



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

Departamento de Física

TESIS PRESENTADA PARA OPTAR AL GRADO DE DOCTOR DE LA FACULTAD DE
CIENCIAS EXACTAS

MODELOS MINIMALES Y TEORÍA DE
LA INFORMACIÓN DE POBLACIONES
NEURONALES

TESISTA: LIC. LISANDRO MONTANGIE

DIRECTOR: DR. FERNANDO FABIÁN MONTANI

2017

Resumen

El cerebro es un sistema dinámico cuyas variables de estado codifican la información sobre el mundo exterior. Entender cómo las neuronas procesan e integran las señales de entrada requiere entender sus acciones cooperativas, en vez de describir únicamente su actividad como elementos individuales. No obstante, construir una aproximación apropiada de la actividad colectiva en la corteza cerebral resulta una tarea difícil. En particular, se han propuesto modelos de máxima entropía para investigar correlaciones entre pares en poblaciones de neuronas, que han logrado describir efectivamente la actividad de ciertas regiones como la retina. Sin embargo, desviaciones de este modelo indican que las correlaciones de alto orden tienen que ser tenidas en cuenta para lograr un modelado realista en general. En esta Tesis, abordamos esta problemática a partir de dos enfoques particulares.

En una primera parte, presentamos un estudio basado en la descomposición de la información transmitida por una población pequeña de neuronas. En el capítulo 3, elaboramos una extensión de la expansión en series para ventanas temporales cortas en el marco previo, para estudiar las contribuciones de los tripletes a la codificación del estímulo. Luego, en el capítulo 4, introducimos los estimadores apropiados para el enfoque desarrollado.

En segunda instancia, profundizamos un modelo poblacional sencillo mediante la introducción de correlaciones de alto orden. En el capítulo 5, presentamos esta extensión obteniendo analíticamente la distribución conjunta de disparos poblacional asintótica en este contexto. Luego, en el capítulo 6, estudiamos cómo se relacionan las estadísticas de las señales de entrada comunes al conjunto de neuronas con las de sus patrones de disparo. Finalmente, en el capítulo 7, evaluamos este modelo funcionalmente a medida que consideramos grupos neuronales de distinto tamaño.

Agradecimientos

Que lo breve de mis palabras no sean medida de lo que todo lo agradecido que estoy.

A Eze, por ser mi compañero en todo.

A Fernando, por ser lo que considero un buen director y guiarme a lo largo de este proceso.

A todas las personas que considero amigos, que por suerte son muchos. Espero haberles transmitido todo lo que significan para mí.

A todas la gente del IFLY que lo hacen un excelente lugar para trabajar (en especial a los geniales compañeros con los que tuve el placer de compartir oficina).

A mi familia.

A la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata.

A CONICET por el soporte económico para realizar esta tesis, con la esperanza de que en un futuro se torne una institución más democrática e inclusiva.

Índice general

Resumen	3
Agradecimientos	5
1. Introducción	12
1.1. Hacia una descripción teórica del código neuronal	14
1.2. Los modelos minimales	16
1.2.1. La inclusión de las correlaciones de alto orden	19
1.3. Objetivos de esta Tesis	20
2. Marco teórico general	21
2.1. Representación de la respuesta neuronal	21
2.2. Teoría de la información en neurociencia	24
2.2.1. Entropía	24
2.2.2. Divergencias entre dos distribuciones	27
2.2.3. Información mutua	28
2.2.4. Descomposición de la información en mecanismos de codificación	30
2.3. Métodos de la geometría de la información	35
2.3.1. Conceptos básicos de geometría diferencial	36
2.3.2. Variedades de familias exponenciales	38
3. Contribuciones de alto orden al código neuronal	40
3.1. Expansión de la información para ventanas temporales cortas	41
3.1.1. Contribuciones de primer y segundo orden de la expansión . .	47
3.1.2. Contribuciones de tercer orden en la expansión	48

3.1.3. Términos de orden superior	51
3.2. Codificación sinérgica y redundante en términos de alto orden	54
3.3. Modelado computacional con estadística de Poisson	58
3.3.1. Grado de sinergia	63
3.4. Modelo de red tipo integrar-y-disparar	64
3.5. Modelo de red tipo Izhikevich	65
4. Restricciones de máxima entropía en las correlaciones entre tripletes	71
4.1. Interacciones neuronales en el marco de la geometría de la información	72
4.1.1. Maximización de la entropía	76
4.1.2. Interacciones entre tripletes	78
4.2. Cuantificación de correlaciones entre tripletes espurias	82
4.2.1. Caso de triplete homogéneo	83
5. Correlaciones de alto orden en una población homogénea	85
5.1. Propuesta del modelo extendido	87
5.1.1. ¿Cómo podemos interpretar el teorema del límite central extendido en el marco teórico de la geometría de la información?	91
5.2. Resultados	92
5.2.1. Distribución de disparos asintótica	92
5.2.2. Estimaciones analíticas	96
5.2.3. Entropía relativa e información de Fisher	96
6. Generación de patrones de actividad poblacionales con entradas correlacionadas	102
6.1. El modelo generador de patrones de actividad poblacionales	103
6.2. Relacionando correlaciones de alto orden en la entrada y la salida	105
6.3. Modelo de población homogénea	106
6.3.1. Resultados numéricos	109
6.3.2. Cálculo de la entropía y la divergencia de Kullback-Leibler del modelo	112
6.3.3. Estudio de la causalidad del modelo	114

7. Propiedades de escala de las entradas con distribuciones con colas pesadas	121
7.1. Modelando potenciales de membrana	122
7.2. Resultados	125
7.2.1. Coeficientes de interacción del modelo q -DG	125
7.2.2. Esfuerzo	127
7.2.3. Dispersión estadística	129
7.2.4. Tasa de entropía y calor específico de la población	131
8. Discusión y Conclusiones	134
A. Cálculo de la información de Fisher para la distribución conjunta de disparos	142
B. Estimaciones numéricas de la función de distribución cumulativa q-gaussiana bivariada	144
C. Cuantificadores de complejidad del modelo de neurona simple de Izhikevich	148
D. Trabajos publicados	151
Bibliografía	152

Índice de figuras

1.1. Patrones de disparo poblacionales	17
1.2. Modelos de máxima entropía de segundo orden	19
2.1. Relación entre las entropías y la información mutua	29
2.2. Sistema de coordenadas de S'	37
3.1. Esquema de tres neuronas interactuantes	50
3.2. I_{SRL} versus correlación de ruido entre tripletes γ_{123} versus correlación de señal entre tripletes ν_{123} , con $f(\bar{\mathbf{r}}) = 0$	59
3.3. I_{SRL} versus correlación de ruido entre tripletes γ_{123} versus correlación de señal entre tripletes ν_{123} , con $f(\bar{\mathbf{r}}) > 0$	60
3.4. I_{SRL} versus correlación de ruido entre tripletes γ_{123} versus correlación de señal entre tripletes ν_{123} , con $f(\bar{\mathbf{r}}) < 0$	61
3.5. Fracción de sinergia	69
3.6. Comportamientos neurocomputacionales relevantes de neuronas biológicas	70
3.7. Componentes en el modelo de Izhikevich	70
4.1. Subconjuntos y coordenadas ortogonales	75
4.2. Distribuciones de probabilidad en la variedad	77
4.3. Correlaciones espurias entre tripletes idénticos	83
5.1. Distribución concentrada de la tasa de disparos	97
5.2. Distribución ampliamente distribuida de la tasa de disparos	97
5.3. Distribución bimodal de la tasa de disparos	98
5.4. Entropía relativa de Tsallis del modelo extendido	99

5.5. Información de Fisher de la distribución	100
6.1. Gráficas rasterizadas de patrones de disparo multineuronales del modelo DG	110
6.2. Gráficas rasterizadas de patrones de disparo multineuronales del modelo q -DG	111
6.3. Distribución conjunta de disparos como función del número de espigas normalizado	112
6.4. Correlación en la salida versus correlación en la entrada	116
6.5. Correlación en la salida como función de la tasa de disparo	117
6.6. Entropía diferencial normalizada	118
6.7. Divergencia de Kullback-Leibler normalizada	118
6.8. Complejidad en función del parámetro de deformación	119
6.9. Complejidad versus entropía para el modelo q -DG	119
6.10. Información de Fisher versus complejidad para el modelo q -DG	120
7.1. Curtosis en exceso de la distribución q -gaussiana en función del parámetro de deformación	124
7.2. Coeficientes de interacción	127
7.3. Valores del esfuerzo de un conjunto de tres neuronas idénticas	129
7.4. Dispersión estadística en función del tamaño de la población	130
7.5. Tasa de entropía en función del tamaño de la población	132
7.6. Calor específico en función del tamaño de la población	133
C.1. Representaciones bidimensionales de los cuantificadores de información causal	149
C.2. Representación tridimensional de los cuantificadores de información causal	150

Capítulo 1

Introducción

Nuestro entendimiento de cómo el mundo sensorial es representado por la actividad eléctrica de los nervios sensoriales se encuentra limitada, en un principio, por nuestra capacidad para registrar esta actividad. Gran parte de lo que conocemos hoy en día como el problema de la codificación neuronal fue establecido en aproximadamente una década por Adrian y colaboradores a principios del siglo XX. Los primeros trabajos clásicos de Adrian están contenidos principalmente en una serie de publicaciones en el año 1926, dos de ellos junto con Zotterman [1–3]. Estos experimentos establecieron tres hechos fundamentales. Primero, se mostró que las neuronas sensoriales individuales producen potenciales de acción estereotipados. Estas respuestas, también conocidas como espigas, son ondas de descargas eléctricas que viajan a lo largo de la membrana celular modificando las distribuciones de carga eléctrica de las mismas. Esto significa que las neuronas individuales sólo pueden transmitir información en el cerebro a través del tiempo relativo de llegada o el número de estas espigas. Segundo, Adrian observó que en respuesta a un estímulo estático, en su caso la carga continua en un receptor de estiramiento de los músculos de una rana, la tasa de disparos de estos potenciales de acción incrementa a medida que el estímulo se vuelve mayor. Por ende, la intensidad del estímulo es la frecuencia de estos disparos. El tercer descubrimiento fue que si el estímulo estático es continuado por un tiempo prolongado, la tasa de disparos comienza a disminuir, propiedad conocida como adaptación. Adrian sugirió que este fenómeno fisiológico se corresponde con el fenómeno perceptual en donde perdemos gradualmente la conciencia de los estímu-

los constantes. En resumen, estos primeros experimentos establecieron el paradigma de la neurona (o circuitos neuronales) como elemento capaz de “computar” señales externas.

Caracterizar la relación que existe entre el estímulo y la respuesta al mismo es difícil, debido a que las señales neuronales son complejas y variables. Las neuronas responden produciendo secuencias de disparos que reflejan tanto la dinámica intrínseca de la neurona como las características temporales del estímulo. De la serie temporal producida puede ser importante la cantidad de potenciales de acción en un cierto tiempo, el tiempo preciso de los mismos o cuánto se relacionan las respuestas de neuronas distintas, entre otros. De esta forma, es posible definir códigos de conteo de espigas, códigos temporales o códigos de correlación. Un código de disparos independientes correspondería al caso en donde la generación de cada potencial de acción es independiente de todas las otras espigas de la secuencia. En contraste, si cada disparo no codifica independientemente del resto, llamamos al código correlacionado, debido a que estos eventos simultáneos entre espigas pueden transmitir información adicional. Es necesario aclarar que con correlación nos referimos a la actividad conjunta sistemática entre grupos de neuronas.

En realidad, la información es probablemente transmitida tanto por disparos individuales como a través de correlaciones, y una línea divisoria arbitraria debe ser trazada para caracterizar el código. Aislar las propiedades de la respuesta que codifican cambios en el estímulo puede resultar difícil, especialmente si la escala temporal para estos cambios es del mismo orden que el intervalo promedio entre potenciales de acción. Además, las respuestas neuronales pueden variar de prueba en prueba, incluso aún cuando el mismo estímulo es presentado repetidas veces. Existen muchas razones posibles que expliquen esta variabilidad, incluyendo niveles variables de excitación, una aleatoriedad asociada con diversos procesos biofísicos que afectan el disparo de las neuronas, y los efectos de otros procesos cognitivos que ocurren durante el ensayo experimental. La complejidad y la variabilidad de las secuencias de los potenciales de acción, entre ensayo y ensayo, provocan que sea poco probable la descripción y predicción determinística del momento preciso de cada disparo. Por tal razón, se busca una representación probabilística de la respuesta. Es decir, poder obtener las probabilidades de la generación de distintas secuencias de

disparos frente ante estímulos específicos. Típicamente, muchas neuronas responden a un dado estímulo, y las características de dicho estímulo son entonces codificadas por las actividades de grandes poblaciones de neuronas. Al estudiar la codificación de poblaciones, se deben examinar no sólo los patrones de disparo individuales de cada neurona, sino también las relaciones que existen entre estos patrones de disparo a lo largo de toda la población de células que generan una respuesta al estímulo.

1.1. Hacia una descripción teórica del código neuronal

A partir de estos y siguientes experimentos, el interés por comprender cómo el cerebro representa y procesa la información ha generado múltiples esfuerzos. Sin embargo, la definición de qué significa la comprensión del código neuronal en una región del sistema nervioso o cómo cuantificamos la noción de que una neurona transmite información sobre el entorno a través de una serie de potenciales de acción resulta central. Por lo tanto, requiere de enfoques interdisciplinarios que participen en distintos niveles de estudio: desde el nivel molecular hasta el nivel celular (neuronas individuales), al nivel de conjuntos reducidos (redes de pequeño mundo) de neuronas (como por ejemplo las columnas corticales), o a escala de grandes subsistemas como el de la percepción visual, que incluye regiones de la corteza cerebral, el cerebelo y el sistema nervioso como un conjunto. Y a su vez, la complejidad inherente al objeto de estudio fomenta el establecimiento de formulaciones matemáticas precisas y del desarrollo de modelos cuantitativos. Los enfoques teóricos en el campo de las neurociencias han ayudado a proponer nuevas ideas y medios para futuras investigaciones [4, 5]. El principal rol de un teórico es el de formular y explorar modelos, evaluarlos y luego validarlos o rechazarlos y postular nuevas hipótesis para ser verificadas por experimentales. Una ventaja importante de la neurociencia teórica es la velocidad y eficiencia con la cual los modelos pueden ser analizados y evaluados. Las técnicas disponibles para poder llevar a cabo análisis matemáticos y computacionales son rápidas en comparación al tiempo de realización de los experimentos.

Para poder entender el cómputo realizado por la neurona o el circuito neuronal, los teóricos formulan una descripción matemática de la relación entre la respuesta

neuronal del sistema considerado y el estímulo que evocó dicha respuesta. Esto lleva a que los modelos sean precisos, aunque frecuentemente abstractos. Un ejemplo es la teoría de la información [6, 7], la cual ha demostrado ser muy exitosa para estudiar las propiedades del procesamiento sensorial de las neuronas. Este marco teórico utiliza promedios de disparos y covarianzas para cuantificar la información codificada por los potenciales de acción. Originalmente, el formalismo de la teoría de la información fue desarrollado por Shannon a fines de la década de 1940 para cuantificar la habilidad de un esquema de codificación o un canal de comunicación para transmitir información. Posteriormente, su potencial para el estudio de datos experimentales en neurociencia fue introducido por Werner y Mountcastle [8], y el uso de esta aproximación ha avanzado considerablemente durante los últimos años. El entendimiento de los principios que puedan llevar al procesamiento de la información sensorial fue ampliamente beneficiado por el uso de la teoría de la información. Este método puede ser extendido a través de la separación de la información total en componentes individuales que reflejan los diferentes aspectos de la actividad neuronal (es decir, las correlaciones entre neuronas, las tasas de disparo, entre otros [9–13]), permitiendo entender exactamente cómo las neuronas procesan información sensorial.

No obstante, únicamente extraer información de un experimento es una visión limitada de las posibilidades de los estudios teóricos. Construir modelos que reproduzcan evidencias biológicas, tales como patrones de actividad de poblaciones de neuronas, ha sido la principal contribución que las ciencias duras han otorgado al campo [4, 5]. Un buen modelo teórico corresponde a aquel capaz de identificar el conjunto mínimo de características necesarias para comprender un fenómeno en particular encontrado en un sistema biológico. Con un número de suposiciones mayor al necesario para una descripción minimal, los modelos se tornan muy complejos para ser analizados y evaluados. Sin embargo, un modelo apropiado debería ser lo suficientemente flexible tal que sus características puedan ser revisadas y reemplazadas por otras, en el caso que sea necesario. Realizar simplificaciones considerables al desarrollar modelos está justificado siempre y cuando estos generen mensajes claros y puedan capturar los fenómenos relevantes del sistema.

Particularmente, construir modelos en neurociencia resulta difícil a causa de diversas razones [14]. Entre ellas, se encuentran principalmente el hecho de que los sis-

temas biológicos están compuestos por un gran número de elementos [15], que dichos elementos típicamente interactúan entre ellos y, finalmente, que se deban comparar con un número acotado de muestras debido a las limitaciones tecnológicas en la adquisición de datos. Por ejemplo, actualmente es posible registrar simultáneamente la actividad de sólo alrededor de cien neuronas y avances tecnológicos modernos permitirán prontamente el registro de miles e incluso decenas de miles de neuronas [16], de las aproximadamente cien billones que constituyen el cerebro humano. Al considerar una representación binaria de la actividad de las neuronas, es decir solamente dos estados son posibles, es necesario estimar 2^N patrones de actividad, donde N es el número de células. Por esta razón, las cantidades que describen a un sistema se deben calcular a partir de pequeños conjuntos. A pesar de este hecho, trabajos recientes [17, 18] han sugerido que descripciones precisas en un marco probabilístico de estos sistemas pueden ser logradas sin tener que analizar todas las posibles configuraciones del mismo. Estos modelos básicos ofrecen potencialmente una simplificación fundamental, pero es necesario debatir si estas consideraciones en subsistemas de menor tamaño pueden ser extrapoladas correctamente a sistemas de mayor dimensión.

1.2. Los modelos minimales

Considerar la menor cantidad de suposiciones, es decir formular modelos sin incorporar conocimientos adicionales a los que surgen de los datos experimentales, nos lleva a los modelos de máxima entropía. En mecánica estadística, el principio de máxima entropía [19, 20] establece que la distribución de probabilidad menos sesgada que se le puede atribuir a un sistema estadístico es aquella que maximiza la entropía, o la desinformación, dadas ciertas condiciones fijas. En otras palabras, en alguna situación de desconocimiento, la distribución estadística con menor sesgo es aquella en la que menos información extrínseca al problema contenga. Por lo tanto, supongamos que la actividad de las neuronas que registramos depende en parte de las interacciones entre ellas pero no es posible acceder a los parámetros detallados de las mismas y entonces no resulta posible una descripción completa del sistema. Entonces, ¿es posible inferir propiedades del sistema usando modelos restringidos por la información disponible? Esto automáticamente genera distribuciones de ti-

po Boltzmann, definiendo un mapa de energía sobre los estados del sistema. Por lo pronto, nos concentraremos en distribuciones estáticas, es decir que no varían con el tiempo. Observar la actividad de una población de neuronas en un cierto instante implica tener un registro de los potenciales de acción en cada ventana temporal, como se muestra en la Figura 1.1. A su vez, para poder modelar el patrón en estudio se procede a binarizar cada ventana temporal, asignando un valor cuando la neurona haya disparado y otro cuando la neurona se mantuvo en silencio.

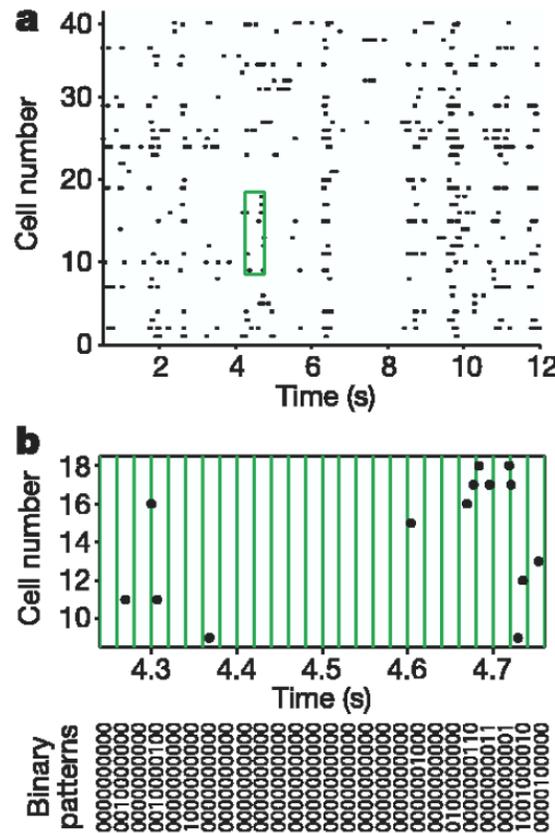


Figura 1.1: **a**: Un segmento de las respuestas simultáneas de 40 células ganglionares de la retina de una salamandra a un estímulo visual natural. Cada punto representa el momento de disparo de un potencial de acción. **b**: Discretización de los trenes de espiga poblacionales en patrones binarios para la región encuadrada (en verde) de **a**. Cada línea (del panel inferior) describe la actividad de las células a un dado instante. Se muestran 10 de 40 células. Imágen adaptada de [17].

Si consideramos como restricciones las actividades medias individuales (promedios de disparos) y las covarianzas entre pares de neuronas, obtenemos modelos completamente análogos al modelo de Ising [21], un modelo matemático de ferromagnetismo ampliamente utilizado en mecánica estadística. En este modelo la actividad de un espín (neurona) i es representada por una variable $x_i = \pm 1$, tal que la distribución de probabilidades del sistema resulta

$$P_{Ising}(x_1, \dots, x_N) = \frac{1}{Z} \exp \left\{ \sum_{i=1}^N h_i x_i + \frac{1}{2} \sum_{i \neq j} J_{ij} x_i x_j \right\}, \quad (1.1)$$

donde h_i es el campo local (sesgo de la neurona i), J_{ij} es la fuerza de acoplamiento simétrico entre espines (entre neuronas) y Z es la función de partición. Estas dos cantidades pueden ser obtenidas a partir de la actividad media de cada neurona individual $\langle x_i \rangle$ y las correlaciones de pares $\langle x_i x_j \rangle$ [22]. Análogamente, la actividad de cada neurona puede ser representada por 0 y 1, o cualquier otra representación binaria mientras se mantengan las normalizaciones correspondientes. Los modelos de máxima entropía de interacción entre pares de neuronas han probado ser aproximaciones válidas en poblaciones de la retina, reduciendo sustancialmente el número de parámetros necesarios [17]. A pesar de la potencial complejidad de las interacciones en la retina, este modelo captura más del 90 % de la estructura en los patrones detallados de los disparos y silencios en la red, en comparación a las predicciones de modelos que asumen independencia entre sus elementos (ver figura 1.2). Debido a que este modelo tiene una cantidad reducida de parámetros, es posible evadir la “maldición” de la dimensionalidad y los problemas de muestreo asociados que comúnmente limitarían la exploración de redes más grandes. El predominio de las interacciones entre pares significa que las reglas de aprendizaje basadas en correlaciones entre pares [23] podrían ser suficientes para generar modelos internos quasi-óptimos para la distribución de combinaciones binarias del código retinal, y por ende permitiendo al cerebro evaluar precisamente nuevos eventos por su grado de sorpresa.

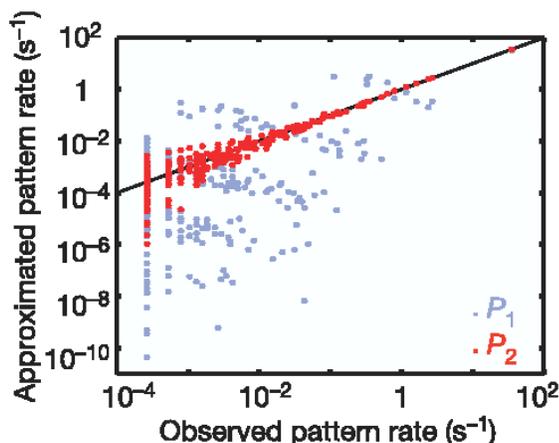


Figura 1.2: Utilizando el mismo grupo de 10 células de la figura 1.1, se graficó la tasa de ocurrencia de cada patrón de disparo predicho por un modelo independiente (indicado por P_1 , en gris) y por el modelo de máxima entropía (indicado por P_2 , en rojo) que toma en cuenta todas las correlaciones entre pares en comparación con la tasa medida. Las tasas de patrones que ocurren comúnmente son predichos con una precisión mejor que el 10%, y la dispersión entre predicciones y observaciones está confinada en su mayoría a eventos raros para los cuales la medición de las tasas en sí no es certera. La línea negra denota igualdad. Figura adaptada de [17].

1.2.1. La inclusión de las correlaciones de alto orden

Sin embargo, estos modelos pueden no ser suficientes para entender redes corticales más complejas. Por esta razón, resulta necesario estudiar cómo las interacciones de órdenes mayores a dos entre neuronas y por ende las llamadas correlaciones de alto orden afectan la descripción de la dinámica de los sistemas neuronales. Trabajos recientes [24, 25], se concentraron en identificar cuándo los modelos que sólo involucraran interacciones entre pares fallan sustancialmente en la representación de poblaciones de la corteza. Efectivamente, detectaron la presencia de interacciones de órdenes superiores en regiones como la corteza sensorial, mediante la comparación con modelos de máxima entropía en marcos más generales. Estos modelos, cuya implementación se encuentra correctamente definida en el marco de la geometría de la

información [26], están caracterizados por

$$P(x_1, \dots, x_N) = \exp \left\{ \sum_{i=1}^N x_i \theta_i + \sum_{i<j} x_i x_j \theta_{ij} + \sum_{i<j<k} x_i x_j x_k \theta_{ijk} + \sum_{i<j<k<l} x_i x_j x_k x_l \theta_{ijkl} + \dots + x_1 \dots x_N \theta_{1\dots N} - \psi \right\}, \quad (1.2)$$

donde existen un total de $2^N - 1$ coeficientes de correlación θ diferentes que identifican interacciones de cualquier orden, y ψ corresponde a la constante de normalización.

La detección de las mismas despertó interés respecto al origen y funcionalidad de estas correlaciones. Este hecho propulsó la formulación de modelos más fenomenológicos que, aunque no estuviesen postulados dentro del principio de entropía máxima, representasen minimalmente la generación de estas estructuras. Diversos trabajos han estudiado la importancia de las correlaciones de alto orden en la función de un circuito neuronal y la codificación [27, 28], y cómo estas interacciones afectan el disparo de una neurona post-sináptica [29]. Un ejemplo en particular que utilizaremos a lo largo de la Tesis es el modelo de Gaussiana Dicotomizada [30–32], el cual ha probado ser efectivo para representar y reproducir patrones de actividad que poseen correlaciones de alto orden [33, 34].

1.3. Objetivos de esta Tesis

Dados los antecedentes previamente mencionados, en esta Tesis trabajaremos dos objetivos generales. En primera instancia, se buscará encontrar, en el marco de la teoría de la información, medidas que permitan evaluar el impacto de las correlaciones de alto orden en la transmisión de la información. Para esto, se utilizará una aproximación en expansión en series para ventanas temporales cortas y poblaciones de neuronas pequeñas. En segundo lugar, se estudiarán estos efectos pero en poblaciones grandes de neuronas, y su impacto en la capacidad de representación del estímulo. Para este objetivo, se formulará la extensión de un modelo de baja dimensionalidad que comprenda las contribuciones con estadística de alto orden, a través de la inclusión de un único parámetro.

Capítulo 2

Marco teórico general

Las neuronas representan y transmiten información mediante secuencias de disparos en distintos patrones temporales. Para poder estudiar la codificación neuronal es necesario caracterizar cómo los atributos de un estímulo están representados por estos potenciales de acción, introduciendo convenientemente a este análisis conceptos de probabilidades y de la teoría de la información. Por lo tanto, en este capítulo desarrollaré brevemente las representaciones y herramientas matemáticas que serán usadas a lo largo de la Tesis.

2.1. Representación de la respuesta neuronal

Los potenciales de acción, o espigas, transmiten información a través de sus ritmos de disparo. Aunque puedan variar en duración, amplitud y forma, son generalmente tratados como eventos idénticos estereotipados al estudiar la codificación neuronal. Si se ignora la corta duración de un potencial de acción (alrededor de 1 ms), un tren de disparos puede ser caracterizado sencillamente por el conjunto de los intervalos de tiempos en los que estos eventos ocurrieron. Para n potenciales de acción, dicho conjunto de tiempos se denota como $\{t_i\}$ con $i = 1, \dots, n$ y el ensayo durante el cual los disparos son medidos se considera de una duración T , tal que $\sum_{i=1}^n t_i = T$. De

esta manera, la secuencia puede ser representada como la suma de deltas de Dirac δ ,

$$\varrho(t) = \sum_{i=1}^n \delta(t - t_i), \quad (2.1)$$

donde $\varrho(t)$ es llamada la función de respuesta neuronal. Trivialmente, se tiene que

$$\int_0^T \varrho(t) dt = n. \quad (2.2)$$

A razón de que la secuencia de potenciales de acción generados por un estímulo dado varía de ensayo a ensayo, las respuestas neuronales son generalmente tratadas estadística o probabilísticamente. En este trabajo, en general las caracterizaremos por las tasas de disparo, en vez de por secuencias de disparo particulares. Más específicamente, están representadas por la tasa de conteo de disparos r , que se obtiene sumando el número de potenciales de acción del ensayo y dividiéndolo por la duración del mismo. Es decir,

$$r = \frac{n}{T} = \frac{1}{T} \int_0^T \varrho(t) dt. \quad (2.3)$$

Esta cantidad puede ser interpretada además como el promedio temporal de la función de respuesta neuronal sobre la duración del ensayo. La tasa de conteo de disparos puede ser determinada de un único ensayo experimental, pero a expensas de la pérdida de toda resolución temporal sobre variaciones de la respuesta durante el curso de la misma. Para superar este problema, se define la tasa de disparo dependiente del tiempo, mediante el conteo de disparos en ventanas temporales Δt pequeñas. Sin embargo, el resultado de cada medida será binario (dependiendo si hubo o no un disparo), dando sólo dos posibles valores. Por esta razón, se define la tasa de disparo dependiente del tiempo promediada sobre múltiples ensayos con el mismo estímulo, tal que,

$$r(t; s) = \frac{1}{\Delta t} \int_t^{t+\Delta t} \langle \varrho(\tau) \rangle_s d\tau, \quad (2.4)$$

donde s es el atributo que caracteriza al estímulo presentado. Pese a que las neuronas responden generalmente a diferentes propiedades de un estímulo, el atributo s es representativo debido a que se mantiene constante durante el ensayo. Para Δt suficientemente pequeños, $r(t)\Delta t$ es el número promedio de potenciales de acción

que ocurrieron en el intervalo $[t, t + \Delta t]$ sobre la cantidad total de ensayos.

De la misma manera que la función de respuesta neuronal $\varrho(t)$ pueda ser promediada en los ensayos para obtener la tasa de disparo $r(t)$, la tasa de conteo de disparos puede ser promediada sobre los ensayos, dando la tasa de disparo media \bar{r} . Por ende, esta cantidad está dada por

$$\bar{r} = \frac{\bar{n}}{T} = \frac{1}{T} \int_0^T \bar{\rho}(t) dt = \frac{1}{T} \int_0^T r(t) dt. \quad (2.5)$$

La tasa de disparo media expresada como una función del estímulo, es decir $\bar{r} = f_{\text{tuning}}(s)$, es llamada la función de ajuste de la respuesta neuronal, dependiendo del parámetro s de dicho estímulo. Estas funciones son importantes para caracterizar la selectividad de las neuronas a una variedad de parámetros del estímulo.

A partir de los datos disponibles obtenidos en un número finito de ensayos experimentales, se puede extraer la información de la tasa de disparo de la secuencia de potenciales de acción. Para ello es necesario primero discretizar el tren de disparos en una secuencia binaria, dividiendo el tiempo total en intervalos de tiempo pequeños, como se ilustra en la figura 1.1. La cantidad computada es la tasa de conteo de disparos sobre la duración del intervalo, ya que la tasa de disparo dentro de un dado intervalo es aproximada por la primera.

Debido a que el conteo de disparos puede tomar sólo valores enteros, las tasas computadas de esta forma serán siempre múltiplos enteros de $\frac{1}{\Delta t}$, y, por lo tanto, tomarán valores discretos. Decrecer el valor de Δt aumenta la resolución temporal de medida, ya que provee una estimación de la tasa de disparo a intervalos de tiempo más cercanos, pero a expensas de decrecer la resolución de discriminar distintas tasas. Además, tomar ventanas temporales pequeñas significa que, al traducir el tren de disparos a una secuencia binaria, se corre el riesgo de contar disparos cercanos como un solo disparo, representando erróneamente el código, además de quitar cualquier dependencia temporal de los parámetros.

2.2. Teoría de la información en neurociencia

La teoría de la información [6, 7] se construyó como un marco general para cuantificar la habilidad de un esquema de codificación o un canal de comunicación para transmitir información. Se asume que el código involucra un conjunto de símbolos, y que los procesos de codificación y transmisión son estocásticos y poseen ruido. Las cantidades que se utilizan para la cuantificación, que son la entropía y la información mutua, dependen de las probabilidades con las cuales estos símbolos, o combinaciones de los mismos, aparecen en el código. Los canales de comunicación, al ser ruidosos, tienen sólo capacidades limitadas para transmitir información. La entropía es una medida de la capacidad teórica de un código de transmitir información, mientras que la información mutua es la medida de cuánta de esa capacidad es realmente usada para describir un conjunto particular de datos. Las técnicas de la teoría de la información son utilizadas para evaluar estos límites y encontrar esquemas de codificación que los saturan.

En neurociencia, los símbolos considerados son las respuestas neuronales, y las características de los estímulos son los conjuntos de datos que describen. Idealmente, los símbolos analizados serían las secuencias de los potenciales de acción (es decir, una lista de los tiempos de disparo de los potenciales de acción). Sin embargo, computar la entropía y la información mutua se torna difícil debido a la gran cantidad de datos requeridos para la determinación de la frecuencia de ocurrencia de muchas secuencias de disparos distintas. Por esta razón, se utilizan descripciones simplificadas de la respuesta de una neurona que reducen el número posible de símbolos a ser considerados.

2.2.1. Entropía

La entropía de Shannon $H(X)$ [6] es una medida de la incerteza de una variable aleatoria. Sea X una variable discreta aleatoria con alfabeto χ y función de probabilidad $P(x) = Pr\{X = x\}$, $x \in \chi$, la entropía de Shannon de X es definida como

$$H(X) = - \sum_{x \in \chi} P(x) \log_2 P(x). \quad (2.6)$$

Se puede observar que la entropía es una función de la distribución de X , pero no depende en los valores de la variable aleatoria X en sí, sólo de sus probabilidades. Como el logaritmo de la ecuación (2.6) es en base dos, la entropía tiene unidades de bits. Si, en cambio, se utiliza el logaritmo natural ('ln') en la definición la unidad de la entropía es nats. Se puede realizar un cambio de unidades de la forma $H_b(X) = (\log_b a)H_a(X)$. La demostración de esta propiedad surge inmediatamente al tener en cuenta la propiedad de cambio de base del logaritmo $\log_b p = \log_b a \log_a p$.

La entropía es una cantidad siempre no negativa, es decir, $H(X) \geq 0$. Esto es debido a que, sabiendo que toda distribución de probabilidad toma valores entre cero y uno, $0 \leq P(x) \leq 1$, esto implica que $\log_2 \frac{1}{P(x)} \geq 0$. La entropía de X también puede ser interpretada como el valor de expectación de la "variable aleatoria" $\log_2 \frac{1}{P(x)}$. Por lo tanto,

$$H(X) = E_{P(x)} \left[\log_2 \frac{1}{P(x)} \right].$$

En el código neuronal, la entropía de las respuestas \mathbf{r} generadas por presentaciones repetidas de un mismo estímulo s es calculada utilizando la probabilidad condicional $P(r | s)$, de la forma

$$H_s(\mathbf{r}) = - \sum_{r \in \mathbf{r}} P(r | s) \log_2 P(r | s).$$

Si esta entropía es promediada sobre el conjunto de estímulos, se obtiene una cantidad llamada entropía de ruido

$$H_{\text{ruido}}(\mathbf{r}; \{s\}) = \langle H_s \rangle_s = - \sum_{s \in \{s\}} \sum_{r \in \{r\}} P(s)P(r | s) \log_2 P(r | s).$$

Es la entropía asociada a la parte de la variabilidad de la respuesta que no resulta de cambios en el estímulo.

Finalmente, la entropía total de un ensamble de neuronas es definida como

$$H(\mathbf{r}) = - \sum_{r \in \mathbf{r}} P(r) \log_2 P(r). \quad (2.7)$$

Cuando las respuestas están caracterizadas por la tasa de conteo de disparos, cuanti-

fica la impredecibilidad asociada con una respuesta en particular, promediada sobre todas las respuestas. Las respuestas con muy baja probabilidad contribuyen poco a la entropía total, a pesar de que el factor $-\log_2 P(r)$ es muy grande, debido a que ocurren raramente. Las respuestas con muy alta probabilidad tampoco contribuyen significativamente a la suma, ya que $-\log_2 P(r) \approx 0$. La entropía es entonces una medida de la variabilidad de la respuesta; sin embargo, no otorga información alguna sobre la fuente de dicha variabilidad, es decir, el estímulo.

Entropía conjunta y entropía condicional

Sea Y una variable discreta aleatoria con alfabeto ψ y función de probabilidad $P(y) = Pr\{Y = y\}$, $y \in \psi$, la entropía de Shannon conjunta $H(X, Y)$ del par de variables aleatorias (X, Y) con distribución de probabilidad conjunta $P(x, y)$ se define como

$$H(X, Y) = - \sum_{x \in \chi} \sum_{y \in \psi} P(x, y) \log_2 P(x, y), \quad (2.8)$$

que también puede ser expresada como $H(X, Y) = -E_{P(x,y)} [\log_2 P(X, Y)]$. Si X e Y son variables aleatorias estadísticamente independientes, se obtiene que $H(X, Y) = H(X) + H(Y)$.

La entropía condicional $H(Y | X)$ de un par de variables aleatorias $(X, Y) \sim P(x, y)$ es definida, teniendo en cuenta que $P(y | x) = \frac{P(x,y)}{P(x)}$, como

$$\begin{aligned} H(Y | X) &= \sum_{x \in \chi} P(x) H(Y | X = x) \\ &= - \sum_{x \in \chi} P(x) \sum_{y \in \psi} P(y | x) \log_2 P(y | x) \\ &= - \sum_{x \in \chi} \sum_{y \in \psi} P(x, y) \log_2 P(y | x) \\ &= -E_{P(x,y)} [\log_2 P(y | x)]. \end{aligned} \quad (2.9)$$

Si X e Y son variables aleatorias estadísticamente independientes, se puede ver que $H(Y | X) = H(Y)$.

A partir de estas dos definiciones surge una “regla de la cadena” para la entropía

de la forma

$$H(X, Y) = H(X) + H(Y | X) = H(Y) + H(X | Y). \quad (2.10)$$

Esta propiedad puede ser demostrada sencillamente utilizando las definiciones de las propiedades conjuntas y condicionales.

2.2.2. Divergencias entre dos distribuciones

Muchas medidas de “distancia” entre dos distribuciones de probabilidad en la teoría de la información han sido introducidas y extensamente estudiadas. Entre las medidas propuestas, la más comúnmente utilizada es la divergencia de Kullback-Leibler [35] o entropía relativa entre dos distribuciones de probabilidades $P(x)$ y $Q(x)$ de una variable X , que se define como

$$\begin{aligned} D[P \parallel Q] &= \sum_x P(x) \log_2 \frac{P(x)}{Q(x)} \\ &= H(P, Q) - H(P). \end{aligned} \quad (2.11)$$

Análogamente,

$$D[P \parallel Q] = E_{P(x)} \left[\log_2 \frac{P(x)}{Q(x)} \right]. \quad (2.12)$$

Provee una “cuasi-distancia” entre dos distribuciones de probabilidad, siendo $D[P \parallel Q] \geq 0$, con $D[P \parallel Q] = 0 \Leftrightarrow P(x) = Q(x)$, pero es no simétrica y no satisface la desigualdad triangular.

Por convención, basándose en argumentos de continuidad, se define que $\lim_{\varepsilon \rightarrow 0} \varepsilon \log_2 \frac{\varepsilon}{Q} = 0$ y $\lim_{\varepsilon \rightarrow 0} P \log_2 \frac{P}{\varepsilon} = \infty$. Por ende, si existe un símbolo $x \in \chi$ tal que $P(x) > 0$ y $Q(x) = 0$ entonces $D[P \parallel Q] = \infty$ [7].

Otra medida ampliamente utilizada, basada en la divergencia de Kullback-Leibler, es la divergencia de Jensen-Shannon. Esta se calcula como

$$\begin{aligned} J[P, Q] &= H[(P + Q)/2] - H[P]/2 - H[Q]/2 \\ &= D[P \parallel P_0]/2 + D[Q \parallel P_0]/2 \end{aligned} \quad (2.13)$$

en donde P_0 es una distribución de referencia. El nombre de esta cantidad se basa en la desigualdad de Jensen [7] y la definición de entropía de Shannon. En comparación a la divergencia de Kullback-Leibler, esta última tiene como propiedades ser simétrica y siempre ser de valor finito y no negativa.

2.2.3. Información mutua

La información mutua $I(X; Y)$ de un par de variables aleatorias (X, Y) con distribución de probabilidad conjunta $P(x, y)$ y probabilidades marginales $P(x)$ y $P(y)$, se define como la entropía relativa entre la distribución de probabilidad conjunta y el producto $P(x)P(y)$, es decir

$$I(X; Y) = D(P(x, y) \parallel P(x)P(y)) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log_2 \frac{P(x, y)}{P(x)P(y)}.$$

Análogamente, se puede obtener como

$$E_{P(x, y)} \left[\log_2 \frac{P(x, y)}{P(x)P(y)} \right].$$

Es una medida de la cantidad de información que una variable aleatoria contiene de otra variable aleatoria. Por ende, es la reducción de la incerteza de una variable aleatoria debido al conocimiento de la otra.

Se puede derivar una relación entre la información mutua y la entropía de la forma

$$I(X; Y) = H(X) - H(X | Y) = H(Y) - H(Y | X).$$

Como ilustrado en la figura 2.1, la simetría en la relación anterior significa que X dice tanto sobre Y como Y de X .

Finalmente, se puede ver que

$$I(X; X) = H(X) - H(X | X) = H(X).$$

Por lo tanto, la información mutua de una variable aleatoria con sí misma es la entropía de dicha variable aleatoria. Por esta razón, la entropía puede ser referida como la “autoinformación” de una variable aleatoria.

