximization

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 J. Phys. A: Math. Theor. 48 425003

(<http://iopscience.iop.org/1751-8121/48/42/425003>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 217.112.157.7

This content was downloaded on 03/07/2016 at 17:45

Please note that [terms and conditions apply](#).

Hamiltonian formalism and path entropy maximization

Sergio Davis¹ and Diego González

Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile

E-mail: sdavis@gnm.cl and dgonzalez@gnm.cl

Received 15 May 2015, revised 10 August 2015

Accepted for publication 27 August 2015

Published 21 September 2015



CrossMark

Abstract

Maximization of the path information entropy is a clear prescription for constructing models in non-equilibrium statistical mechanics. Here it is shown that, following this prescription under the assumption of arbitrary instantaneous constraints on position and velocity, a Lagrangian emerges which determines the most probable trajectory. Deviations from the probability maximum can be consistently described as slices in time by a Hamiltonian, according to a nonlinear Langevin equation and its associated Fokker–Planck equation. The connections unveiled between the maximization of path entropy and the Langevin/Fokker–Planck equations imply that missing information about the phase space coordinate never decreases in time, a purely information-theoretical version of the second law of thermodynamics. All of these results are independent of any physical assumptions, and thus valid for any generalized coordinate as a function of time, or any other parameter. This reinforces the view that the second law is a fundamental property of plausible inference.

Keywords: path entropy, Langevin, Fokker–Planck, maximum caliber

1. Introduction

Jaynes' principle of path entropy maximization (also known as the maximum caliber principle, or MaxCal for short) [1] is a clear prescription for the construction of dynamical models, both in non-equilibrium statistical mechanics [2, 3] as well as for any dynamical process [4, 5].

In this work, we derive consequences of the validity of the maximum caliber principle for the problem of estimating the trajectory of a coordinate, for instance, the position of a particle

¹ Author to whom any correspondence should be addressed.

as a function of time. This framework will allow us to describe either classical mechanical systems under uncertainty (e.g. under the influence of random forces) or stochastic signals such as time series. Position and time are familiar concepts to us, but the results of this work apply for inferences about any quantity $q(s)$ parameterized by a continuous index s , for instance, in geometrical problems under uncertainty.

In the context of classical mechanics it is unavoidable to connect the maximum caliber formalism, in which the probability of paths is proportional to the exponential of an action, $P[x()] \propto \exp(A[x()])$, with Feynman's path integral formalism in quantum mechanics [6]. A related attempt to bridge path integrals and classical mechanics is found in [7], although in this case the probability distribution is explicitly constructed to suppress non-classical paths. Connections between thermodynamics and dynamical systems (although without invoking the idea of probabilities of paths) is found in [8] and [9].

The aim of this work is twofold: in the first place, it is interesting to explore to what extent the structure of the dynamical framework of classical (in the sense of non-quantum) physics is already contained in the simple idea of maximization of path entropy. However, the results of this work may also result in powerful tools applicable to the continuous maximum caliber formalism.

2. The maximum caliber principle

Suppose for an unknown coordinate x described as a function of time t we only know the expectation of a function $f(x, \dot{x}; t)$ (over the distribution of possible trajectories $x(t)$) as a function of time $F(t)$, that is,

$$\langle f(x, \dot{x}; t) \rangle_I = F(t), \quad (1)$$

for every instant in the interval $[t_i, t_f]$. According to the maximum caliber principle, the optimal assignment of probability for the trajectories maximizes

$$\mathcal{S} = - \int Dx() P[x()] \ln \frac{P[x()|I]}{P[x()|I_0]}, \quad (2)$$

where I denotes all the given information about the problem, in particular the constraint in equation (1), and I_0 is the complete ignorance state. The path probability $P[x()|I_0]$ is the *a priori* measure needed for a consistent definition of entropy in continuous systems. Imposing this constraint we can write the resulting probability as

$$P[x()|I] = \frac{1}{Z[\lambda()]} P[x()|I_0] \exp\left(- \int_{t_i}^{t_f} dt \lambda(t) f(x(t), \dot{x}(t); t)\right), \quad (3)$$

where λ is a Lagrange multiplier function. In the following we will assume, for simplicity of the analysis, that the prior measure $P[x()|I_0]$ is flat. If this is not the case, it can be absorbed as an extra term in the action $A[x()]$, to be defined below.

If we define a Lagrangian

$$L(x, \dot{x}; t) = \alpha \lambda(t) f(x, \dot{x}; t), \quad (4)$$

we can rewrite equation (3) as

$$P[x()|I] = \frac{1}{Z[\lambda()]} \exp\left(- \frac{1}{\alpha} A[x()]\right), \quad (5)$$

where the functional A is the action, defined as

$$A[x()] = \int_{t_i}^{t_f} dt L(x(t), \dot{x}(t); t), \quad (6)$$

and the constant α , with dimensions of action, is extracted only to ensure the exponent in equation (5) is adimensional. Equation (5) is analogous to the probability amplitude assigned to a trajectory in the Feynman path integral formalism [6] (except the exponent here is real-valued), with α being a constant analogous to Planck's constant. We see that the most probable path is automatically prescribed by the principle of minimum action, and thus is a solution of the Euler–Lagrange equation

$$\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}. \quad (7)$$

It is also immediately clear from this formalism that geometric constraints of the form $g_k(x(t); t) = 0$ will only add terms to the Lagrangian, as prescribed by classical mechanics [10],

$$\begin{aligned} L(x, \dot{x}; t) &= \alpha \lambda(t) f(x, \dot{x}; t) + \sum_k \mu_k(t) g_k(x(t); t) \\ &= \alpha \lambda(t) \left[f(x, \dot{x}; t) + \sum_k \tilde{\mu}_k(t) g_k(x(t); t) \right], \end{aligned} \quad (8)$$

where the $\tilde{\mu}_k = \mu_k/(\alpha \lambda)$ are (rescaled) Lagrange multipliers. Thus, from the point of view of maximum caliber, there is no conceptual distinction between the original Lagrangian and the terms from holonomic constraints: the entire Lagrangian always arises from constraints.

A striking example of this loss of distinction occurs when the quantity F which is known and imposed as a constraint is the instantaneous joint probability $P(x(t), \dot{x}(t)|I)$. This constraint can be written as

$$\langle \delta(x(t) - x) \delta(\dot{x}(t) - \dot{x}) \rangle_I = P(x(t), \dot{x}(t)|I), \quad (9)$$

i.e., the constraining function $f(x, \dot{x}; t)$ is the product of delta functions. If the probability in the right hand side is known for all x and \dot{x} in the state space (x, \dot{x}) where the probability is non-zero, the Lagrange multiplier function will now depend on (x, \dot{x}, t) and the (normalized) action is

$$\begin{aligned} \frac{A}{\alpha} &= \int_{t_i}^{t_f} dt \int dx d\dot{x} \lambda(x, \dot{x}; t) \delta(x(t) - x) \delta(\dot{x}(t) - \dot{x}) \\ &= \int_{t_i}^{t_f} dt \lambda(x(t), \dot{x}(t); t), \end{aligned} \quad (10)$$

i.e. the Lagrange multiplier function itself becomes the Lagrangian. This means, although in principle the Lagrangian arising from equation (4) is arbitrary (determined by the choice of constraining function), a particular Lagrangian is singled out, the Lagrange multiplier of the probability.

3. Expectation of functionals

In order to obtain additional relations for expectation values of arbitrary functionals over the distribution given by equation (3), we apply the finite difference method [11] and map the continuous dynamical problem into a maximum entropy problem with finite number of

degrees of freedom by discretizing time, that is, replacing the continuous trajectory $x(t)$ by a vector of N components, $\vec{x} = (x_1, \dots, x_N)$. equation (3) then reduces to a maximum entropy solution

$$P(\vec{x}|I) = \frac{1}{Z} \exp\left(-\frac{1}{\alpha}A(\vec{x})\right) \quad (11)$$

with the action now replaced by a scalar field $A(\vec{x})$, given by

$$A(\vec{x}) = \Delta t \sum_{j=1}^N L_j(x_j, (x_{j+1} - x_j)/\Delta t). \quad (12)$$

There is a connection between expectation values provided by the conjugate variables theorem [12],

$$\langle \nabla \cdot \vec{v}(\vec{x}) \rangle_I = \frac{1}{\alpha} \langle \vec{v} \cdot \nabla A(\vec{x}) \rangle_I. \quad (13)$$

Let us choose $\vec{v} = \hat{e}_k W(\vec{x})$ where \hat{e}_k the unit vector in along the k th coordinate and the trial field W is of the form

$$W(\vec{x}) = \Delta t \sum_{j=1}^N \omega_j(x_j, (x_{j+1} - x_j)/\Delta t). \quad (14)$$

Then,

$$\left\langle \frac{\partial}{\partial x_k} W(\vec{x}) \right\rangle_I = \frac{1}{\alpha} \left\langle W \frac{\partial}{\partial x_k} A(\vec{x}) \right\rangle_I. \quad (15)$$

Now, for both functionals $W(\vec{x})$ and $A(\vec{x})$ we have [11],

$$\frac{\partial}{\Delta t \partial x_k} \rightarrow \frac{\delta}{\delta x(t)}, \quad (16)$$

and the functional version of the conjugate variables theorem is given by

$$\left\langle \frac{\delta W}{\delta x(t)} \right\rangle_I = \frac{1}{\alpha} \left\langle W[x(t)] \frac{\delta A}{\delta x(t)} \right\rangle_I, \quad (17)$$

with $W[x(t)]$ a trial functional with the form

$$W[x(t)] = \int_{t_i}^{t_f} dt \omega(x(t), \dot{x}(t); t). \quad (18)$$

This is the equivalent of the identity (7.30) in [6] for the quantum mechanical path integrals.

By explicitly replacing the Lagrangian L and the trial function ω , we can compactly write equation (17) as

$$\langle \hat{E}_t \omega \rangle_I = \frac{1}{\alpha} \langle W[x(t)] \hat{E}_t L \rangle_I, \quad (19)$$

where we have introduced the operator \hat{E}_t as

$$\hat{E}_t G(x, \dot{x}; t) = \left(\frac{\partial}{\partial x} - \frac{d}{dt} \frac{\partial}{\partial \dot{x}} \right) G(x, \dot{x}; t). \quad (20)$$

Choosing $W[x()] = 1$ tells us that

$$\langle \hat{E}_t L \rangle_I = 0 \quad (21)$$

for all instants t , i.e., the Euler–Lagrange equation is valid in expectation over the ensemble of trajectories, not only for the most probable one. This suggests treating $\hat{E}_t L$ itself as a random variable with zero expectation for the purpose of constructing a Langevin equation. However, we need to establish the predicted correlation $\langle \hat{E}_t L \cdot \hat{E}_{t'} L \rangle_I$ of the quantity $\hat{E}_t L$. Now we use $\omega = \hat{E}_t L$ in equation (19) to show that

$$\langle \hat{E}_t L \cdot \hat{E}_{t'} L \rangle_I = \left\langle \hat{E}_t \left(\hat{E}_{t'} L \right) \right\rangle_I = \frac{1}{\alpha} \left\langle \frac{\delta^2 A}{\delta x(t) \delta x(t')} \right\rangle_I. \quad (22)$$

As the action has the form in equation (6), its second functional derivative is

$$\begin{aligned} \frac{\delta^2 A}{\delta x(t) \delta x(t')} &= \frac{\delta}{\delta x(t')} \left[\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right] \Big|_t \\ &= \frac{\partial}{\partial x'} \left(\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right) - \frac{d}{dt'} \frac{\partial}{\partial \dot{x}'} \left(\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right) \end{aligned} \quad (23)$$

and this is only non-zero for $t = t'$, thus in all generality we can write it as

$$\langle \hat{E}_t L \cdot \hat{E}_{t'} L \rangle_I = 2R(t) \delta(t - t') \quad (24)$$

with $R(t)$ a function of time to be determined. Here we note that, by taking the limit $t \rightarrow t'$ in equation (24), it follows that $R(t)$ cannot be negative. From equations (21) and (24), we now formally write the stochastic differential equation (SDE),

$$\hat{E}_t L = \frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} = \sqrt{2R(t)} \xi(t) \quad (25)$$

with $\xi(t)$ an unbiased, uncorrelated ‘noise’,

$$\langle \xi(t) \rangle_I = 0 \quad (26)$$

$$\langle \xi(t) \xi(t') \rangle_I = \delta(t - t'). \quad (27)$$

It is crucial to note at this point that $R(t)$ does not necessarily represent a ‘random’ physical influence: as the probability interpretation when using the maximum entropy and maximum caliber principles is unavoidably Bayesian, $R(t)$ represents inaccessible or hidden information. Without this uncertainty, the system follows the trajectory which minimizes the action (equation (25) reduces to the Euler–Lagrange equation, equation (7)).

Equation (25) is, in general, a second-order SDE, but we can obtain a system of two first-order SDEs by expressing it in canonical form [11]. For this we define the Hamiltonian \mathcal{H} corresponding to the Lagrangian L as the Legendre transformation

$$\mathcal{H}(x, p) = p\dot{x} - L(x, \dot{x}; t) = \alpha \lambda(t) \left(\dot{x} \frac{\partial f}{\partial \dot{x}} - f \right) \Big|_{x,p}, \quad (28)$$

together with the momentum $p = \partial L / \partial \dot{x}$. Now we have defined a phase space (x, p) and by virtue of the Legendre transformation the following relations hold, independently of the form of the Lagrangian,

$$\frac{\partial \mathcal{H}}{\partial x} = - \frac{\partial L}{\partial x}, \quad (29)$$

$$\frac{\partial \mathcal{H}}{\partial p} = \dot{x}. \quad (30)$$

By introducing the Poisson brackets

$$\{F, G\} = \frac{\partial F}{\partial x} \frac{\partial G}{\partial p} - \frac{\partial F}{\partial p} \frac{\partial G}{\partial x} \quad (31)$$

we see that the required system of first-order SDEs is a nonlinear Langevin equation in phase space,

$$\dot{\Gamma}_i = \{\Gamma_i, \mathcal{H}\} + R_{ij}(t)\xi_j(t), \quad (32)$$

where we have used the Einstein summation convention, $\vec{\Gamma} = (x, p)$ and the only non-zero component of the uncertainty matrix R_{ij} is $R_{22}(t) = \sqrt{2R(t)}$ with $\xi_2(t) = \xi(t)$.

Equations (25) and (32) give a strong theoretical basis to the use of the Langevin formalism, independent of physical considerations. Most importantly, they support the assumption of uncorrelated noise common in the literature for non-physical Langevin forces [13] as an inevitable requirement if the problem consists only of instantaneous information about position and velocity. Assuming any other kind of noise will be inconsistent with the information one has by means of equation (1).

4. Time evolution of the phase space probability

We can obtain the time evolution of the probability distribution in phase space, $P(\vec{\Gamma}(t)|I) = P$, by means of the Kramers–Moyal expansion [14],

$$\partial_t P = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \partial_{j_1 \dots j_n}^n \left(D_{j_1 \dots j_n}^{(n)}(\vec{\Gamma}; t) P \right), \quad (33)$$

where we again used the Einstein summation convention. If we truncate this expansion at $n = 2$, one is left with a Fokker–Planck equation,

$$\partial_t P + \partial_i (D_i^{(1)} P) = \frac{1}{2} \partial_{ij}^2 (D_{ij}^{(2)} P), \quad (34)$$

with coefficients

$$D_i^{(1)} = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \langle \Delta \Gamma_i \rangle_{\vec{\Gamma}} = \{\Gamma_i, \mathcal{H}\}, \quad (35)$$

$$D_{ij}^{(2)} = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \langle \Delta \Gamma_i \Delta \Gamma_j \rangle_{\vec{\Gamma}} = R_{ik}(t) R_{jk}(t), \quad (36)$$

where $\Delta \Gamma_i = \Gamma_i(t + \epsilon) - \Gamma_i(t)$. Therefore we can analyze the time evolution of the phase space probability P by solving

$$\frac{\partial P}{\partial t} + \{P, \mathcal{H}\} = R(t) \frac{\partial^2 P}{\partial p^2}. \quad (37)$$

This is a Boltzmann transport equation with collision term proportional to the uncertainty $R(t)$. Note that, for an arbitrary Hamiltonian, in the case of complete certainty ($R(t) \rightarrow 0$) we recover the Liouville equation. It is useful to write equation (37) in logarithmic form, namely,

$$\frac{\partial}{\partial t} \ln P + \{\ln P, \mathcal{H}\} = R(t) \left[\frac{\partial^2}{\partial p^2} \ln P + \left(\frac{\partial}{\partial p} \ln P \right)^2 \right]. \quad (38)$$

Multiplying by an arbitrary function $G(x, p; t)$ and taking expectation over P we have

$$\begin{aligned} & \left\langle G \frac{\partial}{\partial t} \ln P \right\rangle_{I,t} + \langle G \{\ln P, \mathcal{H}\} \rangle_{I,t} \\ &= R(t) \left\langle G \left[\frac{\partial^2}{\partial p^2} \ln P + \left(\frac{\partial}{\partial p} \ln P \right)^2 \right] \right\rangle_{I,t} \end{aligned} \quad (39)$$

FIXME

Here we apply the general ‘fluctuation-dissipation’ relation for an arbitrary parameter α of a distribution P ,

$$\frac{\partial}{\partial \alpha} \langle G \rangle_{I,\alpha} = \left\langle \frac{\partial G}{\partial \alpha} \right\rangle_{I,\alpha} + \left\langle G \frac{\partial}{\partial \alpha} \ln P \right\rangle_{I,\alpha} \quad (40)$$

with $\alpha = t$, and the conjugate variables theorem [12], as

$$\left\langle \frac{\partial G}{\partial \Gamma_i} \right\rangle_I = - \left\langle G \frac{\partial}{\partial \Gamma_i} \ln P \right\rangle_I. \quad (41)$$

After some algebraic manipulations and repeated use of equation (40) and 41 to eliminate $\ln P$, we obtain a classical (non-quantum) analog of the Ehrenfest theorem [15],

$$\frac{d}{dt} \langle G \rangle_{I,t} = \left\langle \frac{\partial G}{\partial t} \right\rangle_{I,t} + \langle \{G, \mathcal{H}\} \rangle_{I,t} + R(t) \left\langle \frac{\partial^2 G}{\partial p^2} \right\rangle_{I,t} \quad (42)$$

which gives us the time evolution of an arbitrary phase space function G , and also provides us with a direct way to compute $R(t)$ from time-dependent averages.

An important remark we address here is the fact that the expectation values over trajectories are completely described by the Lagrangian (equation (19)), while the instantaneous properties, representing the system as successive ‘slices’ in time, are in turn described by the Hamiltonian (equation (42)).

5. Second law of thermodynamics

Replacing $G = \ln P$ in equation (42) and taking into account that, because of equation (40) and equation (41), $\langle \partial_t \ln P \rangle_{I,t} = 0$ and $\langle \{\ln P, \mathcal{H}\} \rangle_{I,t} = 0$, we see that

$$\frac{dS}{dt} = R(t) \left\langle \left(\frac{\partial}{\partial p} \ln P \right)^2 \right\rangle_{I,t} \geq 0, \quad (43)$$

with $S(t) = -\langle \ln P \rangle_{I,t}$ the information entropy of the distribution P at time t . Therefore, missing information about the location in phase space never decreases with time, and its increase is solely due to the uncertainty. In other words, irreversibility is equivalent to uncertainty over the trajectory followed by the system, whatever its origin (e.g. sensibility to initial conditions, external driving).

6. Concluding remarks

In summary, we have explored the consequences of constructing a maximum caliber dynamical model using only instantaneous information about position and velocity. In the first place, the problem of prediction of the most probable trajectory reduces to standard classical mechanics, complete with a Lagrangian, a phase space and a Hamiltonian. For the description of deviations around the most probable trajectory, the formalism naturally leads to a nonlinear Langevin equation with uncorrelated noise in phase space, and its corresponding Fokker–Planck equation. This Fokker–Planck equation is completely determined by the Hamiltonian of the problem.

Given the striking resemblance of equations (5), (17) and (42) to real-valued versions of the Feynman path integral formulation of quantum mechanics and the Ehrenfest theorem, respectively, it remains to be seen if the idea of path entropy maximization could lead to their quantum mechanical (complex-valued) versions under suitable constraints, as well as its connection with Nelson’s stochastic mechanics [16, 17].

Acknowledgments

The authors gratefully acknowledge funding from FONDECYT grant number 1140514. DG also acknowledges funding from CONICYT PhD fellowship 21140914.

References

- [1] Jaynes E T 1980 *Ann. Rev. Phys. Chem.* **31** 579
- [2] Dewar R 2003 *J. Phys.: Math. Gen.* **36** 631
- [3] Grandy W T 2008 *Entropy and the Time Evolution of Macroscopic Systems* (Oxford: Oxford University Press)
- [4] Haken H 1986 *Z. Phys. B* **63** 505
- [5] Pressé S, Ghosh K, Lee J and Dill K A 2013 *Rev. Mod. Phys.* **85** 1115
- [6] Feynman R P and Hibbs A R 2005 *Quantum Mechanics and Path Integrals* (New York: McGraw-Hill)
- [7] Gozzi E, Reuter M and Thacker W D 1989 *Phys. Rev. D* **40** 3363
- [8] Grmela M 2013 *Comput. Math. Appl.* **65** 1457
- [9] Grmela M 2014 *Entropy* **16** 1652
- [10] Lanczos C 1970 *The Variational Principles of Mechanics* (New York: Dover)
- [11] Gelfand I M and Fomin S V 2000 *Calculus of Variations* (New York: Dover)
- [12] Davis S and Gutiérrez G 2012 *Phys. Rev. E* **86** 051136
- [13] Yura Y, Takayasu H, Sornette D and Takayasu M 2014 *Phys. Rev. Lett.* **112** 098703
- [14] Risken H 1996 *The Fokker–Planck Equation: Methods of Solution and Applications* (Berlin: Springer)
- [15] Shankar R 1994 *Principles of Quantum Mechanics* (New York: Plenum)
- [16] Nelson E 1966 *Phys. Rev.* **150** 1079
- [17] González D, Davis S and Gutiérrez G 2014 *Found. Phys.* **44** 923

A Bayesian Interpretation of First-Order Phase Transitions

Sergio Davis¹ · Joaquín Peralta² · Yasmín Navarrete¹ ·
Diego González¹ · Gonzalo Gutiérrez¹

Received: 12 May 2015 / Accepted: 19 October 2015
© Springer Science+Business Media New York 2015

Abstract In this work we review the formalism used in describing the thermodynamics of first-order phase transitions from the point of view of maximum entropy inference. We present the concepts of transition temperature, latent heat and entropy difference between phases as emergent from the more fundamental concept of internal energy, after a statistical inference analysis. We explicitly demonstrate this point of view by making inferences on a simple game, resulting in the same formalism as in thermodynamical phase transitions. We show that analogous quantities will inevitably arise in any problem of inferring the result of a yes/no question, given two different states of knowledge and information in the form of expectation values. This exposition may help to clarify the role of these thermodynamical quantities in the context of different first-order phase transitions such as the case of magnetic Hamiltonians (e.g. the Potts model).

