ANÁLISIS DE GEODÉSICAS PRINCIPALES SOBRE EL ESPACIO DE LAS MATRICES DE COVARIANZA PARA LA DESCRIPCIÓN DE ACCIONES EN VIDEO

SANTIAGO NIÑO CAMPOS

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
BUCARAMANGA
2021

ANÁLISIS DE GEODÉSICAS PRINCIPALES SOBRE EL ESPACIO DE LAS MATRICES DE COVARIANZA PARA LA DESCRIPCIÓN DE ACCIONES EN VIDEO

SANTIAGO NIÑO CAMPOS

Trabajo de Grado para optar al título de
Matemático

Director
Fabio Martínez Carrillo
Ph.D en Ingeniería de Sistemas y Computación

Codirector
Juan Galvis
Ph.D en Matemáticas

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
BUCARAMANGA
2021

AGRADECIMIENTOS

El autor expresa su agradecimiento:

En primer lugar a Dios de quien proceden la sabiduría, inteligencia y ciencia.

A mis padres, Abdón y Lucia, por su inmenso amor y apoyo único e incondicional, quienes me enseñaron que con esfuerzo, disciplina y valentía lograré alcanzar las metas que me he propuesto, por mí, por ellos.

A los profesores Fabio Martínez y Juan Galvis. Muchas gracias por confíar en mí para desarrollar este trabajo, por su paciencia, dedicación, orientación y formación profesional y personal. Profesores que me brindaron una excelente guía y de quienes admiro su pasión y entrega por la ciencia y la academia. Sus enseñanzas son de un valor incalculable para mi vida y sin sus consejos y dirección no habría sido posible la culminación de este trabajo.

Al matemático, compañero y amigo Juan Andrés Olmos por enseñarme y motivarme a iniciar en este nuevo camino del saber, por brindarme su apoyo, ayuda y conocimiento cuando lo necesité.

A mi mejor amiga Laura Sofía por su apoyo en todo momento, su amistad y cariño, quien siempre busca lo mejor de mí y me impulsa a continuar adelante sin importar las adversidades demostrandome lo orgullosa que está de mí.

Al grupo de investigación BivL^{2}ab de la escuela de ingeniería de sistemas por acogerme y enseñarme nuevas ramas de la ciencia, a la escuela de matemáticas, todos sus docentes y en general, a la Universidad Industrial de Santander (UIS) por toda la formación integra y profesional que me han brindado.

A todos aquellos que de alguna u otra forma contribuyeron en este trabajo.

Muchas gracias.

CONTENIDO

pág.

INTRODUCCIÓN **12**

1. OBJETIVOS **16**

2. PRELIMINARES **17**

2.1. EL ESPACIO DE LAS MATRICES DE COVARIANZA 17

2.2. GEOMETRÍA RIEMANNIANA 19

2.3. ANÁLISIS DE COMPONENTES PRINCIPALES 29

3. ANÁLISIS DE GEODÉSICAS PRINCIPALES **37**

3.1. VARIANZA 37

3.2. SUBVARIEDAD GEODÉSICA 38

3.3. PROYECCIÓN 39

3.4. ANÁLISIS DE GEODÉSICAS PRINCIPALES 42

4. ALGORITMOS ESTUDIADOS **44**

4.1. ALGORITMO DE FLETCHER 44

4.2. ALGORITMO DE YUCHEN XIE 47

5. MODELO PROPUESTO COMO DESCRIPTOR DE VIDEO **51**

5.1. DESCRIPTOR DE COVARIANZA POR CUADRO 51

5.2. DESCRIPTOR DE VIDEO USANDO COMPONENTES PRINCIPALES 54

6. CONFIGURACIÓN EXPERIMENTAL **58**

6.1. BASE DE DATOS 58

6.2. FLUJO ÓPTICO Y REDES NEURONALES CONVOLUCIONALES 59

6.3. CLASIFICACIÓN 61
6.4. VALIDACIÓN EXPERIMENTAL 62

7. EVALUACIÓN Y RESULTADOS 63

8. CONCLUSIONES 70

BIBLIOGRAFÍA 72

ANEXOS 75

LISTA DE FIGURAS

pág.

Figura 1. Visualización de direcciones principales en dos conjuntos dadas por el ACP. 14

Figura 2. Representación del espacio $PD(2)$ en $\mathbb{R}^3$. 19

Figura 3. Interpretación geométrica de las aplicaciones exponencial y logaritmo. 24

Figura 4. Proyección de un punto x_i sobre un subespacio lineal v_1. 30

Figura 5. Proyección de el punto x_i sobre un subespacio lineal v_2 ortogonal a v_1. 31

Figura 6. Geodésicas entre A y B en S^2 (curva azul) y $\mathbb{R}^3$ (línea roja). 38

Figura 7. Aproximación del operador $\pi_H(x)$. 39

Figura 8. Ilustración de la aproximación en la esfera unitaria S^2. 41

Figura 9. Ilustración algoritmo de Fletcher. 44

Figura 10. Ilustración del algoritmo Fletcher en el espacio $PD(2)$. 46

Figura 11. Ilustración algoritmo de Yuchen. 48

Figura 12. Ilustración del algoritmo Yuchen en el espacio $PD(2)$. 49

Figura 13. Comparación entre componentes obtenidas por los algoritmos. 50

Figura 14. Representación del descriptor de covarianza. 52

Figura 15. Ejemplo del cálculo de la matriz de covarianza a una imagen. 52

Figura 16. Modelo experimental, proceso inicial. 55

Figura 17. Modelo experimental del algoritmo de Fletcher. 56

Figura 18. Modelo experimental del algoritmo de Yuchen. 57

Figura 19. Primera secuencia del conjunto de datos UT-Interaction. 59

Figura 20. Ejemplo del flujo óptico calculado a un video de la secuencia 1. 64

Figura 21. Gráficas número de componentes vs varianza respectiva para cada clase del conjunto de datos. 66

Figura 22. Gráficas número de componentes vs logaritmo de la varianza respectiva para cada clase del conjunto de datos. 67

Figura 23. Ejemplo de las 15 características seleccionadas para una imagen con MobileNetV2. 69

Figura 24. Comparación de descriptores. 75

Figura 25. Resultados para las cinco diferentes formas de clasificar. 78

LISTA DE TABLAS

pág.

Tabla 1. Resultados del experimento inicial. 64

Tabla 2. Resultados del experimento con las características dadas por la MobileNetV2. 69

LISTA DE ANEXOS

pág.

Anexo A. ¿Es conveniente adjuntar componentes a la media para la descripción? 75

Anexo B. ¿Es buena la forma en que se está clasificando? 77

RESUMEN

TÍTULO: ANÁLISIS DE GEODÉSICAS PRINCIPALES SOBRE EL ESPACIO DE LAS MATRICES DE COVARIANZA PARA LA DESCRIPCIÓN DE ACCIONES EN VIDEO. *

AUTOR: SANTIAGO NIÑO CAMPOS **

PALABRAS CLAVE: VARIANZA, PROYECCIÓN, DIRECCIONES PRINCIPALES, SUBVARIEDAD GEODÉSICA.