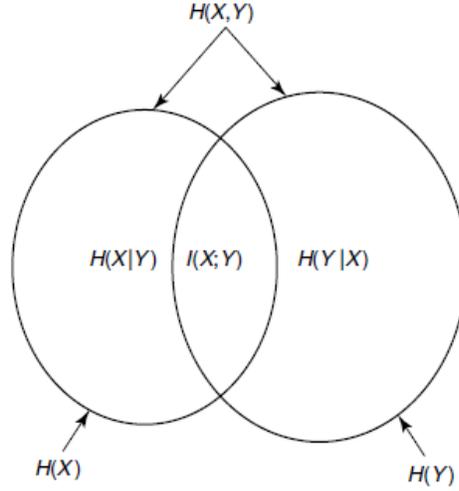


Figura 2.1: Relación entre las entropías y la información mutua (Ilustración extraída de [7]).

Para el código neuronal, debido a que

$$P(r) = \sum_{s \in \{s\}} P(s)P(r | s), \quad (2.14)$$

es posible reescribir la información mutua como

$$I(\mathbf{r}; \{s\}) = \sum_{s \in \{s\}} \sum_{r \in \mathbf{r}} P(s)P(r | s) \log_2 \left(\frac{P(r | s)}{P(r)} \right). \quad (2.15)$$

Otra forma de definir la información mutua se puede derivar utilizando la propiedad de la probabilidad conjunta

$$P(r; s) = P(s)P(r | s) = P(r)P(s | r), \quad (2.16)$$

obteniendo de esta manera

$$I(\mathbf{r}; \{s\}) = \sum_{s \in \{s\}} \sum_{r \in \{r\}} P(r; s) \log_2 \left(\frac{P(r; s)}{P(r)P(s)} \right). \quad (2.17)$$

Esta definición revela que la información mutua es simétrica con respecto al intercambio de \mathbf{r} y $\{s\}$. Esto significa que la información que transmite un conjunto de respuestas sobre un conjunto de estímulos es idéntica a la información que transmite un conjunto de estímulos sobre el conjunto de respuestas.

Además, a partir de la definición de la entropía de ruido, se puede obtener la información mutua de un ensamble de neuronas como

$$\begin{aligned} I(\mathbf{r}; \{s\}) &= H_s(\mathbf{r}) - H_{\text{ruido}}(\mathbf{r}; \{s\}) \\ &= - \sum_{r \in \mathbf{r}} P(r) \log_2 P(r) + \sum_{s \in \{s\}} \sum_{r \in \mathbf{r}} P(s)P(r | s) \log_2 P(r | s). \end{aligned} \quad (2.18)$$

Para transmitir información sobre un conjunto de estímulos, las respuestas neuronales deben ser distintas para diferentes estímulos. Una neurona puede proveer información sobre un estímulo sólo si la variabilidad de su respuesta está correlacionada con cambios en ese estímulo, en vez de ser puramente aleatoria. Las respuestas que son informativas sobre la identidad del estímulo deberían exhibir mayor variabilidad para ensayos involucrando distintos estímulos que para ensayos experimentales que usen el mismo estímulo en repetición. La información mutua es una medida dependiente de la entropía que se basa en esta idea, siendo la diferencia entre la entropía de respuesta total y la entropía de respuesta promediada sobre ensayos que involucren la presentación del mismo estímulo repetidas veces. Substraer de la entropía total la entropía cuando el estímulo no cambia, remueve de esta cantidad la contribución de la variabilidad de respuesta que no está asociada con la identidad del estímulo.

2.2.4. Descomposición de la información en mecanismos de codificación

Para una dada elección de código, siguiendo [6], podemos escribir la información mutua transmitida por la respuesta poblacional sobre el conjunto completo de estímulos $\{s\}$ como

$$I(\mathbf{r}; \{s\}) = \left\langle \sum_{\mathbf{r} \in \{r\}} P(\mathbf{r}|s) \log_2 \left[\frac{P(\mathbf{r}|s)}{P(\mathbf{r})} \right] \right\rangle_s, \quad (2.19)$$

donde \mathbf{r} denota el espacio de respuestas. Los paréntesis angulares indican el promedio sobre distintos estímulos, $\langle A(s) \rangle_s \equiv \sum_{s \in \{s\}} P(s)A(s)$, siendo $P(s)$ la probabilidad de que el estímulo s esté presente. $P(\mathbf{r}|s)$ es la probabilidad de observar una respuesta en particular \mathbf{r} condicionada al estímulo s , y $P(\mathbf{r}) = \langle P(\mathbf{r}|s) \rangle_s$ es su promedio a lo largo de todas las presentaciones de estímulo. La probabilidad $P(\mathbf{r}|s)$ puede ser determinada experimentalmente repitiendo cada estímulo de exactamente la misma forma a lo largo de varias pruebas experimentales, mientras se registran las respuestas neuronales. La información mutua cuantifica cuán bien un observador ideal de dichas respuestas puede discriminar entre todos los posibles estímulos diferentes, basándose en una única prueba experimental. En la ecuación (2.19) la suma se realiza sobre todas las posibles respuestas poblacionales. Esta ecuación cuantifica la información total transmitida por la actividad de la población de células. Sin embargo, no nos dice nada sobre las contribuciones específicas de las correlaciones cruzadas entre neuronas a la información total transmitida o si las correlaciones cruzadas hacen el código redundante (contribuciones negativas a la información) o sinérgico (contribuciones positivas a la información).

Por ende, para describir exactamente el impacto de la correlación en la información, es necesario definir una medida de correlación relacionada con la respuesta. Esta medida se denota usualmente como $\gamma(\mathbf{r}|s)$, y cuantifica cuánto más alta es la probabilidad que las neuronas emitan una respuesta en comparación al caso no correlacionado, normalizada a la probabilidad de un evento con tasas de disparo esperadas en el caso no-correlacionado [9, 11, 12]. Es nombrada como la “correlación de ruido” en la literatura previa [36], dado que mide correlaciones en la variabilidad de la respuesta dadas repeticiones de pruebas experimentales con el mismo estímulo presente. Valores positivos de esta cantidad indican correlación y valores negativos indican anticorrelación.

Una definición matemática natural [12] para la correlación cruzada normalizada de la respuesta poblacional \mathbf{r} (la correlación de ruido) es la siguiente:

$$\gamma(\mathbf{r}|s) = \begin{cases} \frac{P(\mathbf{r}|s)}{P_{ind}(\mathbf{r}|s)} - 1 & \text{si } P_{ind}(\mathbf{r}|s) \neq 0, \\ 0 & \text{si } P_{ind}(\mathbf{r}|s) = 0. \end{cases} \quad (2.20)$$

La probabilidad de que las respuestas de la población sean independientes $P_{ind}(\cdot)$

puede ser obtenida tomando el producto de las distribuciones de probabilidad de las neuronas individuales:

$$P_{ind}(\mathbf{r}|s) = \prod_{i=1}^N P(r_i|s). \quad (2.21)$$

Ya que $\gamma(\mathbf{r}|s)$ es una cantidad dependiente de la respuesta, tiene que satisfacer (para cada estímulo s) un vínculo de suma cero:

$$\sum_{\mathbf{r}} P_{ind}(\mathbf{r}|s) \gamma(\mathbf{r}|s) = 0. \quad (2.22)$$

Este vínculo surge directamente de la normalización a la unidad de las probabilidades de respuesta. Nombramos $P_{ind}(\mathbf{r}) = \langle P_{ind}(\mathbf{r}|s) \rangle_s$.

Otro coeficiente importante para describir la codificación poblacional que cuantifica la similaridad en la modulación del estímulo en las respuestas de neuronas individuales es el coeficiente de “similaridad de señal” o de “correlación de señal” $\nu(\mathbf{r})$ [36]. Es decir, otorga una idea de las correlaciones en las medias de las respuestas de las neuronas a lo largo de todo el conjunto de estímulos. Es distinto de cero únicamente si las señales que provienen de las neuronas individuales están correlacionadas positivamente o negativamente. Tanto $\nu(\mathbf{r})$ como $\gamma(\mathbf{r}|s)$ pueden variar entre -1 y ∞ , con 0 indicando falta de correlación.

El coeficiente de similaridad en la señal (correlación de señal) se define como:

$$\nu(\mathbf{r}) = \begin{cases} \frac{P_{ind}(\mathbf{r})}{\prod_{i=1}^N P(r_i)} - 1 & \text{si } \prod_{i=1}^N P(r_i) \neq 0, \\ 0 & \text{si } \prod_{i=1}^N P(r_i) = 0. \end{cases} \quad (2.23)$$

Depende únicamente de las probabilidades de respuesta de las neuronas individuales y, por ende, satisface una condición de normalización similar a la de la ecuación (2.22), dada por la expresión siguiente:

$$\sum_{\mathbf{r}} \left(\prod_{i=1}^N P(r_i) \right) \nu(\mathbf{r}) = 0. \quad (2.24)$$

Como una proporción de las probabilidades completas y las independientes, transmiten cuán probable sería encontrar eventos correlacionados en el conjunto de neuronas (en comparación a eventos independientes). Aunque definidas de manera similar, la diferencia principal entre estas dos medidas es la modulación por parte del estímulo. Además, como principio general, si las correlaciones de ruido y de señal poseen el mismo signo la información se degrada; por el contrario, si tienen signos opuestos, la última es realizada [10].

Al considerar estas cantidades es posible escribir la información total en componentes, cada una reflejando la contribución de un mecanismo de codificación distinto, como en [9, 12]:

$$I(\mathbf{r}; \{s\}) = I_{lin} + I_{sig-sim} + I_{cor-ind} + I_{cor-dep}. \quad (2.25)$$

Las componentes de la descomposición exacta de la información son dadas a continuación, como en [12, 13]. La actividad de C células es observada y cada respuesta poblacional posible \mathbf{r} se compone por las respuestas de las células individuales r_1, \dots, r_C . La información total que fuera a ser transmitida si todas las células fuesen independientes, I_{lin} , resulta:

$$I_{lin} = \frac{1}{\ln 2} \sum_c \sum_{r_c} \langle P(r_c|s) \ln \frac{P(r_c|s)}{P(r_c)} \rangle_s. \quad (2.26)$$

Aquí las sumas son realizadas para cada célula c y cada respuesta posible de dicha neurona, dado que ninguna suposición adicional fue hecha sobre la probabilidad de respuesta. El término de similaridad en la señal, $I_{sig-sim}$, se encuentra dado por:

$$I_{sig-sim} = \frac{1}{\ln 2} \sum_{\mathbf{r}} \left[\prod_c P(r_c) \right] \left\{ \nu(\mathbf{r}) + [1 + \nu(\mathbf{r})] \ln \left[\frac{1}{1 + \nu(\mathbf{r})} \right] \right\}. \quad (2.27)$$

La cantidad total de información atribuible a la actividad correlacionada sobre el código neuronal (I_{cor}) se define como $I_{cor} = I_{cor-ind} + I_{cor-dep}$. La componente independiente del estímulo, $I_{cor-ind}$, se calcula como:

$$I_{cor-ind} = \sum_{\mathbf{r}} \langle P_{ind}(\mathbf{r}|s) \gamma(\mathbf{r}|s) \rangle_s \log_2 \left[\frac{1}{1 + \nu(\mathbf{r})} \right]. \quad (2.28)$$

Por otro lado la componente de correlación que depende del estímulo, $I_{cor-dep}$, se calcula como:

$$I_{cor-dep} = \sum_{\mathbf{r}} \langle P_{ind}(\mathbf{r}|s) [1 + \gamma(\mathbf{r}|s)] \log_2 \left\{ \frac{\langle P_{ind}(\mathbf{r}|s') \rangle_{s'} [1 + \gamma(\mathbf{r}|s)]}{\langle P_{ind}(\mathbf{r}|s') \rangle_{s'} [1 + \gamma(\mathbf{r}|s')] } \right\} \rangle_s. \quad (2.29)$$

Por medio de esta descomposición, es posible separar las contribuciones aditivas a la información de las contribuciones de neuronas como unidades independientes. Contribuciones de mecanismos de codificación positivas, negativas o nulas surgirían al considerar a las neuronas como unidades dentro de una población. La sinergia (adición positiva) sería entonces evidencia de un código poblacional entre las células interactuantes y la redundancia (contribución negativa) resultaría en valores menores de información que la resultante por combinación aditiva [37, 38].

El primer término de la descomposición de la información I_{lin} es la información obtenida si cada célula fuese a transmitir información independientemente, es decir, no se considera sinergia o redundancia. La información total transmitida por la población es entonces la suma lineal de la información expresada por cada neurona individual [39]. El segundo término, $I_{sig-sim}$, toma en cuenta la redundancia que podría llegar a surgir incluso en la ausencia de correlaciones cruzadas, cuando existen similitudes en la distribución a lo largo de los estímulos en las probabilidades de respuesta estímulo-dependientes de las células individuales. Este término es siempre menor o igual a cero, entonces no puede contribuir sinérgicamente.

Si las respuestas poblacionales son estadísticamente independientes, estas son las únicas contribuciones no nulas a la descomposición de la información. En este sentido, la suma de I_{lin} y $I_{sig-sim}$ cuantifica cuánta de esta información puede ser obtenida de las neuronas evaluadas individualmente, sin ninguna referencia a la actividad simultánea de las mismas con otras neuronas.

Los próximos dos términos en la descomposición de la información son componentes dependientes de las correlaciones: una componente estímulo-dependiente $I_{cor-dep}$ y una componente independiente del estímulo $I_{cor-ind}$. Por ende, expresan cualquier otro efecto que las correlaciones cruzadas puedan llegar a tener más allá de aquellos tenidos en cuenta por las propiedades de las neuronas individuales. En torno a eso,

la suma de ambos términos correlacionales $I_{cor-ind} + I_{cor-dep}$ cuantifica la cantidad de información verdaderamente disponible de la actividad correlacionada de toda la población [11].

Puede ser probado mediante desigualdades básicas de la teoría de la información que la componente correlacional estímulo-dependiente ($I_{cor-dep}$) es no-negativa, y que es cero si y sólo si, para cualquier respuesta \mathbf{r} dada, la fuerza de correlación $\gamma(\mathbf{r}|s)$ es independiente del estímulo. Por lo tanto, este término mide cuán bien la identidad del estímulo se encuentra “etiquetada” por las diferencias en las correlaciones de espiga de una prueba experimental a otra a lo largo del conjunto de estímulos [12]. $I_{cor-dep}$ es no-nulo sólo si la correlación de ruido $\gamma(\mathbf{r}|s)$ es distinta a cero para alguna respuesta \mathbf{r} o estímulo s .

Para una dada respuesta \mathbf{r} , la componente de correlación estímulo-independiente ($I_{cor-ind}$) es positiva (sinérgica) cuando la similaridad en la señal y la correlación cruzada tienen signos opuestos, y negativa (redundante) en caso contrario. Esto es, incluso cuando no se encuentran moduladas por el estímulo, las correlaciones cruzadas pueden afectar todavía el código neural a través de la interacción entre la correlación entre células y las similaridades en la señal [40, 41]. Además, en contraposición a la componente de correlación dependiente del estímulo $I_{cor-dep}$, $I_{cor-ind}$ no desaparece si $\gamma(\mathbf{r}|s)$ o $\nu(\mathbf{r})$ no son iguales a cero.

2.3. Métodos de la geometría de la información

La geometría de la información [26] estudia una familia de distribuciones de probabilidad mediante el uso de la geometría diferencial. Por lo tanto, provee un nuevo marco teórico para el análisis matemático y relevante para una amplia gama de dominios. A su vez, aporta una nueva perspectiva para ver la estructura de sistemas de interés, como el conjunto de disparos en una población de neuronas en nuestro caso. La geometría de la información permite describir modelos estocásticos de trenes múltiples de espigas mediante una familia de distribuciones de probabilidad, proveyendo no sólo un entendimiento intuitivo del objeto de estudio pero también de herramientas útiles para el análisis de diversos datos neuronales. Estudiando la estructura geométrica de una familia de probabilidades en el espacio curvo (variedad

de Riemann), esta descripción proporciona diversas relaciones entre distribuciones de probabilidad tal como la divergencia o discrepancia entre dos distribuciones de probabilidad, y la ortogonalidad de dos distribuciones. En este sentido, la propiedad más atractiva de la geometría de la información es que permite descomponer la distribución conjunta de probabilidades de múltiples variables estocásticas en estructuras jerárquicas de forma ortogonal. En esta sección, presentaré el marco matemático para el análisis de la actividad de una población de neuronas desde el punto de vista de esta teoría, introduciendo aspectos fundamentales para la comprensión del mismo.

2.3.1. Conceptos básicos de geometría diferencial

El objetivo principal de la geometría diferencial moderna es caracterizar las propiedades globales de las variedades [26]. Para la geometría de la información, los aspectos más importantes de la geometría diferencial son aquellos que permiten tomar problemas de la teoría de información y visualizarlos geoméricamente. Actualmente, en el campo de la geometría de la información, caracterizar las propiedades locales de las variedades, en general, es un requerimiento suficiente.

Una variedad es un concepto matemático que denota una abstracción de objetos geométricos, como ser curvas y superficies, en un espacio de cierta dimensión. Una variedad S es un conjunto con un sistema de coordenadas y , y por lo tanto, al ser un conjunto posee elementos, también llamados puntos de S . En este contexto, los puntos de S son distribuciones de probabilidad de variables aleatorias. El sistema de coordenadas de S debe permitir un mapeo uno a uno de S (o uno de sus subconjuntos) a \mathbb{R}^n , tal que se pueda especificar cada punto de S usando un vector de n números reales. De esta manera, estos vectores serán las coordenadas de los puntos de S , y n será la dimensión de S . Una subvariedad M de S , es también un subconjunto de los elementos de S , y mantiene todas las propiedades locales de la variedad.

Sea S una variedad y sea $\varphi : S \rightarrow \mathbb{R}^n$ el sistema de coordenadas de S (ver figura 2.2). Entonces φ mapea cada punto p de S a n números reales: $\varphi(p) = [\xi^1(p), \dots, \xi^n(p)] = [\xi^1, \dots, \xi^n]$, y estas son las coordenadas del punto p . Cada ξ^i puede ser visto como una función $p \rightarrow \xi^i(p)$ que mapea cada punto a su i -ésima coordenada; estas funciones $\xi^i : S \rightarrow \mathbb{R}$, con $i = 1, \dots, n$, se denominan funciones coordenadas.

Sea $\zeta = [\rho^i]$ otro sistema de coordenadas para S . Las coordenadas de $[\rho^i]$ pueden

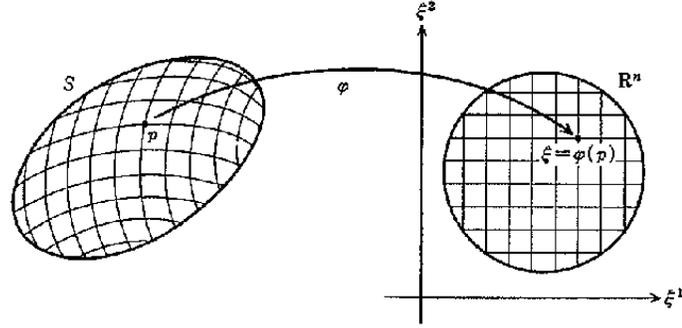


Figura 2.2: Ejemplo de un sistema de coordenadas de S (Ilustración extraída de [42]).

ser obtenidas a partir de $[\xi^i]$ mediante la transformación de coordenadas en \mathbb{R}^n dada por $\zeta \circ \varphi^{-1} : [\xi^i] \rightarrow [\rho^i]$. Se debe tener en cuenta que al cambiar de coordenadas pueden cambiar las propiedades de la variedad en cuestión, afectando el estudio de la misma. Por lo tanto, se restringirán los sistemas de coordenadas a aquellos que permitan transformaciones de coordenadas suaves entre ellos.

Cuando los elementos de una variedad S son distribuciones de probabilidad, tal que $S = \{P_\xi \mid \xi \in \mathbb{R}^n\}$, se dice que S es un modelo estadístico n -dimensional. Dado un punto ξ , la matriz de información de Fisher de S en ξ es la matriz $n \times n$ $C(\xi) = [g_{ij}(\xi)]$, donde el (i, j) -ésimo elemento está dado por:

$$g_{ij}(\xi) = E_\xi \left[\frac{\partial}{\partial \xi_i} l_\xi \frac{\partial}{\partial \xi_j} l_\xi \right], \quad (2.30)$$

donde $l_\xi = \ln P_\xi$. En particular, si $n = 1$ se denomina Información de Fisher. g_{ij} resulta la base natural del sistema de coordenadas $[\xi^i]$, y determina una métrica Riemanniana, llamada métrica de Fisher o métrica de información. Esta métrica es invariante sobre los cambios de coordenadas, y provee el medio para calcular las direcciones de cambio de la variedad con respecto a sus coordenadas.

La distancia entre dos distribuciones P_ξ y $P_{\xi+d\xi}$ en un espacio de Riemman curvo se encuentra dado por la forma cuadrática llamada distancia geodésica

$$ds^2 = \sum_1^n g_{ij} d\xi_i d\xi_j. \quad (2.31)$$

Es posible demostrar que esta distancia al cuadrado tiene una relación cercana con la divergencia de Kullback-Leibler (ver ecuación (2.12)) de la forma

$$ds^2 \equiv 2 D[P(x; \xi) \parallel P(x; \xi + d\xi)]. \quad (2.32)$$

En este sentido, se pueden definir dos direcciones como ortogonales cuando sus correspondientes variables aleatorias no se encuentran correlacionadas. Es decir, cuando las componentes de la matriz de información de Fisher se anulan.

2.3.2. Variedades de familias exponenciales

Denominamos a una variedad como e -plana (exponencialmente plana) cuando existe un sistema de coordenadas parametrizado por θ tal que para todo i, j y k

$$E \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_{\theta} \frac{\partial}{\partial \theta_k} l_{\theta} \right] = 0, \quad (2.33)$$

Tales parámetros son llamados coordenadas e -afines. Una variedad e -plana típica es la ampliamente conocida familia de distribuciones exponenciales

$$P(\mathbf{x}, \boldsymbol{\theta}) = \exp \left\{ \sum_i \theta_i k_i(\mathbf{x}) - \psi(\boldsymbol{\theta}) \right\} \quad (2.34)$$

donde $k_i(\mathbf{x})$ representa la función dada y ψ es un factor de normalización. Un ejemplo típico de la familia exponencial es la distribución gaussiana que, en la representación de la ecuación (2.34), posee los parámetros

$$\begin{aligned} k(x) &= \{x, x^2\} \\ \boldsymbol{\theta} &= \left\{ \frac{\mu}{\sigma^2}, -\frac{1}{\sigma^2} \right\} \\ \psi(\boldsymbol{\theta}) &= \frac{\mu}{\sigma^2} + \frac{\ln 2\pi}{2} + \ln \theta \end{aligned}$$

donde μ y σ corresponden a la media y la varianza de la distribución respectivamente. En las coordenadas e -planas, la relación de ortogonalidad se mantiene debido a que

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j} \ln P = -\frac{\partial^2}{\partial\theta_i\partial\theta_j} \psi(\theta) \quad (2.35)$$

es independiente de x y $E[\frac{\partial}{\partial\theta} \ln P] = 0$.

De la misma manera, se dice que una variedad es m -plana (plana mezcla) cuando existe un sistema de coordenadas $\boldsymbol{\eta}$ tal que

$$E \left[\frac{1}{P(x, \boldsymbol{\eta})} \frac{\partial^2}{\partial\eta_i\partial\eta_j} l_{\boldsymbol{\eta}} \frac{\partial}{\partial\eta_k} l_{\boldsymbol{\eta}} \right] = 0. \quad (2.36)$$

$\boldsymbol{\eta}$ es llamado coordenadas m -afines.

En la teoría de la geometría de la información, se ha demostrado que una familia exponencial es automáticamente m -plana aunque no sea necesariamente una familia mezcla [26]. Las coordenadas m -afines ($\boldsymbol{\eta}$) de una familia exponencial se encuentran definidas por

$$\eta_i = E[k_i(x)] = \frac{\partial}{\partial\eta_i} \psi(\boldsymbol{\theta}), \quad (2.37)$$

también conocidos como parámetros de esperanza. La transformación de coordenadas entre las dos coordenadas $\boldsymbol{\eta}$ y $\boldsymbol{\theta}$ están dadas por la siguiente regla

$$\theta_i = \frac{\partial\phi(\boldsymbol{\eta})}{\partial\eta_i} \quad (2.38)$$

considerando que $\phi(\boldsymbol{\eta}) = E[l_{\boldsymbol{\eta}}]$. Los vectores tangentes a las curvas de coordenadas η_i y θ_i son ortogonales en cualquier punto. Dado el interés en las distribuciones generadas en el marco de la mecánica estadística no extensiva [43], sería posible además realizar una descripción geométrica de la familia de distribuciones q -exponenciales de forma análoga a lo desarrollado previamente para la familia de distribuciones exponenciales [44]. Sin embargo, este desarrollo excede al alcance de esta Tesis y no es utilizado para el desarrollo de los trabajos presentados en la misma.

La ecuación (2.34) constituye la representación central de los modelos de máxima entropía. En los capítulos siguientes, expondremos resultados obtenidos en este marco como referencia.

Capítulo 3

Contribuciones de alto orden al código neuronal

Experimentos contemporáneos en neurociencia han desafiado el interrogante si la información transmitida por la actividad de un conjunto de neuronas se encuentra únicamente determinada por el número de potenciales de acción disparados por cada célula independientemente o si las correlaciones presentes en las secuencias de espigas además juegan un papel en la codificación [8, 10, 11, 13, 45–52]. Además, diversos estudios han cuestionado la relevancia de las correlaciones de alto orden en secuencias de disparos binarizados [14, 24, 34, 53, 54], dado que se ha demostrado que incluso correlaciones entre tripletes impactan profundamente el procesamiento de la información en poblaciones pequeñas de neuronas [28].

Si la información de distintas células fuese independiente, consecuentemente incrementaría linealmente con el número de neuronas de la población. Si ese no es el caso, entonces dos regímenes son posibles. Por un lado, si muchas células en la muestra transmiten información similar, entonces el código se denomina redundante ya que la información de toda la población sería sobreestimada si fuese considerada independiente [5]. Por otro lado, parte de la información surge solamente al tomar en cuenta interacciones (o respuestas simultáneas) de neuronas diferentes. Por ejemplo, información sólo disponible de los tiempos relativos de las espigas. Este código se dice sinérgico, y una mayor información se transmite en una población que aquella obtenida por la simple suma de la información de cada neurona independientemente.

Resulta importante entonces conseguir un mejor entendimiento sobre las correlaciones entre las células en el cerebro y su impacto en la codificación poblacional. Esto es, las modalidades con las cuales las correlaciones contribuyen a la transmisión de la información, en vez de la simple cuantificación de la información total transmitida por el conjunto [12]. Una medida válida para evaluar el rol de las correlaciones en el código neuronal es verificar si la información es codificada sinérgica, redundante o independientemente [13, 36, 37, 39, 55–63].

3.1. Expansión de la información para ventanas temporales cortas

Consideramos un periodo de tiempo de corta duración durante el cual la actividad de N células es observada. La respuesta poblacional de dichas neuronas al estímulo presentado, en la ventana temporal post-estímulo, se denota como el vector \mathbf{r} , con cada elemento r_1, \dots, r_N del vector describiendo la respuesta individual de cada neurona. Cada estímulo diferente se denota como s , y los estímulos considerados son puramente abstractos; el formalismo detallado en este capítulo se puede aplicar a una amplia variedad de paradigmas experimentales. La respuesta de cada célula puede ser entonces descripta de diversas maneras, dependiendo de las preguntas a ser respondidas. Por ejemplo, en un código de conteo de espigas, r_i sería simplemente el conteo de disparos de la neurona i medidos en la ventana de tiempo post-estímulo en una dada prueba experimental con el estímulo s presente. En este capítulo, describiremos la respuesta por la tasa de disparos de cada neurona, es decir el conteo de espigas dividido por la duración de la ventana post-estímulo, y la llamaremos $r_i(s)$ para establecer explícitamente su dependencia con el estímulo en cuestión. Por otro lado, si estuviésemos interesados en un código temporal de espigas, la respuesta r_i sería una secuencia de tiempos de cada disparo $\{t_j^i\}$, donde t_j^i denota el tiempo del disparo j -ésimo de la i -ésima neurona para una dada prueba experimental, como usado en [11]. Todas las definiciones y las ecuaciones derivadas a continuación son válidas para cualquier elección de código neuronal \mathbf{r} , incluyendo códigos temporales de disparos con la notación apropiada.

Desde un punto de vista probabilístico, es posible interpretar a la cantidad $r(t)\Delta t$

como la probabilidad de ocurrencia de un potencial de acción durante un intervalo de corta duración Δt alrededor de un tiempo t . Por ende, la respuesta neuronal y su relación con el estímulo, están completamente caracterizadas por la distribución de probabilidad de los tiempos de disparo de los potenciales de acción como función del estímulo. La tasa de disparo de la neurona i , cuando el estímulo s es presentado, es

$$r_i(t_i^j; s) = \sum_j \frac{\delta_{t, t_i^j}}{\Delta t},$$

y la probabilidad de observar un disparo de la misma neurona en el intervalo de tiempo de duración Δt centrado en t_i^a , cuando el estímulo s está presente, promediada sobre todos los posibles patrones de potenciales de acción que ocurren en otros intervalos de tiempo es

$$P(t_i^j; s) = r_i(t_i^j; s)\Delta t.$$

Se opta por estudiar el código de conteo de disparos, y por ende estas cantidades son calculadas en toda la ventana temporal y no para cada intervalo de tiempo. De esta manera, la notación para la probabilidad individual de disparo utilizada será

$$P(r_i; s) = r_i(s)\Delta t.$$

La probabilidad del conjunto de respuestas neuronales en adelante será denotada como $P(r)$, la probabilidad conjunta de una respuesta evocada y la aparición de un estímulo como $P(r; s)$ y la probabilidad condicional de una respuesta a causa de un estímulo particular como $P(r | s)$.

Sin embargo, la tasa de disparos individual no otorga ninguna información con respecto a las interacciones entre neuronas, factor central en el estudio de la importancia de las mismas en la codificación neuronal. Por ende, es necesario introducir parámetros que describan las correlaciones entre disparos, es decir, que tengan en cuenta la relación entre patrones de espigas de distintas neuronas. Existen dos tipos de correlaciones que influyen sobre la información del conjunto, las correlaciones de ruido y las correlaciones de señal [36] (descriptas previamente en el capítulo 2). Las correlaciones de ruido representan la variabilidad de las respuestas con respecto al promedio para cada estímulo. Las correlaciones de señal, en cambio, son indepen-

dientes de cada estímulo en sí, dado que representan la variabilidad de las respuestas promediadas en el conjunto total de estímulos considerados. Las definiciones de estas cantidades para órdenes particulares (y no en general) se explicitan a continuación.

Se definen las correlaciones de ruido segundo orden como

$$\gamma_{ij}(s) = \frac{\overline{r_i(s)r_j(s)}}{\bar{r}_i(s)\bar{r}_j(s)} - 1 \quad (3.1)$$

y de tercer orden de la forma

$$\gamma_{ijk}(s) = \frac{\overline{r_i(s)r_j(s)r_k(s)}}{\bar{r}_i(s)\bar{r}_j(s)\bar{r}_k(s)} - 1. \quad (3.2)$$

Utilizamos la notación con un delineado superior para denotar cantidades promediadas sobre repeticiones, y para distinguirlos de promedios sobre los estímulos. Tanto γ_{ij} como γ_{ijk} varían de -1 a ∞ , con 0 indicando falta de correlación. Estas correlaciones son importantes para definir las probabilidades condicionales entre disparos de neuronas.

Las correlaciones de señal de segundo y tercer orden están dadas por, respectivamente,

$$\nu_{ij} = \frac{\langle \bar{r}_i(s)\bar{r}_j(s) \rangle_s}{\langle \bar{r}_i(s) \rangle_s \langle \bar{r}_j(s) \rangle_s} - 1, \quad (3.3)$$

y

$$\nu_{ijk} = \frac{\langle \bar{r}_i(s)\bar{r}_j(s)\bar{r}_k(s) \rangle_s}{\langle \bar{r}_i(s) \rangle_s \langle \bar{r}_j(s) \rangle_s \langle \bar{r}_k(s) \rangle_s} - 1. \quad (3.4)$$

Al igual que las correlaciones de ruido, tanto ν_{ij} como ν_{ijk} varían de -1 a ∞ , en donde 0 indica la falta de correlación.

Para evaluar cómo interacciones de órdenes específicos afectan la codificación de la información total de una población neuronal, es necesario depender del método de la expansión en series de la información mutua, introducido en [9–11, 64]. Permite cuantificar separadamente la información disponible si cada neurona fuese a transmitir información independientemente de las desviaciones que surgen de efectos de sinergia y redundancia, considerando hasta correlaciones de pares entre neuronas. Estos efectos no lineales resultan de diferentes contribuciones que pueden ser agrupadas como contribuciones redundantes debido a similitudes en los perfiles de respues-

ta promediados de distintas neuronas, un término de correlaciones independiente del estímulo que puede ser tanto sinérgico como redundante, o una contribución de correlaciones dependiente del estímulo sinérgica [13, 37, 39], como descrito en el capítulo 2. Es importante notar que estas componentes tienen el mismo significado tanto en la descomposición exacta de la información como en la aproximación de expansión en series al considerar contribuciones de distintos órdenes.

El método de la expansión en serie está basado en la suposición de que la información es transmitida en ventanas temporales post-estímulo que son cortas comparadas con los intervalos inter-espiga típicos, tal que el número promedio de espigas emitidas en cada presentación de estímulo sea bajo. En este capítulo, presento una extensión de esta aproximación hasta contribuciones de tercer orden y se propone una dependencia general para cualquier orden. Se obtiene una expresión analítica exacta que nos permite estimar una corrección para la componente de correlación estímulo dependiente de la información mutua dentro del marco de la expansión en series.

Cuando correlaciones de trenes de espigas son considerados en un sistema de tres neuronas por ejemplo, existe una limitación al analizar únicamente interacciones de pares. Dos tipos distintos de estímulos externos pueden generar correlaciones de pares o de tripletes, con análisis de autocorrelación y correlación cruzada idénticos en ambos casos [65]. En otras palabras, la conectividad de pares en la entrada de una red podría generar correlaciones de alto orden que fuesen indistinguibles de aquellas generadas por conectividad de tripletes en la entrada de la red. No sería posible diferenciar, en dicho escenario, las dos actividades de la población neuronal incluso mediante un análisis de autocorrelación y correlación cruzada, y las diferencias entre estas dos configuraciones no son captadas por meras correlaciones entre pares. Son las correlaciones de alto orden que determinan si los disparos coincidentes entre dos neuronas son también coincidentes con las espigas de una tercer neurona. Las correlaciones de pares por lo tanto no son suficientes para proveer descripciones confiables de los sistemas neuronales en general [24, 34].

Primero, se asume que las probabilidades de disparo individuales son proporcionales a la duración temporal t de la manera

$$P(r_i | r_j, r_k \cdots, r_n; s) \sim t, \quad (3.5)$$

y usando la “regla de la cadena”

$$P(r_i, \dots, r_n; s) = P(r_i | r_j, \dots, r_n; s) P(r_j | r_k, \dots, r_n; s) P(r_k | \dots, r_n; s) \dots P(r_n; s), \quad (3.6)$$

la probabilidad de que n espigas hayan sido disparadas es el producto de las probabilidades condicionales de cada disparo dada la presencia de otros potenciales de acción en el patrón. Entonces, como cada término es proporcional a t , la probabilidad de tener n espigas resulta proporcional de la forma

$$P(r_i, r_j, r_k, \dots, r_n; s) \sim t^n. \quad (3.7)$$

Esta es una suposición natural, debido a que sólo implica que la probabilidad de observar un disparo en una ventana temporal es proporcional a la resolución temporal de medida. Únicamente es violada en el caso de que dos disparos ocurran en el mismo momento con una precisión temporal infinita, caso poco probable.

Como una extensión propuesta de la expansión en series utilizada en las referencias [9–11], podemos escribir la probabilidad de tener tres neuronas disparando (rotuladas como i , j y k) como

$$P(r_i, r_j, r_k; s) = \bar{r}_i(s) \bar{r}_j(s) \bar{r}_k(s) t^3 [1 + \gamma_{ijk}(s)] + \mathcal{O}(t^4). \quad (3.8)$$

Las ecuaciones (3.5-3.7) son la base principal del método de la expansión en series. Por ejemplo, en el marco teórico de las interacciones de a pares, la probabilidad de que no se haya disparado ninguna espiga en un cierto tiempo t es

$$P(0; s) = 1 - \sum_{i=1}^N \bar{r}_i(s) t + \frac{1}{2} \sum_{i,j=1}^N \bar{r}_i(s) \bar{r}_j(s) [1 + \gamma_{ij}(s)] t^2 + \mathcal{O}(t^3), \quad (3.9)$$

y la probabilidad de observar un único potencial de acción disparado por la neurona i , $P(r_i; s)$, se puede obtener como

$$P(r_i; s) = \bar{r}_i(s) t + \bar{r}_i(s) t^2 \sum_{j=1}^N \bar{r}_j(s) [1 + \gamma_{ij}(s)] + \mathcal{O}(t^3), \quad (3.10)$$

donde N denota el número de neuronas. Y en este contexto podemos escribir la probabilidad de que dos neuronas (r_i, r_j) dispares un par de espigas como:

$$P(r_i, r_j; s) = \bar{r}_i(s)\bar{r}_j(s)t^2[1 + \gamma_{ij}(s)] + \mathcal{O}(t^3). \quad (3.11)$$

Dadas estas definiciones, las siguientes suposiciones adicionales son hechas para poder computar la expansión en series. Primero, que la información mutua es analítica en el tiempo. Segundo, es necesario que diferentes pruebas experimentales (una medida o simulación) sean realizaciones aleatorias del mismo proceso, siendo este un requerimiento propio de un buen experimento. La validez de estas suposiciones fue examinada previamente [9–11]. Entonces, la información mutua puede ser expandida como una serie de la ventana de tiempo post-estímulo como

$$I = I_0 + tI_1 + \frac{t^2}{2}I_2 + \frac{t^3}{6}I_3 + \dots, \quad (3.12)$$

done I_1 es la tasa de información instantánea y I_i con $i \geq 2$ es la derivada i -ésima de la información. Luego, se insertan las probabilidades de respuesta de distintos órdenes, contenidas en la secuencia de espigas, en la sumatorias sobre las respuestas en la ecuación (2.19). Por cada término en la sumatoria sobre las respuestas, usamos la expansión en potencias del logaritmo como en [9]. Luego, agrupamos todos los términos de la sumatoria que tengan la misma potencia en t , y usando la ecuación (3.12), obtenemos las expresiones para cada derivada de la información. Al realizar las sumas de las probabilidades de órdenes mayores a uno se deben dividir las mismas por un factor que tiene en cuenta permutaciones equivalentes.

En otras palabras, las derivadas de la información pueden ser calculadas en términos de las probabilidades dadas por las ecuaciones (3.8), (3.9), (3.10) y (3.11) de la siguiente manera. Primero, uno inserta las probabilidades definidas por las ecuaciones (3.8), (3.9), (3.10) y (3.11) en la suma sobre las respuestas de la ecuación (2.19). Luego, para cada término de la suma sobre las respuestas, uno utiliza la expansión en potencias del logaritmo como función de t :

$$\log_2(1 - tx) = -\frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{(tx)^i}{i} \quad (3.13)$$

como en [9–11]. Finalmente, agrupando todos los términos de la suma que tienen la misma potencia en t , y utilizando la ecuación (3.12), uno obtiene las expresiones para las derivadas de la información y los componentes de codificación de todos los órdenes.

3.1.1. Contribuciones de primer y segundo orden de la expansión

Los términos correspondientes al primer y segundo orden pueden ser encontrados en [9], donde son separados en componentes de codificación de la misma manera como posteriormente fue hecho en [12]. El primer orden de la expansión (y también el primer término de la descomposición de la información) corresponde a I_{lin} , esto es, no se considera ninguna sinergia o redundancia. El segundo orden se descompone en los siguientes componentes (denotados por el subíndice 2 para distinguirlos de aquellos correspondientes a órdenes superiores) como en [12]: $I_{sig-sim,2}$, toma en cuenta la redundancia que pueda surgir incluso en la ausencia de correlaciones cruzadas, $I_{cor-ind,2}$ y $I_{cor-dep,2}$ que miden la cantidad de información verdaderamente disponibles de la actividad correlacionada entre pares de neuronas de toda la población. Cada expresión de estas componentes en la expansión en serie hasta segundo orden son dadas a continuación. A primer orden,

$$\begin{aligned} I_1 &= I_{lin}. \\ &= \sum_{i=1}^N \langle \bar{r}_i(s) \log_2 \left[\frac{\bar{r}_i(s)}{\langle \bar{r}_i(s') \rangle_{s'}} \right] \rangle_s. \end{aligned} \quad (3.14)$$

A segundo orden, la contribución no lineal es

$$I_2 = I_{sig-sim,2} + I_{cor-ind,2} + I_{cor-dep,2} \quad (3.15)$$

en donde

$$I_{sig-sim,2} = \frac{1}{2 \ln 2} \sum_{i,j=1}^N \langle \bar{r}_i(s) \rangle_s \langle \bar{r}_j(s) \rangle_s \left[\nu_{ij} + (1 + \nu_{ij}) \ln \left(\frac{1}{1 + \nu_{ij}} \right) \right], \quad (3.16)$$

$$I_{cor-ind,2} = \frac{1}{2 \ln 2} \sum_{i,j=1}^N \langle \bar{r}_i(s) \bar{r}_j(s) \gamma_{ij}(s) \rangle_s \ln \left(\frac{1}{1 + \nu_{ij}} \right), \quad (3.17)$$

y

$$I_{cor-dep,2} = \frac{1}{2 \ln 2} \sum_{i,j=1}^N \langle \bar{r}_i(s) \bar{r}_j(s) [1 + \gamma_{ij}(s)] \ln \left\{ \frac{\langle \bar{r}_i(s') \bar{r}_j(s') \rangle_{s'} [1 + \gamma_{ij}(s)]}{\langle \bar{r}_i(s') \bar{r}_j(s') [1 + \gamma_{ij}(s')] \rangle_{s'}} \right\} \rangle_s, \quad (3.18)$$

Resulta interesante notar que la descomposición de esta expansión para ventanas temporales cortas hasta segundo orden es exactamente análoga a la previamente detallada en la descomposición exacta en componentes de codificación de la información de una población de neuronas (ver capítulo 2). A pesar de esta analogía, sólo los momentos de primer y segundo orden de las distribuciones de las probabilidades de respuesta son consideradas en las ecuaciones aproximadas de la expansión en serie a segundo orden. Por lo tanto, la aproximación puede ser mejorada al insertar órdenes de más alto orden sucesivamente.

3.1.2. Contribuciones de tercer orden en la expansión

Sin embargo, si la expansión es continuada hasta órdenes más altos, una componente extra debe ser considerada. Dicha componente resulta relacionada con la componente de correlación estímulo dependiente. Al extender este método para considerar contribuciones hechas por interacciones de tercer orden, obtenemos las componentes $I_{sig-sim,3}$, $I_{cor-ind,3}$ y $I_{cor-dep,3}$, como es esperado. Además, obtenemos un término adicional que refleja la posibilidad de “etiquetar” erróneamente un estímulo con una estadística de tripletes cuando en realidad corresponde a disparos coincidentes de tres neuronas que pueda surgir por casualidad de correlaciones de a pares [66], al que denotamos como $I_{cor-ch,3}$.

