



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE COMPUTACIÓN

Modelo numérico de resolución a gran escala para sistemas lineales mixtos mediante multiplicadores alternados: Aplicaciones al análisis anatomofuncional del cerebro humano

Tesis presentada para optar al título de
Licenciado en Ciencias de la Computación

Gastón Ezequiel Zanitti

Director: Demian Wassermann

Codirector: Guillermo Gallardo Diez

Buenos Aires, 2019

RESUMEN

La neurociencia moderna sostiene que existe una relación entre la funcionalidad cerebral y la conectividad extrínseca. Trabajos anteriores en el área han demostrado que dado un conjunto de sujetos y usando sólo conectividad estructural, medida a través de imágenes de difusión, somos capaces de predecir la activación funcional de diferentes tipos de respuestas en el cerebro. Sin embargo, ninguno de ellos ha considerado la diferencia de conectividad intra-sujeto, probablemente por la complejidad de los modelos que toman en cuenta esta información.

Para abordar este problema proponemos modelar la relación entre funcionalidad y conectividad utilizando un modelo lineal mixto. Este enfoque nos da la ventaja de poder modelar la conectividad intra-sujeto separando los efectos aleatorios propios de cada sujeto de los efectos fijos de la población.

Sin embargo, el enfoque de máxima verosimilitud utilizado en la mayoría de paquetes de software tiene la restricción de un alto consumo de memoria. En el aporte principal de este trabajo, proponemos una nueva forma de resolver estos modelos basados en el método de los multiplicadores alternados (o ADMM, por sus siglas en inglés). Este método nos proporciona la ventaja práctica de resolver los problemas de cálculo del enfoque de máxima verosimilitud mientras que nos permite una optimización de escala arbitraria y descentralizada, resolviendo el problema de ajustar un modelo a grandes conjuntos de datos. Finalmente, ilustramos esta mejora comparando nuestros resultados con el trabajo previo realizado por Osher et al.

Palabras claves: Neurociencias, ADMM, Método de multiplicadores alternados, Modelo Lineal Mixto, Lasso.

ABSTRACT

Modern neuroscience holds that there is a relationship between brain functionality and extrinsic connectivity. Previous works in the area have shown that using only structural connectivity, as measured through diffusion weighted imaging, we are able to predict the functional activation of different types of responses in the brain, given a set of subjects. However, neither of them has considered the difference intra-subject connectivity, probably due to the complexity in the models that take this information into account. To address this problem we model the relationship between functionality and connectivity, using a linear mixed model. This approach give us the advantage of being able to separate random effects, modeling intra-subjects connectivity, from the fixed effects of the population. However, the maximum likelihood approach used in the majority of the software package has the constraint of a large memory consumption. In the principal contribution of this work, we propose a new way to solve this model based on the alternating direction method of multipliers (ADMM) with the practical advantage of solving the computation problems of the maximum likelihood approach providing us with arbitrary-scale and decentralized optimization, solving the problem of fitting a model against huge datasets. Finally, we illustrate this improvement by comparing our results against Osher et al. previous work.

Keywords: Neuroscience, ADMM, Alternating direction method of multipliers, Linear mixed model, Lasso.

AGRADECIMIENTOS

A mis padres, Luis y Claudia, porque todo lo que soy y pueda ser se lo debo a ellos, mis dos gigantes.

A Ugo, que estaría orgulloso de saber lo ‘alto’ que llegó su nieto.

A Alicia, por siempre haber fomentado mi curiosidad y mi aprendizaje con cuanto libro/revista/enciclopedia me interesaba.

A Matias, Pablo, Luciana y Ariadna, por seguir todo este proceso desde el primer día, y siempre estar ahí, aunque sea para un chiste.

A Carol, por todo el amor y el apoyo incondicional. Por soportar mis tristezas y mis desilusiones. Por estar ahí para festejar las conquistas y las victorias. Por querer ser parte de toda esta aventura.

A Patricio, porque además de todo el conocimiento y las puertas que me abrió la universidad, también me dejó un gran amigo.

A Mati L. y Leo, por todas las apuestas y las cervezas pos final, los chistes y todo el empuje mutuo que nos dimos desde los primeros días y durante todos estos años para seguir avanzando.

A Demian por la oportunidad, la paciencia y los consejos.

A Guillermo, por hacerme sentir como en casa, incluso a miles de kilómetros de distancia y por todo el apoyo y los consejos.

A todos mis profesores y ayudantes, que con su pasión y conocimiento me inspiraron siempre a seguir avanzando.

A todos aquellos con los que compartí largas horas de estudio, TP interminables y eternas vueltas de mate.

A Luis y Claudia

Índice general

1..	Introducción	1
2..	Neurociencias: Introducción	3
2.1.	Anatomía básica del cerebro	3
2.2.	Introducción a la Resonancia Magnética	5
2.2.1.	Resonancia Magnética Funcional	5
2.3.	Resonancia Magnética de Difusión	6
3..	Modelos Lineales	7
3.1.	Análisis de regresión	7
3.2.	Modelo lineal	7
3.2.1.	Modelo lineal con LASSO	8
3.3.	Modelo lineal mixto	9
3.4.	Método de los multiplicadores alternados (ADMM)	11
3.4.1.	Doble ascenso	11
3.4.2.	Descomposición Doble	12
3.4.3.	Langrangianos aumentados y el método de los multiplicadores	13
3.4.4.	Algoritmo	13
3.4.5.	Forma escalada	14
4..	Relacionando estructura y función en el cerebro	15
4.1.	Relación entre estructura y función en el cerebro	15
5..	Contribución metodológica	19
5.1.	Resolución mediante ADMM	22
5.2.	Independencia entre sujetos	26
5.3.	Regularización L1	28
6..	Experimentos y resultados	29
6.1.	Experimentos en datos sintéticos	29
6.1.1.	Datos	29
6.1.2.	Diseño experimental	30
6.2.	Experimentos sobre sujetos del Human Connectome Project	30
6.2.1.	Participantes	30
6.2.2.	Diseño experimental	31
6.2.3.	Prediciendo activaciones funcionales en toda la corteza cerebral	32
6.2.4.	Prediciendo activaciones funcionales en regiones específicas	32
6.2.5.	Resultados	32
7..	Discusión	43
7.1.	Características del modelo a tener en cuenta	43
7.2.	Análisis de los resultados en datos sintéticos	43
7.3.	Análisis de los resultados en sujetos del Human Connectome Project	44

7.4. Observaciones finales	45
8.. Conclusión	47
Apéndices	49
Apéndice A.. Calculo de derivadas	51
A.1. Caso general	51
A.1.1. Derivada respecto a θ_1	51
A.1.2. Derivada respecto a x	51
A.1.3. Derivada respecto a θ_2	53
A.2. Independencia entre sujetos	61
A.2.1. Derivada de β_i	61
A.2.2. Derivada de β_*	63

1. INTRODUCCIÓN

El manejo simultáneo de efectos a nivel poblacional y de sujeto es una necesidad común en la neurociencia moderna [p.e. 17]. No obstante, los modelos estadísticos utilizados comúnmente, como el modelo de efectos aleatorios [p.e. 20], no escalan bien con grandes volúmenes de datos. Esto presenta un problema en la actualidad donde no solo ha incrementado el tamaño de las poblaciones estudiadas sino también la cantidad de datos por sujeto [p.e. 22]. En este trabajo abordamos esta problemática mediante la resolución de un modelo de efectos aleatorios utilizando métodos proximales [23]. Nuestra implementación nos permite reducir drásticamente los requisitos de memoria, abrir las puertas a la computación en paralelo e incluir sin problema estrategias de regularización [p.e. 31]. Ilustramos todas estas ventajas en uno de los problemas más importantes de la neurociencia: el estudio de la relación entre estructura y función en el cerebro humano.

En la neurociencia moderna existe la teoría de que la funcionalidad de una región del cerebro está determinada por su conectividad estructural. Sur et al. [30] presentan una primera evidencia de esto mostrando que la conectividad estructural dirige el desarrollo funcional en ciertas regiones del cerebro de hurones. A su vez, diversos estudios han demostrado que ocurren cambios en la conectividad estructural en los límites de regiones funcionales identificadas mediante citoarquitectura [p.e. 19, 26]. En un trabajo reciente, Saygin et al. [27] afirman que mediante la conectividad estructural se pueden localizar las áreas especializadas en detección de caras en el giro fusiforme. Este trabajo es luego extendido por Osher et al. [22] a la localización de otras categorías visuales en la corteza cerebral. Tanto el trabajo de Saygin et al. [27] como el de Osher et al. [22] se basan en un modelo lineal que relaciona las respuestas funcionales en la corteza cerebral y la conectividad estructural. Sin embargo, dicho modelo no tiene en cuenta la alta variabilidad entre sujetos, tanto a nivel de conectividad estructural como de respuestas funcionales. Este tipo de estudio podría verse beneficiado con el uso de modelos lineales mixtos.

Los modelos lineales mixtos son una extensión de los modelos lineales ordinarios que permiten modelar tanto efectos fijos como efectos aleatorios. Esto es, efectos sobre la población promedio y efectos específicos de cada sujeto. Una particularidad fundamental de los modelos lineales mixtos es que se basan en la identificación explícita de las características individuales y de la población que queremos modelar. En este modelo, los efectos individuales siguen una misma distribución de probabilidad, pero los parámetros de dicha distribución varían entre individuos [20]. Al mismo tiempo, todos los sujetos comparten un mismo efecto poblacional. Dado el nivel de muestreo necesario para modelar la información del cerebro de cada sujeto, incluir la información de los efectos aleatorios hace crecer de forma multiplicativa la cantidad de información a procesar. Es por ello que el enfoque de máxima verosimilitud [21] utilizado en la mayoría de los paquetes de software [29], no es capaz de lidiar con todos estos parámetros de forma simultánea. En consecuencia, proponemos una nueva forma de resolver estos modelos basados en el método de multiplicadores alternados [7] (ADMM), un meta-algoritmo de optimización de la familia de los algoritmos proximales.

ADMM es una técnica para resolver problemas de optimización convexa que consiste en separar a éstos en piezas más pequeñas, cada una de las cuales es más fácil de manejar, para aproximar la solución de forma iterativa. Este enfoque nos proporciona una

optimización escalable y descentralizada, resolviendo el problema de ajustar un modelo a grandes conjuntos de datos. ADMM se basa en algoritmos no paralelos existentes. Puede ser visto como un coordinador modular que incentiva la colaboración de algoritmos más simples para resolver problemas mucho más grandes [7].

Otro beneficio de ADMM es que es simple de implementar a la vez que permite aplicar técnicas de regularización de manera sencilla. Estas técnicas son importantes para mejorar la interpretabilidad de nuestros modelos y evitar el sobreajuste. En nuestro caso, decidimos implementar Lasso [31] en base a la suposición de que sólo áreas específicas del cerebro están activas durante la ejecución de diferentes tareas. Lasso reduce algunos coeficientes y fija otros en 0, y por lo tanto intenta mantener las buenas características de selección de subconjuntos y ridge regression, dándonos también un modelo más consistente con la realidad.

Este trabajo está dividido en ocho capítulos. El capítulo 2 es una breve introducción teórica a la anatomía básica del cerebro, la resonancia magnética funcional y la resonancia magnética de difusión. El capítulo 3 es una introducción a los modelos lineales y al método de multiplicadores alternados. En el capítulo 4 abordamos un análisis de la relación existente entre estructura y función en el cerebro presentando los trabajos previos en el área. El capítulo 5 muestra nuestra contribución: Explicamos el algoritmo para calcular modelos lineales mixtos mediante la aplicación de multiplicadores alternados. Finalmente los capítulos 6, 7 y 8 presentan la discusión de los resultados obtenidos seguido de las conclusiones y las ideas para futuros trabajos.

2. NEUROCIENCIAS: INTRODUCCIÓN

En esta tesis estamos interesados en aplicar nuestra contribución a problemas en el área de la neurociencia, por lo cual creemos necesario incluir una introducción a la misma. En este capítulo presentamos una introducción a la anatomía básica del cerebro, así como también una breve introducción a la resonancia magnética de imágenes funcional y de difusión. Este capítulo está basado en los libros Diffusion MRI [4] y Neuroscience [16]. En caso de querer profundizar en este tema, por favor referirse a ellos.

2.1. Anatomía básica del cerebro

En el cuerpo humano, el sistema nervioso es el mecanismo encargado de analizar e integrar estímulos internos y externos, así como generar la reacción y los ajustes del organismo necesarios en respuesta a ellos. El sistema nervioso se divide en dos partes: el sistema nervioso central y el sistema nervioso periférico. El sistema nervioso central se compone del cerebro y la médula espinal. El sistema nervioso periférico se encarga de transmitir estímulos desde el ambiente hacia el sistema nervioso central, donde son integrados con estímulos internos para producir una respuesta. Esta respuesta viaja nuevamente por el sistema nervioso periférico y se traduce en movimientos del cuerpo o ajustes de órganos internos.

A nivel microscópico, el cerebro está compuesto mayormente por dos tipos de células: las neuronas y las glías. Las neuronas son células especializadas para la comunicación, que intercambian señales eléctricas entre sí mediante sus axones. Las células gliales sirven de soporte para las neuronas y se encargan del mantenimiento de los axones. A nivel macroscópico se pueden distinguir dos tipos de tejidos: la materia blanca y la gris. La materia blanca está mayormente compuesta por axones protegidos por una vaina de mielina, que proporciona aislamiento y permite transmitir señales nerviosas más rápidamente. La mielina es responsable de la apariencia blanca de este tipo de tejido. Por su parte, la materia gris, de la cual está compuesta la corteza cerebral, se encuentra formada mayormente por los cuerpos neuronales.

Morfológicamente hablando, el cerebro tiene una apariencia rugosa conformada por pliegues denominados circunvoluciones, separados por depresiones denominadas surcos. Las circunvoluciones y los surcos cerebrales cumplen una función muy importantes, estos pliegues ayudan a aumentar la superficie de la corteza cerebral dentro de un mismo volumen. Esto permite aumentar la capacidad del cerebro para procesar información manteniendo el mismo volumen del cráneo.

A nivel anatómico, los surcos forman divisiones cerebrales creando dos hemisferios y lóbulos dentro de estos. En cada hemisferio se distinguen cuatro lóbulos: el frontal, el parietal, el occipital y el temporal. Desde una perspectiva lateral (Fig. 2.2), el lóbulo frontal se encuentra en la región anterior de la corteza cerebral. El lóbulo parietal y el temporal se encuentran posteriores al frontal, con el lóbulo parietal situado por encima del lóbulo temporal. Finalmente el lóbulo occipital se sitúa en la región posterior de la corteza cerebral.

Cada uno de estos lóbulos cerebrales es responsable de varias funciones importantes. El lóbulo frontal se encarga de funciones motoras, producción del lenguaje, planeamiento

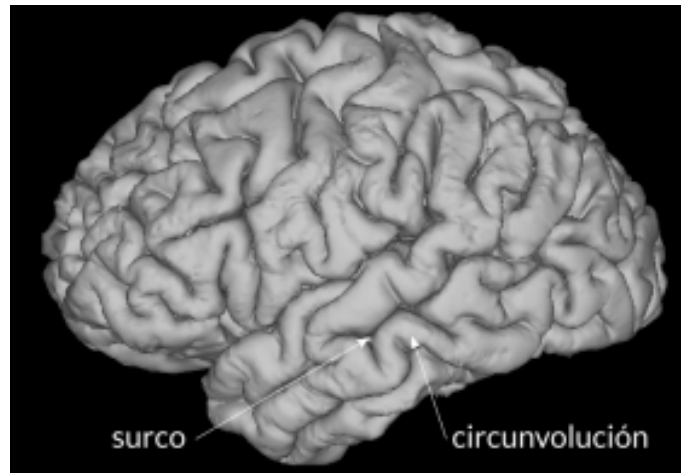


Fig. 2.1: Los surcos y circunvoluciones cerebrales

y otras funciones cognitivas de alto orden. El lóbulo parietal se ocupa de la comprensión del lenguaje, la orientación y percepción espacial y los sentidos somáticos, como el tacto y la temperatura. El lóbulo occipital está relacionado mayormente con el procesamiento visual. El lóbulo temporal está involucrado en la audición, la comprensión del habla y el reconocimiento visual.

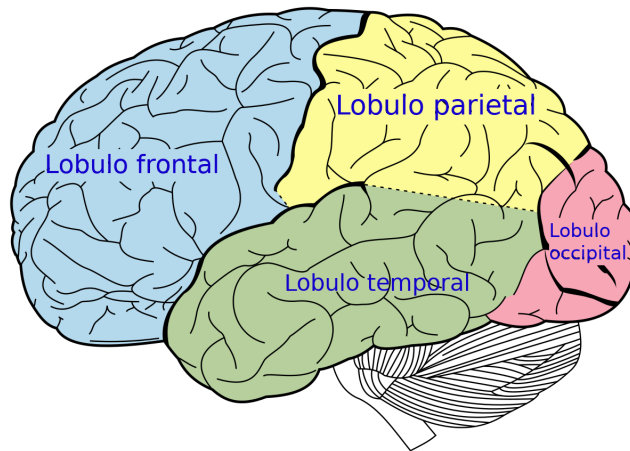


Fig. 2.2: Visión lateral de los lóbulos cerebrales [10]

Más allá de la división en lóbulos, distintos atlas cerebrales coexisten actualmente. Un atlas del cerebro es una forma de organizar las distintas regiones de éste en base a características particulares. Algunos se basan en características meramente anatómicas, por ejemplo, utilizando los surcos de mayor profundidad como delimitadores. Otros, en cambio, se basan en características funcionales al dividir el cerebro en base a, por ejemplo, las respuestas obtenidas a ciertos estímulos. También, existen atlas basados en la citoarquitectura (la composición de la corteza a nivel celular). En esta tesis en particular utilizaremos los atlas anatómicos de Desikan et al. [11] y Destrieux et al. [12]. Ambos atlas delimitan regiones basados en los surcos cerebrales. La principal diferencia entre ellos es que el atlas de Destrieux es una versión mucho más refinada y densa en cantidad de regiones que el

primero.

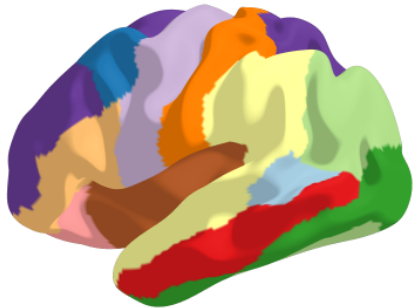


Fig. 2.3: Atlas Desikan [11]

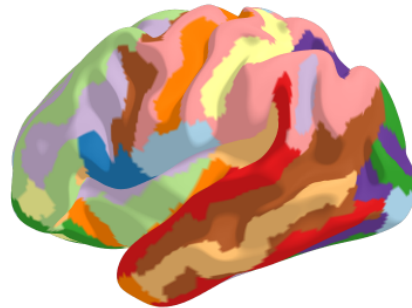


Fig. 2.4: Atlas Destrieux [12]

2.2. Introducción a la Resonancia Magnética

Las imágenes de resonancia magnética reflejan principalmente el agua y concentración de grasa; específicamente, reflejan la señal de los núcleos de hidrógeno. Los núcleos de hidrógeno poseen un dipolo magnético. Estos dipolos magnéticos, conocidos como spins, se alinean mediante la aplicación de un campo magnético externo. El campo magnético de rotación induce corriente en las bobinas receptoras colocado cerca del paciente y esta corriente es la señal utilizada para generar imágenes de MR, que se refleja como brillo en cada píxel de una imagen.

Una escaneo por resonancia magnética comienza con la excitación de los núcleos de hidrógeno con un pulso de radiofrecuencia que inclina su vector de magnetización en 90 grados a lo largo del campo magnético principal. Los datos de MRI se obtienen girando ciertos campos dentro y fuera en una secuencia prescrita, referida como una "secuencia de pulsos". La secuencia de pulsos determina cómo se verá la imagen, incluido el tipo de contraste de imagen, y su resolución y calidad.

2.2.1. Resonancia Magnética Funcional

La imagen por resonancia magnética funcional (fMRI) es un procedimiento no invasivo que permite estimar qué regiones cerebrales se 'activan' al ejecutar una tarea determinada. Cuando una región del cerebro es activada por una tarea cognitiva, la activación local de las neuronas genera un mayor requerimiento de oxígeno. Dado que las neuronas no cuentan con reservas de oxígeno, el cerebro responde ajustando su suministro de sangre a la región en un proceso conocido como respuesta hemodinámica. Este aumento en el flujo sanguíneo produce un aumento en la proporción de hemoglobina oxigenada en esa área específica. fMRI permite medir los niveles locales y transitorios de oxígeno en la sangre (en inglés Blood-oxygen-level dependent contrast imaging o abreviado BOLD-contrast imaging). Por lo tanto, esta técnica obtiene imágenes que reflejan la activación de distintas zonas del cerebro a partir del aumento en el flujo de oxígeno.

2.3. Resonancia Magnética de Difusión

La resonancia magnética de difusión se basa en el concepto de que las moléculas dentro de un fluido no se encuentran estáticas, sino que se mueven de forma aleatoria. A este fenómeno físico se lo conoce como difusión. De acuerdo a la primera ley de Fick, cuando un espécimen contiene diferentes regiones con diferentes concentraciones de moléculas éstas tienden, en promedio, a pasar de las regiones más concentradas a las menos concentradas.

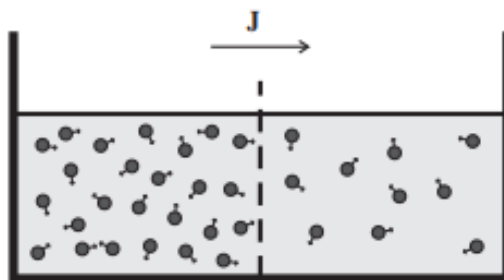


Fig. 2.5: Primera ley de Fick

A nivel molecular, la mezcla difusiva es el resultado de las colisiones entre moléculas o átomos en el líquido o gas. El incesante movimiento de las moléculas de agua de un tejido es influenciado por una variedad de factores, como restricciones en las membranas celulares, cito esqueletos y macromoléculas. La resonancia magnética ofrece una oportunidad única de cuantificar las características difusionales de una amplia gama de especímenes. Dado que los procesos de difusión son influenciados por la geometría estructural, la resonancia magnética puede ser usada para probar características estructurales del entorno de forma no invasiva. Mediante el empleo de la comprensión de cómo las características micro estructurales contribuyen al proceso de difusión general, puede ser posible obtener información valiosa sobre la micro estructura biológica simplemente observando el movimiento de las moléculas de agua que la componen. En particular, dentro de estructuras fibrosas como la materia blanca, las moléculas de agua tienden a difundir en la dirección de la fibra en la cual se encuentran. Las imágenes de difusión entonces permiten cuantificar la preferencia de orientación de la difusión y relacionarla con la orientación de los axones.

3. MODELOS LINEALES

El objetivo de esta tesis es presentar un modelo de regresión lineal capaz de modelar tanto efectos fijos como efectos aleatorios. En este capítulo presentamos una breve introducción a los modelos lineales, modelos lineales regularizados, modelos lineales mixtos. Finalmente, presentamos el método de los multiplicadores alternados, que permite resolver problemas de optimización convexos, y el cual adaptaremos a la resolución de modelos lineales mixtos.

3.1. Análisis de regresión

El análisis de regresión es un método estadístico utilizado para estimar la relación entre distintas variables. Dados un vector de parámetros desconocidos β , un conjunto de variables independientes A y un conjunto de variables dependientes Y , el análisis de regresión explora cuáles de las variables independientes están relacionadas con las variables dependientes, y las formas de estas relaciones. Más específicamente, ayuda a entender como cambia el valor de una variable independiente cuando algunas de las variables dependientes se modifican y las demás se dejan fijas.

Siendo A e Y variables conocidas, el análisis de regresión ofrece una forma de minimizar la distancia entre los valores medidos y los predichos de la variable dependiente Y a través de una función f conocida como función de regresión, quedando el problema modelado de la siguiente manera

$$Y \approx f(A, \beta, \epsilon) \quad (3.1)$$

donde ϵ representa un posible error o ruido introducido por la aproximación realizada por la función f bajo ciertas restricciones. Suponiendo que el error ϵ esta distribuido de forma Gaussiana, dicha aproximación puede ser formalizada como:

$$E(Y|A) = f(A|\beta), \quad (3.2)$$

donde la función f debe ser especificada. En otras palabras, el análisis de regresión estima la esperanza condicional de la variable dependiente dadas las variables independientes. Algunas veces la forma de la función propuesta se basa en el conocimiento o la hipótesis previa acerca de la relación entre las variables dependientes e independientes y en caso de no disponer de tales conocimientos, es preferible elegir una forma flexible para f .

3.2. Modelo lineal

Un modelo lineal predice el valor de una variable a través de otras suponiendo que la relación entre éstas es lineal y modelando dicha relación mediante una función de la forma

$$y = a\beta + b \quad (3.3)$$

la cual es posible resolver mediante el método de cuadrados mínimos: Sean $(a_k, y_k)_{k=1}^n$ un conjunto de puntos tal que $a_k, y_k \in \mathbb{R}^2$ y sea $\{f_j(x)\}_{j=1}^m$ una base de M funciones

linealmente independientes. Nuestro objetivo consiste, entonces, en encontrar una función $f'(a)$ que sea combinación lineal de dichas funciones, de modo tal que $f(a_k) \approx y_k$, esto es:

$$f'(a) = \sum_{j=1}^m c_j f_j(a) \quad (3.4)$$

Por tanto, se trata de hallar los M coeficientes c_j que hagan que la función $f'(a)$ ofrezca la mejor aproximación para los puntos dados (a_k, y_k) , donde ‘mejor aproximación’ hace referencia a minimizar el error acumulado en cada punto respecto al error total.