Keywords Maximum entropy · Bayesian inference · Phase transitions

1 Introduction

Jaynes' proposal of the principle of maximum entropy (PME) as a general tool of probabilistic inference [1,2] is remarkable in that it is both widely used [3] and somewhat controversial [4,5]. It asserts that the most unbiased probability distribution P given some fixed knowledge \mathcal{I} is the one that maximizes Shannon's information entropy,

✉ Sergio Davis
sdavis@gnm.cl

¹ Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile

² Departamento de Física, Facultad de Ciencias Físicas, Universidad Andrés Bello, Santiago, Chile

$$S[P(x|\mathcal{I})] = - \sum_x P(x|\mathcal{I}) \log_2 P(x|\mathcal{I}) \quad (1)$$

while being consistent with said knowledge. This is because $S[P]$ is a measure of uncertainty [6] or lack of knowledge about the degrees of freedom (represented collectively by x in the notation above) and, maximizing it leads to the probabilistic model containing the least amount of information, but nevertheless able to reproduce the features one demands of it. As this is a process of inference it cannot be deductive: predictions derived from the maximum entropy model may be proved wrong by subsequent measurements, and this reflects an incompleteness of the fixed knowledge used to constrain the maximization.

Jaynes' interpretation of the formalism of statistical mechanics sees it as just the application of this principle of maximum entropy, valid in all statistical inference, to the case of a macroscopic number of particles (and degrees of freedom). In this situation the predictions are almost perfectly sharp, with uncertainties vanishing as $1/\sqrt{N}$, with N the number of degrees of freedom. This is all well described for the case of thermodynamic equilibrium of a single phase. However, how this information-theoretical interpretation manifests itself in the case of the study of phase transitions, and what can we learn from this, is an issue which has not been so extensively clarified. For instance, in his book on probability theory [2] (p. 602), Jaynes wrote in a somewhat cryptic footnote, that

“... in statistical mechanics the relative probability P_j/P_k of two different phases, such as liquid and solid, is the ratio of their partition functions Z_j/Z_k , which are the normalization constants for the sub-problems of prediction within one phase. In Bayesian analysis, the data are indifferent between two models when their normalization constants become equal; in statistical mechanics the temperature of a phase transition is the one at which the two partition functions become equal...”

This suggests that the problem of liquid-solid phase transition, or in fact, any phase transition, can be posed as a model comparison problem, and therefore the transition temperature and the free energy can be given an information-theoretical meaning. In this work, we attempt to clarify (“demistify”, one might even say) the interpretation of these quantities, showing that they are consequences of the maximum entropy inference rule under known internal energy averages. In this sense, only this internal energy is fundamental, whereas temperature, entropy and free energy are derived quantities, of a statistical nature. This statistical character of temperature is understood, for instance, in terms of the kinetic theory of gases; however, we will show that its meaning is much wider in the context of information theory.

In order to remove all particularities of thermodynamics from the treatment of first-order phase transitions, we present a parallel of the formalism used in first-order phase transitions based entirely on the application of the PME. We introduce a simple game, the “disc throwing” game, and answer two questions related to it by means of the PME. In the answers to these questions we will recover the concepts of transition temperature, Helmholtz free energy and the rule that imposes its equality for the two phases at the coexistence point.

The rest of the paper is organized as follows. In Sect. 2 we review the main features of the maximum entropy formalism. Section 3 shows an illustration of PME inference, while Sect. 4 describes and solves the disc throwing game problem. In Sect. 5 we expose the perfect parallel between the solution of this problem and that of the coexistence of two phases in thermodynamical equilibrium. Finally we conclude with some remarks.

2 Maximum Entropy Inference

Consider a system having N discrete degrees of freedom $\mathbf{x} = (x_1, \dots, x_N)$ and being fully described in statistical terms by a function $f(\mathbf{x})$ with known expectation value f_0 . Knowledge of f_0 is symbolically represented by \mathcal{I} . According to the PME, the most unbiased model is the one that maximizes the Gibbs-Shannon entropy functional

$$S = - \sum_{\mathbf{x}} P(\mathbf{x}|\mathcal{I}) \log_2 P(\mathbf{x}|\mathcal{I}) \tag{2}$$

subject to the constraint \mathcal{I} , i.e., to

$$\langle f(\mathbf{x}) \rangle = f_0. \tag{3}$$

Maximization under this constraint, and the always implicit constraint of proper normalization of the probability, is achieved by the inclusion of Lagrange multipliers λ and μ respectively, after which the problem reduces to the maximization of the augmented function

$$\begin{aligned} \tilde{S} = & - \sum_{\mathbf{x}} P(\mathbf{x}|\mathcal{I}) \log_2 P(\mathbf{x}|\mathcal{I}) + \lambda(f_0 - \sum_{\mathbf{x}} P(\mathbf{x}|\mathcal{I}) f(\mathbf{x})) \\ & + \mu(1 - \sum_{\mathbf{x}} P(\mathbf{x}|\mathcal{I})). \end{aligned} \tag{4}$$

This leads to the well-known maximum entropy (MaxEnt) model

$$P(\mathbf{x}|\lambda) = \frac{1}{Z(\lambda)} \exp(-\lambda f(\mathbf{x})) \tag{5}$$

in which we have changed the notation from the purely abstract $P(\mathbf{x}|\mathcal{I})$ to the more concrete $P(\mathbf{x}|\lambda)$, given that the parameter λ distinguishes between all the possible states of knowledge compatible with the possible values of f_0 . The function Z ,

$$Z(\lambda) = \sum_{\mathbf{x}} \exp(-\lambda f(\mathbf{x})). \tag{6}$$

is known as the partition function. The Lagrange multiplier λ is usually determined as the unique solution of

$$- \frac{\partial}{\partial \lambda} \ln Z(\lambda) = f_0. \tag{7}$$

The procedure just outlined could in principle be performed using so-called generalized entropies in place of the Gibbs-Shannon entropy (Eq. 2), such as the Tsallis [7] or Rényi [8] entropies. Although these entropies may be valid tools in describing the complexity of non-extensive systems, their use in statistical inference has been shown to lead to inconsistencies [9–12]. If the degrees of freedom contained in \mathbf{x} are continuous, Shannon entropy needs to be replaced with the relative entropy

$$S = - \int d\mathbf{x} P(\mathbf{x}|\mathcal{I} \wedge I_0) \log_2 \frac{P(\mathbf{x}|\mathcal{I} \wedge I_0)}{P(\mathbf{x}|I_0)} \tag{8}$$

where I_0 denotes an “initial” state of knowledge. The solution to the maximum entropy problem is now

$$P(\mathbf{x}|\mathcal{I} \wedge I_0) = \frac{1}{Z(\lambda)} P(\mathbf{x}|I_0) \exp(-\lambda f(\mathbf{x})) \tag{9}$$

with

$$Z(\lambda) = \int d\mathbf{x} P(\mathbf{x}|I_0) \exp(-\lambda f(\mathbf{x})). \tag{10}$$

In both cases (discrete and continuous degrees of freedom), the maximized entropy has a value

$$S = \ln Z(\lambda) + \lambda f_0. \tag{11}$$

Now, we have just described the formalism of the canonical ensemble if we think of the system as composed by n particles with position \mathbf{r}_i and momentum \mathbf{p}_i (with $i=1,\dots,n$) and the descriptor function as the Hamiltonian $f = \mathcal{H}(\mathbf{r}_1, \dots, \mathbf{r}_n, \mathbf{p}_1, \dots, \mathbf{p}_n)$. Then Eq. 5 is the canonical distribution where we identify $\lambda = \beta = 1/(k_B T)$.

In thermodynamic notation, Eq. 11 reads,

$$S(\beta)/k_B = \ln Z(\beta) + \beta E(\beta) \tag{12}$$

If we introduce the Helmholtz free energy $\beta F(\beta) = -\ln Z(\beta)$, Eq. 12 reduces to

$$S(\beta)/k_B = \beta(E(\beta) - F(\beta)) \tag{13}$$

i.e., using $\beta = 1/k_B T$,

$$F(T) = E(T) - TS(T). \tag{14}$$

3 An Illustration of the Maximum Entropy Formalism

Suppose we have a swimming pool full of plastic balls (all spherical) of different radii. The average volume of a ball is V . What is the average radius?

We have the constraint,

$$\left\langle \frac{4}{3}\pi r^3 \right\rangle = V, \tag{15}$$

which is equivalent to

$$\langle r^3 \rangle = \frac{3}{4\pi} V, \quad (16)$$

from which the most unbiased model for r is

$$P(r|\lambda) = \frac{1}{Z(\lambda)} \exp(-\lambda r^3) \Theta(r). \quad (17)$$

The partition function is given by

$$Z(\lambda) = \int_0^\infty dr \exp(-\lambda r^3) = \Gamma(4/3) \lambda^{-1/3}, \quad (18)$$

therefore, the value of λ is determined from

$$-\frac{\partial}{\partial \lambda} \ln Z(\lambda) = \frac{1}{3\lambda} = \frac{3}{4\pi} V, \quad (19)$$

i.e., $\lambda = 4\pi/(9V)$. Note that from inspection of Eq. 17 and the fact that λ is positive, the most probable radius is zero and the probability monotonically decreases with r . The expectation of r is then

$$\begin{aligned} \langle r \rangle &= 7 \frac{1}{Z(\lambda)} \int_0^\infty dr r \exp(-\lambda r^3) = \frac{3^{1/3}}{3} \frac{\Gamma(2/3)}{\Gamma(4/3)} \\ &\cdot \left(\frac{3V}{4\pi} \right)^{1/3} \approx 0.729011 \sqrt{3} \langle r^3 \rangle. \end{aligned} \quad (20)$$

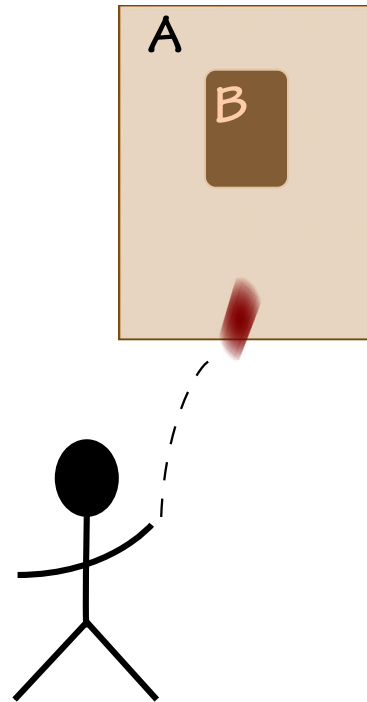
From this example two interesting things emerge. First, the expected radius is less than the naïve estimate $r_0 = \sqrt{3} \langle r^3 \rangle$, valid in the case where all the balls have the same radius. Second, the Lagrange multiplier λ is larger for small V , and this is expected given that the smaller V is, the possible radii are more concentrated around zero and therefore there is less uncertainty about the value of the radius. This means the constraint of known V (Eq. 15) has greater “weight” for smaller V . As the distribution function $P(r|\lambda)$ decreases from $r = 0$ onward, there are more balls with $r \leq r_0$ than with $r > r_0$ and thus the estimate $\langle r \rangle$ is skewed towards zero.

4 A Simple Disc Throwing Game

Suppose a player can throw a disc into a surface A (with area Σ_A), containing within it a smaller surface B (with area $\Sigma_B < \Sigma_A$). We consider A and B to be disjoint regions, as shown in Fig. 1. A successful hit within B gives the player n_B points, whereas a hit inside A (outside B) gives n_A points to the player (as hitting B is more difficult, $n_B > n_A$). This is similar to the game “rayuela” as is known in some South American countries.

We can present two questions about this game:

Fig. 1 Schematic representation of the disc throwing game



- (a) With only the information laid out above, and particularly without knowing anything about the performance of the player, what probability should one assign to hitting B ?
- (b) Now consider the player has obtained an average score of \bar{n} in the past (over enough trials to be considered a reliable average). What probability should one assign now to hitting B ?

In (a) the intuitive answer is that the probabilities of hitting either A or B are completely determined by their areas. In fact, considering each landing point as a coordinate inside A , and because such points are mutually exclusive, exhaustive alternatives and there is symmetry under exchange, we can easily see that

$$\frac{P(A|\mathcal{I}_1)}{P(B|\mathcal{I}_1)} = \frac{\Sigma_A}{\Sigma_B} \quad (21)$$

From this, given that landing in A or B constitute mutually exclusive and exhaustive propositions, $P(A|\mathcal{I}_1) + P(B|\mathcal{I}_1) = 1$. Therefore,

$$P(\alpha|\mathcal{I}_1) = \frac{\Sigma_\alpha}{\Sigma_A + \Sigma_B} \quad (22)$$

with $\alpha = A, B$. The predicted score of the player, with just the information we have in (a), is then

$$\bar{n} = \frac{\Sigma_A n_A + \Sigma_B n_B}{\Sigma_A + \Sigma_B}. \quad (23)$$

We see that probabilities are governed only by the ratio Σ_A/Σ_B , and we can conclude that always $P(A|\mathcal{I}_1) > P(B|\mathcal{I}_1)$, given that the area of B is considerably

smaller. Now, what happens in (b) is that we have to constrain the inference to this new information, given in the form of an expectation value. We invoke the law of large numbers and assume $\langle n \rangle = \bar{n}$, then the most unbiased probability for either result given \bar{n} , according to the PME, is (using Eq. 9),

$$P(\alpha|\mathcal{I}_2) = \frac{1}{Z(\lambda)} \Sigma_\alpha \exp(-\lambda n_\alpha) \tag{24}$$

with

$$Z(\lambda) = \Sigma_A \exp(-\lambda n_A) + \Sigma_B \exp(-\lambda n_B), \tag{25}$$

and

$$-\frac{\partial}{\partial \lambda} \ln Z(\lambda) = \bar{n}. \tag{26}$$

After explicitly using the result of Eq. 25 in Eq. 26 and some algebra, we have that

$$\Sigma_A (\bar{n} - n_A) \exp(-\lambda n_A) = \Sigma_B (n_B - \bar{n}) \exp(-\lambda n_B) \tag{27}$$

from which it follows that λ is given by

$$\lambda(\bar{n}) = -\frac{1}{n_B - n_A} \left[\ln \Sigma_A - \ln \Sigma_B + \ln \frac{\bar{n} - n_A}{n_B - \bar{n}} \right]. \tag{28}$$

In order to simplify notation, let us introduce

$$\Delta n = n_B - n_A, \tag{29}$$

$$\Delta S = S_B - S_A = \ln \Sigma_B - \ln \Sigma_A. \tag{30}$$

Then Eq. 28 reads,

$$\lambda \Delta n - \Delta S = \ln \frac{n_B - \bar{n}}{\bar{n} - n_A} \tag{31}$$

It is clear that, when $\lambda = 0$, Eq. 31 implies

$$\frac{n_B - \bar{n}}{\bar{n} - n_A} = \frac{\Sigma_A}{\Sigma_B}. \tag{32}$$

which is nothing but the result of section (a), Eq. 23. This happens when the reported average score \bar{n} is the same as predicted from the area information alone. This reflects a complete lack of ability from the player to control the hitting spot, because the results do not differ from pure “random” shots. However, if \bar{n} is not consistent with Eq. 23, then $\lambda \neq 0$ and the ratio between probabilities is not simply the ratio of the respective areas, but it is given by

$$\frac{P(A|\mathcal{I}_2)}{P(B|\mathcal{I}_2)} = \frac{\Sigma_A}{\Sigma_B} \exp(-\lambda(n_A - n_B)) \tag{33}$$

i.e., defining $\Delta \ln P = \ln P(B|\mathcal{I}_2) - \ln P(A|\mathcal{I}_2)$,

$$\Delta \ln P = \Delta S - \lambda \Delta n, \quad (34)$$

or, if we define $F_\alpha = n_\alpha - S_\alpha/\lambda$, we have

$$\Delta \ln P = -\lambda \Delta F. \quad (35)$$

Therefore, the most probable outcome (A or B) would be the one with lowest value of F .

After comparing Eqs. 34 and 31, the ratio of probabilities is given by

$$\frac{P(A|\mathcal{I}_2)}{P(B|\mathcal{I}_2)} = \frac{n_B - \bar{n}}{\bar{n} - n_A}. \quad (36)$$

There will be an interesting value of \bar{n} , namely the average $(n_A + n_B)/2$, where $P(A|\mathcal{I}_2) = P(B|\mathcal{I}_2)$. In this case we are maximally uncertain with respect to which region the player will hit, i.e., we have “canceled out” all the information we had from the areas by using the average score. This situation corresponds to a “critical value” of the Lagrange multiplier,

$$\lambda_0 = \lambda \left(\frac{n_A + n_B}{2} \right) = \frac{\Delta S}{\Delta n}. \quad (37)$$

5 Bayesian Thermodynamics

Perhaps it will be striking to the reader (at first) to notice that we have replicated the formalism used to study first-order phase transitions in thermodynamical systems. Imagine the two regions A and B of the game introduced previously, as regions in phase space corresponding, for instance, to liquid and solid, respectively. We can relate the area of each region Σ to the volume in phase space occupied by each of the thermodynamic phases, and in this sense, the quantity

$$S = \ln \Sigma \quad (38)$$

is readily interpreted as the Boltzmann entropy (taking $k_B=1$). Therefore the most probable phase (i.e., the most stable phase in thermodynamical terms) is, in absence of any other information, the one with the largest value of entropy. This is the same situation as in the microcanonical ensemble [13].

When we have information about the expected (or average) score \bar{n} , analogous to the measured internal energy E of a thermodynamical system (n_A and n_B are then the internal energies for the liquid and solid phases, respectively), what decides the most probable phase is, according to Eq. 35, the difference in the quantity

$$F = n - S/\lambda \quad (39)$$

Table 1 Equivalences between concepts arising in the analysis of the throwing game and thermodynamical concepts

Throwing game	Thermodynamics
Logarithm of area ($\ln \Sigma$)	Entropy (S)
Game average score (\bar{n})	Internal energy (E)
Score difference (Δn)	Latent heat (L)
Critical multiplier ($1/\lambda_0$)	Transition temperature (T_0)

which is precisely the Helmholtz free energy (under the identification $\lambda = \beta = 1/T$),

$$F = E - TS. \quad (40)$$

If we are given a low enough value of energy (close to the energy of the ideal solid) then, despite the fact that the liquid phase has a larger entropy, we are forced to conclude that the system is in one of the (relatively) few solid phase points. Because this reversal of our prediction after knowing \bar{n} is strikingly unexpected, this situation is described by a large value of the Lagrange multiplier λ which, in the context of thermodynamics, corresponds to a low value of temperature T .

The limiting situation when we cannot claim to know the most probable phase happens when $\Delta F = 0$, which is the condition of thermodynamic phase coexistence. The Lagrange multiplier then is $\lambda_0 = \Delta S/\Delta n$, or, in thermodynamic notation,

$$T_0 = L/\Delta S(T_0), \quad (41)$$

where L is the latent heat associated with the first-order phase transition and $\Delta S(T_0)$ is the entropy difference at the transition temperature T_0 .

All these equivalences are summed up in Table 1.

6 Concluding Remarks

We have shown that, because to every yes/no question we can associate a change in evidence introduced by a new fact, there exist analogous quantities to the free energy difference between phases and the transition temperature, that are closely connected to this change in evidence. When the evidence is strong enough to completely cancel out our initial judgments about the probability of one phase over another and leave us undecided, the “weight” of this evidence is proportional to the transition inverse temperature.

Thus, in this view, the problem of thermodynamic equilibrium between phases is seen as answering the question: *is the system in phase A if we know that its average energy is \bar{E} ?* in a Bayesian/maximum entropy formalism. The concepts of transition temperature and free energy arise naturally as consequences of this inference framework, and therefore are not intrinsic properties of the systems or the phases.

Acknowledgments SD and JP gratefully acknowledge funding from FONDECYT grant 1140514. JP acknowledges partial support from FONDECYT 11130501. GG acknowledges partial support from FONDECYT 1120603. DG acknowledges funding from CONICYT PhD fellowship 21140914.

References

1. Jaynes, E.T.: Information theory and statistical mechanics. *Phys. Rev.* **106**, 620 (1957)
2. Jaynes, E.T.: *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge (2003)
3. Pressé, S., Ghosh, K., Lee, J., Dill, K.A.: Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **85**, 1115 (2013)
4. Uffink, J.: Can the maximum entropy principle be explained as a consistency requirement? *Stud. Hist. Philos. Sci. Part B* **26**, 223–261 (1995)
5. Dias, P.M.C., Shimony, A.: A critique of Jaynes' maximum entropy principle. *Adv. Appl. Math.* **2**, 172–211 (1981)
6. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948)
7. Tsallis, C.: Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **52**, 479–487 (1988)
8. Rényi, A.: In: *Proceedings of Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 547–561. University of California Press, Berkeley (1961)
9. Shore, J.E., Johnson, R.W.: Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory* **26**, 26–37 (1980)
10. Skilling, J.: The axioms of maximum entropy. In: Erickson, G.J., Smith, C.R. (eds.) *Maximum Entropy and Bayesian methods in Science and Engineering*, pp. 173–187. Kluwer Academic Publishers, Dordrecht (1988)
11. Caticha, A., Giffin, A.: Updating probabilities. *AIP Conf. Proc.* **872**, 31 (2006)
12. Pressé, S., Ghosh, K., Lee, J., Dill, K.A.: Nonadditive entropies yield probability distributions with biases not warranted by the data. *Phys. Rev. Lett.* **111**, 180604 (2013)
13. Callen, H.: *Thermodynamics and an Introduction to Thermostatistics*. Wiley, New York (1985)

PAPER • OPEN ACCESS

The Maximum Caliber principle applied to continuous systems

To cite this article: Diego González and Sergio Davis 2016 *J. Phys.: Conf. Ser.* **720** 012006

View the [article online](#) for updates and enhancements.