Escribimos las probabilidades definidas en el marco de las interacciones de tercer orden dadas por las ecuaciones (3.8), (3.9), (3.10) y (3.11) en la sumatoria sobre \mathbf{r} de la ecuación (2.19). Adicionalmente, para cada término de la suma consideramos la expansión en potencias de la ecuación (3.13) como función de t . Agrupando los términos de potencia tres en t y tomando en cuenta la ecuación (3.12) para orden

tres, esto es los términos que contienen t^3 obtenemos

$$\begin{aligned}
 I_3 = & \frac{1}{6 \ln 2} \sum_{i,j,k=1}^N \langle \bar{r}_i(s) \rangle_s \langle \bar{r}_j(s) \rangle_s \langle \bar{r}_k(s) \rangle_s \left[\nu_{ijk} + (1 + \nu_{ijk}) \ln \left(\frac{1}{1 + \nu_{ijk}} \right) \right] \\
 & + \langle \bar{r}_i(s) \bar{r}_j(s) \bar{r}_k(s) \gamma_{ijk}(s) \rangle_s \ln \left(\frac{1}{1 + \nu_{ijk}} \right) \\
 & + \langle \bar{r}_i(s) \bar{r}_j(s) \bar{r}_k(s) [1 + \gamma_{ijk}(s)] \ln \left\{ \frac{\langle \bar{r}_i(s') \bar{r}_j(s') \bar{r}_k(s') \rangle_{s'} [1 + \gamma_{ijk}(s)]}{\langle \bar{r}_i(s') \bar{r}_j(s') \bar{r}_k(s') [1 + \gamma_{ijk}(s')] \rangle_{s'}} \right\} \rangle_s \\
 & - 3 \langle \bar{r}_i(s) \bar{r}_j(s) \bar{r}_k(s) [1 + \gamma_{ijk}(s)] \ln \left\{ \frac{\bar{r}_i(s) \bar{r}_j(s) [1 + \gamma_{ij}(s)]}{\langle \bar{r}_i(s') \bar{r}_j(s') [1 + \gamma_{ij}(s')] \rangle_{s'}} \right\} \rangle_s. \quad (3.19)
 \end{aligned}$$

Entonces a tercer orden, la contribución no lineal es $I_3 = I_{sig-sim,3} + I_{cor-ind,3} + I_{cor-dep,3} + I_{cor-ch,3}$. La contribución de la similaridad en la señal es

$$I_{sig-sim,3} = \frac{1}{6 \ln 2} \sum_{i,j,k=1}^N \langle \bar{r}_i(s) \rangle_s \langle \bar{r}_j(s) \rangle_s \langle \bar{r}_k(s) \rangle_s \left[\nu_{ijk} + (1 + \nu_{ijk}) \ln \left(\frac{1}{1 + \nu_{ijk}} \right) \right]. \quad (3.20)$$

La cantidad total de información atribuible a la actividad correlacionada de tercer orden sobre la codificación neuronal en general está dada por la contribución correlacional independiente del estímulo

$$I_{cor-ind,3} = \frac{1}{6 \ln 2} \sum_{i,j,k=1}^N \langle \bar{r}_i(s) \bar{r}_j(s) \bar{r}_k(s) \gamma_{ijk}(s) \rangle_s \ln \left(\frac{1}{1 + \nu_{ijk}} \right) \quad (3.21)$$

y la componente de correlación estímulo-dependiente

$$\begin{aligned}
 I_{cor-dep,3} = & \frac{1}{6 \ln 2} \sum_{i,j,k=1}^N \langle \bar{r}_i(s) \bar{r}_j(s) \bar{r}_k(s) [1 + \gamma_{ijk}(s)] \\
 & \times \ln \left\{ \frac{\langle \bar{r}_i(s') \bar{r}_j(s') \bar{r}_k(s') \rangle_{s'} [1 + \gamma_{ijk}(s)]}{\langle \bar{r}_i(s') \bar{r}_j(s') \bar{r}_k(s') [1 + \gamma_{ijk}(s')] \rangle_{s'}} \right\} \rangle_s. \quad (3.22)
 \end{aligned}$$

El término adicional calculado resulta

$$\begin{aligned}
 I_{cor-ch,3} = & -\frac{1}{2 \ln 2} \sum_{i,j,k=1}^N \langle \bar{r}_i(s) \bar{r}_j(s) \bar{r}_k(s) [1 + \gamma_{ijk}(s)] \\
 & \times \ln \left\{ \frac{\bar{r}_i(s) \bar{r}_j(s) [1 + \gamma_{ij}(s)]}{\langle \bar{r}_i(s') \bar{r}_j(s') [1 + \gamma_{ij}(s')] \rangle_{s'}} \right\} \rangle_s. \quad (3.23)
 \end{aligned}$$

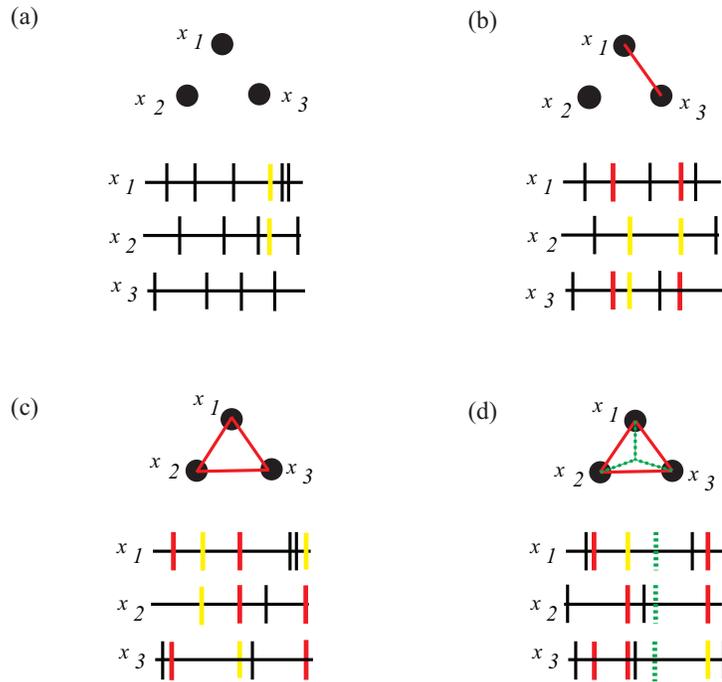


Figura 3.1: Esquema de tres neuronas interactuantes. (a)-(d) Cuatro casos característicos descritos en el texto. Notar que las entradas de pares y tripletes se encuentran representadas en rojo y verde respectivamente. Las correlaciones por casualidad se marcan con amarillo.

En lo siguiente, discutiremos el significado del término $I_{cor-ch,3}$ cuando sólo se especifican las correlaciones de a pares. Consideremos el caso de tres neuronas con cuatro escenarios posible. La Figura 3.1 (a) ilustra el caso en donde tres neuronas no comparten entradas comunes, pero sin embargo aún pueden llegar a existir coincidencias entre potenciales de acción entre pares de neuronas (marcadas en amarillo). En la Figura 3.1 (b), sólo dos de las neuronas poseen entradas comunes, y por lo tanto

existe correlación de orden dos (marcada en rojo). Notar que en este caso, podrían existir coincidencias entre espigas de tripletes de neuronas debido a que dos neuronas se encuentran correlacionadas y una tercera se correlaciona por casualidad con dicho par. La Figura 3.1 (c) describe el caso en el que las tres neuronas comparten entradas comunes (entre cada par), y los tripletes coincidentes son más probables de ocurrir que aquellos correspondientes a la Figura 3.1 (b). Finalmente, la Figura 3.1 (d) ilustra el caso en el que las neuronas en cuestión comparten entradas comunes, tanto entre pares como entre el triplete, respectivamente denotadas por rojo y verde. Sin embargo, correlaciones de orden dos y de orden tres pueden llegar a aparecer por casualidad. Por lo tanto, las diferencias aparentes entre estas dos poblaciones (es decir, entre los casos (c) y (d)) no son capturadas por meras correlaciones entre pares. Este problema surgiría incluso si estuviésemos considerando el caso de una población de neuronas que sólo poseen entradas comunes entre cada par, donde pueden haber correlaciones por casualidad entre el triplete, como hemos comentado previamente. Si ese fuese el caso, necesitaríamos del término $I_{cor-ch,3}$, que cuantifica la cantidad de información provista por el conjunto de correlaciones verdaderas entre los pares y correlaciones por casualidad entre el triplete. En la siguiente sección discutiremos cómo tener en cuenta el término $I_{cor-ch,3}$ en la expansión en series de la información mutua puede cambiar el escenario de sinergia/redundancia al considerar una población de neuronas.

3.1.3. Términos de orden superior

La componente $I_{cor-ch,3}$ puede ser encontrada consistentemente cuando interacciones de órdenes mayores a dos son consideradas, y depende de correlaciones de ruido de órdenes menores. Notar que resulta posible extender las definiciones previas para la correlación de ruido de orden n de la manera

$$\gamma_{i_1 \dots i_n}(s) = \frac{\overline{r_{i_1}(s) \cdots r_{i_n}(s)}}{\bar{r}_{i_1}(s) \cdots \bar{r}_{i_n}(s)} - 1. \quad (3.24)$$

Similarmente, el n -ésimo orden de la correlación de señal está dado por:

$$\nu_{i_1 \dots i_n} = \frac{\langle \bar{r}_{i_1}(s) \cdots \bar{r}_{i_n}(s) \rangle_s}{\langle \bar{r}_{i_1}(s) \rangle_s \cdots \langle \bar{r}_{i_n}(s) \rangle_s} - 1. \quad (3.25)$$

La probabilidad (hasta cuarto orden) de tener cuatro neuronas disparando se escribe como

$$P(r_i, r_j, r_k, r_l; s) = \bar{r}_i(s)\bar{r}_j(s)\bar{r}_k(s)\bar{r}_l(s)t^4 [1 + \gamma_{ijkl}(s)] + \mathcal{O}(t^5), \quad (3.26)$$

y también se puede generalizar para N neuronas que disparan n espigas simultáneamente (permitiendo interacciones de orden n) de la forma

$$P(r_{i_1}, r_{i_2}, \dots, r_{i_n}; s) = \bar{r}_{i_1}(s) \cdots \bar{r}_{i_n}(s)t^n [1 + \gamma_{i_1 \dots i_n}(s)] + \mathcal{O}(t^{n+1}). \quad (3.27)$$

Las suposiciones que deben ser hechas para poder computar la expansión en series de la información hasta cualquier orden son las mismas que las previamente detalladas. Específicamente, a cuarto orden las contribuciones a la expansión en series pueden ser calculadas como

$$I_4 = I_{sig-sim,4} + I_{cor-ind,4} + I_{cor-dep,4} + I_{cor-ch,4}. \quad (3.28)$$

Las contribuciones de similitud en la señal, de correlación independiente del estímulo y la dependiente del estímulo resultan, respectivamente,

$$I_{sig-sim,4} = \frac{1}{24 \ln 2} \sum_{i,j,k,l=1}^N \langle \bar{r}_i(s) \rangle_s \cdots \langle \bar{r}_l(s) \rangle_s \left[\nu_{ijkl} + (1 + \nu_{ijkl}) \ln \left(\frac{1}{1 + \nu_{ijkl}} \right) \right], \quad (3.29)$$

$$I_{cor-ind,4} = \frac{1}{24 \ln 2} \sum_{i,j,k,l=1}^N \langle \bar{r}_i(s) \cdots \bar{r}_l(s) \gamma_{ijkl}(s) \rangle_s \ln \left(\frac{1}{1 + \nu_{ijkl}} \right), \quad (3.30)$$

$$I_{cor-dep,4} = \frac{1}{24 \ln 2} \sum_{i,j,k,l=1}^N \langle \bar{r}_i(s) \cdots \bar{r}_l(s) [1 + \gamma_{ijkl}(s)] \rangle_s \times \ln \left\{ \frac{\langle \bar{r}_i(s') \cdots \bar{r}_l(s') [1 + \gamma_{ijkl}(s)] \rangle_{s'}}{\langle \bar{r}_i(s') \cdots \bar{r}_l(s') [1 + \gamma_{ijkl}(s')] \rangle_{s'}} \right\}_s. \quad (3.31)$$

Finalmente, $I_{cor-ch,4}$ se calcula

$$\begin{aligned}
 I_{cor-ch,4} = & -\frac{1}{24 \ln 2} \sum_{i,j,k,l=1}^N \langle \bar{r}_i(s) \cdots \bar{r}_l(s) [1 + \gamma_{ijkl}(s)] \\
 & \times \left\{ 6 \ln \left(\frac{\bar{r}_i(s) \bar{r}_j(s) [1 + \gamma_{ij}(s)]}{\langle \bar{r}_i(s') \bar{r}_j(s') [1 + \gamma_{ij}(s')] \rangle_{s'}} \right) \right. \\
 & \left. + 4 \ln \left(\frac{\bar{r}_i(s) \bar{r}_j(s) \bar{r}_k(s) [1 + \gamma_{ijk}(s)]}{\langle \bar{r}_i(s') \bar{r}_j(s') \bar{r}_k(s') [1 + \gamma_{ijk}(s')] \rangle_{s'}} \right) \right\}_s. \quad (3.32)
 \end{aligned}$$

Tener en cuenta que depende tanto de las correlaciones de cuarto orden, como de las correlaciones de segundo y tercer orden.

Luego, extendiendo estos cálculos a órdenes superiores, derivamos el orden n -ésimo de la componente que corresponde a las contribuciones de correlaciones espurias $I_{cor-ch,n}$. La contribución de similitud de señal de cualquier orden se lee como

$$\begin{aligned}
 I_{sig-sim,n} = & \frac{1}{n! \ln 2} \sum_{i_1, \dots, i_n=1}^N \langle \bar{r}_{i_1}(s) \rangle_s \cdots \langle \bar{r}_{i_n}(s) \rangle_s \\
 & \times \left[\nu_{i_1 \dots i_n} + (1 + \nu_{i_1 \dots i_n}) \ln \left(\frac{1}{1 + \nu_{i_1 \dots i_n}} \right) \right]. \quad (3.33)
 \end{aligned}$$

La componente de correlación independiente del estímulo resulta

$$I_{cor-ind,n} = \frac{1}{n! \ln 2} \sum_{i_1, \dots, i_n=1}^N \langle \bar{r}_{i_1}(s) \cdots \bar{r}_{i_n}(s) \gamma_{i_1 \dots i_n}(s) \rangle_s \ln \left(\frac{1}{1 + \nu_{i_1 \dots i_n}} \right), \quad (3.34)$$

y la componente de correlación estímulo-dependiente

$$\begin{aligned}
 I_{cor-dep,n} = & \frac{1}{n! \ln 2} \sum_{i_1, \dots, i_n=1}^N \langle \bar{r}_{i_1}(s) \cdots \bar{r}_{i_n}(s) [1 + \gamma_{i_1 \dots i_n}(s)] \\
 & \times \ln \left\{ \frac{\langle \bar{r}_{i_1}(s') \cdots \bar{r}_{i_n}(s') \rangle_{s'} [1 + \gamma_{i_1 \dots i_n}(s)]}{\langle \bar{r}_{i_1}(s') \cdots \bar{r}_{i_n}(s') [1 + \gamma_{i_1 \dots i_n}(s')] \rangle_{s'}} \right\}_s. \quad (3.35)
 \end{aligned}$$

La componente I_{cor-ch} correspondiente al orden n -ésimo de la expansión en series

puede ser escrita como:

$$\begin{aligned}
 I_{cor-ch,n} = & -\frac{1}{n! \ln 2} \sum_{i_1, \dots, i_n=1}^N \langle \bar{r}_{i_1}(s) \cdots \bar{r}_{i_n}(s) [1 + \gamma_{i_1 \dots i_n}(s)] \\
 & \times \sum_{m=2}^{n-1} \binom{n}{m} \ln \left\{ \frac{\bar{r}_{i_1}(s) \cdots \bar{r}_{i_m}(s) [1 + \gamma_{i_1 \dots i_m}(s)]}{\langle \bar{r}_{i_1}(s') \cdots \bar{r}_{i_m}(s') [1 + \gamma_{i_1 \dots i_m}(s')] \rangle_{s'}} \right\}_s, \quad (3.36)
 \end{aligned}$$

donde $\gamma_{i_1 \dots i_m}$ es el coeficiente de correlación de ruido de orden m .

3.2. Codificación sinérgica y redundante en términos de alto orden

Para investigar más extensamente cómo las correlaciones de alto orden pueden contribuir al procesamiento de la información en el cerebro, es necesario evaluar cómo la componente I_{cor-ch} en cada término de orden mayor a dos conforma la codificación neuronal cuando se considera que dentro de una población las neuronas interactúan entre ellas. Como previamente explicado en [9], la expansión en series de la información mutua muestra que las correlaciones generales en la distribución de las respuestas promedio solamente, esto es las correlaciones de señal, solo pueden contribuir redundantemente. Para lograr una codificación sinérgica de la información, las correlaciones en la variabilidad de las respuestas (o correlaciones de ruido) son necesarias. Puede ser probado que incluso cuando las últimas son independientes del estímulo, aún es posible obtener sinergia al tomar en consideración el signo de la redundancia de Shannon (obtenida de la ecuación (2.19)) al substraer la información transmitida por la población de la suma de aquella transportada por cada célula individual [9].

Cuando las correlaciones son tomadas como independientes del estímulo elegido, esto es $\gamma(\mathbf{r}|s) = \gamma(\mathbf{r})$, es posible derivar cuantitativamente los valores de γ y ν para los cuales el sistema es redundante o sinérgico. Para ello en los próximos párrafos definiremos una medida I_{SRL} que nos permite delimitar los distintos escenarios posibles de redundancia y sinergia al considerar contribuciones de tercer orden.

Consideremos un triplete de neuronas, en donde la correlación de ruido es toma-

da como independiente del estímulo, es decir $\gamma_{123}(s) = \gamma_{123}$. Bajo esta suposición, es posible calcular analíticamente las curvas $(\gamma_{123}, \nu_{123})$ para las cuales la información mutua cambia de signo. Para esto, consideremos primero el caso $I_3 = 0$, ya que constituye el valor para el cual las contribuciones de tercer orden pasan de ser sinérgicas a redundantes y viceversa ($I_3 > 0$ contribuye sinérgicamente y $I_3 < 0$ redundantemente). Por lo tanto, en este primer caso ($I_3 = 0$):

$$\begin{aligned}
 0 &= \langle \bar{r}_1(s) \rangle_s \langle \bar{r}_2(s) \rangle_s \langle \bar{r}_3(s) \rangle_s \left[\nu_{123} + (1 + \nu_{123}) \ln \left(\frac{1}{1 + \nu_{123}} \right) \right] \\
 &+ \langle \bar{r}_1(s) \bar{r}_2(s) \bar{r}_3(s) \rangle_s \gamma_{123} \ln \left(\frac{1}{1 + \nu_{123}} \right) + 0 \\
 &+ \langle \bar{r}_1(s) \bar{r}_2(s) \bar{r}_3(s) [1 + \gamma_{123}] \\
 &\times \ln \left[\frac{\bar{r}_1(s) \bar{r}_2(s)}{\langle \bar{r}_1(s') \bar{r}_2(s') \rangle_{s'}} \frac{\bar{r}_1(s) \bar{r}_3(s)}{\langle \bar{r}_1(s') \bar{r}_3(s') \rangle_{s'}} \frac{\bar{r}_2(s) \bar{r}_3(s)}{\langle \bar{r}_2(s') \bar{r}_3(s') \rangle_{s'}} \right] \rangle_s. \quad (3.37)
 \end{aligned}$$

Como es esperado cuando las correlaciones son tomadas independientes del estímulo elegido, $I_{cor-dep,3}$ se anula bajo esta suposición. Sin embargo, los demás términos se mantienen relevantes (ver sección 2.2.4). Luego,

$$\gamma_{123} = \frac{\frac{\nu_{123}}{1 + \nu_{123}} - \left[\ln(1 + \nu_{123}) + f(\bar{\mathbf{r}}) \right]}{\left[\ln(1 + \nu_{123}) + f(\bar{\mathbf{r}}) \right]}, \quad (3.38)$$

donde definimos la función $f(\bar{\mathbf{r}})$ como

$$\begin{aligned}
 f(\bar{\mathbf{r}}) &= f(\bar{r}_1(s), \bar{r}_2(s), \bar{r}_3(s)). \\
 &= \frac{1}{\langle \bar{r}_1(s) \bar{r}_2(s) \bar{r}_3(s) \rangle_s} \langle \bar{r}_1(s) \bar{r}_2(s) \bar{r}_3(s) \\
 &\times \ln \left[\frac{\bar{r}_1(s) \bar{r}_2(s)}{\langle \bar{r}_1(s') \bar{r}_2(s') \rangle_{s'}} \frac{\bar{r}_1(s) \bar{r}_3(s)}{\langle \bar{r}_1(s') \bar{r}_3(s') \rangle_{s'}} \frac{\bar{r}_2(s) \bar{r}_3(s)}{\langle \bar{r}_2(s') \bar{r}_3(s') \rangle_{s'}} \right] \rangle_s, \quad (3.39)
 \end{aligned}$$

en donde $f(\bar{\mathbf{r}})$ es un promedio normalizado del producto de las tasas de disparo del triplete de neuronas pesado por el logaritmo natural del producto de las tasas de disparo normalizadas de los pares de neuronas correspondientes [67]. Es necesario notar que si los pares de neuronas no codifican para el estímulo considerado entonces

$f(\bar{\mathbf{r}})$ es nulo. Por ende, estudiamos en lo siguiente el signo de los componentes de la información de tercer orden en las regiones delimitadas por las curvas determinadas por la ecuación (3.38). Consideremos ahora el caso en donde $I_3 \neq 0$. Si I_3 es positivo, entonces su contribución es sinérgica; en caso contrario, cuando I_3 es negativo, es redundante. Estimamos luego la contribución sinérgica de los tripletes como $I_{SRL} = I - (I_{lin} + I_2)$, calculado como la información mutua total I (ecuación (3.12)) menos la suma lineal de la información transmitida por cada célula individual (I_{lin}) y menos las contribuciones de orden dos a la información I_2 , considerando sólo tres neuronas e imponiendo la restricción de que todas las correlaciones son tomadas como independientes de los estímulos elegidos. Luego, la sinergia I_{SRL} resulta

$$\begin{aligned}
 I_{SRL} = & \langle \bar{r}_1(s) \rangle_s \langle \bar{r}_2(s) \rangle_s \langle \bar{r}_3(s) \rangle_s \left\{ \nu_{123} - (1 + \nu_{123})(1 + \gamma_{123}) \left[\ln(1 + \nu_{123}) \right. \right. \\
 & \left. \left. + f(\bar{r}_1(s), \bar{r}_2(s), \bar{r}_3(s)) \right] \right\}. \tag{3.40}
 \end{aligned}$$

Si I_{SRL} es positivo entonces, como es esperado, la contribución de los tripletes al código es referida como sinérgica; de lo contrario, cuando I_{SRL} es negativo, es redundante. Si esta cantidad es cero, entonces el código es como mucho no lineal a segundo orden. Notar que en la ecuación (3.40) el factor fuera de la llave dado por $\langle \bar{r}_1(s) \rangle_s \langle \bar{r}_2(s) \rangle_s \langle \bar{r}_3(s) \rangle_s$ es estrictamente positivo ya que depende exclusivamente de las tasas medias de disparo de las neuronas consideradas.

Las Figuras 3.2 a 3.4 ilustran los distintos regímenes de codificación para el tercer orden en un ejemplo de tres células, cuando se analiza I_{SRL} versus γ_{123} versus ν_{123} . Cuando el término $I_{cor-ch,3}$ es nulo, entonces el análisis se reduce al mismo realizado previamente por [9] para el ejemplo de dos células al considerar únicamente correlaciones de a pares, y que corresponde a las Figuras 3.2 (a) y (b). En este sentido, de los resultados mostrados en las Figuras 3.2 (a) y (b) tomamos que $f(\bar{\mathbf{r}}) = 0$, y en este caso en el que consideramos además el coeficiente de correlación de ruido de tercer orden γ_{123} versus el coeficiente de correlación de señal de tercer orden ν_{123} , las zonas correspondientes a la redundancia y sinergia son exactamente iguales a las mostradas en el caso de correlaciones de pares puras (o verdaderas) [9]. Por lo tanto, en tal escenario, cuando $f(\bar{\mathbf{r}}) = 0$, las zonas de sinergia y redundancia se encuentran definidas solamente por correlaciones de segundo orden. Si las respuestas medias

de cada neurona a los distintos estímulos no están correlacionadas, tenemos que $\nu_{123} = 0$, y luego la redundancia es exactamente cero (independientemente del valor de la correlación de ruido γ_{123}) como es mostrado a lo largo de la línea correspondiente al eje vertical de la Figura 3.2 (a) (ver también la línea punteada paralela al eje de γ_{123} en la Figura 3.2 (b)). Si las neuronas anti-covarían en sus perfiles de respuesta frente a los estímulos considerados entonces, para poder obtener sinergia, deben tener un coeficiente de correlación de ruido de valor positivo por sobre el umbral establecido. Por el contrario, si las células tienen una correlación de señal mayor a cero, entonces las coincidencias deben ser activamente suprimidas por una correlación de ruido mayor que el valor del umbral correspondiente. Cuando las correlaciones de señal y ruido poseen el mismo signo, uno siempre obtiene redundancia en el límite de la escala de ventanas temporales cortas. La Figura 3.2 (b) muestra, además de esto, que el gradiente en la cantidad de redundancia o sinergia se vuelve más pronunciado con respecto al valor de la correlación de ruido y al valor de la correlación de señal, en comparación a la falta de variabilidad cuando la correlación de señal es pequeña sin importar el valor de la correlación de ruido. Cuando las neuronas covarían en el número medio de espigas que disparan para cada estímulo, manteniendo todas las demás cantidades constantes, obtenemos una situación redundante. Pero, si existe suficiente anticorrelación en el ruido, o viceversa si existe anticorrelación en la señal y suficiente correlación positiva en el ruido, entonces la codificación pasa a ser sinérgica.

Sin embargo, cuando el término $I_{cor-ch,3}$ es no nulo (tanto positivo como negativo) existe una modificación sustancial de las dinámicas de correlación. Analíticamente, encontramos que este cambio depende del valor de $f(\bar{\mathbf{r}})$, una función que depende únicamente de las tasas de disparo medias de las tres neuronas en cuestión. Esto significa que es una cantidad modelo-dependiente y/o fuertemente variable con el conjunto de datos considerados. Aún cuando el valor de esta función sea pequeño, posee un gran impacto en la interacción entre las correlaciones de ruido y de señal. Por un lado, como puede verse en la Figura 3.3 (a) y (b), cuando $f(\bar{\mathbf{r}})$ es positivo, entonces la región de redundancia se incrementa ampliamente para valores esperados de correlación. Sin embargo, continúa siendo posible alcanzar sinergia para correlaciones de señal positivas y correlaciones de ruido negativas por debajo de un cierto valor umbral. Por otra parte, cuando el valor de $f(\bar{\mathbf{r}})$ es negativo, como se muestra

en la Figura 3.4 (a) y (b), entonces correlaciones pequeñas de cualquier signo llevan a una contribución sinérgica, mientras que es posible obtener un código redundante superando un umbral para correlaciones de ruido y de señal de signos opuestos. Para ambos casos, la cantidad total de redundancia o sinergia es mucho más sensible a la variación cuando tanto la correlación de ruido como la de señal son grandes. Resulta importante remarcar que en ambos casos presentados en las Figuras 3.3 y 3.4, cuando $f(\bar{\mathbf{r}}) \neq 0$ ($I_{cor-ch,3} \neq 0$), las correlaciones entre tripletes redefinen completamente los escenarios posible de sinergia y redundancia.

Existe evidencia que la computación cooperativa caracteriza las interacciones neuronales en diversas escalas temporales y espaciales [68–71]. Evidencia directa muestra que el análisis de correlaciones neuronales de a pares no resuelve dicha dinámica de cooperación poblacional [24, 25, 54, 62, 66, 72–75]. Hemos mostrado la existencia de un término correlacional mixto que define un nuevo escenario al analizar correlaciones de alto orden, poniendo de manifiesto las limitaciones de sólo considerar correlaciones de segundo orden en secuencias de potenciales de acción en un sistema de tres neuronas. Utilizamos una extensión de la expansión en serie para ventanas temporales cortas de la información mutua para estudiar explícitamente un término mixto entre correlaciones de segundo y tercer orden a través de $I_{cor-ch,3}$, el cual modifica significativamente la interacción entre la correlación en el ruido y la correlación en la señal que puede llevar tanto a sinergia como a redundancia dependiendo de las tasas de disparo promediadas sobre distintas repeticiones y estímulos.

3.3. Modelado computacional con estadística de Poisson

Un proceso estocástico que genera una secuencia de eventos, como un tren de espigas por ejemplo, se denomina un proceso de Poisson si la probabilidad de un evento que ocurre a un tiempo en particular es independiente del evento que lo precede inmediatamente. A pesar de esta limitación evidente, los procesos de Poisson han demostrado proveer una aproximación extremadamente útil a la generación de disparos neuronales estocásticos de una población cuando es extendido apropiadamente. La metodología que utilizamos para realizar nuestro modelado computacional

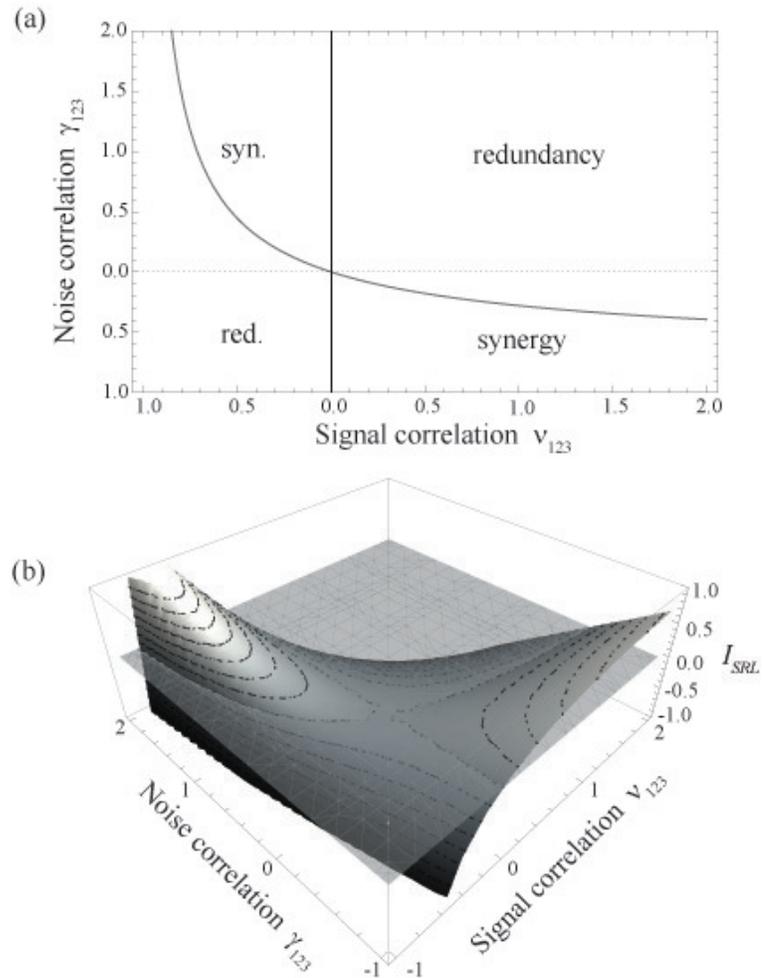


Figura 3.2: I_{SRL} (en unidades de bits) versus correlación de ruido (‘Noise correlation’) entre tripletes γ_{123} versus correlación de señal (‘Signal correlation’) entre tripletes ν_{123} . La consideración de un triplete de neuronas, en donde la correlación de ruido es tomada como independiente del estímulo, ilustra el rango de posibilidades para la dinámica cuando $f(\bar{\mathbf{r}}) = 0$ (o $I_{cor-ch,3} = 0$). (a) Regiones donde los valores positivos de I_{SRL} corresponden a sinergia (‘synergy’) y valores negativos corresponden a redundancia (‘redundancy’), tomando el plano γ_{123} versus ν_{123} . (b) El gráfico en tres dimensiones de I_{SRL} versus γ_{123} versus ν_{123} correspondiente al plano anterior. Como esperado, valores por sobre del eje cero indican codificación sinérgica, mientras que valores por debajo del cero indican redundancia.

de poblaciones pequeñas de neuronas correlacionadas es similar a la usada en las re-

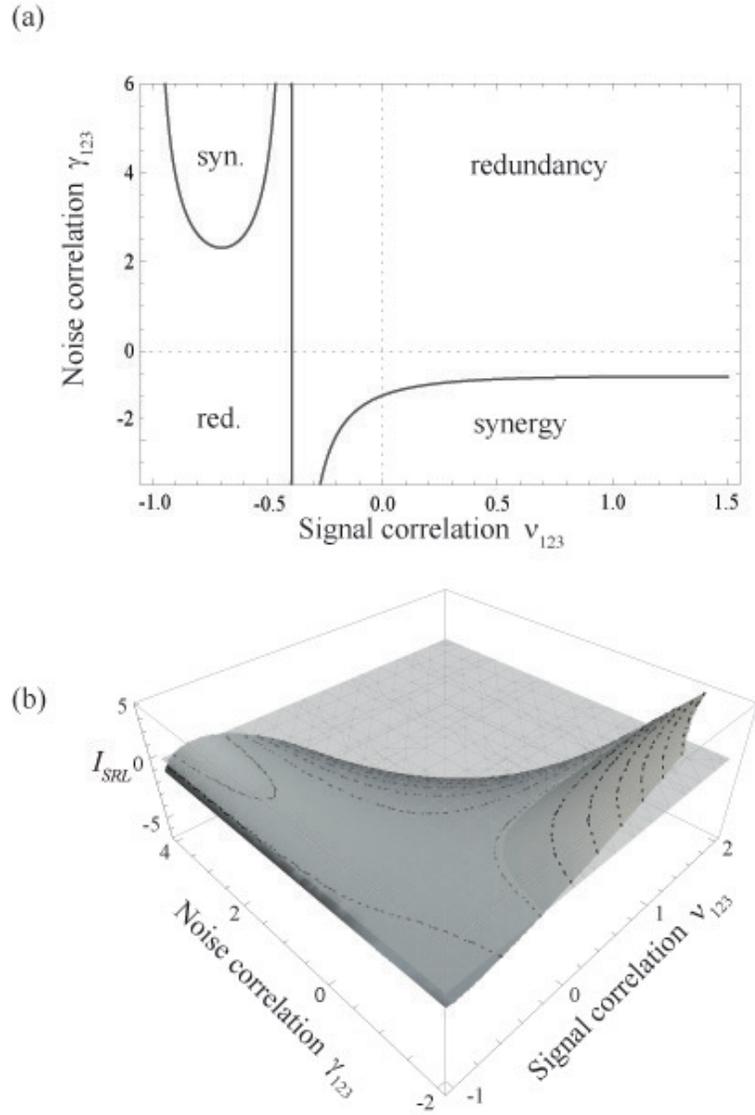


Figura 3.3: I_{SRL} (en unidades de bits) versus correlación de ruido (‘Noise correlation’) entre tripletes γ_{123} versus correlación de señal (‘Signal correlation’) entre tripletes ν_{123} . Igual que en la Figura 3.2 (a) y (b) pero considerando $f(\bar{\mathbf{r}}) > 0$ (entonces $I_{cor-ch,3} \neq 0$).

ferencias [65, 76] y ha sido previamente aplicada en [9–11, 13, 39]. En lo siguiente, consideramos un proceso de Poisson inhomogéneo, el cual produce variaciones en la tasa de disparos. Para lograr esto, la probabilidad $P(t_i)$ de tener N espigas en una

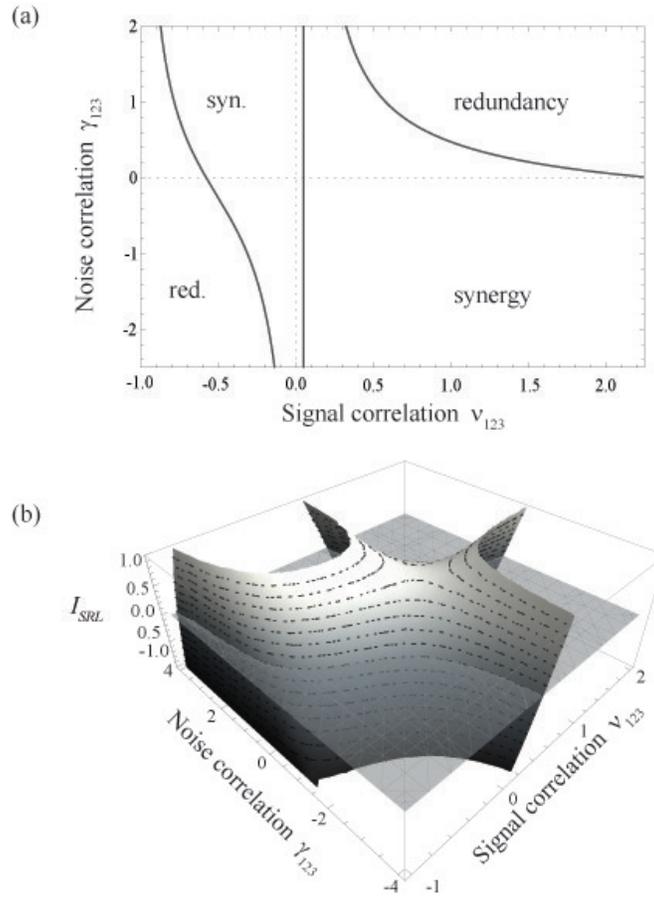


Figura 3.4: I_{SRL} (en unidades de bits) versus correlación de ruido (‘Noise correlation’) entre tripletes γ_{123} versus correlación de señal (‘Signal correlation’) entre tripletes ν_{123} . Igual que en la Figura 3.2 (a) y (b) pero considerando $f(\bar{\mathbf{r}}) < 0$ (entonces $I_{cor-ch,3} \neq 0$).

ventana temporal t_i (el índice i varía dese 1 a M), tal que $0 \leq t_1 \leq t_2 \leq \dots \leq t_M \leq T$) se encuentra dada por

$$P(t_i) = \frac{[\Lambda_\theta(t_i)]^N}{N!} \exp[-\Lambda_\theta(t_i)] \quad (3.41)$$

donde Λ_θ a un dado t_i se construye mediante la incorporación de diferentes condiciones de estímulo en ese momento, y es definida como

$$\Lambda_\theta(t_i) = \{m + a \exp [b (\cos ((\theta + \varphi) - \theta_{pref}) - 1)]\} t_i. \quad (3.42)$$

Esto es, Λ_θ es la función predicha del estímulo a un dado tiempo t_i . Notar que el proceso se define como inhomogéneo dado que la tasa de disparos media se permite variar con el tiempo. Para el modelo computacional elegimos valores similares a aquellos adoptados en [77]: $m = 15$ es un ajuste de comienzo, $a = 1,3$ determina la altura de la curva de ajuste, $b = 3$ especifica el ancho de banda del ajuste (cuantificado como el ancho del ajuste en el punto medio entre las respuestas evocadas mínimas y máximas), $\theta_{pref} = \frac{\pi}{4}$ es la orientación preferida, y θ es la orientación del estímulo (varía desde 0 a 2π). La modulación a lo largo de las neuronas de la tasa de disparo es producida tomando una dependencia de tipo función de von Mises [77] e incorporando un corrimiento de fase en la dependencia del estímulo $\varphi = \frac{\pi}{10}$ entre las distintas células. En la práctica, esto es realizado para cada célula a lo largo de números aleatorios elegidos de una distribución de Poisson con parámetro alimentado por $\Lambda(t_i)$, a un dado momento t_i . Considerando diferentes repeticiones, construimos las probabilidades de tener una respuesta particular \mathbf{r} condicional al estímulo $s = \theta$. Esto genera procesos de Poisson inhomogéneos donde las correlaciones entre las neuronas son producidas por las distintas modulaciones [65, 76].

La idea principal detrás de este modelado computacional es que neuronas cercanas comparten entradas comunes y poseen campos de recepción similares [13]. Por ende, puede ser asumido que la tasa de disparos para cada neurona aproximadamente sigue una estadística de Poisson inhomogénea, y al tomar un corrimiento de fase en la dependencia con el estímulo entre cada una de las células las modulaciones en la actividad son generadas. Esto tiene sustento experimental dado que se ha probado recientemente que las modulaciones en la tasa pueden ciertamente incrementar las correlaciones entre neuronas [78].

En lo siguiente consideramos el caso mostrado en la Figura 3.1 (d) que ilustra el caso en que tres neuronas comparten entradas comunes de segundo y tercer orden. Realizamos el modelado computacional descrito previamente con el conjunto de dos y tres células que dispara siguiendo una distribución de Poisson inhomogénea,

en donde las espigas no son independientes debido a la modulación de la tasa en sus entradas. Consideramos tres conjuntos de pares de neuronas dados por los pares $\{x_1, x_2\}$, $\{x_1, x_3\}$ y $\{x_2, x_3\}$, que poseen correlaciones entre pares dadas las modulaciones de las tasas. Esto se logra truncando apropiadamente el espacio de respuestas del conjunto $\{x_1, x_2, x_3\}$. Finalmente, consideramos las correlaciones entre tripletes entre $\{x_1, x_2, x_3\}$, que son tres células generadas con el método descrito anteriormente.

3.3.1. Grado de sinergia

Para computar el grado de sinergia (o fracción de sinergia SF) [13] en la población para el rango de ventanas temporales, calculamos el índice de sinergia como

$$SF = 1 - \frac{I_{lin}}{I}. \quad (3.43)$$

Valores de SF mayores a 0 indican codificación sinérgica, mientras que valores menores a 0 indican redundancia. Se debe notar que, como marcamos anteriormente, si $I_{cor-ch,3} \neq 0$ existe una modificación sustancial en las regiones de sinergia y redundancia, debido a que las correlaciones de tercer orden modifican la dinámica en comparación al caso en donde sólo las correlaciones puras de segundo orden son consideradas. Las correlaciones entre tripletes en tal caso pueden llevar tanto a sinergia como a redundancia.

La Figura 3.5 muestra la fracción de sinergia al considerar la simulación de tres neuronas distintas como en la Figuras 3.1 (c)-(d) cuyas estadísticas de disparo siguen una distribución de Poisson inhomogénea, con un paso temporal de 1 *ms*, 1000 repeticiones y 100 estímulos diferentes con una curva de ajuste de tipo von Mises. En particular, de la Figura 3.5 (a) es posible ver que a medida que la ventana temporal se incrementa, la redundancia aumenta cuando las correlaciones entre los pares del conjunto neuronal $\{x_1, x_2\}$, $\{x_2, x_3\}$ y $\{x_1, x_3\}$ son consideradas, con el tipo de interacción mostrado en la Figura 3.1 (d). Esto se ilustra más claramente en la Figura 3.5 (b), donde el efecto de agregar correlaciones estímulo-dependientes de tercer orden incrementa la fracción de sinergia a medida que la ventana temporal aumenta. Este resultado puede ser explicado mediante el hecho de que existe una

cantidad significativa de correlaciones independientes del estímulo a tercer orden, como se explicita en la Figura 3.5 (c). A su vez, este efecto se vuelve más relevante a medida que alargamos la duración de la ventana temporal. Notar además de la Figura 3.5 (c) que $I_{cor-ch,3} \neq 0$, y por lo tanto cuando consideramos las interacciones descritas en la Figura 3.1 (d), las correlaciones entre tripletes se vuelven relevantes. Como discutido previamente, la posibilidad de que $I_{cor-ch,3} \neq 0$ permite redefinir las “regiones permitidas” de sinergia/redundancia al ser comparado con el caso en el cual sólo las correlaciones entre pares son tomadas en cuenta.

