Análisis en componentes principales como técnica de reducción de dimensionalidad aplicada al problema del control motor

Vanina Camacho Docentes: Germán Pequera, María José González, Franco Simini

Resumen—In the existing literature, techniques have been tried to understand the problem of motor control mainly in the most common types of locomotion in humans, walk and run. However, there is a locomotion that is preferred in specific situations, for example, when descending steep slopes, which is the gallop. In this work, electromyographic (EMG) signals of seven muscles concentrated in the legs collected during the human gallop are used to explain muscular activation using the recognized technique of dimensional reduction of principal component analysis (PCA). It was found that muscle activation for the human gallop is explained in four components with 99 % accuracy.

I. INTRODUCCIÓN

El sistema nervioso se encarga de controlar los movimientos del cuerpo humano de forma coordinada. Una única tarea motora puede realizarse de distintas maneras con un resultado similar. Por lo que el sistema nervioso tiene que tener una estrategia para reducir los grados de libertad para ejecutar los movimientos. Ya que el control individual de cada músculo para realizar cada tarea motora específica o sus variaciones implicaría una muy alta complejidad del cerebro[1]. En [2] se menciona que una teoría para entender cómo el sistema nervioso supera la complejidad que implica cada movimiento, podría ser la del control modular. En la que la complejidad elevada de la diversidad de los movimientos se reduce a combinar un menor número de patrones básicos de activación de los músculos para lograr una tarea motora específica[3]. También se menciona en [2] que es común que éstas regularidades en el sistema motor pueden ser representadas mediante sinergias musculares. Las sinergias musculares representan la activación de grupos de músculos como unidades individuales [4]. Los movimientos son producidos por una activación coordinada de éstas sinergias. Existe evidencia de que la selección, activación y combinación flexible de estos patrones básicos, denominados sinergias musculares, son de origen neural y siguen una organización modular[3]. Cada sinergia muscular representa un perfil invariante en el tiempo de activación a través de los músculos, activada por un coeficiente dependiente del tiempo (módulo). Cuando las sinergias y los módulos, se contemplan juntos, los patrones de activación muscular se reconstruyen totalmente^[4]. Es decir que mediante el control de un grupo de módulos y de sinergias se pueden controlar movimientos

mas complejos. Desde el punto de vista matemático esto es un problema de reducción de dimensionalidad. Por lo que se pueden aplicar técnicas como el análisis de componentes principales¹ (PCA). Es una técnica muy utilizada en distintas aplicaciones para la reducción de la dimensionalidad, para compresión y para la extracción de características.

La locomoción es un importante aspecto de la vida humana. Caminar minimiza los costos metabólicos para transportarse a bajas velocidades y el trote es preferido para altas velocidades. Estos tipos de locomoción son simétricas. Otro tipo de locomoción para los humanos que es asimétrica es el galope. Aunque no es una elección común hay situaciones en las que sí se elige espontáneamente, por ejemplo, los niños al jugar, en algunos deportes, al bajar escaleras o pendientes muy pronunciadas[5].

En [6], [3], [7] se trata este problema con sujetos caminando y/o trotando, en este trabajo se trata el problema con el galope humano como tipo de locomoción. En este trabajo se aplica PCA para la reducción de la dimensión al problema del control motor basado en la activación modular y sinergias musculares para explicar el galope humano.

II. DESARROLLO DEL MÉTODO

II-A. PCA

Se define como "la proyección ortogonal de los datos en un espacio lineal de menor dimensión, conocido como subespacio principal, de manera que la varianza de los datos proyectados se maximice"[8]. La figura 1 ilustra esta definición proyectando datos en dimensión dos en la recta definida por la dirección de máxima varianza de dimensión uno.

El objetivo de PCA radica en derivar un nuevo grupo de variables ordenadas según la importancia (definida por la cantidad de varianza que concentran). Esto puede traer la desventaja de que las nuevas características sean difíciles de interpretar.