Related content

- [Hamiltonian formalism and path entropy maximization](#)
Sergio Davis and Diego González
- [Numerical study on the matching law between charge caliber and delay time of the rod-shaped explosively formed projectile](#)
H.M. Shen, W.B. Li, X.M. Wang et al.
- [Research of Dust Field Optimization Distribution Based on Parameters Change of Air Duct Outlet in Fully Mechanized Excavation Face of Coal Mine](#)
Xiao-Yan Gong, Zhi-Xin Xia, Yue Wu et al.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

The Maximum Caliber principle applied to continuous systems

Diego González, Sergio Davis

Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile
E-mail: dgonzalez@gmm.cl

Abstract. We give a brief presentation of the maximization of path entropy (Maximum Caliber) formalism as an approach to inference over trajectories. We develop its continuous-time version involving path integrals and present an identity between expectations over the most unbiased probability distribution of trajectories. This relation between expectations can be used to solve for the Lagrange multipliers in a Maximum Caliber problem without the need for the dynamical partition function.

1. Introduction

Although theoretically equilibrium is the fundamental concept in Thermodynamics, in practice most systems studied in a laboratory are systems out of equilibrium. This closeness to reality is what gives tremendous importance to the development of a formalism for describing these systems. Finding a formalism based on as few as possible principles and consistent with the principle of maximum entropy [1] in the case of equilibrium, which is still valid for a time-dependent treatment of statistical properties of a system, is the goal of much ongoing research. A proposal which has recently gained visibility in the statistical physics community is the principle of maximum path entropy or maximum *caliber* [2, 3].

In this framework, a probability functional associates a probability to every possible path the system can take when going from point A to point B . This probability functional can be directly used to predict expectations of functional properties and also instantaneous properties (through a *time-slicing* procedure), in what promises to become a useful tool for non-equilibrium systems. It is also expected that the non-equilibrium formalism by Evans *et al* [4] can be recovered from this single principle.

2. The Maximum Caliber principle

In the same way that the principle of Maximum Entropy (MaxEnt for short) produces the most unbiased probability distribution consistent with some constraint $\langle f(\vec{x}) \rangle = F$ by maximizing the Gibbs-Shannon entropy functional

$$S[p] = - \int d\vec{x} p(\vec{x}) \ln p(\vec{x}) \quad (1)$$

under said constraints, the principle of Maximum Caliber (MaxCal) replaces the microstates \vec{x} with micro-trajectories $x(t)$. Then, the functional corresponding to the Gibbs-Shannon entropy is called the *caliber*,



$$C[p] = - \int Dx() p[x()] \ln p[x()] \quad (2)$$

where the integral is now a functional integral over all trajectories (denoted by $Dx()$) and $p[x()]$ is a probability functional.

When the caliber in Eq. 2 is maximized under the constraints of normalization,

$$\int Dx() P[x()] = \langle 1 \rangle = 1, \quad (3)$$

and a known expectation of a functional

$$\langle F[x()] \rangle = \int Dx() P[x()] F[x()] = F_0, \quad (4)$$

the most unbiased distribution of trajectories is of the form

$$P[x()] = \frac{1}{Z} \exp(-\lambda F[x()]), \quad (5)$$

with Z a normalization constant, given by

$$Z = \int Dx() \exp(-\lambda F[x()]). \quad (6)$$

The Lagrange multiplier λ can be obtained, as in MaxEnt, from

$$-\frac{\partial}{\partial \lambda} \ln Z(\lambda) = F_0. \quad (7)$$

Equivalently, if an instantaneous property is known in expectation for each time t in an interval, $t \in [0, T]$,

$$\langle G(x(t), \dot{x}(t), t) \rangle = \int Dx() P[x()] G[x(); t] = g(t) \quad (8)$$

the probability functional is of the form

$$P[x()] = \frac{1}{\eta} \exp \left(- \int_0^T dt \lambda(t) G(x(t), \dot{x}(t); t) \right). \quad (9)$$

This is equivalent to the first kind of constraint (Eq. 4), because clearly if we integrate Eq. 8 in t from 0 to T we see that we are constraining the expectation of a *functional*

$$F[x()] = \int_0^T dt \lambda(t) G(x(t), \dot{x}(t); t), \quad (10)$$

in the sense of Eq. 4.

In the case of constraints as expressed in Eq. 8, the Lagrange multiplier function can be obtained, in an analogous way to Eq. 7, from

$$-\frac{\delta}{\delta \lambda(t)} \ln Z[\lambda()] = g(t). \quad (11)$$

If the instantaneous property G depends only on the instantaneous position $x(t)$ and velocity $\dot{x}(t)$ (and possible, explicitly on time), the functional F has the form of the classical action $A[x()]$ in mechanics

$$A[x()] = \int_0^T dt \mathcal{L}(x(t), \dot{x}(t); t), \quad (12)$$

and, in fact, if $\lambda > 0$ the most probable trajectory is the one which minimizes $A[x()]$, i.e., the principle of minimum action [5].

3. Conjugate variables theorem (CVT) for functionals

Once given the form of the probability functional (Eq. 5), we look for analytical tools that allows us to evaluate expectations of different properties. In MaxEnt there is such an identity, known as the conjugate variables theorem (CVT) [6], itself a direct consequence of the divergence theorem.

For an arbitrary distribution $p(\vec{x})$, we have

$$\langle \nabla \cdot \vec{v}(\vec{x}) \rangle + \langle \vec{v}(\vec{x}) \cdot \nabla \ln p \rangle = 0. \quad (13)$$

where the expectations are taken under the distribution p , and $\vec{v}(\vec{x})$ is an arbitrary (differentiable) vector field. We look for a generalization for MaxCal, i.e., when $\vec{x} \rightarrow x(t)$.

Consider the functional derivative of a product of functionals,

$$\frac{\delta}{\delta x(t)} (P[x()]G[x()]) = \frac{\delta P[x()]}{\delta x(t)} G[x()] + \frac{\delta G[x()]}{\delta x(t)} P[x()]. \quad (14)$$

If we rewrite the functional derivative of $P[x()]$ as

$$\frac{\delta P[x()]}{\delta x(t)} = P[x()] \frac{\delta}{\delta x(t)} \ln P[x()], \quad (15)$$

and replace in Eq. 14, we get

$$\frac{\delta}{\delta x(t)} (P[x()]G[x()]) = P[x()] \left[G[x()] \frac{\delta}{\delta x(t)} \ln P[x()] + \frac{\delta G[x()]}{\delta x(t)} \right]. \quad (16)$$

Now we integrate over all trajectories $x()$,

$$\int Dx() \frac{\delta}{\delta x(t)} (P[x()]G[x()]) = \int Dx() P[x()] G[x()] \frac{\delta}{\delta x(t)} \ln P[x()] + \int Dx() P[x()] \frac{\delta G[x()]}{\delta x(t)}, \quad (17)$$

and, as the probability functional P can be normalized, the “surface integral” on the left hand side vanishes. Therefore, we arrive at the following identity, generalization of Eq. 13 for functionals,

$$\left\langle \frac{\delta G[x()]}{\delta x(t)} \right\rangle + \left\langle G[x()] \frac{\delta}{\delta x(t)} \ln P[x()] \right\rangle = 0, \quad (18)$$

where $G[x()]$ corresponds to an arbitrary functional. For a Maximum Caliber probability functional (Eq. 5), this reduces to

$$\left\langle \frac{\delta G[x()]}{\delta x(t)} \right\rangle = \lambda \left\langle G[x()] \frac{\delta}{\delta x(t)} F[x()] \right\rangle. \quad (19)$$

This means we can avoid Eq. 7 to solve for λ , simply obtaining it from

$$\lambda = \frac{\left\langle \frac{\delta}{\delta x(t)} G[x()] \right\rangle}{\left\langle G[x()] \frac{\delta}{\delta x(t)} F[x()] \right\rangle}. \quad (20)$$

Using Eq. 19 with a constant functional, $G[x()] = 1$, we see that

$$\left\langle \frac{\delta}{\delta x(t)} F[x(\cdot)] \right\rangle = \left\langle \frac{\delta}{\delta x(t)} A[x(\cdot)] \right\rangle = 0, \quad (21)$$

that is, the principle of minimum action also holds in expectation. A particular case of this relation is the result in Ref. [5],

$$\langle \dot{p}(t) \rangle = -\langle \Phi'(x(t)) \rangle, \quad (22)$$

which is the expectation of Newton's second law.

4. Conclusions

We have shown that the Maximum Caliber principle is a powerful tool to study dynamical systems in a probabilistic framework. A generalization of the conjugate variables theorem (CVT) is presented for probability functionals, from which we prove that the minimum action principle holds also in expectation over the ensemble of trajectories. This “functional CVT” can also be used to find novel relations between expectations of functionals.

Acknowledgments

Both authors acknowledge funding from FONDECYT grant 1140514. DG also acknowledges funding from CONICYT PhD fellowship 21140914.

References

- [1] Jaynes E T 1957 *Physical Review* **106** 620–630
- [2] Jaynes E T 1980 *Ann. Rev. Phys. Chem.* **31** 579–601
- [3] Pressé S, Ghosh K, Lee J and Dill K A 2013 *Reviews of Modern Physics* **85** 1115–1141
- [4] Evans D J, Searles D J and Williams S R 2008 *J. Chem. Phys.* **128** 014504
- [5] González D, Davis S and Gutiérrez G 2014 *Found. Phys.* **44** 923
- [6] Davis S and Gutiérrez G 2012 *Phys. Rev. E* **86** 051136

Liouville's theorem from the principle of maximum caliber in phase space

Diego González, and Sergio Davis

Citation: [AIP Conference Proceedings](#) **1757**, 020003 (2016);

View online: <https://doi.org/10.1063/1.4959044>

View Table of Contents: <http://aip.scitation.org/toc/apc/1757/1>

Published by the [American Institute of Physics](#)

Articles you may be interested in

[Communication: Maximum caliber is a general variational principle for nonequilibrium statistical mechanics](#)

[The Journal of Chemical Physics](#) **143**, 051104 (2015); 10.1063/1.4928193

[Why square roots of probabilities?](#)

[AIP Conference Proceedings](#) **1757**, 020001 (2016); 10.1063/1.4959042

[Applications of the divergence theorem in Bayesian inference and MaxEnt](#)

[AIP Conference Proceedings](#) **1757**, 020002 (2016); 10.1063/1.4959043

[Maximum entropy derivation of quasi-Newton methods](#)

[AIP Conference Proceedings](#) **1757**, 020004 (2016); 10.1063/1.4959045

Liouville's Theorem from the Principle of Maximum Caliber in Phase Space

Diego González^{1,a)} and Sergio Davis^{1,b)}

¹*Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile.*

^{a)}Corresponding author: dgonzalez@gnm.cl

^{b)}sdavis@gnm.cl

Abstract. One of the cornerstones in non-equilibrium statistical mechanics (NESM) is Liouville's theorem, a differential equation for the phase space probability $\rho(q, p; t)$. This is usually derived considering the flow in or out of a given surface for a physical system (composed of atoms), via more or less heuristic arguments.

In this work, we derive the Liouville equation as the partial differential equation governing the dynamics of the time-dependent probability $\rho(q, p; t)$ of finding a "particle" with Lagrangian $L(q, \dot{q}; t)$ in a specific point (q, p) in phase space at time t , with $p = \partial L / \partial \dot{q}$. This derivation depends only on considerations of inference over a space of continuous paths. Because of its generality, our result is valid not only for "physical" systems but for any model depending on constrained information about position and velocity, such as time series.

INTRODUCTION

Equilibrium Statistical Mechanics was unified and shown to be an inference procedure by E. T. Jaynes in 1957, who introduced the (now well-established) principle of Maximum Entropy [1]. This principle provides a systematic procedure to construct a probability distribution and, from it, to obtain macroscopic properties of thermodynamic systems.

For non-equilibrium systems there are few exact results, most expressed in terms of differential equations for time-dependent properties and probability distributions. As a starting point in many derivations of classical statistical mechanics lies the Liouville equation

$$\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \left\{ \rho, \mathcal{H} \right\} = 0, \quad (1)$$

which describes the time evolution of the probability density of microstates with constant energy. Besides this fundamental equation, there is no unifying axiomatic approach, in strong contrast with the equilibrium theory.

As such unifying proposal, Jaynes in 1980 introduced a straightforward extension of his Maximum Entropy principle, the *Maximum Caliber* Principle [2], to dynamical systems. In this proposal, the path entropy or *caliber* is maximized subject to constraints to obtain the most unbiased path probability distribution.

We have previously [3] shown that Maximum Caliber predicts Newton's second law $F = m \cdot a$ for paths from informational constraints on position and velocity, and recently [4] that a Boltzmann transport equation arises naturally from the formalism in phase space for the same constraints.

In this work we continue in the same line of reasoning, and show that the Liouville equation is the fundamental equation that describes the instantaneous probability distribution $\rho(q, p; t)$ in phase space in a problem of inference over an ensemble of trajectories $\vec{q}(t)$ and $\vec{p}(t)$, described by a probability functional $P[\vec{q}, \vec{p}]$.

The Maximum Caliber Principle

The Maximum Caliber principle can be seen as a generalization of the Maximum Entropy principle for assigning a probability distribution to an ensemble of trajectories (or paths) given some information in the form of expectations.

Consider a dynamical system described by a path $\bar{x}(t) \in \mathbb{X}$, where \mathbb{X} represents the space of continuous paths $\bar{x}(t)$ with $t \in [0, \tau]$ and fixed boundary conditions $x(0) = x_0$ and $x(\tau) = x_\tau$. If the information we have about the system is given as the known value G_0 of the expectation of a functional $G[\bar{x}(t)]$, that is,

$$\langle G[\bar{x}(t)] \rangle = \int_{\mathbb{X}} \mathcal{D}\bar{x} P[\bar{x}(t)] G[\bar{x}(t)] = G_0, \quad (2)$$

then the most unbiased probability functional $P[\bar{x}(t)]$ compatible with this constraint and the implicit constraint of normalization over \mathbb{X} ,

$$\langle 1 \rangle = \int_{\mathbb{X}} \mathcal{D}\bar{x} P[\bar{x}(t)] = 1, \quad (3)$$

is the one that maximizes the *caliber*

$$C[P] = - \int_{\mathbb{X}} \mathcal{D}\bar{x} P[\bar{x}(t)] \ln \frac{P[\bar{x}(t)]}{\pi[\bar{x}(t)]}. \quad (4)$$

which is in fact the Shannon-Jaynes entropy in path space with $\pi[\bar{x}(t)]$ the invariant measure of \bar{x} . The probability functional which maximizes $C[P]$ is of the form

$$P[\bar{x}(t)|G_0] = \pi[\bar{x}(t)] \frac{\exp(-\beta G[\bar{x}(t)])}{Z(\beta)}, \quad (5)$$

with β a Lagrange multiplier, determined by the constraint equation

$$-\frac{\partial}{\partial \beta} \ln Z(\beta) = G_0. \quad (6)$$

This is the approach most similar to maximum entropy. On the other hand, if the information we have on the system is given as an infinite set of instantaneous constraints

$$\langle L[\bar{x}(t); t] \rangle = L_0(t) \quad (7)$$

for each value of $t \in [0, \tau]$, maximization of $C[P]$ leads to the probability functional

$$P[\bar{x}(t)|L_0(t)] = \frac{\exp\left(-\int_0^\tau dt \beta(t) L[\bar{x}(t); t]\right)}{Z[\beta(t)]}, \quad (8)$$

where now the Lagrange multiplier $\beta(t)$ is a function of time, determined by

$$-\frac{\delta}{\delta \beta(t)} \ln Z[\beta(t)] = L_0(t). \quad (9)$$

This is equivalent to the first kind of constraint (Eq. 2) if we identify the functional G as

$$\beta G[\bar{x}(t)] = \int_0^\tau dt \beta(t) L[\bar{x}(t); t]. \quad (10)$$

Time-slicing of the path distribution in phase space and the continuity equation for probability

Although the Maximum Caliber principle is promising as the definitive solution for inference of time-independent systems (including non-equilibrium Statistical Mechanics), it comes expressed in the language of paths $\bar{x}(t)$ and probability functionals $P[\bar{x}(t)]$. In practice it is often needed to express the information in terms of the instantaneous probability density of position $\rho(x; t)$ and the joint probability distribution of position and velocity, $\rho(x, \dot{x}; t)$. The procedure that discards information when going from the functional representation to the instantaneous representation we will call *time-slicing*. The *time-slicing problem* is, then, to define the general procedure by which to obtain instantaneous probability densities for arbitrary quantities $G[\bar{x}(t); t]$ at arbitrary times $t \in [0, \tau]$.

For a classical mechanical system with a well-defined Lagrangian $\mathcal{L}(q, \dot{q}; t)$ a phase space can be constructed by defining the momentum p as

$$p = \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right). \quad (11)$$

In the same way, each coordinate trajectory $\bar{q}(t)$ has an associated ‘‘momentum trajectory’’ $\bar{p}(t)$ obtained by applying the same rule for each time $t \in [0, \tau]$,

$$\bar{p}(t) = \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \Big|_i. \quad (12)$$

If we have a path space Γ with fixed initial position and momentum, for instance, $q(0) = q_0$ and $p(0) = p_0$, then the normalization condition for the probability functional is

$$\langle 1 \rangle = \int_{\Gamma} D\bar{q} D\bar{p} P[\bar{q}(), \bar{p}()] = 1. \quad (13)$$

Now we want to determine the probability $\rho(q', p'; t)$ of being at a point (q', p') in phase space at a posterior time $t > 0$. This will be given by

$$\rho(q', p', t) = \langle \delta(\bar{q}(t) - q') \delta(\bar{p}(t) - p') \rangle. \quad (14)$$

If we take a partial derivative with respect to t of the above expression, then we obtain

$$\frac{\partial \rho(q', p', t)}{\partial t} + \frac{\partial}{\partial q'} \langle \delta(\bar{q}(t) - q') \delta(\bar{p}(t) - p') \frac{d\bar{q}}{dt}(t) \rangle + \frac{\partial}{\partial p'} \langle \delta(\bar{q}(t) - q') \delta(\bar{p}(t) - p') \frac{d\bar{p}}{dt}(t) \rangle = 0, \quad (15)$$

and, rewriting the expectations of the delta functions in the form

$$\langle \delta(\bar{q}(t) - q') \delta(\bar{p}(t) - p') \frac{d\bar{q}}{dt}(t) \rangle = \rho(q', p'; t) \left\langle \frac{d\bar{q}}{dt}(t) \right\rangle_c \quad (16)$$

$$\langle \delta(\bar{q}(t) - q') \delta(\bar{p}(t) - p') \frac{d\bar{p}}{dt}(t) \rangle = \rho(q', p'; t) \left\langle \frac{d\bar{p}}{dt}(t) \right\rangle_c, \quad (17)$$

where the subscripted expectation $\langle \cdot \rangle_c$ denotes a time-sliced expectation, conditional to $\bar{q}(t) = q'$ and $\bar{p}(t) = p'$, we find that the probability density ρ for the phase space coordinate (q', p') follows a continuity equation,

$$\frac{\partial \rho(q', p'; t)}{\partial t} + \frac{\partial}{\partial q'} (\rho(q', p'; t) U_q(q', p'; t)) + \frac{\partial}{\partial p'} (\rho(q', p'; t) U_p(q', p'; t)) = 0, \quad (18)$$

with current velocity components

$$U_q = \left\langle \frac{d\bar{q}}{dt}(t) \right\rangle_c, \quad (19)$$

$$U_p = \left\langle \frac{d\bar{p}}{dt}(t) \right\rangle_c. \quad (20)$$

In order to obtain a particular partial differential equation for ρ we must first determine the explicit form of this current velocity as a function of the phase space coordinate. In order to do this in the general case, we will need to introduce some assumptions about the particular form of the probability functional $P[\bar{q}(), \bar{p}()]$, and this we will achieve by invoking the maximum caliber principle.