3.4. Modelo de red tipo integrar-y-disparar

En una red de neuronas biológicamente realista, usualmente toma múltiples entradas para que una neurona pueda propagar una señal. Una representación simple de esto es considerar numerosas señales de entrada conectando una neurona a la siguiente, incrementando o disminuyendo el efecto del disparo de la neurona individual. Esto puede ser logrado ajustando los pesos de conexión entre cada neurona y, por ende en una simulación computacional de una red neuronal, podemos imitar la dinámica neuronal a través de una metodología conocida como “integrar-y-disparar” [4]. Cuando el potencial de membrana de la neurona alcanza su umbral, la célula dispara una espiga y la señal es transmitida hacia otra neurona. Luego la señal propagada puede inducir un potencial de acción en esta nueva neurona. Si la neurona no dispara, su potencial puede ser incrementado y cuando recibe otra nueva señal de entrada dentro de un periodo relativamente corto, existiran más chances que alcance el umbral y se produzca una espiga.

Para estudiar la hipótesis de los regímenes de sinergia y redundancia cuando las correlaciones son independientes del estímulo elegido, se simularon secuencias de espigas utilizando una red pequeña, enteramente interconectada de neuronas del tipo integrar-y-disparar. Para evaluar nuestro marco teórico, realizamos simulaciones utilizando este modelo para obtener valores numéricos para las distintas variables compatibles con las suposiciones y aproximaciones hechas. Es importante resaltar que cuando el término $I_{cor-ch,3}$ es no nulo (tanto positivo o negativo), existe una modificación sustancial de las dinámicas de correlación. Esta alteración depende del

valor de $f(\bar{\mathbf{r}})$, que es función de las tasas de disparo de las tres neuronas en cuestión. Para un triplete de neuronas tomadas de la población, una contribución redundante ($I_{SRL} < 0$) se obtuvo con $f(\bar{\mathbf{r}}) \approx 0,24$, y $\nu_{123} = 0,07$ con $\gamma_{123} = 0,000001$ cercana a cero. El intercambio entre las correlaciones en el ruido y en la señal, en este caso, lleva a la redundancia como en la Figura 3.3 (a)-(b).

3.5. Modelo de red tipo Izhikevich

Las neuronas disparan potenciales de acción cuando se encuentran cerca de una bifurcación entre el estado de reposo a activo, y es el delicado balance entre el ruido, las corrientes dinámicas y la condición inicial que determina el diagrama de fases de la actividad neuronal. Mientras que existe un número enorme de mecanismos iónicos de excitabilidad y generación de espigas posibles, solamente cuatro mecanismos de bifurcación pueden resultar en dicha transición. Estas bifurcaciones dividen a las neuronas en cuatro categorías: integradores o resonadores, monoestables o biestables [79]. Metodologías de bifurcaciones nos permiten reproducir precisamente las propiedades biofísicas de los modelos neuronales tipo Hodgkin-Huxley considerando sólo un sistema de ecuaciones diferenciales ordinarias de dos dimensiones y cuatro parámetros diferentes. Este modelo se llama “modelo simple de neurona que dispara” [80], y el sistema de ecuaciones diferenciales ordinarias está dado por:

$$\frac{dv}{dt} = 0,04 v^2 + 5 v + 140 u + i_c \quad (3.44)$$

$$\frac{du}{dt} = a (b v - u) \quad (3.45)$$

con el reseteo post-espiga auxiliar

$$\text{si } v \geq +30 \text{ mV, luego } \begin{cases} v \leftarrow c \\ u \leftarrow u + d, \end{cases} \quad (3.46)$$

donde v es el potencial de membrana de la neurona, y u es la variable de recuperación de la membrana, que toma en cuenta la activación de las corrientes iónicas K_+ y la inactivación de las corrientes iónicas Na_+ , y otorga retroalimentación negativa a v .

Por ende estamos considerando un sistema 2–dimensional de ecuaciones diferenciales ordinarias de las dos variables u and v , y todos los “tipos” conocidos de neuronas pueden ser reproducidos tomando valores diferentes para los cuatro parámetros a , b , c y d . Luego que la espiga llega a su punto máximo en $+30\text{ mV}$ (no debe ser confundido con el umbral de disparo), el voltaje de membrana y la variable de recuperación son reseateados según la ecuación (3.46). La variable i_c corresponde a la corriente talámica ruidosa del modelo [80–82].

En la Figura 3.6 utilizamos el modelo simple de la actividad neuronal para reproducir veinte de las propiedades neurocomputacionales más fundamentales de las neuronas biológicas, cambiando los diferentes parámetros a , b , c , d e I [83]. Dicha figura muestra comportamientos neuronales diferentes luego de considerar las corrientes de entrada: (A) disparo tónico; (B) disparo fásico; (C) ráfagas tónicas; (D) ráfagas tónicas; (E) modelo mixto (ráfagas luego disparos); (F) adaptación en la frecuencia de disparos; (G) excitabilidad de clase 1; (H) excitabilidad de clase 2; (I) latencia de disparos; (J) oscilaciones sub-umbral; (K) preferencia de frecuencia y resonancia; (L) integración y detección de coincidencia; (M) espiga de rebote; (N) ráfaga de rebote; (O) variabilidad de umbral; (P) bi-estabilidad de estados de reposo y disparo; (Q) depolarización post-potencial de acción; (R) acomodación; (S) disparos inducidos por inhibición; y (T) ráfagas inducidas por inhibición. Los valores de parámetros correspondientes para los casos denotados por (A) hasta (T) en la Figura 3.6 están dados en el Cuadro 3.1.

En resumen, cada neurona puede ser también descripta por un modelo simple que permite reproducir varias de las características más fundamentales de las neuronas biológicas [81]. En esta sección, consideramos un modelo de simulación de redes en el cual el número de neuronas interconectadas es un parámetro bajo control, y cada neurona puede estar interconectada aleatoriamente con dos o más neuronas [80–82]. Las contribuciones inhibitorias hiperpolarizan el potencial y lo mueven más lejos del umbral de disparo. En contraste, las contribuciones excitatorias depolarizan el potencial de membrana (esto es, lo llevan más cerca del umbral).

Consideramos además una simulación que tiene en cuenta los disparos de neuronas corticales con retrasos en la conducción axonal y plasticidad dependiente del tiempo de disparo como la regla de aprendizaje sináptica Hebbiana [4]. La mag-

	a	b	c	d		a	b	c	d
A	0,02	0,2	-65	6	K	0,1	0,26	-60	-1
B	0,02	0,25	-65	6	L	0,02	-0,1	-55	6
C	0,02	0,2	-50	2	M	0,03	0,25	-60	4
D	0,02	0,25	-55	0,05	N	0,03	0,25	-52	0
E	0,02	0,2	-55	4	O	0,03	0,25	-60	4
F	0,01	0,2	-65	8	P	0,1	0,26	-60	0
G	0,02	-0,1	-55	6	Q	1	0,2	-60	-21
H	0,2	0,26	-65	0	R	0,02	1	-55	4
I	0,02	0,2	-65	6	S	-0,02	-1	-60	8
J	0,05	0,26	-60	0	T	-0,026	-1	-45	-2

Cuadro 3.1: Los distintos parámetros a , b , c , d para reproducir los veinte comportamientos neurocomputacionales más fundamentales de las neuronas biológicas.

nitudo de los pesos sinápticos entre las neuronas pre-sináptica depende del tiempo relativo de las espigas. El peso de las conexiones sinápticas entre las neuronas pre-a post-sináptica incrementa como $0,1 \times \exp -(t/t_0)$ si la post-neurona dispara luego del disparo presináptico, donde $t_0 = 20 \text{ ms}$ y t es el intervalo entre espigas. De otra forma, si el orden de los disparos es revertido, disminuye como $0,12 \times \exp (t/t_0)$ [82].

Tomamos en cuenta diferentes configuraciones de estímulos cambiando el parámetro a que describe la escala temporal de la variable de recuperación (para la cual siempre tomamos valores $a > 0$). El parámetro a describe la escala temporal de la variable de recuperación u (valores más pequeños resultan en una recuperación más lenta). Para el modelado computacional tomamos diez estímulos diferentes s como diez valores equiespaciados de a que varían desde 0,01 a 0,1 (un valor típico es $a = 0,02$ [80]). Con el fin de realizar nuestra estimación de teoría de la información consideramos 10000 repeticiones, 10 estímulos diferentes y realizados sobre una ventana temporal promedio de 10 ms. Entonces, sobre la base de los datos generados por nuestro modelo computacional, construimos las probabilidades de obtener una respuesta en particular \mathbf{r} igual a 0 o 1 (donde 0 denota eventos sin ningún potencial de acción y 1 que una espiga es disparada en la ventana temporal considerada) condicional al estímulo s descrito previamente.

Sea una red que consiste en tres neuronas, tomando en cuenta dos casos: (i) uno

en el cual las neuronas tengan únicamente entradas comunes entre pares en sus sinapsis excitatorias ($M = 2$), y (ii) otro en el cual se permitan las entradas comunes entre tripletes en las mismas ($M = 3$). La Figura 3.7 muestra las distintas componentes de correlación dependientes del estímulo considerando estas dos redes Hebbianas en donde las sinapsis excitatorias por neurona son iguales a $M = 2$ y $M = 3$, respectivamente. El número de sinapsis excitatorias por neurona fue implementado de la misma manera que en la referencia [82], que es la forma estándar para implementar este tipo de modelado. Así, ilustramos cómo incrementar la conectividad de entrada de la red puede afectar la correlación dependiente del estímulo. Notar que $I_{cor-ch,3}$ contribuye en ambas posibilidades con una cantidad relevante de información. En este caso, los tripletes de espigas por “casualidad” posibles que puedan haber surgido de correlaciones de segundo orden pueden contribuir a una subestimación de la información si no fuesen a ser tenidos en cuenta. En este modelo, las correlaciones dependientes del estímulo de segundo y tercer orden incrementan la cantidad de información total a medida que la conectividad de entrada es aumentada. Además, la conectividad entre tripletes es reflejada en valores mayores de correlaciones estímulo-dependientes.

Los resultados mostrados en este capítulo explicitan un claro efecto de las correlaciones entre tripletes en la transmisión de la información. En el siguiente capítulo, se desarrolla cómo es posible obtener correlaciones espurias entre tripletes de neuronas.

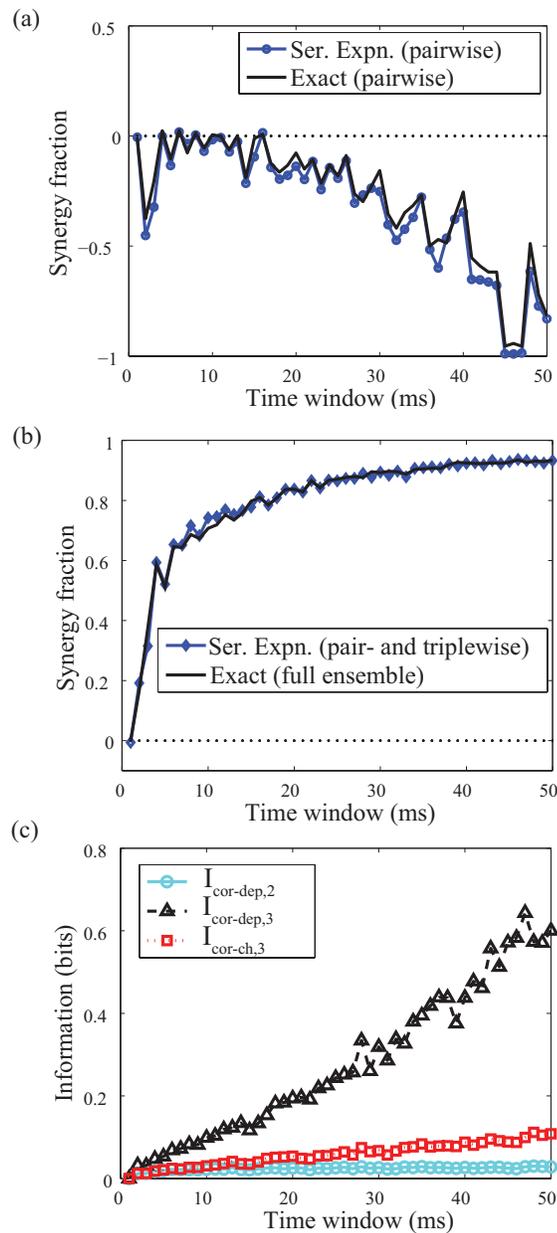


Figura 3.5: (a) Fracción de sinergia ('Synergy fraction') al considerar únicamente correlaciones entre pares como en la Figura 3.1 (c). La línea azul muestra el grado de sinergia usando la información mutua total I obtenida utilizando la expansión en serie, mientras que la línea negra indica uno calculado utilizando la definición dada por la ecuación (2.19). (b) Igual que (a), pero estimando la fracción de sinergia cuando los términos de correlación de tercer orden son incluidos como en la Figura 3.1 (d). (c) Contribuciones de las componentes de correlación dependientes del estímulo.

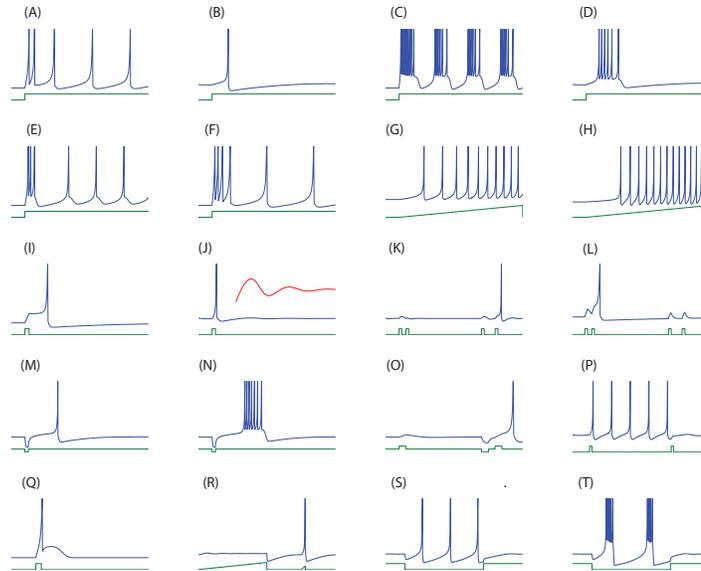


Figura 3.6: (A)-(T): Comportamientos neurocomputacionales relevantes de neuronas biológicas, como presentados en la referencia [81]. La respuesta neuronal (potencial de membrana, mV) y corriente de entrada I se encuentran por debajo de cada sub-gráfico. Notar que el agregado en la sub-figura (J) representa las oscilaciones sub-umbral. Imagen presentada en [81].

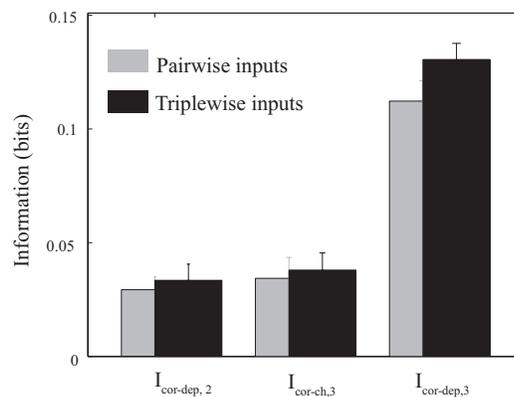


Figura 3.7: Componentes en el modelo de Izhikevich. Las barras grises y negras representan redes con conectividad de entrada entre pares ('Pairwise inputs') y entre tripletes ('Triplewise inputs'), respectivamente.

Capítulo 4

Restricciones de máxima entropía en las correlaciones entre tripletes

El término “correlación espacial” ha sido utilizado para denotar la actividad sincronizada de un grupo de neuronas, mientras que el término “correlación temporal” ha sido usado para indicar cadenas de eventos de disparo a intervalos temporales específicos. A su vez, términos como par o triplete han sido utilizados para denotar patrones espaciotemporales de dos o tres neuronas disparando simultáneamente o en secuencia [84, 85]. Establecer la presencia de dichos patrones no es directo: tres neuronas, por ejemplo, pueden disparar en conjunto más frecuentemente que lo esperado por casualidad sin exhibir una interacción de tercer orden auténtica. Tripletes verdaderos y, en general, correlaciones de alto orden auténticas deben entonces ser distinguidos de interacciones que pueden ser explicados en términos de correlaciones de ordenes menores [86].

Cuando se consideran correlaciones neuronales en secuencias de espigas, existe una limitación en sólo analizar las interacciones entre pares. En un sistema de tres neuronas, por ejemplo, dos tipos de entradas externas pueden generar correlaciones de segundo y tercer orden, con análisis de autocorrelación y correlación cruzada idénticos en ambos casos. En otras palabras, la conectividad entre pares de la entrada de la red puede generar correlaciones de alto orden indistinguibles de aquellas generadas por una conectividad entre tripletes. No sería posible distinguir, en ese escenario, las dos actividades poblacionales distintas a través de análisis de autoco-

relación y correlación cruzada, y las diferencias entre esas dos configuraciones no son capturadas por meras correlaciones de segundo orden. Son las correlaciones de alto orden que determinan si espigas coincidentes de dos neuronas también son coincidentes con las espigas de una tercer célula. A tercer orden, se toma en cuenta la posibilidad de tripletes por “casualidad” que pueden haber surgido de correlaciones entre pares y puedan contribuir a una sub- o sobre-estimación de la información total, como analizado en el capítulo 3.

Como expuesto en el capítulo 2, la geometría de la información [26] permite introducir en la teoría de la información herramientas de geometría diferencial para analizar las correlaciones en patrones de disparo de neuronas. Se utilizan para este fin conceptos basados en la ortogonalidad de parámetros de coordenadas, caracterizando en este contexto el principio de máxima entropía [19]. Se busca caracterizar el sistema en este estado, debido a que la distribución de probabilidad que mejor representa el conocimiento del mismo es aquella con la mayor entropía (en el sentido de la teoría de información). La distribución de probabilidad menos sesgada que se le puede atribuir a un sistema estadístico es aquella en la que, dadas ciertas condiciones fijas, maximiza la entropía. En una situación de desconocimiento de información, la distribución menos sesgada es aquella en que menos información extrínseca al problema contenga. En este marco, estudiaremos las interacciones de hasta tercer orden, obteniendo las restricciones (o vínculos) de máxima entropía en función de las correlaciones definidas en el capítulo 3. De esta manera, calculamos una aproximación para las correlaciones espurias.

4.1. Interacciones neuronales en el marco de la geometría de la información

Introducimos brevemente la aproximación para analizar espigas neuronales [42, 87] en el marco de la geometría de la información, descrito en el capítulo 2. En este enfoque, los modelos logarítmicamente lineales son propuestos para representar las frecuencias de disparos de un conjunto de neuronas y mostrar que los coeficientes no nulos de estos modelos son una medida natural para los disparos sincrónicos. Los patrones de actividad de una población de tamaño N se encuentra representada

como un vector binario $\mathbf{x} = (x_1, \dots, x_N)$, donde $x_i = 0$ si la neurona i se encuentra en silencio en una ventana temporal corta Δt y $x_i = 1$ si dispara. Entonces, para una dada unidad de tiempo, consideramos el conjunto de distribuciones de probabilidad de vectores binarios, $\{P(\mathbf{x})\}$, que consiste de 2^N probabilidades

$$P(\mathbf{x}) = \text{Prob}\{x_1 = i_1, \dots, x_N = i_N\} = p_{i_1 \dots i_N} \quad (4.1)$$

sujetas a la normalización $\sum_{i_1, \dots, i_N=0,1} p_{i_1 \dots i_N} = 1$.

Como propuesto por Amari y colaboradores [26, 42], el conjunto de todas las distribuciones de probabilidad $\{P(\mathbf{x})\}$ forma una variedad $(2^N - 1)$ -dimensional \mathbf{S}_N . Esta aproximación utiliza la ortogonalidad de los parámetros naturales y de valor esperado en la familia de distribuciones exponenciales. Además, resulta práctica para analizar los disparos neuronales de manera sistemática basándose en la teoría de la información. Tales distribuciones de probabilidad pueden ser unívocamente determinadas utilizando un “sistema de coordenadas”.

Un sistema de coordenadas posible es aquel dado por el conjunto de los $2^N - 1$ valores de probabilidad marginales:

$$\eta_i = E[x_i] = \text{Prob}\{x_i = 1\}, \quad i = 1, \dots, N \quad (4.2)$$

$$\eta_{ij} = E[x_i x_j] = \text{Prob}\{x_i = x_j = 1\}, \quad i < j \quad (4.3)$$

⋮

$$\eta_{123 \dots N} = E[x_1 \dots x_N] = \text{Prob}\{x_1 = x_2 = \dots = x_N = 1\}, \quad (4.4)$$

donde $E[\cdot]$ cuantifica el valor esperado. Estas son llamadas coordenadas- $\boldsymbol{\eta}$ [26]. Además, provisto que $P(\mathbf{x}) \neq 0$, cualquier distribución de ese tipo puede ser expandido de la forma

$$\begin{aligned} P(\mathbf{x}) = \exp \left\{ \sum_{i=1}^N x_i \theta_i + \sum_{i < j} x_i x_j \theta_{ij} + \sum_{i < j < k} x_i x_j x_k \theta_{ijk} \right. \\ \left. + \sum_{i < j < k < l} x_i x_j x_k x_l \theta_{ijkl} + \dots + x_1 \dots x_N \theta_{1 \dots N} - \psi \right\}, \quad (4.5) \end{aligned}$$

donde existen un total de $2^N - 1$ coeficientes θ diferentes que pueden ser usados para

determinar unívocamente la distribución de probabilidad. Un significado práctico de los mismos será discutido en los siguientes párrafos. θ forma a un sistema de coordenadas, llamado coordenadas- θ . El parámetro de normalización ψ puede ser también referido como θ_0 , pero puede prestarse a confusión si las neuronas son etiquetadas de forma distinta. La geometría de la información cuantifica medidas de cualquier orden de manera directa, lo que resulta una ventaja de este enfoque. Una vez que θ es definido, se tiene un subconjunto $\mathbf{E}(\theta)$ para cada θ , que es una familia de distribuciones que poseen el mismo valor de θ , y además, el conjunto de todas las distribuciones de probabilidad cuyos marginales sean los mismos se denota como $\mathbf{M}(\eta)$.

En el modelo de neuronas binarias especificado anteriormente, existen 2^N estados posibles para N neuronas. Si todas las neuronas disparan independientemente unas de las otras, la probabilidad de la configuración \mathbf{x} esta dada por $P(x_1, \dots, x_N) = P(x_1) \cdots P(x_N)$. Las neuronas se dicen ser correlacionadas si no disparan independientemente y, por lo tanto, los métodos para detectar correlaciones generalmente buscan desviaciones de este caso. Una correlación entre dos neuronas, la neurona 1 y la neurona 2 por ejemplo, se expresa matemáticamente como $P(x_1, x_2) \neq P(x_1)P(x_2)$. Esto se corresponda con la idea de que dos neuronas están correlacionadas si la distribución de probabilidades conjunta para sus estados no puede ser determinada a partir de las dos distribuciones de probabilidad individuales $P(x_1)$ y $P(x_2)$.

En términos de coordenadas es posible entender además esta descripción, como se puede observar en la Figura 4.1. El subconjunto $\mathbf{E}(\theta)$ es una subvariedad dos-dimensional en donde (η_1, η_2) puede variar libremente, pero la coordenada θ se mantiene fija. Se define el origen $\theta = 0$ en el caso de que no exista correlación (es decir, cuando $\eta_{12} = \eta_1\eta_2$) y entonces, $\mathbf{E}(0)$ es el conjunto de todas las distribuciones independientes. De la misma manera, el subconjunto $\mathbf{M}(\eta_1, \eta_2)$ forma una subvariedad de dimensión uno en \mathbf{S}_2 , donde el par (η_1, η_2) está fijo y θ es libre. La dirección tangencial de $\mathbf{M}(\eta_1, \eta_2)$ representa la dirección en la cual sólo la coordenada de interacción pura cambia, mientras que las direcciones tangenciales a $\mathbf{E}(\theta)$ son las direcciones en las cuales solamente η_1 y η_2 cambian pero θ se mantiene fijo. Luego, se requiere que $\mathbf{E}(\theta)$ y $\mathbf{M}(\eta_1, \eta_2)$ sean ortogonales en cualquier punto, es decir, que las direcciones de cambio de las componentes marginales y de interacción sean

mutuamente ortogonales.

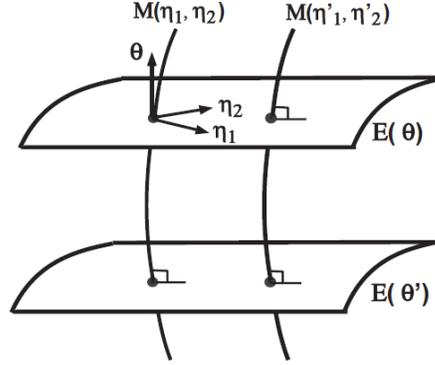


Figura 4.1: Subconjuntos y coordenadas ortogonales (Ilustración adaptada de [26]).

Las direcciones de pequeños cambios en las coordenadas η_i ($i = 1, 2$) y θ se encuentran representadas, respectivamente, por los elementos de la matriz de información de Fisher (ver capítulo 2)

$$\frac{\partial}{\partial \eta_i} l(\mathbf{x}; \eta_1, \eta_2, \theta), \text{ con } i = 1, 2 \quad (4.6)$$

$$\frac{\partial}{\partial \theta} l(\mathbf{x}; \eta_1, \eta_2, \theta) \quad (4.7)$$

donde $l(\mathbf{x}; \eta_1, \eta_2, \theta) = \ln P(\mathbf{x}; \eta_1, \eta_2, \theta)$ es llamada la función verosimilitud. Estas funciones son variables aleatorias que denotan cómo la probabilidad logarítmica varía debido a pequeños cambios en los parámetros en sus respectivas direcciones. Las últimas se dicen ortogonales cuando las variables aleatorias correspondientes no están correlacionadas, y entonces

$$E\left[\frac{\partial}{\partial \eta_i} l(\mathbf{x}; \eta_1, \eta_2, \theta) \frac{\partial}{\partial \eta_i} l(\mathbf{x}; \eta_1, \eta_2, \theta)\right] = 0. \quad (4.8)$$

La distribución de probabilidad de la respuesta del par puede ser caracterizado tanto especificando sus probabilidades individuales en un sistema de coordenadas denominado coordenadas- p , como por sus marginales, en un sistema de coordenadas llamado coordenadas- η . Sin embargo, para poder estudiar como interactúan las neuronas

de una población, es conveniente, bajo la condición de que $P(\mathbf{x}) \neq 0$ para todo \mathbf{x} , expandir $P(\mathbf{x})$ en coordenadas θ de la forma

$$P(\mathbf{x}) = \exp \left\{ \sum_{i=1,2} \theta_i x_i + \theta x_1 x_2 - \psi \right\}, \quad (4.9)$$

donde ψ es la constante de normalización. Dado que x_i toma valores binarios, esta expansión es exacta. Las coordenadas θ son los parámetros naturales en la familia de las probabilidades exponenciales en general.

Debido a que la coordenada θ es definida ortogonal a las coordenadas marginales η_1 y η_2 , se dice que θ representa la interacción pura independiente de los marginales. Cumpliendo con la condición de ortogonalidad, se define de la siguiente forma

$$\theta = \ln \frac{p_{11} p_{00}}{p_{10} p_{01}}. \quad (4.10)$$

El resto de los coeficientes toman los valores

$$\begin{aligned} \theta_1 &= \ln \frac{p_{10}}{p_{00}}, \\ \theta_2 &= \ln \frac{p_{01}}{p_{00}}, \\ \psi &= -\ln p_{00}. \end{aligned} \quad (4.11)$$

4.1.1. Maximización de la entropía

Sea \bar{p} la distribución independiente más cercana (a una “distancia” mínima) a $p \equiv P(\mathbf{x})$, tal que

$$\bar{p} = \operatorname{argmin}_{q \in \mathbf{E}(0)} D[p \parallel q], \quad (4.12)$$

donde q es una distribución de probabilidad cualquiera y D es la divergencia de Kullback-Leibler (ver capítulo 2). Las coordenadas de p son las coordenadas mixtas (η_1, η_2, θ) y las coordenadas de \bar{p} están dadas por $(\eta_1, \eta_2, 0)$, tal que $\bar{p} = P(x_1; \eta_1)P(x_2; \eta_2)$, siendo $p(x_i; \eta_i)$ la distribución marginal de \bar{p} .

Sea p_0 la distribución uniforme cuyas coordenadas mixtas son, por ejemplo,

$(1/2, 1/2, 0)$, luego la distribución pertenece a $\mathbf{E}(0)$. Entonces,

$$\bar{p} = \operatorname{argmin}_{q \in \mathbf{M}(\eta_1, \eta_2)} D[q \parallel p_0]. \quad (4.13)$$

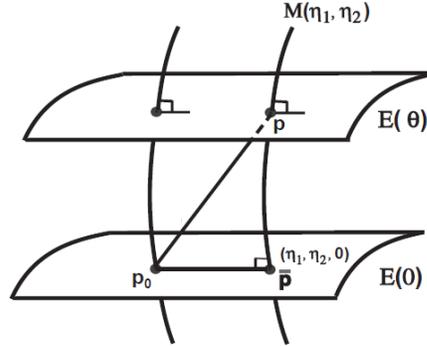


Figura 4.2: Distribuciones de probabilidad en la variedad (Ilustración adaptada de [26]).

Luego, teniendo en cuenta que la divergencia entre dos distribuciones de probabilidad está dada por la entropía relativa, se obtiene que

$$D[q \parallel p_0] = H(p_0) - H(q). \quad (4.14)$$

$H(q)$ es la entropía de q y $H_0 = 2 \log_2 2$ (para la distribución ejemplo $(1/2, 1/2, 0)$) es constante, ya que corresponde a la distribución de probabilidad uniforme. Reescribiendo, se concluye que para maximizar la entropía

$$H(q) = H(p_0) - D[q \parallel p_0], \quad (4.15)$$

debe ocurrir que $q \equiv \bar{p}$, debido a que $H(\bar{p}) = H(p_0) - D[\bar{p} \parallel p_0]$ y \bar{p} minimiza esta “distancia” entre aquellas distribuciones pertenecientes a $\mathbf{M}(\eta_1, \eta_2)$. En conclusión, se vincula la entropía máxima con ausencia de correlación, es decir que establecer $\theta = 0$ es análogo al principio de máxima entropía en el marco de la geometría de la información.

4.1.2. Interacciones entre tripletes

Consideremos ahora un conjunto de tres neuronas, rotuladas como las neuronas 1, 2 y 3. Extender la idea de correlación previamente descrita introduce algunas complicaciones, además de que no resulta posible una ilustración del caso como para dos neuronas. No es simplemente suficiente comparar la probabilidad conjunta $P(x_1, x_2, x_3)$ con el producto $P(x_1)P(x_2)P(x_3)$ y declarar la existencia de un triplete cuando las dos no son iguales. Esto se prestaría a confundir a los tripletes “auténticos” con pares o disparos independientes solapados. Por ende, las neuronas 1, 2, y 3 pueden llegar disparar más frecuentemente que lo predicho por el modelo independiente debido a que, por ejemplo, las neuronas 1 y 2 se encuentran involucradas en una interacción binaria. Las distribuciones de probabilidad para las tres configuraciones de pares posibles son $P(x_1, x_2)$, $P(x_2, x_3)$, y $P(x_1, x_3)$. Una interacción “genuina” de tercer orden entre las tres neuronas ocurriría si fuese imposible determinar la distribución conjunta para esta configuración de células utilizando sólo información de estas distribuciones de segundo orden. Una forma canónica de construir una distribución conjunta de tres neuronas a partir de las distribuciones de los pares es la de maximizar la entropía sujeta a la restricción que las distribuciones marginales de cada par se encuentran dadas por $P(x_1, x_2)$, $P(x_2, x_3)$, y $P(x_1, x_3)$. La distribución resultante es aquella que adiciona la menor información más allá de la contenida en las distribuciones conjuntas de los pares. Por lo tanto, en el marco de la geometría de la información podemos escribir la distribución conjunta obtenida $P(\mathbf{x})$ entonces como

$$P(x_1, x_2, x_3) = \exp \left\{ \sum_{i=1,2,3} x_i \theta_i + \sum_{i < j} x_i x_j \theta_{ij} - \psi \right\}. \quad (4.16)$$

En este modelo, existe un parámetro para cada neurona individual y para cada par de neuronas. Aquí los parámetros θ pueden ser pensados como coeficientes que miden la tendencia de la neurona (o las neuronas) a las que están asociados de estar activas. Incrementando el valor de θ_i sin cambiar cualquier otro θ aumenta la probabilidad de que la neurona i se encuentre en su estado activo. Análogamente, incrementar el valor de θ_{ij} de la misma manera favorece el disparo simultáneo de las neuronas i y j . Se debe notar que existe una correlación de segundo orden entre

las neuronas i y j en el sentido de las definiciones previas precisamente cuando $\theta_{ij} \neq 0$. El parámetro de normalización ψ representa la probabilidad de que todas las neuronas se encuentren inactivas en simultáneo. Sin embargo, podemos incluir además un parámetro adicional en la ecuación (4.5), esto es,

$$P(x_1, x_2, x_3) = \exp \left\{ \sum_{i=1,2,3} x_i \theta_i + \sum_{i < j} x_i x_j \theta_{ij} + x_1 x_2 x_3 \theta_{123} - \psi \right\}. \quad (4.17)$$

Manteniendo todos los otros coeficientes fijos e incrementando el valor de θ_{123} aumenta las probabilidades de que las tres neuronas disparen simultáneamente. La ecuación (4.16) corresponde al caso especial de 4.17 en donde θ_{123} es igual a cero, de la misma manera que la independencia corresponde al caso especial de la ecuación (4.16) en donde todos los términos de segundo orden θ_{ij} son iguales a cero. Por lo tanto, este marco de trabajo proveería los medios para diferencias correlaciones de alto orden (en este caso, de tercer orden) “puras” (o verdaderas) de correlaciones “por casualidad” (introducidas en el capítulo 3).

La distribución de probabilidad conjunta $P(\mathbf{x})$, con $\mathbf{x} = (x_1, x_2, x_3)$, está dada por

$$p_{abc} = \text{Prob}\{x_1 = a; x_2 = b; x_3 = c\},$$

donde a , b y c pueden ser tanto 0 como 1. De las ocho probabilidades posibles, las cuales son denotadas por p_{111} , p_{100} , p_{010} , p_{001} , p_{110} , p_{101} , p_{011} , p_{000} , solamente siete de ellas son libres, y por ende el conjunto de distribuciones de x forma una variedad de dimensión siete S_7 . Como ocurre en el caso general de N neuronas, la distribución $P(\mathbf{x})$ puede ser descompuesta en componentes marginales (η) y de interacción (θ).

Las distribuciones marginales de x_i se encuentran especificadas por las cantidades

$$\eta_i = E[x_i] = \text{Prob}\{x_i = 1\}, \quad i = 1, \dots, N \quad (4.18)$$

$$\eta_{ij} = E[x_i x_j] = \text{Prob}\{x_i = x_j = 1\}, \quad i < j \quad (4.19)$$

$$\vdots$$

$$\eta_{123} = E[x_1 x_2 x_3] = \text{Prob}\{x_1 = x_2 = x_3 = 1\}. \quad (4.20)$$

Además, por definición, las componentes marginales son

$$\begin{aligned}
 \eta_1 &= p_{100} + p_{110} + p_{101} + p_{111} \\
 \eta_2 &= p_{010} + p_{110} + p_{011} + p_{111} \\
 \eta_3 &= p_{001} + p_{101} + p_{011} + p_{111} \\
 \eta_{12} &= p_{110} + p_{111} \\
 \eta_{13} &= p_{101} + p_{111} \\
 \eta_{23} &= p_{011} + p_{111} \\
 \eta_{123} &= p_{111}.
 \end{aligned}$$

A su vez, el coeficiente de interacción entre el triplete para este modelo está dado por

$$\theta_{123} = \ln \frac{p_{111}p_{100}p_{010}p_{001}}{p_{110}p_{101}p_{011}p_{000}}, \quad (4.21)$$

el cual resulta ortogonal a cualquier variación en los marginales individuales y de segundo orden. El resto de los coeficientes toman los siguientes valores

$$\begin{aligned}
 \theta_1 &= \ln \frac{p_{100}}{p_{000}} \\
 \theta_2 &= \ln \frac{p_{010}}{p_{000}} \\
 \theta_3 &= \ln \frac{p_{001}}{p_{000}} \\
 \theta_{12} &= \ln \frac{p_{110}p_{000}}{p_{100}p_{010}} \\
 \theta_{13} &= \ln \frac{p_{101}p_{000}}{p_{100}p_{001}} \\
 \theta_{23} &= \ln \frac{p_{011}p_{000}}{p_{010}p_{001}} \\
 \psi &= -\ln p_{000}.
 \end{aligned}$$

Todas las cantidades previamente especificadas fueron definidas en [42].

Alternativamente, podríamos utilizar una notación adicional como el superíndice (\mathcal{O}, N) de estos coeficientes para representar el orden del modelo logarítmicamente lineal \mathcal{O} y el número de neuronas de la red N [88, 89]. En caso previo, los coeficientes

de correlación podrían ser nombrados como

$$\boldsymbol{\theta}^{(3,3)} = (\theta_1^{(3,3)}, \theta_2^{(3,3)}, \theta_3^{(3,3)}; \theta_{12}^{(3,3)}, \theta_{13}^{(3,3)}, \theta_{23}^{(3,3)}; \theta_{123}^{(3,3)}). \quad (4.22)$$

Esto resulta de interés dado que trabajos previos [90] han demostrado que si el número de neuronas en una red es conocido, el modelo logarítmicamente lineal de ese mismo orden provee una estimación más precisa de la fuerza de conexión. Es decir, para el sistema de tres neuronas descrito previamente, $\theta_{12}^{(3,3)}$ proveería una mejor estimación del parámetro de interacción entre el par que la medida

$$\theta_{12}^{(2,3)} = \ln \frac{p_{11*} p_{00*}}{p_{10*} p_{01*}}, \quad (4.23)$$

donde $*$ representa la marginalización sobre la tercer neurona (“escondida”). Por ende, $p_{x_1 x_2 *}$ constituiría una distribución de probabilidad marginal de x_1 y x_2 en este sistema.

Imponiendo $\theta_{123} = 0$ (considerando una extensión del principio de máxima entropía hasta tercer orden), es posible obtener

$$\eta_{123} = \frac{(\eta_{12} - \eta_{123})(\eta_{13} - \eta_{123})(\eta_{23} - \eta_{123})}{(\eta_1 - \eta_{12} - \eta_{13} + \eta_{123})(\eta_2 - \eta_{12} - \eta_{23} + \eta_{123})(\eta_1 - \eta_{13} - \eta_{23} + \eta_{123})} \times (1 - \eta_1 - \eta_2 - \eta_3 + \eta_{12} + \eta_{13} + \eta_{23} - \eta_{123}) \quad (4.24)$$

Esta cantidad también ha sido referida como “esfuerzo” [54], debido a que cuantifica cuanto es presionado un modelo de máxima entropía de segundo orden para ajustarse a un sistema que presenta interacciones entre tripletes. Esta medida identifica un patrón consistente a estas interacciones: cuando las interacciones de tercer orden se encuentran presentes, el esfuerzo es casi siempre negativo, indicando que existe menos sincronía que la que fuese esperada de las interacciones entre pares solamente. Es decir, cuando los patrones de actividad observados difieren de aquellos predichos por un modelo de segundo orden, esta cantidad indica la naturaleza de esta diferencia, tanto si existe exceso o deficiencia de sincronización. En comparación a la divergencia Kullback-Leibler (ver capítulo 2), la diferencia más importante entre estas dos medidas es que, dado que es una coordenada y no una distancia, puede ser negativa. Por lo tanto, puede indicarnos no solo cuándo el modelo de segundo orden falla sino

también cómo falla.

4.2. Cuantificación de correlaciones entre tripletes espurias

En el capítulo 3 estudiamos las contribuciones de tercer orden y el efecto de las correlaciones espurias entre tripletes en la transmisión de la información por secuencias de potenciales de acción para duraciones temporales cortas. De esta manera, utilizando las definiciones para las probabilidades en esta aproximación, dadas por las ecuaciones (3.8), (3.9), (3.10) y (3.11) es posible obtener una medida del mejor estimador para la correlación de ruido entre tripletes que proviene únicamente de las interacciones de menor orden. Resolviendo la ecuación (4.24) para un sistema de tres neuronas (rotuladas como 1, 2 y 3) se obtiene la ecuación de tercer orden (en correlación de ruido espuria $\gamma_{123}^{ch}(s)$) :

$$\begin{aligned}
 0 = & (1 + \gamma_{123}^{ch}(s))^3 \prod_{i=1}^3 \bar{r}_i(s)t + (1 + \gamma_{123}^{ch}(s))^2 \left\{ 3 \prod_{i=1}^3 \bar{r}_i(s)t \left[\sum_{i<j=1}^3 (1 + \gamma_{ij}(s)) \right] \right. \\
 & \left. + \sum_{i<j=1}^3 \bar{r}_i(s)\bar{r}_j(s)t^2 [(2(\bar{r}_i(s) + \bar{r}_j(s))t - 1)(1 + \gamma_{ij}(s)) - 1] \right\} \\
 & + (1 + \gamma_{123}^{ch}(s)) \left\{ 1 - 3 \prod_{i=1}^3 \bar{r}_i(s)t \prod_{i<j=1}^3 (1 + \gamma_{ij}(s)) + \left[1 - \sum_{i=1}^3 \bar{r}_i(s)t \right] \right. \\
 & \times \left[\sum_{i \neq j \neq 1}^3 \bar{r}_i(s)t(1 + \gamma_{ij}(s))(1 + \gamma_{ik}(s)) \right] + \left[\sum_{i<j=1}^3 (1 + \gamma_{ij}(s)) \right] \\
 & \times \left[\sum_{i<j=1}^3 \bar{r}_i(s)\bar{r}_j(s)t^2(1 + \gamma_{ij}(s)) \right] + \sum_{i \neq j \neq 1}^3 \bar{r}_i(s)t [(1 + \gamma_{ij}(s)) + (1 + \gamma_{ik}(s))] \left. \right\} \\
 & - \left[\prod_{i<j} (1 + \gamma_{ij}(s)) \right] \left[1 - \sum_{i=1}^3 \bar{r}_i(s)t + \sum_{i<j=1}^3 \bar{r}_i(s)\bar{r}_j(s)t^2(1 + \gamma_{ij}(s)) \right].
 \end{aligned} \tag{4.25}$$

Es importante remarcar que $r_i(s)$ es la tasa de disparo de la neurona i sujeta al estímulo s presente, y que $\bar{\cdot}$ representa el promedio con respecto a las repeticiones de la prueba. Como definido en el capítulo 3, la correlación de ruido de segundo orden puede ser calculada de la forma dada por la ecuación (3.1).