DESCRIPCIÓN:

En el análisis de video las matrices de covarianza han sido una forma compacta y robusta de representar acciones u actividades cotidianas, tales representaciones suelen ser redundantes y pueden ser expresadas de una forma más compacta utilizando elementos característicos como su media y sus direcciones de mayor variación. El Análisis de Componentes Principales (ACP) es un proceso estadístico que permite reducir la dimensión de un conjunto de datos mediante proyecciones sobre las direcciones de mayor varianza, este proceso en espacios Euclidianos es de gran utilidad, sin embargo, en espacios curvos este método resulta ineficiente ya que tales direcciones no necesariamente son lineales. En este trabajo se estudia el espacio de las matrices de covarianza formulado como una variedad Riemanniana equipado de una métrica conocida como afín-invariante la cual nos otorga propiedades importantes y necesarias como la existencia y unicidad de geodésicas, a su vez se estudia desde un enfoque geométrico el ACP generalizando los conceptos de varianza, subvariedad geodésica y proyección. Todo ello permite formular el análisis de geodésicas principales (AGP) como una aproximación mediante el espacio tangente a la variedad, de allí se estudian dos algoritmos que computan este cálculo. De lo anterior se propone un descriptor dado por la concatenación de la media y las direcciones de mayor varianza mediante el uso de dos algoritmos: Fletcher y Yuchen. En el experimento inicial se obtienen una exactitud del $67.79\,\%$ para el algoritmo de Fletcher y un $68.84\,\%$ en el algoritmo de Yuchen, en este último se evidencia un menor número de componentes utilizadas.

---

* Trabajo de grado

** Facultad de Ciencias. Escuela de Matemáticas. Director: Fabio Martínez Carrillo, Ph.D. Codirector: Juan Galvis, Ph.D.

ABSTRACT

TITLE: PRINCIPAL GEODESIC ANALYSIS ONTO THE SPACE OF COVARIANCE MATRICES FOR DESCRIPTION OF ACTIONS IN VIDEO. *

AUTHOR: SANTIAGO NIÑO CAMPOS **

KEYWORDS: VARIANCE, PROJECTION, PRINCIPAL DIRECTIONS, GEODESIC SUBMANIFOLD.

DESCRIPTION:

In video analysis covariance matrices have been a compact and robust way to represent actions or daily activities, such representations are usually redundant and can be expressed in a more compact way using features elements such as their mean and their directions of greatest variation. The Principal Component Analysis (PCA) is a statistical process that allows reducing the dimension of a data set through projections on the directions of greatest variance, this process in Euclidean spaces is very useful, however, in curved spaces this method is inefficient since such directions are not necessarily linear. In this work it studies the space of the covariance matrices formulated as a Riemannian manifold equipped with a metric known as affine-invariant which gives us important and necessary properties such as the existence and uniqueness of geodesics, in turn, the PCA is studied from a geometric approach generalizing the concepts of variance, geodesic submanifold and projection. All this allows formulating the Principal Geodesic Analysis (PGA) as an approximation through the tangent space to the manifold, from there two algorithms that compute this calculation are studied. From the above, a descriptor is proposed given by the concatenation of the mean and the directions with the greatest variance through the use of two algorithms: Fletcher and Yuchen. In the initial experiment, an accuracy of $67.79\,\%$ is obtained for the Fletcher algorithm and $68.84\,\%$ for the Yuchen algorithm, the latter there is evidence of a lower number of components used.

---

* Bachelor Thesis

** Faculty of Science. School of Mathematics. Advisor: Fabio Martínez Carrillo, Ph.D. Co-advisor: Juan Galvis, Ph.D.

INTRODUCCIÓN

Las secuencias de video son una fuente de información relevante para la captura de patrones espaciales y dinámicos, permitiendo describir nociones de alto nivel como lo es el reconocimiento de acciones. Lo anterior ha llevado al aumento en su uso y aplicación en actividades cotidianas, dando lugar a una alta demanda en diferentes áreas como la vigilancia, los deportes, la biología entre muchos otros[1].

No obstante, el análisis de video, modelamiento de información multidimensional y cuantificación de patrones es una tarea compleja. De hecho se hace necesario mejorar la representación eficiente que capture de forma compacta las propiedades importantes de la información contenida en una secuencia. Particularmente, las representaciones de video son redundantes y pueden ser expresadas como combinaciones lineales de otras dimensiones. Un método clásico muy usado y que aprovecha este hecho es el Análisis de Componentes Principales (ACP o PCA por su nombre en inglés) introducida por Pearson a finales del siglo XIX y estudiada por Hotelling en los años 30 del siglo XX [2].

La idea principal del ACP es reducir la dimensionalidad del conjunto de datos en variables incorrelacionadas entre sí, conservando en lo más posible la varianza original del conjunto. Este principio se consigue proyectando los datos sobre las direcciones en que varía el conjunto. Tales direcciones se conocen como componentes principales donde las primeras componentes conservan la mayor variación del conjunto original. El estudio del ACP ha mostrado que el problema de encontrar tales

---

[1] Hà Quang MINH y Vittorio MURINO. *Algorithmic Advances in Riemannian Geometry and Applications For Machine Learning, Computer Vision, Statistics, and Optimization.* Springer International, 2016.

[2] I.T. JOLLIFFE. *Principal Component Analysis (2nd Ed.)* Springer series in statistics, 2002.

direcciones junto con sus respectivas magnitudes de varianza, se reduce a determinar los vectores y valores propios de la matriz de covarianza del conjunto de datos[3]. Las representaciones logradas con el método del ACP han resultado efectivas en múltiples tareas que trabajan datos multivariantes, permitiendo, entre otras, el procesamiento de imágenes y el reconocimiento de acciones en video. Este procesamiento es en general realizado sobre datos crudos, que co-existen en espacios Euclidianos, sin embargo los cálculos de momentos estadísticos no necesariamente pueden ser operados de forma convencional debido a la estructura que conforma el espacio donde residen estos datos [4]. Por otra parte, recientes trabajos han utilizado esquemas compactos de procesamiento de características, utilizando descriptores de covarianza [5], cuya idea principal es modelar la imagen mediante correlaciones entre distintas características de bajo nivel por ejemplo color, gradiente de intensidad, patrones texturales, entre muchos otros. En estos enfoques, se hace primero una representación compacta de cada instancia (imagen o video) por medio de la matriz de covarianza de diferentes características para luego ser pos-procesadas en tareas relacionadas como clasificación, reconocimiento, seguimiento, entre otras.

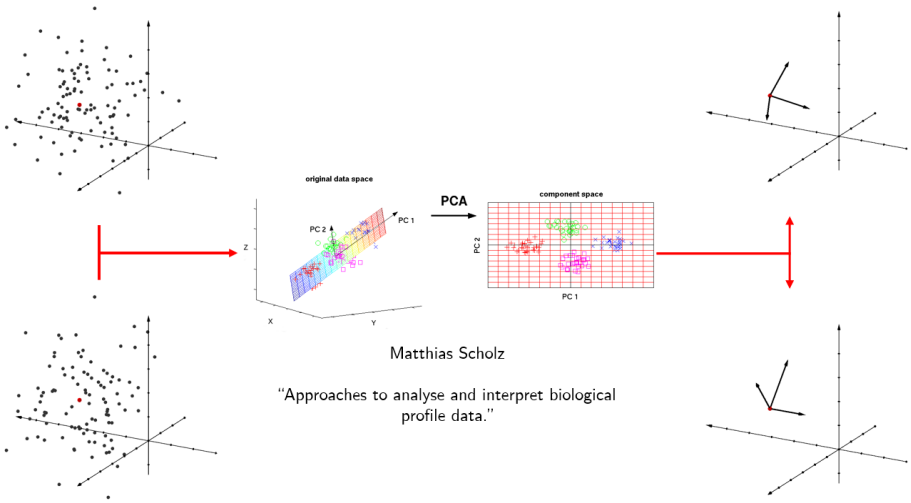
---

[3] César SÁNCHEZ. *Análisis de componentes principales*. Máster en Técnicas Estadísticas. Universidade de Santiago de Compostela, 2008-2009.