Probability Functional for Paths in Phase Space

If the classical action is constrained for a physical system by

$$\left\langle \int_0^\tau dt \mathcal{L}(\bar{q}(t), \frac{d\bar{q}}{dt}(t); t) \right\rangle = \int \mathcal{D}q P[\bar{q}()] S[\bar{q}()] = S_0 \quad (21)$$

then, according to the maximum caliber principle, the most unbiased probability functional for the coordinate path \bar{q} that agrees with that information will be given by

$$P[\bar{q}()|S_0] = \frac{\exp(-\beta \int_0^\tau dt \mathcal{L}(\bar{q}(t), \frac{d\bar{q}}{dt}(t); t))}{Z(\beta)}, \quad (22)$$

with β a Lagrange multiplier. From this we can construct the probability functional for the phase space path $P[\bar{q}(), \bar{p}()]$ using the product rule of probability theory,

$$P[\bar{q}(), \bar{p}()|S_0] = P[\bar{q}()|S_0] P[\bar{p}()|\bar{q}(), S_0], \quad (23)$$

and the fact that complete knowledge of $\bar{q}()$ implies complete knowledge of $\bar{p}()$, so $P[\bar{p}()|\bar{q}(), S_0]$ must be a functional Dirac's delta,

$$P[\bar{p}()|\bar{q}(), \mathcal{I}] = \delta \left[\frac{\partial L}{\partial \dot{q}} - \bar{p}() \right] \approx \prod_i \delta \left(\frac{\partial L}{\partial \dot{q}(t_i)} - p(t_i) \right), \quad (24)$$

for any proposition \mathcal{I} not in conflict with $\bar{q}()$. This strong constraint, which fixes $\bar{p}(t)$ for every instant $t \in [0, \tau]$, can be translated into a single delta constraint, $P[\bar{p}()|\bar{q}()] = \delta(\epsilon_p[\bar{q}()])$, where we introduce a new functional ϵ_p ,

$$\epsilon_p[\bar{q}()] = \int_0^\tau dt \left[\bar{p}(t) - \frac{\partial L}{\partial \dot{q}(t)} \right]^2. \quad (25)$$

Now our phase space probability functional can be written as

$$P[\bar{q}(), \bar{p}()|S_0] = \frac{1}{\eta(\beta)} \delta(\epsilon_p[\bar{q}()]) \exp \left(-\beta \int_0^\tau dt \mathcal{L}(\bar{q}(t), \frac{d\bar{q}}{dt}(t); t) \right) \quad (26)$$

where η is a normalization factor,

$$\eta(\beta) = \int \mathcal{D}\bar{q} \mathcal{D}\bar{p} \delta(\epsilon_p[\bar{q}()]) \exp \left(-\beta \int_0^\tau dt \mathcal{L}(\bar{q}(t), \frac{d\bar{q}}{dt}(t); t) \right), \quad (27)$$

and where the delta function on ϵ_p plays the role of a prior probability on phase space. The quadratic form of the integrand in ϵ_p makes it relatively easy to evaluate its functional derivatives. It only remains to write the exponent in terms of $\bar{q}(t)$ and $\bar{p}(t)$, and for this we use the Legendre transformation

$$\mathcal{H}(q, p) = p\dot{q} - \mathcal{L}(q, \dot{q}; t) \quad (28)$$

and insert the Hamiltonian \mathcal{H} , leading to

$$P[\bar{q}(), \bar{p}()|S_0] = \frac{1}{\eta(\beta)} \delta(\epsilon_p[\bar{q}()]) \exp \left(-\beta \int_0^\tau dt \left[\bar{p}(t) \frac{d\bar{q}}{dt}(t) - \mathcal{H}(\bar{q}(t), \bar{p}(t), t) \right] \right). \quad (29)$$

We will extract information about the current velocity in the continuity equation (Eq. 18) by using a recently derived general identity between expectations known as the conjugate variables theorem [5, 4].

Functional form of the Conjugate Variables Theorem

The conjugate variables theorem [5] (CVT) for N -dimensional state space is the identity

$$\left\langle \frac{\partial \omega(\vec{x})}{\partial x_i} \right\rangle + \left\langle \omega(\vec{x}) \frac{\partial}{\partial x_i} \ln \rho(\vec{x}) \right\rangle = 0. \quad (30)$$

where $\omega(x)$ is an arbitrary, differentiable function of the state x , and $\rho(x)$ is the time-independent probability distribution of x . Its natural extension to probability functionals is

$$\left\langle \frac{\delta W[x(t)]}{\delta x(t)} \right\rangle + \left\langle W[x(t)] \frac{\delta}{\delta x(t)} \ln P[x(t)] \right\rangle = 0, \quad (31)$$

as put forward in Ref. [4]. For our probability functional in phase space (Eq. 29), the CVT gives us two independent identities,

$$\begin{aligned} \left\langle \frac{\delta W[\bar{q}(t), \bar{p}(t)]}{\delta \bar{q}(t)} \right\rangle &= -\beta \left\langle W[\bar{q}(t), \bar{p}(t)] \left(\frac{d\bar{p}}{dt}(t) + \frac{\partial \mathcal{H}}{\partial q} \right) \right\rangle - \left\langle W[\bar{q}(t), \bar{p}(t)] \frac{\delta}{\delta \bar{q}(t)} \ln \delta(\epsilon_p[\bar{q}(t)]) \right\rangle, \\ \left\langle \frac{\delta W[\bar{q}(t), \bar{p}(t)]}{\delta \bar{p}(t)} \right\rangle &= \beta \left\langle W[\bar{q}(t), \bar{p}(t)] \left(\frac{d\bar{q}}{dt}(t) - \frac{\partial \mathcal{H}}{\partial p} \right) \right\rangle - \left\langle W[\bar{q}(t), \bar{p}(t)] \frac{\delta}{\delta \bar{p}(t)} \ln \delta(\epsilon_p[\bar{q}(t)]) \right\rangle. \end{aligned} \quad (32)$$

Note that in the expressions above we use the δ symbol to denote both the functional derivative $\delta/\delta \bar{q}$ and Dirac's delta function $\delta(\cdot)$, each use we hope is clear in context. The functional derivatives of $\delta(\epsilon_p[\bar{q}(t)])$ involve the functional derivatives of $\epsilon_p[\bar{q}(t)]$,

$$\frac{\delta}{\delta \bar{q}(t)} \epsilon_p[\bar{q}(t)] = -2 \left[\bar{p}(t) - \frac{\partial \mathcal{L}}{\partial \dot{q}} \right] \left(\frac{\partial^2 \mathcal{L}}{\partial \dot{q} \partial q} - \frac{d}{dt} \left(\frac{\partial^2 \mathcal{L}}{\partial \dot{q}^2} \right) \right), \quad (33)$$

$$\frac{\delta}{\delta \bar{p}(t)} \epsilon_p[\bar{q}(t)] = 2 \left[\bar{p}(t) - \frac{\partial \mathcal{L}}{\partial \dot{q}} \right], \quad (34)$$

through the chain rule, and these derivatives are zero when evaluated on the condition itself. Therefore, the second terms in the right-hand side of Eqs. 32 both vanish for any W . The CVT reduces then to

$$\begin{aligned} \left\langle \frac{\delta W[\bar{q}(t), \bar{p}(t)]}{\delta \bar{q}(t)} \right\rangle &= -\beta \left\langle W[\bar{q}(t), \bar{p}(t)] \left(\frac{d\bar{p}}{dt}(t) + \frac{\partial \mathcal{H}}{\partial q} \right) \right\rangle, \\ \left\langle \frac{\delta W[\bar{q}(t), \bar{p}(t)]}{\delta \bar{p}(t)} \right\rangle &= \beta \left\langle W[\bar{q}(t), \bar{p}(t)] \left(\frac{d\bar{q}}{dt}(t) - \frac{\partial \mathcal{H}}{\partial p} \right) \right\rangle. \end{aligned} \quad (35)$$

As an example of the use of this general identity, we take the simplest non-trivial choice $W = 1$, and readily obtain Hamilton's equations in expectation,

$$\left\langle \frac{d\bar{q}}{dt}(t) \right\rangle = \left\langle \left(\frac{\partial H}{\partial p} \right) \Big|_t \right\rangle, \quad (36)$$

$$\left\langle \frac{d\bar{p}}{dt}(t) \right\rangle = \left\langle \left(\frac{\partial H}{\partial q} \right) \Big|_t \right\rangle, \quad (37)$$

which are in agreement with Newton's second law in the particular case obtained in Ref. [3] from Maximum Caliber.

Derivation of the Liouville equation

Now we are equipped to compute the components of the current velocity U_q and U_p in the continuity equation (Eq. 18). To construct these time-sliced expectations we take Eq. 35 and choose $W = \delta(\bar{q}(t) - q') \delta(\bar{p}(t) - p')$. After some more or less straightforward computation we obtain

$$\begin{aligned}\left\langle \frac{d\bar{q}}{dt}(t) \right\rangle_c &= U_q = \frac{1}{\beta} \frac{\partial}{\partial p'} \ln \rho + \frac{\partial H}{\partial p'}, \\ \left\langle \frac{d\bar{p}}{dt}(t) \right\rangle_c &= U_p = -\frac{1}{\beta} \frac{\partial}{\partial q'} \ln \rho - \frac{\partial H}{\partial q'}.\end{aligned}\tag{38}$$

We are ready to place these expressions into the continuity equation (Eq. 18) from which we readily find that the divergence of the current velocity is zero:

$$\frac{\partial}{\partial q'} U_q + \frac{\partial}{\partial p'} U_p = 0.\tag{39}$$

Therefore the continuity equation becomes the Liouville equation:

$$\frac{\partial \rho}{\partial t} + \left\{ \rho, H \right\} = 0.\tag{40}$$

CONCLUSIONS

We have obtained the Liouville equation as a direct consequence of the Maximum Caliber formalism for any system with classical action given by the integral of a Lagrangian function dependent on position and velocity. The derivation is self-contained: it does not assume anything besides the validity of the Maximum Caliber formalism and the Legendre transformation. This makes the use of phase space ideas in inference for dynamical systems natural and independent of physical considerations, being for instance suited to the study of time series.

ACKNOWLEDGMENTS

The authors acknowledge support from FONDECYT grant 1140514. DG acknowledges funding from CONICYT PhD fellowship 21140914.

REFERENCES

- [1] E. T. Jaynes, [Physical Review](#) **106**, 620–630 (1957).
- [2] E. T. Jaynes, [Ann. Rev. Phys. Chem.](#) **31**, 579–601 (1980).
- [3] D. González, S. Davis, and G. Gutiérrez, [Found. Phys.](#) **44**, p. 923 (2014).
- [4] S. Davis and D. González, [J. Phys. A: Math. Theor.](#) **48**, p. 425003 (2015).
- [5] S. Davis and G. Gutiérrez, [Phys. Rev. E](#) **86**, p. 051136 (2012).

Continuity equation for probability as a requirement of inference over paths

Diego González^{1,3}, Daniela Díaz², and Sergio Davis^{3,a}

¹ Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

² Facultad de Física, Pontificia Universidad Católica de Chile, Santiago, Chile

³ Comisión Chilena de Energía Nuclear, Casilla 188-D, Santiago, Chile

Received 17 May 2016 / Received in final form 19 July 2016

Published online 3 October 2016 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2016

Abstract. Local conservation of probability, expressed as the continuity equation, is a central feature of non-equilibrium Statistical Mechanics. In the existing literature, the continuity equation is always motivated by heuristic arguments with no derivation from first principles. In this work we show that the continuity equation is a logical consequence of the laws of probability and the application of the formalism of inference over paths for dynamical systems. That is, the simple postulate that a system moves continuously through time following paths implies the continuity equation. The translation between the language of dynamical paths to the usual representation in terms of probability densities of states is performed by means of an identity derived from Bayes' theorem. The formalism presented here is valid independently of the nature of the system studied: it is applicable to physical systems and also to more abstract dynamics such as financial indicators, population dynamics in ecology among others.

1 Introduction

The study of non-equilibrium systems is an active area of research with important applications in biology [1], materials science [2], fluid [3] and plasma physics, financial modelling and other complex systems, as well as in more fundamental research aiming to understand phenomena such as irreversibility and the origin of the second law of Thermodynamics. Unlike equilibrium statistical mechanics, which is a well-established and unified theory that can be derived in its entirety from a single fundamental principle, namely the principle of Maximum Entropy (MaxEnt) [4], non-equilibrium Statistical Mechanics [5,6] currently lacks this kind of axiomatic formulation: rather it is presented as a set of partial differential equations and identities, each one with restricted validity (for instance, some of them are only valid in Onsager's linear regime). However, there is one very promising candidate for a unifying principle in non-equilibrium Statistical Mechanics, the principle of Maximum Caliber [7–9], which is the application of MaxEnt to the space of dynamical trajectories. The success of the Maximum Caliber approach suggests that a powerful formulation of non-equilibrium SM relies on the concept of paths, and therefore a framework for path probabilities is required to advance in this field. More details about this principle are explained in the Appendix.

It is immediate to note that all the different partial differential equations for the time-dependent probability $\rho(\mathbf{x}; t)$ appearing in non-equilibrium statistical mechanics are, in fact, continuity equations of the form

$$\partial_t \rho + \nabla(\rho \mathbf{v}) = 0. \quad (1)$$

Therefore it is possible to obtain the full set of partial differential equations as particular cases of a continuity equation with different *flow velocities* \mathbf{v} . In turn, the continuity equation, which represents local conservation of the probability, is usually obtained via *ad hoc* arguments [10] involving fluxes in and out of a given surface. This is of course valid, but it is not clear what is the range of validity of this kind of derivation for more abstract systems.

Taking the idea of path inference as fundamental, in this work we show that the continuity equation (Eq. (1)) is a direct consequence of performing statistical inference in the space of paths from a point A to a point B . In the derivation of this result we employ what we call the *time-slicing* equation, a rule which extracts information about the probability of different microscopic states on a slice of constant time from an ensemble of dynamical paths.

2 Continuity equation and Non-equilibrium Statistical Mechanics

A crucial element in Non-equilibrium Statistical Mechanics is the time-dependent probability density of microstates $\rho(\mathbf{x}; t)$. This is typically obtained by solving a

^a e-mail: sdavis@gnm.c1

partial differential equation appropriate for the system, for instance the Fokker-Planck equation for several kinds of Brownian motion models, and the Liouville Equation in the case of time evolution in phase space. All these equations can be recast as particular cases of the continuity equation (Eq. (1)). Let us write this equation using index notation as:

$$\partial_t \rho(\mathbf{x}; t) + \partial_\alpha (\rho(\mathbf{x}; t) v^\alpha(\mathbf{x}, t)) = 0, \quad (2)$$

where $\partial_\alpha = \partial/\partial x_\alpha$ with $\alpha = 1, 2, 3$, and $\partial_t = \partial/\partial t$. Here the different forms of the probability current $\mathbf{v}(\mathbf{x}, t)$ give us the different equations for the probability density ρ . As an example, consider the case where $v^\alpha(\mathbf{x}, t)$ is given by:

$$v^\alpha(\mathbf{x}, t) = \mu^\alpha(\mathbf{x}, t) - \frac{1}{2} \partial_\beta D^{\alpha\beta}(\mathbf{x}, t) - \frac{1}{2} D^{\alpha\beta}(\mathbf{x}, t) \partial_\beta \ln \rho(\mathbf{x}; t), \quad (3)$$

with μ^α the components of a drift vector and $D^{\alpha\beta}$ the components of a diffusion tensor. Replacing in equation (2) we recover the Fokker-Planck equation [11] in the Ito representation,

$$\partial_t \rho(\mathbf{x}; t) + \partial_\alpha (\rho(\mathbf{x}; t) \mu^\alpha(\mathbf{x}, t)) - \frac{1}{2} \partial_\alpha \partial_\beta (D^{\alpha\beta}(\mathbf{x}, t) \rho(\mathbf{x}; t)) = 0. \quad (4)$$

Alternatively, the probability current given by

$$v^\alpha(\mathbf{x}, t) = \mu^\alpha(\mathbf{x}, t) - \frac{1}{4} \partial_\beta D^{\alpha\beta}(\mathbf{x}, t) - \frac{1}{2} D^{\alpha\beta}(\mathbf{x}, t) \partial_\beta \ln \rho(\mathbf{x}; t). \quad (5)$$

Leads to the Stratonovich form of the Fokker-Planck equation,

$$\begin{aligned} \partial_t \rho(\mathbf{x}; t) + \partial_\alpha (\rho(\mathbf{x}; t) \mu^\alpha(\mathbf{x}, t)) \\ - \frac{1}{2} \partial_\alpha \partial_\beta (D^{\alpha\beta}(\mathbf{x}, t) \rho(\mathbf{x}; t)) \\ + \frac{1}{2} \partial_\alpha \left(\frac{1}{2} \partial_\beta D^{\alpha\beta} \right) = 0. \end{aligned} \quad (6)$$

Similarly, the Liouville equation

$$\partial_t \rho(\mathbf{r}, \mathbf{p}; t) + \{ \rho, \mathcal{H} \} = 0, \quad (7)$$

with $\{A, B\}$ the Poisson bracket,

$$\{A, B\} = \sum_{i=1}^{3N} \frac{\partial A}{\partial x_i} \frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial B}{\partial x_i}, \quad (8)$$

can be recovered with the choice

$$\mathbf{v}(\mathbf{r}, \mathbf{p}, t) = - \sum_{i=1}^{3N} \hat{e}_i \frac{\partial \mathcal{H}}{\partial x_i} + \sum_{i=3N+1}^{6N} \hat{e}_i \frac{\partial \mathcal{H}}{\partial p_i}. \quad (9)$$

It seems that establishing the formal basis for Non-equilibrium Statistical Mechanics requires a solid justification for the nature of the continuity equation and its range of applicability. In the following sections we will deduce the general continuity equation as in equation (1) purely from considerations of statistical inference over dynamical paths.

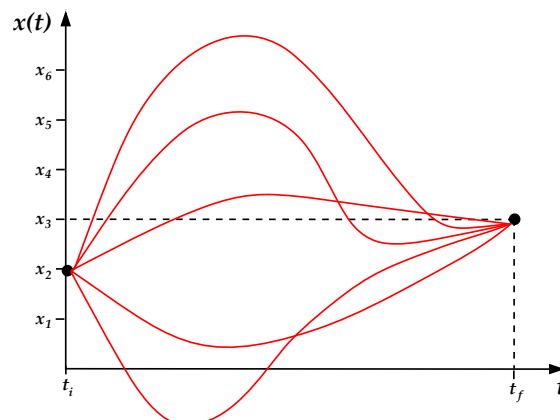


Fig. 1. Some continuous paths in a path space \mathbb{X} with boundary conditions $X(t_i) = x_2$ and $X(t_f) = x_3$.

3 Inference over the space of paths and the time-slicing equation

Dynamical systems (parameterized by time t) follow paths which are smooth curves (although not necessarily differentiable), such that the coordinates at a time t are given by vector functions $\mathbf{X}(t)$. We will assume that there is some uncertainty about the path the system will take when going from a point A to a point B , and that justifies the use of probabilities.

Before proceeding, let us make a few remarks about notation. In the following, we will denote a complete path by its function $\mathbf{X}()$ without explicitly writing its argument. A functional G of the path $\mathbf{X}()$ will be denoted by $G[\mathbf{X}()]$. In contrast, a state of the system will be denoted by \mathbf{x} , in lowercase, so that, for instance, $\mathbf{x}_0 = \mathbf{X}(t_0)$ is the condition that the system is in state \mathbf{x}_0 at time t_0 when following the path $\mathbf{X}()$.

Imposing fixed boundary conditions defines a space of smooth paths \mathbb{X} such that the actual evolution of a system in a given realization of the process is described by some path $\mathbf{X}() \in \mathbb{X}$. An example of such a path space \mathbb{X} in one spatial dimension, where paths have fixed boundary conditions $X(t_i) = x_2$ and $X(t_f) = x_3$, is sketched in Figure 1.

Now, in order to do inference over dynamical properties we can introduce the concept of a *probability functional* $\rho[\mathbf{X}()]$, which gives the probability density assigned to every possible path $\mathbf{X}() \in \mathbb{X}$. If this functional is known, we can in principle estimate any observable $G[\mathbf{X}()]$ by computing its expectation

$$\langle G \rangle = \int_{\mathbb{X}} \mathcal{D}\mathbf{X}() \rho[\mathbf{X}()] G[\mathbf{X}()]. \quad (10)$$

These expectations are given by path integrals. However, we usually need to determine the expectation of instantaneous quantities $g(\mathbf{X}(t))$ which should be estimated using the instantaneous (or time-dependent) probability density

of the states $\rho(\mathbf{x}; t)$, as:

$$\langle g(\mathbf{X}(t)) \rangle = \langle g(\mathbf{x}) \rangle_t = \int d\mathbf{x} \rho(\mathbf{x}; t) g(\mathbf{x}). \quad (11)$$

The connection between the two formalisms is given by the representation of $\rho(\mathbf{x}; t)$ as an expectation of a Dirac delta functional over the path distribution,

$$\rho(\mathbf{x}; t) = \left\langle \delta(\mathbf{X}(t) - \mathbf{x}) \right\rangle = \int_{\mathbb{X}} \mathcal{D}\mathbf{X}() \rho[\mathbf{X}()] \delta(\mathbf{X}(t) - \mathbf{x}). \quad (12)$$

This equality defines the *time-sliced* or instantaneous probability, and we will refer to it as the *time-slicing equation*. Notice that we employ the same symbol ρ for the probability density of states and for the probability functional of paths, each case should be clear in context by the presence of square brackets to signal functional evaluation.

4 Properties of the time-slicing equation

When performing operations on probabilities, a more explicit notation is preferable, in which we use the *probability operator* (denoted by the capital letter P). The operator P is a binary operator, which takes two assertions (or logical propositions) A and B and combines them giving a real, non-negative number $P(A|B)$. This should be read as “the probability of A being true given that B is true”. The probabilities we have employed so far, $\rho(\mathbf{x}; t)$ and $\rho[\mathbf{Y}()]$, can be written explicitly as:

$$\rho(\mathbf{x}; t) = P(\mathbf{X}(t) = \mathbf{x}|I), \quad (13)$$

$$\rho[\mathbf{Y}()] = P(\mathbf{X}() = \mathbf{Y}()|I), \quad (14)$$

where $\mathbf{X}()$ represents the real path followed by the system, and I is a *prior* state of knowledge. Using these definitions and the rules for probability, we will prove a more general relation,

$$\left\langle \delta(\mathbf{X}(t) - \mathbf{x}) G[\mathbf{X}()] \right\rangle_I = \rho(\mathbf{x}; t) \left\langle G[\mathbf{X}()] \right\rangle_{\mathbf{x}, t}, \quad (15)$$

where $\langle G \rangle_{\mathbf{x}, t}$ is the *time-sliced expectation* of the functional G , i.e., the expectation that considers only the paths where $\mathbf{X}(t) = \mathbf{x}$. This may be considered a *general time-slicing equation*, which allows to “slice” any functional G by taking expectation with the appropriate delta function. To see why this relation is true, let us write the expectation in the left-hand side of equation (15) as a path integral,

$$\left\langle \delta(\mathbf{X}(t) - \mathbf{x}) G[\mathbf{X}()] \right\rangle_I = \int_{\mathbb{X}} \mathcal{D}\mathbf{X}() P(\mathbf{X}()|I) \times \delta(\mathbf{X}(t) - \mathbf{x}) G[\mathbf{x}()], \quad (16)$$

and recognize the Dirac delta function as the probability of the state at a specific time given the path, that is,

$$P(\mathbf{X}(t) = \mathbf{x}|\mathbf{X}()) = \delta(\mathbf{X}(t) - \mathbf{x}). \quad (17)$$

Invoking Bayes’ theorem as:

$$P(\mathbf{X}(t) = \mathbf{x}|\mathbf{X}()) = \frac{P(\mathbf{X}(t) = \mathbf{x}|I) P(\mathbf{X}()|\mathbf{X}(t) = \mathbf{x}, I)}{P(\mathbf{X}()|I)} \quad (18)$$

and replacing in equation (15), we finally obtain

$$\begin{aligned} \left\langle \delta(\mathbf{X}(t) - \mathbf{x}) G[\mathbf{X}()] \right\rangle_I &= \int_{\mathbb{X}} \mathcal{D}\mathbf{X}() P(\mathbf{X}(t) = \mathbf{x}|I) \\ &\times P(\mathbf{X}()|\mathbf{X}(t) = \mathbf{x}, I) G[\mathbf{X}()] = \rho(\mathbf{x}; t) \\ &\times \int_{\mathbb{X}} \mathcal{D}\mathbf{X}() P(\mathbf{X}()|\mathbf{X}(t) = \mathbf{x}, I) G[\mathbf{X}()] = \rho(\mathbf{x}; t) \langle G \rangle_{\mathbf{x}, t}. \end{aligned} \quad (19)$$

Two trivial cases of the identity in equation (15) are: (a) the case with $G = G_0$ (a constant functional) which recovers the time-slicing equation, $\langle \delta(\mathbf{X}(t) - \mathbf{x}) \rangle = \rho(\mathbf{x}; t)$, and (b) the case with $G = G(\mathbf{X}(t))$, for which the identity holds immediately because the functional G is constant for all the points where $\mathbf{X}(t) = \mathbf{x}$ and therefore it drops out of the expectation in the left-hand side.