Este esquema es fácilmente extensible a notación matricial, donde esta idea se corresponde con la de resolver el siguiente sistema de ecuaciones:

$$Y = A\beta \quad (3.5)$$

donde A es una matriz de $M \times N$, β es el vector de parámetros desconocidos de $N \times 1$ e Y es un vector solución de $M \times 1$. En particular, una forma de resolver este problema es mediante el método de cuadrados mínimos. Este método se basa en la idea de resolver el sistema de ecuaciones mencionado computando un vector β que minimice la función de costo:

$$\operatorname{argmin}_{\beta} \|A\beta - Y\|_2^2 \quad (3.6)$$

Esta función se deriva de la noción de error. El error de la función $f(A_i)$ para un valor, (A_k, Y_k) , se define como

$$e_k = Y_k - A_k\beta \quad (3.7)$$

donde, como definimos anteriormente $f(A_k)$ es equivalente a $A_k\beta$, A partir de esto, podemos definir el error cuadrático medio:

$$E_c(f) = \sqrt{\frac{\sum_{k=1}^n (e_k)^2}{n}} \quad (3.8)$$

La aproximación por mínimos cuadrados se basa, entonces, en la minimización del error cuadrático medio y tiene solución general cuando se trata de un problema de aproximación lineal (lineal en sus coeficientes c_j) cualesquiera que sean las funciones base $f_j(a)$.

3.2.1. Modelo lineal con LASSO

La regresión aplicando Lasso [31] (least absolute shrinkage and selection operator, por sus siglas en inglés), es un método de análisis de regresión que realiza selección de variables. Esto mejora la exactitud e interpretabilidad de modelos estadísticos al seleccionar solamente un subconjunto de las variables provistas para usar en el modelo final. Este procedimiento se realiza mediante la aplicación de una regularización ℓ_1 sobre los parámetros propuestos. Es una relajación del problema implícito de usar una norma ℓ_0 . Esto quiere decir que agrega un factor de suma absoluta a los coeficientes con el fin de descartar aquellos que influyen en menor medida en la solución de nuestro modelo. Luego, nuestro sistema lineal que anteriormente estaba constituido como

$$\operatorname{argmin}_{\beta} (\|A\beta - Y\|_2^2) \quad (3.9)$$

se convierte en uno de la forma

$$\operatorname{argmin}_{\beta} (\|A\beta - Y\|_2^2 + \alpha \|\beta\|_1) \quad (3.10)$$

En otras palabras, resuelve la misma minimización de cuadrados mínimos usada con anterioridad, pero agregando una penalización con $\alpha \|\beta\|_1$, donde α es una constante y $\|\beta\|_1$ la norma ℓ_1 de los parámetros a calcular. La regularización ℓ_1 permite que ciertos valores se aproximen a cero, a diferencia de la regularización ℓ_2 que solo ‘encoge’ la diferencia entre los distintos parámetros:

$$\operatorname{argmin}_{\beta} (\|A\beta - Y\|_2^2 + \alpha \|\beta\|_2) \quad (3.11)$$

Esto se debe principalmente a la diferencia de forma entre ambas regularizaciones, la cual es introducida por el hecho de tomar la norma de los valores de β en lugar del cuadrado como en el caso ℓ_2 . A partir de la figura 3.1 [2], se puede ver que la región de

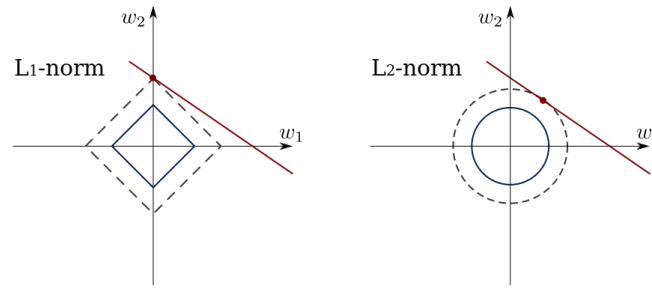


Fig. 3.1: Interpretación geométrica de las normas ℓ_1 y ℓ_2 : un objeto convexo que yace tangente al límite, como la línea que se muestra, es probable que en el caso ℓ_1 se encuentre con una esquina del cuadrado. Mientras tanto, en el caso de la norma ℓ_2 , los puntos en el límite no se distinguen unos de otros, haciendo que la probabilidad de tener contacto con cualquiera de ellos sea la misma [2].

restricción definida por la norma ℓ_1 es un cuadrado rotado de manera que sus esquinas descansan sobre los ejes, mientras que la región definida por la norma ℓ_2 es un círculo, que es invariable en rotación y, por lo tanto, no tiene esquinas. Como consecuencia, al aproximar un objeto hacia el límite, como la línea que se muestra en la figura 3.1, es probable que en el caso ℓ_1 ésta se encuentre con una esquina del cuadrado (o un hipercubo para una situaciones de mayor dimensión). Éstas, al estar situadas sobre los ejes, hacen que valores próximos a cero sean forzados hacia éste. Mientras tanto, en el caso de la norma ℓ_2 , los puntos en el límite no se distinguen unos de otros, haciendo que la probabilidad de tener contacto con cualquiera de ellos sea la misma.

3.3. Modelo lineal mixto

El termino ‘modelos mixtos’ refiere a modelos que usan tanto efectos fijos como aleatorios en el mismo análisis. Los efectos fijos son aquellos de interés general en el conjunto de datos, y que están presentes a lo largo de cada uno de las mediciones. Los efectos aleatorios modelan características que son producto de las mediciones específicas de cada elemento del conjunto, como por ejemplo, las variaciones específicas de cada sujeto, así como también las variaciones entre mediciones de la misma característica tomadas en distinto momento.

Los modelos lineales mixtos proveen, entonces, un enfoque flexible y general para resolver problemas que combinan ambas necesidades. Nos proporcionan todas las ventajas de los modelos lineales, mientras que nos permiten definir características sobre la estructura de los datos. Al incluir patrones de correlación de forma explícita nos permiten modelar relaciones internas entre los datos al mismo tiempo que nos posibilitan acceder a una mejor interpretación de los mismos.

A partir de lo expuesto por Laird et al. [20], podemos extender un modelo lineal a un modelo lineal mixto de la siguiente manera:

$$Y_i = A_i\beta + Z_i\mathcal{B}_i + \epsilon_i \quad i = 1, \dots, M \quad (3.12)$$

donde:

- ▷ Y_i es un vector de observaciones conocidas para el sujeto i .
- ▷ β es un vector de efectos fijos desconocido de la población.
- ▷ \mathcal{B}_i es un vector de efectos aleatorios para el sujeto i .
- ▷ ϵ_i es un vector de errores aleatorios.
- ▷ A y Z son matrices de diseño que relacionan a Y_i con β y con u_i respectivamente

Basándonos en Lindstrom et al. [21] podemos expresar las mediciones disponibles como

$$Y_i | \mathcal{B}_i \sim \mathcal{N}(A_i\beta + Z_i\mathcal{B}_i, \sigma^2\Lambda_i) \quad i = 1, \dots, M \quad (3.13)$$

donde Y_i es independiente de Y_j para todo $i \neq j$, Λ_i no depende de i excepto porque su tamaño debe ser de $n_i \times n_i$, β es un vector de efectos fijos de tamaño p y los \mathcal{B}_i (un vector de efectos aleatorios de tamaño q asociados con el individuo i) están distribuidos como $\mathcal{N}(0, \sigma^2 D)$, independientes uno del otro. Además D es una matriz de covarianza definida positiva y se supone que ϵ tiene distribución $\mathcal{N}(0, R_i)$ donde R_i es una matriz de covarianza definida positiva.

Podemos entonces reescribir (3.13) de forma matricial tomando:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_M \end{bmatrix}, \quad A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_M \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} \mathcal{B}_1 \\ \mathcal{B}_2 \\ \vdots \\ \mathcal{B}_M \end{bmatrix},$$

$$\Sigma = \text{diag}(\Sigma_1, \Sigma_2, \dots, \Sigma_M) \quad \bar{D} = \text{diag}(D, D, \dots, D) \quad (3.14)$$

$$Z = \text{diag}(Z_1, Z_2, \dots, Z_M) \quad \Lambda = \text{diag}(\Lambda_1, \Lambda_2, \dots, \Lambda_M)$$

Para luego, unificando todas las observaciones, obtener un modelo que queda definido como

$$Y | \mathcal{B} \sim \mathcal{N}(A\beta + Z\mathcal{B}, \sigma^2\Lambda) \quad (3.15)$$

donde

$$\mathcal{B} \sim \mathcal{N}(0, \sigma^2 \bar{D}) \quad (3.16)$$

y la distribución marginal de Y es

$$Y \sim \mathcal{N}(A\beta, \Sigma) \quad \Sigma = \sigma^2(\Lambda + Z\bar{D}Z^t) \quad (3.17)$$

Cuando Λ y D son conocidas, el estimador estándar para β y \mathcal{B} es el estimador de mínimos cuadrados:

$$\hat{\beta}(\theta) = (A^t V^{-1} A)^{-1} A^t V^{-1} Y \quad \hat{\mathcal{B}}(\theta) = \bar{D} Z^t V^{-1} (Y - A \hat{\beta}(\theta)) \quad (3.18)$$

donde

$$V = \Lambda + Z D Z^t \quad (3.19)$$

y el vector θ contiene los elementos únicos de D y los parámetros en Λ .

Como se mencionó con anterioridad las matrices A y Z se corresponden a matrices de diseño que relacionan los parámetros a estimar en el modelo. Como se puede observar en 3.18, la resolución de nuestro modelo lineal mixto se compone ahora de una cantidad de parámetros que crece de forma exponencial.

3.4. Método de los multiplicadores alternados (ADMM)

El método de los multiplicadores alternados (desde ahora ADMM) es un algoritmo simple pero poderoso que se adapta bien a la optimización de problemas convexos. Se plantea como un procedimiento de descomposición en coordenadas en el que las soluciones a pequeños subproblemas locales se coordinan para encontrar una solución a un problema global. Fue introducido por primera vez a mediados de 1970 por Gabay y Mercier [13] y por Glowinski y Marrocco [15]. El hecho de que ADMM haya sido desarrollado con tanta anticipación a la actual disponibilidad de sistemas de computación distribuidos a gran escala y problemas de optimización masiva puede explicar por qué no es tan ampliamente conocido hoy como creemos que debería ser. La principal contribución de este trabajo será, entonces, la adaptación de este algoritmo a los modelos lineales mixtos, motivados por la necesidad de poder obtener resultados en conjuntos de datos de gran escala. El algoritmo ADMM se basa en algunos conceptos que, si bien no son necesarios para el entendimiento del algoritmo principal, aportan información útil. Se presenta un pequeño resumen de ellos a continuación:

3.4.1. Doble ascenso

Consideremos el problema:

$$\begin{aligned} &\text{minimizar } f(x) \\ &\text{sujeto a } Ax = b \end{aligned} \quad (3.20)$$

donde $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ y $f : \mathbb{R} \rightarrow \mathbb{R}$ es una función convexa. El Lagrangiano del problema (3.20) se define como

$$L(x, y) = f(x) + y^T (Ax - b) \quad (3.21)$$

y la función dual es

$$g(y) = \inf_x L(x, y) = -f^*(-A^T y) - b^T y \quad (3.22)$$

donde y es el multiplicador de Lagrange. El problema de optimización de descenso doble se transforma entonces en

$$\operatorname{argmax}_y g(y) \quad (3.23)$$

pudiendo recuperar x^* a partir de y^* como

$$x^* := \operatorname{argmin}_x L(x, y^*) \quad (3.24)$$

En el método de ascenso doble (o dual) resolvemos este problema mediante el descenso por gradiente. Suponiendo que g es diferenciable, primero buscamos $x^+ = \operatorname{argmin}_x L(x, y)$ y luego obtenemos el residuo $\nabla g(y) = Ax^+ - b$. El método consiste entonces en iterar y actualizar sobre

$$\begin{aligned} x^{k+1} &:= \operatorname{argmin}_x L(x, y^k) \\ y^{k+1} &:= y^k + \alpha^k (Ax^{k+1} - b) \end{aligned} \quad (3.25)$$

donde $\alpha^k > 0$ es el tamaño del paso de iteración y k es el contador de iteraciones. Para más información del método, referirse a [7, Capítulo 2.1]

3.4.2. Descomposición Doble

Una de las mayores ventajas del método de ascenso dual es que, en algunos casos, puede dar lugar a algoritmos descentralizados. Sea f una función separable

$$f(x) = \sum_{i=1}^N f_i(x_i) \quad (3.26)$$

donde $x = (x_1, \dots, x_N)$ y las variables $x_i \in \mathbb{R}$. Si particionamos la matriz A en

$$A = [A_1 \dots A_N] \quad \text{tal que} \quad Ax = \sum_{i=1}^N A_i x_i \quad (3.27)$$

el Lagrangiano entonces puede ser escrito como:

$$L(x, y) = \sum_{i=1}^N L_i(x_i, y) = \sum_{i=1}^N (f_i(x_i) + y^T A_i x_i - (1/N)y^T b), \quad (3.28)$$

el cual es separable en x . Gracias a esto, la minimización del paso (3.25) puede ser dividida en N problemas separables, los cuales pueden ser resueltos en paralelo tal que:

$$\begin{aligned} x_i^{k+1} &:= \operatorname{argmin}_{x_i} L_i(x_i, y^k) \\ y^{k+1} &:= y^k + \alpha^k (Ax^{k+1} - b) \end{aligned} \quad (3.29)$$

Cuando esto ocurre, el método de descenso dual se conoce como descomposición dual. Para más información del método, referirse al capítulo 2.2 del trabajo de Boyd et al. [7]

3.4.3. Langrangianos aumentados y el método de los multiplicadores

El método de los Lagrangianos aumentados surgió como una forma de ofrecer robustez al método de ascenso dual, ofreciendo convergencia sin la necesidad de realizar suposiciones sobre f . El Lagrangiano aumentado para (3.20) es

$$L_\rho(x, y) = f(x) + y^T(Ax - b) + (\rho/2) \|Ax - b\|_2^2 \quad (3.30)$$

donde $\rho > 0$ es un parámetro de penalización. Invirtiendo el razonamiento, el Lagrangiano aumentado puede verse como una forma de resolver el problema

$$\begin{aligned} &\text{minimizar } f(x) + (\rho/2) \|Ax - b\|_2^2 \\ &\text{sujeto a } Ax = b \end{aligned} \quad (3.31)$$

el cual es equivalente al problema (3.20) ya que para cualquier posible x el término agregado es cero. El beneficio de incluir esta penalización es que $g_\rho(y) = \inf_x L_\rho(x, y)$ presenta mayor diferenciabilidad mientras que el gradiente del método se conserva. Si bien este método presenta una mejor convergencia que el descenso dual, cuando f es separable, el Lagrangiano aumentado L_ρ no lo es, por lo que la minimización de x no puede realizarse en paralelo para cada x_i . Esto significa que el método de los multiplicadores aumentados no puede aplicarse en una descomposición doble. Para abordar esta problemática, necesitamos un paso mas.

3.4.4. Algoritmo

ADMM es un algoritmo que fue pensado con la idea de combinar la descomposición dual con las propiedades de convergencia del método de los multiplicadores. Para conseguir este objetivo, lleva el problema del descenso doble a uno de la forma:

$$\begin{aligned} &\text{minimizar } f(x) + g(z) \\ &\text{sujeto a } Ax + Bz = c \end{aligned} \quad (3.32)$$

A diferencia de (3.20) la variable original x ha sido dividida en dos, llamadas x y z . El valor óptimo de este problema, queda entonces definido como:

$$p^* = \inf\{f(x) + g(z) | Ax + Bz = c\} \quad (3.33)$$

Al igual que en el método de los multiplicadores, el Lagrangiano aumentado está dado por

$$L_\rho(x, y, z) = f(x) + g(z) + y^T(Ax + Bz - c) + (\rho/2) \|Ax - Bz - c\|_2^2 \quad (3.34)$$

ADMM consiste entonces en la aplicación sucesiva de las siguientes iteraciones

$$x^{k+1} := \underset{x}{\operatorname{argmin}} L_\rho(x, z^k, y^k) \quad (3.35)$$

$$z^{k+1} := \underset{z}{\operatorname{argmin}} L_\rho(x^{k+1}, z, y^k) \quad (3.36)$$

$$y^{k+1} := y^k + \rho(Ax^{k+1} + Bz^{k+1} - c) \quad (3.37)$$

El algoritmo es muy similar al descenso dual, solo que éste consiste en dos pasos de minimización, uno para la variable x (3.35) y otro para la variable z (3.36) y la consiguiente

actualización de la variable dual (3.37) usando el mismo parámetro para definir el tamaño del paso, ρ , que en el Lagrangiano aumentado. El método de los multiplicadores para (3.32) se define como

$$\begin{aligned} (x^{k+1}, z^{k+1}) &:= \underset{x, z}{\operatorname{argmin}} L_\rho(x, z, y^k) \\ y^{k+1} &:= y^k + \rho(Ax^{k+1} + Bz^{k+1} - c) \end{aligned} \quad (3.38)$$

donde el Lagrangiano es minimizado de forma conjunta respecto de ambas variables. En ADMM, por otro lado, x y z son alternados de forma secuencial, lo que da cuenta de la parte de ‘alternados’ en el nombre del algoritmo.

3.4.5. Forma escalada

ADMM puede escribirse de una forma levemente distinta, la cual suele ser muchas veces mas conveniente, combinando los términos lineales y cuadráticos en el Lagrangiano aumentado y escalando la variable dual. Dado el residuo $r = Ax + Bz - c$, tenemos:

$$y^T r + (\rho/2)\|r\|_2^2 = (\rho/2)\|r + (1/\rho)y\|_2^2 - (1/2\rho)\|y\|_2^2 \quad (3.39)$$

$$= (\rho/2)\|r + u\|_2^2 - (\rho/2)\|u\|_2^2 \quad (3.40)$$

donde $u = (1/\rho)y$ es la variable dual escalada, lo que nos permite reescribir (3.32) como

$$x^{k+1} := \underset{x}{\operatorname{argmin}} \left(f(x) + (\rho/2)\|Ax + Bz^k - c + u^k\|_2^2 \right) \quad (3.41)$$

$$z^{k+1} := \underset{z}{\operatorname{argmin}} \left(g(z) + (\rho/2)\|Ax^{k+1} + Bz - c + u^k\|_2^2 \right) \quad (3.42)$$

$$u^{k+1} := u^k + Ax^{k+1} + Bz^{k+1} - c \quad (3.43)$$

Tanto las expresiones definidas en (3.35-3.37) como las definidas en (3.41-3.43) son equivalentes, pero las expresiones en la versión escalada suelen más compactas.

4. RELACIONANDO ESTRUCTURA Y FUNCIÓN EN EL CEREBRO

En esta tesis estamos interesados en utilizar un modelo lineal mixto para estudiar la relación entre estructura y función en el cerebro humano. En este capítulo presentamos una introducción al estado del arte, así como también a las hipótesis biológicas que respaldan dichos estudios.

4.1. Relación entre estructura y función en el cerebro

Desde la introducción de la teoría de la neurona [9] por el español Santiago Ramón y Cajal a finales del siglo XIX hasta la actualidad, se han realizado grandes avances en el entendimiento de la composición y la función del sistema nervioso. Como el elemento físico en el que todos los procesos de información neuronal suceden, es relevante entender cómo la arquitectura del cerebro manifiesta los diversos comportamientos que genera. Lamentablemente, los datos anatómicos y de conectividad completos son imposibles de obtener en la actualidad. Alcanzar un mapa detallado de todas las neuronas y sus interconexiones en un cerebro de mamífero está simplemente fuera del alcance tecnológico actual. Motivado por las técnicas experimentales disponibles, surgió y ha prevalecido a lo largo del tiempo, la noción de que diferentes partes se especializan en procesos particulares y que existe una relación entre la estructura y la funcionalidad [6, 14, 8, 24]. Esto implica que, conociendo cómo una región está conectada con el resto del cerebro, se podría predecir su función.

Penfield et al. [25] fue uno de los pioneros en mostrar que ciertas regiones del cerebro están funcionalmente especializadas. Intentando explicar y curar ciertas enfermedades neurológicas como la epilepsia y aprovechando el hecho de que el cerebro no siente dolor, Penfield et al. [25] aplicaba descargas eléctricas en las distintas áreas cerebrales y les preguntaba a sus pacientes, que estaban despiertos, qué sentían. Al aplicar estas descargas, descubrió una zona cerebral donde están mapeadas distintas funciones motoras y sensoriales. Este mapa sensorial reflejaba la sensibilidad de cada una de las partes de nuestra anatomía. Penfield decidió representar este área como si de una forma humana se tratara, dando lugar al Homúnculo de Penfield.

Sin embargo, desde los experimentos realizados por Penfield et al. [25] al día de hoy, los tiempos han cambiado. La necesidad de técnicas de análisis no invasivas propició la utilización de imágenes funcionales de resonancia magnética (fMRI) e imágenes de difusión (dFMI), aportando a la vez nuevas formas de investigar la relación entre la conectividad y la función cerebral in vivo [18]. Las imágenes funcionales nos permiten obtener una imagen representando la respuesta cerebral a distintos estímulos externos. Respecto a las imágenes de difusión, permiten estimar cómo las partículas de agua difunden dentro de la materia blanca. Utilizando algoritmos de tractografía, es posible entonces estimar la conectividad estructural entre distintos puntos del cerebro. Esto se hace generalmente mediante un proceso de monte carlo, en el cual se estiman cuantas partículas de agua que comienzan en un punto del cerebro llegan a otro.

Actualmente distintos trabajos sostienen que el patrón de conectividad estructural

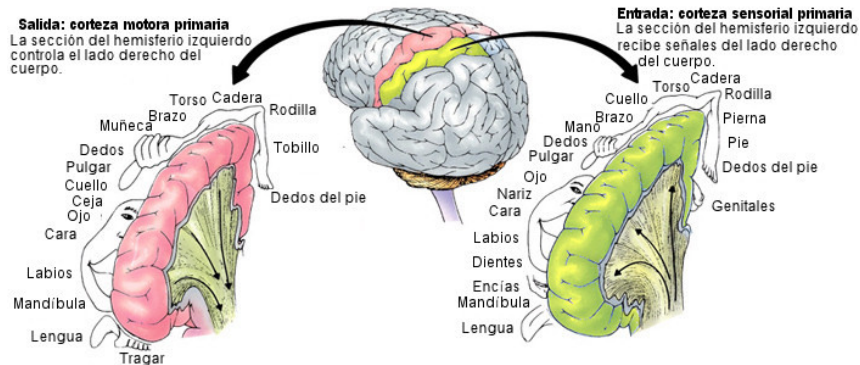


Fig. 4.1: Homúnculo de Penfield [1]: Es una representación pictórica de las divisiones anatómicas de la corteza motora primaria y la corteza somatostésica primaria desarrollada por el Dr. Wilder Penfield.

contiene información suficiente para predecir contrastes funcionales específicos [28, 22]. Un trabajo a destacar es el de Osher et al. [22], en el cual se predice la localización de distintas categorías visuales en la corteza cerebral. Para lograr esto, Osher et al. computan la conectividad de distintos puntos en la corteza cerebral (representados como voxels) a cada una de las 148 regiones del atlas anatómico de Destrieux [12]. De esta manera, cada voxel dentro de la corteza cerebral es descrito por un vector de probabilidades de conexión a las otras regiones del cerebro. Por el lado de la resonancia magnética funcional, Osher et al. [22] le presenta a cada uno de los participantes de su experimento, estímulos que consisten en pequeños videoclips de 3 segundos que intercalan caras, cuerpos, escenarios, objetos e imágenes con ruido. Por cada uno de estos estímulos obtienen una imagen de fMRI, que presenta el grado de activación de cada voxel durante la tarea. Mediante un análisis estadístico, Osher et al. transforman dichos mapas de activación en mapas de z-scores, que muestran qué tan correlacionado está cada voxel con la tarea.

Luego de escanear 26 participantes, Osher et al. [22] utilizan la información de difusión y funcional para predecir función a partir de conectividad, mediante una rutina de validación cruzada leave-one-subject-out (LOOCV). En dicha rutina, entrenan modelos lineales utilizando los voxels dentro de cada región dejando a uno de los participantes del grupo fuera, y utilizando a este último para validar el entrenamiento, tratando de realizar una predicción de la respuesta funcional esperada en cada región. Esto es repetido para cada uno de los participantes, obteniendo así predicciones independientes entre sujetos.

El problema planteado por Osher et al. [22] se modela de la siguiente manera: Para cada voxel de cada región modelada se tienen asociados dos valores distintos:

1. Un valor de contraste funcional representando la activación del voxel durante la tarea elegida.
2. Un vector de 148 dimensiones representando la probabilidad de conexión de esa región con cada una de las regiones restantes.

Se toman entonces los valores obtenidos para cada uno de los voxels de una región específica y se concatenan con el fin de producir:

- Un vector y de largo N con las respuestas funcionales de cada voxel para cada sujeto

del experimento (donde N representa la suma total de todas los voxels para la región específica de los sujetos disponibles).

- Una matriz A de vectores de conectividad para cada voxel en la región seleccionada ($N \times 148$ donde las filas representan el voxels en sí y las columnas son la probabilidad de conexión de dicho voxel con las demás regiones disponibles).

Estos datos son utilizados para modelar una regresión lineal (ver figura 4.2) entre la respuesta de contraste y el vector de conectividad de cada voxel:

$$y = Ax$$

Osher et al. muestran que utilizando este modelo lineal, se puede predecir la activación funcional de cada sujeto mejor que utilizando la activación promedio del grupo. Dado que el modelo utilizado por Osher et al. [22] no tiene en cuenta las diferencias entre sujetos, consideramos que pierde una capa de análisis importante. Este tipo de trabajos podrían verse beneficiados en utilizar modelos lineales mixtos, caracterizando las variaciones específicas de cada sujeto.