[4] P. Thomas FLETCHER y Sarang JOSHI. "Riemannian Geometry for the Statistical Analysis of Diffusion Tensor Data". En: *National Alliance for Medical Image Computing (NAMIC)* 87 (feb. de 2007), págs. 250 -262.

[5] Hà Quang MINH y Vittorio MURINO. *Covariances in Computer Vision and Machine Learning*. Morgan & Claypool Publishers, 2018.

Figura 1. Visualización de direcciones principales en dos conjuntos dadas por el ACP.



Un video puede considerarse como una secuencia de imágenes donde cada una de ellas tiene dimensión, digamos $H \times W$ y pueden ser representadas mediante un conjunto de k características como las ya mencionadas: color, gradiente de intensidad, patrones texturales, entre otras; las cuales poseen dimensión $H \times W$ que mediante una reordenación podemos obtener una expresión equivalente de tamaño $M = H \cdot W$. De esta manera, si tenemos una secuencia de video y representamos cada imagen mediante una matriz de covarianza que represente la información de las k características, llegamos entonces al problema de describir este conjunto de una forma óptima, similar al objetivo del ACP. No obstante, el espacio donde pertenecen las matrices de covarianza no forma un espacio vectorial con lo cual el uso del ACP original no es un enfoque efectivo, pues ignora propiedades especiales que poseen las matrices de covarianza, lo que terminaría en una mala descripción del conjunto de datos[4,6].

---

[6] Inbal HOREV, Florian YGER y Masashi SUGIYAMA. "Geometry-Aware Principal Component Analysis for Symmetric Positive Definite Matrices". En: *Asian Conference on Machine Learning*.

Adicionalmente, la estructura del espacio de las matrices de covarianza ha dificultado particularmente estudios estadísticos, pese a ello se han obtenido resultados en el cálculo de la media lo cual ha permitido describir conjuntos de matrices de covarianza. Aunque esta medida es un avance en la representación de los elementos sobre este espacio sigue siendo insuficiente para una descripción óptima. Podría pensarse entonces en acompañar la media por las direcciones en que varían los datos junto con sus respectivas magnitudes de varianza, una característica del ACP. No muchos estudios se han hecho sobre la generalización del ACP a este espacio, tal generalización se ha denominado Análisis de Geodésicas Principales (AGP) y se quiere profundizar en su estudio e impacto que tiene en aplicaciones como el análisis de video.

A partir de la ejecución del presente trabajo de grado, se busca estudiar y analizar el principal interrogante: ¿De qué forma se aplica la generalización del ACP al espacio de las matrices definidas positivas? De esta manera abordaremos el concepto del AGP mediante un enfoque geométrico-descriptivo del ACP. A su vez, se estudiará el espacio de las matrices de covarianza para que, de una forma natural, se puedan obtener los resultados que nos ofrece el ACP. Esto nos permitirá comprender nociones estadísticas y operaciones aplicables a este espacio no Euclidiano, con lo cual se darán herramientas más robustas en el problema de la descripción de video. Este trabajo parte de experiencias previas realizadas en el grupo de investigación donde se ha estudiado el cálculo de la media en un conjunto de matrices de covarianza. Estos trabajos son el soporte y base de conocimiento para el desarrollo de esquemas de descripción de video utilizando los componentes principales sobre el espacio de matrices de covarianza.

Vol. 45. Proceedings of Machine Learning Research. Hong Kong, 2016, págs. 1-16.

1. OBJETIVOS

Objetivo general

Estudiar el análisis de geodésicas principales sobre un conjunto de matrices de covarianza para el análisis de video.

Objetivos específicos

- Estudiar los conceptos y propiedades del análisis de componentes principales en el espacio Euclidiano $\mathbb{R}^n$.
- Establecer la estructura del espacio de las matrices simétricas definidas positivas y sus características.
- Estudiar el algoritmo presentado en *Principal Geodesic Analysis on Symmetric Spaces: Statistics of Diffusion Tensors* de P. Thomas Fletcher and Sarang Joshi.
- Implementar el algoritmo mediante ejemplos de baja dimensión, observando sus propiedades en una aplicación de clasificación de acciones.

2. PRELIMINARES

En este capítulo se presentan el marco teórico y los conceptos fundamentales relacionados con el trabajo. En este sentido se desarrollan temas y se exponen los resultados previos de geometría Riemanniana, y el proceso clásico del ACP necesarios para la construccion teórica que sigue en la generalización del mismo.

2.1. EL ESPACIO DE LAS MATRICES DE COVARIANZA

Iniciaremos estudiando las matrices de covarianza y el espacio en que ellas están expresadas. Para una matriz X de k variables con n mediciones

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}$$

donde cada x_i^T es un vector columna $k \times 1$ que contiene los valores de las k variables sobre la medición i. Se define la matriz de covarianza asociada a la matriz X como

$$S_X = \frac{1}{n} \sum_{i=1}^{n} (x_i^T - \overline{x})(x_i^T - \overline{x})^T, \tag{1}$$

donde $\overline{x} = (\overline{y_1},\ ...,\ \overline{y_k})^T$ representa el vector media de X donde cada componente i de $\overline{x}$ es la media de la variable i. La componente ij de la matriz S_X representará la covarianza entre las variables i y j con $1 \leq i, j \leq k$. Observe que si la media de X es cero su matriz de covarianza viene dada por

$$S_X = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T.$$

En la sección 5.1 se ilustrará mediante un ejemplo el cálculo de la matriz de covarianza a un conjunto de datos.
Las propiedades mas relevantes que permiten caracterizar el espacio donde coexisten las matrices de covarianza proceden del hecho que la matriz de covarianza S_X es una matriz

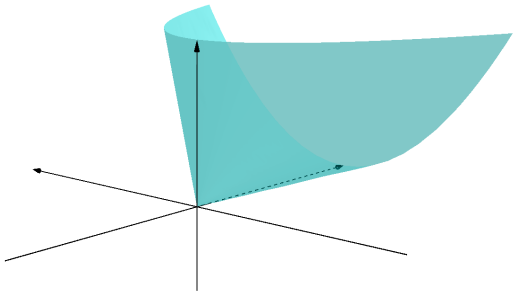
1. Cuadrada de tamaño $k \times k$,
2. Simétrica, $S_X = S_X^T$,
3. Semidefinida positiva, $y^T S_X y \geq 0$ para todo $y \in \mathbb{R}^n$ no nulo.

De la tercera propiedad tenemos que los valores propios de S_X son mayores o iguales a cero. Para evitar posteriores errores numéricos en cálculos nos interesa trabajar con matrices simétricas definidas positivas esto es, matrices cuyos valores propios sean estrictamente mayores a cero, para ello se aplica una regularización a nuestra matriz de covarianza S_X. Sea $\epsilon > 0$, si $v,\ \lambda$ son el vector y valor propio de una matriz A entonces $v,\ \lambda+\epsilon$ son el vector y valor propio de la matriz $A+diag(\epsilon)$. De esta manera, sumaremos una matriz $diag(\epsilon)$ con $\epsilon > 0$ a nuestra matriz de covarianza S_X con lo cual aseguramos que sus valores propios serán estrictamente mayores a cero. Denotaremos así, al espacio de todas las matrices $n \times n$ simétricas definidas positivas como $PD(n)$.
Por ejemplo, si consideramos el espacio $PD(2)$, una matriz $M \in PD(2)$ es de la forma

$$M = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad ac - b^2 > 0, \quad a > 0,$$

de esta manera podemos representar a M como un punto $(a, b, c) \in \mathbb{R}^3$ tal que $ac - b^2 > 0$ y $a > 0$, con lo cual tenemos una representación de $PD(2)$ en $\mathbb{R}^3$ (ver figura 2).