5 Continuity equation from the time-slicing equation

Now we will use equation (12) to derive the continuity equation, showing its fundamental role in a theory of dynamical systems. We take the partial derivative with respect to time on both sides,

$$\partial_t \rho(\mathbf{x}; t) = \int_{\mathbb{X}} \mathcal{D}\mathbf{X}() \rho[\mathbf{X}()] \partial_t \delta(\mathbf{X}(t) - \mathbf{x}), \quad (20)$$

and use the chain rule on the Dirac delta as

$$\partial_t \delta(\mathbf{X}(t) - \mathbf{x}) = \frac{\partial}{\partial \mathbf{X}(t)} \delta(\mathbf{X}(t) - \mathbf{x}) \dot{\mathbf{X}}(t). \quad (21)$$

Noting that the gradient of any function $\phi(\mathbf{X}(t) - \mathbf{x})$ with respect to $\mathbf{X}(t)$ can be expressed as the negative gradient with respect to \mathbf{x} , we obtain

$$\partial_t \rho(\mathbf{x}; t) = -\frac{\partial}{\partial \mathbf{x}} \int_{\mathbb{X}} \mathcal{D}\mathbf{X}() \rho[\mathbf{X}()] \delta(\mathbf{X}(t) - \mathbf{x}) \dot{\mathbf{X}}(t), \quad (22)$$

which we evaluate using the general time-slicing equation, equation (15), leading to the continuity equation (Eq. (1))

$$\partial_t \rho(\mathbf{x}; t) + \frac{\partial}{\partial \mathbf{x}} \left(\rho(\mathbf{x}; t) \langle \dot{\mathbf{X}}(t) \rangle_{\mathbf{x}, t} \right) = 0 \quad (23)$$

with flow velocity

$$\mathbf{v}(\mathbf{x}, t) = \langle \dot{\mathbf{X}}(t) \rangle_{\mathbf{x}, t}. \quad (24)$$

This reveals that the flow velocity is a conditional expectation over paths (in fact, over the *sliced* probability distribution of paths). Because of that, it is a functional of $\rho[\mathbf{X}()]$ and this gives the possibility of different particular forms of the continuity equation for different systems, depending on the kind of paths it explores; for instance, the form of \mathbf{v} could depend on the physical, macroscopic constraints it is subjected to.

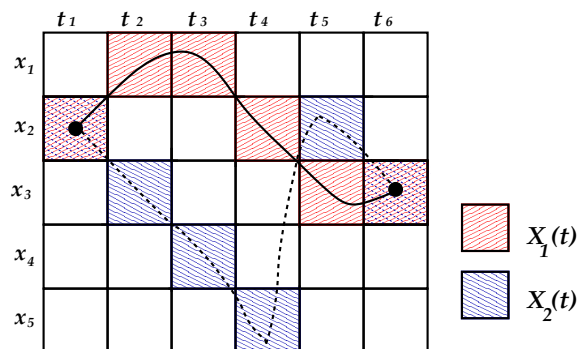


Fig. 2. Discretized space and time with $N = 5$ and $M = 6$. Two paths are shown, the solid line represents a smooth path while the path depicted with the dotted line is discontinuous.

6 Time-slicing in a discrete model

In order to explain the time-slicing process, let us consider a simple model where we have discretized both space and time. In this case, $t \rightarrow t_i = i\Delta t$ and $x \rightarrow x_j = j\Delta x$, with $i = 1, \dots, M$ and $j = 1, \dots, N$. An example of this discretization with $N = 5$ and $M = 6$ is shown in Figure 1.

Here paths can be represented as sequences of positions, $X = \{X_1, X_2, \dots, X_M\}$, where $X_i = X(t_i) \in \{x_1, \dots, x_N\}$. As an example, the solid and dotted paths in Figure 2 are given by $\{x_2, x_1, x_1, x_2, x_3, x_3\}$ and $\{x_2, x_3, x_4, x_5, x_2, x_3\}$, respectively. The probability functional for paths, $\rho[X()]$, reduces to a joint probability of discrete values, $\rho(X) = P(X_1, X_2, \dots, X_M|I)$.

Once we have proposed a model for the path probability $\rho(X)$, derived from some inference method, we can use the time-slicing equation to obtain the probability for the system to be at a discrete position x_i at each time t_j , which we will denote $\rho_{ij} = P(X_j = x_i|I)$. The connection between $\rho(X)$ and ρ_{ij} is of course given by the discrete version of the time-slicing equation,

$$\sum_X \rho(X) \delta(X_j, x_i) = \rho_{ij}, \quad (25)$$

where $\delta(a, b)$ is the Kronecker delta and the sum is performed over all possible discrete paths X . The calculation of the state probability ρ_{ij} then reduces to a simple counting of the number of paths X crossing the point x_i at the time t_j , weighted by the path probability.

7 Concluding remarks

We have shown that the idea of performing inference over a path space \mathbb{X} is a promising foundation for the study of the dynamics of non-equilibrium systems. Its fundamental role is manifest through the use of what we call the time-slicing equations (Eqs. (12) and (15)), which connect a probability functional obtained from a variational principle (such as the Maximum Caliber principle) in path space, with a time-dependent probability density for the

states. The time-slicing equation, which is a definition of a “slice” in time consistent with the laws of probability, is a mathematical identity requiring no underlying physical principle to be valid, and thus applicable to different conceptual systems under different levels of abstraction.

A direct consequence of the time-slicing equation is the continuity equation for the time-dependent probability density (Eq. (1)). This reveals the possibility of obtaining different partial differential equations governing the non-equilibrium statistical mechanics of a variety of systems, simply by “plugging in” the correct form of the probability current $\langle \dot{X}(t) \rangle_{x,t}$, which is dependent on the particular details of each system. We propose that this information must be encoded into the probability functional $\rho[X()]$ using a principle such as Maximum Caliber.

D.G. and S.D. thankfully acknowledge funding from FONDECYT grant 1140514 and PIA-CONICYT ACT-1115. D.G. acknowledges funding from CONICYT Ph.D. fellowship 21140914. DG and SD also wish to thank Gonzalo Gutiérrez for fruitful discussions.

Appendix: Construction of a probability functional using the Maximum Caliber principle

The principle of Maximum Caliber, suggested by Jaynes in 1980, postulates that the most unbiased probability distribution of paths is the one that maximizes their Shannon entropy. This entropy of paths is sometimes called the “caliber” of the system, and is given by the path integral

$$\mathcal{S} = - \int \mathcal{D}\mathbf{X}() \rho[\mathbf{X}()] \ln \frac{\rho[\mathbf{X}()]}{\Pi[\mathbf{X}()]}, \quad (\text{A.1})$$

where $\Pi[\mathbf{X}()]$ is an invariant measure of paths, which usually is taken as a constant. Under the macroscopic constraint on the expectation of an instantaneous function f ,

$$\langle f(\mathbf{X}(t), \dot{\mathbf{X}}(t), t) \rangle = F(t), \quad (\text{A.2})$$

for each instant t in an interval $[0, T]$, the normalized probability that maximizes \mathcal{S} is

$$P(\mathbf{X}()|\lambda()) = \frac{1}{Z[\lambda()]} \exp \left(- \int_0^T dt \lambda(t) f(\mathbf{X}(t), \dot{\mathbf{X}}(t), t) \right) \quad (\text{A.3})$$

where $Z[\lambda()]$ is a *partition functional* imposing normalization, and given by

$$Z[\lambda()] = \int \mathcal{D}\mathbf{X}() \exp \left(- \int_0^T dt \lambda(t) f(\mathbf{X}(t), \dot{\mathbf{X}}(t), t) \right). \quad (\text{A.4})$$

The Maximum Caliber path probabilities have interesting properties. For instance, renaming $\mathcal{L}(x, \dot{x}, t) = \lambda(t) f(x, \dot{x}, t)$ we see that the exponent in equation (A.3) has the form the action of a classical system

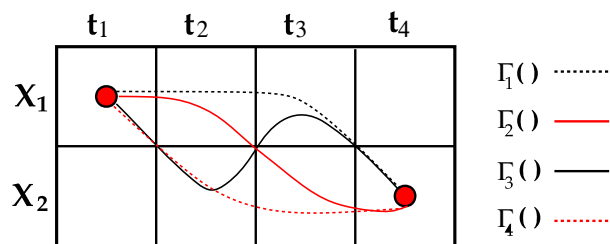


Fig. A.1. Discretized model of 2×4 with the 4 possible paths following the boundary conditions.

Table A.1. The four possible paths joining x_1 in t_1 and x_2 in t_4 for a 2×4 discrete model, together with their action and probability values.

Path	Action	Probability
$\Gamma_1 = \{x_1, x_1, x_1, x_2\}$	-2	0.1050
$\Gamma_2 = \{x_1, x_1, x_2, x_2\}$	0	0.0142
$\Gamma_3 = \{x_1, x_2, x_1, x_2\}$	-4	0.7758
$\Gamma_4 = \{x_1, x_2, x_2, x_2\}$	-2	0.1050

with Lagrangian \mathcal{L} [12,13]. Thus it is useful to rewrite it as:

$$P[\mathbf{X}()|\lambda()] = \frac{1}{Z[\lambda()]} \exp(-A[\mathbf{X}()]) \quad (\text{A.5})$$

with

$$A[\mathbf{X}()] = \int_0^T dt \mathcal{L}(\mathbf{X}(t), \dot{\mathbf{X}}(t); t). \quad (\text{A.6})$$

Equations (A.3) and (A.4) give us a tool for making inferences about paths $\mathbf{X}()$ under known information in the form of time-dependent expectation values.

In order to see how the Maximum Caliber principle and the time-slicing equation work together, consider a discretized model with 2 possible positions x_1 and x_2 , and 4 possible times t_1, t_2, t_3 and t_4 . There are 4 allowed paths from x_1 in t_1 to x_2 in t_4 , described by the vectors $X = (x_1, X_2, X_3, x_2)$ with $X_2, X_3 \in \{x_1, x_2\}$. All these paths are shown in Figure A.1.

Let us assign a simple Lagrangian \mathcal{L} to this discretized space, such as $\mathcal{L}(x_k, t_j) = 2\delta_{kj} - 1$. Here the Lagrangian is independent of velocity, just for clarity. Then, the action for each path can be calculated exactly; these values are shown in Table A.1.

In the same manner, the partition function can also be computed exactly using equation (A.4),

$$Z = \exp(2) + \exp(0) + \exp(4) + \exp(2) \approx 70.37626.$$

With this normalization, the probability for each path is shown in Table A.1. It is also possible to employ the time-slicing equation to obtain the time-dependent probability of visiting each state x_i at the instant t_j , ρ_{ij} , we just

Table A.2. Probability matrix elements for the positions x_1 to x_4 at each instant t_1 and t_2 .

Probability	Value
ρ_{11}	0.1192
ρ_{21}	0.8808
ρ_{12}	0.8808
ρ_{22}	0.1192

compute the sum of the probabilities of each path $\rho(\Gamma)$ passing through a position x_j at time t_i , where $\rho(\Gamma)$ is given by equation (A.3). For example,

$$\rho_{11} = \sum_{\Gamma} \rho(\Gamma) \delta(X_1, x_1) = \rho(\Gamma_1) + \rho(\Gamma_2) \approx 0.1192. \quad (\text{A.7})$$

The full probability matrix ρ_{ij} are given in Table A.2.

We can check that the instantaneous probability is properly normalized for each time, i.e., that

$$\sum_{j=1}^2 \rho_{ji} = 1 \quad \forall i. \quad (\text{A.8})$$

It is important to notice that most probable path (which corresponds to the path with minimum action), Γ_3 , is the path that joins the most probable points for each time.

References

1. D. Collin, F. Ritort, C. Jarzynski, S.B. Smith, I. Tinoco, C. Bustamante, *Nature* **437**, 231 (2005)
2. E.H. Trepagnier, C. Jarzynski, F. Ritort, G.E. Crooks, C.J. Bustamante, J. Liphardt, *Proc. Natl. Acad. Sci.* **101**, 15038 (2004)
3. J. Cardy, G. Falkovich, K. Gawedzki, S. Nazarenko, *Nonequilibrium statistical mechanics and turbulence* (London Mathematical Society, 2009)
4. E.T. Jaynes, *Phys. Rev.* **106**, 620 (1957)
5. S.R. de Groot, P. Mazur, *Nonequilibrium thermodynamics* (Dover Publications Inc., New York, 1984)
6. R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, 2001)
7. E.T. Jaynes, *Ann. Rev. Phys. Chem.* **31**, 579 (1980)
8. S. Pressé, K. Ghosh, J. Lee, K.A. Dill, *Rev. Mod. Phys.* **85**, 1115 (2013)
9. D. González, S. Davis, *J. Phys.: Conf. Ser.* **720**, 012006 (2016)
10. D.J. Evans, G.P. Morriss, *Statistical Mechanics of Nonequilibrium Liquids* (ANU E press, Canberra, 2007)
11. H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, 1996)
12. D. González, S. Davis, G. Gutiérrez, *Found. Phys.* **44**, 923 (2014)
13. S. Davis, D. González, *J. Phys. A* **48**, 425003 (2015)

Jarzynski equality in the context of maximum path entropy

Diego González, and Sergio Davis

Citation: [AIP Conference Proceedings](#) **1853**, 080003 (2017);

View online: <https://doi.org/10.1063/1.4985368>

View Table of Contents: <http://aip.scitation.org/toc/apc/1853/1>

Published by the [American Institute of Physics](#)

Articles you may be interested in

[Maximum entropy PDF projection: A review](#)

[AIP Conference Proceedings](#) **1853**, 070001 (2017); 10.1063/1.4985362

[A method for density estimation based on expectation identities](#)

[AIP Conference Proceedings](#) **1853**, 110001 (2017); 10.1063/1.4985376

[Ergodic statistical models: Entropic dynamics and chaos](#)

[AIP Conference Proceedings](#) **1853**, 100001 (2017); 10.1063/1.4985374

[Deriving proper uniform priors for regression coefficients, part II](#)

[AIP Conference Proceedings](#) **1853**, 050005 (2017); 10.1063/1.4985358

[The classical limit of entropic quantum dynamics](#)

[AIP Conference Proceedings](#) **1853**, 090001 (2017); 10.1063/1.4985370

[Elements of the cognitive universe](#)

[AIP Conference Proceedings](#) **1853**, 040002 (2017); 10.1063/1.4985353

Jarzynski Equality In The Context Of Maximum Path Entropy

Diego González^{1,2,a)} and Sergio Davis¹

¹*Comisión Chilena de Energía Nuclear, Casilla 188-D, Santiago, Chile.*

²*Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile*

a)Corresponding author: dgonzalez@gnm.cl

Abstract. In the global framework of finding an axiomatic derivation of nonequilibrium Statistical Mechanics from fundamental principles, such as the maximum path entropy – also known as Maximum Caliber principle – , this work proposes an alternative derivation of the well-known Jarzynski equality, a nonequilibrium identity of great importance today due to its applications to irreversible processes: biological systems (protein folding), mechanical systems, among others. This equality relates the free energy differences between two equilibrium thermodynamic states with the work performed when going between those states, through an average over a path ensemble.

In this work the analysis of Jarzynski's equality will be performed using the formalism of inference over path space. This derivation highlights the wide generality of Jarzynski's original result, which could even be used in non-thermodynamical settings such as social systems, financial and ecological systems.

INTRODUCTION

Following Jaynes' idea about extending the Maximum Entropy principle [1] to time-dependent systems via the Maximum Caliber Principle [2, 3, 4], it is possible to recover a large part of the structure of non-equilibrium statistical mechanics. In particular we can recover all the PDEs that are particular cases of the continuity equation [5], such as the Liouville equation [6], the Fokker-Planck equation, among others. The resulting structure consists of general nonequilibrium relations known as fluctuation theorems.

In the past two decades several non-equilibrium relations have appeared that describe some properties of processes connecting two equilibrium states. Among them, arguably the most renowned is the Jarzynski equality (JE) [7]. This equality goes beyond the statement of the second law of thermodynamics as

$$W \geq \Delta F, \quad (1)$$

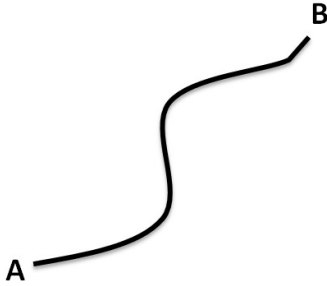
where ΔF is the difference of Helmholtz free energy between equilibrium states and W is the macroscopic work needed to move the system from one state to the other.

SETTING OF THE PROBLEM

In order to state the Jarzynski equality we will first set up our notation. A microscopic state is denoted by $\Gamma \in \mathbb{R}^N$, whereas a microscopic path $\Gamma() \in \mathbb{X}$ is a continuous function of time, such that $\Gamma(t)$ is the microscopic state at time t . The function itself will be denoted by $\Gamma()$ (without its argument) and a functional G of $\Gamma()$ will be written as $G[\Gamma()]$. An expectation of a functional G over a path ensemble characterized by \mathcal{I} is given by the path integral

$$\left\langle G[\Gamma()] \right\rangle_{\mathcal{I}} = \int_{\mathbb{X}} \mathcal{D}\Gamma() P[\Gamma()|\mathcal{I}] G[\Gamma()], \quad (2)$$

where $P[\Gamma()|\mathcal{I}]$ is the path probability functional in the state of knowledge \mathcal{I} . The integration $\int_{\mathbb{X}} \mathcal{D}\Gamma()$ is performed over all paths in \mathbb{X} .

$$\begin{aligned}
& \langle H_A(\Gamma_A) \rangle_{I,\beta} = U_A \\
& A: (T, U_A, P_A, V_A, \dots) \\
& \langle H_B(\Gamma_B) \rangle_{I,\beta} = U_B \\
& B: (T, U_B, P_B, V_B, \dots)
\end{aligned}$$


$$\langle e^{-\beta W[\Gamma]} \rangle_{I,\beta} = e^{-\beta \Delta F}$$

FIGURE 1. Schematic representation of the setup for the Jarzynski equality.

Let us consider a system in a thermodynamic equilibrium state $A(T, U_A, P_A, V_A, \dots)$ which is described by the Hamiltonian $H_A(\Gamma_A)$, where Γ_A is a microstate in the subset A of all microstates. From this point (corresponding to time $t = 0$) the system evolves into a new thermodynamical equilibrium state $B(T, U_B, P_B, V_B, \dots)$ at $t = t_F$ described by the Hamiltonian $H_B(\Gamma_B)$, at the same temperature T (but of course with different thermodynamical properties and different microstates). The non-equilibrium process associated with this evolution from $t = 0$ to $t = t_F$ usually involves random forces and also external (e.g. mechanical) forces, as in the case of stretching of a molecule [8]. Note that we are describing the thermodynamic states A and B of the system by (in general) different Hamiltonian functions, and this allows the case when one “system” (let us say, the ideal gas) turns into another (an interacting gas) by switching on the interaction as time progresses. In this way, we can measure ΔF between different systems via computer simulation of the Jarzynski equality, although they really are different thermodynamic states of a more general system that includes both Hamiltonians as special cases. The Hamiltonian for such a general system could be written, for instance, as

$$H(\Gamma_A, \Gamma_B; t) = (1 - \lambda(t))H_A(\Gamma_A) + \lambda(t)H_B(\Gamma_B), \quad (3)$$

with $\lambda(0) = 0$ and $\lambda(t_F) = 1$.

Jarzynski’s equality (JE) asserts that the difference in Helmholtz free energy, $\Delta F = F_B - F_A$ between the macroscopic states A and B , is related to the expectation over paths of an exponential function of the work required to go from A to B along the microscopic path $\Gamma()$,

$$\left\langle \exp(-\beta W[\Gamma]) \right\rangle_{\beta, I} = \exp(-\beta \Delta F). \quad (4)$$

Here $W[\Gamma()]$ is the microscopic work needed to transit from a particular microstate in A to the corresponding microstate in B along the path $\Gamma()$, and the expectation $\langle \cdot \rangle_{\beta, I}$ is taken over the path ensemble at constant inverse temperature $\beta = 1/k_B T$ such that it also agrees with specified boundary conditions for the paths.

This equality is remarkable in that it has been extensively verified in experiments on individual proteins [9, 10, 8] by performing repeated cycles of stretching at different rates in order to obtain the free energy difference between the folded and unfolded thermodynamic states. Jarzynski equality was originally derived for the canonical ensemble at constant volume, but equivalent identities were later found for the microcanonical [11] and isothermic-isobaric [9] ensembles. In this work, for reasons of simplicity, we will focus on the original, constant-volume derivation.