4.2.1. Caso de triplete homogéneo

Para poder interpretar las soluciones posibles, es conveniente estudiar el caso homogéneo y obtener estimaciones analíticas. Esto es, el caso en el cual las tres neuronas disparan con la misma frecuencia y las correlaciones entre pares son iguales. A pesar de no ser el escenario más interesante, ni el más probable, otorga una exploración útil de la ecuación (4.25).

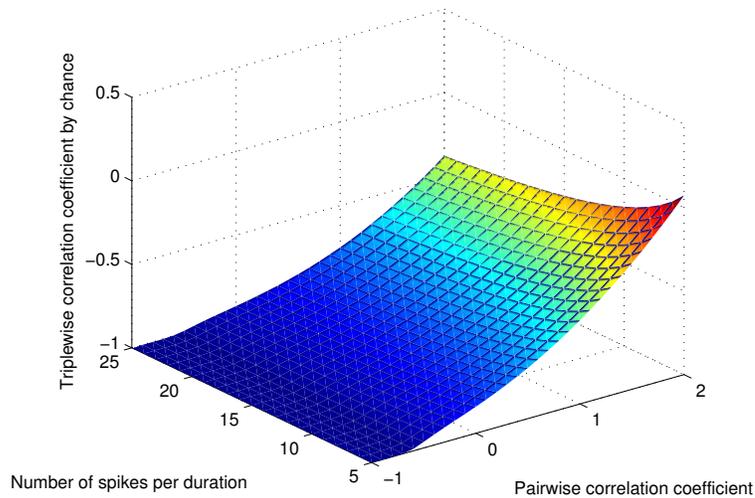


Figura 4.3: Valores numéricos para las correlaciones espurias entre tripletes (‘Triplewise correlation coefficient by chance’) idénticos para rangos comunes de tasas de disparo (‘Number of spikes per duration’) y correlaciones de ruido entre pares (‘Pairwise correlation coefficient’).

La Figura 4.3 ilustra los valores correspondientes a las correlaciones espurias (o “por casualidad”) entre tripletes para rangos comunes de tasas de disparo y correlaciones de ruido entre pares. Estos resultados se encuentran en concordancia con la literatura previa [10, 12]. Esta figura muestra la fuerte dependencia de esta medida con las interacciones entre pares, incluso para valores pequeños.

Sin embargo, los resultados expuestos son válidos para poblaciones neuronales pequeñas. En los capítulos siguientes desarrollaremos otra metodología complementaria para poder estudiar poblaciones más grandes de neuronas.

Capítulo 5

Correlaciones de alto orden en una población homogénea

En términos neurofisiológicos, las neuronas individuales se consideran que hacen contribuciones pequeñas y comprensibles al comportamiento animal [17]. No obstante, la mayor parte de los comportamientos involucran un gran número de neuronas, miles o incluso millones [91]. Dado que estas neuronas puedan estar conectadas a otras neuronas mediante hasta 1000 trillones de conexiones sinápticas, las formas en que la información pueda ser transmitida entre ellas es extremadamente compleja. Estas neuronas se encuentran organizadas en capas o regiones, tal que las neuronas cercanas tienen propiedades similares en sus respuestas. Alrededor de la mitad de las entradas excitatorias a cualquier neurona provienen de neuronas cercanas que caen aproximadamente dentro de un cilindro de un radio de entre 100-200 μm , estructurado como una columna frecuentemente denominada “mini-columna” [92–94]. Esto sugiere que las neuronas corticales reciben abundantes entradas excitatorias y se encuentran embebidas en una red de señales altamente convergentes. Debido a que estas redes son de naturaleza recurrente, es probable que dichas neuronas reciban entradas similares y emitan potenciales de acción bajo condiciones similares. Esto significa que es esperable que las condiciones que producen una respuesta en cualquier neurona involucren la actividad considerable de un número grande de sus entradas más allá de las estadísticas de segundo orden. Por lo tanto, es razonable pensar que gran cantidad de espigas de entrada lleguen en sincronía dentro de una

ventana temporal pequeña.

Con esta idea en mente, Amari y colaboradores desarrollaron un modelo matemáticamente simple en donde las pequeñas contribuciones comunes que excitan a una neurona son sumadas para generar un disparo [31]. Sin tener en cuenta una dependencia de estímulo particular y consideradas independientes, siguiendo el teorema del límite central, contribuyen a una dinámica de caminante aleatorio del potencial de membrana de la neurona. Por ende, se trata de un modelo que consiste de una variable binaria cuyo proceso latente sigue una distribución normal.

Por otra parte, las interacciones de alto orden entre neuronas, es decir aquellas que no pueden ser reducidas a correlaciones entre pares, han sido identificadas en la activación neuronal simultánea. En este capítulo, proveemos un modelo simple matemáticamente tratable capaz de considerar correlaciones de alto orden en la distribución conjunta de disparos de una población neuronal. En nuestro modelo, las correlaciones entre neuronas surgen de entradas q -gaussianas en neuronas con umbral. Por ende, es una extensión del modelo propuesto por Amari [31], donde las entradas al modelo son distribuidas gaussianamente y por lo tanto, no poseen interacciones más allá del segundo orden. Nuestro formalismo se basa en avances recientes hechos en el teorema extendido del límite central, y entonces usando herramientas matemáticas de la mecánica estadística no extensiva [43, 95–100] proveemos un enfoque que cuantifica el grado de interacciones de alto orden. Presentamos una solución analítica exacta de la distribución conjunta de disparos incluyendo patrones de correlaciones neuronales de todos los ordenes a través de una población. Esto es, estimamos por medio de un modelo analíticamente soluble la cantidad de correlación de orden mayor a dos en un “pool” (es decir, un conjunto homogéneo) de neuronas a través de una aplicación directa de una distribución q -gaussiana de las entradas sinápticas comunes proveyendo la expresión para la distribución conjunta, la entropía y la información de Fisher. Testeamos la robustez de nuestro método usando un conjunto simulado de neuronas correlacionadas e independientes. Utilizando nuestro modelo, investigamos distintas soluciones analíticas al considerar tres distribuciones típicas: concentrada, ampliamente desparramadas y bimodales. Estudiamos las propiedades emergentes de la información de Fisher y la entropía en una población grande de neuronas, y mostramos su impacto en la eficiencia de la

codificaciones poblacional.

5.1. Propuesta del modelo extendido

Representamos la actividad de las neuronas en una población de tamaño N mediante un vector binario $\mathbf{x} = (x_1, \dots, x_N)$, donde $x_i = 0$ si la neurona i se encuentra inactiva en alguna ventana temporal Δt y $x_i = 1$ si está disparando una espiga. Sea la distribución de probabilidad de esos vectores binarios, $\{P(\mathbf{x})\}$ que consiste de 2^N probabilidades

$$P(\mathbf{x}) = \text{Prob}\{x_1 = i_1, \dots, x_N = i_N\} = p_{i_1 \dots i_N} \quad (5.1)$$

sujetas a la normalización

$$\sum_{i_1, \dots, i_N=0,1} p_{i_1 \dots i_N} = 1. \quad (5.2)$$

Se denomina “límite termodinámico” cuando el número de partículas consideradas en un sistema alcanza el límite $N \rightarrow \infty$ y el volumen del sistema también crece en proporción al número de partículas. El límite termodinámico es aproximado asintóticamente en la mecánica estadística usando el teorema del límite central [101]. Dicho teorema asegura que la función de distribución de probabilidad de cualquier cantidad medible es una distribución gaussiana normal, siempre y cuando sean consideradas un número suficientemente grande de variables aleatorias independientes con exactamente la misma media y varianza. Importantemente, el teorema del límite central no se sostiene si las correlaciones entre las variables aleatorias no pueden ser despreciadas.

En particular, el teorema del límite central ha sido utilizado por Amari y colaboradores [31] con el fin de obtener el modelo de entradas correlacionadas para estimar la distribución conjunta de probabilidades de disparo en una población neuronal considerando el límite de un número muy grande de neuronas. En su enfoque, las correlaciones entre pares son cuantificadas a través de la covarianza $\langle u_i u_j \rangle$ de la suma pesada de entradas gaussianas (debido al teorema del límite central) u_i y u_j de dos pares de neuronas dadas of neurons ($i \neq j$, $i = 1 \dots N$ y $j = 1 \dots N$). Esto es $u_i = \sum_{j=1}^m w_{ij} - h$, donde w_{ij} es es peso de conexión desde la entrada j -ésima a la i -ésima neurona (donde $h = E[u_i]$ denota la media).

Sin embargo, en publicaciones recientes [97–100], se ha propuesto una generalización del teorema del límite central para el caso cuando un sistema con variables aleatorias débil o fuertemente correlacionadas está siendo considerado. Si juntamos un número suficientemente grande de dichos sistemas, la distribución de probabilidad convergerá a una distribución q -gaussiana. Este hecho se encuentra en acuerdo con teoremas recientemente probados por Amari y Ohara [44], que permiten la introducción de una estructura q -geométrica a una familia de distribuciones de probabilidad arbitraria, y garantiza que la familia de todas las distribuciones de probabilidad pertenecen a la familia de distribuciones q -exponenciales.

Utilizaremos la “extensión natural” del teorema del límite central propuesto en [100], que toma en cuenta los casos en donde las correlaciones entre variables aleatorias no son despreciables. Esto resulta en las llamadas distribuciones q -gaussianas (en lugar de gaussianas) como las funciones de densidad de probabilidad límites:

$$\mathcal{G}_q(x; \beta) = \frac{\sqrt{\beta}}{C_q} \exp_q(-\beta x^2) \quad (5.3)$$

donde \exp_q es definida como la función q -exponencial [96]

$$\exp_q(x) = \begin{cases} \exp(x) & \text{if } q = 1 \\ [1 + (1 - q)x]^{\frac{1}{1-q}} & \text{si } q \neq 1 \text{ y } 1 + (1 - q)x > 0 \\ 0 & \text{si } q \neq 1 \text{ y } 1 + (1 - q)x \leq 0 \end{cases} \quad (5.4)$$

donde el factor C_q es

$$C_q = \begin{cases} \frac{2\sqrt{\pi}\Gamma(\frac{1}{1-q})}{(3-q)\sqrt{1-q}\Gamma(\frac{3-q}{2q-2})} & \text{para } -\infty < q < 1, \\ \sqrt{\pi} & \text{para } q = 1, \\ \frac{\sqrt{\pi}\Gamma(\frac{3-q}{2q-2})}{\sqrt{1-q}\Gamma(\frac{1}{1-q})} & \text{para } 1 < q < 3. \end{cases} \quad (5.5)$$

donde β es un número real positivo y q es un índice real positivo (problema-dependiente). Consideramos en principio el caso $1 < q < 3$. El límite de $q = 1$ se recupera una distribución gaussiana normal, ya que $\lim_{N \rightarrow \infty} (1 + \frac{1}{N})^N = e$, que puede ser reescrito como $\lim_{q \rightarrow 1} (1 + (1 - q))^{\frac{1}{(1-q)}} = e$. En otras palabras, el teorema del límite central

es recuperado cuando $q \rightarrow 1$.

Consideremos ahora la probabilidad de tener exactamente $k = N \cdot r$ neuronas disparando (y entonces $r = \frac{k}{N}$, siendo r la tasa de disparo poblacional) dentro de una dada ventana temporal ΔT dentro de una población de N neuronas [31, 74]. En el marco del modelo de población homogénea tenemos que

$$P_r(r = \frac{k}{N}) = P_r\{x_1 = x_2 = \dots = x_k = 1, x_{k+1} = \dots = x_N = 0\}, \quad (5.6)$$

donde la neurona i se encuentra sujeta a la suma pesada de entradas u_i , por lo tanto $x_i = 1$ si y sólo si $u_i > 0$ y $x_i = 0$ si $u_i \leq 0$. Siguiendo [74], el pool neuronal recibe entradas comunes correlacionadas más allá del segundo orden s_1, s_2, \dots, s_M , y u_i es pesada por de la forma $u_i = \sum_{j=1}^M w_{ij}s_j - h$, donde w_{ij} son pesos de conexión aleatoriamente asignados. Considerando que las u_i se encuentran sujetas a una distribución q -gaussiana (debido al teorema extendido del límite central) $N_q(-h, 1)$, definimos en analogía a [31] $u_i = \sqrt{1 - \alpha} v_i + \sqrt{\alpha} \varepsilon - h$, para $i = 1, \dots, N$. Tomamos $\alpha = E_q[u_i u_j]$ como la q -varianza, $h = E_q[u_i]$ como la q -media, y dos variables aleatorias, una q -gaussiana v_i sujeta a $N_q(0, 1)$ y una gaussiana ε sujeta a $N(0, 1)$.

La distribución conjunta de disparos puede entonces ser estimada utilizando una aproximación de “punto silla de montar” [102] (también conocida como aproximación de Laplace), de la forma

$$Q_q(r) = \sqrt{\frac{1}{r(1-r)|z_q''(\varepsilon_0)|}} \frac{1}{\sqrt{2\pi}} \exp[Nz_q(\varepsilon_0) - \frac{\varepsilon_0^2}{2}], \quad (5.7)$$

donde $z_q(\varepsilon) = r \log(\frac{F_q(\varepsilon)}{r}) + (1-r) \log(\frac{1-F_q(\varepsilon)}{1-r})$. En el contexto de la aproximación de punto silla de montar: $\varepsilon_0 = \arg \max_{\varepsilon \in \mathbb{R}} [z_q(\varepsilon)]$ and $\frac{dz(\varepsilon)}{d\varepsilon} = 0$. La solución es $\varepsilon_0 = F_q^{-1}(r)$, que implica $r = F_q(\varepsilon_0)$, donde r varía en el intervalo $[0, 1]$ y ε_0 está definida para todos los números reales. Adicionalmente, $\varepsilon_0 = F_q^{-1}(r)$ depende del grado de correlación de la arquitectura de la red, que es cuantificada por q .

Notar que las funciones $F_q(\varepsilon)$ y $z_q(\varepsilon)$ se encuentran definidas como en la referencia

[74]:

$$\begin{aligned}
 F_q(\varepsilon) &\equiv P_r(u > 0|\varepsilon) = P_r(u_i > \frac{h - \sqrt{\alpha}\varepsilon}{\sqrt{1-\alpha}}) \\
 &= \frac{1}{\sqrt{2\pi}} \int_{\frac{h-\sqrt{\alpha}\varepsilon}{\sqrt{1-\alpha}}}^{\infty} \exp_q\left(-\frac{v^2}{2}\right) dv
 \end{aligned} \tag{5.8}$$

y

$$z_q(\varepsilon) = r \ln \left[\frac{F_q(\varepsilon)}{r} \right] + (1-r) \ln \left[\frac{1-F_q(\varepsilon)}{1-r} \right]. \tag{5.9}$$

Luego, la ecuación (5.8) se puede escribir como

$$F_q(\varepsilon) = \frac{1}{2\sqrt{2\pi}\sqrt{\frac{q-1}{2}}} B\left(\frac{1}{1+\xi_0(\varepsilon)}, \frac{1}{q-1}, \frac{1}{2}\right). \tag{5.10}$$

donde $B(\cdot)$ es la función Beta y

$$\xi_0(\varepsilon) = \frac{(q-1)(h-\sqrt{\alpha}\varepsilon)^2}{2(1-\alpha)}. \tag{5.11}$$

Si consideramos el límite en el marco de trabajo del teorema del límite central (es decir, $q = 1$), la ecuación (5.10) se reduce a

$$F_{q=1}(\varepsilon) = \frac{1}{2} \text{Erfc}\left(\frac{1}{\sqrt{2}} \frac{h-\sqrt{\alpha}\varepsilon}{\sqrt{1-\alpha}}\right) \tag{5.12}$$

donde $\text{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-t^2) dt$ denota la función de error complementaria. Sin embargo, si el efecto de las correlaciones de orden mayor a dos no es despreciable, entonces, de acuerdo al teorema del límite central extendido, q debe ser mayor a 1. Es posible testear por la presencia de correlaciones de alto orden midiendo la distribución de actividad de registros de unidad múltiple y ajustando q , que representa la presencia de correlaciones de alto orden en la distribución de disparos. Uno puede mostrar por simple comparación cuán estadísticamente distinta del caso $q = 1$ la distribución medida es.

5.1.1. ¿Cómo podemos interpretar el teorema del límite central extendido en el marco teórico de la geometría de la información?

Por un lado, la representación completa de cada parámetro de interacción (incluyendo órdenes mayores a dos) planteada por la ecuación (4.5), consiste en utilizar $2^N - 1$ coeficientes. Por otro lado, debido a que nuestra extensión se logra con la inclusión de un único parámetro adicional, el parámetro de deformación q , el modelado de la población neuronal se mantiene en un espacio de dimensión baja.

Para relacionar estos dos enfoques, proveemos una explicación intuitiva sobre la conexión entre el teorema extendido del límite central y el marco teórico que provee la geometría de la información. Consideremos una distribución q -gaussiana de x , con media μ y varianza σ . En la literatura, las últimas pueden ser también encontradas como q -media y q -varianza. El coeficiente ψ denota una constante de normalización. Entonces,

$$\begin{aligned} \mathcal{G}_q(x; \mu, \sigma) &= \exp_q \left[-\beta \left(\frac{x - \mu}{\sigma} \right)^2 - \psi \right] \\ &= \left\{ 1 + (q - 1) \left[\beta \left(\frac{x - \mu}{\sigma} \right)^2 - \psi \right] \right\}^{\frac{1}{1-q}}, \end{aligned} \quad (5.13)$$

puede ser reescrita como

$$\mathcal{G}_q(x; \mu, \sigma) = \exp \left\{ \frac{1}{1-q} \ln \left[1 + (q - 1) \left(\beta \left[\frac{x - \mu}{\sigma} \right]^2 - \psi \right) \right] \right\} \quad (5.14)$$

y utilizando la definición en series para la función logarítmica

$$\ln(1 + x) = - \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n \quad (5.15)$$

luego

$$\begin{aligned}
 \mathcal{G}_q(x; \mu, \sigma) &= \exp \left\{ - \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} (q-1)^{n-1} \left[\beta \left(\frac{x-\mu}{\sigma} \right)^2 - \psi \right]^n \right\} \\
 &= \exp \left[-\beta \left(\frac{x-\mu}{\sigma} \right)^2 \right] \\
 &\quad \times \exp \left\{ \psi - \sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{n} (q-1)^{n-1} \left[\beta \left(\frac{x-\mu}{\sigma} \right)^2 - \psi \right]^n \right\}.
 \end{aligned} \tag{5.16}$$

Esta expresión puede ser reacomodada como

$$\mathcal{G}_q(x; \mu, \sigma) = \exp \left\{ - \left[\beta \left(\frac{x-\mu}{\sigma} \right)^2 \right] + \tilde{\Psi}(x, q) \right\}, \tag{5.17}$$

donde

$$\tilde{\Psi}(x, q) = \psi - \sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{n} (q-1)^{n-1} \left[\beta \left(\frac{x-\mu}{\sigma} \right)^2 - \psi \right]^n. \tag{5.18}$$

Por ende, esto sugiere que en el marco teórico del teorema extendido del límite central las correlaciones son tomadas en cuenta a través del parámetro q . Este coeficiente actúa entonces como un “ruido de color” en $\tilde{\Psi}(x, q)$.

5.2. Resultados

5.2.1. Distribución de disparos asintótica

En lo siguiente, obtendremos la expresión analítica exacta de la distribución descrita previamente. Para hacerlo, primero tomemos la derivada de $F_q(\varepsilon)$ con respecto

a ε . Usando la regla de la cadena para la derivada, podemos escribir

$$F'_q(\varepsilon) = \frac{1}{2\sqrt{2\pi}\sqrt{\frac{q-1}{2}}} \frac{dB\left(\frac{1}{1+\xi_0(\varepsilon)}; \frac{1}{q-1}, \frac{1}{2}\right) d\left(\frac{1}{1+\xi_0(\varepsilon)}\right)}{d\left(\frac{1}{1+\xi_0(\varepsilon)}\right) d\varepsilon}. \quad (5.19)$$

Considerando que

$$\frac{d}{dz}B(z; a, b) = (1-z)^{b-1}z^{a-1}, \quad (5.20)$$

luego

$$\frac{dB\left(\frac{1}{1+\xi_0(\varepsilon)}; \frac{1}{q-1}, \frac{1}{2}\right)}{d\left(\frac{1}{1+\xi_0(\varepsilon)}\right)} = \left[1 - \frac{1}{1+\xi_0(\varepsilon)}\right]^{-\frac{1}{2}} \left[\frac{1}{1+\xi_0(\varepsilon)}\right] \left(\frac{1}{q-1}\right)^{-1} \quad (5.21)$$

y

$$\frac{d\left(\frac{1}{1+\xi_0(\varepsilon)}\right)}{d\varepsilon} = \frac{(q-1)\sqrt{\alpha}(h-\sqrt{\alpha}\varepsilon)}{(1-\alpha)[1+\xi_0(\varepsilon)]^2}. \quad (5.22)$$

Por lo tanto, reescribimos

$$F'_q(\varepsilon) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{\alpha}{1-\alpha}} [1+\xi_0(\varepsilon)]^{\left(\frac{q-2}{q-1} - \frac{3}{2}\right)} \quad (5.23)$$

Tomando la primera derivada de $z_q(\varepsilon)$ con respecto a ε ,

$$z'_q(\varepsilon) = \left\{ \frac{r}{F_q(\varepsilon)} - \frac{1-r}{1-F_q(\varepsilon)} \right\} \quad (5.24)$$

y derivamos nuevamente de la misma manera, obtenemos

$$z''_q(\varepsilon) = - \left\{ \frac{r}{F_q^2(\varepsilon)} + \frac{1-r}{[1-F_q(\varepsilon)]^2} \right\} [F'_q(\varepsilon)]^2 + \left\{ \frac{r}{F_q(\varepsilon)} - \frac{1-r}{1-F_q(\varepsilon)} \right\} F''_q(\varepsilon). \quad (5.25)$$

Como estamos trabajando dentro de la aproximación de punto silla de montar [26, 74, 102], podemos definir un parámetro ε_0 que maximiza $z_q(\varepsilon)$, y por lo tanto,

$$\varepsilon = \varepsilon_0 \implies z'_q(\varepsilon_0) = 0 \wedge F_q(\varepsilon_0) = r. \quad (5.26)$$

Luego,

$$\begin{aligned} z''_q(\varepsilon_0) &= - \left[\frac{1}{r} - \frac{1}{1-r} \right] [F'_q(\varepsilon_0)]^2 \\ &= - \frac{1}{r(1-r)} \frac{\alpha}{2\pi(1-\alpha)} [1 + \xi_0(\varepsilon_0)] \left[2 \left(\frac{q-2}{q-1} \right)^{-3} \right]. \end{aligned} \quad (5.27)$$

La distribución conjunta de disparos puede ser entonces calculada como

$$Q_q(r) \simeq \sqrt{\frac{1-\alpha}{\alpha}} \left\{ \exp_q \left[- \frac{(h - \sqrt{\alpha} F_q^{-1}(r))^2}{2(1-\alpha)} \right] \right\}^{-\frac{(q+1)}{2}} \exp \left[- \frac{(F_q^{-1}(r))^2}{2} \right], \quad (5.28)$$

donde

$$F_q^{-1}(r) = \frac{1}{\sqrt{\alpha}} \left\{ h - \sqrt{\frac{2(1-\alpha)}{q-1} \left[B_{inverse}^{-1} \left(r; \frac{1}{q-1}, \frac{1}{2} \right) - 1 \right]} \right\}. \quad (5.29)$$

Finalmente,

$$Q_q(r) \simeq \sqrt{\frac{1-\alpha}{\alpha}} \left[B_{inverse} \left(r; \frac{1}{q-1}, \frac{1}{2} \right) \right]^{-\frac{(q+1)}{2(q-1)}} \exp \left[- \frac{(F_q^{-1}(r))^2}{2} \right]. \quad (5.30)$$

Equivalentemente, la ecuación anterior puede ser expresada en términos de q -exponenciales de la forma

$$\begin{aligned} Q_q(r) &\simeq \sqrt{\frac{1-\alpha}{\alpha}} \left\{ \exp_q \left[B_{inverse}^{-1} \left(r; \frac{1}{q-1}, \frac{1}{2} \right) - 1 \right] \right\}^{-\frac{(q+1)}{2}} \\ &\times \exp \left[- \frac{1}{2\alpha} \left\{ h - \sqrt{\frac{2(1-\alpha)}{q-1} \left[B_{inverse}^{-1} \left(r; \frac{1}{q-1}, \frac{1}{2} \right) - 1 \right]} \right\}^2 \right]. \end{aligned} \quad (5.31)$$

En el límite $q \rightarrow 1$,

$$\begin{aligned} \lim_{q \rightarrow 1} Q_q(r) &= \sqrt{\frac{1-\alpha}{\alpha}} \lim_{q \rightarrow 1} \left\{ \exp_q \left[-\frac{(h - \sqrt{\alpha} F_q^{-1}(r))^2}{2(1-\alpha)} \right] \right\}^{-\frac{(q+1)}{2}} \\ &\quad \times \exp \left[-\frac{(F_q^{-1}(r))^2}{2} \right], \end{aligned} \quad (5.32)$$

considerando

$$\lim_{q \rightarrow 1} F_q^{-1}(r) = F^{-1}(r) \quad (5.33)$$

y

$$\lim_{q \rightarrow 1} \exp_q(x) = \exp(x) \quad (5.34)$$

obtenemos

$$\lim_{q \rightarrow 1} Q_q(r) = \sqrt{\frac{1-\alpha}{\alpha}} \exp \left[\frac{(h - \sqrt{\alpha} F^{-1}(r))^2}{2(1-\alpha)} - \frac{(F^{-1}(r))^2}{2} \right] = Q_1(r), \quad (5.35)$$

que resulta acorde con lo obtenido previamente por Amari y colaboradores [31].

Por ende, por medio de un modelo analítico podemos testear la presencia de correlaciones de alto orden midiendo la distribución de la actividad, y luego ajustando q . Cuando las correlaciones en la entrada común presináptica son solo de a pares, q tiende a 1 y la expresión se reduce a la ecuación (5.35), es decir la fórmula propuesta por Amari y colaboradores. Sin embargo, cuando las correlaciones en el sistema son fuertes, q se vuelve mayor a 1 para “sesgar” las probabilidades de que microestados particulares ocurran, tomando en cuenta entonces correlaciones de alto orden en la ecuación (5.31). Esto es, el parámetro q es entonces una manera de caracterizar el grado de correlación de la arquitectura de la red dentro de $Q_q(r)$, y en particular cuan fuertes son. Por lo tanto, al estimar la media h y la varianza α de una distribución y el grado de correlaciones de alto orden q podemos obtener la distribución de probabilidad conjunta de disparos $Q_q(r)$.

5.2.2. Estimaciones analíticas

En lo siguiente, proveemos estimaciones analíticas de cómo la distribución de probabilidad conjunta se comporta para diferentes conjuntos de los valores h , α y q . Consideramos tres distribuciones típicas como función de r : concentrada, ampliamente distribuida, y bimodal. Primero, la Figura 5.1 muestra una distribución concentrada. Esta concentración es causada por un pequeña cantidad de varianza $\alpha = 0,01$ y tal pico se corre a una mayor cantidad de neuronas sincronizadas a medida que el parámetro q aumenta. Segundo, la Figura 5.2 ilustra el caso de una distribución conjunta ampliamente distribuida considerando $\alpha = 0,2$. Se debe notar que cuando se incrementa el valor de q , la probabilidad de tener una mayor cantidad de espigas sincrónicas aumenta. A medida que q crece el pico de la distribución se corre hacia la derecha, lo que indica otra vez un número mayor de células sincronizadas. Finalmente, la Figura 5.3 muestra una distribución bimodal, al tomar $\alpha = 0,8$. En tal caso, las neuronas disparan simultáneamente en un momento y se inactivan en otros cuando el valor de q es cercano a 1. Al adicionar una mayor correlación a través de q , la proporción de células en silencio (o quiescentes) se vuelven activas llevando a una mayor cantidad de potenciales de acción en sincrónicos a distintos tiempos y a cualquier orden de correlación. Esto puede reflejar el impacto de la correlación en el ruido en el cerebro, y nos provee con una retroalimentación importante: correlaciones de alto orden en la entrada pequeñas, e incluso indetectables, pueden llegar a tener un impacto importante a nivel poblacional. Por ende, nuestra expresión analítica nos permite evaluar cómo las correlaciones de ordenes superiores a dos dependen de la estructura de la interacción en la señal de entrada a la población. Muestra además, que la distribución de las tasas de disparo se vuelve más sesgado a medida que q toma valores mayores, induciendo excursiones más grandes de espigas simultáneas.

5.2.3. Entropía relativa e información de Fisher

Una generalización de la divergencia de Kullback-Leibler (ver ecuación (2.12) en el capítulo 2), en el marco de la termodinámica no extensiva, es la entropía relativa de Tsallis o entropía q -relativa. Esta entropía relativa es una medida de la ineficiencia

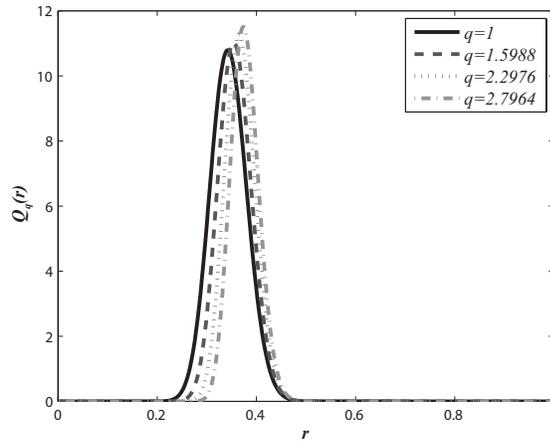


Figura 5.1: Distribución concentrada de la tasa de disparos $Q_q(r)$ para varios q , con $\alpha = 0,01$ y $h = 0,4$ fijos.

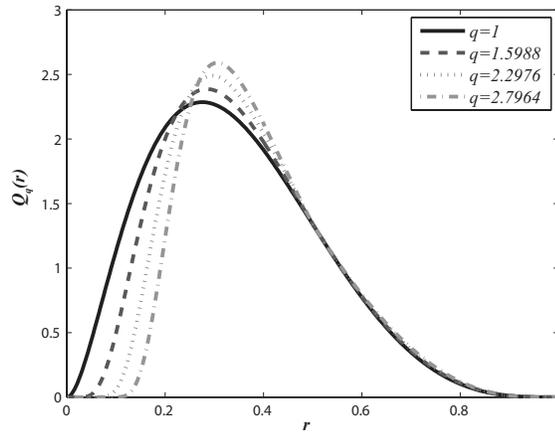


Figura 5.2: Distribución ampliamente distribuida de la tasa de disparos $Q_q(r)$ para varios q , con $\alpha = 0,2$ y $h = 0,4$ fijos.

en asumir que la distribución es P_0 cuando la verdadera distribución es P [103, 104]:

$$\begin{aligned}
 T_q[P \parallel P_0] &\equiv \int P(x) \frac{[P(x)/P_0(x)]^{1-q} - 1}{1-q} dx \\
 &= \frac{1}{q-1} \left[\int P(x) \left(\frac{P(x)}{P_0(x)} \right)^{q-1} dx - 1 \right] \\
 &= \frac{1}{q-1} \left[\int P(x)^q P_0(x)^{1-q} dx - 1 \right].
 \end{aligned} \tag{5.36}$$

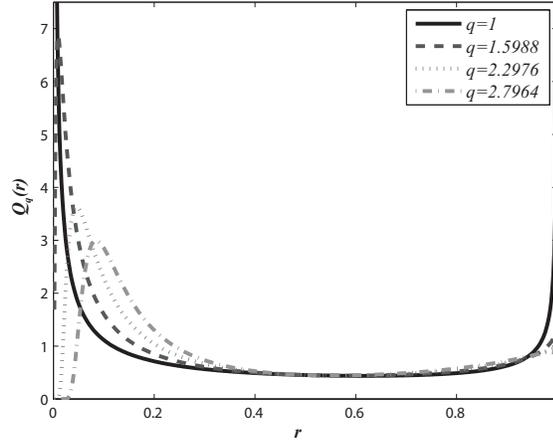


Figura 5.3: Distribución bimodal de la tasa de disparos $Q_q(r)$ para varios q , con $\alpha = 0,8$ y $h = 0,4$ fijos.

Calculamos la entropía relativa utilizando una distribución gaussiana como la función densidad de probabilidad de referencia. Por lo tanto, es positivo para el caso en que la distribución de probabilidad es descrita por una distribución de codificación óptima Q_q en vez de una distribución de referencia Q_1 . En el límite $q \rightarrow 1$, la entropía relativa de Tsallis se vuelve la divergencia de Kullback-Leibler divergence. Ambas entropías relativas son no simétricas, *i.e.*, $T_q[Q_q \parallel Q_1] \neq T_q[Q_1 \parallel Q_q]$. Esta medida nos permite comparar cuantitativamente la distribución de actividad extendida Q_q con la de referencia (Q_1).

Otra medida que también es particularmente interesante en este contexto es testear la aplicabilidad de la ecuación (5.7), dado que fue utilizada en la referencia [105] para obtener una prueba dentro en el contexto de la teoría de la información del teorema del límite central, es la información de Fisher. Esta cantidad [106, 107] constituye una medida del contenido gradiente de una distribución $P(x)$:

$$I(P) = \int \frac{|\vec{\nabla} P(x)|^2}{P(x)} dr, \quad (5.37)$$

y por lo tanto es suficientemente sensible incluso a pequeñas perturbaciones locales. La información de Fisher puede ser interpretada como el alcance de la habilidad de estimar un parámetro, como la cantidad de información que puede ser extraída de

un conjunto de medidas, y además como una medida del estado de desorden de un sistema o fenómeno [107, 108]. Su principal propiedad resulta la llamada “cota de Cramer-Rao”, que expresa una cota inferior para la varianza de un estimador no sesgado. El operador gradiente afecta significativamente la contribución de variaciones $-p$ locales al valor de la información de Fisher, por lo que se denomina “local”. La sensibilidad local es útil en escenarios cuyas descripciones necesitan apelar a un noción de “orden” [109–111]. La información de Fisher también resulta práctica para detectar cambios dinámicos en la función densidad de probabilidad. Es decir, una función de distribución de probabilidades más angosta tiende a tener una mayor información de Fisher que una distribución más ampliamente distribuida.

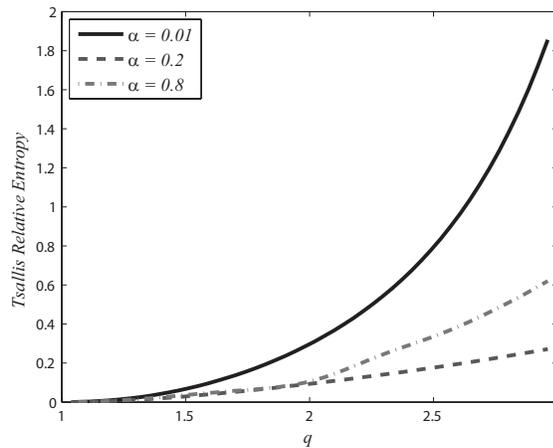


Figura 5.4: Entropía relativa de Tsallis (‘Tsallis Relative Entropy’) del modelo gaussiano $Q_1(r)$ y del modelo extendido $Q_q(r)$ en función de q , for various α and $h = 0,4$ fijo.

La Figura 5.4 muestra la entropía relativa de Tsallis como función del parámetro q . La función crece a medida de que el grado de correlación aumenta, cuantificando la ineficiencia de asumir incorrectamente que la distribución es aquella correspondiente a $Q_1(r)$ cuando la distribución verdadera es $Q_q(r)$. Esto significa que el total de la información en una población no puede ser computado sin tener conocimiento de la estructura correlacional. Si nuestro sistema se encuentra en un estado muy ordenado y por ende está representado por una densidad de probabilidad muy angosta, obtenemos una medida de la información de Fisher máxima. Por otro lado, cuando

el sistema bajo estudio se encuentra en un estado muy desordenado, uno obtiene una distribución de probabilidad chata y la información de Fisher es cercana a cero [112]. En lo siguiente, tomamos ventaja de nuestra solución analítica propuesta para la distribución conjunta de probabilidades al considerar correlaciones en los disparos neuronales de todos los órdenes para investigar cómo la información se comporta a medida que la correlación en el ruido aumenta.

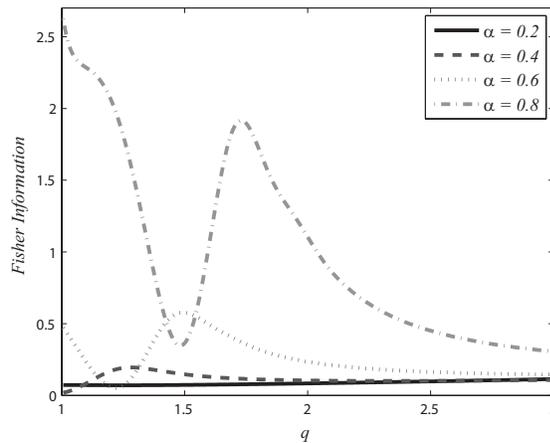


Figura 5.5: Información de Fisher (‘Fisher Information’) de la distribución $Q_q(r)$ versus q .

La Figura 5.5 corresponde a la gráfica de la información de Fisher en función de q , teniendo en cuenta el resultado del apéndice A. Mostramos que es posible cuantificar la cantidad óptima del parámetro q que maximiza la información. El desempeño en la codificación puede ser mejorado mediante la modificación de la salida mediante las correlaciones de alto orden en la entrada y sus no-linealidades. Además, la información satura a medida que el nivel de correlaciones en el ruido aumenta. Esto concuerda con la intuición general necesitada para entender cómo las correlaciones puedan afectar la información en un código poblacional: para correlaciones positivas, la información satura a medida que el número de neuronas correlacionadas aumenta y estas tienen un gran efecto a nivel poblacional [58]. Sin embargo, para valores muy pequeños de α la información no llega a mostrar saturación.

Este modelo además permite relacionar analíticamente los parámetros que caracterizan la estadística de la distribución de entrada y la distribución de la salida.

En el capítulo siguiente se desarrollará dicha metodología, para definir un modelo generador de disparos y de esta manera vincular analíticamente la estructura de la red y la correlación observada en la actividad de la población.

Capítulo 6

Generación de patrones de actividad poblacionales con entradas correlacionadas

En este capítulo, caracterizamos estadísticamente la actividad de disparos de una población obtenida de registros simultáneos de neuronas a lo largo de todas las capas de una microcolumna cortical simulada. Ponemos un énfasis en cómo realizar estimaciones cuando consideramos correlaciones de alto orden en la conectividad de la entrada a dicho conjunto de células, proveyendo un marco teórico matemático apropiado para cuantificar correlaciones de alto orden en la salida de espigas de la población neuronal. Proveemos además, en este contexto, un modelo generador de patrones de disparo con una estadística dada. Es decir, imponiendo momentos fijos para la actividad de la población, podemos calcular los momentos necesarios del proceso latente.

Aprovechamos el hecho de que las distribuciones q -gaussianas, en la región de colas pesadas, son equivalentes a la distribución $-t$ de Student con un mapeo directo entre el parámetro de deformación q de la distribución q -gaussiana y el grado de libertad de la distribución $-t$ de Student. Esto nos permite escribir la distribución conjunta de probabilidades de los disparos de la población de neuronas en términos de la función de distribución acumulativa de la distribución $-t$ de Student tomando en cuenta las correlaciones más allá de aquellas de segundo orden en sus entradas y,

más importantemente, proveer una estimación robusta de la estadística de la salida en términos de la distribución q -gaussiana bivariada. Este esquema resalta la importancia tanto de las entradas comunes no gaussianas como de las correlaciones de alto orden en la estadística de la salida, al describir la estructura de las redes corticales.

Presentamos una comparación cuidadosa del presente método con el modelo de Gaussiana Dicotomizada (DG) [32]. Esto resulta importante para investigar cómo la cuantificación de las correlaciones de alto orden en la entrada de una población neuronal del modelo q -DG difiere de aquella cuya entradas no poseen interacciones de orden mayor a dos. Caracterizar el grado de correlaciones de alto orden a través del parámetro q puede ayudarnos a ganar un mayor entendimiento de cómo la información es procesada en redes corticales y sobre la dinámica no lineal de disparo dentro del grupo neuronal.

6.1. El modelo generador de patrones de actividad poblacionales

Una población de N neuronas, si asumimos que los trenes de espigas son discretizados en ventanas temporales suficientemente pequeñas tal que cada ventana temporal contiene a lo sumo un disparo, puede ser representado como una colección de N secuencias binarias. Esto es, cada neurona puede generar o no un potencial de acción a un dado tiempo y por lo tanto la población puede estar en cualquiera de los 2^N estados diferentes. Estimar la distribución conjunta completa sobre estos 2^N patrones se vuelve rápidamente inviable al incrementar el número de neuronas N , a razón de la explosión combinatoria. Sin embargo, frecuentemente es posible medir los primeros y segundos momentos de dicha distribución. Estos son la tasa de disparos de las células individuales y las correlaciones entre pares de las mismas. El modelo de Gaussiana Dicotomizada permite generar trenes de espigas de una población de N neuronas donde estos momentos fueron especificados. A pesar de esto, la distribución de la suma de las entradas a esa población se asume gaussiana y, por ende, no puede incluir posibles correlaciones de ordenes más allá al segundo [26, 32, 33]. Entonces, sería conveniente poder relajar los vínculos no gaussianos en la distribución de entradas comunes para de esta manera llegar a tener en cuenta el impacto

de las correlaciones de alto orden.

Modelamos entonces una variable aleatoria multivariada, N -dimensional $X \in \{0, 1\}^N$ con media $\boldsymbol{\mu}$ y matriz de covarianza $\boldsymbol{\Sigma}$. Una muestra de la distribución q -DG es obtenida luego de primero tomar una muestra de una variable aleatoria N -dimensional q -gaussiana \mathbf{U} y luego aplicarle un umbral dando 0 o 1. Esto es,

$$X_i = 1 \text{ si y sólo si } U_i > 0 \text{ donde } U \sim \mathcal{G}_q(\mathbf{h}, \mathbf{A}), \quad (6.1)$$

o cero en el caso contrario.

Dicha operación con umbral cambiará los momentos de la salida, luego X no tendrá, en general, la misma media y covarianza que \mathbf{U} . Los momentos de la salida están dados por las medias $\mu_i = \langle X_i \rangle$ y covarianzas de pares $\Sigma_{ij} = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle$. Sin embargo, el efecto de truncado puede ser calculado [113], como lo es también posible en el caso del modelo de Gaussiana Dicotomizada, y corregido. Muestras de X pueden ser construídas mediante la generación de muestras de \mathbf{U} (con medias \mathbf{h} y matriz de covarianza \mathbf{A}), y binarizadas en 0 y 1 mediante un umbral, i. e. definiendo $X_i = 1$ si y sólo si $U_i > 0$.