II-A1. Método: Dado el conjunto de datos $x = \{x_n\}$ con n = 1, ..., N, donde x_n tiene dimensión D. Se quieren proyectar los datos en un espacio de dimensión p < D. Se

¹Principal component analysis en inglés



Figura 1. Ejemplo de PCA en 2 dimensiones. Imagen de. Imagen tomada de Bishop [8]

asume que los datos están centrados en el origen, de no estarlo es necesario restar la media. Se definen los datos proyectados $y = \{y_i\}$ de dimensión p están dados por la ecuación (1).

$$y_i = \sum_{j=1}^p a_{ij} x_j \qquad y = A^T x \tag{1}$$

Siendo A la matriz de coeficientes.Se quiere demostrar que las direcciones que maximizan la varianza de los datos proyectados están dadas por las direcciones de los vectores propios. En la ecuación (2) se tiene la expresión de la varianza de los datos proyectados para el caso en que se proyecta sobre la dirección a_1 ($a_1^T a_1 = 1$).

$$var\{y_1\} = E\{a_1^T x x^T a_1\} - E\{a_1^T x\} E\{x^T a_1\}$$

= $a_1^T (E\{x x^T\} - E\{x\} E\{x^T\}) a_1$ (2)
= $a_1^T \Sigma a_1$

donde se usó que los datos originales están centrados en 0 $(E\{x\} = 0)$ y que $\Sigma = E\{xx^T\}$ es la matriz de covarianza de los datos.

Se quiere entonces maximizar $var\{y_1\} = a_1^T \Sigma a_1$. Este problema se resuelve con multiplicadores de Lagrange, obteniendo que el vector a_1 debe ser vector propio de Σ asociado al valor propio λ máximo. Como a_1 es vector propio de Σ :

$$var\{y_1\} = \lambda a_1^T a_1$$

$$= \lambda$$
(3)

Se demuestra por inducción que la proyección lineal óptima, en el sentido de maximizar la varianza, sobre un espacio p-dimensional es aquella definida por p vectores propios de Σ asociados a los p mayores valores propios.

Se asumió en la descripción del método que la dimensión p era conocida. Por lo general no se conoce y hay distintas técnicas para encontrarla. En este trabajo se elige según la cantidad de vectores que capturen la mayor varianza en los datos. Esto es, que agregar componentes no cambie sustancialmente la varianza de los datos proyectados.

II-B. Datos

Se tienen datos provenientes de registros de señales electromiográficas (EMG) que representan una medida indirecta de la contracción muscular.

Se tienen señales provenientes de 7 músculos en un ciclo de galope humano brindados por el tutor². Los músculos medidos son: Sóleo, Gastrocnemio Lateral, Tibial Anterior, Bíceps Femoral, Vasto Lateral, Recto Femoral y Glúteo Máximo. En la figura 2 se muestra un diagrama de los músculos mencionados.



Figura 2. Diagrama de músculos medidos

El galope es un tipo de marcha que tiene una fase de doble apoyo (los dos pies apoyados) y una de vuelo (los dos pies en el aire). En la figura 3 se muestra un diagrama de la cinemática del galope humano. Las líneas verdes representan la pierna derecha y las rojas la izquierda. Las lineas punteadas representan las piernas en la fase de vuelo y las continuas el contacto con el piso.



Figura 3. Diagrama del galope humano. Imagen tomada de Fiers[5]

²Germán Pequera

³Imagen modificada de http://humananatomylibrary.com/

Antes de aplicar PCA (descripto en la sección II-A) es necesario pre-procesar los datos. El pre-procesamiento implica amplificar y rectificar la señal para luego filtrarla con un pasabajos sin retardo (zero-lag) de frecuencia de corte 10Hz, de forma de obtener la envolvente de la señal como se describe en [6].

II-C. PCA aplicado al problema del control motor

El objetivo principal de utilizar PCA en este problema es representar los datos EMG originales como una combinación lineal de p = 4 componentes como determina la ecuación (4).

$$E = WC^T \tag{4}$$

Siendo W los pesos o sinergias y C los módulos o componentes de activación[6], que resultan ser los datos proyectados según las direcciones dadas por PCA.

II-D. Resultados

En la figura 4 se pueden ver las EMG pre-procesadas de los 7 músculos de acuerdo a [6] junto a las señales originales. La diferencia de alturas en algunos casos se debe a que se aplico una etapa de ganancia a las señales para procesarlas. Esto no afecta al resto del trabajo.