PROOF OF THE JARZYNSKI EQUALITY

We start from the definition of the expectation on the left-hand side of Eq. 4. This is given by

$$\left\langle \exp(-\beta W[\Gamma]) \right\rangle_{\beta, I} = \int_{\mathcal{X}} D\Gamma() P[\Gamma()|\beta, I] \exp(-\beta W[\Gamma]) \quad (5)$$

Now we wish to include into the probability distribution $P[\Gamma()|\beta, \mathcal{I}]$ the assumption that the states A and B are in thermodynamical equilibrium at the start and at the end of the non-equilibrium process. In fact, the system has to be found in a microstate Γ_A at $t = 0$ belonging to the subset $A \subset \mathbb{R}^N$, and similarly, at time $t = t_F$ it must be found in a microstate $\Gamma_B \in B$ such that $B \subset \mathbb{R}^N$. The subsets A and B contain all the microstates compatible with the external conditions $(T, U_A, P_A, V_A, \dots)$ and $(T, U_B, P_B, V_B, \dots)$, respectively.

In order to incorporate these boundary conditions, we use the marginalization rule of probability [12] as follows. First, let us write

$$P[\Gamma()|\beta, \mathcal{I}] = \int_A \int_B d\Gamma_A d\Gamma_B P(\Gamma(), \Gamma_A, \Gamma_B|\beta, \mathcal{I}), \quad (6)$$

where $\Gamma_A = \Gamma(0) \in A$ and $\Gamma_B = \Gamma(t_F) \in B$ are the microscopic boundary conditions for the path $\Gamma()$. Separating the probability according to the product rule, we obtain

$$P[\Gamma()|\beta, \mathcal{I}] = \int_A \int_B d\Gamma_A d\Gamma_B P(\Gamma_A, \Gamma_B|\beta, \mathcal{I}) P[\Gamma()|\Gamma_A, \Gamma_B, \beta, \mathcal{I}]. \quad (7)$$

It follows that the path expectation is of the form

$$\begin{aligned} \left\langle \exp(-\beta W[\Gamma()]) \right\rangle_{\beta, \mathcal{I}} &= \int_A \int_B \int \mathcal{D}\Gamma() d\Gamma_A d\Gamma_B P(\Gamma_A, \Gamma_B|\beta, \mathcal{I}) P[\Gamma()|\Gamma_A, \Gamma_B, \beta, \mathcal{I}] \exp(-\beta W[\Gamma()]) \\ &= \left\langle \left\langle \exp(-\beta W[\Gamma()]) \right\rangle_{\Gamma_A, \Gamma_B, \beta, \mathcal{I}} \right\rangle_{\beta, \mathcal{I}}. \end{aligned} \quad (8)$$

According to the First Law of Thermodynamics, the work involved in a process is given by $W = \Delta U - Q$. If the system evolves from A to B without any heat exchange (that is, if the interaction with the heat bath is disconnected during the non-equilibrium process), then $W = \Delta U$, where W is the macroscopic work, defined as the expectation $W = \langle W[\Gamma()] \rangle$ of the microscopic work $W[\Gamma()] = H_B(\Gamma_B) - H_A(\Gamma_A)$ with $U_A = \langle H_A(\Gamma_A) \rangle_{\beta, \mathcal{I}}$, $U_B = \langle H_B(\Gamma_B) \rangle_{\beta, \mathcal{I}}$.

We can write Equation 8 as

$$\left\langle \exp(-\beta W[\Gamma()]) \right\rangle_{\beta, \mathcal{I}} = \left\langle \left\langle \exp(-\beta[H_B(\Gamma_B) - H_A(\Gamma_A)]) \right\rangle_{\Gamma_A, \Gamma_B, \beta, \mathcal{I}} \right\rangle_{\beta, \mathcal{I}}, \quad (9)$$

where the inner expectation is immediately solved as the boundary conditions Γ_A and Γ_B are given. Therefore it can be written as

$$\left\langle \exp(-\beta W[\Gamma()]) \right\rangle_{\beta, \mathcal{I}} = \left\langle \exp(-\beta[H_B(\Gamma_B) - H_A(\Gamma_A)]) \right\rangle_{\beta, \mathcal{I}}. \quad (10)$$

This means the right-hand side is independent of the path *ensemble* $P[\Gamma()|\Gamma_A, \Gamma_B, \beta, \mathcal{I}]$; it only depends on the joint probability of the boundary conditions $P(\Gamma_A, \Gamma_B|\beta, \mathcal{I})$. Using the product rule of probability to separate it into two factors we get

$$\left\langle \exp(-\beta W[\Gamma()]) \right\rangle_{\beta, \mathcal{I}} = \int_A \int_B d\Gamma_A d\Gamma_B P(\Gamma_A|\beta, \mathcal{I}) P(\Gamma_B|\Gamma_A, \beta, \mathcal{I}) \exp(-\beta H_B(\Gamma_B)) \exp(\beta H_A(\Gamma_A)). \quad (11)$$

As the system started in an equilibrium thermodynamic state A described by the Hamiltonian $H_A(\Gamma_A)$, the probability distribution for the initial microstate Γ_A is, according to the maximum entropy principle, given by the canonical ensemble

$$P(\Gamma_A|\beta) = \frac{\exp(-\beta H_A(\Gamma_A))}{Z_A(\beta)}, \quad (12)$$

where the factor $\exp(-\beta H_A(\Gamma_A))$ cancels out the factor $\exp(\beta H_A(\Gamma_A))$ in the expectand in Equation 11. Finally, we get

$$\left\langle \exp(-\beta W[\Gamma()]) \right\rangle_{\beta, \mathcal{I}} = \frac{1}{Z_A(\beta)} \int_{\Gamma_A} \int_{\Gamma_B} d\Gamma_A d\Gamma_B P(\Gamma_B|\Gamma_A, \beta, \mathcal{I}) \exp(-\beta H_B(\Gamma_B)), \quad (13)$$

which we can rewrite to introduce the canonical distribution at B ,

$$P(\Gamma_B|\beta) = \frac{\exp(-\beta H_B(\Gamma_B))}{Z_B(\beta)}, \quad (14)$$

by multiplying by $Z_B(\beta)/Z_A(\beta)$. This is because the microscopic states belonging to B can also be described using the maximum entropy principle, due to the assumption of thermodynamic equilibrium at t_f . Now the right-hand side is of the form

$$\begin{aligned} \left\langle \exp(-\beta W[\Gamma(\cdot)]) \right\rangle_{\beta, \mathcal{I}} &= \frac{Z_B(\beta)}{Z_A(\beta)} \int_A \int_B d\Gamma_A d\Gamma_B P(\Gamma_B|\Gamma_A, \beta, \mathcal{I}) P(\Gamma_B|\beta) \\ &= \frac{Z_B(\beta)}{Z_A(\beta)} \left\langle \int_A d\Gamma_A P(\Gamma_B|\Gamma_A, \beta, \mathcal{I}) \right\rangle_{\beta}. \end{aligned} \quad (15)$$

Up to this point we have not assumed the form of the dynamical evolution of the microstates from Γ_A to Γ_B . If the system follows a differential equation

$$\frac{d\Gamma(t)}{dt} = \dot{\Gamma}(t) = G(\Gamma(t); \theta), \quad (16)$$

where θ are arbitrary parameters, then the probability of the microstate Γ_B given the initial microstate Γ_A is given by a Dirac delta,

$$P(\Gamma_B|\Gamma_A, \beta, \mathcal{I}) = P(\Gamma_B|\Gamma_A, \mathcal{I}) = \delta(\Gamma_B - \Gamma^*), \quad (17)$$

where Γ^* is the final microstate of the evolution with initial condition Γ_A . According to Equation 16, which is compatible with both deterministic and stochastic dynamics (depending on the role of θ), the final state Γ^* can be written as

$$\Gamma^* = \Gamma_A + \int_0^{t_f} dt G(\Gamma(t); \theta). \quad (18)$$

Replacing $P(\Gamma_B|\Gamma_A, \mathcal{I})$ in the right-hand side of Equation 15, we have

$$\left\langle \exp(-\beta W[\Gamma(\cdot)]) \right\rangle_{\beta, \mathcal{I}} = \frac{Z_B(\beta)}{Z_A(\beta)} \left\langle \int_A d\Gamma_A \delta(\Gamma_B - \Gamma_A - \int_0^{t_f} dt G(\Gamma(t); \theta)) \right\rangle_{\beta} = \frac{Z_B(\beta)}{Z_A(\beta)}, \quad (19)$$

because we can integrate the delta function with respect to Γ_A , regardless of the value of θ . Using the definition of free energy [13], $F = -\beta^{-1} \ln Z$, we see that the ratio Z_B/Z_A is equal to $\exp(-\beta \Delta F)$ and finally, we obtain

$$\left\langle \exp(-\beta W[\Gamma(\cdot)]) \right\rangle_{\beta, \mathcal{I}} = \exp(-\beta \Delta F). \quad (20)$$

This is the Jarzynski equality, and when the Jensen inequality, $\langle \exp(g) \rangle \geq \exp(\langle g \rangle)$ for any g , is applied to it, we recover the well-known inequality for macroscopic work,

$$\Delta F \leq \left\langle W[\Gamma] \right\rangle_{\beta}. \quad (21)$$

JARZYNSKI EQUALITY AND MODEL COMPARISON

As we have seen, the JE is just a consequence of two ‘‘equilibrium states’’ (MaxEnt distributions) connected via a nonequilibrium process that follows a dynamics of the form given in Equation 16. We can use this formalism, in principle, to study the problem of Bayesian model comparison, through the computation of the Bayes factor (BF) [14]. The BF is given by

$$\frac{P(M_B|D, I)}{P(M_A|D, I)} = \frac{P(D|M_B, I) P(M_B|I)}{P(D|M_A, I) P(M_A|I)} \quad (22)$$

and can be reduced to the computation of the ratio

$$R = \frac{P(D|M_B, I)}{P(D|M_A, I)} \quad (23)$$

if we assume the prior ratio $P(M_B|I)/P(M_A|I)$ as known. The probabilities of the data given each model, appearing in R , can be written explicitly as integrals over the parameters of the model, that is

$$\begin{aligned} P(D|M_A, I) &= \int d\theta P(\theta|M_A, I)P(D|\theta, M_A, I), \\ P(D|M_B, I) &= \int d\phi P(\phi|M_B, I)P(D|\phi, M_A, I). \end{aligned} \quad (24)$$

Each integral is equivalent to the computation of the normalization factor (or partition function) in the following way. Let us define the fictitious Hamiltonians

$$\begin{aligned} H_A(\theta) &= -\ln P(\theta|M_A, I) - \ln P(D|\theta, M_A, I), \\ H_B(\phi) &= -\ln P(\phi|M_B, I) - \ln P(D|\phi, M_B, I), \end{aligned} \quad (25)$$

and take $\beta = 1$. Then

$$\begin{aligned} P(D|M_A, I) &= \int d\theta \exp(-\beta H_A(\theta)) = Z_A(\beta), \\ P(D|M_B, I) &= \int d\phi \exp(-\beta H_B(\phi)) = Z_B(\beta), \end{aligned} \quad (26)$$

and we see that the ratio R is equivalent to

$$R = \frac{Z_B(\beta)}{Z_A(\beta)} = \exp(-\beta \Delta F) = \left\langle \exp(-\beta W[\theta(), \phi()]) \right\rangle_\beta, \quad (27)$$

according to the JE, with

$$W[\theta(), \phi()] = \int_0^{t_F} dt \lambda \frac{d}{d\lambda} H(\theta(t), \phi(t); t) \quad (28)$$

where the extended Hamiltonian $H(\theta, \phi; t)$ is given by

$$H(\theta, \phi; t) = (1 - \lambda(t))H_A(\theta) + \lambda(t)H_B(\phi). \quad (29)$$

With this choice of H the “work” functional W simplifies to

$$W[\theta(), \phi()] = \int_0^{t_F} dt \left(\frac{d\lambda(t)}{dt} \right) [H_B(\phi(t)) - H_A(\theta(t))]. \quad (30)$$

We just need to postulate a fictitious non-equilibrium process that joins the models M_A and M_B through some evolution equation compatible with Equation 16. For instance, in discrete time, an evolution map given by

$$\theta_{t+\Delta t} = \theta_t + \Delta t G_1(\theta_t, \phi_t), \quad (31)$$

$$\phi_{t+\Delta t} = \phi_t + \Delta t G_2(\theta_t, \phi_t), \quad (32)$$

is of this type, and this includes the case where G_1 and G_2 are stochastic generators like Markov-Chain Monte Carlo or Gibbs updating.

Conclusions

Because the JE follows from the assumption of initial and final equilibrium states described as MaxEnt distributions, and a non-equilibrium dynamics described in a quite general form, it is possible to extend the validity of the result to the Bayesian inference framework almost unmodified, being plausible to apply the equality in different subjects of Science such as signal analysis, biology, economy, among others. In fact, as recently shown by the authors [15], arbitrary inference problems with binary choices are described by the formalism of thermodynamical first-order phase transitions and have associated with them concepts such as free energy differences ΔF .

It is important to note that the validity of this result is independent of the details of the intermediate process, the only important element being that evolution from A to B (both equilibrium states) given by Equation 16.

Understanding these and other non-equilibrium relationships from the point of view of Bayesian inference, is a key goal to aim for in the larger scheme of formalizing a theory for non-equilibrium Statistical Mechanics, being the JE a first step into this territory. It shows that the formalism for inference over paths, i.e., the use of the Bayesian probability framework in path space, and the maximum entropy principle (including the maximum path entropy principle) seems to be sufficient for the completion of this objective.

ACKNOWLEDGMENTS

DG acknowledges support from CONICYT PhD fellowship 21140914. SD acknowledges support from FONDECYT grant 1140514.

REFERENCES

- [1] E. T. Jaynes, [Physical Review](#) **106**, 620–630 (1957).
- [2] E. T. Jaynes, [Ann. Rev. Phys. Chem.](#) **31**, 579–601 (1980).
- [3] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, [Reviews of Modern Physics](#) **85**, 1115–1141 (2013).
- [4] S. Davis and D. González, [J. Phys. A: Math. Theor.](#) **48**, p. 425003 (2015).
- [5] D. González, D. Díaz, and S. Davis, [Eur. Phys. J. B](#) **89**, p. 214 (2016).
- [6] D. González and S. Davis, [AIP Conf. Proc.](#) **1757**, p. 20003 (2016).
- [7] C. Jarzynski, [Phys. Rev. Lett.](#) **78**, 2690–2693 (1997).
- [8] C. Bustamante, J. Liphardt, and F. Ritort, [Physics Today](#) **58**, p. 43 (2005).
- [9] G. Hummer and A. Szabo, [Proc. Nac. Acad. Sci. USA](#) **98**, p. 3658 (2001).
- [10] J. Liphardt, S. Dumont, S. B. Smith, I. T. Jr., and C. Bustamante, [Science](#) **296**, p. 1832 (2002).
- [11] A. B. Adib, [Phys. Rev. E](#) **71**, p. 056128 (2005).
- [12] D. S. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial* (Oxford University Press, 2006).
- [13] H. Callen, *Thermodynamics and an Introduction to Thermostatistics* (Wiley, 1985).
- [14] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, 2003).
- [15] S. Davis, J. Peralta, Y. Navarrete, D. González, and G. Gutiérrez, [Found. Phys.](#) **46**, 350–359 (2016).

PAPER • OPEN ACCESS

Expectation values of general observables in the Vlasov formalism

To cite this article: Diego González *et al* 2018 *J. Phys.: Conf. Ser.* **1043** 012008

View the [article online](#) for updates and enhancements.

Related content

- [Universal time evolution of a Rydberg lattice gas with perfect blockade](#)
B Olmos, R González-Férez, I Lesanovsky *et al.*
- [THE LINEAR THEORY OF LANDAU DAMPING](#)
V P Maslov and M V Fedoryuk
- [Hamiltonian formalism and path entropy maximization](#)
Sergio Davis and Diego González



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Expectation values of general observables in the Vlasov formalism

Diego González^{1,2}, Abiam Tamburrini¹, Sergio Davis², Jalaj Jain²,
Gonzalo Gutiérrez¹

¹Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile.

²Comisión Chilena de Energía Nuclear, Casilla 188-D, Santiago.

E-mail: dgonzalez@gnm.cl, abiamtamburrini@gmail.com, sdavis@cchen.cl

Abstract. Collisionless plasmas in an arbitrary dynamical state are described by the Vlasov equation, which gives the time evolution of the probability density $\rho(\mathbf{x}, \mathbf{v})$. In this work we introduce a new analytical procedure to generate particular partial differential equations (PDEs) for an arbitrary macroscopic observable $w(\mathbf{x}, \mathbf{v})$ that can be expressed as a function of positions and velocities, without solving the time evolution of the probability itself. This technique, which we will call the “Ehrenfest procedure” (as it produces relations that are analogous to Ehrenfest’s theorem in Quantum Mechanics), is based on the iterative application of the fluctuation-dissipation theorem and the recently proposed conjugate variables theorem (CVT) in order to eliminate the explicit dependence on ρ . In particular, we show how this formalism is applied to the Vlasov equation for collisionless plasmas, and derive a general evolution equation for the fluctuations of any macroscopic property w in this kind of plasma.

1. Introduction

The standard treatment of nonequilibrium systems in Statistical Mechanics involves the solution of a partial differential equation (PDE) for the time-dependent probability density of microstates of the system. This is a formidable problem for systems with large number of degrees of freedom, and most of the time it is approximated using numerical methods on a discretized time and space grid. After one solves for the probability density, it is possible to take expectations to compute the nonequilibrium values of macroscopic quantities.

An interesting problem in plasma physics approached from the point of view of Statistical Mechanics is to obtain properties of collisionless plasmas [1] which are described by the Vlasov equation. This equation is a particular case of the Liouville theorem of conservation of the volume of phase space,

$$\frac{\partial \rho}{\partial t} + \{\rho, H\} = 0. \quad (1)$$

where $\rho(\mathbf{x}, t)$ is the time-dependent probability density of microstates \mathbf{x} . In the particular case of collisionless plasma, the Hamiltonian involves kinetic terms and the interaction of charged particles with electromagnetic fields, and Liouville’s theorem becomes the Vlasov equation

$$\frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla_r \rho + \frac{q}{m} (\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{B}) \cdot \nabla_v \rho = 0. \quad (2)$$



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

This equation is used to model collisionless plasmas out of equilibrium, where the force on each particle is given by the Lorentz force, with time-dependent electric and magnetic fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ respectively, and where those \mathbf{E} and \mathbf{B} are consistent with Maxwell's equations.

There is no general analytical solution to this problem for arbitrary electromagnetic fields (or arbitrary charge and current densities) and the numerical methods are extremely expensive in terms of computational cost, as they require the discretization of space and time with a grid size as small as the desired resolution, which depends of the highest frequency of modes of oscillation in the plasma, typically the frequency associated with the electrons [2].

In this work we show a general formalism that allows the construction of the particular PDE associated to an arbitrary, time-dependent macroscopic property $w(\mathbf{r}, \mathbf{v}, t)$ from the original PDE describing the time evolution of the probability density ρ . The development of this procedure, which we will call the ‘‘Ehrenfest procedure’’ (in reference to the Ehrenfest theorem in Quantum Mechanics), requires the application of two mathematical identities of probability theory, namely the fluctuation-dissipation theorem and the conjugate variable theorem, which we will introduce in the next section.

2. Mathematical Tools

As we have explained in the previous section, the objective of this procedure is to move from the problem of solving a PDE for the probability density function to the (usually much simpler) problem of solving a PDE for one of its macroscopic properties. In order to achieve this, we will employ two theorems of probability theory for continuous variables, that are of general validity. These are the Fluctuation-Dissipation Theorem (FDT) and the recently introduced Conjugate Variable Theorem (CVT), which will be used to eliminate the explicit dependence of the probability density function from the PDEs to be solved. In the following we will review the mathematical statement of these theorems.

2.1. Fluctuation-Dissipation Theorem

According to the definition of expectation, any time-dependent macroscopic property W_t can be written as

$$W_t = \langle w(x, t) \rangle_t = \int dx \rho(x, t) w(x, t), \quad (3)$$

where $w(x, t)$ is a macroscopic property, $\rho(x, t)$ is the probability density function and $x = (x^1, x^2, \dots, x^n)$ is an instantaneous state of the system. The subindex t in the expectation $\langle \cdot \rangle_t$ and W_t indicates the temporal dependence of those quantities. Taking partial derivatives of the definition of expectation with respect to time we get

$$\frac{\partial}{\partial t} \langle w(x, t) \rangle_t = \int dx [w(x, t) \frac{\partial}{\partial t} \rho(x, t) + \rho(x, t) \frac{\partial}{\partial t} w(x, t)],$$

and rewriting $\frac{\partial \rho(x, t)}{\partial t}$ as $\rho(x, t) \frac{\partial}{\partial t} \ln \rho(x, t)$ is possible to rewrite each term as an expectation. This leads to the so-called fluctuation-dissipation theorem,

$$\frac{\partial}{\partial t} \langle w(x, t) \rangle_t - \left\langle \frac{\partial}{\partial t} w(x, t) \right\rangle_t = \left\langle w(x, t) \frac{\partial}{\partial t} \ln \rho(x, t) \right\rangle_t. \quad (4)$$

This identity is valid for any expectation that depends parametrically on variables, as in this case the time t .

2.2. Conjugate Variable Theorem

In a similar way to the FDT, we can construct an identity involving spatial derivatives of an arbitrary observable $w(\mathbf{x}, t)$ and of the probability density $\rho(\mathbf{x}, t)$, the recently introduced conjugate variables theorem [3]. Its N -dimensional version reads

$$\left\langle \nabla \cdot \mathbf{w}(\mathbf{x}, t) \right\rangle = - \left\langle \mathbf{w}(\mathbf{x}, t) \cdot \nabla \ln \rho(\mathbf{x}, t) \right\rangle, \quad (5)$$

where $\nabla = \frac{\partial}{\partial x_i} \hat{x}_i$ and $\mathbf{w}(\mathbf{x}, t)$ is a macroscopic vector property dependent of the state of the system and time.