Es posible elegir la media \mathbf{h} y la matriz de covarianza \mathbf{A} de \mathbf{U} tal que después del truncado, \mathbf{X} posea los momentos deseados $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$. Para una implementación algorítmica más sencilla, explotamos la equivalencia, cuando las condiciones son cumplidas, entre las distribuciones q -gaussianas y las distribuciones $-t$ de Student. Asumiendo (sin perder la generalidad) varianzas unitarias para \mathbf{U} , las probabilidades de disparo medias $\boldsymbol{\mu}$ y covarianza entre pares de neuronas $\boldsymbol{\Sigma}$ de \mathbf{X} están dados por

$$\mu_i = \Phi_q(h_i), \quad (6.2)$$

$$\Sigma_{ii} = \Phi_{q,2}(h_i, -h_i, 0), \quad (6.3)$$

$$\Sigma_{ij} = \Phi_{q,2}(h_i, h_j, A_{ij}) - \Phi_{q,2}(h_i, h_i, 0). \quad (6.4)$$

$\Phi_q(\cdot)$ es la función de distribución acumulativa de una variable q -gaussiana univariada con media 0 y varianza unitaria (de aquí en adelante referida como la distribución q -Normal), y $\Phi_{q,2}(\cdot, \cdot; \rho)$ la función de distribución acumulativa de una distribución q -gaussiana bivariada con varianzas unitarias y coeficiente de correlación ρ (q -normal bivariada). Ambas distribuciones se encuentran caracterizadas además por el paráme-

tro de deformación q . En el apéndice B mostramos como generar eficientemente distribuciones q -gaussianas bivariadas.

Para que sea posible definir distribuciones q -normales (es decir, que la varianza de una distribución q -gausiana sea unitaria) el parámetro de deformación q debe ser restringido al intervalo $[1, 5/3)$. Para estos valores, las distribuciones anteriores pueden ser directamente mapeadas a t -distribuciones de Student escaleadas \mathcal{T}_ν . Para que la varianza sea 1, $\beta = \frac{1}{5-3q}$. Bajo esta condición, el grado de libertad seguiría siendo $\nu = \frac{3-q}{q-1}$ y el parámetro de escaleo es $\zeta^2 = \frac{5-3q}{3-q}$, tal que

$$\begin{aligned} \mathcal{N}_q(x) &= \mathcal{G}_q\left(x; \frac{1}{5-3q}\right) = \frac{1}{\zeta} \mathcal{T}_\nu\left(\frac{x}{\zeta}\right) \\ &\equiv \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\zeta^2\nu}\Gamma\left(\frac{\nu}{2}\right)} \left[1 - \frac{1}{\nu} \left(\frac{x}{\zeta}\right)^2\right]^{-\frac{\nu+1}{2}}. \end{aligned} \quad (6.5)$$

6.2. Relacionando correlaciones de alto orden en la entrada y la salida

Desarrollaremos en lo siguiente un análisis de la relación entre los parámetros que caracterizan la estadística de las señales de entrada a las neuronas y sus respectivas actividades. Las ecuaciones (6.2 - 6.4) proveen una descripción matemática de las mismas para el caso considerado de neuronas binarias con umbral de disparo. Los valores medios h_i pueden ser calculados invirtiendo la ecuación (6.2). Esto es podemos computar estas cantidades utilizando

$$h_i = \Phi_q^{-1}(\mu_i), \quad (6.6)$$

dado que puede ser demostrado que $\Phi_q(\cdot)$ es invertible en el rango considerado.

Determinar Σ_{ij} puede ser logrado encontrando los valores apropiados tal que

$$\Sigma_{ij} - [\Phi_{q,2}(h_i, h_j, A_{ij}) - \Phi_{q,2}(h_i, h_j, 0)] = 0. \quad (6.7)$$

A pesar de esta aparente simplicidad, un análisis cuidadoso del rango de las matrices de covarianza posibles es necesario. Al lidiar con variables aleatorias simuladas

con matrices de covarianza que no son directamente estimadas de datos experimentales sino que son construídas mediante otras consideraciones, no cualquier matriz simétrica definida positiva puede ser usada como la matriz de covarianza de una distribución binaria multivariada.

En comparación a la distribución gaussiana, la función de la distribución acumulativa q -gaussiana bivariada independiente no es el producto de las funciones de distribución acumulativas individuales. Por lo tanto, las posibles covarianzas no están restringidas solamente por los valores de las medias de las salidas de las neuronas sino también por el parametro de deformación de la distribución de la entrada. Por ejemplo, para dos variables aleatorias binarias X_1 y X_2 con medias μ_1 y μ_2 , la covarianza está restringida por los valores [32]

$$\max \{-\mu_{1,2}, \mu_1 + \mu_2 - (1 + \mu_{1,2})\} \leq \text{Cov}(X_1, X_2) \leq \min \{\mu_1 - \mu_{1,2}, \mu_2 - \mu_{1,2}\}. \quad (6.8)$$

donde $\mu_{1,2} = \Phi_q([\Phi_q^{-1}(\mu_1), \Phi_q^{-1}(\mu_2)], 0)$. Esto provee un conjunto diferente de vínculos para la covarianza. Para demostrar que la ecuación (6.4) posee una solución para cualquier μ_i, μ_j y $\Sigma_{ij} = \text{Cov}(X_i, X_j)$ que satisfaga las condiciones establecidas anteriormente, es necesario mostrar que podemos encontrar un α tal que

$$\Phi_{q,2}(h_i, h_j, \alpha) = \text{Cov}_{ij} + \Phi_{q,2}(h_i, h_i, 0). \quad (6.9)$$

Dado que $\Phi_{q,2}$ es continuo en α , y que satisface que

$$\Phi_{q,2}(h_i, h_j, -1) \leq \text{Cov}_{ij} + \Phi_{q,2}(h_i, h_i, 0) \leq \Phi_{q,2}(h_i, h_j, 1) \quad (6.10)$$

hemos establecido correctamente los límites de la función de distribución acumulativa de la distribución q -gaussiana bivariada.

6.3. Modelo de población homogénea

Proveemos una derivación alternativa a la desarrollada en el capítulo 5 para la distribución asintótica de conteo de espigas de una población homogénea grande (debido a la tratabilidad analítica) de neuronas para el modelo q -DG en términos de

la distribución $-t$ de Student. Esto provee una implementación más sencilla de los algoritmos para cuantificar las correlaciones de alto orden.

Consideramos un grupo de N neuronas donde cada unidad posee un potencial de membrana u_i sujeto a la distribución conjunta q -Normal. Dada la variable aleatoria q -gaussiana N -dimensional U , donde $A = \mathbb{I}(1 - \alpha) + \alpha \mathbb{1}_N \mathbb{1}_N^T$, entonces

$$u_i = h + \sqrt{1 - \alpha}v_i + \sqrt{\alpha}\varepsilon. \quad (6.11)$$

Las variables v_i son variables aleatorias independientes sujetas a N_q y ε corresponde a ruido gaussiano. Luego, las variables u_i son independientes para un dado ε .

En el marco del modelo q -DG, los patrones binarios X son generados por $X = 1$. La estadística de la entrada es elegida tal que la salida X tiene media μ y covarianza Σ . Dado que nos concentramos en una población de neuronas idénticas, $\mu_i = \mu$ y $\Sigma_{ij} = \sigma$, y tomamos el coeficiente de correlación de segundo orden como

$$\rho = \frac{\sigma}{\mu(1 - \mu)}. \quad (6.12)$$

Por simetría, todos los patrones de actividad con el mismo número de espigas serán equiprobables. Como vimos en el capítulo 5, tal población se encuentra caracterizada completamente por el número de neuronas que se activan simultáneamente k . En este sentido, podemos definir en una población de N neuronas la tasa de disparo poblacional o la proporción de neuronas disparando como $r = k/N$, y entonces $k = N \cdot r$. Por lo tanto, la probabilidad de que k neuronas generen un potencial de acción en un momento dado es

$$P(k) = Prob\{x_1 = x_2 = \dots = x_k = 1, x_{k+1} = \dots = x_N = 0\}, \quad (6.13)$$

donde el disparo de cada neurona x_i se encuentra restringido por su potencial de membrana como previamente explicado. En la ecuación (6.13) presentamos una forma compacta de la probabilidad, no obstante todas las combinaciones posibles son tomadas en cuenta para computar la probabilidad como en [74, 75]. La probabilidad de tener un cierto patrón con k espigas está dada por el valor esperado tomado con

respecto a la variable aleatoria ε , y $P\{\cdot|\varepsilon\}$ es la probabilidad condicional para ε . Esto nos permite calcular la probabilidad de que la tasa de disparos poblacional sea $r = k \cdot N$, separando la contribución de las neuronas activas $P\{u > 0|\varepsilon\}^k$ de aquellas en silencio $P\{u \leq 0|\varepsilon\}^{N-k}$, como

$$\begin{aligned} P(k) &= E_\varepsilon [P\{u_1; \dots; u_k > 0, u_{k+1}; \dots; u_N \leq 0|\varepsilon\}] \\ &\equiv E_\varepsilon \left[\binom{N}{k} (P(u > 0|\varepsilon))^k (P(u \leq 0|\varepsilon))^{N-k} \right]. \end{aligned} \quad (6.14)$$

Entonces, para cada neurona,

$$\begin{aligned} P(u_i > 0) &= P\left(v_i < \frac{\sqrt{\alpha}\varepsilon + h}{\sqrt{1-\alpha}}\right) \\ &= \Phi_q\left(\frac{\sqrt{\alpha}\varepsilon + h}{\sqrt{1-\alpha}}\right) \\ &\equiv \frac{1}{\zeta} T_\nu\left(\frac{\sqrt{\alpha}\varepsilon + h}{\zeta\sqrt{1-\alpha}}\right) = F_q(\varepsilon). \end{aligned} \quad (6.15)$$

En esta ecuación, T_ν es la función de distribución acumulativa de una distribución $-t$ de Student con grado de libertad ν y parámetro de escala ζ . La distribución asintótica de la tasa de disparo para la población homogénea se calcula utilizando el mismo método de la aproximación silla de montar [102]. Considerando la ecuación (6.15), podemos calcular fácilmente cantidades anteriores en términos de la distribución $-t$ de Student dado que su derivada es comúnmente conocida. Por lo tanto, la densidad de probabilidad para una población homogénea generada con el modelo q -DG es

$$f_\nu(r) = \zeta \sqrt{\frac{1-\alpha}{\alpha}} \left[1 + \frac{(\mathbb{T}_\nu^{-1}(\zeta r))^2}{\nu} \right]^{\frac{\nu+1}{2}} \exp \left[-\frac{\zeta}{2} \frac{1-\alpha}{\alpha} \left(\mathbb{T}_\nu^{-1}(\zeta r) - \frac{h}{\zeta\sqrt{1-\alpha}} \right)^2 \right], \quad (6.16)$$

o, equivalentemente en términos del parámetro de deformación q ,

$$f_q(r) = \sqrt{\frac{(1-\alpha)(5-3q)}{\alpha(3-q)}} \exp_q^{-1} \left[-\left(\frac{q-1}{3-q}\right) \left(T_{\left(\frac{3-q}{q-1}\right)}^{-1} \left(\sqrt{\frac{5-3q}{3-q}} r \right) \right)^2 \right] \\ \times \exp \left[-\frac{\sqrt{5-3q}(1-\alpha)}{2\alpha(3-q)} \left(T_{\left(\frac{3-q}{q-1}\right)}^{-1} \left(\sqrt{\frac{5-3q}{3-q}} r \right) - \frac{h\sqrt{3-q}}{\sqrt{(5-3q)(1-\alpha)}} \right)^2 \right]. \quad (6.17)$$

6.3.1. Resultados numéricos

Estudiamos el régimen asintótico para investigar el efecto de las correlaciones de alto orden en las entradas comunes y la actividad de salida de una población de neuronas simples. Las Figuras 6.1 y 6.2 (a)-(d) muestran las gráficas rasterizadas de la población generada considerando $q = 1$ (modelo DG) y en el caso del modelo q -DG con parámetro de deformación $q = 1,35$, respectivamente, para valores de coeficiente de correlación $\rho = 0$, $\rho = 0,05$, $\rho = 0,1$ y $\rho = 0,2$. Consideramos que poseemos correlaciones más allá de entre pares en la entrada cuando los valores de q son mayores a uno. Esto produce mayores cantidades de sincronización entre espigas y el efecto se vuelve más pronunciado a medida que ρ aumenta. Esto es, tener $q = 1,35$ induce un número más considerable de disparos simultáneos, lo que puede ser observado por simple comparación de los patrones de actividad en las Figuras 6.1 y 6.2.

Las últimas proveen una primera evaluación de cómo la estructura de las correlaciones en la entrada pueden afectar la actividad de salida. Sin embargo, proveer un análisis más preciso es necesario para cuantificar el efecto de las correlaciones de alto orden. La Figura 6.3 (a)-(b) muestra las distribuciones conjuntas de disparo (dadas por la ecuación (6.17)) cuando consideramos una distribución gaussiana ($q = 1$) y una distribución no gaussiana del potencial de membrana con $q = 1,35$, respectivamente, tomando además los valores $\rho = 0$, $\rho = 0,05$, $\rho = 0,1$ y $\rho = 0,2$. Como esperado, un mayor parámetro de deformación promueve mayores excursiones de espigas sincronizadas [74, 75].

Segundo, analizamos precisamente cómo las correlaciones de alto orden afectan las

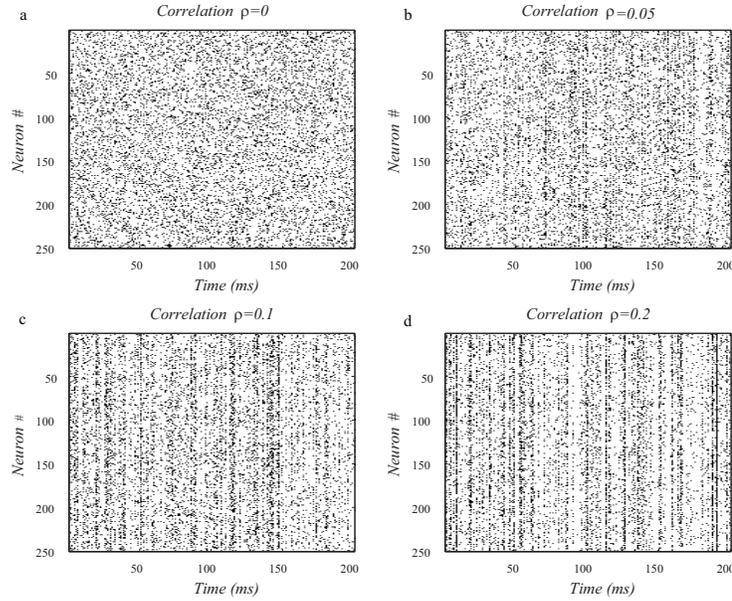


Figura 6.1: Gráficas rasterizadas de patrones de disparo multineuronales dentro del formalismo DG ($q = 1$) y considerando distintos coeficientes de correlación de la salida ρ . (a) $\rho = 0$, (b) $\rho = 0,05$, (c) $\rho = 0,1$ y (d) $\rho = 0,2$.

correlaciones en la salida. Más específicamente, cuantificamos cómo las correlaciones entre los potenciales de acción son afectados por las correlaciones de alto orden en sus entradas, y si es posible que produzcan desviaciones del modelado de segundo orden puro (*i.e.* el modelo DG). Para realizar esto, proveemos un análisis detallado de cómo las correlaciones en la actividad de salida se desvían del modelo DG a medida que el parámetro de deformación q se incrementa al considerar distintos valores para la media y la varianza de las entradas neuronales. La ecuación (6.4) explicita la relación característica que existe entre las correlaciones entre pares en la entrada y las correlaciones de segundo orden en la salida de neuronas binarias que poseen un umbral. Dado el mismo grado de correlación en la entrada α , la correlación en la salida ρ crecerá monótonamente con sus tasas de disparo [78]. La Figura 6.4 (a), (b) y (c) muestra la dependencia de ρ con respecto a α , al tomar en cuenta valores diferentes de la tasas de disparo poblacional $\mu = 0,1$, $\mu = 0,25$ y $\mu = 0,5$, respectivamente, para $q = 1$, $q = 1,1$, $q = 1,3$ y $q = 1,3$. Al incrementar la cantidad de correlaciones de alto orden en la conectividad de la entrada, se producen números ligeramente

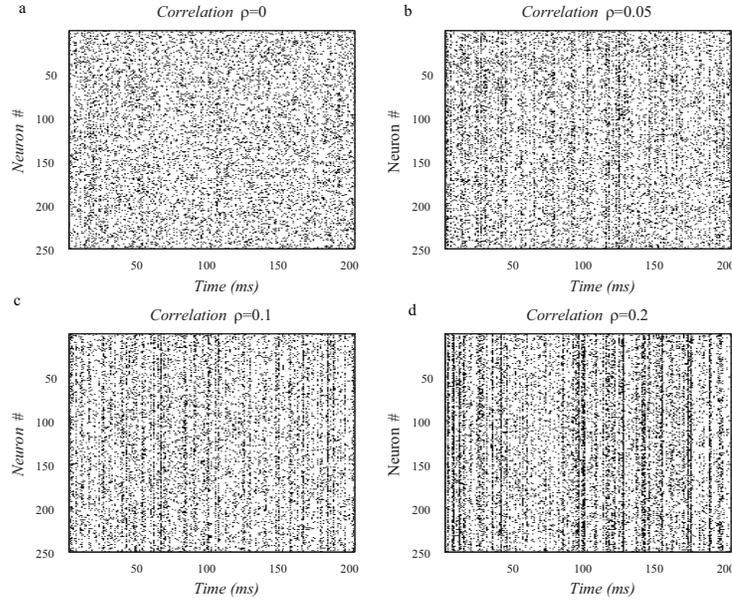


Figura 6.2: Ídem a la Figura 6.1 pero considerando el formalismo q -DG con parámetro de deformación $q = 1,35$. Los coeficientes de correlación de la salida se toman como (a) $\rho = 0$, (b) $\rho = 0,05$, (c) $\rho = 0,1$ y (d) $\rho = 0,2$.

menores de espigas simultáneas en los patrones de actividad. No obstante, valores mayores en la tasa de disparo aumentan la sincronización.

Con el objetivo de entender mejor cómo las neuronas corticales transforman la correlación entre sus entradas de corrientes sinápticas en correlación entre sus secuencias de espigas de sus salidas, necesitamos además investigar el coeficiente de correlación ρ como función de la tasa de disparo μ . La ecuación (6.4) también provee un medio para obtener esta relación cuando la varianza en la entrada se mantiene constante. La Figura 6.5 (a)-(c) ilustra cómo la correlación en la salida ρ varía como función de la tasa de disparo μ , para distintos valores de la correlación entre pares de la entrada $\alpha = 0,1$, $\alpha = 0,2$ y $\alpha = 0,3$, para parámetros de deformación $q = 1$, $q = 1,1$, $q = 1,2$ y $q = 1,3$. Como puede ser observado en la figura, las correlaciones de alto orden en las entradas producen un menor valor de sincronía al considerar tasas de disparo pequeñas ($\mu < 0,15$). Cuando $\mu = 0,15$ la correlación es igual a la obtenida por el modelo DG. En contrastem cuando la tasa de disparo toma valores mayores ($\mu > 0,15$), esto lleva a mayores excursiones de espigas simultáneas a medida

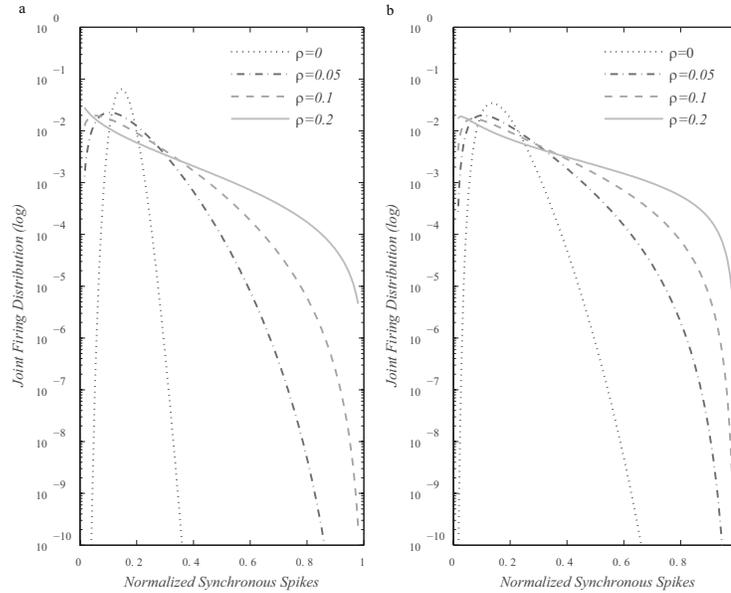


Figura 6.3: Distribución conjunta de disparos (‘Joint Firing Distribution’) como función del número de espigas normalizado (‘Normalized Synchronous Spikes’), en escala semi-logarítmica en el eje Y. (a) Considerando $q = 1$ y $\mu = 0,15$ para $\rho = 0$, $\rho = 0,05$, $\rho = 0,1$ y $\rho = 0,2$. (b) Considerando $q = 1,35$ y $\mu = 0,15$ para $\rho = 0$, $\rho = 0,05$, $\rho = 0,1$ y $\rho = 0,2$.

que aumenta q . La relación entre la correlación de segundo orden de la actividad y la tasa de disparo es sensible a las heterogeneidades en la entrada y, incluso cuando las correlaciones de alto orden en la entrada se mantienen fijas, la correlación aumenta con la tasa de disparo. Sin embargo, muestra un comportamiento complejo a medida que el orden de la correlación en la entrada se incrementa. Es decir, las correlaciones de alto orden inducen mayor silencio simultáneo [114] cuando $\mu < 0,15$, mientras que para mayores valores de μ el grado de sincronía crece no linealmente.

6.3.2. Cálculo de la entropía y la divergencia de Kullback-Leibler del modelo

La entropía, en la teoría de la información [6], es una cantidad que mide cuán interesante resulta un conjunto de respuestas. Dada la distribución de respuestas neuronales, si cada una de las mismas es idéntica o si solamente aparecen unas pocas

respuestas diferentes el conjunto de datos es relativamente poco interesante. Computar la entropía entonces otorga un sentido de la variabilidad en la distribución de tasas de disparo. En general, la entropía de una variable aleatoria X siguiendo una función densidad de probabilidad $P(x)$ está definida por la ecuación (2.6). Dado que estamos interesados en cuantificar diferencias de entropía, definimos una entropía diferencial normalizada ΔS_q en términos del parámetro de deformación q de la forma

$$\Delta S_q(X) = \frac{S_q(X) - S_1(X)}{S_1(X)}. \quad (6.18)$$

Es la diferencia entre el valor de la entropía de la distribución para un dado q y el valor de la entropía para el caso $q = 1$, normalizado al segundo.

Utilizando esta medida, determinamos el efecto de las correlaciones de alto orden en la entrada común en la entropía de la población. La Figura 6.6 muestra que el grado de caos resulta fuertemente dependiente de las correlaciones de orden más allá del segundo. Esto es, la Figura 6.6 ilustra que el incremento perceptual de entropía crece con el parámetro q , mediante la ecuación (6.18).

Un método disponible para testear la presencia de correlaciones de alto orden es midiendo la distribución de la actividad poblacional en registros de multi-unidades y ajustando el parámetro de deformación q de tal distribución. Luego, podemos computar la divergencia de Kullback-Leibler como distancia entre esta y la función no correlacionada estándar. Al igual que con el cálculo de la entropía, consideramos una versión normalizada de la divergencia de Kullback-Leibler. La definimos como la diferencia entre la divergencia de Kullback-Leibler de la distribución f_q para una dada correlación entre pares y el caso independiente de segundo orden para un dado valor de q y la divergencia de Kullback-Leibler entre una distribución f_1 de la correlación entre pares dada y el caso independiente para $q = 1$. f_q corresponde a la función definida en 6.17. Esto es,

$$\Delta_q = D[f_q(\rho) \parallel f_q(0)] - D[f_1(\rho) \parallel f_1(0)]. \quad (6.19)$$

La Figura 6.7 muestra cuán estadísticamente diferentes las distribuciones de disparo conjuntas son al considerar correlaciones de alto orden, según la ecuación (6.19). La divergencia de Kullback-Leibler relativa aumenta a medida que el grado de corre-

lación se vuelve más alto, cuantificando la ineficiencia de asumir incorrectamente que la distribución es aquella que corresponde a $f_1(\rho)$ cuando la distribución verdadera es $f_q(\rho)$. Entonces, la cantidad de información en la salida poblacional no puede ser computada sin conocer la estructura correlacional de sus entradas, cuantificada por el parámetro de deformación q .

6.3.3. Estudio de la causalidad del modelo

En lo siguiente, consideramos además la medida de complejidad estadística MPR [115] de una distribución P , dado que es capaz de cuantificar detalles críticos respecto a los procesos dinámicos que subyacen a la distribución. Los extremos opuestos de orden perfecto (por ejemplo, una secuencia periódica) y la aleatoriedad máxima (un lanzamiento de moneda justo) pueden ser descriptos muy fácilmente porque no poseen estructura alguna. La complejidad estadística en ambos casos sería nula. A una dada distancia de estos extremos, existen un amplio rango de estructuras ordinales posibles. La complejidad estadística, entonces, nos permite cuantificar esta variedad de comportamiento [116]. Basándose en la noción avanzada por López-Ruiz y colaboradores [117], esta cantidad se define a través del producto

$$C_{JS}(P) = \mathcal{Q}_J[P \parallel P_e] \cdot H_{norm}(P). \quad (6.20)$$

\mathcal{Q}_J corresponde al “desequilibrio”, definido en términos de la divergencia de Jensen-Shannon (ver capítulo 2). Esto es,

$$\mathcal{Q}_J[P \parallel P_e] = Q_0 J[P \parallel P_e] \quad (6.21)$$

con

$$J[P \parallel P_e] = H[(P + P_e)/2] - H(P)/2 - H(P_e)/2. \quad (6.22)$$

P_e es una distribución uniforme, es decir $P_e = p_j = 1/N, \forall j = 1, \dots, N$, y se considera la distribución de referencia para esta medida, con N el número posible de estados del sistema en estudio. Q_0 es una constante de normalización ($0 \leq \mathcal{Q}_J \leq 1$), igual a la inversa del valor máximo posible de $J[P \parallel P_e]$. Este valor se obtiene cuando una de las componentes de P , por ejemplo p_m , es igual a 1 y el resto de los p_j son iguales

a cero. $H_{norm}(P)$ es la entropía de Shannon normalizada como

$$H_{norm}(P) = H(P)/H_{max}, \quad (6.23)$$

con $H_{max} = H(P_e) = \ln N$, ($0 \leq \mathcal{H}_S \leq 1$). Para un dado valor de H_{norm} , el rango posible de valores de C_{JS} varía entre un mínimo C_{min} y un máximo C_{max} , restringiendo los valores alcanzables en un plano de complejidad-entropía dado [118].

En esta sección, mostraremos resultados preliminares utilizando esta medida para la distribución dada por el modelo q -DG. Además, con el fin de completar la explicación, en el apéndice C se muestran estos cuantificadores para el modelo de neurona simple de Izhikevich (introducidos en el capítulo 3) utilizando la metodología de Bandt-Pompe [119].

En el contexto del modelo q -DG, P_e correspondería al caso en el que cada patrón de disparo es equiprobable. Esto significa que la actividad de la población de neuronas no resulta informativa en absoluto con respecto a la entrada que reciben. Es decir, que la complejidad de la distribución generada por el modelo q -DG indicaría la capacidad de codificación del conjunto de células. Por ejemplo, en la Figura 6.8 se puede observar un máximo en la complejidad para el valor $q = 1,187$, habiendo utilizado los parámetros $\mu = 0,08$ y $\rho = 0,7$. Eso significa que ese valor provee la mejor “estructura”, proveyendo la mayor distancia al caso de eventos equiprobables. Además, puede corroborarse, estudiando el plano $C \times H$, es decir complejidad MPR versus entropía, para estos mismos valores como ilustrado en la Figura 6.9.

Por otro lado, podemos estudiar cómo se relaciona la información de Fisher respecto a la complejidad. Al graficar el plano $F \times C$ para estos mismos valores, se puede ver que la complejidad encuentra también un valor óptimo respecto a la información de Fisher.

Por lo tanto, comparando las Figuras 6.9 y 6.10 se puede concluir que el sistema posee un estado en donde la capacidad de transmisión de información resulta óptima.

Sin embargo, las propiedades de la población modelada pueden variar con respecto al tamaño de la misma. En el siguiente capítulo, estudiaremos estos efectos de tamaño finito, discutiendo además la relevancia funcional de estas observaciones.

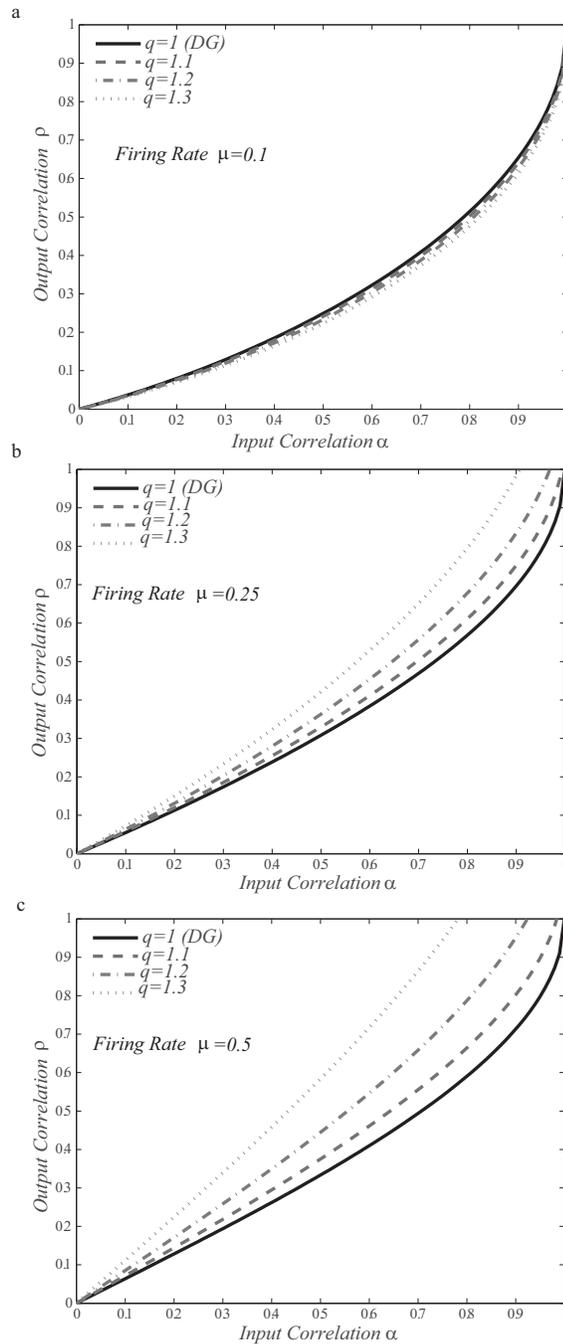


Figura 6.4: Correlación en la salida ('Output Correlation') ρ versus correlación en la entrada ('Input Correlation') α , tomando $q = 1$, $q = 1,1$, $q = 1,2$ y $q = 1,3$, con la tasa de disparos μ fija. (a) Considerando una tasa de disparo $\mu = 0,1$. (b) Tasa de disparo $\mu = 0,25$. (c) Tasa de disparo $\mu = 0,5$.

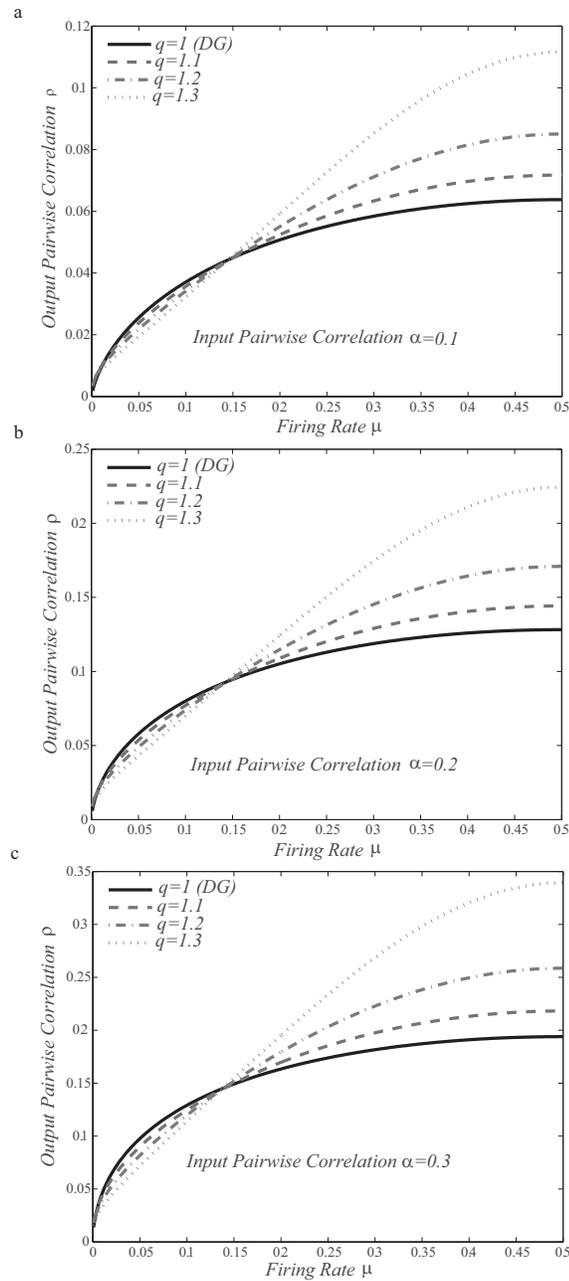


Figura 6.5: Correlación en la salida (‘Output Pairwise Correlation’) ρ como función de la tasa de disparo (‘Firing Rate’) μ , tomando los valores $q = 1$, $q = 1,1$, $q = 1,2$ y $q = 1,3$. (a) Considerando una correlación de entrada fija $\alpha = 0,1$. (b) Correlación en la entrada $\alpha = 0,2$. (c) Correlación en la entrada $\alpha = 0,3$.

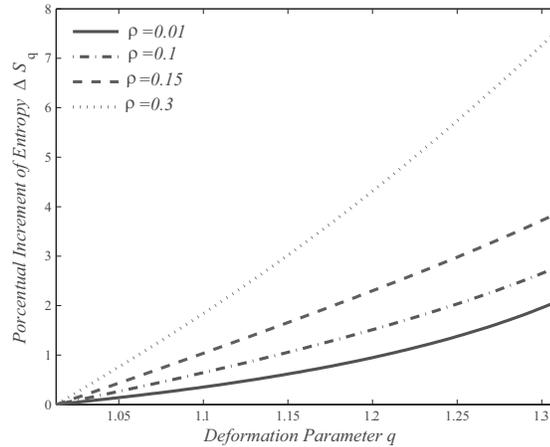


Figura 6.6: Entropía diferencial normalizada. Incremento porcentual de la entropía (‘Percentual Increment of Entropy’) como función del parámetro de deformación (‘Deformation Parameter’) q , como definida en la ecuación (6.18) considerando diferentes valores de correlación en la salida $\rho = 0,01$, $\rho = 0,1$, $\rho = 0,15$ y $\rho = 0,3$.

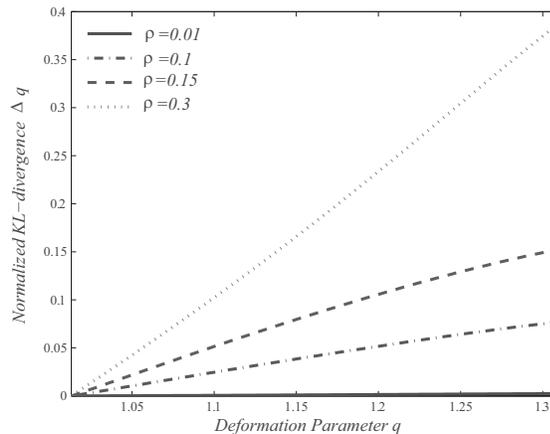


Figura 6.7: Divergencia de Kullback-Leibler normalizada (‘Normalized KL-divergence’) versus el parámetro de deformación (‘Deformation Parameter’) q , dados los valores de correlación en la salida $\rho = 0,01$, $\rho = 0,1$, $\rho = 0,15$ y $\rho = 0,3$. La distribución referencia se toma con $\rho = 0$ para cada valor de ρ , y es normalizada a la distribución con $q = 1$ como en la ecuación (6.19).

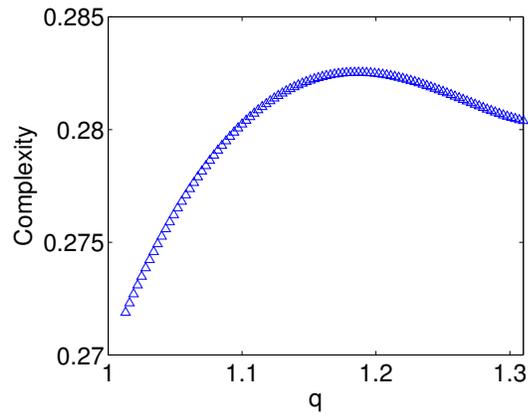


Figura 6.8: Complejidad ('Complexity') en función del parámetro de deformación q , para el rango de valores utilizados en este capítulo. Se utilizaron para este ejemplo $\mu = 0,08$ y $\rho = 0,7$

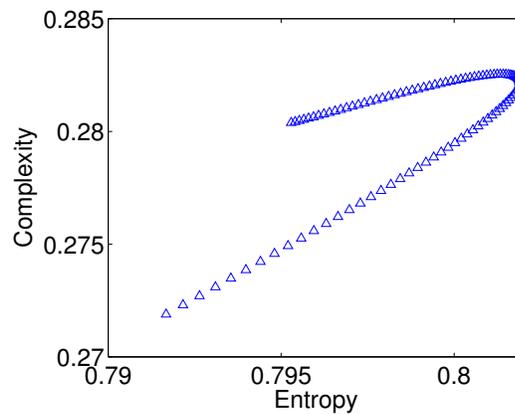


Figura 6.9: Complejidad ('Complexity') versus entropía ('Entropy'). Cada punto se corresponde para con un valor distinto de q . Se utilizaron para este ejemplo $\mu = 0,08$ y $\rho = 0,7$

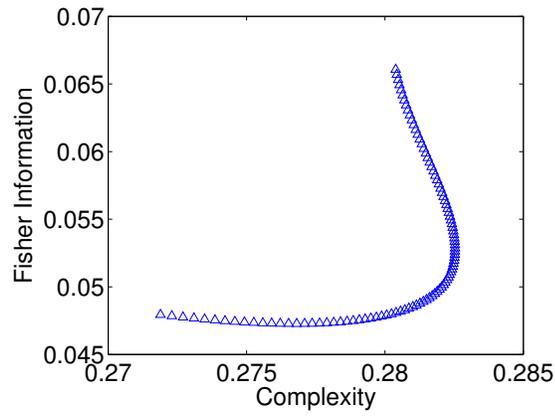


Figura 6.10: Información de Fisher ('Fisher Information') versus complejidad ('Complexity'). Cada punto se corresponde para con un valor distinto de q . Se utilizaron para este ejemplo $\mu = 0,08$ y $\rho = 0,7$

Capítulo 7

Propiedades de escala de las entradas con distribuciones con colas pesadas

A pesar de la extensa utilidad del enfoque simple del modelo DG, una pregunta particular resulta eludida: ¿qué sucede cuando la estadística de la suma de las entradas a una población de neuronas (o del potencial de membrana de dichas neuronas) no sigue una distribución gaussiana? Registros experimentales *in vivo* de fluctuaciones sub-umbral del potencial de membrana en la corteza auditiva primaria de rata, tanto en preparaciones con y sin anestesia, mostraron dinámicas que consistían de excursiones mucho más extensas que las pequeñas fluctuaciones predichas por modelos de caminata aleatoria [120, 121]. Por otra parte, la distribución de la actividad del potencial de campo local medido en [34] resultó poseer colas más pesadas que lo esperado por el ajuste a una distribución gaussiana utilizado para el modelado. Por lo tanto, la extensión propuesta para el modelo DG [74, 75, 122] parece una buena candidata para estudiar esas desviaciones con la inclusión de un único parámetro libre.

En este capítulo, nos concentraremos en la curtosis estadística en la distribución de la entrada como medida de no-gaussianidad. Definimos tres regímenes distintivos de la estadística del potencial de membrana: gaussiano (*I*), no gaussiano con curtosis grande pero finita (*II*), y no gaussiano con curtosis infinita (*III*). Encontramos que

diferentes dinámicas implican cambios sustanciales en las estrategias de codificación, dependiendo del tamaño de la población considerada. Además, lo caracterizamos en términos de cantidades físicas ampliamente utilizadas para estudiar estadísticas poblacionales y que previamente han sido utilizadas en el campo de la neurociencia.

7.1. Modelando potenciales de membrana

Gran parte de lo que sabemos sobre la estadística de las actividades poblacionales en la corteza proviene de observar las secuencias de potenciales de acción en el registro de las mismas. Las tasas de disparos medias, las correlaciones entre pares e incluso las correlaciones de alto orden han sido minuciosamente estudiadas de esta manera. Sin embargo, cómo esta actividad se relaciona con las señales de entrada recibidas por dichas poblaciones es de gran relevancia. Una neurona cortical típica recibe usualmente entradas de aproximadamente otras 10^3 neuronas que pueden estar tanto dentro de la misma columna cortical como en partes más distales de la corteza y en otros lados. Esto sugiere que las neuronas corticales se encuentran embebidas en una red de señales altamente convergentes y que identificar estas fuentes numerosas y altamente dispersas puede ser una tarea abrumadora. Alternativamente, se podría inferir la actividad de esta sub-población mediante el registro del potencial de membrana de la neurona individual que recibe esta entrada. Específicamente, la dinámica del potencial de membrana establece restricciones en el grado en que el conjunto presináptico coopera para excitar la neurona registrada para que dispare [120]. Un enfoque teórico muy utilizado (y particularmente por el modelo DG) es el de considerar que la entrada a esta neurona consiste de una gran cantidad de contribuciones pequeñas sumadas entre ellas. En este caso, de acuerdo al teorema del límite central el potencial de membrana sigue una distribución gaussiana (asumiendo varianza finita) y sus fluctuaciones siguen un caminante aleatorio. No obstante, el teorema del límite central no resulta válido si las correlaciones entre variables aleatorias no son despreciables. Este sería el caso en el que la población presináptica se encuentre altamente correlacionada: las neuronas estarían inactivas la mayor parte del tiempo, excepto durante momentos breves cuando grandes grupos disparan de manera sincronizada. Resultaría imposible distinguir entre ambas posibilidades (o

incluso casos intermedios) estudiando solamente los patrones de actividad de la población, dado que ambos modelos podrían tomar en cuenta esencialmente cualquier tren de espigas observado.