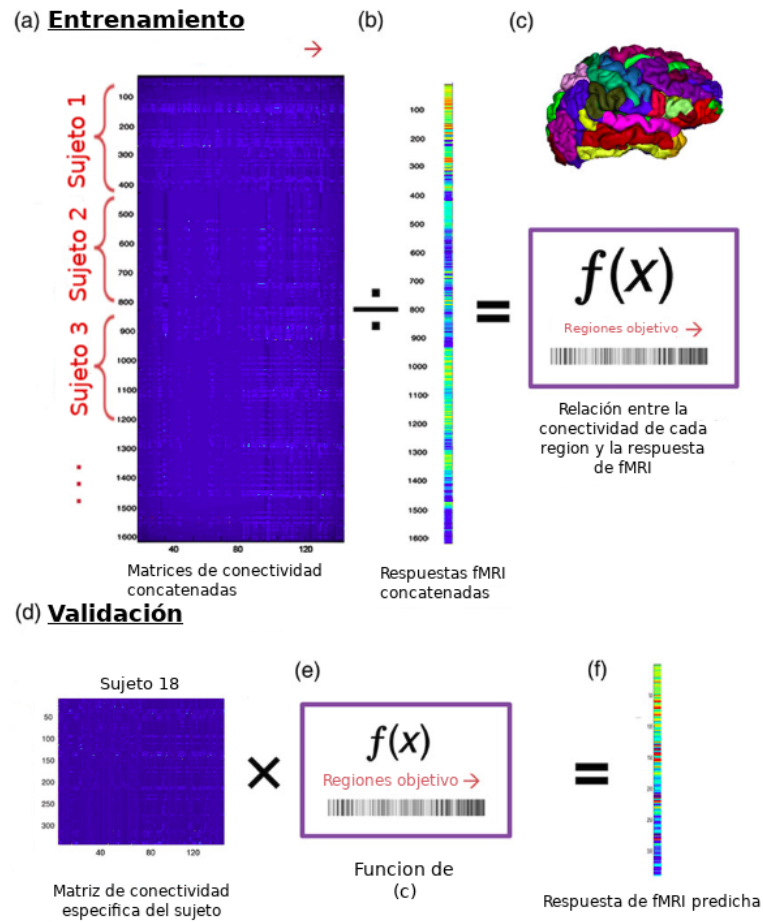


Fig. 4.2: Modelo propuesto por Osher et al. [22]. (a) Se calcula la conectividad estructural de cada sujeto y se concatenan uno sobre el otro (excluyendo los sujetos de validación) dejando los voxels sobre cada una de las filas y las regiones en las columnas. (b) Se concatena el contraste de respuesta de fMRI. (c) Se modela la relación entre (a) y (b) mediante una regresión lineal. (d) Los sujetos de validación también cuentan con una matriz de conectividad utilizada en (e) para, mediante la misma función modelada en (c), predecir la información de fMRI correspondiente a este sujeto (f).

5. CONTRIBUCIÓN METODOLÓGICA

Como mencionamos en el capítulo 4, Osher et al. [22] demostraron que dado un conjunto de sujetos y usando solo conectividad estructural, medida a través de imágenes de difusión, es posible predecir la activación funcional de diferentes tipos de respuestas en el cerebro. Sin embargo, no tuvieron en cuenta la alta variabilidad entre sujetos, tanto en conectividad como en activaciones funcionales. Nuestra hipótesis es que la respuesta media depende al mismo tiempo de la combinación de efectos poblacionales como de efectos específicos a nivel sujeto, por lo que al tenerlos en cuenta dentro de nuestro modelo, deberíamos ser capaces de mejorar nuestra estimación. Tiene sentido, entonces, abordar esta situación modelando dicha relación con un modelo lineal mixto.

Los modelos lineales mixtos [20] son una extensión de los modelos lineales que permiten tanto el modelo de efectos fijos como de efectos aleatorios, para referirse respectivamente al promedio poblacional y a los efectos específicos de cada sujeto.

Sea $\beta \in \mathbb{R}^{p \times 1}$ un vector que denota parámetros desconocidos de la población y sea $A_i \in n_i \times p$ la matriz de conexión para el i -ésimo sujeto que relaciona β con su respectiva matriz de activación funcional y_i . Sea \mathcal{B}_i un vector de $k \times 1$ efectos individuales desconocidos y Z_i una matriz conocida de $n_i \times k$ que modela la distribución de los voxels de activación del i -ésimo sujeto. Entonces, para cada sujeto i tenemos el siguiente modelo:

$$y_i = A_i\beta + Z_i\mathcal{B}_i + \epsilon_i \in \mathbb{R}^{n_i}, \quad \mathcal{B}_i \in \mathbb{R}^k \sim \mathcal{N}(\mu_i, \Sigma_i), \quad \epsilon_i \sim \mathcal{N}(0, R_i) \quad (5.1)$$

donde cada \mathcal{B}_i se supone independiente al igual que ϵ_i [20].

Como en los modelos lineales, podemos entonces describir a los modelos lineales mixtos mediante la distribución de una variable-vector de respuesta aleatoria \mathcal{Y} , cuyo valor observado es y_{obs} y \mathcal{B} , su vector de efectos aleatorios. Entonces, si en un modelo lineal la distribución de \mathcal{Y} es normal multivariada

$$\mathcal{Y} \sim \mathcal{N}(A\beta, \sigma^2 W^{-1}) \quad (5.2)$$

con W una matriz diagonal de muestras de las variables independientes y parámetros en β y σ , en un modelo lineal mixto la distribución *condicional* de \mathcal{Y} dado $\mathcal{B} = b$ es de la forma

$$(y|\mathcal{B} = b) \sim \mathcal{N}(A\beta + Zb, \sigma^2 W^{-1}), \quad (5.3)$$

y la distribución de \mathcal{B} es también normal multivariada, con media cero y matriz de covarianza Σ de $k \times k$ tal que

$$\mathcal{B} \sim \mathcal{N}(\mu_b, \Sigma) \quad (5.4)$$

Dado que Σ es una matriz de covarianza debe ser definida positiva, por lo que es conveniente expresar el modelo en términos de un factor de covarianza Λ_θ , el cual es una matriz de $k \times k$, y con parámetro θ . De esta manera, es posible expresar la matriz de covarianza Σ como

$$\Sigma_{\sigma\theta} = \sigma^2 \Lambda_\theta \Lambda_\theta^t \quad (5.5)$$

donde σ es la misma varianza que en la distribución condicional de la ecuación (5.3). Desde la perspectiva de la estabilidad computacional y la eficiencia, es conveniente reformular el

modelo para que θ solo aparezca en la distribución condicional del vector respuesta dados los efectos aleatorios b . La reformulación se realiza mediante la definición de una variable de efectos aleatorios esférica, \mathcal{U} con distribución

$$\mathcal{U} \sim \mathcal{N}(0, \sigma^2 I) \quad (5.6)$$

y suponiendo que cada y_i está centrada, obtenemos

$$\mathcal{B} - \mu_b = \Lambda_\theta \mathcal{U} \quad (5.7)$$

donde μ_b es la media de \mathcal{B} , ecuación (5.4). Entonces \mathcal{B} tendrá la distribución $\mathcal{N}(\mu_b, \Sigma_\theta)$ deseada. La distribución condicional, ecuación (5.3), del vector respuesta dado un conjunto de efectos aleatorios, ahora puede ser formulada como:

$$\begin{aligned} (y|\mathcal{U} = u) &\sim \mathcal{N}(\mu_{y|\mathcal{U}=u}, \sigma^2 W^{-1}) \\ \text{donde } \mu_{y|\mathcal{U}=u} &= A\beta + Z(\Lambda_\theta u + \mu_b) \end{aligned} \quad (5.8)$$

es un vector de predictores lineales que pueden ser interpretados como una media condicional. En una forma similar, definimos también $\mu_{\mathcal{U}|y=y_{obs}}$ como la media condicional de los efectos aleatorios esféricos, dado el valor observado del vector de respuesta.

Un problema común en los estudios de neuroimágenes es que a menudo hay muchas más variables potenciales que necesitan ser consideradas que casos clínicos en el estudio, lo que dificulta el modelado estadístico. Una forma de resolver esta problemática aplicada a la ecuación (5.1) es mediante el método de mínimos cuadrados penalizados [5]. En particular, este método consiste en minimizar la suma de cuadrados residual ponderada por una penalización, definido como:

$$r^2(\theta, \beta, u) = \rho^2(\theta, \beta, u) + \|\Lambda_\theta u\|_2^2 \quad (5.9)$$

donde

$$\rho^2(\theta, \beta, u) = \|W^{1/2}[y_{obs} - \mu_{y|\mathcal{U}=u}]\|_2^2 \quad (5.10)$$

es la suma de cuadrados. Esta función objetivo es obtenida a partir de la formulación probabilística de nuestro problema en ecuación (5.8): de ecuación (5.8) y ecuación (5.6) tenemos

$$p(y|u) \propto -\exp(\rho^2(\theta, \beta, u, \mu_b)), \quad p(u) \propto -\exp(\|u\|^2) \quad (5.11)$$

Entonces, la ecuación (5.9) es obtenida usando la definición de probabilidad condicional:

$$p(y, u) = p(y|u) p(u) \rightarrow -\log(p(y, u)) = -\log(p(y|u)) - \log(p(u)). \quad (5.12)$$

El término “penalizado” de la ecuación (5.8) hace referencia al término $\|\Lambda_\theta u\|_2^2$ en la ecuación (5.9) que implica una probabilidad a priori sobre b , que nos permite marginalizarlo. Bates et al. [5] define la misma probabilidad sobre u , pero nosotros decidimos incluir Λ_θ porque esto nos va a permitir obtener iteraciones para el algoritmo de ADMM con pasos analíticos. Si bien, tanto u como b responden a distribuciones normales, no necesariamente estaríamos regularizando sobre la misma distribución, sin embargo, dado que definimos Λ_θ como una identidad por una constante, la regularización es la misma, solo que pondera combinaciones lineales de u .

Finalmente, si expresamos la ecuación (5.9) en notación matricial de bloques, tenemos que:

$$r^2(\theta, \beta, u) = \left\| \begin{bmatrix} W^{1/2}(y_{obs} - Z\mu_b) \\ 0 \end{bmatrix} - \begin{bmatrix} W^{1/2}Z\Lambda_\theta & W^{1/2}A \\ \Lambda_\theta & 0 \end{bmatrix} \begin{bmatrix} u \\ \beta \end{bmatrix} \right\|^2 \quad (5.13)$$

La notación en bloques nos permite mostrar que el problema de cuadrados mínimos penalizado puede ser visto como un problema de cuadrados mínimos ordinario dado un vector extendido implicando que el valor de minimización, $\begin{bmatrix} \mu_{\mathcal{U}|y=y_{obs}} \\ \hat{\beta}_\theta \end{bmatrix}$, satisface las ecuaciones normales

$$\begin{bmatrix} \Lambda_\theta^t Z^t W (y_{obs} - o) \\ A^t W (y_{obs} - o) \end{bmatrix} = \begin{bmatrix} \Lambda_\theta^t Z^t W Z \Lambda_\theta + \Lambda_\theta^2 & \Lambda_\theta^t Z^t W A \\ A^t W Z \Lambda_\theta & X^t W A \end{bmatrix} \begin{bmatrix} \mu_{\mathcal{U}|y=y_{obs}} \\ \hat{\beta}_\theta \end{bmatrix} \quad (5.14)$$

Luego, podemos descomponer el factor izquierdo de la multiplicación mediante una factorización de Cholesky:

$$\begin{bmatrix} \Lambda_\theta^t Z^t W Z \Lambda_\theta + \Lambda_\theta^2 & \Lambda_\theta^t Z^t W A \\ A^t W Z \Lambda_\theta & A^t W A \end{bmatrix} = \begin{bmatrix} L_\theta & 0 \\ R_{ZA}^t & R_A^t \end{bmatrix} \begin{bmatrix} L_\theta^t & R_{ZA} \\ 0 & R_A \end{bmatrix} \quad (5.15)$$

lo que nos permite reescribir nuestra ecuación de cuadrados mínimos penalizados como:

$$r^2(\theta, \beta, u) = r^2(\theta, \hat{\beta}_\theta, \mu_{\mathcal{U}|y=y_{obs}}) + \left\| L_\theta^t (u - \mu_{\mathcal{U}|y=y_{obs}}) + R_{ZA}(\beta - \hat{\beta}_\theta) \right\|^2 + \left\| R_A(\beta - \hat{\beta}_\theta) \right\|^2 \quad (5.16)$$

A partir de la ecuación (5.13), también podemos expresar el problema como el logaritmo de una probabilidad a ser maximizada, dada por:

$$\operatorname{argmax} \mathcal{L}(\theta, \beta, \hat{\beta}_\theta, \mu_{\mathcal{U}|y=y_{obs}}) = \log f_y(y_{obs}) \quad (5.17)$$

donde

$$f_y(y_{obs}) = \int f_{y,\mathcal{U}}(y_{obs}, u) du \quad (5.18)$$

donde marginalizamos u dado que no nos interesa observarla, y f se define como

$$f_y(y_{obs}) = \frac{|W|^{1/2}}{(2\pi\sigma^2)^{\frac{(n+q)}{2}}} \exp \left[\frac{-r^2(\theta, \hat{\beta}_\theta, \mu_{\mathcal{U}|y=y_{obs}}) - \left\| R_A(\beta - \hat{\beta}_\theta) \right\|^2}{2\sigma^2} \right] \int \exp \left[\frac{-\left\| L_\theta^t (u - \mu_{\mathcal{U}|y=y_{obs}}) + R_{ZA}(\beta - \hat{\beta}_\theta) \right\|^2}{2\sigma^2} \right] du \quad (5.19)$$

a la cual, como muestra Bates et al. [5], podemos aplicarle el cambio de variable

$$v = L_\theta^t (u - \mu_{\mathcal{U}|y=y_{obs}}) + R_{ZA}(\beta - \hat{\beta}_\theta) \quad (5.20)$$

para obtener

$$f_y(y_{obs}) = \frac{|W|^{1/2}}{(2\pi\sigma^2)^{\frac{(n+q)}{2}}} \exp \left[\frac{-r^2(\theta, \hat{\beta}_\theta, \mu_{\mathcal{U}|y=y_{obs}}) - \|R_A(\beta - \hat{\beta}_\theta)\|^2}{2\sigma^2} \right] \int \exp \left[\frac{-\|v\|^2}{2\sigma^2} \right] |L_\theta|^{-1} du \quad (5.21)$$

Esto nos permite resolver la integral definida sobre u en la ecuación (5.18) con facilidad y obtener

$$\frac{|W|^{1/2}|L_\theta|^{-1}}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left[\frac{-r^2(\theta, \hat{\beta}_\theta, \mu_{\mathcal{U}|y=y_{obs}}) - \|R_A(\beta - \hat{\beta}_\theta)\|^2}{2\sigma^2} \right] \quad (5.22)$$

Dado que nuestra intención es resolver una minimización de los parámetros, podemos aplicar logaritmos para despejar simplificar los cálculos y transformar todo en una minimización. En particular tomamos el doble de la verosimilitud del logaritmo, que puede escribirse de la siguiente manera

$$-2\mathcal{L}(\theta, \beta, \hat{\beta}_\theta, \mu_{\mathcal{U}|y=y_{obs}}) = \log \frac{|L_\theta|^2}{|W|} + n \log(2\pi\sigma^2) + \frac{r^2(\theta, \hat{\beta}_\theta, \mu_{\mathcal{U}|y=y_{obs}})}{\sigma^2} + \frac{\|R_A(\beta - \hat{\beta}_\theta)\|^2}{\sigma^2} \quad (5.23)$$

Entonces, el problema que buscamos resolver puede escribirse como

$$\operatorname{argmin}_{\theta, \beta, \hat{\beta}_\theta, \mu_{\mathcal{U}|y=y_{obs}}} \log \frac{|L_\theta|^2}{|W|} + n \log(2\pi\sigma^2) + \frac{r^2(\theta, \hat{\beta}_\theta, \mu_{\mathcal{U}|y=y_{obs}})}{\sigma^2} + \frac{\|R_A(\beta - \hat{\beta}_\theta)\|^2}{\sigma^2} \quad (5.24)$$

donde $\theta > \theta_{min}$

Dado que vamos a tomar parámetros relacionados a la varianza de los sujetos, definiremos θ como una constante por sujeto y Λ_θ una matriz identidad. Debido a esto, el término de regularización propuesto en la ecuación (5.9) agrega también un Λ_θ que actúa sobre u en la minimización. Por eso debemos introducir la desigualdad en la condición de θ dado que sin esta restricción, la varianza del sistema tiende a cero.

5.1. Resolución mediante ADMM

Presentaremos a partir de esta sección nuestra contribución metodológica, mediante la implementación de una solución para el problema de minimización visto en la ecuación (5.24) mediante el método de los multiplicadores alternados. Como vimos en la sección 3.4, el método de los multiplicadores alternados, que a partir de ahora llamaremos simplemente ADMM, es una adaptación del algoritmo de descomposición dual. Este método permite, dada una función $h(y)$ tal que sus variables puedan ser desacopladas de forma que $h(y) = f(j(y)) + g(k(y))$, obtener un algoritmo iterativo y alternado entre las minimizaciones de $j(y)$ y $k(y)$. Si aplicamos esta idea de descomposición de variables al problema

de minimización descripto en la ecuación (5.24), tenemos que:

$$\underset{\theta_1, \theta_2, \hat{\beta}_{\theta_1}, \hat{\beta}_{\theta_2}, \beta, \mu_{\mathcal{U}|y=y_{obs}}}{\operatorname{argmin}} \quad f_1(\theta_1) + n \log(2\pi\sigma^2) + f_2(\theta_2, \hat{\beta}_{\theta_1}, \mu_{\mathcal{U}|y=y_{obs}}) + f_3(\beta, \hat{\beta}_{\theta_2}) \quad (5.25)$$

donde $\theta_1 > \theta_{min}$, $\theta_2 > \theta_{min}$ y $S_\beta x = \hat{\beta}_{\theta_2}$

Entonces nuestra función objetivo tiene la forma

$$\underset{\theta_1, \theta_2, \hat{\beta}_{\theta_1}, \hat{\beta}_{\theta_2}, \beta, \mu_{\mathcal{U}|y=y_{obs}}}{\operatorname{argmin}} \quad f_1(\theta_1) + n \log(2\pi\sigma^2) + f_2(\theta_2, \hat{\beta}_{\theta_1}, \mu_{\mathcal{U}|y=y_{obs}}) + f_3(\beta, \hat{\beta}_{\theta_2}) + (1 - \mathbb{1}_{\theta_{min}}(\theta_*)) \quad (5.26)$$

donde $\theta_1 = \theta_2 = \theta_*$ y $S_\beta x = \hat{\beta}_{\theta_2}$

donde $\mathbb{1}_{\theta_{min}}(\theta)$ es la función indicadora de θ tal que

$$\mathbb{1}_{\theta_{min}}(\theta) = \begin{cases} 1 & \theta > \theta_{min} \\ 0 & \text{otro caso} \end{cases} \quad (5.27)$$

Esta función define una proyección a partir del conjunto $\mathcal{C} = \{x | x > \theta_{min}\}$ que genera, dada por:

$$\Pi_{\mathcal{C}}(x) = \begin{cases} x & x > \theta_{min} \\ \theta_{min} & x = \theta_{min} \end{cases} \quad (5.28)$$

Podemos entonces agregar los Lagrangianos necesarios para las variables desacopladas

$$\begin{aligned} L(\theta_1, \theta_2, \hat{\beta}_{\theta_1}, \mu_{\mathcal{U}|y=y_{obs}}, \hat{\beta}_{\theta_2}, \alpha) &= f_1(\theta_1) + f_2(\theta_2, \hat{\beta}_{\theta_1}, \mu_{\mathcal{U}|y=y_{obs}}) + f_3(\beta, \hat{\beta}_{\theta_2}) \\ &\quad + \alpha_\theta^T (\theta_1 - \theta_2) + (\tau/2) \|\theta_1 - \theta_2\|_2^2 \\ &\quad + \alpha_\beta^T (\hat{\beta}_{\theta_1} - \hat{\beta}_{\theta_2}) + (\tau/2) \|\hat{\beta}_{\theta_1} - \hat{\beta}_{\theta_2}\|_2^2 + 1 - \mathbb{1}_{\theta_{min}}(\theta_*) \end{aligned} \quad (5.29)$$

Dado que nuestra idea es minimizar cada variable de forma alternada, vamos a necesitar derivar cada una de las f 's que definen nuestro problema. En particular, en los casos de f_2 y f_3 , al estar compuestas por más de una variables objetivo, necesitamos realizar una minimización conjunta de todas ellas. Proponemos entonces, una modificación para el caso puntual de f_2 que nos permite simplificar el despeje necesario para alcanzar dicho mínimo. Definimos entonces una variable x tal que

$$x \mapsto x(\mu_{\mathcal{U}|y=y_{obs}}, \hat{\beta}_{\theta_1}) = \begin{bmatrix} \mu_{\mathcal{U}|y=y_{obs}} \\ \hat{\beta}_{\theta_1} \end{bmatrix} \quad (5.30)$$

y dos proyectores de forma tal que nos permitan tomar

$$S_\mu x(\mu_{\mathcal{U}|y=y_{obs}}, \hat{\beta}_{\theta_1}) = \mu_{\mathcal{U}|y=y_{obs}} \quad \text{y} \quad S_\beta x(\mu_{\mathcal{U}|y=y_{obs}}, \hat{\beta}_{\theta_1}) = \hat{\beta}_{\theta_1} \quad (5.31)$$

A partir de esta modificación, podemos reescribir la ecuación (5.29) como

$$\begin{aligned} L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \alpha) &= f_1(\theta_1) + f_2(\theta_2, x) + f_3(\beta, \hat{\beta}_{\theta_2}) \\ &\quad + \alpha_\theta^T (\theta_1 - \theta_2) + (\tau/2) \|\theta_1 - \theta_2\|_2^2 \\ &\quad + \alpha_\beta^T (S_\beta x - \hat{\beta}_{\theta_2}) + (\tau/2) \|S_\beta x - \hat{\beta}_{\theta_2}\|_2^2 \\ &\quad + 1 - \mathbb{1}_{\theta_{min}}(\theta_*) \end{aligned} \quad (5.32)$$

Y finalmente, aplicando la forma escalada vista en la ecuación (3.40), si tomamos $r_\theta = \theta_1 - \theta_2$ obtener

$$\alpha^t r + (\tau/2)\|r\|_2^2 = (\tau/2)\|r + (1/\tau)\alpha\|_2^2 - (1/2\tau)\|\alpha\|_2^2 \quad (5.33)$$

$$(\tau/2)\|r + z_\theta\|_2^2 - (\tau/2)\|z_\theta\|_2^2 \quad (5.34)$$

donde $z_\theta = (1/\tau)\alpha_\theta$. Si realizamos el mismo procedimiento sobre $\hat{\beta}_\theta$, podemos reescribir nuestro problema en su versión final como

$$\begin{aligned} L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, z) &= f_1(\theta_1) + f_2(\theta_2, x) + f_3(\beta, \hat{\beta}_{\theta_2}) \\ &+ \frac{\tau_\theta}{2}\|\theta_1 - \theta_2 + z_\theta\|_2^2 + \frac{\tau_\beta}{2}\|S_\beta x - \hat{\beta}_{\theta_2} + z_\beta\|_2^2 \\ &+ 1 - \mathbb{1}_{\theta_{min}}(\theta_*) \end{aligned} \quad (5.35)$$

De esta manera, el problema de resolver un modelo lineal mixto, queda separado en funciones con variables desacopladas, las cuales podemos minimizar y luego iterar como propone el algoritmo de ADMM en (3.41-3.43). Es nuestro caso, dichas iteraciones están definidas como

$$\theta_1^{x+1} := \underset{\Pi_C(\theta_1)}{\operatorname{argmin}} L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, z^k) \quad (5.36)$$

$$\hat{\beta}_{\theta_2}^{x+1} := \underset{\{\hat{\beta}_{\theta_2}, \beta\}}{\operatorname{argmin}} L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, z^k) \quad (5.37)$$

$$\beta^{x+1} := \underset{\{\hat{\beta}_{\theta_2}, \beta\}}{\operatorname{argmin}} L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, z^k) \quad (5.38)$$

$$x^{k+1} := \underset{\{\Pi_C(\theta_2), x\}}{\operatorname{argmin}} L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, z^k) \quad (5.39)$$

$$\theta_2^{x+1} := \underset{\{\Pi_C(\theta_2), x\}}{\operatorname{argmin}} L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, z^k) \quad (5.40)$$

$$z_\theta^{k+1} := z_\theta^k + \theta_1^{k+1} - \theta_2^{k+1} \quad (5.41)$$

$$z_\beta^{k+1} := z_\beta^k + \hat{\beta}_{\theta_1}^{k+1} - \hat{\beta}_{\theta_2}^{k+1} \quad (5.42)$$

Para realizar cada una de las minimizaciones necesarias, vamos a derivar cada una de las funciones respecto a las variables involucradas, con el fin de obtener soluciones analíticas para cada función. Primero buscamos un mínimo para f_1 , derivando $L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, z)$. Como vimos en la sección 3.4, el algoritmo de ADMM requiere que las funciones de la descomposición sean convexas, por lo que derivaremos respecto de la precisión, $\frac{1}{\theta_1}$. Por lo tanto, la ecuación (5.36) queda reescrita como

$$\underset{\Pi_C(\theta_1)}{\operatorname{argmin}} L(\theta_1, \theta_2, \hat{\beta}_{\theta_1}, \hat{\beta}_{\theta_2}, z) = \left\{ \theta_1 : \frac{\partial \log \frac{|L_{\theta_1}|^2}{|W|} + \frac{\tau_\theta}{2} \left\| \frac{1}{\theta_1} - \frac{1}{\theta_2} + z_\theta \right\|_2^2}{\partial \theta_1} = 0 \right\} \quad (5.43)$$

para, luego de derivar e igualar a cero (ver apéndice), obtener

$$-2k\theta_1^2 + \tau_\theta\theta_1\left(\frac{1}{\theta_2} - z_\theta\right) - \tau_\theta = 0 \quad (5.44)$$