Both theorems (FDT and CVT) are expectation identities which are valid for states arbitrarily far from equilibrium, and they connect arbitrary expectations involving $w(\mathbf{x}, t)$ and their derivatives, with expectations that include derivatives of $\ln \rho(\mathbf{x}, t)$. In the following we will use this interesting feature of these theorems in order to produce differential equations for the expectation of $w(\mathbf{r}, \mathbf{v}, t)$, where we consider the state space as $\mathbf{x} = \mathbf{r} \otimes \mathbf{v}$.

3. Ehrenfest Procedure

Now we are able to present the analytical procedure that we will call the ‘‘Ehrenfest procedure’’, first employed in Ref. [4]. The idea is to obtain a PDE for the time-dependent expectation $\langle w \rangle_t$ of a macroscopic quantity $w(\mathbf{x}, t)$ from the original PDE for the probability density $\rho(\mathbf{x}, t)$.

In order to fix ideas, we will demonstrate the procedure on the Vlasov equation, which is a PDE for $\rho(\mathbf{r}, \mathbf{v}, t)$ of the form

$$\frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla_r \rho + \frac{\mathbf{F}}{m} \cdot \nabla_v \rho = 0, \quad (6)$$

where $\rho = \rho(\mathbf{x}, t)$ with $\mathbf{x} = \mathbf{r} \otimes \mathbf{v}$ and \mathbf{F} is the Lorentz force, given by $\mathbf{F} = q(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{B})$.

The procedure is divided in three stages. First, we rewrite the PDE for ρ as a PDE for $\ln \rho$ by using the relations

$$\frac{\partial}{\partial t} \rho = \rho \frac{\partial}{\partial t} \ln \rho, \quad (7)$$

$$\nabla_x \rho = \rho \nabla_x \ln \rho. \quad (8)$$

Once we have put the original PDE in logarithmic form,

$$\rho \frac{\partial}{\partial t} \ln \rho + \rho \mathbf{v} \cdot \nabla_x \ln \rho + \rho \frac{\mathbf{F}}{m} \cdot \nabla_v \ln \rho = 0. \quad (9)$$

we introduce expectations by applying the operator $\int d\mathbf{r} d\mathbf{v} w \cdot$ (with the test function $w(\mathbf{x}, \mathbf{v}, t)$ defined over the state space) and using the definition of expectation,

$$\left\langle w(\mathbf{r}, \mathbf{v}, t) \right\rangle_t = \int d\mathbf{r} d\mathbf{v} w(\mathbf{r}, \mathbf{v}, t) \rho(\mathbf{r}, \mathbf{v}, t), \quad (10)$$

we obtain

$$\int d\mathbf{r} d\mathbf{v} \rho \left[w \frac{\partial}{\partial t} \ln \rho + w \mathbf{v} \cdot \nabla_x \ln \rho + w \frac{\mathbf{F}}{m} \cdot \nabla_v \ln \rho \right] = 0,$$

that is,

$$\left\langle w \frac{\partial}{\partial t} \ln \rho + w \mathbf{v} \cdot \nabla_x \ln \rho + w \frac{\mathbf{F}}{m} \cdot \nabla_v \ln \rho \right\rangle = 0. \quad (11)$$

Finally, we use CVT and FDT to eliminate $\ln \rho$ from each expectation, in order to reduce them to expectations involving derivatives (spatial or temporal) of w . More precisely, we eliminate $\partial_t \ln \rho$ using the FDT, and $\nabla \ln \rho$ using CVT. In this case, we obtain

$$\frac{\partial}{\partial t} \langle w \rangle_t = \left\langle \frac{\partial w}{\partial t} + \partial_{\mathbf{x}}(w\mathbf{v}) + \partial_{\mathbf{v}}\left(w\frac{\mathbf{F}}{m}\right) \right\rangle,$$

which can be expanded as

$$\frac{\partial}{\partial t} \langle w \rangle_t = \left\langle \frac{\partial w}{\partial t} \right\rangle_t + \left\langle \mathbf{v} \partial_{\mathbf{x}} w + w \partial_{\mathbf{x}} \mathbf{v} \right\rangle_{\mathbf{x}} + \left\langle \frac{\mathbf{F}}{m} \partial_{\mathbf{v}} w + w \partial_{\mathbf{v}} \frac{\mathbf{F}}{m} \right\rangle_{\mathbf{v}}. \quad (12)$$

Here we notice that we can simplify this result even further. As \mathbf{v} is not dependent of \mathbf{x} ($\nabla_{\mathbf{x}} \mathbf{v} = 0$) and for the case of the Lorentz force $\nabla_{\mathbf{v}} \cdot \mathbf{F} = 0$, we arrive at our main result, which is a PDE for an arbitrary property $\langle w \rangle_t$,

$$\frac{\partial}{\partial t} \langle w \rangle_t = \left\langle \frac{\partial w}{\partial t} \right\rangle_t + \left\langle \mathbf{v} \cdot \partial_{\mathbf{x}} w \right\rangle_t + \left\langle \frac{\mathbf{F}}{m} \cdot \partial_{\mathbf{v}} w \right\rangle_t. \quad (13)$$

This resembles a classical analog of Ehrenfest's theorem in Quantum Mechanics [5],

$$\frac{d}{dt} \langle \hat{\Omega} \rangle = \left\langle \frac{\partial \hat{\Omega}}{\partial t} \right\rangle - \frac{i}{\hbar} \langle [\hat{\Omega}, \hat{H}] \rangle. \quad (14)$$

where the Poisson bracket has to be used instead of the quantum-mechanical commutator.

4. Results

Using the main result of this work, namely Eq. 13, it is possible to obtain the PDE for the fluctuations of an arbitrary quantity $w(\mathbf{r}, \mathbf{v}, t)$. We will illustrate this by considering the velocity fluctuations $\langle (\delta v)^2 \rangle_t$, and for this we will use $w = v$ and $w = v^2$ separately in Eq. 13. First, we obtain the time evolution of \mathbf{v} as

$$\frac{\partial}{\partial t} \langle \mathbf{v} \rangle_t = \left\langle \frac{\partial \mathbf{v}}{\partial t} \right\rangle_t + \left\langle \mathbf{v} \partial_{\mathbf{x}} \mathbf{v} \right\rangle_t + \left\langle \frac{\mathbf{F}}{m} \partial_{\mathbf{v}} \mathbf{v} \right\rangle_t, \quad (15)$$

but $\partial_t \mathbf{v} = 0$ and $\partial_{\mathbf{x}} \mathbf{v} = 0$, so we simplify it as

$$\frac{\partial}{\partial t} \langle \mathbf{v} \rangle_t = \left\langle \frac{\mathbf{F}}{m} \right\rangle_t. \quad (16)$$

This has a clear meaning: the change in the mean value of the velocity is given by the mean acceleration. Now, we obtain the evolution equation for $w = v^2$,

$$\frac{\partial}{\partial t} \langle v^2 \rangle_t = \frac{2}{m} \langle \mathbf{F} \cdot \mathbf{v} \rangle_t. \quad (17)$$

We can join Eqs. 16 and 17 by taking

$$\frac{\partial}{\partial t} \langle (\delta v)^2 \rangle_t = \frac{\partial}{\partial t} \langle v^2 \rangle_t - 2 \langle \mathbf{v} \rangle_t \cdot \frac{\partial}{\partial t} \langle \mathbf{v} \rangle_t, \quad (18)$$

and we can write a PDE for $\langle (\delta v)^2 \rangle_t = \langle v^2 \rangle_t - \langle \mathbf{v} \rangle_t^2$ of the form

$$\frac{\partial}{\partial t} \langle (\delta v)^2 \rangle_t = \frac{2}{m} \left\{ \langle q\mathbf{E} \cdot \mathbf{v} \rangle_t - \langle \mathbf{v} \rangle_t \langle q\mathbf{E} \rangle_t - \frac{q}{c} \langle \mathbf{v} \times \mathbf{B} \rangle_t \langle \mathbf{v} \rangle_t \right\}. \quad (19)$$

Here is interesting to note that, for the case where electromagnetic fields \mathbf{E} and \mathbf{B} are spatially homogeneous, the right-hand side vanishes and we have that the size of the velocity fluctuations is kept constant in time.

5. Concluding remarks

We have shown the application of our “Ehrenfest procedure” to the Vlasov equation for collisionless plasmas. Our main result here was a general relation, Eq. 13, which gives the time evolution of any macroscopic property in a Vlasov system. From the general Ehrenfest relation we have derived the partial differential equation for the velocity fluctuations, Eq. 19, which indicates that for a collisionless plasma under spatially homogeneous electromagnetic fields, its initial velocity distribution does not widen with time. We can see that the Ehrenfest procedure is a general tool applicable whenever we have a system that follows a continuity equation for probability. It allows to quickly obtain dynamical evolution equations for particular properties, and thus appears as a powerful addition to the study of dynamical systems and non-equilibrium Statistical Mechanics.

Testing of Eq. 13 and 19 on different configurations of fields remains as future work, to be connected with numerical simulations of Vlasov systems.

Acknowledgements

SD and DG gratefully acknowledge funding from FONDECYT grant 1140514. DG acknowledges funding from CONICYT PhD Fellowship 21140914.

References

- [1] Bellan P M 2006 *Fundamentals of Plasma Physics* (Cambridge University)
- [2] Birdsall C and Langdon A 1991 *Plasma Physics via Computer Simulation*
- [3] Davis S and Gutiérrez G 2012 *Phys. Rev. E* **86** 051136
- [4] Davis S and González D 2015 *J. Phys. A: Math. Theor.* **48** 425003
- [5] Shankar R 1994 *Principles of Quantum Mechanics* (Plenum Press)

Article

Probabilistic Inference for Dynamical Systems

Sergio Davis ^{1,2,*}, Diego González ^{1,3}  and Gonzalo Gutiérrez ³¹ Comisión Chilena de Energía Nuclear, Casilla 188-D Santiago, Chile; dgonzalez@gnm.cl² Departamento de Física, Facultad de Ciencias Exactas, Universidad Andres Bello, Sazié 2212, Piso 7, 8370136 Santiago, Chile³ Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile, Casilla 653 Santiago, Chile; gonzalo@fisica.ciencias.uchile.cl

* Correspondence: sergio.davis@cchen.cl

Received: 30 April 2018; Accepted: 6 September 2018; Published: 12 September 2018



Abstract: A general framework for inference in dynamical systems is described, based on the language of Bayesian probability theory and making use of the maximum entropy principle. Taking the concept of a path as fundamental, the continuity equation and Cauchy's equation for fluid dynamics arise naturally, while the specific information about the system can be included using the maximum caliber (or maximum path entropy) principle.

Keywords: dynamical systems; bayesian inference; fluid equations

1. Introduction

Dynamical system models are widely used to describe complex physical systems (e.g., the weather), as well as social and economic systems (e.g., the stock market). These systems are usually subject to high levels of uncertainty, either in their initial conditions and/or in their interactions with their environment. From the point of view of constructing predictive models, the optimal description of the time-dependent state of such a system given external constraints is a challenge with promising applications in both fundamental and applied science. This is of course an *inference problem* in which we must choose the most likely solution out of the (possibly infinite) alternatives compatible with the given information we have about the system.

Given all this, it seems that a unified framework for performing inference on dynamical systems may open new possibilities in several areas, including non-equilibrium statistical mechanics and thermodynamics, hydrodynamics (including magnetohydrodynamics), and classical mechanics under stochastic forces, among other possible fields of application. Of course, this vision of inference applied to dynamical systems is not new—the clearest exposition of the ideas that we aim to extend here was given by E. T. Jaynes [1,2], followed by several others [3–5].

In this work, we present some elements for a general framework of inference in dynamical systems, written in the language of Bayesian probability. The first is the master equation, which is shown as a direct consequence of the laws of probability. Next, we develop the treatment of inference over paths from which we obtain the continuity equation and Cauchy's equation for fluid dynamics, and discuss their range of applicability. Finally, we close with some concluding remarks.

2. Why Bayesian Inference?

Unlike the standard (“frequentist”) interpretation of probability theory, in which probabilities are frequencies of occurrence of repeatable events, Bayesian probability can be understood as the natural extension of classical logic in the case of *uncertainty* [6,7]. Bayesian probability deals with unknown quantities rather than identical repetitions/copies of an event or system, and is able to include *prior* information when needed.

The conceptual framework of Bayesian probability provides an elegant language to describe dynamical systems under uncertainty. A straightforward advantage of the Bayesian framework is that one does not need to assume an ensemble of “many identical copies of a system”. A single system with uncertain initial conditions and/or forces is sufficient to construct a theory. The probability of finding this particular system in a given state at a given time would not be a frequency, but rather a degree of plausibility conditioned on the known information. In fact, we can lift even the common assumption of “many degrees of freedom”. The motion of a single particle could be used to construct an internally consistent theory, where the equations of motion for time-dependent probability densities are similar to the ones of hydrodynamics. We will describe in detail both features of the Bayesian formulation in the following sections.

A brief overview of Bayesian notation used in this work follows. We will take $P(Q|\mathcal{I})$ as the probability of a particular *proposition* Q being true given knowledge \mathcal{I} . On the other hand, $\langle G \rangle_{\mathcal{I}}$ will denote the expected value of an arbitrary *quantity* G given knowledge \mathcal{I} , and will be given by

$$\langle G \rangle_{\mathcal{I}} = \sum_{\mathbf{u}} P(\mathbf{u}|\mathcal{I})G(\mathbf{u}), \tag{1}$$

where \mathbf{u} represents one of the possible states of the system.

3. Dynamical Evolution of Probabilities

Consider a discrete-time system that can transit between n possible states $\{x_1, \dots, x_n\}$ at different times. If we denote by $X(t)$ the state of the system at time t , we have that the joint probability of being in state a at time t and in state b at time $t + \Delta t$ is given by

$$P(X(t) = a, X(t + \Delta t) = b|\mathcal{I}) = P(X(t + \Delta t) = b|X(t) = a, \mathcal{I}) \cdot P(X(t) = a|\mathcal{I}). \tag{2}$$

By summing over a in Equation (2) and taking $b = x$, we have

$$P(X(t + \Delta t) = x|\mathcal{I}) = \sum_a P(X(t + \Delta t) = x|X(t) = a, \mathcal{I}) \cdot P(X(t) = a|\mathcal{I}), \tag{3}$$

while by summing over b and taking $a = x$, we have

$$P(X(t) = x|\mathcal{I}) = P(X(t) = x|\mathcal{I}) \sum_b P(X(t + \Delta t) = b|X(t) = x, \mathcal{I}). \tag{4}$$

From these two identities (Equations (3) and (4)), we can construct the discrete-time difference,

$$\begin{aligned} \frac{P(X(t + \Delta t) = x|\mathcal{I}) - P(X(t) = x|\mathcal{I})}{\Delta t} &= \sum_{x'} \left[\frac{P(X(t + \Delta t) = x|X(t) = x', \mathcal{I})}{\Delta t} \right] P(X(t) = x'|\mathcal{I}) \\ &\quad - P(X(t) = x|\mathcal{I}) \sum_{x'} \left[\frac{P(X(t + \Delta t) = x'|X(t) = x, \mathcal{I})}{\Delta t} \right]. \end{aligned} \tag{5}$$

This equation is a discrete-time form of the celebrated “master equation” [8–10], involving time-dependent transition probabilities. The case where $P(X(t + \Delta t) = x|X(t) = x', \mathcal{I})$ is independent of t (i.e., a function only of the initial state x' and final state x) is more commonly known as the master equation in the literature. In the continuous-time limit when $\Delta t \rightarrow 0$, we can write this equation as:

$$\frac{\partial \rho}{\partial t} = \sum_{x'} W_t(x' \rightarrow x)\rho(x'; t) - \rho(x; t) \sum_{x'} W_t(x \rightarrow x'), \tag{6}$$

where the instantaneous density ρ is given by $\rho(a; t) := P(X(t) = a|\mathcal{I})$ and we have defined the (continuous-time) *transition rate* W_t as

$$W_t(a \rightarrow b) := \lim_{\Delta t \rightarrow 0} \frac{P(X(t + \Delta t) = b | X(t) = a, \mathcal{I})}{\Delta t}. \tag{7}$$

In this sense, the master equation as written in Equation (5) is a direct consequence of the laws of probability, and its validity is universal whenever we have transitions between states. Please note that there is no requirement for the system to be Markovian: regardless of the form of the joint probability

$$P(X(t + \Delta t) = x, X(t) = x', X(t - \Delta t) = x'', \dots | \mathcal{I}),$$

there is always a *marginal* model

$$P(X(t + \Delta t) = x, X(t) = x' | \mathcal{I}) = \int dx'' \dots P(X(t + \Delta t) = x, X(t) = x', X(t - \Delta t) = x'', \dots | \mathcal{I}) \tag{8}$$

which yields the transition probability in Equation (7) as:

$$P(X(t + \Delta t) = x | X(t) = x', \mathcal{I}) = \frac{P(X(t + \Delta t) = x, X(t) = x' | \mathcal{I})}{\int d\eta P(X(t + \Delta t) = \eta, X(t) = x' | \mathcal{I})}. \tag{9}$$

It is for this probability that Equation (5) holds. In general, the transition rate W_t will most probably be time-dependent, due to the fact that it captures the dependence of the previous history of the system up to t . It follows from this that all probabilities of time-dependent quantities must evolve in time according to Equation (5) (or (6) in the case of continuous time) for some (possibly time-dependent) transition probability (rate). This continuous-time master equation (Equation (6)) is more general than the continuity equation, as it includes the case where some quantities can be created or destroyed during a process. However, time evolution under global and local conservation laws is a fundamental case that can also be readily obtained from the Bayesian formalism, as we will see in the following sections. As is well-known, the continuous-time master equation can be approximated in the limit of infinitesimally small transitions to obtain the Fokker–Planck equation [9,11], but in the next section we start from the existence of continuous paths as a postulate.

4. Fluid Theories in a Bayesian Formulation

We will now consider a dynamical system that follows a path $X(t) \in \mathbb{X}$ in time, where \mathbb{X} denotes the space of all paths consistent with given boundary conditions. The path $X(t)$ is not completely known, and we only have access to partial information denoted by \mathcal{I} .

In this setting, Bayesian theory defines a functional $P[X|\mathcal{I}]$ that is the probability density of the path $X(t)$ being the “true path” under the known information. For any arbitrary *functional* $\mathcal{F}[X]$ of the path, we can then write its expected value as a path integral:

$$\langle \mathcal{F} \rangle_{\mathcal{I}} = \int_{\mathbb{X}} \mathcal{D}X P[X|\mathcal{I}] \mathcal{F}[X]. \tag{10}$$

On the other hand, the expected value of any instantaneous *quantity* $A(x; t)$ is given by

$$\langle A \rangle_{t, \mathcal{I}} = \int_{\mathbb{X}} \mathcal{D}X P[X|\mathcal{I}] A(X(t); t) = \int_V dx \rho(x; t) A(x; t), \tag{11}$$

where $\rho(x; t) := P(X(t) = x | \mathcal{I})$ is the instantaneous probability density at time t .