Figura 2. Representación gráfica del espacio $PD(2)$ como el interior de un cono en $\mathbb{R}^3$.



2.2. GEOMETRÍA RIEMANNIANA

El espacio $PD(n)$ es un tipo particular de variedad descrita como un espacio Riemanniano, concepto que construiremos a partir de los siguientes resultados matemáticos, los cuales proceden de [7,8,9,10,11].

Definición 1 *Una variedad diferenciable de dimensión n es un espacio topológico M, Hausdorff, $2-$numerable, y una familia de aplicaciones inyectivas $x_\alpha : U_\alpha \subset \mathbb{R}^n \to M$ de conjuntos abiertos U_α de $\mathbb{R}^n$ en M tal que:*

1. $\bigcup_\alpha x_\alpha(U_\alpha) = M$.

---

[7] William M. BOOTHBY. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Second Edition. ACADEMIC PRESS, INC., 1986.

[8] Sigurdur HELGASON. *Differential Geometry, Lie Groups, and Symmetric Spaces*. ACADEMIC PRESS, INC., 1978.

[9] Manfredo DO CARMO y Francis FLAHERTY. *Riemannian Geometry*. Birkhäuser, 1992.

[10] John M. LEE. *Introduction to Smooth Manifolds*. Second Edition. Springer, 2012.

[11] Jürgen JOST. *Riemannian Geometry and Geometric Analysis*. Seventh Edition. Springer, 2017.

2. *Para cada par* α, β, *con* $x_\alpha(U_\alpha) \cap x_\beta(U_\beta) = W \neq \emptyset$, *los conjuntos* $x_\alpha^{-1}(W)$ *y* $x_\beta^{-1}(W)$ *son conjuntos abiertos de* $\mathbb{R}^n$ *y la aplicación* $x_\beta^{-1} \circ x_\alpha$ *es diferenciable de clase* C^∞.

3. *La familia* $\{(U_\alpha,\ x_\alpha)\}$ *es maximal respecto a (1) y (2).*

El par $(U_\alpha,\ x_\alpha)$ es llamado una carta coordenada (o carta) en M, al conjunto U_α lo llamaremos un dominio coordenado, la familia $\{(U_\alpha,\ x_\alpha)\}$ que satisface la condiciones (1) y (2) se denomina una estructura diferenciable en M. El ítem (3) es una condición técnica, si (U, x) es una carta tal que $x \circ x_\alpha^{-1}$ y $x_\alpha \circ x^{-1}$ son de clase C^∞ para todo α entonces $(U, x) \in \{(U_\alpha,\ x_\alpha)\}$.
Como consecuencia de la anterior definición podemos extender la idea de funciones diferenciables entre variedades diferenciables

Definición 2 *Sean* M_1 *y* M_2 *variedades diferenciables de dimensión* n *y* m *respectivamente. Una función* $\varphi : M_1 \to M_2$ *es diferenciable en* $p \in M_1$ *si dada una parametrización* $\gamma_2 : V \subset \mathbb{R}^m \to M_2$ *en* $\varphi(p)$ *existe una parametrización* $\gamma_1 : U \subset \mathbb{R}^n \to M_1$ *en* p *tal que* $\varphi(\gamma_1(U)) \subset \gamma_2(V)$ *y la función*

$$\gamma_2^{-1} \circ \varphi \circ \gamma_1 : U \subset \mathbb{R}^n \to \mathbb{R}^m,$$

es diferenciable en $\gamma_1^{-1}(p)$.

De igual manera la idea de vector tangente a una variedad diferenciable puede extenderse como sigue. Dada una variedad diferenciable M, una función diferenciable $\alpha : (-\epsilon, \epsilon) \to M$ es llamada una curva (diferenciable) en M. Un vector tangente en $p \in M$ es un vector tangente en $t = 0$ de alguna curva $\alpha : (-\epsilon, \epsilon) \to M$ con $\alpha(0) = p \in M$. Llamaremos al conjunto de todos los vectores tangentes a un punto $p \in M$ como el espacio tangente a M en p y lo denotaremos por T_pM. El espacio tangente T_pM por su definición es un espacio vectorial el cual puede dotarse de un producto interno[9] (forma simétrica, bilinear y definida positiva).

Definición 3 *Una forma bilinear en un espacio vectorial V sobre $\mathbb{R}$ es una función $\phi : V \times V \to \mathbb{R}$ que satisface*

- $\phi(\alpha u_1 + \beta u_2, v) = \alpha\phi(u_1, v) + \beta\phi(u_2, v)$,
- $\phi(u, \alpha v_1 + \beta v_2) = \alpha\phi(u, v_1) + \beta\phi(u, v_2)$.

Una forma bilinear es llamada simétrica si $\phi(u, v) = \phi(v, u)$ y definida positiva si $\phi(u, u) \geq 0$ con igualdad si y solo si $u = 0$, en este caso, llamaremos a ϕ un producto interno en V.
Con la idea de espacio tangente se pueden extender a variedades diferenciables la noción del diferencial de una función diferenciable [9].

Proposición 1 *Sean M_1 y M_2 variedades diferenciables de dimensión n y m respectivamente, y sea $\varphi : M_1 \to M_2$ una función diferenciable. Para todo $p \in M_1$ y para cada $v \in T_pM_1$ sea $\alpha : (-\epsilon, \epsilon) \to M_1$ tal que $\alpha(0) = p$ y $\alpha'(0) = v$. Defina $\beta = \varphi \circ \alpha$. La función $d\varphi_p : T_pM_1 \to T_{\varphi(p)}M_2$ dada por $d\varphi_p(v) = \beta'(0)$ es una función lineal que no depende de la elección de α.*

La función lineal $d\varphi_p$ es llamada el diferencial de φ en p, lo que nos permite definir el concepto de subvariedad diferenciable.

Definición 4 *Sean H y M variedades diferenciables de dimensión n y m respectivamente. Una función diferenciable $\varphi : H \to M$ se dice una inmersión si $d\varphi_p : T_pH \to T_{\varphi(p)}M$ es inyectiva para todo $p \in H$. Si además φ es un homeomorfismo sobre $\varphi(H) \subset M$ donde $\varphi(H)$ tiene la topología inducida por M, decimos que φ es una incrustación. Si $H \subset M$ y la inclusión $i : H \subset M$ es una incrustación se dice que H es una subvariedad diferenciable de M.*

Observe que si $\varphi : H \to M$ es una inmersión entonces $n \leq m$. No toda subvariedad será útil en el presente trabajo, en la sección 3.2 se dará explicación del porqué y una caracterización sobre ellas.

El espacio tangente también permite tener la noción de medida sobre una variedad diferenciable M mediante una métrica conocida como métrica Riemanniana la cual es una correspondencia que asocia a cada punto de $p \in M$ un producto interno $\langle \cdot, \cdot \rangle_p$ en el espacio tangente T_pM que varia de forma diferenciable con el punto p. Entenderemos entonces una variedad Riemanniana como una variedad diferenciable con una métrica Riemanniana asociada.