En segundo lugar, vamos a probar que podemos minimizar las variables de f_2 de manera conjunta hasta alcanzar un mínimo teórico. Dado que f_2 es una función cuadrática, vamos a probar, de manera analítica, que existe un punto donde las dos derivadas se hacen cero, mediante sustitución. Tenemos, entonces, un sistema de ecuaciones compuesto por los gradientes de f_2 :

$$\operatorname{argmin}_{\{\Pi_{\mathcal{C}(\theta_2), x}\}} L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z) = \left\{ x, \theta_2 : \begin{cases} \frac{\partial f_2(\theta_2, x) + \frac{\tau_\beta}{2} \|S_\beta x - \hat{\beta}_{\theta_2} + z_\beta\|_2^2}{\partial x} = 0 \\ \frac{\partial f_2(\theta_2, x) + \frac{\tau_\theta}{2} \|\theta_1 - \theta_2 + z_\theta\|_2^2}{\partial \theta_2} = 0 \end{cases} \right\} \quad (5.45)$$

Veamos que este sistema de ecuaciones tiene solución analítica. Para resolver el gradiente de x tenemos

$$L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z) = \frac{r^2(\theta, x)}{\sigma^2} + \frac{\tau_\beta}{2} \|S_\beta x - \hat{\beta}_{\theta_2} + z_\beta\|_2^2 \quad (5.46)$$

donde si derivamos respecto de x e igualamos a cero, obtenemos

$$x = \begin{bmatrix} \Lambda_{\theta_2}^t Z^t W Z \Lambda_{\theta_2} + \Lambda_{\theta_2}^2 & \Lambda_{\theta_2}^t Z^t W A \\ A^t W Z \Lambda_{\theta_2} & A^t W A + \sigma^2 \tau_\beta I_\beta \end{bmatrix}^{-1} \begin{bmatrix} \Lambda_{\theta_2}^t Z^t W (y_{obs} - o) \\ \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o) \end{bmatrix} \quad (5.47)$$

Luego, podemos resolver el gradiente de θ_2 . En este caso también necesitamos derivar respecto a $\frac{1}{\theta_2}$ asegurando que la función minimizada sea convexa. Luego, reemplazando x obtenemos una solución para nuestro sistema de ecuaciones:

$$\frac{x^t \begin{bmatrix} -2\theta_2^{-3} (Z^t W Z + I) & -\theta_2^{-2} Z^t W A \\ -\theta_2^{-2} A^t W Z & 0 \end{bmatrix} x - 2x^t \begin{bmatrix} -\theta_2^{-2} W^{1/2} Z & 0 \\ -\theta_2^{-2} & 0 \end{bmatrix}^t \begin{bmatrix} W^{1/2} (y_{obs} - o) \\ 0 \end{bmatrix}}{\sigma^2} - \tau_\theta \theta_2^{-3} + \tau_\theta \theta_2^{-2} (\theta_1^{-1} + z_\theta) = 0 \quad (5.48)$$

donde terminamos resolviendo una función cuántica para la cual necesitamos calcular sus raíces (ver apéndice). Por último, utilizamos la misma lógica para la solución conjunta de f_3 . Por lo tanto, podemos tomar la derivada conjunta e igualar a cero para tratar de buscar la expresión que define su mínimo.

$$\operatorname{argmin}_{\{\beta, \hat{\beta}_{\theta_2}\}} L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z) = \left\{ \beta, \hat{\beta}_{\theta_2} : \begin{cases} \frac{\partial f_3(\beta, \hat{\beta}_{\theta_2}) + \frac{\tau}{2} \|S_\beta x - \hat{\beta}_{\theta_2} + z_\beta\|_2^2}{\partial \hat{\beta}_{\theta_2}} = 0 \\ \frac{\partial f_3(\beta, \hat{\beta}_{\theta_2})}{\partial \beta} = 0 \end{cases} \right\} \quad (5.49)$$

Luego

$$\frac{\partial L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z)}{\partial \beta} = 0 \quad (5.50)$$

$$\beta = \hat{\beta}_{\theta_2} \quad (5.51)$$

Y haciendo lo mismo respecto a $\hat{\beta}_{\theta_2}$:

$$L(\theta_1, \theta_2, \hat{\beta}_{\theta_1}, \hat{\beta}_{\theta_2}, z) = \frac{\|R_A(\beta - \hat{\beta}_{\theta_2})\|^2}{\sigma^2} + \frac{\tau}{2} \|\hat{\beta}_{\theta_1} - \hat{\beta}_{\theta_2} + z_\beta\|_2^2 \quad (5.52)$$

$$\frac{-2R_A^t R_A \beta + 2R_A^t R_A \hat{\beta}_{\theta 2}}{\sigma^2} - \tau \hat{\beta}_{\theta 1} - \tau z_\beta + \tau \hat{\beta}_{\theta 2} = 0 \quad (5.53)$$

Donde podemos remplazar el resultado de la ecuación (5.51) y despejar:

$$\frac{-2R_A^t R_A \beta + 2R_A^t R_A \beta}{\sigma^2} - \tau \hat{\beta}_{\theta 1} - \tau z_\beta + \tau \beta = 0 \quad (5.54)$$

$$\hat{\beta}_{\theta 2} = \beta = z_\beta - \hat{\beta}_{\theta 1} \quad (5.55)$$

Finalmente, una vez calculadas las soluciones analíticas para cada uno de los pasos de ADMM definidos en (5.36-5.42), obtenemos un algoritmo definido de forma iterativa que nos permite resolver modelos lineales mixtos. Sin embargo, si bien es cierto que algunos pasos pueden realizarse en paralelo con la ventaja en tiempo de cómputo que esto conlleva, todavía no resolvimos el problema que implicaba tener que cargar en memoria enormes matrices de diseño. Analizaremos una solución a este problema a continuación.

5.2. Independencia entre sujetos

Una de las ventajas de expresar el problema de suma de cuadrados como densidades de probabilidad que realizan Bates et al. [5] es que nos permite, suponiendo que los datos son independientes para cada sujeto, separar las operaciones definidas en (5.36-5.42) de forma que podemos realizar cada uno de estos cálculos de manera independiente. Esta separación combinada con ADMM nos permite obtener una solución distribuida y escalable para nuestro modelo. De esta manera resolvemos la problemática dada por la necesidad de cargar todas las matrices de diseño propias de un modelo lineal mixto. Sea

$$\Gamma \sim \mathcal{N}(0, \sigma^2 I) \quad (5.56)$$

$$y_{\mathcal{U}} - A\beta + Zb \sim W^{-1/2}\Gamma \quad (5.57)$$

donde, si los sujetos son mediciones independientes idénticamente distribuidas, W no tiene términos cruzados entre sujetos distintos. Además, dado que Z no tiene términos cruzados entre sujetos, entonces, podemos expresar la densidad de f_y en ecuación (5.19) como:

$$f_y(y_{obs}) = f_{y_1, \dots, y_n}(y_{obs_1}, \dots, y_{obs_n}) \quad (5.58)$$

$$f_y(y_{obs}) = f_{y_1}(y_{obs_1}) \times \dots \times f_{y_n}(y_{obs_n}) \quad (5.59)$$

$$\prod_{i=1}^n \frac{|W_i|^{1/2} |L_{\theta_i}|^{-1}}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp \left[\frac{-r^2(\theta_i, \hat{\beta}_{\theta_i}, \mu_{\mathcal{U}_i|y_i=y_{obs_i}}) - \|R_A(\beta - \hat{\beta}_{\theta_i})\|^2}{2\sigma^2} \right] \quad (5.60)$$

La cual, como vimos en la ecuación (5.17), podemos volver a expresar como una minimización mediante la aplicación de logaritmos a ambos lados:

$$\mathcal{L}(\theta_i, \beta, \hat{\beta}_{\theta_i}, \mu_{\mathcal{U}_i|y_i=y_{obs_i}}) = \log f_{y_1, \dots, y_n}(y_{obs_1}, \dots, y_{obs_n}) \quad (5.61)$$

y obtener

$$\sum_{i=1}^n \left(\log \frac{|L_{\theta_i}|^2}{W} + n \log(2\pi\sigma^2) + \frac{r^2(\theta_i, \hat{\beta}_{\theta_i}, \mu_{\mathcal{U}_i|y_i=y_{obs_i}})}{\sigma^2} + \frac{\|R_A(\beta - \hat{\beta}_{\theta_i})\|^2}{\sigma^2} \right) - 2\mathcal{L}(\theta, \beta, \hat{\beta}_{\theta}, \mu_{\mathcal{U}|y=y_{obs}}) = \quad (5.62)$$

Luego, puedo volver a aplicar los mismos procedimientos utilizados con anterioridad. Esto nos permite volver a definir la misma variable x definida en la ecuación (5.30), y desacoplar haciendo

$$\operatorname{argmin}_{\{\theta_{1_1}, \theta_{2_1}, x_1, \beta_{2_1}, \beta_1\}, \dots, \{\theta_{1_n}, \theta_{2_n}, x_n, \beta_{2_n}, \beta_n\}} \sum_{i=1}^n f_1(\theta_{1_i}) + n \log(2\pi\sigma^2) + f_2(\theta_{2_i}, x_i) + f_3(\beta_i, \hat{\beta}_{\theta_{2_i}}) \quad (5.63)$$

$$\text{donde } \theta_{1_i} = \theta_{2_i} = \theta_*, S_{\beta_i} x_i = \hat{\beta}_{\theta_{2_i}} \text{ y } \beta_1 = \dots = \beta_n = \beta_*$$

donde la diferencia con el caso visto anteriormente está en que también es necesario desacoplar β_1, β_n . Defino entonces las L_i de cada sujeto que voy a minimizar

$$L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z, v) = \sum_{n=1}^n f_1(\theta_{1_i}) + f_2(\theta_{2_i}, x_i) + f_3(\beta_i, \hat{\beta}_{\theta_{2_i}}) + \frac{\tau_\theta}{2} \|\theta_{1_i} - \theta_{2_i} + z_{\theta_i}\|_2^2 + \frac{\tau_\beta}{2} \|S_{\beta_i} x - \hat{\beta}_{\theta_{2_i}} + z_{\beta_i}\|_2^2 + \frac{\tau}{2} \|\beta_i - \beta_* + v_i\|_2^2 \quad (5.64)$$

Luego, puedo definir las ecuaciones de ADMM para cada uno de los i sujetos como:

$$\theta_{1_i}^{x+1} := \operatorname{argmin}_{\theta_{1_i}} L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, \beta_*, z^k, v^k) \quad (5.65)$$

$$\beta_i^{x+1} := \operatorname{argmin}_{\{\hat{\beta}_{\theta_2}, \beta\}} L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, \beta_*, z^k, v^k) \quad (5.66)$$

$$\hat{\beta}_{\theta_{2_i}}^{x+1} := \operatorname{argmin}_{\{\hat{\beta}_{\theta_2}, \beta\}} L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, \beta_*, z^k, v^k) \quad (5.67)$$

$$x_i^{x+1} := \operatorname{argmin}_{\{\theta_2, x\}} L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, \beta_*, z^k, v^k) \quad (5.68)$$

$$\theta_{2_i}^{x+1} := \operatorname{argmin}_{\{\theta_2, x\}} L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, \beta_*, z^k, v^k) \quad (5.69)$$

$$\beta_*^{k+1} := \frac{1}{n} \sum_{i=1}^n \beta_i^{k+1} - v_i^k \quad (5.70)$$

$$z_{\theta_i}^{k+1} := z_{\theta_i}^k + \theta_{1_i}^{k+1} - \theta_{2_i}^{k+1} \quad (5.71)$$

$$z_{\beta_i}^{k+1} := z_{\beta_i}^k + \hat{\beta}_{\theta_{1_i}}^{k+1} - \hat{\beta}_{\theta_{2_i}}^{k+1} \quad (5.72)$$

$$v_i^{k+1} := v_i^k + \beta_i^{k+1} - \beta_*^{k+1} \quad (5.73)$$

A diferencia del resultado obtenido en la ecuación (5.51), esta vez la derivada de β_i incluye el lagrangiano. Por lo tanto, ya no es posible realizar la misma simplificación sobre f_3 y la ecuación (5.49) ahora tiene la forma

$$\operatorname{argmin}_{\{\beta, \hat{\beta}_{\theta_2}\}} L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, \beta_*, z, v) = \left\{ \begin{array}{l} \beta_i, \hat{\beta}_{\theta_{2_i}} : \\ \frac{\partial f_3(\beta_i, \hat{\beta}_{\theta_{2_i}}) + \frac{\tau_\beta}{2} \|S_{\beta_i} x - \hat{\beta}_{\theta_{2_i}} + z_{\beta_i}\|_2^2}{\partial \hat{\beta}_{\theta_{2_i}}} = 0 \\ \frac{\partial f_3(\beta_i, \hat{\beta}_{\theta_{2_i}}) + \frac{\tau_{\beta_*}}{2} \|\beta_i - \beta_* + v_i\|_2^2}{\partial \beta_i} = 0 \end{array} \right\} \quad (5.74)$$

El resto de cálculos se realiza de forma análoga a los vistos en el caso general, con la diferencia de que ahora es necesario considerar variables $\theta_{1_i}, \theta_{2_i}, \hat{\beta}_{\theta_{1_i}}, \hat{\beta}_{\theta_{2_i}}$ para cada sujeto.

5.3. Regularización L1

Una de las ventajas del método de ADMM, es que permite la implementación de regularizaciones de forma sencilla. Un caso importante es el de las regularizaciones L1 que ya presentamos en la sección 3.2.1. Dado que aprovecharemos la independencia entre sujetos vista en 5.2, tomando la ecuación (5.64) tengo que:

$$L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z, v) = \sum_{n=1}^n f_1(\theta_{1_i}) + f_2(\theta_{2_i}, x_i) + f_3(\beta_i, \hat{\beta}_{\theta_{2_i}}) \quad (5.75)$$

$$+ \frac{\tau_\theta}{2} \|\theta_{1_i} - \theta_{2_i} + z_{\theta_i}\|_2^2 + \frac{\tau_\beta}{2} \|S_{\beta_i} x - \hat{\beta}_{\theta_{2_i}} + z_{\beta_i}\|_2^2 + \frac{\tau}{2} \|\beta_i - \beta_* + v_i\|_2^2$$

donde puedo agregar los términos regularizados. En nuestro caso particular, aplicaremos una regularización sobre el término β_* global a todos los sujetos.

$$L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z, v) = \sum_{n=1}^n f_1(\theta_{1_i}) + f_2(\theta_{2_i}, x_i) + f_3(\beta_i, \hat{\beta}_{\theta_{2_i}}) \quad (5.76)$$

$$+ \frac{\tau_\theta}{2} \|\theta_{1_i} - \theta_{2_i} + z_{\theta_i}\|_2^2 + \frac{\tau_\beta}{2} \|S_{\beta_i} x - \hat{\beta}_{\theta_{2_i}} + z_{\beta_i}\|_2^2 + \frac{\tau}{2} \|\beta_i - \beta_* + v_i\|_2^2$$

$$+ \lambda \|\beta_*\|_1$$

luego, las iteraciones de ADMM se definen como:

$$\theta_{1_i}^{x+1} := \operatorname{argmin}_{\theta_{1_i}} L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, \beta_*, z^k, v^k) \quad (5.77)$$

$$\beta_i^{x+1} := \operatorname{argmin}_{\{\hat{\beta}_{\theta_2}, \beta\}} L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, \beta_*, z^k, v^k) \quad (5.78)$$

$$\hat{\beta}_{\theta_{2_i}}^{x+1} := \operatorname{argmin}_{\{\hat{\beta}_{\theta_2}, \beta\}} L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, \beta_*, z^k, v^k) \quad (5.79)$$

$$x_i^{x+1} := \operatorname{argmin}_{\{\theta_2, x\}} L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, \beta_*, z^k, v^k) \quad (5.80)$$

$$\theta_{2_i}^{x+1} := \operatorname{argmin}_{\{\theta_2, x\}} L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, \beta, \beta_*, z^k, v^k) \quad (5.81)$$

$$\beta_*^{k+1} := \frac{1}{n} \sum_{i=1}^n S_{\lambda/\tau N} (\beta_i^{k+1} - v_i^k) \quad (5.82)$$

$$z_{\theta_i}^{k+1} := z_{\theta_i}^k + \theta_{1_i}^{k+1} - \theta_{2_i}^{k+1} \quad (5.83)$$

$$z_{\beta_i}^{k+1} := z_{\beta_i}^k + \hat{\beta}_{\theta_{1_i}}^{k+1} - \hat{\beta}_{\theta_{2_i}}^{k+1} \quad (5.84)$$

$$v_i^{k+1} := v_i^k + \beta_i^{k+1} - \beta_*^{k+1} \quad (5.85)$$

donde S es el operador de soft-thresholding definido como:

$$S_k(a) = \begin{cases} a - k & a > k \\ 0 & |a| \leq k \\ a + k & a < -k \end{cases} \quad (5.86)$$

6. EXPERIMENTOS Y RESULTADOS

En los capítulos previos presentamos el marco teórico y el estado del arte en modelos lineales. Mas específicamente, en el capítulo 3 realizamos una introducción a los modelos lineales mixtos y al método de los multiplicadores alternados que utilizamos. En el capítulo 4 realizamos una breve introducción al estado del arte en el estudio de la relación entre estructura y función. Y finalmente en el capítulo 5 presentamos nuestro método para resolver modelos lineales mixtos mediante el algoritmo de ADMM. En este capítulo presentamos distintos experimentos en los que mostramos la aplicabilidad de nuestro método y sus resultados. Nuestros experimentos fueron realizados tanto sobre datos sintéticos como sobre sujetos extraídos de la base de datos Human Connectome Project [32].

En esta tesis presentamos un método para la resolución de modelos lineales mixtos, esto es, los modelos de la forma:

$$Y = A\beta + Z\mathcal{B} \quad (6.1)$$

donde $Y \in R^{s*v}$ representa la activación funcional del cerebro a través de todos los sujetos, siendo s el número de sujetos y v el número de vértices. $\beta \in R^r$ es un vector que representa los efectos fijos, siendo r el número de regiones. $\mathcal{B} \in R^j$ es un vector que representa los efectos aleatorios, donde j es la cantidad de sujetos. $A \in R^{s*v \times r}$ es una matriz de conectividad que relaciona vértices con regiones mientras que $Z \in R^{s*v \times j}$ es una matriz que modela la relación entre los vértices y sujetos capturada por los efectos aleatorios, con el fin de separar la varianza de cada uno de éstos y obtener una activación base.

6.1. Experimentos en datos sintéticos

6.1.1. Datos

Para nuestro experimento en datos sintéticos, generamos un conjunto de datos simulando tres sujetos. Generamos una matriz de conectividad compuesta por veinticinco vectores, cada uno dividido en cinco regiones distintas. Planificamos la información de los sujetos de forma tal que la matriz sea simétrica, simulando la relación entre vértices y regiones que encontraríamos en nuestro experimento. También generamos un vector de efectos fijos $\beta \in R^r$ donde r es el numero de regiones, tal que

$$\beta \sim \mathcal{U}(0, 1) \quad (6.2)$$

Para simular los efectos aleatorios de casa sujeto generamos un conjunto de vectores $\{b\}_i, i \in [0, 2]$ a partir de las variables θ y Λ tal que

$$\Lambda \sim \mathcal{N}(0, \sigma) \text{ y } \theta \sim \mathcal{U}(0, 1) \quad (6.3)$$

donde $\sigma \sim \mathcal{U}(0, 0.01)$, calculando

$$b_i = \Lambda_i \cdot \theta_i \quad (6.4)$$

para cada uno de los sujetos. Nuestra matriz de diseño Z , fue diseñada para relacionar los vértices de la región analizada con su sujeto correspondiente. Combinamos toda esta

información de acuerdo al esquema de un modelo lineal mixto definido en la ecuación (6.1) y le agregamos un error $e \sim \mathcal{N}(0, \sigma)$, a partir del cual reconstruimos la activación funcional Y .

También generamos vectores con distribución uniforme para los valores de inicialización de nuestros sujetos, calculando entonces un β_0 y un θ_0 para cada uno de ellos.

6.1.2. Diseño experimental

Estudiamos la estabilidad de nuestro algoritmo realizando varias ejecuciones con distintos valores iniciales. En todas las ejecuciones los valores estimados de θ y β lograron converger a un valor, demostrando así la estabilidad del algoritmo (figura 6.1). Destacamos un fallo en la convergencia de θ_2 para uno de los sujetos, que rápidamente se estabiliza a una distancia constante de su valor objetivo, producto probablemente de un mínimo local. El resto de variables converge de manera correcta.

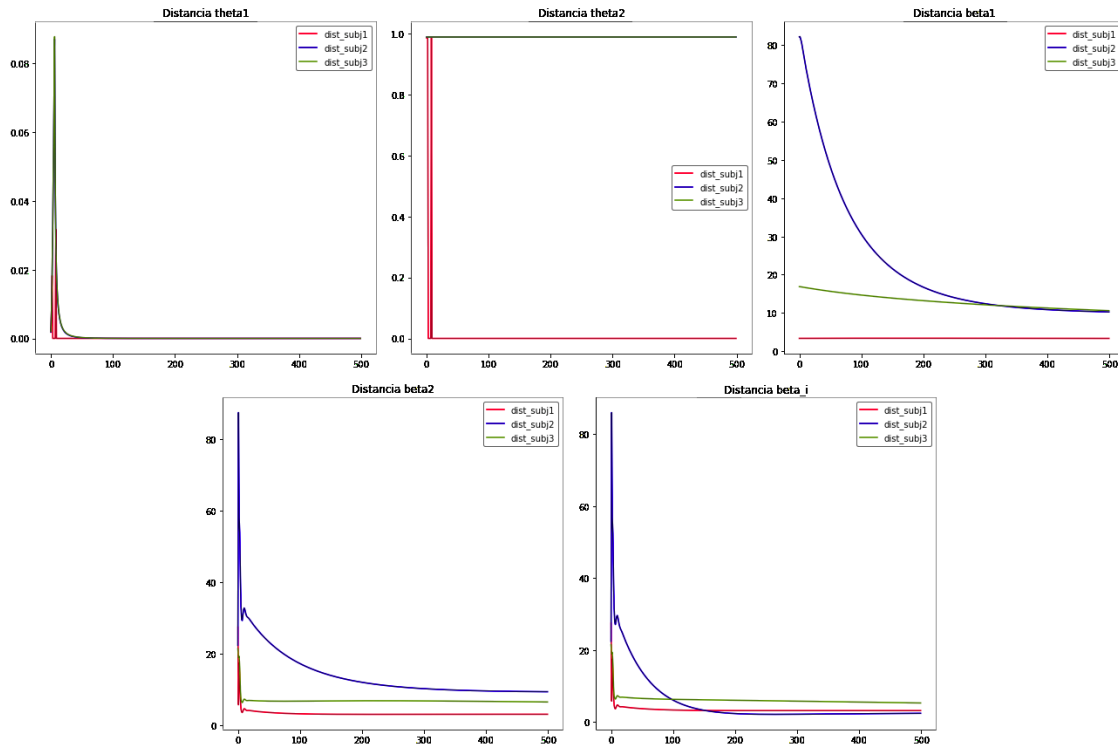


Fig. 6.1: Convergencia de las distintas variables a sus valores esperados en el análisis realizado sobre el ejemplo sintético. Salvo por una convergencia de θ_2 hacia un mínimo local que no es el buscado para uno de los sujetos, todas las variables presentan una convergencia estable.

6.2. Experimentos sobre sujetos del Human Connectome Project

6.2.1. Participantes

Descargamos datos anatómicos, funcionales y de difusión de 137 sujetos del Human Connectome Project (HCP) [32]. El conjunto de datos está compuesto por información

proveniente de 73 mujeres y 64 hombres, todos diestros entre 31 y 35 años. Una ventaja importante de utilizar sujetos de HCP es que todos poseen una superficie compuesta por 32 mil puntos representando su corteza cerebral. Los vértices de dichas superficies están coregistradas a través de todos los sujetos, esto quiere decir que, dado un vértice en el cerebro de un sujeto, podemos saber donde está en el cerebro de otro. Otra ventaja de este conjunto de datos es que cada sujeto posee el atlas anatómico de Desikan [11] y Destrieux [12] ya calculado. Respecto a la información funcional, se les pidió a los sujetos mover la mano derecha, y luego se computó un contraste entre la activación mientras la mano se movía y la señal fMRI promedio. Durante nuestros experimentos tanto la información funcional como la estructural fue manipulada sin hacer ninguna distinción durante el proceso de validación cruzada.

6.2.2. Diseño experimental

Utilizando la información de difusión simulamos el movimiento aleatorio de partículas de agua a través de la materia blanca mediante un proceso de Markov. Este proceso es conocido como tractografía. Repetimos este proceso desde cada vértice del cerebro de cada sujeto. Estimar la probabilidad de que dos vértices estén conectados calculando la cantidad de caminos aleatorios que conectaban ambos puntos sobre el total de caminos generados. Como consecuencia, obtuvimos matrices de conectividad de dimensión $v \times v$, donde v representa la cantidad de vértices en la superficie de cada sujeto. Para reducir el tamaño de las matrices de conectividad, calculamos la conectividad de cada vértice a las regiones anatómicas definidas en el atlas de Desikan (68 regiones) y el de Destrieux (148 regiones). La conectividad a una región se calculó tomando el promedio de la conectividad a los vértices que componían dicha región. Esto nos permitió obtener matrices de conectividad de dimensión $v \times 148$ y $v \times 68$ respectivamente.