By using a quantity $A(x'; t) = \delta(x' - x)$, we see that the probability density itself has a path integral representation:

$$\rho(x; t) = \int_{\mathbb{X}} \mathcal{D}X P[X|\mathcal{I}] \delta(X(t) - x). \tag{12}$$

By differentiating Equation (12) with respect to time, we obtain the continuity equation for the instantaneous probability density [12]:

$$\begin{aligned}
 \frac{\partial \rho}{\partial t} &= \int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \frac{\partial}{\partial t} \delta(\mathbf{X}(t) - \mathbf{x}) \\
 &= \int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \left(\nabla \delta(\mathbf{X}(t) - \mathbf{x}) \right) \cdot \dot{\mathbf{X}}(t) \\
 &= -\nabla_{\mathbf{x}} \cdot \left(\int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \delta(\mathbf{X}(t) - \mathbf{x}) \dot{\mathbf{X}}(t) \right) \\
 &= -\nabla_{\mathbf{x}} \cdot \left(\rho \cdot \mathbf{v}(\mathbf{x}; t) \right),
 \end{aligned} \tag{13}$$

where $\mathbf{v}(\mathbf{x}; t)$ is the velocity field that describes the flow of probability, given by

$$\mathbf{v}(\mathbf{x}; t) = \left\langle \dot{\mathbf{X}}(t) \right\rangle_{\mathbf{X}(t)=\mathbf{x}, \mathcal{I}} = \frac{1}{\rho(\mathbf{x}; t)} \left\langle \dot{\mathbf{X}}(t) \delta(\mathbf{X}(t) - \mathbf{x}) \right\rangle_{\mathcal{I}}. \tag{14}$$

This equation describes the global and local conservation of the probability of finding the system in a given state \mathbf{x} at a time t , and is guaranteed to hold for any system moving continuously in time through paths $\mathbf{X}(t) \in \mathbb{X}$. In the same way, it is possible to derive a dynamical equation for the velocity field $\mathbf{v}(\mathbf{x}; t)$ itself, by differentiating it with respect to time. We have:

$$\begin{aligned}
 \frac{\partial}{\partial t} (\rho \cdot v_{\mu}) &= \frac{\partial}{\partial t} \int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \delta(\mathbf{X}(t) - \mathbf{x}) \dot{X}_{\mu}(t) \\
 &= \int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \frac{\partial}{\partial t} \delta(\mathbf{X}(t) - \mathbf{x}) \dot{X}_{\mu}(t) \\
 &= \int_{\mathbb{X}} \mathcal{D}\mathbf{X}P[\mathbf{X}|\mathcal{I}] \left(\delta(\mathbf{X}(t) - \mathbf{x}) \ddot{X}_{\mu}(t) + \partial_{\nu} \delta(\mathbf{X}(t) - \mathbf{x}) \dot{X}_{\mu}(t) \dot{X}_{\nu}(t) \right) \\
 &= \left\langle \delta(\mathbf{X}(t) - \mathbf{x}) \ddot{X}_{\mu}(t) \right\rangle_{\mathcal{I}} - \partial_{\nu} \left\langle \delta(\mathbf{X}(t) - \mathbf{x}) \dot{X}_{\mu}(t) \dot{X}_{\nu}(t) \right\rangle_{\mathcal{I}} \\
 &= \rho a_{\mu}(\mathbf{x}; t) - \partial_{\nu} (\rho [C_{\mu\nu} + v_{\mu}v_{\nu}]),
 \end{aligned} \tag{15}$$

where in the last line we have defined the acceleration field as

$$\mathbf{a}(\mathbf{x}; t) := \left\langle \ddot{\mathbf{X}}(t) \right\rangle_{\mathbf{X}(t)=\mathbf{x}, \mathcal{I}} \tag{16}$$

and the velocity covariance matrix

$$C_{\mu\nu}(\mathbf{x}; t) = \left\langle \dot{X}_{\mu}(t) \dot{X}_{\nu}(t) \right\rangle_{\mathbf{X}(t)=\mathbf{x}, \mathcal{I}} - v_{\mu}(\mathbf{x}; t)v_{\nu}(\mathbf{x}; t). \tag{17}$$

By using the continuity equation (Equation (13)) to rewrite the left-hand side as

$$\begin{aligned}
 \frac{\partial}{\partial t} (\rho \cdot v_{\mu}) &= \frac{\partial \rho}{\partial t} v_{\mu} + \rho \frac{\partial v_{\mu}}{\partial t} \\
 &= -\nabla_{\mathbf{x}} \cdot (\rho \cdot \mathbf{v}) v_{\mu} + \rho \frac{\partial v_{\mu}}{\partial t} \\
 &= -\partial_{\nu} (\rho v_{\nu}) v_{\mu} + \rho \frac{\partial v_{\mu}}{\partial t},
 \end{aligned} \tag{18}$$

we obtain, dividing both sides by ρ and using Equation (15), that

$$\frac{\partial v_{\mu}}{\partial t} - (\partial_{\nu} v_{\nu}) v_{\mu} - v_{\nu} v_{\mu} (\partial_{\nu} \ln \rho) = a_{\mu} - \partial_{\nu} (C_{\mu\nu} + v_{\mu}v_{\nu}) - (C_{\mu\nu} - v_{\mu}v_{\nu}) \partial_{\nu} \ln \rho. \tag{19}$$

The term $v_{\nu} v_{\mu} \partial_{\nu} \ln \rho$ cancels, and we have:

$$\frac{\partial v_{\mu}}{\partial t} - (\partial_{\nu} v_{\nu}) v_{\mu} = a_{\mu} - \partial_{\nu} C_{\mu\nu} - \partial_{\nu} (v_{\mu}v_{\nu}) - C_{\mu\nu} \partial_{\nu} \ln \rho, \tag{20}$$

from which we now cancel the term $(\partial_\nu v_\nu)v_\mu$, arriving at

$$\frac{\partial v_\mu}{\partial t} = a_\mu - \partial_\nu C_{\mu\nu} - v_\nu \partial_\nu v_\mu - C_{\mu\nu} \partial_\nu \ln \rho. \tag{21}$$

Rearranging the derivatives of v_μ in the left-hand side, we have

$$\begin{aligned} \left(\frac{\partial}{\partial t} + v_\nu \partial_\nu\right) v_\mu &= a_\mu - \partial_\nu C_{\mu\nu} - C_{\mu\nu} \partial_\nu \ln \rho \\ &= a_\mu - \frac{1}{\rho} \partial_\nu (\rho C_{\mu\nu}) \\ &= a_\mu + \frac{1}{\rho} \partial_\nu \sigma_{\mu\nu}, \end{aligned} \tag{22}$$

which is the Cauchy momentum equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla\right) \mathbf{v} = \frac{D\mathbf{v}}{Dt} = \mathbf{a}(\mathbf{x}; t) + \frac{1}{\rho} \nabla \cdot \overleftrightarrow{\sigma}, \tag{23}$$

with D/Dt the *convective derivative* and $\overleftrightarrow{\sigma} = -\rho \cdot \overleftrightarrow{\mathbf{C}}$ the stress tensor.

Equations (13) and (23) form a closed coupled system of equations for $\rho(\mathbf{x}; t)$ and $\mathbf{v}(\mathbf{x}; t)$, needing as their only external input the velocity covariance matrix $C_{\mu\nu}$. These equations are then *built-in features* of inference over paths. In a Bayesian approach, they are valid for any system that moves continuously in time. The Cauchy momentum equation includes most notably the Navier–Stokes equation as a particular case [13].

5. Including Particular Knowledge into Our Models

At this point, we have developed a generic framework where no particular details about a system have been included. Clearly all those details have to be contained in $P[\mathbf{X}|\mathcal{I}]$, or rather, in the covariance matrix $C_{\mu\nu}(\mathbf{x}; t)$ which can be derived from it. The question remains about how to incorporate these details in the most unbiased manner. In principle, we could start from the *null* assumption of equiprobable paths,

$$P[\mathbf{X}|\mathcal{I}_0] = \text{constant},$$

and add new information \mathcal{R} later on, by updating our probability functional $P[\mathbf{X}|\mathcal{I}_0]$ to a new $P[\mathbf{X}|\mathcal{I}]$, where $\mathcal{I} = (\mathcal{I}_0, \mathcal{R})$. There are essentially two equivalent methods to achieve this, and depending on the actual form of \mathcal{R} , one of them may be more directly applicable than the other.

- (1) **Bayes’ theorem:** the posterior distribution $P(\mathbf{u}|\mathcal{I}_0, \mathcal{R})$ is given in terms of the prior $P(\mathbf{u}|\mathcal{I}_0)$ by

$$P(\mathbf{u}|\mathcal{I}_0, \mathcal{R}) = \frac{P(\mathbf{u}|\mathcal{I}_0) \cdot P(\mathcal{R}|\mathbf{u}, \mathcal{I}_0)}{P(\mathcal{R}|\mathcal{I}_0)}.$$

This method is most useful when \mathcal{R} is comprised of statements about the states \mathbf{u} (e.g., boundary conditions).

- (2) **Principle of maximum entropy:** the posterior distribution $p(\mathbf{u})$ is the one that maximizes

$$\mathcal{S}[p_0 \rightarrow p] = - \int_V d\mathbf{u} p(\mathbf{u}) \ln \frac{p(\mathbf{u})}{p_0(\mathbf{u})},$$

where $p_0(\mathbf{u})$ is the prior distribution. This method is most useful when \mathcal{R} consists of constraints on the final model $p(\mathbf{u})$, usually expressed as fixed expected values.

In Reference [2], Jaynes assumes the continuity equation from the start and derives the flux $\mathbf{J}(\mathbf{x}; t) = \rho \cdot \mathbf{v}(\mathbf{x}; t)$ from symmetry considerations, the central limit theorem, and Bayes theorem.

In our classification, this corresponds to method (1). In Reference [5], Gull recovers Brownian motion by essentially performing discrete-time maximum caliber inference under constraints over location and particle speed, hence corresponding to an application of method (2).

6. The Maximum Caliber Principle

The function $p(\mathbf{u})$ that is closest to our prior probability $p_0(\mathbf{u})$ and is consistent with the constraints \mathcal{R} is the one that maximizes the relative entropy [14,15]

$$\mathcal{S}[p_0 \rightarrow p] = - \int_V d\mathbf{u} p(\mathbf{u}) \ln \frac{p(\mathbf{u})}{p_0(\mathbf{u})}$$

among the set of functions p that are compatible with \mathcal{R} . The negative of this relative entropy, known as the Kullback–Leibler divergence, is commonly used to measure the “informational distance” from p_0 to p . It is important to note that this is a rule of inference and not a physical principle, and therefore it is not bounded by the meaning assigned to the states \mathbf{x} , as long as we can write (Bayesian) probabilities over them.

For the general case of m constraints of the form

$$\langle f_j(\mathbf{u}) \rangle = F_j, \quad (j = 1, \dots, m),$$

the maximum entropy solution starting from $P(\mathbf{u}|\mathcal{I}_0)$ is obtained through the use of m Lagrange multipliers (one for each constraint),

$$P(\mathbf{u}|\mathcal{I}_0, \mathcal{R}) = P(\mathbf{u}|\mathcal{I}_0) \left[\frac{\exp\left(-\sum_{j=1}^m \lambda_j f_j(\mathbf{x})\right)}{Z(\boldsymbol{\lambda})} \right], \tag{24}$$

where $Z(\boldsymbol{\lambda})$ is the partition function. This is compatible with Bayesian updating, as this posterior distribution is proportional to the prior. The Lagrange multipliers are solutions of the constraint equations in terms of Z :

$$F_j = \frac{\partial}{\partial \lambda_j} \ln Z(\boldsymbol{\lambda}) \quad (j = 1, \dots, m). \tag{25}$$

In exactly the same way, the path (relative) entropy (sometimes known as the *caliber*) is defined as the path integral [1,16–22]:

$$\mathcal{S}[p_0 \rightarrow p] = - \int_{\mathbb{X}} D\mathbf{X} p[\mathbf{X}] \ln \frac{p[\mathbf{X}]}{p_0[\mathbf{X}]},$$

where $p_0[\mathbf{X}] := P[\mathbf{X}|\mathcal{I}_0]$ is the prior path probability. The use of this generalization is justified based on the fact that we can write any path $\mathbf{X}(t)$ in terms of a complete orthonormal basis $\{\mathbf{B}_i\}$,

$$\mathbf{X}(t) = \sum_{n=0}^{N-1} \gamma_n \mathbf{B}_n(t), \quad N \rightarrow \infty, \tag{26}$$

and then there is a one-to-one correspondence between every path $\mathbf{X}(t)$ and its coordinates $(\gamma_0, \gamma_1, \dots, \gamma_{N-1})$. Inference over paths \mathbf{X} then becomes completely equivalent to inference over the coefficients γ , which form a system with N degrees of freedom.

In summary, for the general maximum caliber inference problem we have m constraints, written as

$$\langle \mathcal{F}_j[\mathbf{X}] \rangle = F_j \quad (j = 1, \dots, m), \tag{27}$$

from which the probability functional obtained is

$$P[\mathbf{X}|\mathcal{I}_0, \mathcal{R}] = \frac{1}{Z(\lambda)} P[\mathbf{X}|\mathcal{I}_0] \exp\left(-\sum_{j=1}^m \lambda_j \mathcal{F}_j[\mathbf{X}]\right). \tag{28}$$

Any such maximum caliber solution can be cast in the “canonical” form, as

$$P[\mathbf{X}|\mathcal{I}] = \frac{1}{Z(\alpha)} \exp\left(-\frac{1}{\alpha} \mathcal{A}[\mathbf{X}]\right), \tag{29}$$

where \mathcal{A} is a functional, analogous to the Hamilton action of a classical system, and $\alpha > 0$ is a constant with the same physical units as \mathcal{A} . By simple inspection of this canonical form, it is straightforward to see that the most probable path is the one with minimum action:

$$\frac{\delta \mathcal{A}[\mathbf{X}]}{\delta \mathbf{X}(t)} = 0, \quad \forall t. \tag{30}$$

7. An Illustration: Newtonian Mechanics of Charged Particles

As an example of the application of this formalism, consider a “particle” with known square speed v^2 , known instantaneous probability density ρ , and known velocity field $\mathbf{v}(\mathbf{x}; t)$ for all times $t \in [0, \tau]$. The corresponding constraints are then

$$\langle |\dot{\mathbf{X}}(t)|^2 \rangle = v^2(t) \quad \forall t, \tag{31}$$

$$\langle \delta(\mathbf{X}(t) - \mathbf{x}) \rangle = \rho(\mathbf{x}; t) \quad \forall \mathbf{x}, t, \tag{32}$$

$$\langle \dot{\mathbf{X}}(t) \delta(\mathbf{X}(t) - \mathbf{x}) \rangle = \rho(\mathbf{x}; t) \mathbf{v}(\mathbf{x}; t) \quad \forall \mathbf{x}, t. \tag{33}$$

The resulting maximum caliber solution is of the form

$$P[\mathbf{X}|\mathcal{I}] = \frac{1}{Z(\alpha)} \exp\left(-\frac{1}{\alpha} \int_0^\tau dt L(\mathbf{X}(t), \dot{\mathbf{X}}(t); t)\right), \tag{34}$$

with the Hamilton action

$$\mathcal{A}[\mathbf{X}] = \int_0^\tau dt L(\mathbf{X}(t), \dot{\mathbf{X}}(t); t), \tag{35}$$

and a Lagrangian defined as

$$L(\mathbf{X}(t), \dot{\mathbf{X}}(t); t) = \lambda_1(t) |\dot{\mathbf{X}}(t)|^2 + \int_V dx (\lambda_2(\mathbf{x}; t) + \lambda_3(\mathbf{x}; t) \cdot \dot{\mathbf{X}}(t)) \delta(\mathbf{X}(t) - \mathbf{x}), \tag{36}$$

where λ_1, λ_2 , and λ_3 are Lagrange multipliers. This Lagrangian can be cast into a more familiar form,

$$L(\mathbf{X}(t), \dot{\mathbf{X}}(t); t) = \frac{1}{2} m(t) \dot{\mathbf{X}}(t)^2 - \Phi(\mathbf{X}(t); t) + \dot{\mathbf{X}}(t) \cdot \mathbf{A}(\mathbf{X}(t); t), \tag{37}$$

by simply renaming the Lagrange multipliers and integrating the delta function [18]. Interestingly, this is none other than the Lagrangian for a particle with time-dependent “mass” $m(t)$ in an external “electromagnetic potential” (Φ, \mathbf{A}) . The most probable path under these constraints is determined by the solution of the Euler–Lagrange equation, which reduces to Newton’s second law under a “Lorenz force”,

$$\frac{d}{dt} (m\mathbf{v}) = -\nabla\Phi(\mathbf{x}; t) - \frac{\partial \mathbf{A}(\mathbf{x}; t)}{\partial t} + \mathbf{v} \times (\nabla \times \mathbf{A}(\mathbf{x}; t)), \tag{38}$$

as shown in the Appendix. In particular, it is important to note that it is the constraint on the squared speed $v^2(t)$ that adds the mass $m(t)$ to the model, as $m = 2\lambda_1$, the constraint on the probability density

$\rho(\mathbf{x}; t)$ adds the scalar potential $\Phi(\mathbf{x}; t)$ to the model, as $\Phi = -\lambda_2$, and finally the constraint on the local velocity field $v(\mathbf{x}; t)$ adds the vector potential $A(\mathbf{x}; t)$ to the model, because $A = \lambda_3$. Nowhere in the derivation of this Lagrangian have we assumed the existence of charges, electromagnetic fields, or the Lorenz force. The structure that is revealed is the most unbiased under the constraints given in Equations (31) to (33), that is, with approximate knowledge of its location (given by ρ) and velocity “field lines” (given by v). This model could be used for people in a busy street crossing, or vehicles in a city.

8. Concluding Remarks

We have shown that it is possible to construct a fluid theory from Bayesian inference of an abstract system with N degrees of freedom moving along paths $X(t)$, and that this theory automatically includes the continuity equation and the Cauchy momentum equation as built-in features. Moreover, through the use of the Maximum Caliber principle, it is possible to formulate the dynamics of such an abstract system in terms of an action that is minimal for the most probable path, resembling the well-known structures of Lagrangian and Hamiltonian mechanics.

By entering the square speed, instantaneous probability density, and velocity field into our model, a Lagrangian of a “particle” under external fields emerges naturally. This “particle” moves on average according to Newton’s law of motion $F = m \cdot a$ under the Lorenz force, with scalar and vector potentials determined by the known information about the location and velocity lines. In this formulation, the only ingredients that we could call *physical* were the existence of an N -dimensional “particle” moving continuously along (unknown) paths. In this application, we see that position and velocity are the only intrinsic (*real* or *ontological*) properties of the particle at a given time t . On the other hand, the time-dependent mass $m(t)$ and the fields $\Phi(\mathbf{x}; t)$ and $A(\mathbf{x}; t)$ are emergent parameters (in fact, Lagrange multipliers) needed to impose the constraints on the known information used to construct the model.

Author Contributions: All authors of this paper contributed to the work. S.D. conceived the structure of the paper while D.G. and G.G. provided analysis of previous results and discussions.

Funding: This research was funded by FONDECYT grant 1171127.

Acknowledgments: S.D. and G.G. gratefully acknowledged funding from FONDECYT 1171127.

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

Appendix A. Derivation of the Lorenz Force from the Lagrangian of a Particle in an Electromagnetic Field

We start from the Euler–Lagrange equation

$$\frac{\partial L}{\partial x_\mu} - \frac{d}{dt} \left(\frac{\partial L}{\partial v_\mu} \right) = 0 \quad (\text{A1})$$

for the Lagrangian in Equation (37),

$$L(\mathbf{x}, \mathbf{v}; t) = \frac{1}{2} m(t) v^2 - \Phi(\mathbf{x}; t) + v_\nu A_\nu(\mathbf{x}; t),$$

which gives

$$-\partial_\mu \Phi + v_\nu \partial_\mu A_\nu - \frac{d}{dt} (m v_\mu + A_\mu) = 0. \quad (\text{A2})$$

Replacing the total time derivative of the vector field $A(\mathbf{x}; t)$,

$$\frac{d}{dt} A_\mu(\mathbf{x}; t) = \partial_t A_\mu + v_\nu \partial_\nu A_\mu,$$

we have

$$\frac{d}{dt}(mv_\mu) + \partial_t A_\mu + v_\nu \partial_\nu A_\mu = -\partial_\mu \Phi + v_\nu \partial_\nu A_\nu, \quad (\text{A3})$$

which can be written as

$$\frac{d}{dt}(mv_\mu) = (-\partial_\mu \Phi - \partial_t A_\mu) + v_\nu (\partial_\mu A_\nu - \partial_\nu A_\mu). \quad (\text{A4})$$

Using the fact that the i -th component of $\mathbf{v} \times (\nabla \times \mathbf{A})$ is given by

$$\begin{aligned} [\mathbf{v} \times (\nabla \times \mathbf{A})]_i &= \epsilon_{ijk} v_j (\epsilon_{klm} \partial_l A_m) \\ &= \epsilon_{ijk} \epsilon_{klm} (v_j \partial_l A_m) \\ &= (\delta_{il} \delta_{jm} - \delta_{im} \delta_{jl}) v_j \partial_l A_m \\ &= v_j (\partial_i A_j - \partial_j A_i), \end{aligned} \quad (\text{A5})$$

we finally obtain

$$\frac{d}{dt}(mv_\mu) = (-\partial_\mu \Phi - \partial_t A_\mu) + [\mathbf{v} \times (\nabla \times \mathbf{A})]_\mu, \quad (\text{A6})$$

which is Equation (38).

References

- Jaynes, E.T. The Minimum Entropy Production Principle. *Ann. Rev. Phys. Chem.* **1980**, *31*, 579–601. [[CrossRef](#)]
- Jaynes, E.T. Clearing up Mysteries—The Original Goal. In *Maximum Entropy and Bayesian Methods: Cambridge, England, 1988*; Springer: Dordrecht, The Netherlands, 1989; pp. 1–27.
- Dewar, R. Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. *J. Phys. A Math. Gen.* **2003**, *36*, 631–641. [[CrossRef](#)]
- Grandy, W.T. *Entropy and the Time Evolution of Macroscopic Systems*; Oxford Science Publications: New York, NY, USA, 2008.
- Gull, S. Some Misconceptions about Entropy. Available online: <http://www.mrao.cam.ac.uk/steve/maxent2009/images/miscon.pdf> (accessed on 11 September 2018).
- Cox, R.T. Probability, frequency and reasonable expectation. *Am. J. Phys.* **1946**, *14*, 1–13. [[CrossRef](#)]
- Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, UK, 2003.
- Risken, H. *The Fokker-Planck Equation: Methods of Solution and Applications*; Springer: Berlin, Germany, 1996.
- Van Campen, N.G. *Stochastic Processes in Physics and Chemistry*; North Holland: Amsterdam, The Netherlands, 2007.
- Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University Press: New York, NY, USA, 2001.
- Van Campen, N.G. The expansion of the master equation. *Adv. Chem. Phys.* **1976**, *34*, 245–309.
- González, D.; Díaz, D.; Davis, S. Continuity equation for probability as a requirement of inference over paths. *Eur. Phys. J. B* **2016**, *89*, 214. [[CrossRef](#)]
- Lamb, H. *Hydrodynamics*; Dover Books on Physics; Dover Publications: New York, NY, USA, 1945.
- Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [[CrossRef](#)]
- Caticha, A.; Giffin, A. Updating Probabilities. *AIP Conf. Proc.* **2006**, *872*, 31.
- Pressé, S.; Ghosh, K.; Lee, J.; Dill, K.A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **2013**, *85*, 1115–1141. [[CrossRef](#)]
- González, D.; Davis, S.; Gutiérrez, G. Newtonian mechanics from the principle of Maximum Caliber. *Found. Phys.* **2014**, *44*, 923. [[CrossRef](#)]
- Davis, S.; González, D. Hamiltonian Formalism and Path Entropy Maximization. *J. Phys. A Math. Theor.* **2015**, *48*, 425003. [[CrossRef](#)]
- Hazoglou, M.J.; Walther, V.; Dixit, P.D.; Dill, K.A. Maximum caliber is a general variational principle for nonequilibrium statistical mechanics. *J. Chem. Phys.* **2015**, *143*, 051104. [[CrossRef](#)] [[PubMed](#)]

20. Wan, H.; Zhou, G.; Voeltz, V.A. A maximum-caliber approach to predicting perturbed folding kinetics due to mutations. *J. Chem. Theory Comput.* **2016**, *12*, 5768–5776. [[CrossRef](#)] [[PubMed](#)]
21. Cafaro, C.; Ali, S.A. Maximum caliber inference and the stochastic Ising model. *Phys. Rev. E* **2016**, *94*, 052145. [[CrossRef](#)] [[PubMed](#)]
22. Dixit, P.; Wagoner, J.; Weistuch, C.; Pressé, S.; Ghosh, K.; Dill, K.A. Perspective: Maximum caliber is a general variational principle for dynamical systems. *J. Chem. Phys.* **2018**, *148*, 010901. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).