Un espacio vectorial dotado de un producto interno nos permite definir la longitud de un vector, de esta manera si $u \in T_pM$ entonces $||u|| = \sqrt{\langle u, u \rangle_p}$ además, si $\alpha : [a, b] \subset I \to M$ es una curva de un abierto I restringida a un intervalo cerrado sobre M, se define la longitud de la curva como

$$L(\alpha) = \int_a^b ||\alpha'(t)|| \, dt.$$

Lo anterior nos permite definir la distancia entre dos puntos sobre una variedad Riemanniana

Definición 5 *La distancia Riemanniana entre dos puntos $A, B \in M$ inducida por la métrica Riemanniana $\langle \cdot, \cdot \rangle_p$ se define como*

$$d(A, B) = \text{ínf}\{L(\alpha) \mid \alpha : [a, b] \to \mathcal{M} \text{ es diferenciable por partes y } \alpha(a) = A, \alpha(b) = B\}. \tag{2}$$

Particularmente, una geodésica es una curva que minimiza la distancia entre dos puntos, más formalmente se define como

Definición 6 *Una curva parametrizada $\alpha : I \to M$ es una geodésica si $\frac{D}{dt}(\alpha'(t)) = 0$ para todo $t \in I$, esto es, la derivada covariante del vector tangente a la curva es cero a lo largo de la curva.*

Si $\alpha : I \to M$ es una geodésica entonces la longitud del vector tangente $\alpha'(t)$ es constante. [9]

Teorema 1 [11] *Sea M una variedad Riemanniana, $p \in M$ y $v \in T_pM$. Entonces existe $\epsilon > 0$ y una única geodésica $\alpha : [0, \epsilon] \to M$ con $\alpha(0) = p$ y $\alpha'(0) = v$ donde α depende diferenciablemente de p y v.*

Se introducen a continuación dos aplicaciones importantes para el trabajo sobre variedades Riemannianas.

Definición 7 *Sea M una variedad Riemanniana, $p \in M$,*

$$V_p = \{v \in T_pM : \alpha_v \text{ es definida en } [0, 1]\},$$
$$Exp_p : V_p \to M$$
$$v \to \alpha_v(1)$$

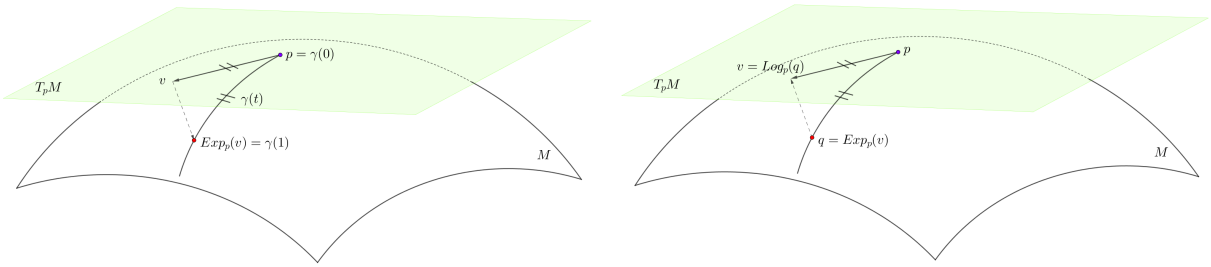
es llamada la aplicación exponencial de M en p.

Note que $||\alpha'_v(t)||$ es contante a lo largo de la geodésica $\alpha_v(t)$ entonces la longitud L desde $\alpha_v(0)$ hasta $\alpha_v(1)$ es

$$L(\alpha_v) = \int_0^1 ||\alpha'_v(t)||\, dt = \int_0^1 ||v||\, dt = ||v||.$$

Geométricamente, $Exp_p(v)$ es el punto en la única geodésica $\alpha_v(t)$ determinada por v cuya distancia desde p a lo largo de la geodésica es la longitud de v [7]. La aplicación exponencial Exp_p es una difeomorfismo de una vecindad de $0 \in T_pM$ sobre una vecindad de $p \in M$ [11], esto implica la existencia de una vecindad de $p \in M$ que puede mapearse difeomorfamente a una vecindad de $0 \in T_pM$ con lo cual se define la aplicación inversa de la aplicación exponencial llamada aplicación logaritmo $Log_p = Exp_p^{-1}$, si $Exp_p(v) = q \in M$ entonces $Log_p(q) = v \iff Exp_p(v) = q$.

Figura 3. Interpretación geométrica de las aplicaciones exponencial y logaritmo. En la figura de la izquierda, Exp_p mapea un elemento $v \in T_pM$ a M mientras que en la figura de la derecha, aplicando la función inversa Log_p lleva de vuelta el elemento en M a T_pM.



Una variedad se dice (geodésicamente) completa si para todo $p \in M$ la aplicación exponencial Exp_p se define en todo T_pM, en tal caso escribimos $Exp_p : T_pM \to M$, de esta manera cualquier geodésica $\alpha(t)$ con $\alpha(0) = p$ es definida para todo $t \in \mathbb{R}$. Nos interesa ahora estudiar la forma de obtener estas propiedades locales de manera global, iniciaremos presentando el teorema de Hopf–Rinow [9,11].

Teorema 2 *(**Hopf–Rinow**) Sea M una variedad Riemanniana, las siguientes afirmaciones son equivalentes*

1. *Los subconjuntos cerrados y acotados de M son compactos,*
2. *M con la distancia definida en (2) es un espacio métrico completo,*
3. *M es geodésicamente completo.*

Adicionalmente, cualquiera de las afirmaciones del teorema anterior implica que para cualquier par de puntos $p, q \in M$ existe una geodésica α que los une y cuya longitud $L(\alpha) = d(p, q)$. Como aplicación del teorema de Hopf–Rinow tenemos el siguiente resultado global [9,12].

---

[12] Serge LANG. *Fundamentals of Differential Geometry*. Springer, 1999.

Teorema 3 *(**Hadamard-Cartan**) Sea M una variedad Riemanniana de dimensión m, completa y simplemente conexa con curvatura seccional no positiva. Entonces M es difeomorfa a $\mathbb{R}^m$, más precisamente $Exp_p : T_pM \to M$ es un difeomorfismo para todo $p \in M$.*

Para una descripción más detallada del concepto de curvatura seccional no positiva puede verse [7,8], de igual forma el concepto de simplemente conexo puede encontrarse en cualquier libro de topología general.

Definición 8 *Sea $\alpha : [0, \epsilon] \to M$ una geodésica. El punto $\alpha(t_0)$ se dice el conjugado de $\alpha(0)$ a lo largo de α, $t_0 \in (0, \epsilon]$, si existe un campo de Jacobi J a lo largo de α no idénticamente cero tal que $J(0) = 0 = J(t_o)$.*

Se recomienda ver [9] para más detalles sobre la definición 8, más precisamente sobre los campos de Jacobi y la forma en que estos se relacionan con los puntos conjugados.

Un punto p de una variedad Riemanniana completa M se llama polo si no posee puntos conjugados, cualquier punto de una variedad completa M con curvatura seccional no positiva es un polo de M es decir, M no posee puntos conjugados, con lo cual existe una única geodésica que conecta a cualquier par de puntos sobre la variedad. Si M es una variedad que satisface el teorema de Hadamard-Cartan dados $p, q \in M$, $v \in T_pM$ será el único vector tal que $Exp_p(v) = q$ con lo cual la geodésica tiene la forma

$$\alpha_v(t) = Exp_p(tv) \text{ donde } v = Log_p(q),$$

y la distancia Riemanniana entre p y q vendrá dada por

$$d(p, q) = L(\alpha_v) = ||v|| = ||Log_p(q)||.$$

El espacio de matrices simétricas de tamaño n denotado por $Sym(n)$ es un ejemplo de variedad diferenciable ya que es un espacio vectorial. Para probar que $PD(n)$ es

una variedad diferenciable basta probar que es un abierto de $Sym(n)$, considere la función $f : \mathbb{R}^{n\times n} \to \mathbb{R}^n$ tal que $f(A) = (det_1(A),\ ...,\ det_n(A))^T$ donde $det_i(A)$ denota el $i-$ésimo menor principal de la matriz A. La función $det_i : \mathbb{R}^{n\times n} \to \mathbb{R}$ es una función polinómica y por tanto continua para cada $i = 1,\ ...,\ n$ así f es continua, además es claro que $f^{-1}(\mathbb{R}^n_+) = PD(n)$ por lo tanto $PD(n)$ es abierto de $Sym(n)$.
Probaremos ahora que para todo $p \in PD(n)$, $T_pPD(n) = Sym(n)$. Una contenencia se tiene por la definición del espacio tangente a p con lo cual $T_PPD(n) \subset Sym(n)$. Para probar la otra contenencia, sea $v \in Sym(n)$ y $p \in PD(n)$ veamos que $v \in T_PPD(n)$, defina la función $\gamma : (-\epsilon, \epsilon) \to PD(n)$ como $\gamma(t) = p+tv$, dado que $PD(n)$ es abierto de $Sym(n)$ existe $\delta > 0$ tal que $\gamma(t)|_{(-\delta,\delta)} \subset PD(n)$ con lo cual tenemos una curva contenida en $PD(n)$ tal que $\gamma(0)|_{(-\delta,\delta)} = p$ y $\gamma'(0)|_{(-\delta,\delta)} = v$ por lo tanto $v \in T_pPD(n)$ [13].