En el caso de la información de fMRI, tenemos un vector por cada sujeto, compuesto por valores (zscores) que representan el nivel de activación para cada vértice de la corteza cerebral.

Tomamos, entonces, la información de cada sujeto y los concatenamos para producir un vector con todas las respuestas de activación (de N vértices de tamaño) y una matriz de conectividad conteniendo los vectores con la probabilidad de conexión a las distintas regiones.

Los modelos lineales mixtos tienen la necesidad de definir una matriz de diseño adicional, Z , con respecto a los modelos lineales ordinarios. En este caso $Z_{ij} = 1$ si el vértice i corresponde a cada una de las regiones de la superficie del sujeto j .

Tomando ventaja de la distribución del modelo hecha en ecuación (5.3) y suponiendo que la activación de cada vértice es independiente de los demás, modelamos la matriz de covarianza W como una matriz identidad.

Sobre la base del conocimiento de que solo áreas específicas del cerebro están activas durante la ejecución de diferentes tareas, sabemos que la representación de la actividad de los vértices en nuestra matriz funcional es mala. En consecuencia, consideramos que un gran porcentaje de los parámetros en el modelo no añadirán una contribución significativa al mismo. Basándonos en esta hipótesis, repetimos el mismo análisis pero utilizando una regularización L1 [31] en ambos modelos, intentando descartar los parámetros que no hacen una contribución significativa y esperando obtener una mejora en nuestros resultados.

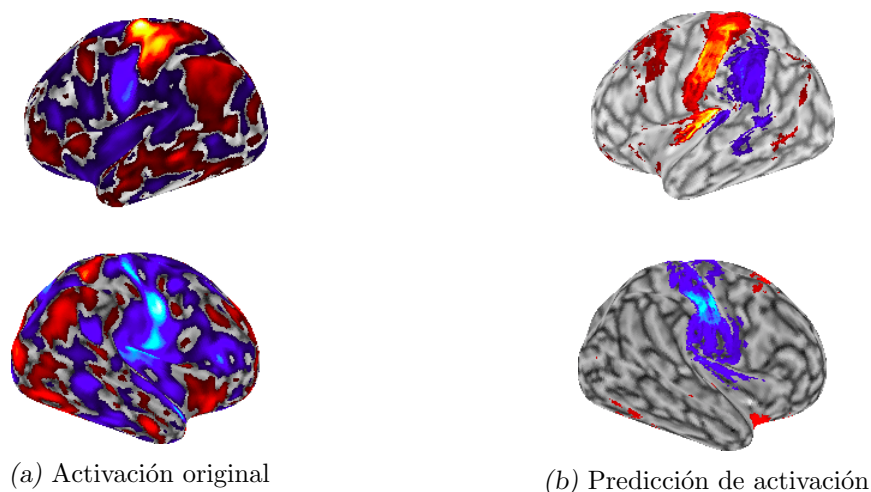


Fig. 6.2: Comparación entre la activación original de un sujeto en toda la corteza cerebral y la predicción realizada mediante nuestro modelo lineal mixto.

6.2.3. Prediciendo activaciones funcionales en toda la corteza cerebral

Habiendo definido los parámetros de nuestro modelo, procedemos a realizar una validación cruzada para ajustar un modelo lineal, y un modelo mixto entre la activación funcional en la corteza cerebral y la conectividad estructural. En particular, utilizamos una rutina de validación cruzada k -fold con $k = 10$, en la que dividimos el conjunto de datos en k subconjuntos consecutivos. Cada subconjunto se utiliza una vez como validación, mientras que los $k - 1$ subconjuntos restantes forman el conjunto de entrenamiento. En nuestro caso, reducimos el fold de entrenamiento seleccionando un subconjunto de 90 sujetos para reducir los tiempos de entrenamiento.

Es importante destacar que, al intentar utilizar el modelo de efectos aleatorios comúnmente utilizado en la literatura para predecir la activación funcional de toda la corteza cerebral, los algoritmos fallaron por quedarse sin memoria. La figura 6.2 ilustra qué, si bien nuestro método permite ajustar un modelo lineal mixto, el mismo realiza una predicción pobre de la activación funcional en toda la corteza cerebral.

6.2.4. Prediciendo activaciones funcionales en regiones específicas

Intentando mejorar la predicción obtenida, así como también buscando comparar los resultados de utilizar nuestro método contra el modelo lineal, decidimos restringir la predicción a regiones específicas del cerebro. Basándonos en Barch et al. [3] realizamos nuestro análisis sobre las tres regiones que presentan una mayor cantidad de vértices activados: la circunvolución postcentral, precentral y el surco precentral del hemisferio izquierdo. En este caso, estudiamos solo la conectividad de los vértices en dichas regiones, e intentamos predecir la activación funcional solo en sus vértices.

6.2.5. Resultados

Vamos a comparar dos modelos con diferentes propiedades, un modelo lineal y un modelo lineal mixto. Los modelos lineales, predicen el valor de una variable mediante el supuesto de que la relación entre ellas es lineal. Los modelos lineales mixtos son una

extensión de los modelos lineales que permiten tanto efectos fijos como efectos aleatorios. Como consecuencia, tenemos un número distinto de parámetros sobre los que ajustar. Mientras que en el modelo lineal tenemos un número de parámetros igual al número de regiones en el atlas seleccionado, en el modelo lineal mixto tenemos que ajustar la misma cantidad de parámetros que en el modelo lineal, los efectos fijos, sumados a los efectos aleatorios los cuales vienen dados por la inclusión de la matriz de diseño Z que agrega una cantidad de parámetros equivalente al número de vértices por el número de sujetos. Debido a esta situación, vamos a presentar tres estadísticos: El error cuadrático medio normalizado (NMSE, por sus siglas en inglés) para comparar el error en el ajuste y Chi-cuadrado (χ^2) para comparar el error en el ajuste con una mejor interpretación de las unidades. Además, utilizaremos el criterio de información de Akaike (AIC) para comparar los resultados teniendo en cuenta los grados de libertad y presentaremos la log-verosimilitud de la predicción realizada (LLH). Tomando ventaja del modelo intrasujeto de efectos aleatorios, también calcularemos estos estadísticos de forma iterativa sobre cada sujeto, promediando el resultado final. Las figuras 6.3-6.8 muestran la predicción de la activación funcional en ambos modelos. Además, presentan el peso promedio de las regiones en el modelado de las activación mediante el modelo lineal mixto e histogramas con la distribución de los valores de θ y μ_b calculados. Los resultados obtenidos para los estimadores NMSE, Chi-cuadrado, AIC y LLH se muestran en las tablas 6.1-6.4. En las tablas 6.9 y 6.10 mostramos el porcentaje de vértices activados en cada una de las regiones analizadas.

Tab. 6.1: Atlas de Desikan - Regresión sobre 90 sujetos.

Región	Estimador	LMM (prom. subj)	LMM	Modelo lineal
Postcentral izq.	NMSE	2.25 (0.12)	2.08 (0.09)	1.98 (0.03)
	χ^2	2194 (56)	161661 (4426)	164456 (712)
	AIC	12179 (598)	1107177 (54889)	1051113 (6231)
	LLH	-4161 (244)	-380027 (22550)	-356901 (2883)
Precentral izq.	NMSE	2.36 (0.05)	2.17 (0.03)	2.04 (0.07)
	χ^2	2225 (59)	157402 (1564)	158751 (5788)
	AIC	12774 (61)	1161747 (5461)	1103928 (53131)
	LLH	-4404 (28)	-402406 (2607)	-378402 (21671)
Supra Marginal izq.	NMSE	2.18 (0.14)	2.02 (0.12)	1.99 (0.1)
	χ^2	2151 (78)	159175 (5057)	156637 (5607)
	AIC	11601 (953)	1054490 (87016)	1051718 (85700)
	LLH	-3940 (371)	-359801 (34046)	-358415 (33375)

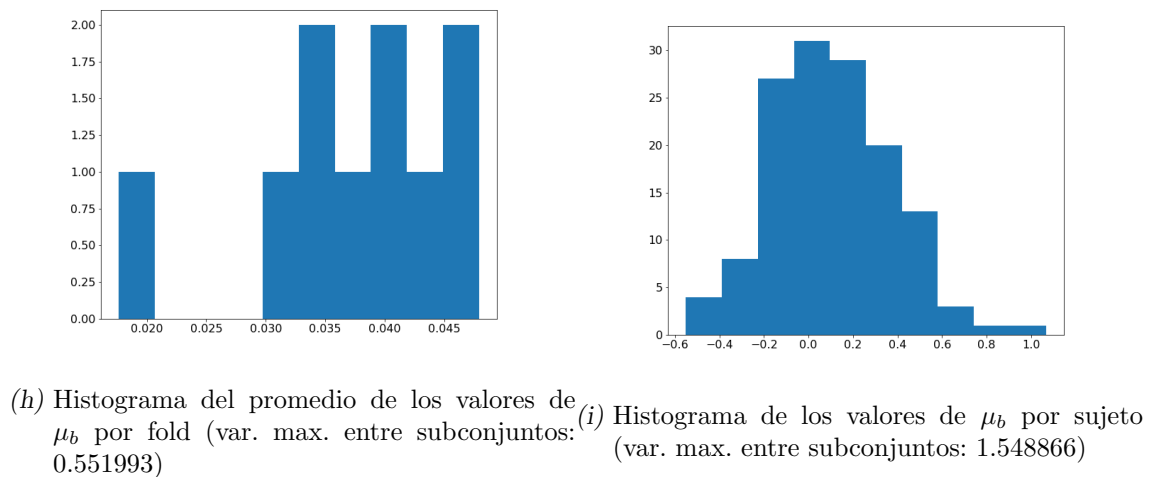
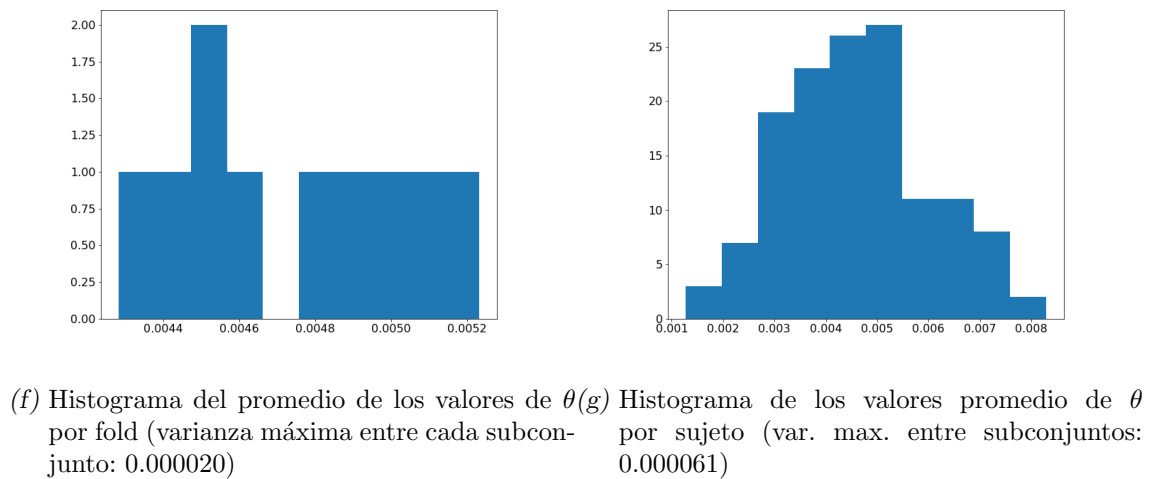
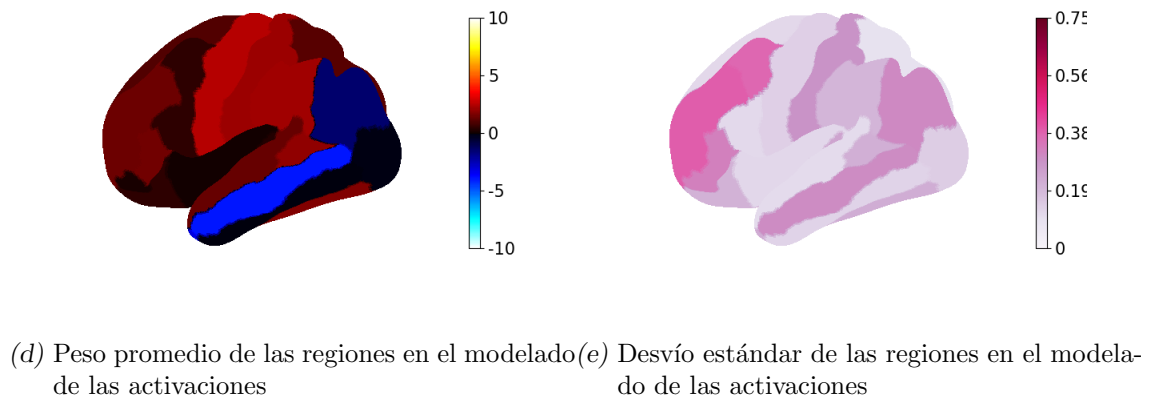
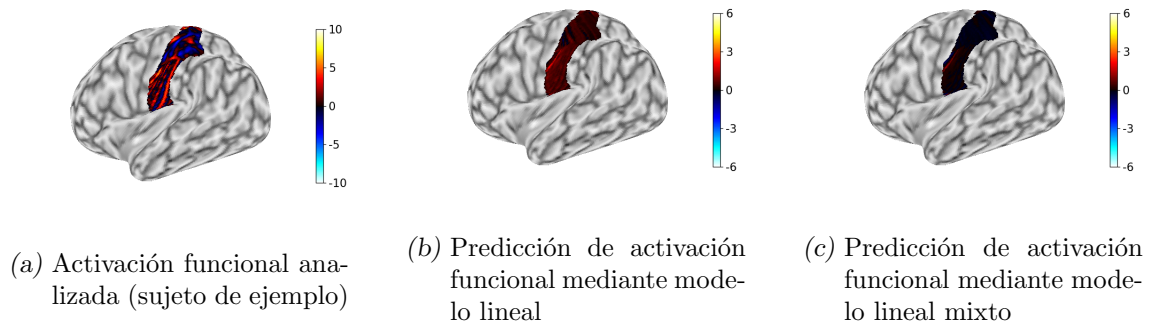
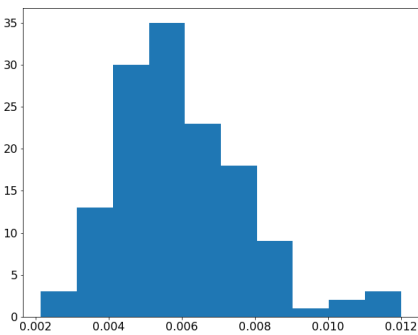
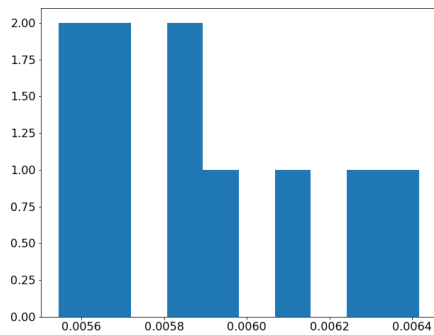
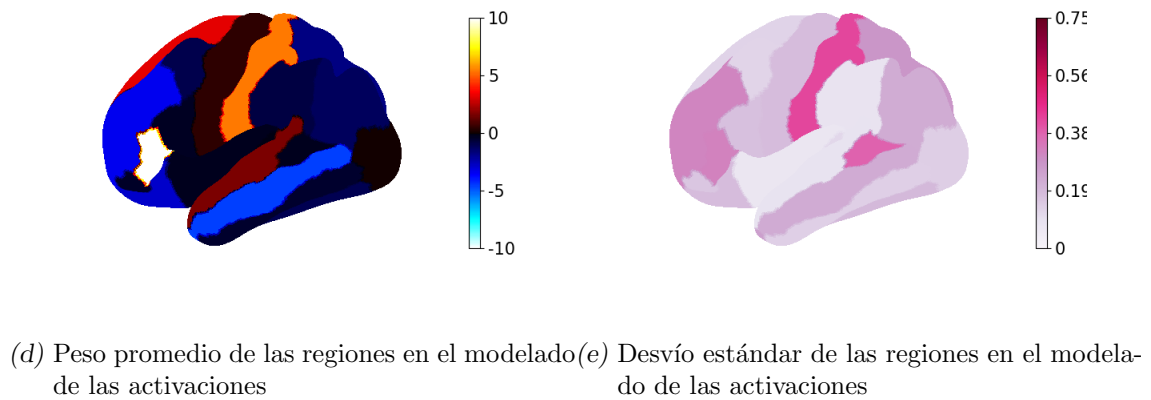
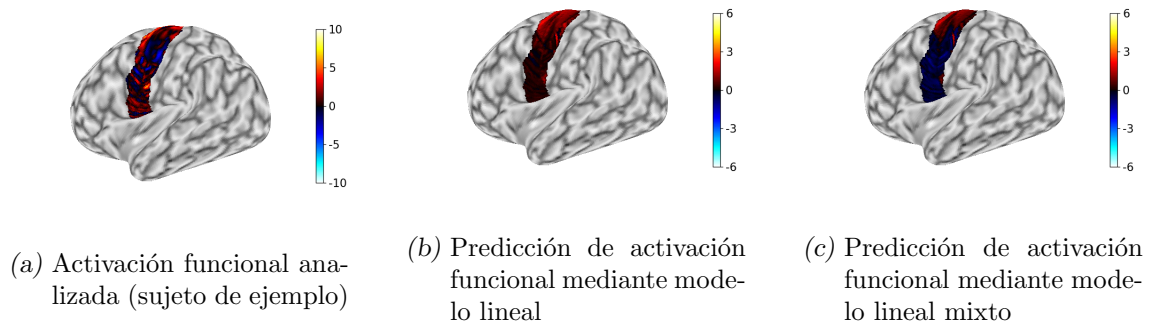
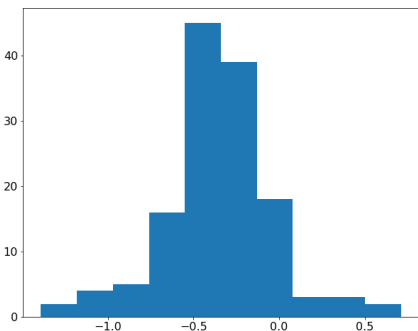
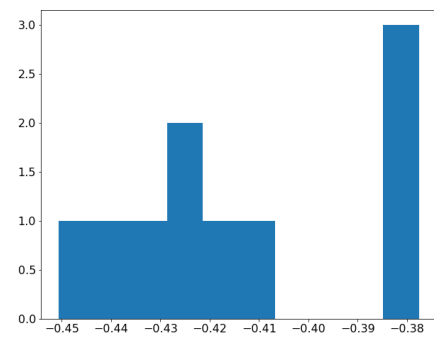


Fig. 6.3: Desikan, región Postcentral izquierda



(f) Histograma del promedio de los valores de $\theta(g)$ por fold (varianza máxima entre cada subconjunto: 0.000021)

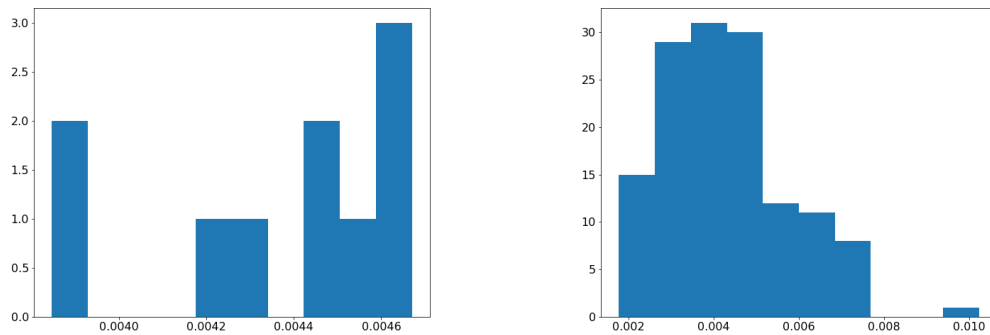
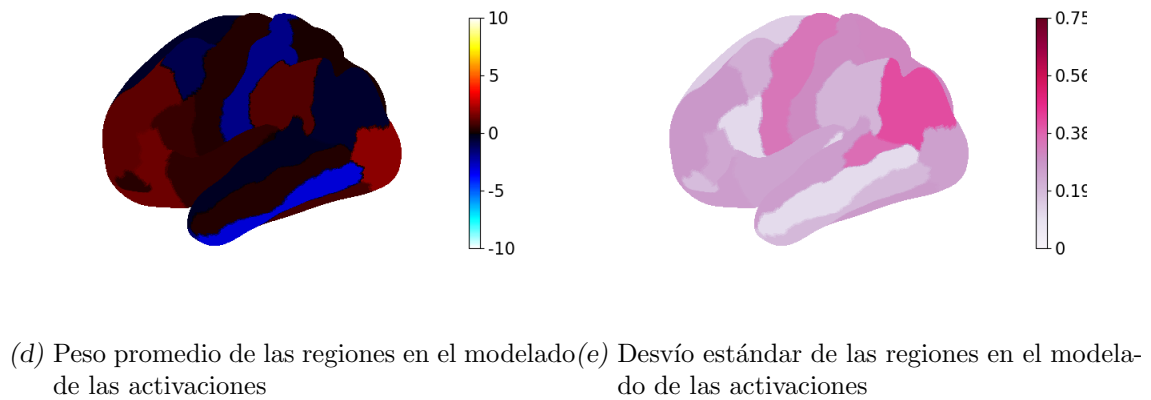
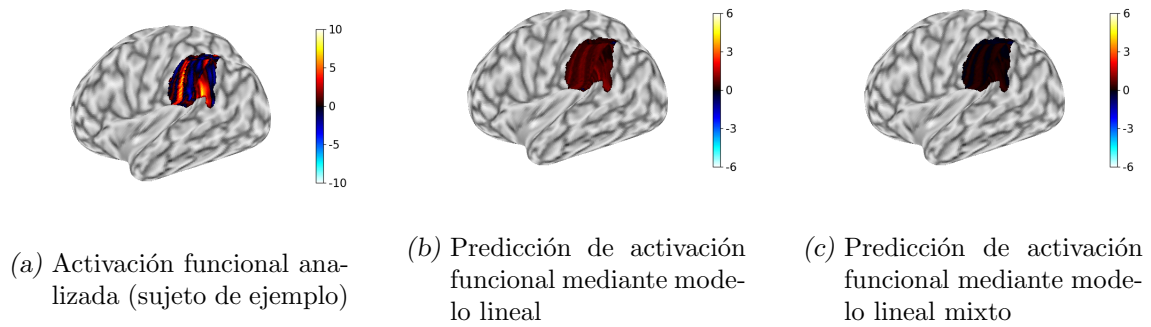
(g) Histograma de los valores promedio de θ por sujeto (var. max. entre subconjuntos: 0.000066)



(h) Histograma del promedio de los valores de μ_b por fold (var. max. entre subconjuntos: 0.790986)

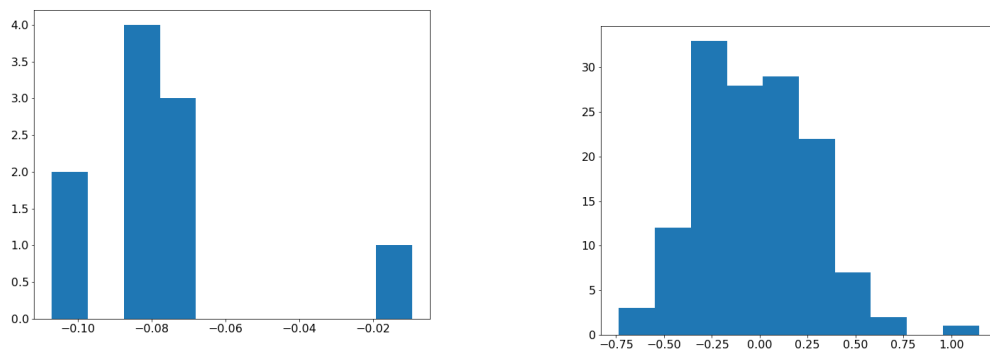
(i) Histograma de los valores de μ_b por sujeto (var. max. entre subconjuntos: 4.243512)

Fig. 6.4: Desikan, región Precentral izquierda



(f) Histograma del promedio de los valores de θ por fold (varianza máxima entre cada subconjunto: 0.000018)