Una característica prominente de esta dinámica no gaussiana está dada por la ocurrencia más frecuente de los valores atípicos, que es la actividad “todo o nada”. Por ende, una distribución que posea colas más pesadas sería requerida para capturar este fenómeno. Esta sobre-representación de valores atípicos en los datos experimentales es sinónimo de dispersión estadística y las señales dispersas han recibido mucha atención [123–125]. La mecánica estadística no extensiva trata con variables aleatorias fuertemente correlacionadas, referida a veces como correlación global [43, 96]. La generalización de la entropía clásica como la base para extender la teoría de Boltzmann-Gibbs, alcanza su máximo en las distribuciones referidas como q -gaussianas. Este hecho, y un número de conjeturas e indicaciones numéricas [97, 99, 100] sugieren la existencia de un teorema del límite central q -análogo. En general, este teorema del límite central parece no ser tratable si nos basámos únicamente en el álgebra clásico. Sin embargo, la mecánica estadística no extensiva utiliza construcciones basadas en un álgebra referido como q -álgebra, y por lo tanto recientemente fue mostrado que, en dicho marco teórico, la correspondiente generalización del teorema del límite central resulta posible y relativamente simple.

Dado el modelo de q -Gaussiana Dicotomizada, ¿cómo podemos definir un “distancia” que cuantifique cuán relevante son las colas de la distribución de la entrada a la población de células?. Por ejemplo, en el análisis de [120] se utilizó la curtosis en exceso, definida como el cociente del cuarto momento central dividido por la cuarta potencia de la desviación estándar, menos tres (que es el valor correspondiente a la curtosis de una distribución gaussiana). En este trabajo, encontraron grandes valores de esta cantidad en las distribuciones de sus registros de potenciales de membrana sub-umbral realizados con la técnica de fijación de membranas en muchos casos. La medida estándar de curtosis, originada por Pearson, se encuentra relacionada a las colas de la distribución: valores más altos de curtosis son el resultado de desviaciones extremas infrecuentes (o valores atípicos), en contraposición a desviaciones moderadas pero frecuentes. Distribuciones con curtosis mayor a tres (o curtosis en exceso positivas) son denominadas como leptocúrticas. Valores predominantemente positi-

vos para las trazas registradas en la corteza auditiva de ratas despiertas reflejan la naturaleza unimodal de las distribuciones de voltaje de membrana con colas largas.

Para las distribuciones q -gaussianas, los excesos de curtosis pueden ser definidos como

$$k_{\mathcal{G}_q} = \begin{cases} 6 \left(\frac{q-1}{7-5q} \right) & \text{para } 1 < q < 7/5, \\ \infty & \text{para } 7/5 < q < 5/3. \end{cases} \quad (7.1)$$

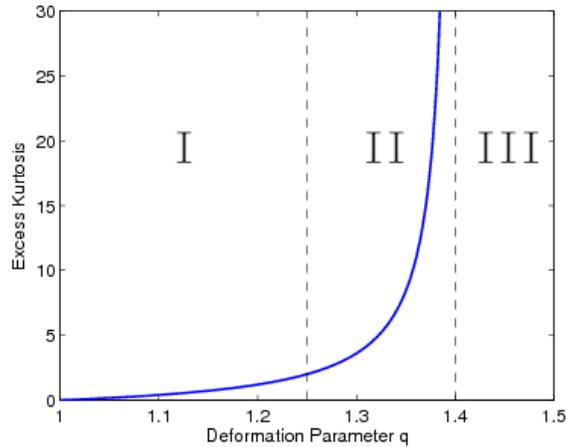


Figura 7.1: Curtosis en exceso (‘Excess Kurtosis’) de la distribución q -gaussiana en función del parámetro de deformación (‘Deformation Parameter’)

Esto mostraría una relación explícita entre el parámetro q y $k_{\mathcal{G}_q}$, y que las distribuciones q -gaussianas son leptocúrticas en este dominio. Sin embargo, esta dependencia no es directamente proporcional: puede establecerse que existen tres regiones definidas. Primero, el valor de curtosis incrementa lentamente a medida que aumenta q y luego comienza a variar abruptamente dado que se acerca a una asíntota. A continuación, para valores de $q > 5/7$ la curtosis diverge dada esta definición. Por ende, nombramos a estas regiones como cuasi-caminata aleatoria ($1 < q < 5/4$), caminata aleatoria mixta ($5/4 < q < 7/5$) y no-caminata aleatoria ($q > 7/5$) respectivamente, en relación a la dinámica esperada para estos valores (ver Figura 7.1). El límite para definir donde termina la primer región y empieza la segunda depende del criterio elegido: i) podemos establecer $q = 5/4$ como límite ya que corresponde con el valor máximo tolerable para que la curtosis en exceso sea considerada parecida a la normal

o ii) evaluando la primer y segunda derivada es posible determinar un punto para el cual la curtosis comienza a cambiar más rápidamente. Además de estos detalles, es necesario tener en cuenta que esta cantidad es una medida estática y no permite distinguir duraciones temporales entre excursiones.

7.2. Resultados

7.2.1. Coeficientes de interacción del modelo q -DG

Recordemos que en el contexto de la geometría de la información (ver capítulo 2), el disparo de las neuronas en una población de tamaño N está representado como un vector binario $\mathbf{x} = (x_1, \dots, x_N)$, donde $x_i = 0$ si la neurona i se encuentra inactiva en una cierta ventana temporal ΔT y $x_i = 1$ si está disparando. Entonces, consideramos el conjunto de distribuciones de probabilidades de vectores binarios, $\{P(\mathbf{x})\}$, que consiste de las 2^N probabilidades

$$P(\mathbf{x}) = \text{Prob}\{x_1 = i_1, \dots, x_N = i_N\} = p_{i_1 \dots i_N} \quad (7.2)$$

sujetas a la normalización $\sum_{i_1, \dots, i_N=0,1} p_{i_1 \dots i_N} = 1$.

Como propuesto por Amari y colaboradores [26], provisto que $P(\mathbf{x}) \neq 0$, cualquier distribución de este tipo puede ser expandida de la forma dada por la ecuación 1.2, donde existen un total de $2^N - 1$ coeficientes de correlación θ diferentes que pueden ser usados para determinar unívocamente la distribución de probabilidad. Dado que estimar todos estos parámetros de interacción resulta computacionalmente demandante debido a la explosión combinatoria, una versión más simple de este modelo de máxima entropía sería considerar que los coeficientes de cada orden son iguales. Esto es, todos los parámetros que caracterizan las propiedades de cada neurona individual y las interacciones entre cualquier grupo de neuronas no dependen de la identidad precisa de las neuronas consideradas, sino del número de neuronas consideradas [24]. Tomando en cuenta esta suposición, la distribución de probabilidades se encuentra ahora caracterizada por sólo N parámetros: dada la simetría de la población, todas las coordenadas θ de un dado orden k son equivalentes y pueden ser representados por un parámetro θ_k . Por ejemplo, todos los coeficientes de interacción de orden 3

θ_{ijk} equivalen a un único parámetro que indicamos como θ_3 . Bajo esta suposición de población homogénea, la ecuación (1.2) está dado ahora por

$$P(\mathbf{x}) = \exp \left\{ \sum_{i=1}^N x_i \theta_1 + \sum_{i<j} x_i x_j \theta_2 + \sum_{i<j<k} x_i x_j x_k \theta_3 + \sum_{i<j<k<l} x_i x_j x_k x_l \theta_4 + \sum_{i<\dots<N} x_1 \cdots x_N \theta_N - \psi \right\}. \quad (7.3)$$

En particular, si consideramos una población homogénea de N neuronas en el modelo q -DG, podemos calcular la probabilidad de que exactamente k neuronas se encuentren disparando en una dada ventana temporal mediante la ecuación (6.14). Por simetría, todos los patrones de actividad con el mismo número de espigas son igualmente probables. Por lo tanto, el modelo es completamente caracterizado por el número de espigas poblacionales k .

Las ecuaciones (7.3) y (6.14) pueden ser igualadas para extraer información de cómo estos coeficientes de correlación dependen de la estadística de la entrada, como fue realizado previamente en [114] para el modelo DG. Por lo tanto, mediante la resolución de ecuaciones lineales, podemos calcular los parámetros θ_i ($k = 1, \dots, N$) y el factor de normalización ψ . Para el modelo DG fue reportado que los coeficientes de interacción mostraban una alternancia de signos con respecto a los órdenes sucesivos. Por ende, este enfoque provee una buena idea de como es la estructura de las correlaciones de la respuesta neuronal.

En este capítulo nos concentramos en el régimen de baja tasa de disparos, que corresponde a los valores reportados en la bibliografía relacionada [120]. Como puede ser observado en la Figura 7.2, una desviación suficiente de la curtosis esperada de un potencial de membrana gaussiano rompe la alternancia en los signos esperada para el caso $q = 1$ [114]. Además, los valores de los coeficientes disminuyen con el parámetro de deformación, sugiriendo una estructura de correlación de alto orden mucho menos organizada. Aún así, este efecto depende del valor de la covarianza en las señales de entrada α en relación a q y a la tasa de disparos de la salida μ (que a su vez está relacionada al umbral de cada neurona).

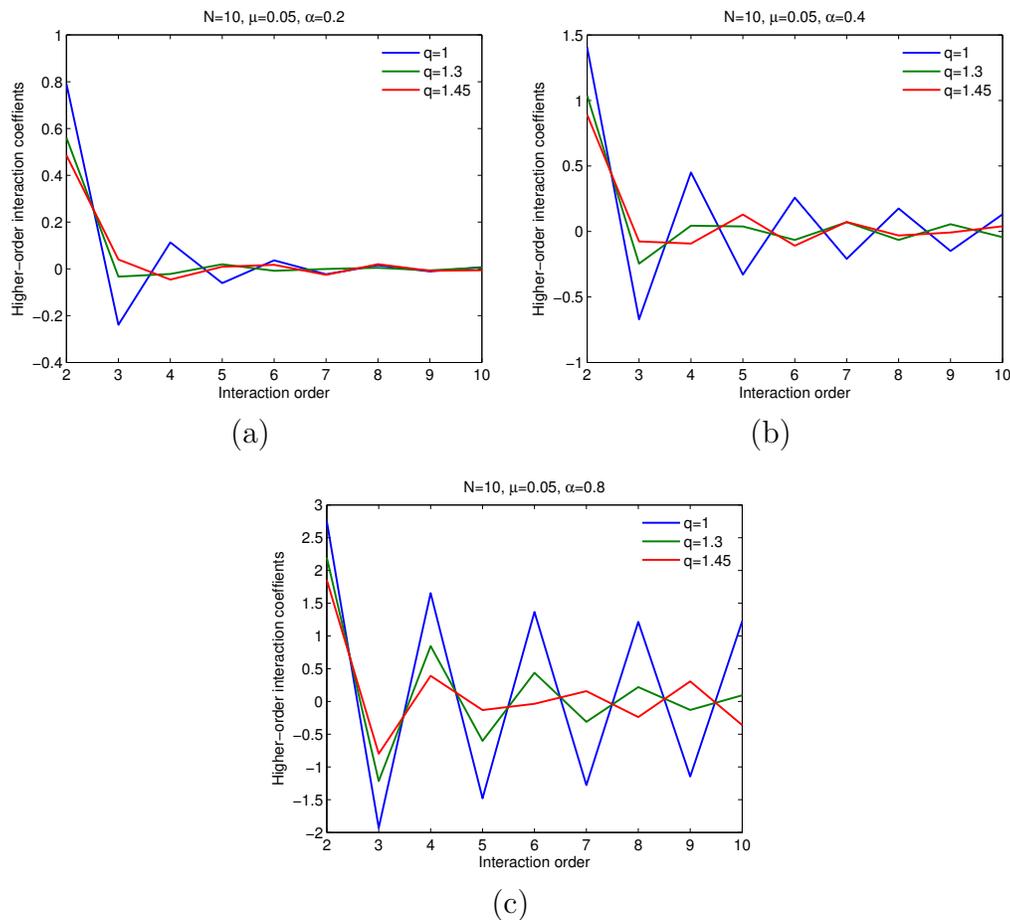


Figura 7.2: Valores de los coeficientes de interacción de alto orden (‘Higher-order interaction coefficients’) para cada orden de interacción (‘Interaction order’) de una población de 10 neuronas, para distintos parámetros de desviación del modelado gaussiano. Se considera el caso de baja tasa de disparos $\mu = 0,05$ con: (a) covarianza en la entrada $\alpha = 0,2$; (b) covarianza en la entrada $\alpha = 0,4$; (c) covarianza en la entrada $\alpha = 0,8$.

7.2.2. Esfuerzo

Particularmente, la componente de máxima entropía de tercer orden (también llamada “esfuerzo”) ha sido extensamente utilizada como medida de desviación de los modelos de máxima entropía de interacciones entre pares en comparación a la divergencia Kullback-Leibler [54] (ver capítulo 4). Entonces, los patrones de disparos de tripletes proveen un entendimiento del comportamiento de la red: si todos los

subconjuntos de tres neuronas poseen un esfuerzo positivo, la actividad sincrónica a través de toda la población resulta facilitada, mientras que si todos los subconjuntos de neuronas tienen un esfuerzo negativo, la sincronización global se encuentra suprimida. Utilizando esta medida, se encontró que los patrones de disparo locales son distintivos: mientras que los patrones de actividad multi-neuronales a grandes distancias pueden ser predichos mediante modelos de interacciones entre pares, los patrones de disparo dentro de grupos locales suelen mostrar evidencia de correlaciones de alto orden [25]. Por ende, estas correlaciones locales implican una organización jerárquica y reducen modestamente la cantidad de información que un grupo transmite, mientras modifican el formato de esta información, creando códigos más dispersos al incrementar los periodos de quiescencia total y concentrando la información en periodos más cortos de actividad común.

El esfuerzo, como definido por [54], de una distribución sobre tres variables aleatorias binarias es el término de interacción de tercer orden en el modelo de máxima entropía de tercer grado y puede ser calculado como:

$$\theta_3 = \ln \frac{p_{111}p_{100}p_{010}p_{001}}{p_{110}p_{101}p_{011}p_{000}}. \quad (7.4)$$

Para un ensamble de tres neuronas idénticas, dada la simetría, resulta suficiente evaluar las probabilidades p_{000} , p_{001} and p_{011} (o cualquier otro conjunto válido de probabilidades, dados uno, dos o ningún disparo). Cada una de estas tres probabilidades pueden ser calculadas numéricamente para el modelo q -DG, dado que no existe una expresión analítica cerrada en general. En trabajos previos, la coordenada de interacción triple del modelo DG fue encontrada ser siempre negativa y decrecer monótonamente al incrementar los coeficientes de correlación entre pares [33, 114].

De la Figura 7.3 podemos observar que las entradas comunes pueden llevar a un esfuerzo cero o incluso positivo si la distribución del potencial de membrana subumbral es suficientemente no gaussiana. Esta condición también depende de la tasa de disparos de la salida y la correlación entre pares de la entrada. A medida que la probabilidad de actividad individual y la covarianza entre unidades respecto a la entrada común disminuyen, estos regímenes se pueden alcanzar. Sin embargo, asumir una dinámica de caminante aleatorio en la actividad latente torna imposible obtener valores no negativos del esfuerzo. Se debe tener en cuenta además que este

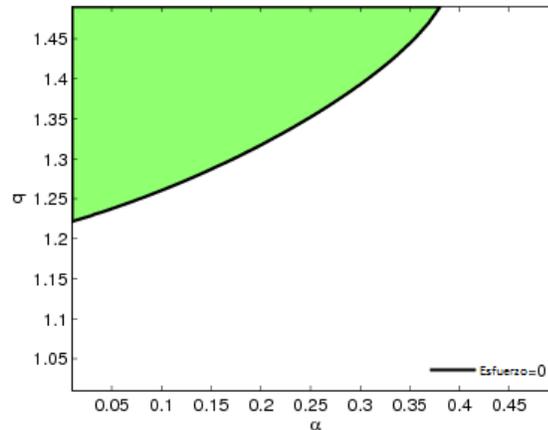


Figura 7.3: Valores del esfuerzo de un conjunto de tres neuronas idénticas, con tasas de disparo baja $\mu = 0,05$. La región verde corresponde a un esfuerzo positivo, mientras que la región blanca a un esfuerzo negativo.

resultado es válido para un modelo de tres neuronas y no es necesariamente válido para ensambles de mayor tamaño.

7.2.3. Dispersión estadística

Por otro lado, si quisiésemos estudiar poblaciones más grandes, dado que el número de parámetros depende del tamaño de la población, se proponen representaciones minimales adicionales de estos rasgos de alto orden. Particularmente, trabajos previos han concluido que estas interacciones de alto orden se pueden resumir concisamente en patrones de silencio simultáneo [114], restringiendo la actividad neuronal y, por lo tanto, influenciando el procesamiento de la información. El silencio simultáneo se refiere a la co-inactivación neuronal: la ocurrencia de periodos breves de silencio durante los cuales todas las neuronas en la red local dejan de disparar. La dispersión estadística es considerada una componente importante en la codificación poblacional [25]. También se demostró que aumenta con las correlaciones [33] y que depende de la inhibición en la red [114]. Para que sea posible evaluar si la presencia de colas pesadas en la distribución del potencial de membrana (que sería una marca distintiva de la dispersión estadística en la señal de entrada a las neuronas) impacta los patrones de

actividad de la población en el modelo q -DG, cuantificamos esta cantidad como la probabilidad de que la población se encuentre inactiva, esto es $P(k = 0) \cdot N$. Para el caso $q = 1$, se encontró que en poblaciones homogéneas las correlaciones incrementan la dispersión [33] y fue conjeturado que en conjuntos heterogéneos de mayor tamaño la dispersión se encuentra todavía dominada por interacciones de alto orden.

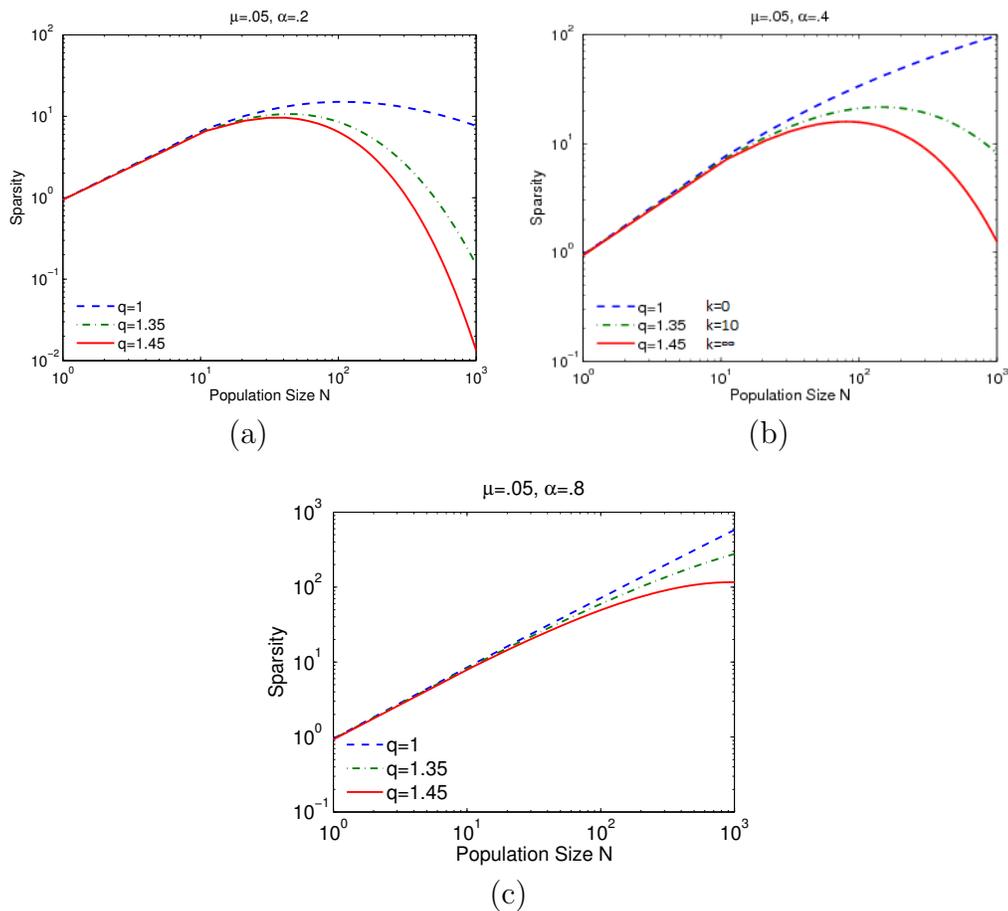


Figura 7.4: Dispersión estadística (‘Sparsity’) en función del tamaño de la población (‘Population Size’), para distintos parámetros de desviación del modelado gaussiano. Se considera el caso de baja tasa de disparos $\mu = 0,05$ con: (a) covarianza en la entrada $\alpha = 0,2$; (b) covarianza en la entrada $\alpha = 0,4$; (c) covarianza en la entrada $\alpha = 0,8$.

Como es posible ver en la Figura 7.4 para valores grandes de q , y por lo tanto desviaciones grandes de dinámicas de caminata aleatoria, la dispersión decrece relevantemente para valores mayores a 10^2 . Este efecto se ve minimizado para valores

extremos de covarianzas entre neuronas con respecto a las señales entrada, como por ejemplo $\alpha \simeq 0,8$, para estos valores de tasas de disparo. Aunque hemos asumido elementos idénticos, y ese no es el caso para una población neuronal real, se sugiere que en este régimen existe la necesidad que las unidades se encuentren fuertemente correlacionadas con respecto al estímulo para que la dispersión de sus actividades, y por lo tanto su eficiencia de codificación, no se reduzca. Como previamente expuesto en el capítulo 6, el parámetro de deformación q afecta el valor del coeficiente de correlación entre pares ρ . Para la tasa de disparo media elegida para este análisis ($\mu = 0,05$), las colas pesadas en la distribución disminuyen la correlación entre pares de neuronas, para una dada covarianza con respecto a la entrada.

Tal disminución en la dispersión estadística para valores pequeños de correlación entre pares fue reportado para el modelo DG [33]. En este marco, la pregunta central sería entonces cuál puede ser la ventaja de la no-gaussianidad en la distribución de los potenciales de membrana.

7.2.4. Tasa de entropía y calor específico de la población

Para responder esta pregunta estudiamos primero cómo la entropía del modelo crece con el tamaño de la población. El comportamiento asintótico de las entropías se calculó previamente en el capítulo 5 para el modelo q -DG, esto es cuando $N \rightarrow \infty$, pero estamos interesados en cómo escala la tasa de entropía con N . Es posible calcular esta última cantidad como

$$S/N = \frac{1}{N} \sum_k P(K = k) \ln \binom{N}{k}, \quad (7.5)$$

donde S es la entropía (ver capítulo 2) y N el tamaño de la población.

En la Figura 7.5, puede ser observado que para valores más altos de curtosis la tasa de entropía aumenta. Esta diferencia no cambia para tamaños mayores a 10^2 , cuando la entropía alcanza su valor asintótico. En cualquier caso, estos valores de entropía más altos en comparación al caso $q = 1$ escalan de forma distinta a la diferencia entre el modelo DG y el modelo de máxima entropía de segundo orden [32, 33], sugiriendo que la razón por la misma es distinta.

Finalmente, estudiamos el impacto de la cantidad q en la capacidad calorífica de

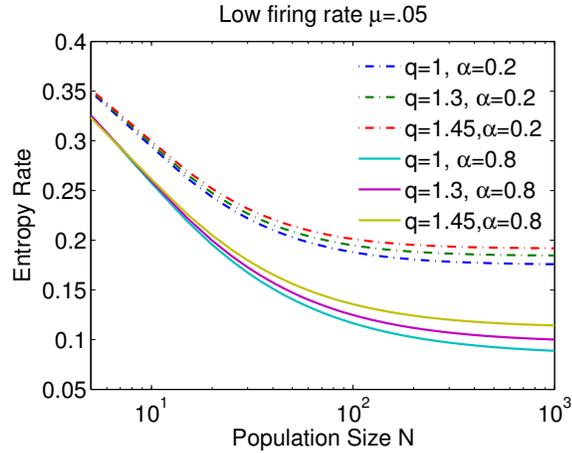


Figura 7.5: Tasa de entropía (‘Entropy Rate’) en función del tamaño de la población (‘Population Size’) con tasa de disparo baja (‘Low firing rate’) $\mu = 0,05$, para distintos valores de q y α .

la población. Calculamos el calor específico (es decir, la capacidad calorífica normalizada por el tamaño de la población) como

$$HC/N = \frac{1}{N} \sum_k P(K = k) \left(\ln P(K = k) - \ln \binom{N}{k} - S \right)^2, \quad (7.6)$$

El modelo DG muestra un calor específico divergente con un pico puntiagudo [33]. Esto es evidencia de que un sistema físico se encuentra en un punto crítico e implica que las fluctuaciones alrededor de la entropía no son despreciables en grandes poblaciones [21]. Además, fue demostrado válido en general para modelos con entradas comunes, siendo capaces de capturar este comportamiento [33]. Adicionalmente, concluyeron que las interacciones de alto orden pueden tener un impacto fuerte en el calor específico: llevan a comportamientos de escala cualitativamente diferentes y criticalidad, influenciando los momentos que lo maximizan [33].

La Figura 7.6 nos indica que el calor específico en el modelo q -DG muestra un comportamiento similar al de la tasa de entropía en relación al parámetro q . Es decir, escala linealmente a medida que aumenta el tamaño de la población, sin importar el valor de curtosis de la distribución de entrada. No obstante, mientras que las pendientes de las rectas permanecen cercanas en magnitud en los valores correspondientes a

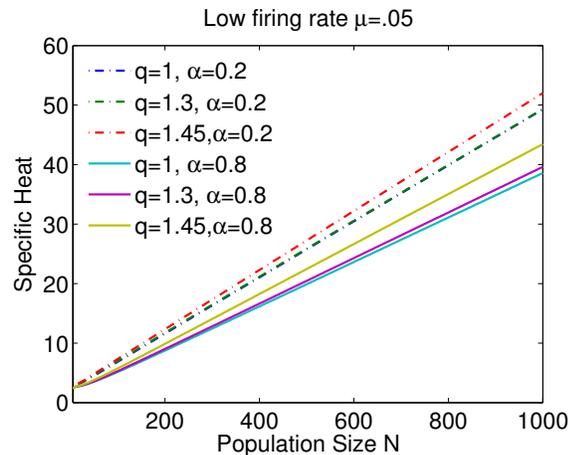


Figura 7.6: Calor específico ('Specific Heat') en función del tamaño de la población ('Population Size') con tasa de disparo baja $\mu = 0,05$, para distintos valores de q y α .

las regiones *I* and *II*, aumentan considerablemente para aquellos parámetros dentro de la región *III*. Este efecto se muestra más evidente para poblaciones mayores y valores de correlación más altos. Por ende, el modelo muestra lo que parece ser una compensación una compensación entre la dispersión estadística y la criticalidad.

Las entradas comunes a una población de neuronas conducen a patrones de actividad dispersos y con estadísticas de alto orden. Esto es un resultado ubicuo al considerar modelos gaussianos de la distribución del potencial de membrana de cada célula. Sin embargo, desviaciones de los mismos caracterizadas por el valor de la curtosis muestran un efecto relevante en la salida. Es decir, que los silencios pueden influir en la estadística de dispersión de manera de convertir el código neuronal en un código correlacionado pero eficiente.

Capítulo 8

Discusión y Conclusiones

La presencia e impacto de las correlaciones de órdenes más allá de dos en la actividad cortical resulta un tema relevante y de alto interés para la comunidad neurocientífica. A medida que los avances tecnológicos permiten el registro en simultáneo de los patrones de disparo de poblaciones de mayor tamaño, desarrollar herramientas y marcos teóricos de análisis de datos experimentales se torna una labor interesante y a la vez desafiante. En la diversa literatura presentada, ha quedado en evidencia que los enfoques tradicionales han demostrado no ser suficientes en este área. Las propiedades emergentes para cualquier sistema complejo no pueden ser reducidas en general a interacciones entre pares [14]. Importantemente, si tripletes o correlaciones de alto orden se encuentran presentes en el conjunto de datos, podrían no ser visibles y llegar a ser percibidos solamente como correlaciones de a pares [126]. En este sentido, a lo largo de esta Tesis Doctoral hemos, a grandes rasgos, presentado dos enfoques diferentes y complementarios para el estudio de interacciones de alto orden.

El primero de los enfoques ha sido la inclusión de términos de grado mayor a dos en la expansión en series para ventanas temporales cortas de la información mutua entre un conjunto de estímulos y la respuesta de una población pequeña [9–11, 86]. A diferencia de otras medidas de correlación más sencillas, la información tiene la ventaja única de capturar todas las dependencias no lineales de cualquier orden estadístico que puedan llegar a estar presentes en los datos. La simple comparación entre la información de un conjunto de neuronas con la suma de la información

provista por cada neurona individual nos permite investigar diversas formas de correlaciones [13, 83]. Resulta natural preguntarse entonces bajo qué condiciones grupos de neuronas representan el conjunto de estímulos y el comportamiento directo en maneras tanto sinérgica como redundante o independientemente. En este sentido, un punto crucial para entender la representación de estímulos externos en la actividad de una población de neuronas es investigar cómo la información transmitida por células individuales se relaciona en conjunto [12].

Como desarrollado en el capítulo 3, la ventaja principal de nuestro método es que nos permite tener en cuenta las variables ocultas vinculadas dinámicamente que no se encuentran accesibles a la medida experimental. Las mismas son expresadas en nuestras estimaciones numéricas mediante un término mixto de correlaciones estímulo-dependientes “por casualidad” de segundo y tercer orden, $I_{cor-ch,3}$. Proveemos, para cualquier orden, la estimación analítica de la corrección que pueda llegar a ser producida por variables ocultas a través del nuevo término $I_{cor-ch,3}$. Se debe notar que el método de expansión en serie no debería ser aplicado para ventanas temporales post-estímulo muy largas o con poblaciones muy grandes. Estas limitaciones fueron previamente exploradas en [67], donde la generalidad de este marco teórico es contestada incluso más aún. Sin embargo, una característica importante de nuestro método es que estamos en absoluto control de cuál orden de correlación domina el conjunto de datos y la cuantificación apropiada de posible correlaciones dependientes del estímulo por casualidad a través de $I_{cor-ch,3}$.

En dicho capítulo además presentamos un estudio exhaustivo y completamente novedoso de las interacciones de alto orden y no un marco teórico mejorado de análisis a segundo orden. En este sentido, nuestro trabajo apunta a cuantificar el efecto de las correlaciones por casualidad en el procesamiento de la información en una población de neuronas. Proponemos que para cuantificar apropiadamente las correlaciones entre pares es también necesario tener en cuenta las correlaciones entre tripletes, ya que el término $I_{cor-ch,3}$ define un nuevo escenario incluso cuando no exista impacto alguno de las correlaciones de tercer orden puras. Más específicamente, dependiendo del valor de $f(\bar{\mathbf{r}})$ (una función de los valores medios de las tasas de disparo de las neuronas consideradas, en este caso un triplete de células, y que se encuentra directamente relacionada al término de correlación mixto dependiente del estímulo), cuando

el término $I_{cor-ch,3}$ es no nulo, existe una modificación sustancial en las dinámicas de correlación. Cuando esta función $f(\bar{\mathbf{r}})$ es positiva, resulta en un régimen de codificación más similar a la intuición tradicional, en donde las correlaciones tienden a resultar en redundancia, y por ende limitan el número de neuronas cuyas salidas puedan llegar a ser útilmente combinadas para representar la variable del estímulo. Por el contrario, si $f(\bar{\mathbf{r}})$ es negativa, lleva a regiones más grandes de sinergia en el plano de correlaciones de ruido versus correlaciones de señal. Importantemente, cuando $f(\bar{\mathbf{r}}) = 0$, que es equivalente a establecer $I_{cor-ch,3} = 0$, las regiones de sinergia y redundancia en el plano de correlaciones de ruido versus correlaciones de señal son cualitativamente equivalentes a las definidas por correlaciones puras de segundo orden en [9].

De esta manera, este estudio realizado para un número discreto de neuronas es el primer intento en proveer una expresión analítica para cuantificar la correlación por casualidad mixta estímulo-dependiente que involucra interacciones de ordenes más allá del segundo. Este trabajo no sólo demuestra rigurosamente que las correlaciones no implican redundancia, pero también que las correlaciones de alto orden pueden incrementar la cantidad de sinergia. Además, como se ha explicado en el capítulo 3, corroboramos diversas hipótesis utilizando resultados numéricos mediante simulaciones con estadística de Poisson, un modelo de integrar-y-disparar y un modelo simple de Izhikevich. Mostramos que las correlaciones dependientes del estímulo no deben ser despreciadas: el efecto de las correlaciones estímulo-dependientes de tercer orden aumentan la fracción de sinergia. Adicionalmente, expresamos que los tripletes por casualidad posibles que puedan haber surgido de correlaciones entre pares contribuyen a una subestimación de la información si no son tomados en cuenta, y que un incremento en la conectividad de la red puede reflejarse en un mayor valor de correlaciones que dependen del estímulo considerado. La posibilidad de sinergia con correlaciones de alto orden es un tema importante ya que prueban ser una propiedad inherente de la dinámica cortical. Nuestro presente método podría ser también una herramienta complementaria muy útil para otros métodos de poblaciones grandes, en el que se estudian las correlaciones sin tener control del orden exacto de las mismas [33, 34, 75].

En el capítulo 4 estudiamos los valores posibles de estas correlaciones de alto

orden por casualidad. Para esto, utilizamos el marco teórico provisto por la geometría de la información [26]. Calculamos la mejor estimación de estas cantidades según el principio de máxima entropía. De esta manera, no sólo podemos estudiar el impacto de las correlaciones de alto orden en la transmisión de la información en una población de neuronas, sino que también podemos testear su presencia y relevancia en los datos experimentales. Como próximo objetivo, realizaremos el análisis de datos de registros neuronales de cientos de neuronas de la corteza visual para determinar con datos reales la cantidad de correlación espúrea cuando consideramos el principio de máxima entropía.

El segundo enfoque corresponde a la extensión de un modelo poblacional simple de neuronas con umbral. Este modelo para poblaciones grandes no pretende ser una descripción biológicamente realista de una red cortical, sino más bien un modelo minimal que refleja la estadística de la misma. El modelo de Gaussiana Dicotomizada refleja la actividad de un conjunto de neuronas que reciben entradas comunes pero que no poseen interacciones más allá del segundo orden [31, 32]. En primera instancia, en el capítulo 5, presentamos un marco teórico en el teorema extendido del límite central [97–100] que permite la consideración de correlaciones de alto orden en la distribución de los potenciales de acción de las neuronas modeladas. Nos basamos en las herramientas provistas por la mecánica estadística no extensiva [43] para estudiar el impacto de la sincronización en la señal de entrada a una población de neuronas binarias, mediante la inclusión de un parámetro de deformación q . El razonamiento principal para justificar esta extensión es que, al tomar un número muy grande de contribuciones neuronales dentro del teorema del límite central, estamos perdiendo información sobre las correlaciones de alto orden. Importantemente, nuestra aproximación converge al modelo DG al considerar el límite $q = 1$. En dicho capítulo, proveemos la expresión analítica exacta de la distribución conjunta de las tasas de disparo de una población de neuronas modelada de esta manera. Además, mostramos utilizando este resultado cómo las correlaciones de alto orden pueden mejorar la codificación de la información en poblaciones grandes de neuronas. Por lo tanto, inferimos las propiedades emergentes de las mismas, exponemos cómo la información satura a medida que el grado de correlación q aumenta. Evaluamos también bajo qué condiciones las correlaciones de alto orden en las entradas pueden llevar a

un número menor o mayor de disparos simultáneos en la actividad poblacional.

Luego, en el capítulo 6, elaboramos cómo poder construir, en este marco, un modelo generador de patrones de actividad con una estadística deseada. Esto es, el proceso latente (es decir, las variaciones sub-umbrales) se modela mediante una distribución q -gaussiana correlacionada con media h y correlación A , que son elegidos tal que las secuencias de disparo de la salida X poseen media μ y varianza entre pares Σ . La generación de patrones de actividad a través de la binarización de una variable aleatoria q -gaussiana multivariada es posible incluso para poblaciones grandes de neuronas y para estadísticas de conteo de espigas con distribuciones marginales arbitrarias. Esto provee un marco de trabajo flexible para la simulación eficiente de trenes de espiga con correlaciones específicas, incluso cuando sus fuentes y orígenes no son conocidos. Adicionalmente, puede ser implementado numéricamente con facilidad dado que, como mostrado previamente, estas distribuciones son equivalentes a las distribuciones $-t$ de Student en las regiones estudiadas, comúnmente utilizadas e incluidas en diversos paquetes numéricos. Nuestra técnica puede ser aplicada a experimentos en neurofisiología incluyendo micro-estimulación de poblaciones de neuronas, expandiendo las posibilidades de la investigación de las estimulaciones naturales que puedan generar diferentes estructuras de correlación y el estudio de sus efectos en conducir respuestas neuronales.

Además, mostramos que nuestro enfoque captura redundancias que aumentan a medida que el parámetro q toma valores más grandes, llevando a una mayor cantidad de aleatoriedad dado que la entropía crece y por ende aumenta el caos. Importante-mente, a través de una medida de la entropía relativa que diverge cuantificamos la ineficiencia de asumir incorrectamente que la distribución se genera como en el modelo DG. Esto demuestra que las correlaciones de alto orden en la entrada poseen un efecto relevante a nivel poblacional. Dada la ubicuidad de las entradas comunes en los sistemas sensoriales, el modelo extendido otorga entonces predicciones cuantitativas para las condiciones bajo las cuales las correlaciones de alto orden serán importantes, cuantificando cuánto se desvían efectivamente los modelos que solamente consideran correlaciones de segundo orden en la entrada común.

Finalmente, en el capítulo 7, estudiamos el efecto funcional que puedan tener las correlaciones de alto orden en las entradas de poblaciones de neuronas, carac-

terizandolas mediante la curtosis de la distribución del potencial de membrana. En un principio, utilizando el marco de trabajo que provee la geometría de la información mostramos cómo la inclusión del parámetro de deformación q (que en este capítulo vinculamos explícitamente a la curtosis) desestructura la actividad de salida. Sin embargo, también permite un mecanismo para lograr valores nulos (e incluso positivos) de esfuerzo, característica poblacional que no era posible alcanzar con el modelo DG. Esto habilita a patrones de actividad más diversos, como consecuencia de la consideración de las correlaciones de alto orden. Por lo tanto, la inclusión de interacciones más allá de entre pares en las entradas comunes a una población tiene un efecto significativo, que varía con el tamaño del conjunto considerado. De esta manera, podemos además consolidar este enfoque con el análisis realizado para tamaños finitos en la aproximación en series de la información mutua para ventanas temporales cortas.

Adicionalmente, la criticalidad ha sido recientemente un punto de discusión en el estudio de poblaciones neuronales [127, 128]. Permanece en debate si se trata de un simple producto colateral de las suposiciones teóricas hechas o si es un aspecto funcional de la actividad de la corteza. Mediante nuestro modelo generador expandimos los límites de la discusión al encontrar un vínculo con las colas de la distribución del potencial de membrana de una neurona. Los sistemas cerca de estados críticos responden óptimamente a ciertas señales externas y esta sensibilidad puede ser útil para la corteza auditiva en particular. En la literatura expuesta [120, 121] se ha propuesto que los potenciales de membrana no gaussianos en la corteza auditiva en sujetos experimentales despiertos puede reflejar una adaptación al rápido procesamiento de los estímulos auditivos. La mayoría de los sistemas que exhiben criticalidad en un sentido termodinámico también exhiben un amplio rango de escalas temporales en sus dinámicas, y entonces la criticalidad puede proveer una estrategia general para los sistemas neuronales para unir el vacío entre la espiga microscópica y el comportamiento macroscópico. Aunque nuestro modelo puede ser aplicado a diversas poblaciones, permanece el interrogante de por qué las dinámicas sub-umbrales en la corteza auditiva difieren con la dinámica de otras cortezas sensoriales. No obstante, una posibilidad intrigante es que esta diferencia se relacione con la habilidad de la corteza auditiva en explotar la temporalidad precisa de los estímulos sensoriales

breves [129].

Por esta razón, en el capítulo mencionado, comparamos cómo varía el calor específico de una población modelada con nuestro método en función del parámetro q . Se ha sugerido que la tasa a la cual diverge la capacidad calorífica por unidad provee un medio para cuantificar la “fuerza” de la criticalidad [128]. Para el modelo DG [33] se habían encontrado formas funcionales similares a las curvas de calor ilustradas en el capítulo 7. Además, demostraron que la variación del calor específico aumenta de forma estrictamente monotonamente creciente con la correlación. Los resultados previamente mostrados en el capítulo 6, en los cuales se especificó que para el régimen de bajas tasas de disparo la correlación no variaba fuertemente, sugieren que la curtosis en la entrada (o q , equivalentemente) es un factor relevante.

Por otro lado, la dispersión estadística es una propiedad que ha sido consistentemente asociada con la codificación óptima y relacionada ampliamente con la capacidad de representación de las poblaciones neuronales [123]. En el marco de nuestro modelo, encontramos que a pesar de que valores mayores de curtosis en la distribución de las entradas (y por ende, colas más pesadas) favorecen el calor específico, también disminuyen la dispersión cuantificada por la probabilidad que todas las neuronas de la población se encuentren inactivas. Esto sugiere que dinámicas diferentes a las de una caminata aleatoria en las variaciones del potencial de membrana son beneficiosos para poblaciones de tamaños menores a cien. Por lo tanto, poblaciones mayores pueden consistir de un número finito de sub-grupos homogéneos, en cuyo caso la entropía y el calor específico estarían dominados por los valores de cada sub-población.

En resumen, a lo largo de esta Tesis hemos estudiado el efecto poblacional de las correlaciones de alto orden utilizando dos metodologías diferentes pero complementarias. El marco teórico provisto por la teoría (y la geometría) de la información constituye una herramienta central para este análisis. Durante el desarrollo del trabajo, se propusieron extensiones válidas de métodos previamente elaborados sin considerar la presencia de correlaciones de alto orden a distintas escalas. Además, consolidamos ambas aproximaciones estudiando cómo varían las propiedades de una población a medida que se modifica su tamaño. Concluimos que el uso de estos enfoques conforman un avance en el estudio de la interdependencia neuronal en la corteza

y resultan relevantes, junto con otras aproximaciones existentes, para investigaciones neurofisiológicas futuras en la capacidad de codificación de poblaciones de neuronas.