En [12,14] se define una métrica Riemanniana (que depende del punto $p \in PD(n)$) sobre el espacio $T_pPD(n)$ mediante el producto interno

$$\langle v, w\rangle_p = tr(vp^{-1}wp^{-1}), \tag{3}$$

definido como la traza del producto de matrices $v, w \in T_pPD(n)$ y la matriz inversa de $p \in PD(n)$, además, la longitud para $v \in T_pPD(n)$ viene dada por $||v|| = \sqrt{tr(vp^{-1}vp^{-1})}$. Este producto define una métrica Riemanniana llamada métrica Afín-Invariante con la cual $PD(n)$ puede describirse como una variedad Riemanniana que satisface el teorema de Hadamard-Cartan y será la métrica que se trabajará en

---

[13] Juan OLMOS. *Cálculo de medias gemoétricas en el cono de las matrices simétricas semidefinidas positivas*. Bucaramanga, Colombia, 2020.

[14] Xavier PENNEC, Pierre FILLARD y Nicholas AYACHE. “A Riemannian Framework for Tensor Computing”. En: *RR-5255, INRIA* (2004), pág. 34.

lo que sigue del trabajo. Una propiedad importante de esta métrica y por la cual se escoge es su invarianza respecto a transformaciones afines. Denotamos por $GL(n)$ al espacio de matrices invertibles $n \times n$, de esta manera el producto interno

$$\langle cvc^T, cwc^T \rangle_{cpc^T} = tr(vp^{-1}wp^{-1}) = \langle v, w \rangle_p \text{ para todo } c \in GL(n).$$

Con esta métrica, dado un punto $p \in PD(n)$ y $v \in T_pPD(n)$ en [15] se define la curva geodésica $\alpha(t)$ tal que $\alpha(0) = p$ y $\alpha'(t) = v$ como

$$\alpha(t) = p^{\frac{1}{2}} \texttt{exp}(t\ p^{-\frac{1}{2}} v p^{-\frac{1}{2}}) p^{\frac{1}{2}}, \quad t \in \mathbb{R}.$$

Donde $p^{\frac{1}{2}}$ denota la raíz cuadrada de p y la función matriz exponencial vendrá dada por la serie infinita

$$\texttt{exp} : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}, \quad \texttt{exp}(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k.$$

De esta manera dados $a, b \in PD(n)$ la única geodésica $\alpha(t) : [0, 1] \to PD(n)$ tal que $\alpha(0) = a$ y $\alpha(1) = b$ es dada por

$$\alpha_{ab}(t) = a^{\frac{1}{2}} \texttt{exp}[t\ \texttt{log}(a^{-\frac{1}{2}} b a^{-\frac{1}{2}})] a^{\frac{1}{2}}, \quad t \in [0, 1].$$

De igual forma se sigue que la aplicación exponencial vendrá dada por

$$Exp_p(v) = \alpha(1) = p^{\frac{1}{2}} \texttt{exp}(p^{-\frac{1}{2}} v p^{-\frac{1}{2}}) p^{\frac{1}{2}}, \quad v \in Sym(n),$$

---

[15] Maher MOAKHER y Mourad ZÉRAÏ. "The Riemannian Geometry of the Space of Positive-Definite Matrices and Its Application to the Regularization of Positive-Definite Matrix-Valued Data." En: *J Math Imaging Vis 40,* (2011).

y su inversa, la aplicación logaritmo

$$Log_p(a) = p^{\frac{1}{2}}\texttt{log}(p^{-\frac{1}{2}}ap^{-\frac{1}{2}})p^{\frac{1}{2}}, \quad a \in PD(n).$$

La distancia Riemanniana entre $a, b \in PD(n)$ estará determinada por la aplicación Log con lo cual

$$\begin{aligned} d(a,b) = L(\alpha_{ab}(t)) &= ||Log_a(b)|| \\ &= ||a^{\frac{1}{2}}\texttt{log}(a^{-\frac{1}{2}}ba^{-\frac{1}{2}})a^{\frac{1}{2}}|| \\ &= \sqrt{tr(\texttt{log}^2(a^{-\frac{1}{2}}ba^{-\frac{1}{2}}))}. \end{aligned}$$

La métrica afín-invariante y los métodos para calcular geodésicas, distancias y la forma de definir las aplicaciones Exp_p y Log_p se derivan naturalmente mediante acciones de grupos de Lie en variedades, se recomienda ver [8,4,16] para obtener una explicación más detallada de estos procesos.

Formulado el espacio $PD(n)$ como una variedad Riemanniana dotada de la métrica afín-invariante podemos definir el concepto estadístico de la media. Fréchet[17] define este valor para un conjunto de variables en un espacio métrico como el punto que minimiza el valor esperado de la suma de las distancias al cuadrado. De esta manera si $A = \{x_1, \ ..., \ x_n\}$ es un conjunto de puntos sobre una variedad Riemanniana M se define la media μ como

$$\mu = \underset{x \in M}{\arg\min} \sum_{i=1}^{n} d(\mu, x_i)^2.$$

---

[16] Jean GALLIER. *Notes on Differential Geometry and Lie Groups*. University of Pennsylvania, Philadelphia, USA, jun. de 2011.

[17] M. FRÉCHET. "Les éléments aléatoires de nature quelconque dans un espace distancié". En: *Annales de l'I. H. P* 10.4 (1948), págs. 215,310.

Dado que la media se define como una minimización, Karcher[18] muestra que para una variedad con curvatura seccional no positiva la media existe y es única, en consecuencia la media para el espacio $PD(n)$ con la métrica Afín-Invariente vendrá dada por

$$\mu = \arg\min_{x \in M} \sum_{i=1}^{n} ||Log_{\mu}(x_i)||^2.$$

Un estudio más detallado sobre este cálculo puede encontrarse en [13].

2.3. ANÁLISIS DE COMPONENTES PRINCIPALES

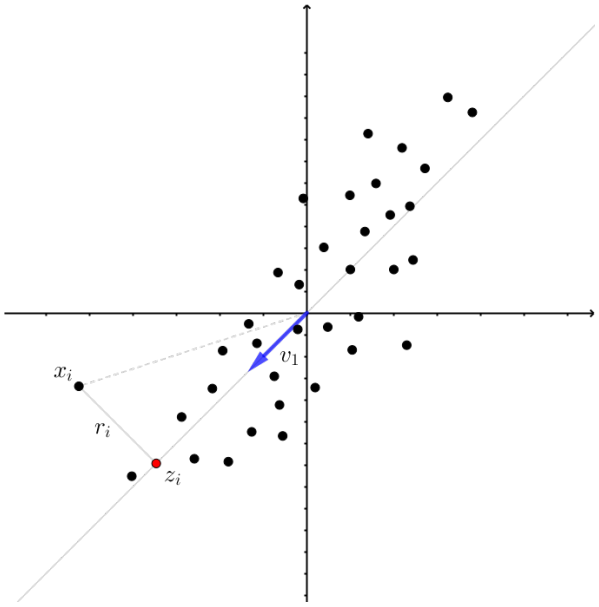
El ACP (comunmente conocido como PCA por sus siglas en inglés de *Principal Component Analysis*) tiene como objetivo transformar y reducir la dimensión de los datos mediante la proyección en subespacios que maximicen su varianza. A continuación se describe en detalle la construcción de estos subespacios y su respectivo valor de varianza.