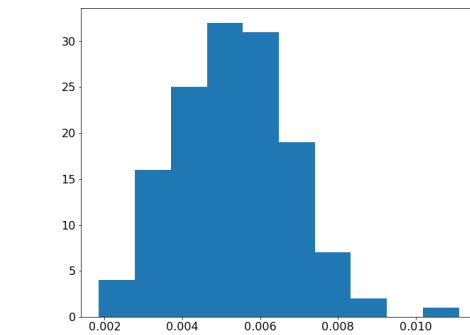
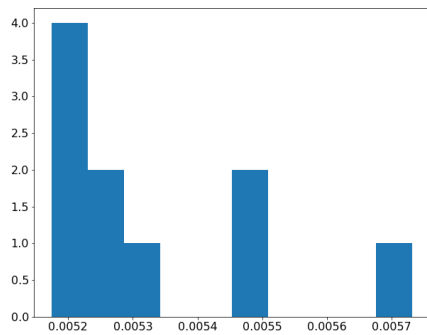
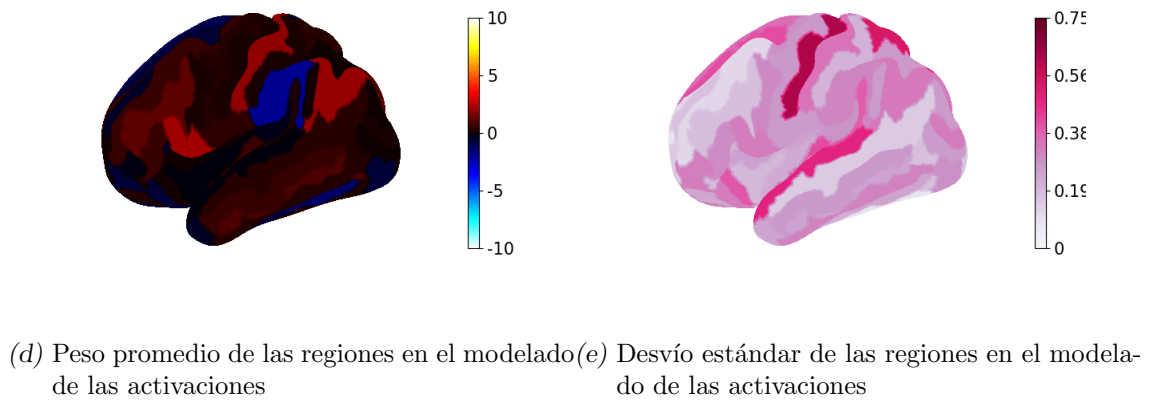
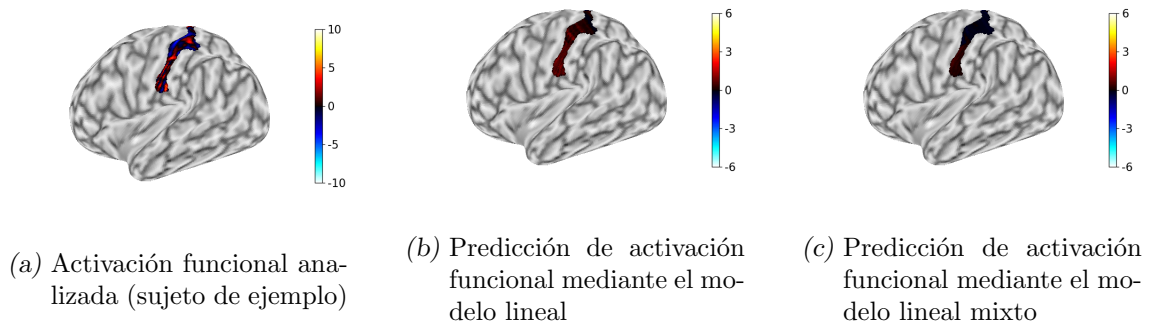
(g) Histograma de los valores promedio de θ por sujeto (var. max. entre subconjuntos: 0.000047)



(h) Histograma del promedio de los valores de μ_b por fold (var. max. entre subconjuntos: 0.573939)

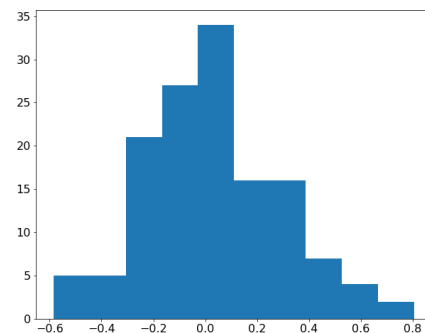
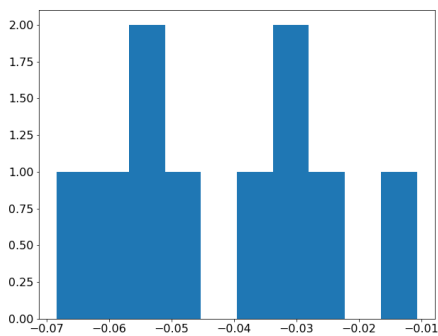
(i) Histograma de los valores de μ_b por sujeto (var. max. entre subconjuntos: 19.688366)

Fig. 6.5: Desikan, región Supra Marginal izquierda



(f) Histograma del promedio de los valores de θ por fold (varianza máxima entre cada subconjunto 0.000018)

(g) Histograma de los valores promedio de θ por sujeto (var. max. entre subconjuntos: 0.000048)



(h) Histograma del promedio de los valores de μ_b por fold (var. max. entre subconjuntos: 0.768816)

(i) Histograma de los valores de μ_b por sujeto (var. max. entre subconjuntos: 7.232169)

Fig. 6.6: Destrieux, Circunvolución Postcentral izquierda

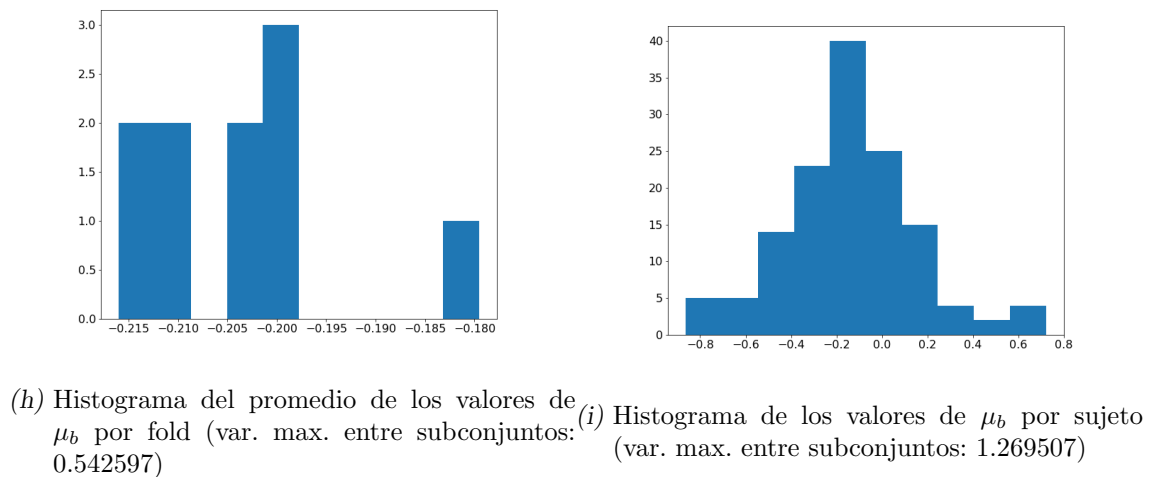
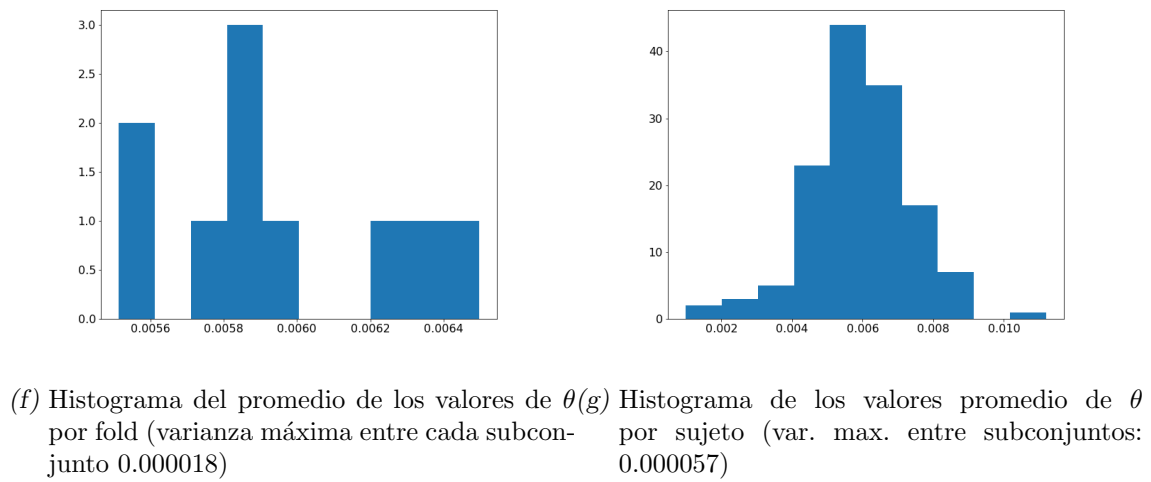
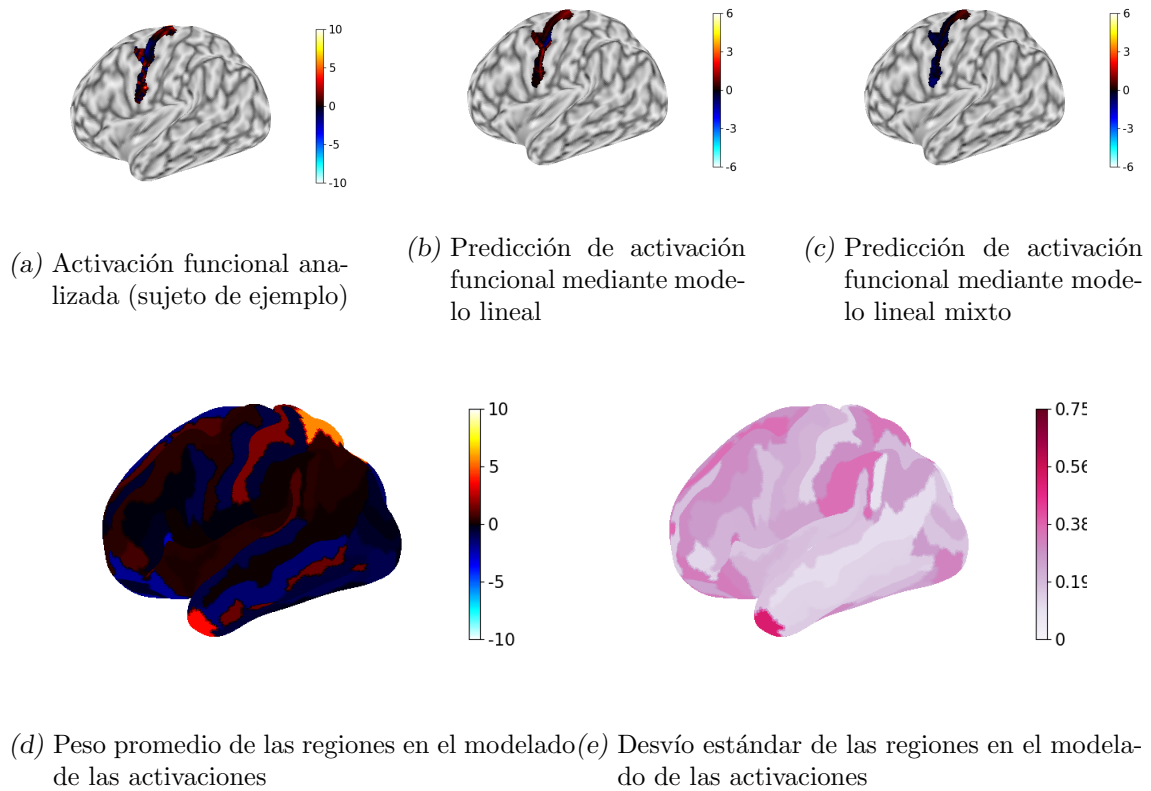
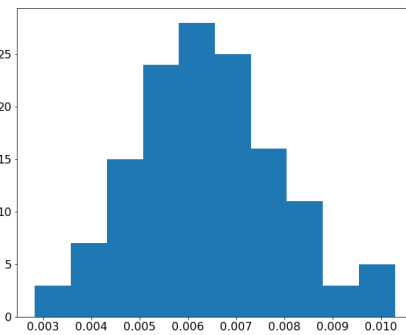
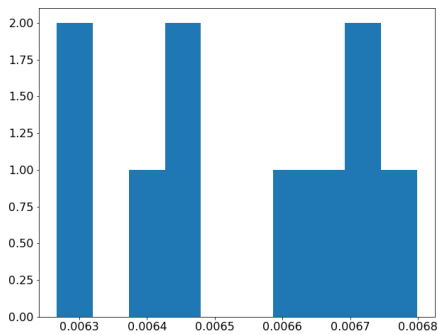
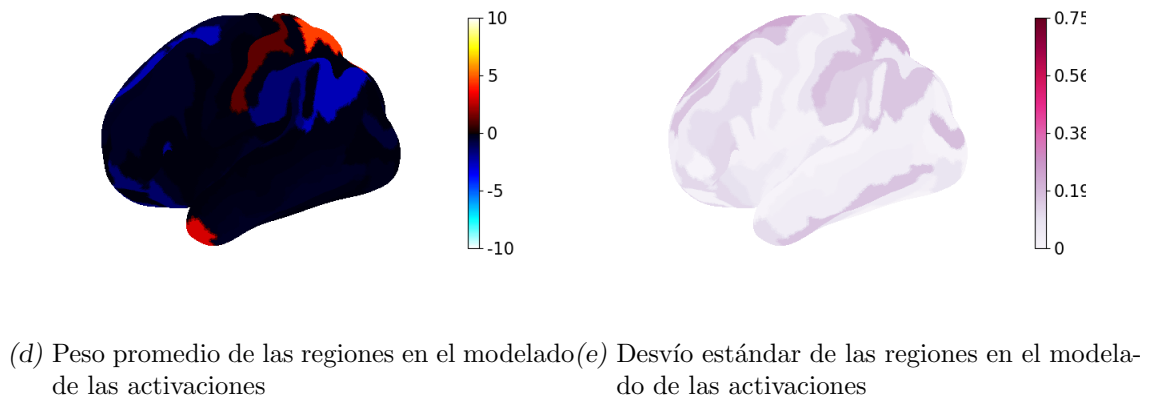
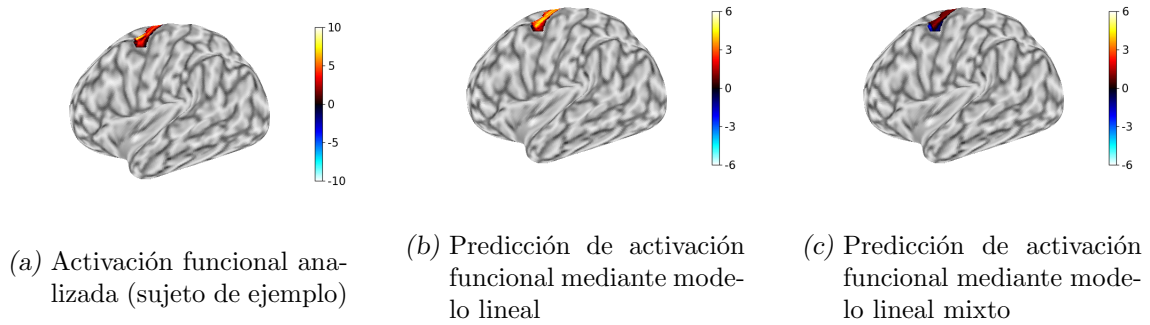
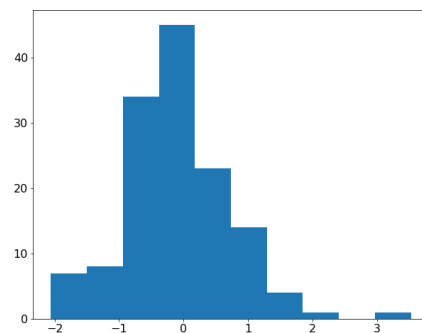
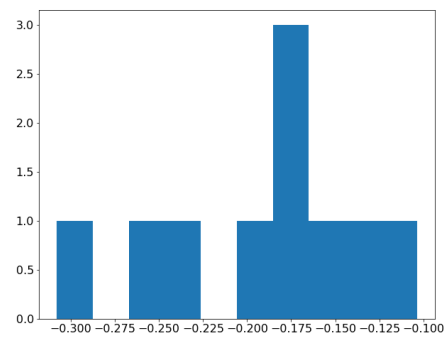


Fig. 6.7: Destrieux, Circunvolución Precentral izquierda



(f) Histograma del promedio de los valores de $\theta(g)$ por fold (varianza máxima entre cada subconjunto 0.000018)

(g) Histograma de los valores promedio de θ por sujeto (var. max. entre subconjuntos: 0.000054)



(h) Histograma del promedio de los valores de μ_b por fold (var. max. entre subconjuntos: 5.338791)

(i) Histograma de los valores de μ_b por sujeto (var. max. entre subconjuntos: 33.499615)

Fig. 6.8: Destrieux, Parte superior del Surco Precentral izquierdo

Tab. 6.2: Atlas de Desikan - Predicción sobre 14 sujetos.

Región	Estimador	LMM (prom. subj)	LMM	Modelo lineal
Postcentral izq.	NMSE	2.21 (0.19)	2.07 (0.13)	2.01 (0.1)
	χ^2	2158 (158)	24726 (1105)	25616 (1101)
Precentral izq.	NMSE	2.33 (0.18)	2.15 (0.11)	2.06 (0.12)
	χ^2	2195 (204)	24153 (936)	24639 (1417)
Supra Marginal izq.	NMSE	2.14 (0.19)	2.0 (0.15)	2.0 (0.14)
	χ^2	2114 (158)	24349 (1182)	24327 (1349)

Tab. 6.3: Atlas de Destrieux - Regresión sobre 90 sujetos.

Región	Estimador	LMM (prom. subj)	LMM	Modelo lineal
Circunvolución Postcentral izq.	NMSE	2.12 (0.09)	1.96 (0.08)	1.83 (0.03)
	χ^2	761 (24)	55499 (494)	53662 (433)
	AIC	4008 (211)	365733 (20196)	343981 (2607)
	LLH	-1352 (81)	-124228 (7893)	-115551 (1064)
Circunvolución Precentral izq.	NMSE	2.2 (0.05)	2.03 (0.03)	1.89 (0.07)
	χ^2	781 (15)	55647 (564)	53672 (513)
	AIC	4218 (29)	385736 (2913)	363765 (19980)
	LLH	-1433 (12)	-132031 (1274)	-123244 (7784)
Parte sup. Surco Precentral izq.	NMSE	2.7 (0.83)	2.29 (0.48)	2.01 (0.32)
	χ^2	714 (72)	46874 (12204)	51687 (25239)
	AIC	3503 (735)	320196 (66481)	356238 (162087)
	LLH	-1197 (229)	-110234 (20816)	-121468 (54819)

Tab. 6.4: Atlas de Destrieux - Predicción sobre 14 sujetos.

Región	Estimador	LMM (prom. subj)	LMM	Modelo lineal
Circunvolución Postcentral izq.	NMSE	2.09 (0.17)	1.95 (0.13)	2.0 (0.1)
	χ^2	752 (48)	8518 (407)	9013 (570)
Circunvolución Precentral izq.	NMSE	2.16 (0.17)	2.01 (0.13)	2.02 (0.13)
	χ^2	772 (57)	8539 (413)	8842 (586)
Parte sup. Surco Precentral izq.	NMSE	2.65 (0.86)	2.28 (0.49)	3.38 (9.2)
	χ^2	701 (113)	7195 (1907)	10619 (15882)

Tab. 6.5: Atlas de Desikan aplicando LASSO - Regresión sobre 90 sujetos.

Región	Estimador	LMM (prom. subj)	LMM	Modelo lineal + LASSO
Postcentral izq.	NMSE	2.26 (0.12)	2.09 (0.09)	1.99 (0.03)
	χ^2	2204 (47)	162193 (4075)	164665 (629)
	AIC	12187 (602)	1107767 (55177)	1051328 (6249)
	LLH	-4165 (246)	-380321 (22692)	-357008 (2897)
Precentral izq.	NMSE	2.38 (0.04)	2.18 (0.03)	2.04 (0.06)
	χ^2	2242 (31)	158211 (946)	158599 (6158)
	AIC	12786 (54)	1162667 (5110)	1103734 (52728)
	LLH	-4410 (25)	-402865 (2434)	-378305 (21470)
Supra Marginal izq.	NMSE	2.19 (0.14)	2.03 (0.12)	1.99 (0.09)
	χ^2	2165 (72)	159907 (4766)	157184 (5444)
	AIC	11611 (952)	1055259 (86929)	1052243 (84649)
	LLH	-3945 (371)	-360186 (34006)	-358678 (32855)

Tab. 6.6: Atlas de Desikan aplicando LASSO - Predicción sobre 14 sujetos.

Región	Estimador	LMM (prom. subj)	LMM	Modelo lineal + LASSO
Postcentral izq.	NMSE	2.22 (0.2)	2.08 (0.14)	1.99 (0.09)
	χ^2	2172 (166)	24817 (1092)	25468 (1080)
Precentral izq.	NMSE	2.35 (0.19)	2.17 (0.11)	2.05 (0.12)
	χ^2	2216 (209)	24287 (840)	24517 (1385)
Supra Marginal izq.	NMSE	2.16 (0.2)	2.01 (0.15)	1.99 (0.13)
	χ^2	2133 (163)	24480 (1201)	24249 (1306)

Tab. 6.7: Atlas de Destrieux aplicando LASSO - Regresión sobre 90 sujetos.

Región	Estimador	LMM (prom. subj)	LMM	Modelo lineal
Circunvolución Postcentral izq.	NMSE	2.05 (0.05)	1.89 (0.04)	1.92 (0.03)
	χ^2	746 (17)	55612 (637)	56439 (354)
	AIC	3803 (30)	345993 (2905)	346829 (2644)
	LLH	-1274 (12)	-116557 (1238)	-116975 (1086)
Circunvolución Precentral izq.	NMSE	2.19 (0.28)	2.02 (0.23)	1.94 (0.04)
	χ^2	791 (110)	57157 (6228)	55114 (1415)
	AIC	4026 (230)	367230 (21737)	365249 (18632)
	LLH	-1361 (91)	-124976 (8719)	-123986 (7112)
Parte sup. Surco Precentral izq.	NMSE	2.82 (0.93)	2.37 (0.54)	2.05 (0.31)
	χ^2	748 (112)	48408 (13389)	52888 (25936)
	AIC	3521 (738)	321646 (66869)	357486 (162617)
	LLH	-1206 (231)	-110959 (21067)	-122093 (55069)

Tab. 6.8: Atlas de Destrieux aplicando LASSO - Predicción sobre 14 sujetos.

Región	Estimador	LMM (prom. subj)	LMM	Modelo lineal + LASSO
Circunvolución Postcentral izq.	NMSE	2.03 (0.14)	1.9 (0.11)	1.9 (0.11)
	χ^2	738 (29)	8553 (448)	8568 (411)
Circunvolución Precentral izq.	NMSE	2.17 (0.33)	2.01 (0.28)	1.96 (0.14)
	χ^2	781 (106)	8783 (1008)	8577 (470)
Parte sup. Surco Precentral izq.	NMSE	2.78 (0.98)	2.16 (0.37)	2.14 (0.42)
	χ^2	737 (141)	7442 (2082)	8333 (3947)

Tab. 6.9: Atlas de Desikan - Porcentaje de vértices activados por región analizadas.

Región	Porcentaje de activación	Cantidad vértices regresión	Cantidad vértices predicción
Post Central izq. (22)	37.60 %	168655	25645
Pre Central izq. (24)	39.00 %	178467	27175
Supra Marginal izq. (31)	30.01 %	155206	23646

Tab. 6.10: Atlas de Destrieux - Porcentaje de vértices activados por región analizadas.

Región	Porcentaje de activación	Cantidad vértices regresión	Cantidad vértices predicción
Circunvolución Post Central izq.	44.37 %	56439	8568
Circunvolución Pre Central izq.	49.87 %	60836	9251
Parte sup. Surco Pre Central izq.	38.37 %	32314	4937

7. DISCUSIÓN

Los modelos de efectos aleatorios, utilizados para estudiar simultáneamente efectos a nivel poblacional e individuales, no escala bien con grandes volúmenes de datos. Esto presenta un problema en áreas de estudio como la neurociencia donde cada vez se utilizan una mayor cantidad de datos de alta resolución. En este trabajo abordamos esta problemática y presentamos un método de resolución de modelos lineal mixto utilizando métodos proximales. Nuestra implementación nos permite reducir drásticamente los requisitos de memoria, abrir las puertas a la computación en paralelo e incluir estrategias de regularización. Para ilustrar estas ventajas, aplicamos nuestro modelo al estudio de la relación entre estructura y función en el cerebro humano. En este capítulo discutimos los resultados obtenidos tanto en experimentos sobre datos sintéticos como en datos provenientes de la base de datos del Human Connectome Project (HCP).

7.1. Características del modelo a tener en cuenta

En primera instancia, como menciona Boyd et al. [7], ADMM puede ser muy lento para converger a una alta precisión. Sin embargo, a menudo se da el caso en que ADMM converge a una precisión modesta -suficiente para muchas aplicaciones- dentro de unas pocas decenas de iteraciones. Este comportamiento hace que ADMM sea similar a algoritmos como el método del gradiente conjugado, por ejemplo, en el sentido de que algunas decenas de iteraciones a menudo producen resultados aceptables de uso práctico. Afortunadamente, este suele ser el caso para muchos tipos de problemas a gran escala. Además, en el caso de problemas estadísticos y de aprendizaje de máquinas, la resolución de un problema de estimación de parámetros con una precisión muy alta a menudo produce una pequeña o nula mejora en el rendimiento real de la predicción. En el caso de nuestro problema se da esta última situación mencionada, donde un aumento considerable en la cantidad de iteraciones no produjo mejoras significativas, pero consideramos que de todas maneras esta situación debe ser tenida en cuenta en futuros experimentos utilizando el método propuesto.

En segundo lugar, creemos que una mejora posible sobre estos resultados, podría obtenerse mediante la modelización de características propias del conjunto de datos que nosotros no tuvimos en cuenta al momento de realizar la experimentación. Por ejemplo, dado que estamos modelando conectividad estructural, y contamos con datos de sujetos de sexo femenino y masculino, podríamos haber utilizado esta distinción en el modelado de nuestros efectos aleatorios en busca de una mejora en la regresión/predicción realizada.