Figura 4. Señales originales y con el pre procesamiento mencionado en [6]

Se aplica PCA a los datos pre-procesados y se grafica la varianza acumulada para determinar la cantidad de módulos que explican el galope humano (ver figura 5).

Se puede ver que a partir de 4 módulos la varianza acumulada representa el 99% de la varianza, por lo que se eligen los 4 módulos mostrados en la figura 6.

La varianza acumulada está dada por la ecuación (5).

$$\sum_{i=1}^{p} var(y_i) = \sum_{i=1}^{p} \lambda_i \tag{5}$$

Por lo que la decisión en este caso de la cantidad de módulos está relacionada directamente con los valores propios de la matriz de covarianza de los datos.



Figura 5. Varianza acumulada



Figura 6. Componentes de activación (módulos) elegidos luego de PCA

La figura 7 muestra los pesos obtenidos para los 7 músculos para las 4 componentes.

Luego se reconstruyen las señales de los 7 músculos a partir de las sinergias y los 4 módulos elegidos como la ecuación (4) y se muestran en la figura 8

Se puede ver que quedan casi completamente determinadas salvo las señales del músculo gastrocnemio lateral y del glúteo máximo donde se aprecian algunas diferencias con la señal original. Esta diferencia se hace menor al agregar mas módulos hasta ser nula si se eligen los 7 módulos posibles.

En el trabajo de [6] se encuentra que tanto la caminata y el trote se pueden explicar a través de 5 módulos, pero parten de 32 señales de músculos. La cantidad de señales que se tienen inicialmente incide en la cantidad de módulos que se eligen.

III. CONCLUSIONES

En particular para el caso del galope como tipo de locomoción humana. Se encontró que puede explicarse a partir de 4 componentes o módulos con un error bajo. El método no se pudo probar en señales de trote y caminata por no tenerlas disponibles, pero se puede ver como la teoría se aplica bien



Figura 7. Pesos (Sinergias) de los 7 músculos para las 4 componentes



Figura 8. Reconstrucción de las señales a partir de los 4 módulos y sinergias elegidos

para el problema del galope, ya que se puede reducir de 7 señales a 4 con sus respectivas sinergias para representar las señales de activación muscular medidas.

REFERENCIAS

- [1] Nikolaj A Bernstein. The co-ordination and regulation of movements. 1967.
- [2] Matthew C Tresch, Vincent CK Cheung, and Andrea d'Avella. Matrix factorization algorithms for the identification of muscle synergies: evaluation on simulated and experimental data sets. *Journal of neurophysiology*, 95(4):2199–2212, 2006.
- [3] Lars Janshen, Alessandro Santuz, Antonis Ekizos, and Adamantios Arampatzis. Modular control during incline and level walking in humans. *Journal of experimental biology*, 220(5):807–813, 2017.
- [4] Vincent CK Cheung, Andrea Turolla, Michela Agostini, Stefano Silvoni, Caoimhe Bennis, Patrick Kasi, Sabrina Paganoni, Paolo Bonato, and Emilio Bizzi. Muscle synergy patterns as physiological markers of motor cortical damage. *Proceedings of the National Academy of Sciences*, 109(36):14652–14656, 2012.
- [5] Pieter Fiers, Dirk De Clercq, Veerle Segers, and Peter Aerts. Biomechanics of human bipedal gallop: asymmetry dictates leg function. *Journal* of Experimental Biology, 216(7):1338–1349, 2013.

- [6] Maria Domenica Cappellini, Alan Cohen, Antonio Piga, Mohamed Bejaoui, Silverio Perrotta, Leyla Agaoglu, Yesim Aydinok, Antonis Kattamis, Yurdanur Kilinc, John Porter, et al. A phase 3 study of deferasirox (icl670), a once-daily oral iron chelator, in patients with β-thalassemia. *Blood*, 107(9):3455–3462, 2006.
- [7] Nadia Dominici, Yuri P Ivanenko, Germana Cappellini, Andrea d'Avella, Vito Mondì, Marika Cicchese, Adele Fabiano, Tiziana Silei, Ambrogio Di Paolo, Carlo Giannini, et al. Locomotor primitives in newborn babies and their development. *Science*, 334(6058):997–999, 2011.
- [8] Christopher M Bishop. Pattern recognition. *Machine Learning*, 128:1–58, 2006.
- [9] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.