Apéndice A

Cálculo de la información de Fisher para la distribución conjunta de disparos

A continuación, calculamos la expresión analítica de la información de Fisher para la distribución conjunta de disparos extendida como

$$\begin{aligned} I(Q_q) &= \int_0^1 \frac{\left[\frac{\partial Q_q(r)}{\partial r} \right]^2}{Q_q(r)} dr \\ &= \int_0^1 Q_q(r) \left\{ F_q^{-1}(r) + \frac{\sqrt{\alpha(q+1)}}{2(1-\alpha)} \frac{(h - \sqrt{\alpha} F_q^{-1}(r))}{\left[1 + \frac{(q-1)}{2(1-\alpha)} (h - \sqrt{\alpha} F_q^{-1}(r))^2 \right]} \right\}^2 \\ &\quad \cdot \left[\frac{\partial F_q^{-1}(r)}{\partial r} \right]^2 dr. \end{aligned} \tag{A.1}$$

Considerando que la derivada de la función Beta $B_{inverse}(z; a, b)$ es

$$\frac{\partial B_{inverse}(z; a, b)}{\partial z} = (1 - B_{inverse}(z; a, b))^{1-b} B_{inverse}(z; a, b)^{1-a} B(a, b), \tag{A.2}$$

obtenemos, luego de algunos cálculos,

$$\begin{aligned}
 I(Q_q) = & \int_0^1 \sqrt{\frac{1-\alpha}{\alpha}} \left(\frac{1-\alpha}{2\alpha(q-1)} \right) B \left(\frac{1}{q-1}, \frac{1}{2} \right) \left\{ \frac{\sqrt{1-\beta} [\beta]^{\left(\frac{q-2}{q-1}\right)}}{\beta^3 - \beta^4} \right\} \\
 & \cdot \left\{ \sqrt{\frac{(q+1)^2}{2(q-1)} \frac{\alpha}{(1-\alpha)} \beta(1-\beta)} + \frac{1}{\sqrt{\alpha}} \left\{ h - \sqrt{\frac{2(1-\alpha)}{q-1} [\beta^{-1} - 1]} \right\} \right\}^2 \\
 & \cdot \left\{ \exp_q [\beta^{-1} - 1] \right\}^{-\frac{(q+1)}{2}} \exp \left[-\frac{1}{2\alpha} \left\{ h - \sqrt{\frac{2(1-\alpha)}{q-1} [\beta^{-1} - 1]} \right\} \right] dr
 \end{aligned} \tag{A.3}$$

donde $\beta = B_{inverse} \left(r; \frac{1}{q-1}, \frac{1}{2} \right)$.

Apéndice B

Estimaciones numéricas de la función de distribución cumulativa q -gaussiana bivariada

La función de distribución cumulativa q -gaussiana bivariada se encuentra definida por

$$\begin{aligned} \Phi_q([a, b], \rho) &= \frac{\Gamma\left(\frac{1}{q-1}\right) (q-1)}{\pi \Gamma\left(\frac{1}{q-1} - 1\right) (6-4q) \sqrt{1-\rho^2}} \\ &\quad \times \int_{-\infty}^a \int_{-\infty}^b dy dx \left[1 + \left(\frac{q-1}{6-4q}\right) \frac{x^2 + y^2 - 2\rho xy}{1-\rho^2} \right]^{\frac{1}{1-q}}, \end{aligned} \tag{B.1}$$

donde $a, b \in \mathbb{R}$ y $q \in [1, 3/2)$. En comparación a la distribución q -gaussiana univariada, su contraparte bivariada no puede ser fácilmente mapeada a una distribución t de Student. Además, esta integral no puede ser resuelta analíticamente. Entonces, como esta función no se encuentra incluida en paquetes numéricos usuales, necesitamos simplificar su expresión para que sea posible obtener valores numéricos precisos.

En primera instancia, utilizamos la llamada “Transformada de Hilhorst” para

reescribir la ecuación (B.1) como

$$\begin{aligned}
 \Phi_q([a, b], \rho) &= \frac{\Gamma\left(\frac{1}{q-1}\right)(q-1)}{\pi\Gamma\left(\frac{1}{q-1}-1\right)(6-4q)\sqrt{1-\rho^2}} \int_0^\infty ds s \left(\frac{s^2}{2}\right)^{\frac{1}{q-1}} e^{-\frac{s^2}{2}} \\
 &\quad \times \int_{-\infty}^a \int_{-\infty}^b dy dx e^{-\frac{s^2}{2} \left(\frac{q-1}{6-4q}\right) \frac{x^2+y^2-2\rho xy}{1-\rho^2}} \\
 &= \frac{2^{2-\frac{1}{q-1}}(q-1)}{\Gamma\left(\frac{1}{q-1}-1\right)(6-4q)} \\
 &\quad \times \int_0^\infty ds s^{\frac{2}{q-1}-1} e^{-\frac{s^2}{2}} \Phi\left(s\sqrt{\frac{q-1}{6-4q}}[a, b], \rho\right).
 \end{aligned} \tag{B.2}$$

$\Phi(\mathbf{x}, \rho)$ corresponde a la función cumulativa gaussiana bivariada de \mathbf{x} con coeficiente de correlación ρ . Entonces, tomando la derivada con respecto a ρ obtenemos

$$\frac{d\Phi_q([a, b], \rho)}{d\rho} = \frac{2^{2-\frac{1}{q-1}}(q-1)}{\Gamma\left(\frac{1}{q-1}-1\right)(6-4q)} \int_0^\infty ds s^{\frac{1}{q-1}} e^{-\frac{s^2}{2}} \frac{e^{-\frac{s^2}{2} \left(\frac{q-1}{6-4q}\right) \frac{a^2+b^2-2\rho ab}{1-\rho^2}}}{2\pi\sqrt{1-\rho^2}}. \tag{B.3}$$

Utilizando el cambio de variables $r = s\sqrt{\left[1 + \left(\frac{q-1}{6-4q}\right)\frac{a^2+b^2-2\rho ab}{1-\rho^2}\right]}$,

$$\begin{aligned} \frac{d\Phi_q([a, b], \rho)}{d\rho} &= \frac{2^{2-\frac{1}{q-1}}(q-1)}{\Gamma\left(\frac{1}{q-1}-1\right)(6-4q)} \frac{\left[1 + \left(\frac{q-1}{6-4q}\right)\frac{a^2+b^2-2\rho ab}{1-\rho^2}\right]^{\frac{1}{1-q}}}{2\pi\sqrt{1-\rho^2}} \\ &\quad \times \int_0^\infty dr r^{\frac{2}{q-1}-1} e^{-\frac{r^2}{2}}. \\ &= \frac{\Gamma\left(\frac{1}{q-1}\right)(q-1)}{\pi\Gamma\left(\frac{1}{q-1}-1\right)(6-4q)\sqrt{1-\rho^2}} \\ &\quad \times \left[1 + \left(\frac{q-1}{6-4q}\right)\frac{a^2+b^2-2\rho ab}{1-\rho^2}\right]^{\frac{1}{1-q}}, \end{aligned} \tag{B.4}$$

debido a que $\int_0^\infty dr r^{\frac{2}{q-1}-1} e^{-\frac{r^2}{2}} = 2^{\frac{1}{q-1}-1}\Gamma\left(\frac{1}{q-1}\right)$. Esta fórmula puede ser integrada con respecto a ρ para producir

$$\begin{aligned} \Phi_q([a, b], \rho) &= \Phi_q([a, b], s) + \frac{\Gamma\left(\frac{1}{q-1}\right)(q-1)}{\pi\Gamma\left(\frac{1}{q-1}\right)(6-4q)} \\ &\quad \times \int_s^\rho dr \frac{\left[1 + \left(\frac{q-1}{6-4q}\right)\frac{a^2+b^2-2rab}{1-r^2}\right]^{\frac{1}{1-q}}}{\sqrt{1-r^2}}, \end{aligned} \tag{B.5}$$

donde $s = \text{sign}(\rho)$ y entonces,

$$\Phi_q([a, b], s) = \begin{cases} \Phi_q(\min\{a, b\}) & \text{if } s = 1 \\ \max\{0, \Phi_q(\min\{a, b\}) - \Phi_q(\min\{a, b\})\} & \text{if } s = -1 \end{cases} \tag{B.6}$$

con $\Phi_q(\cdot)$ definida como la función acumulativa de la distribución *q*-gaussiana univariada. Para implementar la ecuación (B.5) numéricamente en un algoritmo, el cambio de variables $r = \sin(\theta)$ es deseado. Entonces, la función acumulativa de la distribución

q-gaussiana bivariada es calculada como

$$\begin{aligned} \Phi_q([a, b], \rho) &= \Phi_q([a, b], s) + \frac{\Gamma\left(\frac{1}{q-1}\right) (q-1)}{\pi \Gamma\left(\frac{1}{q-1}\right) (6-4q)} \\ &\times \int_{s\pi/2}^{\sin^{-1}\rho} dr \left[1 + \left(\frac{q-1}{6-4q}\right) \frac{a^2 + b^2 - 2ab \sin \theta}{\cos^2 \theta} \right]^{\frac{1}{1-q}}. \end{aligned} \tag{B.7}$$

Apéndice C

Cuantificadores de complejidad del modelo de neurona simple de Izhikevich

Utilizamos la metodología de permutación de Bandt-Pompe [119] para la evaluación de la función de distribución de probabilidad asociada con la serie temporal del modelo de neurona. Basándonos en la cuantificación de las “estructuras” ordinales presentes en las oscilaciones del potencial de membrana de la neurona y su influencia local en la distribución de probabilidades asociada, incorporamos la propia causalidad de la serie temporal a través de un algoritmo de fácil implementación, detallado en [83]. Específicamente, se propone cuantificar las veinte patrones oscilatorios neuronales, mostrados en el capítulo 3 utilizando el modelo simple de neurona de Izhikevich, considerando medidas sutiles que toman en cuenta información causal: entropía, información de Fisher y complejidad MPR de permutación. Esta aproximación permite estimar las “propiedades de agrupación” de estas distintas estructuras de la dinámica neuronal a través de representaciones de dos y tres dimensiones.

La Figura C.1 (a) muestra la complejidad MPR versus la entropía, dada por el plano- $H \times C$, para los veinte comportamientos (A)-(T). La complejidad estadística MPR crece linealmente a medida que aumenta la entropía al considerar distintos tipos de dinámica. El grado de orden decrece al aumentar la entropía, y por ende un sistema de menor grado de entropía se caracteriza por un grado mayor de orden.

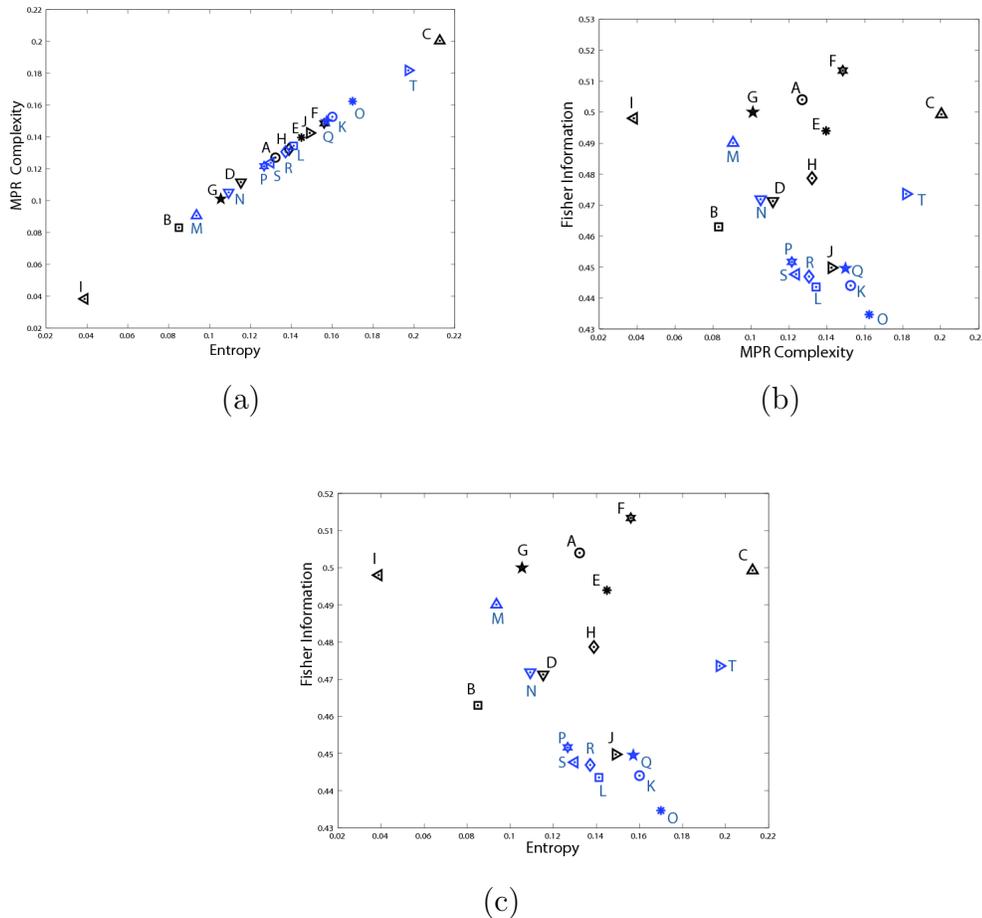


Figura C.1: Representaciones bidimensionales de los cuantificadores de información causal: (a) Complejidad MPR (‘MPR Complexity’) versus entropía (‘Entropy’) (plano- $H \times C$), (b) Información de Fisher (‘Fisher Information’) versus complejidad MPR (‘MPR Complexity’) (plano- $F \times C$), (c) Información de Fisher (‘Fisher Information’) versus entropía (‘Entropy’) (plano- $F \times H$), para los veinte comportamientos más relevantes de las neuronas biológicas (nomenclatura igual a la utilizada en el capítulo 3). Todos los cuantificadores fueron evaluados utilizando la función de distribución de probabilidad obtenida con la metodología Bandt-Pompe con un largo de patrón (dimensión de incrustación) $D = 6$ y retraso temporal $\tau = 1$.

La Figura C.1 (b) corresponde al plano de información causal $F \times C$, es decir a los valores de información de Fisher versus la complejidad MPR, para los mismos modelos neurocomputacionales. En comparación a la figura anterior, presenta una mejor distinción entre los veinte ritmos neuronales. Esto es, la información de Fisher

se comporta no linealmente como función de la complejidad MPR. Una dependencia similar puede ser observada en la Figura C.1 (c), que muestra la información de Fisher versus la entropía, el plano causal $F \times H$. Este resultado es razonable dado que observamos que la complejidad MPR crece linealmente a medida que crece la entropía.

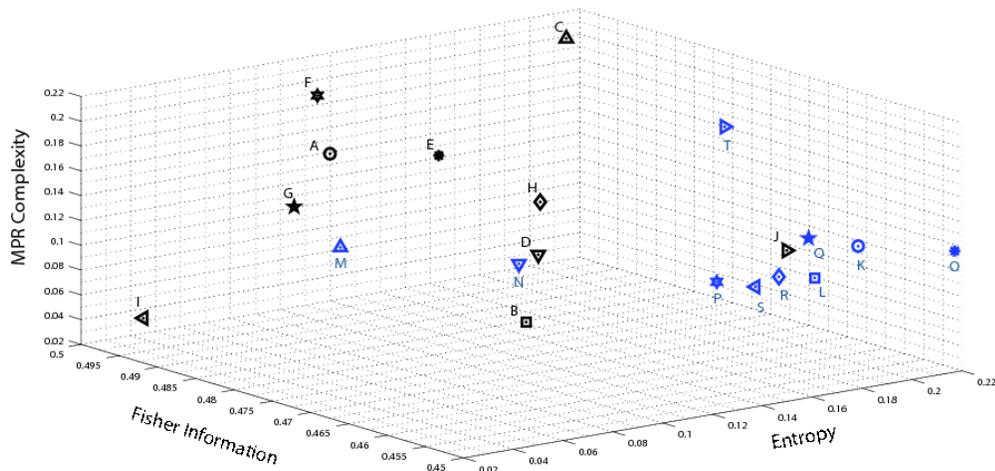


Figura C.2: Representación tridimensional de los cuantificadores de información causal: Complejidad MPR (‘MPR Complexity’) \times Información de Fisher (‘Fisher Information’) \times Entropía (‘Entropy’) para los veinte comportamientos más relevantes de las neuronas biológicas (nomenclatura igual a la utilizada en el capítulo 3). Todos los cuantificadores fueron evaluados utilizando la función de distribución de probabilidad obtenida con la metodología Bandt-Pompe con un largo de patrón (dimensión de incrustación) $D = 6$ y retraso temporal $\tau = 1$.

Por otro lado, la Figura C.2 muestra las propiedades dinámicas emergentes de la información causal del sistema a través de una representación tridimensional: la complejidad estadística MPR \times información de Fisher \times entropía. Notablemente, esta aproximación nos permite clasificar los comportamientos dinámicos por medio del uso de un enfoque en teoría de la información. Cuantificamos la causalidad de la señal e inferimos las propiedades emergentes de cada patrón oscilatorio de actividad dentro de una representación informacional al asignar un “grupo” a cada una de estas caacterísticas. Por lo tanto, podemos distinguir con precisión los diferentes casos (A)-(T) generados por el modelo de neurona simple de Izhikevich.

Apéndice D

Trabajos publicados

Durante el desarrollo de la presente Tesis Doctoral fueron publicados los resultados expuestos en los capítulos 3, 5 (junto con el apéndice A) y 6 (junto con el apéndice B) en la referencias [86], [75] y [122] respectivamente. Adicionalmente, el capítulo 3 y el apéndice C contienen análisis y resultados de la publicación [83]. Estas publicaciones se especifican a continuación:

- Montangie L., Montani F. (2016), Higher-order correlations in common input shapes the output spiking activity of a neural population, *Physica A: Statistical Mechanics and its Applications*. doi:10.1016/j.physa.2016.12.002
- Montangie L., Montani F. (2016), Effect of interacting second- and third-order stimulus dependent correlations produce population coding asymmetries, *Physical Review E*. doi: 10.1103/PhysRevE.94.042303
- Montani F., Baravalle R., Montangie L., Rosso O. A.(2015), Causal information quantification of prominent dynamical features of biological neurons, *Philosophical Transaction of the Royal Society A* . doi: 10.1098/rsta.2015.0109
- Montangie L., Montani F.(2015), Quantifying higher-order correlations in a neuronal pool, *Physica A: Statistical Mechanics and its Applications* 421:388-400 . doi:10.1016/j.physa.2014.11.046

Además, se encuentran en preparación otros trabajos relativos a los estudios detallados en los capítulos 4, 6 (cuantificadores de complejidad) y 7.

Bibliografía

- [1] E. D. Adrian, “The impulses produced by sensory nerve endings: Part I,” *The Journal of Physiology*, vol. 61, pp. 49–72, mar 1926.
- [2] E. D. Adrian and Y. Zotterman, “The impulses produced by sensory nerve-endings: Part II. The response of a Single End-Organ,” *The Journal of Physiology*, vol. 61, pp. 151–171, apr 1926.
- [3] E. D. Adrian and Y. Zotterman, “The impulses produced by sensory nerve-endings: Part III. Impulses set up by Touch and Pressure,” *The Journal of Physiology*, vol. 61, pp. 465–483, aug 1926.
- [4] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 2001.
- [5] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code (Computational Neuroscience)*. A Bradford Book, 1999.
- [6] C. E. Shannon, *The mathematical theory of communication*. Urbana, IL: The University of Illinois Press, 1949.
- [7] T. Cover and J. Thomas, *Elements of Information Theory*, vol. 1. Wiley, 1991.
- [8] G. Werner and V. B. Mountcastle, “Improving data quality in neuronal population recordings,” *Journal of Neurophysiology*, vol. 28, pp. 359–397, mar 1965.

-
- [9] S. R. Schultz, S. Panzeri, A. Treves, and E. T. Rolls, “Correlated firing and the information represented by neurons in short epochs,” *Neurocomputing*, vol. 26-27, pp. 499–504, jun 1999.
- [10] S. Panzeri, R. S. Petersen, S. R. Schultz, M. Lebedev, and M. E. Diamond, “The role of spike timing in the coding of stimulus location in rat somatosensory cortex,” *Neuron*, vol. 29, pp. 769–777, mar 2001.
- [11] S. Panzeri and S. R. Schultz, “A unified approach to the study of temporal, correlational, and rate coding,” *Neural Computation*, vol. 13, pp. 1311–1349, jun 2001.
- [12] G. Pola, A. Thiele, K.-P. Hoffmann, and S. Panzeri, “An exact method to quantify the information transmitted by different mechanisms of correlational coding,” *Network: Computation in Neural Systems*, vol. 14, pp. 35–60, jan 2003.
- [13] F. Montani, A. Kohn, M. A. Smith, and S. R. Schultz, “The role of correlations in direction and contrast coding in the primary visual cortex,” *Journal of Neuroscience*, vol. 27, pp. 2338–2348, feb 2007.
- [14] Y. Roudi, S. Nirenberg, and P. E. Latham, “Pairwise maximum entropy models for studying large biological systems: When they can work and when they can’t,” *PLoS Computational Biology*, vol. 5, p. e1000380, may 2009.
- [15] S. H. Nirenberg and J. D. Victor, “Analyzing the activity of large populations of neurons: how tractable is the problem?,” *Current Opinion in Neurobiology*, vol. 17, pp. 397–400, aug 2007.
- [16] K. D. Harris, R. Q. Quiroga, J. Freeman, and S. L. Smith, “Improving data quality in neuronal population recordings,” *Nature Neuroscience*, vol. 19, pp. 1165–1174, aug 2016.
- [17] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, “Weak pairwise correlations imply strongly correlated network states in a neural population,” *Nature*, vol. 440, pp. 1007–1012, apr 2006.

-
- [18] J. Shlens, G. D. Field, J. L. Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky, “The structure of multi-neuron firing patterns in primate retina,” *Journal of Neuroscience*, vol. 26, pp. 8254–8266, aug 2006.
- [19] E. T. Jaynes, “Information theory and statistical mechanics,” *Physical Review*, vol. 106, pp. 620–630, may 1957.
- [20] A. Katz, *Principles of statistical mechanics: the information theory approach*. W. H. Freeman, 1967.
- [21] G. Parisi, *Statistical Field Theory (Advanced Book Classics)*. Westview Press, 1998.
- [22] G. Tkačik, O. Marre, T. Mora, D. Amodei, M. J. Berry II, and W. Bialek, “The simplest maximum entropy model for collective behavior in a neural network,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2013, p. P03011, mar 2013.
- [23] G. Bi and M. Poo, “Synaptic modification by correlated activity: Hebb’s postulate revisited,” *Annual Review of Neuroscience*, vol. 24, pp. 139–166, mar 2001.
- [24] F. Montani, R. A. A. Ince, R. Senatore, E. Arabzadeh, M. E. Diamond, and S. Panzeri, “The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, pp. 3297–3310, jul 2009.
- [25] I. E. Ohiorhenuan, F. Mechler, K. P. Purpura, A. M. Schmid, Q. Hu, and J. D. Victor, “Sparse coding and high-order correlations in fine-scale cortical networks,” *Nature*, vol. 466, pp. 617–621, jul 2010.
- [26] S. Amari and H. Nagaoka, *Methods of Information Geometry (Translations of Mathematical Monographs)*. Amer Mathematical Society, 2001.

-
- [27] A. K. Barreiro, J. Gjorgjieva, F. Rieke, and E. Shea-Brown, “When do microcircuits produce beyond-pairwise correlations?,” *Frontiers in Computational Neuroscience*, vol. 8, 2014.
- [28] N. A. Cayco-Gajic, J. Zylberberg, and E. Shea-Brown, “Triplet correlations among similarly tuned cells impact population coding,” *Frontiers in Computational Neuroscience*, vol. 9, may 2015.
- [29] J. Zylberberg and E. Shea-Brown, “Input nonlinearities can shape beyond-pairwise correlations and improve information transmission by neural populations,” *Physical Review E*, vol. 92, dec 2015.
- [30] D. R. Cox and N. Wermuth, “On some models for multivariate binary variables parallel in complexity with the multivariate gaussian distribution.,” *Biometrika*, vol. 89, no. 2, pp. 462–9, 2002.
- [31] S. Amari, H. Nakahara, S. Wu, and Y. Sakai, “Synchronous firing and higher-order interactions in neuron pool,” *Neural Computation*, vol. 15, pp. 127–142, jan 2003.
- [32] J. H. Macke, P. Berens, A. S. Ecker, A. S. Tolias, and M. Bethge, “Generating spike trains with specified correlation coefficients,” *Neural Computation*, vol. 21, pp. 397–423, feb 2009.
- [33] J. H. Macke, M. Opper, and M. Bethge, “Common input explains higher-order correlations and entropy in a simple model of neural population activity,” *Physical Review Letters*, vol. 106, may 2011.
- [34] S. Yu, H. Yang, H. Nakahara, G. S. Santos, D. Nikolic, and D. Plenz, “Higher-order interactions characterized in cortical activity,” *Journal of Neuroscience*, vol. 31, pp. 17514–17526, nov 2011.
- [35] S. Kullback, *Information Theory and Statistics*. Dover Publications, 1959.
- [36] T. Gawne and B. Richmond, “How independent are the messages carried by adjacent inferior temporal cortical neurons?,” *Journal of Neuroscience*, vol. 13, pp. 2758–2771, jul 1993.

-
- [37] E. Schneidman, S. Still, M. J. Berry, and W. Bialek, “Network information and connected correlations,” *Physical Review Letters*, vol. 91, p. 238701, dec 2003.
- [38] M. Crumiller, B. Knight, Y. Yu, and E. Kaplan, “Estimating the amount of information conveyed by a population of neurons,” *Frontiers in Neuroscience*, vol. 5, 2011.
- [39] F. Montani, A. Oliynyk, and L. Fadiga, “Superlinear summation of information in premotor neuron pairs,” *International Journal of Neural Systems*, vol. 27, p. 1650009, mar 2017.
- [40] L. F. Abbott and P. Dayan, “The effect of correlated variability on the accuracy of a population code,” *Neural Computation*, vol. 11, pp. 91–101, jan 1999.
- [41] M. Oram, N. Hatsopoulos, B. Richmond, and J. Donoghue, “Excess synchrony in motor cortical neurons provides redundant direction information with that from coarse temporal measures,” *Journal of Neurophysiology*, vol. 86, pp. 1700–1716, oct 2001.
- [42] H. Nakahara and S. Amari, “Information-geometric measure for neural spikes,” *Neural Computation*, vol. 14, pp. 2269–2316, oct 2002.
- [43] C. Tsallis, “Possible generalization of Boltzmann-Gibbs statistics,” *Journal of Statistical Physics*, vol. 52, pp. 479–487, jul 1988.
- [44] S. Amari and A. Ohara, “Geometry of q-exponential family of probability distributions,” *Entropy*, vol. 13, pp. 1170–1185, jun 2011.
- [45] D. J. Tolhurst, “The amount of information transmitted about contrast by neurones in the cat’s visual cortex,” *Vis. Neurosci.*, vol. 2, no. 4, pp. 409–413, 1989.
- [46] A. K. Kreiter and W. Singer, “Oscillatory neuronal responses in the visual cortex of the awake macaque monkey,” *European Journal of Neuroscience*, vol. 4, pp. 369–375, apr 1992.

-
- [47] A. K. Kreiter and W. Singer, “Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey,” *J. Neurosci.*, vol. 16, pp. 2381–2396, Apr 1996.
- [48] R. C. deCharms and M. M. Merzenich, “Primary cortical representation of sounds by the coordination of action-potential timing,” *Nature*, vol. 381, pp. 610–613, jun 1996.
- [49] T. J. Gawne, T. W. Kjaer, J. A. Hertz, and B. J. Richmond, “Adjacent visual cortical complex cells share about 20 information,” *Cereb. Cortex*, vol. 6, no. 3, pp. 482–489, 1996.
- [50] P. R. Roelfsema, A. K. Engel, P. König, and W. Singer, “Visuomotor integration is associated with zero time-lag synchronization among cortical areas,” *Nature*, vol. 385, pp. 157–161, jan 1997.
- [51] A. Kohn, “Stimulus dependence of neuronal correlation in primary visual cortex of the macaque,” *Journal of Neuroscience*, vol. 25, pp. 3661–3673, apr 2005.
- [52] M. I. Chelaru and V. Dragoi, “Negative correlations in visual cortical networks,” *Cerebral Cortex*, vol. 26, pp. 246–256, sep 2014.
- [53] E. Ganmor, R. Segev, and E. Schneidman, “Sparse low-order interaction network underlies a highly correlated and learnable neural population code,” *Proceedings of the National Academy of Sciences*, vol. 108, pp. 9679–9684, may 2011.
- [54] I. E. Ohiorhenuan and J. D. Victor, “Information-geometric measure of 3-neuron firing patterns characterizes scale-dependence in cortical networks,” *Journal of Computational Neuroscience*, vol. 30, pp. 125–141, jul 2010.
- [55] S. Nirenberg, S. M. Carcieri, A. L. Jacobs, and P. E. Latham *Nature*, vol. 411, pp. 698–701, jun 2001.

-
- [56] B. B. Averbeck, M. V. Chafee, D. A. Crowe, and A. P. Georgopoulos, “Neural activity in prefrontal cortex during copying geometrical shapes,” *Experimental Brain Research*, vol. 150, pp. 127–141, apr 2003.
- [57] B. B. Averbeck, “Effects of noise correlations on information encoding and decoding,” *Journal of Neurophysiology*, vol. 95, pp. 3633–3644, feb 2006.
- [58] B. B. Averbeck, P. E. Latham, and A. Pouget, “Neural correlations, population coding and computation,” *Nature Reviews Neuroscience*, vol. 7, pp. 358–366, may 2006.
- [59] S. Nirenberg and P. E. Latham, “Decoding neuronal spike trains: How important are correlations?,” *Proceedings of the National Academy of Sciences*, vol. 100, pp. 7348–7353, may 2003.
- [60] P. E. Latham, “Synergy, redundancy, and independence in population codes, revisited,” *Journal of Neuroscience*, vol. 25, pp. 5195–5206, may 2005.
- [61] J. L. Puchalla, E. Schneidman, R. A. Harris, and M. J. Berry, “Redundancy in the population code of the retina,” *Neuron*, vol. 46, pp. 493–504, may 2005.
- [62] E. Schneidman, J. L. Puchalla, R. Segev, R. A. Harris, W. Bialek, and M. J. Berry, “Synergy from silence in a combinatorial neural code,” *Journal of Neuroscience*, vol. 31, pp. 15732–15741, nov 2011.
- [63] N. S. Narayanan, “Redundancy and synergy of neuronal ensembles in motor cortex,” *Journal of Neuroscience*, vol. 25, pp. 4207–4216, apr 2005.
- [64] M. DeWeese, “Optimization principles for the neural code,” *Network: Computation in Neural Systems*, vol. 7, pp. 325–331, may 1996.
- [65] S. Grün and S. Rotter, eds., *Analysis of Parallel Spike Trains*. Springer US, 2010.
- [66] L. Martignon, G. Deco, K. Laskey, M. Diamond, W. Freiwald, and E. Vaadia, “Neural coding: Higher-order temporal patterns in the neurostatistics of cell assemblies,” *Neural Computation*, vol. 12, pp. 2621–2653, nov 2000.

- [67] M. Bezzi, M. E. Diamond, and A. Treves, “Redundancy and synergy arising from pairwise correlations in neuronal ensembles,” *J Comput Neurosci*, vol. 12, no. 3, pp. 165–174, 2002.
- [68] E. Salinas and T. J. Sejnowski, “Correlated neuronal activity and the flow of neural information,” *Nature Reviews Neuroscience*, vol. 2, pp. 539–550, aug 2001.
- [69] R. Lestienne, “Spike timing, synchronization and information processing on the sensory side of the central nervous system,” *Prog. Neurobiol.*, vol. 65, pp. 545–591, Dec 2001.
- [70] T. Womelsdorf and P. Fries, “The role of neuronal synchronization in selective attention,” *Current Opinion in Neurobiology*, vol. 17, pp. 154–160, apr 2007.
- [71] A. Kohn, A. Zandvakili, and M. A. Smith, “Correlations and brain states: from electrophysiology to functional imaging,” *Current Opinion in Neurobiology*, vol. 19, pp. 434–438, aug 2009.
- [72] M. Bethge and P. Berens, “Near-maximum entropy models for binary neural representations of natural images,” *Advances in Neural Information Processing Systems* 20, 2008.
- [73] S. M. Bohte, H. Spekreijse, and P. R. Roelfsema, “The effects of pair-wise and higher order correlations on the firing rate of a post-synaptic neuron,” *Neural Comput*, vol. 12, pp. 153–179, Jan 2000.
- [74] F. Montani, E. Phoka, M. Portesi, and S. R. Schultz, “Statistical modelling of higher-order correlations in pools of neural activity,” *Physica A: Statistical Mechanics and its Applications*, vol. 392, pp. 3066–3086, jul 2013.
- [75] L. Montangie and F. Montani, “Quantifying higher-order correlations in a neuronal pool,” *Physica A: Statistical Mechanics and its Applications*, vol. 421, pp. 388–400, mar 2015.

- [76] M. Y. Yim, A. Kumar, A. Aertsen, and S. Rotter, “Impact of correlated inputs to neurons: modeling observations from in vivo intracellular recordings,” *Journal of Computational Neuroscience*, vol. 37, pp. 293–304, may 2014.
- [77] S. C. Wissig and A. Kohn, “The influence of surround suppression on adaptation effects in primary visual cortex,” *Journal of Neurophysiology*, vol. 107, pp. 3370–3384, mar 2012.
- [78] J. de la Rocha, B. Doiron, E. Shea-Brown, K. Josić, and A. Reyes, “Correlation between neural spike trains increases with firing rate,” *Nature*, vol. 448, pp. 802–806, aug 2007.
- [79] E. M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting (Computational Neuroscience)*. The MIT Press, 2006.
- [80] E. Izhikevich, “Simple model of spiking neurons,” *IEEE Transactions on Neural Networks*, vol. 14, pp. 1569–1572, nov 2003.
- [81] E. Izhikevich, “Which model to use for cortical spiking neurons?,” *IEEE Transactions on Neural Networks*, vol. 15, pp. 1063–1070, sep 2004.
- [82] E. M. Izhikevich, “Polychronization: Computation with spikes,” *Neural Computation*, vol. 18, pp. 245–282, feb 2006.
- [83] F. Montani, R. Baravalle, L. Montangie, and O. A. Rosso, “Causal information quantification of prominent dynamical features of biological neurons,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 373, p. 20150109, nov 2015.
- [84] M. Abeles, H. Bergman, E. Margalit, and E. Vaadia, “Spatiotemporal firing patterns in the frontal cortex of behaving monkeys,” *J. Neurophysiol.*, vol. 70, pp. 1629–1638, Oct 1993.
- [85] S. Grün, M. Diesmann, and A. Aertsen, “Unitary events in multiple single-neuron spiking activity: I. detection and significance,” *Neural Computation*, vol. 14, pp. 43–80, jan 2002.

-
- [86] L. Montangie and F. Montani, “Effect of interacting second- and third-order stimulus-dependent correlations on population-coding asymmetries,” *Physical Review E*, vol. 94, p. 042303, oct 2016.
- [87] M. Tatsuno, J. Fellous, and S. Amari, “Information-geometric measures as robust estimators of connection strengths and external inputs,” *Neural Computation*, vol. 21, pp. 2309–2335, aug 2009.
- [88] Y. Nie and M. Tatsuno, “Information-geometric measures for estimation of connection weight under correlated inputs,” *Neural Computation*, vol. 24, pp. 3213–3245, dec 2012.
- [89] Y. Nie, J.-M. Fellous, and M. Tatsuno, “Information-geometric measures estimate neural interactions during oscillatory brain states,” *Frontiers in Neural Circuits*, vol. 8, 2014.
- [90] M. Tatsuno and M. Okada, “Investigation of possible neural architectures underlying information-geometric measures,” *Neural Computation*, vol. 16, pp. 737–765, apr 2004.
- [91] S. Herculano-Houzel, “The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost,” *Proceedings of the National Academy of Sciences*, vol. 109, pp. 10661–10668, jun 2012.
- [92] R. Desimone, “Face-selective cells in the temporal cortex of monkeys,” *Journal of Cognitive Neuroscience*, vol. 3, pp. 1–8, jan 1991.
- [93] R. A. Andersen, “Encoding of intention and spatial location in the posterior parietal cortex,” *Cereb. Cortex*, vol. 5, no. 5, pp. 457–469, 1995.
- [94] S. R. Cobb, E. H. Buhl, K. Halasy, O. Paulsen, and P. Somogyi, “Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons,” *Nature*, vol. 378, pp. 75–78, nov 1995.
- [95] C. Tsallis, M. Gell-Mann, and Y. Sato, “Asymptotically scale-invariant occupancy of phase space makes the entropy S_q extensive,” *Proceedings of the National Academy of Sciences*, vol. 102, pp. 15377–15382, oct 2005.

-
- [96] J. M. López and J. M. Sánchez, eds., *New trends in magnetism, magnetic materials, and their applications*. New York: Springer Science + Business Media, 1994.
- [97] S. Umarov, C. Tsallis, and S. Steinberg, “On a q-central limit theorem consistent with nonextensive statistical mechanics,” *Milan Journal of Mathematics*, vol. 76, pp. 307–328, mar 2008.
- [98] S. Umarov, C. Tsallis, M. Gell-Mann, and S. Steinberg, “Generalization of symmetric α -stable levy distributions for $q > 1$,” *Journal of Mathematical Physics*, vol. 51, no. 3, p. 033502, 2010.
- [99] S. Umarov and C. Tsallis, “The limit distribution in the q-CLT for $q \geq 1$ is unique and can not have a compact support,” *Journal of Physics A: Mathematical and Theoretical*, vol. 49, no. 41, p. 415204, 2016.
- [100] C. Vignat and A. Plastino, “Central limit theorem and deformed exponentials,” *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 45, p. F969, 2007.
- [101] S. Ross, *First Course in Probability*, A. Pearson, 2015.
- [102] R. W. Butler, *Saddlepoint Approximations with Applications*. Cambridge University Press, 2007.
- [103] L. Borland, A. R. Plastino, and C. Tsallis, “Information gain within nonextensive thermostatics,” *Journal of Mathematical Physics*, vol. 39, pp. 6490–6501, dec 1998.
- [104] J. Prehl, C. Essex, and K. H. Hoffmann, “Tsallis relative entropy and anomalous diffusion,” *Entropy*, vol. 14, pp. 701–716, apr 2012.
- [105] J. V. Linnik, “An information-theoretic proof of the central limit theorem with lindeberg conditions,” *Theory of Probability & Its Applications*, vol. 4, pp. 288–299, jan 1959.

-
- [106] R. A. Fisher, “On the mathematical foundations of theoretical statistics,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 222, pp. 309–368, jan 1922.
- [107] B. R. Frieden, *Science from Fisher Information: A Unification*. Cambridge University Press, 2004.
- [108] A. L. Mayer, C. W. Pawlowski, and H. Cabezaz, “Fisher information and dynamic regime changes in ecological systems,” *Ecological Modelling*, vol. 195, pp. 72–82, may 2006.
- [109] O. A. Rosso, L. D. Micco, A. Plastino, and H. A. Larrondo, “Infoquantifiers’ map-characterization revisited,” *Physica A: Statistical Mechanics and its Applications*, vol. 389, pp. 4604–4612, nov 2010.
- [110] F. Olivares, A. Plastino, and O. A. Rosso, “Ambiguities in Bandt-Pompe’s methodology for local entropic quantifiers,” *Physica A: Statistical Mechanics and its Applications*, vol. 391, pp. 2518–2526, apr 2012.
- [111] F. Olivares, A. Plastino, and O. A. Rosso, “Contrasting chaos with noise via local versus global information quantifiers,” *Physics Letters A*, vol. 376, pp. 1577–1583, apr 2012.
- [112] F. Montani, E. B. Deleglise, and O. A. Rosso, “Efficiency characterization of a large neuronal network: A causal information approach,” *Physica A: Statistical Mechanics and its Applications*, vol. 401, pp. 58–70, may 2014.
- [113] H. W. Block and Z. Fang, “A multivariate extension of hoeffding’s lemma,” *The Annals of Probability*, vol. 16, pp. 1803–1820, oct 1988.
- [114] H. Shimazaki, K. Sadeghi, T. Ishikawa, Y. Ikegaya, and T. Toyozumi, “Simultaneous silence organizes structured higher-order interactions in neural populations,” *Scientific Reports*, vol. 5, apr 2015.
- [115] P. Lamberti, M. Martin, A. Plastino, and O. Rosso, “Intensive entropic non-triviality measure,” *Physica A: Statistical Mechanics and its Applications*, vol. 334, pp. 119–131, mar 2004.

-
- [116] D. P. Feldman, C. S. McTague, and J. P. Crutchfield, “The organization of intrinsic computation: Complexity-entropy diagrams and the diversity of natural information processing,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 18, p. 043106, dec 2008.
- [117] R. López-Ruiz, H. Mancini, and X. Calbet, “A statistical measure of complexity,” *Physics Letters A*, vol. 209, pp. 321–326, dec 1995.
- [118] M. T. Martin, A. Plastino, and O. A. Rosso, “Generalized statistical complexity measures: Geometrical and analytical properties,” *Physica A: Statistical Mechanics and its Applications*, vol. 369, pp. 439–462, sep 2006.
- [119] C. Bandt and B. Pompe, “Permutation entropy: A natural complexity measure for time series,” *Physical Review Letters*, vol. 88, apr 2002.
- [120] M. R. DeWeese and A. M. Zador, “Non-gaussian membrane potential dynamics imply sparse, synchronous activity in auditory cortex,” *Journal of Neuroscience*, vol. 26, pp. 12206–12218, nov 2006.
- [121] T. Hromádka, A. M. Zador, and M. R. DeWeese, “Up states are rare in awake auditory cortex,” *Journal of Neurophysiology*, vol. 109, pp. 1989–1995, jan 2013.
- [122] L. Montangie and F. Montani, “Higher-order correlations in common input shapes the output spiking activity of a neural population,” *Physica A: Statistical Mechanics and its Applications*, vol. 471, pp. 845–861, apr 2017.
- [123] D. J. Field, “What is the goal of sensory coding?,” *Neural Computation*, vol. 6, pp. 559–601, jul 1994.
- [124] B. Willmore and D. J. Tolhurst, “Characterizing the sparseness of neural codes,” *Network*, vol. 12, pp. 255–270, Aug 2001.
- [125] B. Olshausen and D. Field, “Sparse coding of sensory inputs,” *Current Opinion in Neurobiology*, vol. 14, pp. 481–487, aug 2004.

-
- [126] D. G. Hernández, D. H. Zanette, and I. Samengo, “Information-theoretical analysis of the statistical dependencies among three variables: Applications to written language,” *Physical Review E*, vol. 92, aug 2015.
- [127] G. Tkačik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, M. J. Berry, and W. Bialek, “Thermodynamics and signatures of criticality in a network of neurons,” *Proceedings of the National Academy of Sciences*, vol. 112, pp. 11508–11513, sep 2015.
- [128] M. Nonnenmacher, C. Behrens, P. Berens, M. Bethge, and J. H. Macke, “Signatures of criticality arise in simple neural population models with correlations,” *ArXiv e-prints*, Feb. 2016.
- [129] Y. Yang, M. R. DeWeese, G. H. Otazu, and A. M. Zador, “Millisecond-scale differences in neural activity in auditory cortex can drive decisions,” *Nature Neuroscience*, vol. 11, pp. 1262–1263, oct 2008.