7.2. Análisis de los resultados en datos sintéticos

Un análisis sobre los resultados obtenidos en la figura 6.1 respecto al ejemplo sintético propuesto en el apartado 6.1, nos muestra que, si bien el método converge de forma estable, presenta dos características importantes.

Por un lado, existe la posibilidad de converger hacia mínimos locales que no son necesariamente mínimos globales de nuestra función. Este es un problema común a todos los métodos basados en gradientes. Si observamos la convergencia de θ_2 en la figura 6.1

vemos que para uno de los sujetos simulados su valor converge a un mínimo que no es el esperado, dado que se aleja y se mantiene a una distancia constante del resultado real. Creemos que una forma de mejorar sustancialmente el modelo es intentando atacar este problema. Esto escapa a los alcances actuales de nuestro trabajo y queda como objeto de futuros estudios.

Por otro lado, si observamos las primeras iteraciones de θ_1 , β_2 y β_i en la figura 6.1, veremos que el método presenta cierta inestabilidad inicial en la convergencia. Posiblemente esto sea producto de la interacción entre las distintas variables al momento de ser actualizadas. La minimización de las variables necesarias para cada iteración de ADMM, depende de valores calculados para otras variables tomadas como constantes (por ejemplo el gradiente de x y θ_2 depende de los valores de θ_1 y β_2). Esto hace que un mismo gradiente se calcule sobre superficies diferentes en sucesivas iteraciones y la convergencia pueda presentar algunas situaciones de inestabilidad como las observadas.

Por fuera de estas dos características, el método no presenta problemas de convergencia y si bien la velocidad de convergencia disminuye rápidamente, como en el caso de β_1 , esto afecta de forma significativa el tiempo de ejecución.

7.3. Análisis de los resultados en sujetos del Human Connectome Project

En primera instancia, nos propusimos mostrar la utilidad de nuestro método aplicándolo en el estudio de la relación lineal entre estructura y función en el cerebro humano. Aprovechando los beneficios provistos por nuestra implementación, quisimos estudiar esta relación sobre toda la corteza cerebral. Es importante remarcar que los modelos lineales mixtos provistos por paquetes de software en la actualidad [29] no fueron capaces de manejar este volumen de datos y fallaron frente al alto consumo de memoria de esta tarea.

Para poder comparar nuestro método contra el modelo lineal nos propusimos entonces realizar un segundo análisis acotando los datos a regiones específicas del cerebro. Un análisis cualitativo de los gráficos de activación predichos (ver figuras 6.3-6.8) nos muestra en primera instancia que, tanto el modelo lineal como el modelo lineal mixto, solo logran explicar poco más de un tercio de la varianza de los datos, con activaciones predichas que no superan los 3.5 (a excepción del surco precentral en el atlas de Destrieux). Podemos observar que tanto en el modelo lineal como en nuestro modelo lineal mixto, a pesar de que ciertas regiones son consistentes en cuanto a la figura de las activaciones predichas (como por ejemplo la figura 6.5 o la figura 6.8), en general no se logran los resultados esperados (ver subfiguras (b) y (c) respecto a subfigura (a)).

En las subfigura (e) de cada región, podemos observar que los parámetros de la predicción de nuestro modelo lineal mixto presentan una baja varianza para la mayoría de las regiones y, como es esperable, en la mayoría de las regiones analizadas la región a predecir también es la que realiza el aporte más significativo en dicha predicción (ver subfigura (d)). Por último, la información provista por los histogramas de las subfiguras (g) e (i) confirman la presencia de distintas varianzas por sujeto. Esta característica apoya nuestra idea de la utilización de un modelo mixto capaz de realizar un manejo simultáneo de efectos poblacionales y de sujeto, que pueda tener en cuenta esta variabilidad.

Analicemos ahora los resultados cuantitativos en las tablas 6.1-6.4 obtenidos mediante la comparación de las columnas *LMM*, siglas en inglés para Modelo Lineal Mixto, y *Modelo lineal* de cada una de las tablas mencionadas. Para el caso del atlas de Desikan, tanto en

la predicción (tabla 6.2) como en la regresión (tabla 6.1), el análisis de los estimadores calculados nos dicen que nuestro modelo no presenta una mejora cuantitativa frente al modelo lineal propuesto por Osher et al.[22]. Sin embargo, si reiteramos el análisis utilizando el atlas de Destrieux [12] para generar la división de la corteza cerebral (ver tabla 6.3 y tabla 6.4) observamos que aunque los estimadores de la regresión no hayan superado al modelo lineal, si existe una mejora en los estimadores de la predicción de las tres regiones seleccionadas. Esta diferencia respecto a los dos atlas propuestos, se debe principalmente al hecho de que el atlas de Destrieux [12] es mucho mas denso en cantidad de regiones, otorgando mas grados de libertad al sistema lineal.

Es importante destacar que no encontramos una ventaja significativa en la aplicación de una regularización L1. Tanto si comparamos la columna de *Modelo lineal* con su contraparte *Modelo Lineal + Lasso* en el atlas de Desikan (tabla 6.1 y tabla 6.2 frente a la tabla 6.5 y la tabla 6.6 respectivamente) y en el de Destrieux (tabla 6.3 y tabla 6.4 frente a tabla 6.7 y tabla 6.8) vemos que no existe una mejora significativa en los estimadores. Los resultados se mantienen consistentes con los obtenidos sin la aplicación de la regularización propuesta.

7.4. Observaciones finales

Consideramos que tanto nuestro modelo lineal mixto como el modelo lineal ordinario no dieron buenos resultados. Quizás una forma de mejorar esto sería utilizando otro tipo de atlas. Dado que ninguno de los dos atlas utilizados basa sus divisiones en la especificidad funcional del cerebro, sino que lo hacen respecto a surcos y circunvoluciones, regiones más pequeñas abarcan una mayor cantidad de vértices relacionados con cierta especificidad funcional. Al mismo tiempo, estas divisiones anatómicas no tienen por qué corresponderse a divisiones basadas en conectividad estructural del cerebro.

Para finalizar este capítulo nos gustaría remarcar el hecho de que si bien el método propuesto no mejora la predicción de activaciones funcionales mediante conectividad estructural, el poder ajustar un modelo lineal mixto en toda la corteza es un avance importante. Esto abre las puertas a nuevos trabajos que quieran estudiar los efectos a nivel poblacional e individual en grandes volúmenes de datos.

8. CONCLUSIÓN

Durante este trabajo, planteamos una solución al problema de aplicar modelos jerárquicos sobre grandes conjuntos de datos. En particular, atacamos el problema de ajustar modelos lineales mixtos tratando de resolver las limitaciones producidas por el enfoque de máxima verosimilitud [21] utilizado en la mayoría de paquetes de software de la actualidad [29], que requiere de la condición de que todos los datos estén disponibles en memoria al momento del cálculo. Para ello, presentamos una solución mediante el método de los multiplicadores alternados [7] (o ADMM por sus siglas en inglés) que nos permite ajustar nuestro modelo lineal mixto de forma descentralizada, permitiéndonos una optimización de escala arbitraria, además de ofrecernos la ventaja práctica de resolver las limitaciones de memoria de otros enfoques, reduciendo el orden de memoria del $O(2nm)$ de los modelos de máxima verosimilitud al $O(2n)$ de nuestra propuesta, donde n representa el tamaño de la matriz A en la ecuación (5.1) a ajustar y m la cantidad de sujetos.

A modo de ejemplificación utilizamos este modelo para el estudio de las relaciones entre la funcionalidad cerebral y la conectividad estructural. Para ellos, replicamos el trabajo realizado por Osher et al. [22] pero utilizando un modelo lineal mixto, con la idea de obtener tanto efectos fijos de la población, como efectos aleatorios de los sujetos. Como dijimos anteriormente, nuestra propuesta elimina la necesidad de que todos los datos se carguen en memoria al momento de realizar el ajuste sobre el modelo, por lo que permite un análisis en la totalidad de la corteza cerebral de cada sujeto, considerando la relación entre las distintas regiones y obteniendo efectos aleatorios para cada una de estas. Esto permite llevar el análisis inter-regional de la población realizado por Osher et al. [22] a un nivel individual, modelado la jerarquía de los datos de manera correspondientes. Si bien en esta aplicación, a primera vista los resultados no presentan una clara mejoría, nuestro modelo lineal mixto mediante multiplicadores alternados, abre la puerta a la realización de nuevos experimentos para los que antes se necesitaba hardware especializado, pudiendo así extender el análisis poblacional a un análisis mucho más profundo que tenga en cuenta las características intra-sujeto que estos modelos permiten incluir. Por todo esto, vemos un futuro interesante para la aplicación de modelos lineales mixtos en el campo de la neurociencia. Y este trabajo es un pequeño paso en esa dirección.

Apéndices

Apéndice A

CALCULO DE DERIVADAS

A.1. Caso general

A.1.1. Derivada respecto a θ_1

$$L(\theta_1, \theta_2, \hat{\beta}_{\theta_1}, \hat{\beta}_{\theta_2}, z) = \log \frac{|L_{\theta_1}|^2}{|W|} + \frac{\tau_\theta}{2} \left\| \frac{1}{\theta_1} - \frac{1}{\theta_2} + z_\theta \right\|_2^2 \quad (\text{A.1})$$

$$L(\dots) = \log |L_{\theta_1}^2| - \log |W| + \frac{\tau}{2} \left[\left(\frac{1}{\theta_1} - \frac{1}{\theta_2} + z_\theta \right)^t \left(\frac{1}{\theta_1} - \frac{1}{\theta_2} + z_\theta \right) \right] \quad (\text{A.2})$$

$$L(\dots) = \log |\Lambda_\theta^t Z^t W Z \Lambda_\theta + \Lambda_\theta^2| - \log |W| + \frac{\tau}{2} [\theta_1^{-2} - 2\theta_1^{-1}(\theta_2^{-1} - z_\theta) + (\theta_2^{-1} - z_\theta)^2] \quad (\text{A.3})$$

$$L(\dots) = \log \left| \frac{1}{\theta_1^2} (Z^t W Z + I)^2 \right| - \log |W| + \frac{\tau}{2} [\theta_1^{-2} - 2\theta_1^{-1}(\theta_2^{-1} - z_\theta) + (\theta_2^{-1} - z_\theta)^2] \quad (\text{A.4})$$

Dado $Z^t W Z$ una matriz de $k \times k$

$$L(\dots) = \log \frac{1}{\theta_1^{2k}} + \log |(Z^t W Z + I)^2| - \log |W| + \frac{\tau}{2} [\theta_1^{-2} - 2\theta_1^{-1}(\theta_2^{-1} - z_\theta) + (\theta_2^{-1} - z_\theta)^2] \quad (\text{A.5})$$

$$\frac{\partial L(\dots)}{\partial \theta_1} = -2k\theta_1^{-1} - \tau_\theta \theta_1^{-3} + \tau_\theta \theta_1^{-2} \left(\frac{1}{\theta_2} - z_\theta \right) \quad (\text{A.6})$$

$$\theta_1^{-3} \left(-2k\theta_1^2 - \tau_\theta + \tau_\theta \theta_1 \left(\frac{1}{\theta_2} - z_\theta \right) \right) = 0 \quad (\text{A.7})$$

$$-2k\theta_1^2 + \tau_\theta \theta_1 \left(\frac{1}{\theta_2} - z_\theta \right) - \tau_\theta = 0 \quad (\text{A.8})$$

A.1.2. Derivada respecto a x

$$L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z) = \frac{r^2(\theta, x)}{\sigma^2} + \frac{\tau_\beta}{2} \|S_\beta x - \hat{\beta}_{\theta_2} + z_\beta\|_2^2 \quad (\text{A.9})$$

$$L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z) = \frac{\left\| \begin{bmatrix} W^{1/2}(y_{obs} - o) \\ 0 \end{bmatrix} - \begin{bmatrix} W^{1/2} Z \Lambda_\theta & W^{1/2} A \\ \Lambda_\theta & 0 \end{bmatrix} x \right\|^2}{\sigma^2} + \frac{\tau_\beta}{2} \left[(S_\beta x - \hat{\beta}_{\theta_2} + z_\beta)^t (S_\beta x - \hat{\beta}_{\theta_2} + z_\beta) \right] \quad (\text{A.10})$$

De forma momentanea y por legibilidad voy a definir:

$$M = \begin{bmatrix} W^{1/2} Z \Lambda_\theta & W^{1/2} A \\ \Lambda_\theta & 0 \end{bmatrix} \quad (\text{A.11})$$

$$q = \begin{bmatrix} W^{1/2}(y_{obs} - o) \\ 0 \end{bmatrix} \quad (\text{A.12})$$

Entonces, tomando solo el primer termino y reemplazando, tengo que

$$\frac{\|q - Mx\|^2}{\sigma^2} \quad (\text{A.13})$$

$$\frac{x^t M^t Mx - 2q^t Mx + q^t q}{\sigma^2} \quad (\text{A.14})$$

donde, si derivo respecto a x , obtengo

$$\frac{\partial \frac{x^t M^t Mx - 2q^t Mx + q^t q}{\sigma^2}}{\partial x} = \frac{2M^t Mx - 2M^t q}{\sigma^2} \quad (\text{A.15})$$

Luego, resolviendo el termino del lagrangiano aumentado obtengo:

$$\frac{\tau_\beta}{2} \left[[S_\beta x - (\hat{\beta}_{\theta_2} - z_\beta)]^t [S_\beta x - (\hat{\beta}_{\theta_2} - z_\beta)] \right] \quad (\text{A.16})$$

$$\frac{\tau_\beta}{2} [x^t S_\beta^t S_\beta x - x^t S_\beta^t (\hat{\beta}_{\theta_2} - z_\beta) - (\hat{\beta}_{\theta_2} - z_\beta)^t S_\beta x + (\hat{\beta}_{\theta_2} - z_\beta)^t (\hat{\beta}_{\theta_2} - z_\beta)] \quad (\text{A.17})$$

$$\tau_\beta S_\beta^t S_\beta x - \tau_\beta S_\beta^t (\hat{\beta}_{\theta_2} - z_\beta) \quad (\text{A.18})$$

Finalmente, combinando (A.15) y (A.18) obtengo:

$$\frac{\partial L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z)}{\partial x} = \frac{2M^t Mx - 2M^t q}{\sigma^2} + \tau_\beta S_\beta^t S_\beta x - \tau_\beta S_\beta^t (\hat{\beta}_{\theta_2} - z_\beta) \quad (\text{A.19})$$

donde si reemplazo nuevamente los valores definidos en (A.11) y (A.12) e igualo a cero, obtengo

$$\left(\frac{2 \begin{bmatrix} W^{1/2} Z \Lambda_\theta & W^{1/2} A \\ \Lambda_\theta & 0 \end{bmatrix}^t \begin{bmatrix} W^{1/2} Z \Lambda_\theta & W^{1/2} A \\ \Lambda_\theta & 0 \end{bmatrix}}{\sigma^2} + \tau_\beta S_\beta^t S_\beta \right) x \quad (\text{A.20})$$

$$= \tau_\beta S_\beta^t (\hat{\beta}_{\theta_2} - z_\beta) + \frac{2 \begin{bmatrix} W^{1/2} Z \Lambda_\theta & W^{1/2} A \\ \Lambda_\theta & 0 \end{bmatrix}^t \begin{bmatrix} W^{1/2} (y_{obs} - o) \\ 0 \end{bmatrix}}{\sigma^2}$$

$$x = \left(\frac{2 \begin{bmatrix} W^{1/2} Z \Lambda_\theta & W^{1/2} A \\ \Lambda_\theta & 0 \end{bmatrix}^t \begin{bmatrix} W^{1/2} Z \Lambda_\theta & W^{1/2} A \\ \Lambda_\theta & 0 \end{bmatrix}}{\sigma^2} + \tau_\beta S_\beta^t S_\beta \right)^{-1} \quad (\text{A.21})$$

$$\left(\tau_\beta S_\beta^t (\hat{\beta}_{\theta_2} - z_\beta) + \frac{2 \begin{bmatrix} W^{1/2} Z \Lambda_\theta & W^{1/2} A \\ \Lambda_\theta & 0 \end{bmatrix}^t \begin{bmatrix} W^{1/2} (y_{obs} - o) \\ 0 \end{bmatrix}}{\sigma^2} \right)$$

Por ultimo, puedo combinar los términos internos para obtener una expresión mas prolija:

$$x = \begin{bmatrix} \Lambda_{\theta_2}^t Z^t W Z \Lambda_{\theta_2} + \Lambda_{\theta_2}^2 & \Lambda_{\theta_2}^t Z^t W A \\ A^t W Z \Lambda_{\theta_2} & A^t W A + \sigma^2 \tau_\beta I_\beta \end{bmatrix}^{-1} \begin{bmatrix} \Lambda_{\theta_2}^t Z^t W (y_{obs} - o) \\ \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o) \end{bmatrix} \quad (\text{A.22})$$

A.1.3. Derivada respecto a θ_2

y resolución

$$L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z) = \frac{r^2(\theta, x)}{\sigma^2} + \frac{\tau_\theta}{2} \left\| \frac{1}{\theta_1} - \frac{1}{\theta_2} + z_\theta \right\|_2^2 \quad (\text{A.23})$$

$$L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z) = \frac{r^2(\theta, x)}{\sigma^2} + \frac{\tau_\theta}{2} [(\theta_1^{-1} - \theta_2^{-1} + z_\theta)^t (\theta_1^{-1} - \theta_2^{-1} + z_\theta)] \quad (\text{A.24})$$

$$L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z) = \frac{\left\| \begin{bmatrix} W^{1/2}(y_{obs} - o) \\ 0 \end{bmatrix} - \begin{bmatrix} W^{1/2}Z\Lambda_\theta & W^{1/2}A \\ \Lambda_\theta & 0 \end{bmatrix} x \right\|^2}{\sigma^2} + \frac{\tau_\theta}{2} [(\theta_1^{-1} - \theta_2^{-1} + z_\theta)^t (\theta_1^{-1} - \theta_2^{-1} + z_\theta)] \quad (\text{A.25})$$

De forma momentanea y por legibilidad, voy a definir

$$M = \begin{bmatrix} W^{1/2}Z\Lambda_\theta & W^{1/2}A \\ \Lambda_\theta & 0 \end{bmatrix} \quad (\text{A.26})$$

$$q = \begin{bmatrix} W^{1/2}(y_{obs} - o) \\ 0 \end{bmatrix} \quad (\text{A.27})$$

Entonces, tomando solamente el primer termino y reemplazando obtengo:

$$\frac{x^t M^t M x - 2q^t M x + q^t q}{\sigma^2} \quad (\text{A.28})$$

Resolvamos la derivada del primer termino de (A.28):

$$\begin{aligned} \frac{\partial x^t M^t M x}{\partial \theta_1} &= \frac{\partial x^t \begin{bmatrix} W^{1/2}Z\Lambda_\theta & W^{1/2}A \\ \Lambda_\theta & 0 \end{bmatrix}^t \begin{bmatrix} W^{1/2}Z\Lambda_\theta & W^{1/2}A \\ \Lambda_\theta & 0 \end{bmatrix} x}{\partial \theta_1} \\ &= x^t \begin{bmatrix} -2\theta_2^{-3}(Z^t W Z + I) & -\theta_2^{-2}Z^t W A \\ -\theta_2^{-2}A^t W Z & 0 \end{bmatrix} x \end{aligned} \quad (\text{A.29})$$

Y ahora la derivada del segundo termino. Notese que derivar el tercer termino respecto de θ_2 es igual a cero por lo que podemos eliminarlo.

$$\frac{\partial 2q^t M x}{\partial \theta_1} = \frac{\partial 2x^t \begin{bmatrix} W^{1/2}Z\Lambda_\theta & W^{1/2}A \\ \Lambda_\theta & 0 \end{bmatrix}^t q}{\partial \theta_1} = 2x^t \begin{bmatrix} -\theta_2^{-2}W^{1/2}Z & 0 \\ -\theta_2^{-2} & 0 \end{bmatrix}^t q \quad (\text{A.30})$$

y por ultimo el termino del lagragiano:

$$\frac{\tau_\theta}{2} [(\theta_1^{-1} - \theta_2^{-1} + z_\theta)^t (\theta_1^{-1} - \theta_2^{-1} + z_\theta)] \quad (\text{A.31})$$

$$\frac{\tau_\theta}{2} [\theta_2^{-2} - 2\theta_2^{-1}(\theta_1^{-1} + z_\theta) + (\theta_1^{-1} + z_\theta)^2] \quad (\text{A.32})$$

$$-\tau_\theta \theta_2^{-3} + \theta_2^{-2}(\theta_1^{-1} + z_\theta) \quad (\text{A.33})$$

Ahora puedo volver a remplazar todas las derivadas calculadas en la funcion original e igualar a cero

$$\frac{x^t \begin{bmatrix} -2\theta_2^{-3}(Z^t W Z + I) & -\theta_2^{-2} Z^t W A \\ -\theta_2^{-2} A^t W Z & 0 \end{bmatrix} x - 2x^t \begin{bmatrix} -\theta_2^{-2} W^{1/2} Z & 0 \\ -\theta_2^{-2} & 0 \end{bmatrix}^t \begin{bmatrix} W^{1/2}(y_{obs} - o) \\ 0 \end{bmatrix}}{\sigma^2} - \tau_\theta \theta_2^{-3} + \tau_\theta \theta_2^{-2} (\theta_1^{-1} + z_\theta) = 0 \quad (\text{A.34})$$

Luego, puedo combinar este resultado con la derivada de x para obtener la derivada conjunta que necesito para f_2 . Por cuestiones de legibilidad, voy a resolver primero la sustitución de x en el numerador de (A.34) para finalmente combinar con la derivada del termino del lagrangiano.

Reescribo los términos para poder distribuir mejor

$$(E - FH^{-1}G)^{-1} = \theta_2^{-2} X_{11} \quad (\text{A.41})$$

$$-(E - FH^{-1}G)^{-1}FH^{-1} = \theta_2^{-1} X_{12} \quad (\text{A.42})$$

$$-H^{-1}G(E - FH^{-1}G)^{-1} = \theta_2^{-1} X_{21} \quad (\text{A.43})$$

$$H^{-1} + H^{-1}G(E - FH^{-1}G)^{-1}FH^{-1} = H^{-1} + X_{22} \quad (\text{A.44})$$

Luego, escribo todo como:

$$\begin{pmatrix} [-2\theta_2^{-3}(Z^tWZ + I) & -\theta^{-2}Z^tWA] \\ \begin{bmatrix} -\theta^{-2}A^tWZ \\ 0 \end{bmatrix} & \begin{bmatrix} \theta_2^{-2}X_{11} & \theta_2^{-1}X_{12} \\ \theta_2^{-1}X_{21} & H^{-1} + X_{22} \end{bmatrix} \end{pmatrix} \begin{bmatrix} \Lambda_{\theta_2}^t Z^t W(y_{obs} - o) \\ \Lambda_{\theta_2}^t Z^t W(y_{obs} - o) \end{bmatrix} - 2 \begin{bmatrix} \begin{bmatrix} \theta_2^{-2}X_{11} & \theta_2^{-1}X_{12} \\ \theta_2^{-1}X_{21} & H^{-1} + X_{22} \end{bmatrix}^t \\ -\theta^{-2}Z^tW(y_{obs} - o) \\ 0 \end{bmatrix} \end{pmatrix} \quad (\text{A.45})$$

Y resuelvo

$$\begin{pmatrix} [-2\theta_2^{-3}(Z^tWZ + I) & -\theta^{-2}Z^tWA] \\ \begin{bmatrix} -\theta^{-2}A^tWZ \\ 0 \end{bmatrix} & \begin{bmatrix} \theta_2^{-2}X_{11} & \theta_2^{-1}X_{12} \\ \theta_2^{-1}X_{21} & H^{-1} + X_{22} \end{bmatrix} \end{pmatrix} \begin{bmatrix} \theta_2(y_{obs} - o)^t W Z^t & (y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t) \\ \Lambda_{\theta_2}^t Z^t W(y_{obs} - o) & \Lambda_{\theta_2}^t Z^t W(y_{obs} - o) \end{bmatrix} - 2 \begin{bmatrix} \begin{bmatrix} \theta_2^{-2}X_{11} & \theta_2^{-1}X_{12} \\ \theta_2^{-1}X_{21} & H^{-1} + X_{22} \end{bmatrix}^t \\ -\theta^{-2}Z^tW(y_{obs} - o) \\ 0 \end{bmatrix} \end{pmatrix} \quad (\text{A.46})$$

$$\begin{pmatrix} [-2\theta_2^{-5}(Z^tWZ + I)X_{11} - \theta_2^{-3}Z^tWAX_{21} & -\theta_2^{-4}(Z^tWZ + I)X_{12} - \theta^{-2}Z^tWA(H + X_{22})] \\ \begin{bmatrix} -\theta_2^{-4}A^tWZX_{11} \\ -\theta_2^{-3}A^tWZX_{12} \end{bmatrix} & \begin{bmatrix} \theta_2(y_{obs} - o)^t W Z & (y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t) \\ \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t) + A^t W(y_{obs} - o) & \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t) \end{bmatrix} \end{pmatrix} \begin{bmatrix} \begin{bmatrix} \theta_2^{-2}X_{11} & \theta_2^{-1}X_{12} \\ \theta_2^{-1}X_{21} & H^{-1} + X_{22} \end{bmatrix}^t \\ \theta_2^{-2}Z^tW(y_{obs} - o) \\ 0 \end{bmatrix} \end{pmatrix} \quad (\text{A.47})$$

$$\begin{pmatrix} [-2\theta_2^{-5}(Z^tWZ + I)X_{11} - \theta_2^{-3}Z^tWAX_{21} & -\theta_2^{-4}(Z^tWZ + I)X_{12} - \theta^{-2}Z^tWA(H + X_{22})] \\ \begin{bmatrix} -\theta_2^{-4}A^tWZX_{11} \\ -\theta_2^{-3}A^tWZX_{12} \end{bmatrix} & \begin{bmatrix} \theta_2^{-1}(y_{obs} - o)^t W Z X_{11}^t + \theta_2^{-1}[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)]X_{12}^t & (y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t) \\ \theta_2^{-1}Z^tW(y_{obs} - o) & \theta_2^{-1}Z^tW(y_{obs} - o) \end{bmatrix} \end{pmatrix} - 2 \begin{bmatrix} \begin{bmatrix} \theta_2^{-2}X_{11} & \theta_2^{-1}X_{12} \\ \theta_2^{-1}X_{21} & H^{-1} + X_{22} \end{bmatrix}^t \\ \theta_2^{-2}Z^tW(y_{obs} - o) \\ 0 \end{bmatrix} \end{pmatrix} \quad (\text{A.48})$$