Considere un conjunto de variables $(x_1,\ x_2,\ ...,\ x_n)$ sobre el plano cartesiano. Intuitivamente, se desea hallar una recta que pase cerca de todos los puntos y las distancias entre ellos se mantengan aproximadamente en su proyección sobre dicha recta. Esto se puede asegurar pidiendo que las distancias entre los puntos originales y sus proyecciones sobre la recta sean lo más pequeñas posibles. Considerese el punto x_i y una dirección definida por un vector normal v_1, entonces la proyección de x_i sobre v_1 es el escalar z_i dado como $z_i = v_1^T x_i$.

---

[18] H. KARCHER. “Riemannian center of mass and mollifier smoothing”. En: *Communications on Pure and Applied Mathematics* 30 (1977).

Figura 4. Proyección de un punto x_i sobre un subespacio lineal v_1.



Sea r_i la distancia entre el punto x_i y su proyección en la dirección v_1 entonces el objetivo de ajustar la recta a todos los puntos vendrá dado por:

$$\text{mín} \sum_{i=1}^{n} r_i^2.$$

Note que el punto x_i, su proyección en la dirección v_1 y el origen coordenado forman un triángulo rectángulo, por el teorema de Pitágoras, tendremos que

$$x_i x_i^T = r_i^2 + z_i^2,$$

así, para todos los puntos:

$$\sum_{i=1}^{n} x_i x_i^T = \sum_{i=1}^{n} r_i^2 + \sum_{i=1}^{n} z_i^2.$$

Es claro que el primer elemento de la anterior expresión es constante por ende minimizar $\sum_{i=1}^{n} r_i^2$ es equivalente a maximizar $\sum_{i=1}^{n} z_i^2$, de esta manera podemos

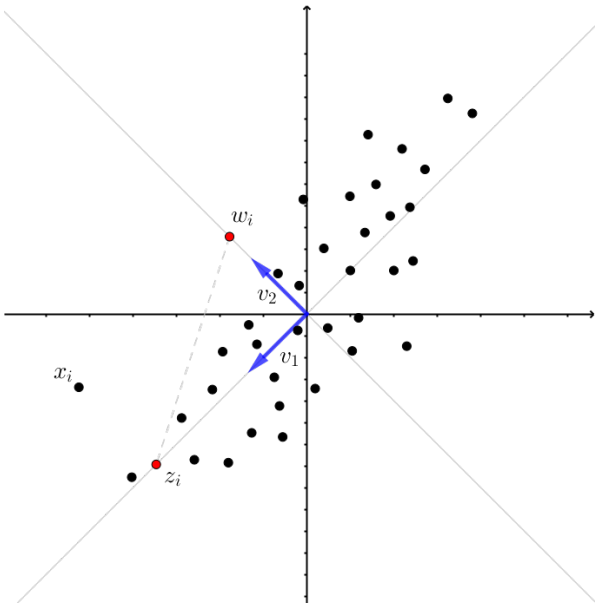
definir la primera componente principal de una forma general como:

$$v_1 = \arg\max_{\|v\|=1} \sum_{i=1}^{n} \langle v, x_i \rangle^2. \tag{4}$$

En cuanto a la segunda componente, se proyecta el punto x_i sobre otro vector normal v_2 donde se obtiene el escalar $w_i = v_2^T x_i$. Observe que esta segunda componente está condicionada respecto a la primera ya que v_1 y v_2 deben ser ortogonales. Ahora para considerar esta condición, se maximiza la distancia entre el punto proyectado sobre la primera componente y la recta perpendicular a ella, nuevamente por el teorema de Pitágoras con la condición dada tendremos un triángulo rectángulo entre los puntos z_i, w_i y el origen coordenado de donde nuestra dirección v_2 vendrá dada por

$$v_2 = \arg\max_{\|v\|=1} \sum_{i=1}^{n} \langle v_1, x_i \rangle^2 + \langle v, x_i \rangle^2.$$

Figura 5. Proyección de el punto x_i sobre un subespacio lineal v_2 ortogonal a v_1.



En general para la k-ésima componente principal se sigue de forma análoga al ar-

gumento usado para la segunda componente teniendo en cuenta que todas estas deben estar incorrelacionadas, esto es, se maximiza la distancia entre los puntos proyectados en las $k-1$ componentes anteriores y la recta ortogonal a todas ellas con lo cual obtenemos

$$v_k = \underset{\|v\|=1}{\arg\max} \sum_{i=1}^{n} \sum_{j=1}^{k-1} \langle v_j, x_i \rangle^2 + \langle v, x_i \rangle^2. \tag{5}$$

La anterior formulación general del PCA se hace mediante un enfoque geométrico el cual usaremos para posteriormente generalizar, particularmente para el espacio Euclidiano $\mathbb{R}^n$ se presenta en [2] una forma de determinar las componentes principales junto con sus respectivas varianzas, proceso el cual se presenta a continuación.

Consideremos nuevamente un conjunto de variables $X = [x_1,\ x_2,\ ...,\ x_n]$, el objetivo será obtener nuevas variables $z_1,\ z_2,\ ...,\ z_n$ incorrelacionadas entre sí que expliquen la variabilidad de X esto se consigue mediante la combinación lineal entre las variables iniciales

$$z_i = v_{i1}x_1 + v_{i2}x_2 + ... + v_{in}x_n = v_i^T X,$$

para algún $i = 1,\ ...,\ n$ donde $v_i = (v_{i1},\ ...,\ v_{in})^T \in \mathbb{R}^n$.
Para determinar la primera componente z_1 debemos encontrar v_1 que maximice su varianza, esto es

$$var(z_1) = \text{máx}\{var(v^T X),\ v \in \mathbb{R}^n,\ v^T v = 1\}.$$

La restricción $v^T v = 1$ es necesaria para mantener la ortogonalidad entre las direcciones v_i^T además, al multiplicar a v^T por una constante podemos aumentar la varianza de z_1, $var(\alpha v^T X) = \alpha^2 var(v^T X)\ \alpha \in \mathbb{R}$.
Sea $\Sigma = cov(X, X)$ la matriz de covarianza de X entonces, si $z_1 = v_1^T X$ tendremos que

$$\begin{aligned} var(z_1) &= cov(v_1^T X, v_1^T X) \\ &= v_1^T cov(X, X) v_1 \\ &= v_1^T \Sigma v_1. \end{aligned}$$

De esta manera obtendremos un problema de optimización

Problema: maximizar $v_1^T \Sigma v_1$ (la varianza)

sujeto a $v_1^T v_1 = 1$.