Resolviendo las multiplicaciones de matrices restantes, obtengo

$$\begin{aligned}
& \{ \theta_2^{-1} (y_{obs} - o)^t W Z X_{11}^t + \theta_2^{-1} [(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t \} \\
& \quad \{ [-2\theta_2^{-5} (Z^t W Z + I) X_{11} - \theta_2^{-3} Z^t W A X_{21}] \theta_2^{-1} Z^t W (y_{obs} - o) \\
& \quad \quad + [-2\theta_2^{-4} (Z^t W Z + I) X_{12}] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)] \\
& \quad \quad + [-\theta_2^{-2} Z^t W A (H^{-1} + X_{22})] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)] \\
& \quad \quad \quad + 2\theta_2^{-2} Z^t W (y_{obs} - o) \} \\
& \quad + \\
& \quad \{ (y_{obs} - o)^t W Z X_{21}^t + [(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] [X_{22}^t + (H^t)^{-1}] \} \\
& \quad \{ -\theta_2^{-5} A^t W Z X_{11} Z^t W (y_{obs} - o) - \theta_2^{-3} A^t W Z X_{12} [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)] \} \\
& \hspace{15em} (A.49)
\end{aligned}$$

Donde todavía puedo hacer algunas multiplicaciones internas para terminar de despejar los términos con θ_2

$$\begin{aligned}
& \{ \theta_2^{-1} (y_{obs} - o)^t W Z X_{11}^t + \theta_2^{-1} [(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t \} \\
& \quad \{ -2\theta_2^{-6} (Z^t W Z + I) X_{11} Z^t W (y_{obs} - o) \\
& \quad \quad - \theta_2^{-4} Z^t W A X_{21} Z^t W (y_{obs} - o) \\
& \quad \quad + [-2\theta_2^{-4} (Z^t W Z + I) X_{12}] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)] \\
& \quad \quad + [-\theta_2^{-2} Z^t W A (H^{-1} + X_{22})] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)] \\
& \quad \quad \quad + 2\theta_2^{-2} Z^t W (y_{obs} - o) \} \\
& \quad + \\
& \quad \{ (y_{obs} - o)^t W Z X_{21}^t + [(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] [X_{22}^t + (H^t)^{-1}] \} \\
& \quad \{ -\theta_2^{-5} A^t W Z X_{11} Z^t W (y_{obs} - o) - \theta_2^{-3} A^t W Z X_{12} [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)] \} \\
& \hspace{15em} (A.50)
\end{aligned}$$

Tanto para el primer término, como para el segundo término de la suma, todos los elementos que componen la primera llave tienen dimensión $1 \times (k \times s)$ y todos los que componen de la segunda llave tienen dimensión $(k \times s) \times 1$. Esto me asegura que el resultado de todas las multiplicaciones van a ser escalares, por lo que procedo a multiplicar los términos

necesarios y a despejar θ_2

$$\begin{aligned}
& -2\theta_2^{-7} \frac{(y_{obs} - o)^t W Z X_{11}^t (Z^t W Z + I) X_{11} Z^t W (y_{obs} - o)}{\sigma^2} \\
& \quad - \theta_2^{-5} \frac{(y_{obs} - o)^t W Z X_{11}^t Z^t W A X_{21} Z^t W (y_{obs} - o)}{\sigma^2} \\
& -2\theta_2^{-5} \frac{(y_{obs} - o)^t W Z X_{11}^t [(Z^t W Z + I) X_{12}] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \\
& -\theta_2^{-3} \frac{(y_{obs} - o)^t W Z X_{11}^t [Z^t W A (H^{-1} + X_{22})] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \\
& \quad + 2\theta_2^{-3} \frac{(y_{obs} - o)^t W Z X_{11}^t Z^t W (y_{obs} - o)}{\sigma^2 \tau_\theta} \\
& -\theta_2^{-7} \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t (Z^t W Z + I) X_{11} Z^t W (y_{obs} - o)}{\sigma^2} \\
& \quad - \theta_2^{-5} \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t Z^t W A X_{21} Z^t W (y_{obs} - o)}{\sigma^2} \\
& -2\theta_2^{-5} \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t [(Z^t W Z + I) X_{12}] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \\
& -\theta_2^{-3} \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t [Z^t W A (H^{-1} + X_{22})] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \\
& \quad + 2\theta_2^{-3} \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t Z^t W (y_{obs} - o)}{\sigma^2} \\
& \quad + \\
& \quad - \theta^{-5} \frac{(y_{obs} - o)^t W Z X_{21}^t A^t W Z X_{11} Z^t W (y_{obs} - o)}{\sigma^2} \\
& \quad - \theta^{-3} \frac{(y_{obs} - o)^t W Z X_{21}^t A^t W Z X_{12} [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \\
& -\theta^{-5} \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] [X_{22}^t + (H^t)^{-1}] A^t W Z X_{11} Z^t W (y_{obs} - o)}{\sigma^2} \\
& -\theta^{-3} \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] [X_{22}^t + (H^t)^{-1}] A^t W Z X_{12} [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \\
& \quad - \tau_\theta \theta_2^{-3} + \tau_\theta \theta_2^{-2} (\theta_1^{-1} + z_\theta) = 0 \tag{A.51}
\end{aligned}$$

Luego, si agrupo los factores comunes:

$$\begin{aligned}
& \theta_2^{-7} \left\{ -2 \frac{(y_{obs} - o)^t W Z X_{11}^t (Z^t W Z + I) X_{11} Z^t W (y_{obs} - o)}{\sigma^2} \right. \\
& \left. - \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t (Z^t W Z + I) X_{11} Z^t W (y_{obs} - o)}{\sigma^2} \right\} \\
& + \theta_2^{-5} \left\{ - \frac{(y_{obs} - o)^t W Z X_{11}^t Z^t W A X_{21} Z^t W (y_{obs} - o)}{\sigma^2} \right. \\
& \left. - 2 \frac{(y_{obs} - o)^t W Z X_{11}^t [(Z^t W Z + I) X_{12}] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \right. \\
& \left. - \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t Z^t W A X_{21} Z^t W (y_{obs} - o)}{\sigma^2} \right. \\
& \left. - 2 \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t [(Z^t W Z + I) X_{12}] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \right. \\
& \left. - \frac{(y_{obs} - o)^t W Z X_{21}^t A^t W Z X_{11} Z^t W (y_{obs} - o)}{\sigma^2} \right. \\
& \left. - \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] [X_{22}^t + (H^t)^{-1}] A^t W Z X_{11} Z^t W (y_{obs} - o)}{\sigma^2} \right\} \\
& + \theta_2^{-3} \left\{ - \frac{(y_{obs} - o)^t W Z X_{11}^t [Z^t W A (H^{-1} + X_{22})] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \right. \\
& \left. + 2 \frac{(y_{obs} - o)^t W Z X_{11}^t Z^t W (y_{obs} - o)}{\sigma^2} \right. \\
& \left. - \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t [Z^t W A (H^{-1} + X_{22})] [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \right. \\
& \left. + 2 \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] X_{12}^t Z^t W (y_{obs} - o)}{\sigma^2} - \tau_\theta \right. \\
& \left. - \frac{(y_{obs} - o)^t W Z X_{21}^t A^t W Z X_{12} [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \right. \\
& \left. - \frac{[(y_{obs} - o)^t W^t A + \sigma^2 \tau_\beta (\hat{\beta}_{\theta_2}^t - z_\beta^t)] [X_{22}^t + (H^t)^{-1}] A^t W Z X_{12} [\sigma^2 \tau_\beta (\hat{\beta}_{\theta_2} - z_\beta) + A^t W (y_{obs} - o)]}{\sigma^2} \right\} \\
& + \tau_\theta \theta_2^{-2} (\theta_1^{-1} + z_\theta) = 0
\end{aligned} \tag{A.52}$$

Comprimiendo variables tengo entonces que:

$$a\theta_2^{-7} + b\theta_2^{-5} + c\theta_2^{-3} + d\theta_2^{-2} = 0 \tag{A.53}$$

$$\theta_2^{-7} (a + b\theta_2^2 + c\theta_2^4 + d\theta_2^5) = 0 \tag{A.54}$$

Y, como $\theta > 0$ puedo despejar obteniendo:

$$a + b\theta_2^2 + c\theta_2^4 + d\theta_2^5 = 0 \tag{A.55}$$

obteniendo una ecuación quintica de la que puedo obtener raíces para el valor de θ_2 que necesito. Donde, en particular, voy a tomar el menor valor positivo en un intento por minimizar la varianza.

A.2. Independencia entre sujetos

A.2.1. Derivada de β_i

$$L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z, v) = \frac{\|R_A(\beta_i - \hat{\beta}_{\theta_2})\|^2}{\sigma^2} + \frac{\tau_*}{2} \|\beta_i - \beta_* + v_i\|_2^2 \quad (\text{A.56})$$

$$L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z, v) = \frac{\beta_i^t R_A^t R_A \beta_i - \beta_i^t R_A^t R_A \hat{\beta}_{\theta_2} - \hat{\beta}_{\theta_2}^t R_A^t R_A \beta_i + \hat{\beta}_{\theta_2}^t R_A^t R_A \hat{\beta}_{\theta_2}}{\sigma^2} + \frac{\tau_*}{2} [(\beta_i - \beta_* + v_i)^t (\beta_i - \beta_* + v_i)] \quad (\text{A.57})$$

$$\frac{\partial L_i(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z, v)}{\partial \beta_i} = \frac{2R_A^2 \beta_i - 2R_A^2 \hat{\beta}_{\theta_2}}{\sigma^2} + \frac{\tau_*}{2} [\beta_i^t \beta_i - \beta_i^t (\beta_* - v_i) - \beta_i (\beta_* - v_i)^t + (\beta_* - v_i)^t (\beta_* - v_i)] \quad (\text{A.58})$$

$$\frac{2R_A^2 \beta_i}{\sigma^2} - \frac{2R_A^2 \hat{\beta}_{\theta_2}}{\sigma^2} + \tau_* \beta_i - \tau_* \beta_* + \tau_* v_i = 0 \quad (\text{A.59})$$

$$\frac{2R_A^2 \beta_i}{\sigma^2} + \tau_* \beta_i = \frac{2R_A^2 \hat{\beta}_{\theta_2}}{\sigma^2} + \tau_* \beta_* - \tau_* v_i \quad (\text{A.60})$$

$$\left(\frac{2R_A^2}{\sigma^2} + \tau_* I \right) \beta_i = \frac{2R_A^2 \hat{\beta}_{\theta_2}}{\sigma^2} + \tau_* \beta_* - \tau_* v_i \quad (\text{A.61})$$

$$\beta_i = \left(\frac{2R_A^2}{\sigma^2} + \tau_* I \right)^{-1} \left(\frac{2R_A^2 \hat{\beta}_{\theta_2}}{\sigma^2} + \tau_* \beta_* - \tau_* v_i \right) \quad (\text{A.62})$$

Igual que antes, la derivada de $\hat{\beta}_{\theta_2}$ se resuelve como

$$\hat{\beta}_{\theta_2} = \left(\frac{2R_A^2}{\sigma^2} + \tau_* I \right)^{-1} \left(\frac{2R_A^2 \beta_i}{\sigma^2} + \tau_* z \beta + \tau_* \hat{\beta}_{\theta_1} \right) \quad (\text{A.63})$$

Y si remplazo este resultado a partir de (A.60) obtengo

$$\frac{2R_A^2 \beta_i}{\sigma^2} + \tau_* \beta_i = \frac{2R_A^2}{\sigma^2} \left(\frac{2R_A^2}{\sigma^2} + \tau_* I \right)^{-1} \left(\frac{2R_A^2 \beta_i}{\sigma^2} + \tau_* z \beta + \tau_* \hat{\beta}_{\theta_1} \right) + \tau_* \beta_* - \tau_* v_i \quad (\text{A.64})$$

$$\frac{2R_A^2 \beta_i}{\sigma^2} + \tau_* \beta_i = \left(\frac{2R_A^2}{\sigma^2} + \tau_* I \right)^{-1} \left(\frac{4R_A^4 \beta_i}{\sigma^4} + \frac{2R_A^2}{\sigma^2} \tau_* z \beta + \frac{2R_A^2}{\sigma^2} \tau_* \hat{\beta}_{\theta_1} \right) + \tau_* \beta_* - \tau_* v_i \quad (\text{A.65})$$

$$\frac{2R_A^2 \beta_i}{\sigma^2} + \tau_* \beta_i = \left(\frac{2R_A^2}{\sigma^2} + \tau_* I \right)^{-1} \frac{4R_A^4 \beta_i}{\sigma^4} + \left(\frac{2R_A^2}{\sigma^2} + \tau_* I \right)^{-1} \frac{2R_A^2}{\sigma^2} \tau_* z \beta + \left(\frac{2R_A^2}{\sigma^2} + \tau_* I \right)^{-1} \frac{2R_A^2}{\sigma^2} \tau_* \hat{\beta}_{\theta_1} + \tau_* \beta_* - \tau_* v_i \quad (\text{A.66})$$

$$\begin{aligned} & \frac{2R_A^2 \beta_i}{\sigma^2} + \tau_* \beta_i - \left(\frac{2R_A^2}{\sigma^2} + \tau_* I \right)^{-1} \frac{4R_A^4 \beta_i}{\sigma^4} \\ &= \left(\frac{2R_A^2}{\sigma^2} + \tau_* I \right)^{-1} \left(\frac{2R_A^2}{\sigma^2} \tau_* z \beta + \frac{2R_A^2}{\sigma^2} \tau_* \hat{\beta}_{\theta_1} \right) + \tau_* \beta_* - \tau_* v_i \end{aligned} \quad (\text{A.67})$$

$$\begin{aligned}
& \left[\frac{2R_A^2}{\sigma^2} + \tau_* I - \left(\frac{2R_A^2}{\sigma^2} + \tau_\beta I \right)^{-1} \frac{4R_A^4}{\sigma^4} \right] \beta_i \\
& = \left(\frac{2R_A^2}{\sigma^2} + \tau_\beta I \right)^{-1} \left(\frac{2R_A^2}{\sigma^2} \tau_\beta z_\beta + \frac{2R_A^2}{\sigma^2} \tau_\beta \hat{\beta}_{\theta 1} \right) + \tau_* \beta_* - \tau_* v_i
\end{aligned} \tag{A.68}$$

$$\begin{aligned}
\beta_i & = \left[\frac{2R_A^2}{\sigma^2} + \tau_* I - \left(\frac{2R_A^2}{\sigma^2} + \tau_\beta I \right)^{-1} \frac{4R_A^4}{\sigma^4} \right]^{-1} \\
& \left[\left(\frac{2R_A^2}{\sigma^2} + \tau_\beta I \right)^{-1} \left(\frac{2R_A^2}{\sigma^2} \tau_\beta z_\beta + \frac{2R_A^2}{\sigma^2} \tau_\beta \hat{\beta}_{\theta 1} \right) + \tau_* \beta_* - \tau_* v_i \right]
\end{aligned} \tag{A.69}$$

lo que me permite resolver β_i y reemplazar en $\hat{\beta}_{\theta 2}$. Para obtener el valor de R_A que necesito puedo hacer lo siguiente: Dado Z matriz de bloques constantes por sujeto y asumiendo la matriz de covarianza Λ_θ diagonal por una constante θ para cada sujeto, entonces tengo que:

$$\begin{bmatrix} \Lambda_\theta^t Z^t W Z \Lambda_\theta + I_k & \Lambda_\theta^t Z^t W A \\ A^t W Z \Lambda_\theta & A^t W A \end{bmatrix} = \begin{bmatrix} L_\theta & 0 \\ R_{ZA}^t & R_A^t \end{bmatrix} \begin{bmatrix} L_\theta^t & R_{ZA} \\ 0 & R_A \end{bmatrix} \tag{A.70}$$

que, haciendo uso de la descomposición LDL de Cholesey puedo re-expresar como:

$$LL^t = LD^{1/2}(LD^{1/2})^t = LD^{1/2}(D^{1/2})^t L^t = LDL \tag{A.71}$$

entonces:

$$\begin{bmatrix} L_\theta & 0 \\ R_{ZA}^t & R_A^t \end{bmatrix} \begin{bmatrix} L_\theta^t & R_{ZA} \\ 0 & R_A \end{bmatrix} \tag{A.72}$$

$$= \begin{bmatrix} I & 0 \\ R_{ZA}^t L_\theta^{-2} & I \end{bmatrix} \begin{bmatrix} L_\theta^2 & 0 \\ 0 & R_A^2 \end{bmatrix} \begin{bmatrix} I & (L_\theta^t)^{-2} R_{ZA} \\ 0 & I \end{bmatrix} \tag{A.73}$$

a partir de este calculo, puedo obtener el valor de R_A de forma analitica dado que:

$$D_j = A_{jj} - \sum_{k=1}^{j-1} L_{jk} D_k L_{jk}^t \tag{A.74}$$

$$L_{ij} = \left(A_{ij} - \sum_{k=1}^{j-1} L_{ik} D_k L_{jk}^t \right) D_j^{-1} \text{ para } i > j \tag{A.75}$$

Luego:

$$L_\theta^2 = \Lambda_\theta^t Z^t W Z \Lambda_\theta + I_k \tag{A.76}$$

$$R_A^2 = A^t W A - R_{ZA}^t L_\theta^{-2} L_\theta^2 (L_\theta^{-2})^t R_{ZA} \tag{A.77}$$

$$R_{ZA}^t L_\theta^{-2} = A^t W Z \Lambda_\theta (\Lambda_\theta^t Z^t W Z \Lambda_\theta + I_k)^{-1} \tag{A.78}$$

Reemplazando

$$R_A^2 = A^t W A - A^t W Z \Lambda_\theta [(\Lambda_\theta^t Z^t W Z \Lambda_\theta + I_k)^{-1}]^t \Lambda_\theta^t Z^t W A \tag{A.79}$$

que me sirve para resolver el valor de β_i y poder despejar $\hat{\beta}_{\theta 2}$

A.2.2. Derivada de β_*

$$L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z, v) = \sum_{i=1}^n f_1(\theta_{1_i}) + f_2(\theta_{2_i}, x_i) + f_3(\beta_i, \hat{\beta}_{\theta_{2_i}}) \quad (\text{A.80})$$

$$+ \frac{\tau}{2} \|\theta_{1_i} - \theta_{2_i} + z_{\theta_i}\|_2^2 + \frac{\tau}{2} \|S_{\beta_i} x - \hat{\beta}_{\theta_{2_i}} + z_{\beta_i}\|_2^2 + \frac{\tau}{2} \|\beta_i - \beta_* + v_i\|_2^2$$

$$L(\theta_1, \theta_2, x, \hat{\beta}_{\theta_2}, z, v) = \sum_{i=1}^n \frac{\tau}{2} \|\beta_i - \beta_* + v_i\|_2^2 \quad (\text{A.81})$$

$$-\beta_* n + \sum_{i=1}^n \beta_i + v_i = 0 \quad (\text{A.82})$$

$$\beta_* = \frac{1}{n} \sum_{i=1}^n \beta_i + v_i \quad (\text{A.83})$$

Bibliografía

- [1] Homunculo (fig).
- [2] Lasso (statistics).
- [3] Deanna M Barch, Gregory C Burgess, Michael P Harms, Steven E Petersen, Bradley L Schlaggar, Maurizio Corbetta, Matthew F Glasser, Sandra Curtiss, Sachin Dixit, Cindy Feldt, Dan Nolan, Edward Bryant, Tucker Hartley, Owen Footer, James M Bjork, Russ Poldrack, Steve Smith, Heidi Johansen-Berg, Abraham Z Snyder, David C Van Essen, and WU-Minn HCP Consortium. Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage*, 80:169–89, 10 2013.
- [4] Peter J. Basser and Evren Özarslan. *Diffusion MRI - From Quantitative Measurement to In-vivo Neuroanatomy*. 2014.
- [5] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48, 10 2015.
- [6] Davi D. Bock, Wei-Chung Allen Lee, Aaron M. Kerlin, Mark L. Andermann, Greg Hood, Arthur W. Wetzel, Sergey Yurgenson, Edward R. Soucy, Hyon Suk Kim, and R. Clay Reid. Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, 471(7337):177–182, 3 2011.
- [7] S. Boyd, Parikh N., Chu E., Peleato B., and Eckstein J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, 3:1–122, 2011.
- [8] Kevin L. Briggman, Moritz Helmstaedter, and Winfried Denk. Wiring specificity in the direction-selectivity circuit of the retina. *Nature*, 471(7337):183–188, 3 2011.
- [9] Theodore H. Bullock, Michael V. L. Bennett, Daniel Johnston, Robert Josephson, Eve Marder, and R. Douglas Fields. The Neuron Doctrine, Redux. *Science*, 310(5749):791–793, 11 2005.
- [10] Henry Vandyke Carter. Lobes of the brain (fig).
- [11] Rahul S. Desikan, Florent Ségonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, Anders M. Dale, R. Paul Maguire, Bradley T. Hyman, Marilyn S. Albert, and Ronald J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, 2006.
- [12] Christophe Destrieux, Bruce Fischl, Anders Dale, and Eric Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1):1–15, 10 2010.
- [13] Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976.

-
- [14] Lindsey L Glickfeld, Mark L Andermann, Vincent Bonin, and R Clay Reid. Cortico-cortical projections in mouse visual cortex are functionally target specific. *Nature Neuroscience*, 16(2):219–226, 1 2013.
- [15] R. Glowinski and A. Marroco. Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 9(R2):41–76, 1975.
- [16] M. W. Halterman. Neuroscience, 3rd Edition. *Neurology*, 64(4):769–769, 2 2005.
- [17] AP Holmes and KJ Friston. Generalisability, random effects and population inference, 1 1998.
- [18] S. Jbabdi, J. F. Lehman, S. N. Haber, and T. E. Behrens. Human and Monkey Ventral Prefrontal Fibers Use the Same Organizational Principles to Reach Their Targets: Tracing versus Tractography. *Journal of Neuroscience*, 33(7):3190–3201, 2 2013.
- [19] H Johansen-Berg, T E J Behrens, M D Robson, I Drobnjak, M F S Rushworth, J M Brady, S M Smith, D J Higham, and P M Matthews. Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36):13335–40, 9 2004.
- [20] Nan M. Laird and James H. Ware. Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4):963, 12 1982.
- [21] Mary J. Lindstrom and Douglas M. Bates. Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *Journal of the American Statistical Association*, 83(404):1014, 12 1988.
- [22] David E. Osher, Rebecca R. Saxe, Kami Koldewyn, John D E Gabrieli, Nancy Kanwisher, and Zeynep M. Saygin. Structural Connectivity Fingerprints Predict Cortical Selectivity for Multiple Visual Categories across Cortex. *Cerebral Cortex*, 26(4):1668–1683, 2016.
- [23] Parikh N. and Boyd S. Proximal Algorithms. In *Foundations and Trends in Optimization*, pages 123–231. 2014.
- [24] Richard E. Passingham, Klaas E. Stephan, and Rolf Kötter. The anatomical basis of functional localization in the cortex. *Nature Reviews Neuroscience*, 3(8):606–616, 8 2002.
- [25] Wilder Penfield, Herbert H. (Herbert Henri) Jasper, and Wilder Seegal to: Penfield. *Epilepsy and the functional anatomy of the human brain*. Little, Brown, Boston, [1st ed.]. edition, 1954.
- [26] Zeynep M. Saygin, David E. Osher, Jean Augustinack, Bruce Fischl, and John D.E. Gabrieli. Connectivity-based segmentation of human amygdala nuclei using probabilistic tractography. *NeuroImage*, 56(3):1353–1361, 6 2011.

-
- [27] Zeynep M Saygin, David E Osher, Kami Koldewyn, Gretchen Reynolds, John D E Gabrieli, and Rebecca R Saxe. Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nature neuroscience*, 15(2):321–7, 12 2011.
- [28] Zeynep M Saygin, David E Osher, Kami Koldewyn, Gretchen Reynolds, John D E Gabrieli, and Rebecca R Saxe. Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nature Neuroscience*, 15(2):321–327, 12 2011.
- [29] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.
- [30] M Sur, P E Garraghty, and A W Roe. Experimentally induced visual projections into auditory thalamus and cortex. *Science (New York, N.Y.)*, 242(4884):1437–41, 12 1988.
- [31] Robert Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.
- [32] D.C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T.E.J. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S.W. Curtiss, S. Della Penna, D. Feinberg, M.F. Glasser, N. Harel, A.C. Heath, L. Larson-Prior, D. Marcus, G. Michalareas, S. Moeller, R. Oostenveld, S.E. Petersen, F. Prior, B.L. Schlaggar, S.M. Smith, A.Z. Snyder, J. Xu, E. Yacoub, and WU-Minn HCP Consortium. The Human Connectome Project: A data acquisition perspective. *NeuroImage*, 62(4):2222–2231, 10 2012.