Una forma de resolver este problema es usando los multiplicadores de Lagrange con lo cual construimos nuestra función L

$$L(v_1) = v_1^T \Sigma v_1 - \lambda_1 (v_1^T v_1 - 1).$$

Ahora, derivando respecto a v_1 e igualando a cero

$$\frac{\partial L}{\partial v_1} = 2\Sigma v_1 - 2\lambda_1 v_1 = 0 \Longrightarrow \Sigma v_1 = \lambda_1 v_1.$$

Con lo cual v_1 es un vector propio de Σ cuyo valor propio asociado es λ_1, además la varianza de z_1 es

$$var(z_1) = var(v_1^T X) = v_1^T \Sigma v_1 = v_1 \lambda_1 v_1 = \lambda_1 v_1^T v_1 = \lambda_1.$$

Por lo tanto para la primera componente z_1 se toma el mayor valor propio de la matriz de covarianza Σ y el vector propio respectivo será la dirección v_1.

Para la segunda componente $z_2 = v_2^T X$ se continúa con un proceso similar, esta vez teniendo en cuenta que z_2 debe estar incorrelacionada con la anterior componente z_1 esto es, $cov(z_2, z_1) = 0$

$$\begin{aligned} cov(z_2, z_1) &= cov(v_2^T X, v_1^T X) \\ &= v_2^T \Sigma v_1 = v_1^T \Sigma v_2 = 0. \end{aligned}$$

Dado que $\Sigma v_1 = \lambda_1 v_1$ entonces $v_2^T \Sigma v_1 = v_2^T \lambda_1 v_1 = \lambda_1 v_2^T v_1 = 0$, con lo cual $v_2^T v_1 = 0$ es decir, v_2 y v_1 son ortogonales, equivalentemente, los vectores están incorrelacionados.

De esta manera se plantea el problema de optimización

$$\begin{aligned} \text{Problema:} \quad \text{maximizar} \quad & v_2^T \Sigma v_2, \\ \text{sujeto a} \quad & v_2^T v_2 = 1, \\ \text{y} \quad & v_2^T v_1 = 0. \end{aligned}$$

Aplicando nuevamente multiplicadores de Lagrange

$$L(v_2) = v_2^T \Sigma v_2 - \lambda_2 (v_2^T v_2 - 1) - \mu (v_2^T v_1).$$

Derivando respecto a v_2 e igualando a cero

$$\frac{\partial L}{\partial v_2} = 2\Sigma v_2 - 2\lambda_2 v_2 - \mu v_1 = 0,$$

si multiplicamos por v_1^T

$$2 \underbrace{v_1^T \Sigma v_2}_{=0} - 2\lambda_2 \underbrace{v_1^T v_2}_{=0} - \mu v_1^T v_1 = 0.$$

Dado que $\mu v_1^T v_1 = 0$ entonces $\mu = 0$ por lo tanto $\Sigma v_2 = \lambda_1 v_2$, de igual forma v_2 es un vector propio de Σ y su valor asociado es λ_2 tal que $var(z_2) = v_2^T \Sigma v_2 = \lambda_2 v_2^T v_2 = \lambda_2$. Se sigue de forma análoga para las demás componentes, la matriz de covarianza Σ es de orden n con lo cual se obtendrán n vectores y valores propios tales que $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \geq 0$, lo anterior se resume en el siguiente teorema [3].

Teorema 4 *Sea $X = [x_1,\ x_2,\ ...,\ x_n]$ un conjunto de variables entonces, las n componentes de X son de la forma $z_j = v_j^T X$, $j = 1,\ ...,\ n$ siendo $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \geq 0$ los n valores propios de la matriz de covarianza Σ de X y $v_1,\ ...,\ v_n$ sus vectores propios asociados normalizados ($\{v_1,\ ...,\ v_n\}$ forma una base ortonormal de $\mathbb{R}^n$) tales que $cov(z_j, z_k) = 0$ si $j \neq k$ y $var(z_j) = \lambda_j$ $j = 1,\ ...,\ n$.*

De esta manera, el ACP en $\mathbb{R}^n$ se reduce a calcular la matriz de covarianza de los datos y posteriormente determinar sus vectores y valores propios [3]. En el proceso del ACP se busca reducir la dimensión del conjunto de datos, esto lleva a la pregunta **¿Cuántas componentes se deben escoger?** de tal forma que se extraiga la mayor cantidad posible de información. A partir de esta pregunta se estudian 4 formas de obtener este valor descritas en [19], nuestra intención será observar si estas formas son un buen indicador en nuestro problema general.

- **Varianza acumulada** Denotaremos l_k la varianza de la $k-$ésima componente principal, si tenemos N componentes entonces el número de componentes m a utilizar será el menor valor tal que

$$t_m = \frac{100 \sum_{k=1}^{m} l_k}{\sum_{k=1}^{N} l_k},$$

  es más grande que un porcentaje de corte fijado t^*, usualmente se fija este procentaje entr el $70\,\%$ y el $90\,\%$.

- **Regla de Kaiser** Henry F. Kaiser sugiere seleccionar todas las componentes principales para las cuales

$$l_k > \bar{l},$$

  donde $\bar{l}$ denota el valor promedio de las varianzas, si se desea un valor más flexible se pueden tomar aquellas componentes para las cuales $l_k > 0.7\bar{l}$.

- **Cattell** Raymond B. Cattell se basa de forma visual, en el plano cartesiano sitúa el número de componentes en el eje x versus su varianza asocida en el eje y esto es k vs l_k y grafica estos puntos unidos linealmente, la idea será

---

[19] Alethea REA y William REA. "How Many Components should be Retained from a Multivariate Time Series PCA?" En: (oct. de 2016).

determinar visualmente un “codo” para el cual los valores siguientes tiendan a un valor específico fijo, este método puede resultar subjetivo.

- **Farmer** S. Farmer hace una pequeña modificación a la forma descrita por Cattell, grafica de igual forma en el eje x el número de componentes mientras que para el eje y usa el logaritmo de la varianza es decir, k vs $log(l_k)$, de igual forma se determina de forma visual un punto para el cual los valores siguientes decrecen linealmente.

Observación: Generalmente en el uso del ACP, se centran los datos restando su respectiva media a cada punto o dato de tal forma que la media posterior sea cero, esto se realiza debido a que la matriz de covarianza no posee una interpretación directa con los datos, sin el proceso de centrar se tendría necesariamente que hacer una transformación afín para interpretar los datos. Este hecho es importante ya que nuestro espacio por su naturaleza curva puede que, al aplicarse directamente la resta de su media se pierdan propiedades necesarias de los puntos como lo es su propiedad definida positiva. Por lo tanto para lo que sigue debemos tener presente la media, cálculo el nos basaremos de [13] quien presenta una forma más precisa de este computo sobre el espacio de las matrices simétricas definidas positivas. La exactitud en el cálculo de la media será de suma importancia en lo que sigue del trabajo ya que este valor será la base a partir de la cual aplicaremos la generalización del ACP.

3. ANÁLISIS DE GEODÉSICAS PRINCIPALES

En este capítulo se presenta la extensión natural del ACP al espacio de las matrices simétricas definidas positivas $PD(n)$ haciendo uso de los resultados presentados en los preliminares.

Consideremos un conjunto de puntos $p_1,\ p_2, ...\ p_n$ sobre una variedad Riemanniana, nuestro objetivo será describir la variabilidad de los puntos de una forma análoga al ACP. Para ello es necesario extender tres conceptos importantes: varianza, subvariedad geodésica y proyección.

3.1. VARIANZA

La varianza σ^2 en $\mathbb{R}^n$ para un conjunto de datos X con media μ se define como el valor esperado de las distancias al cuadrado de los datos respecto a su media y viene dada por la fórmula $\sigma^2 = E[(X-\mu)^2]$. Usaremos la definición de varianza dada por Fréchet [17] quien define la varianza de una variable aleatoria en un espacio métrico como el valor esperado de la distancia al cuadrado respecto de la media.

$$\sigma^2 = E[d(X,\mu)^2].$$

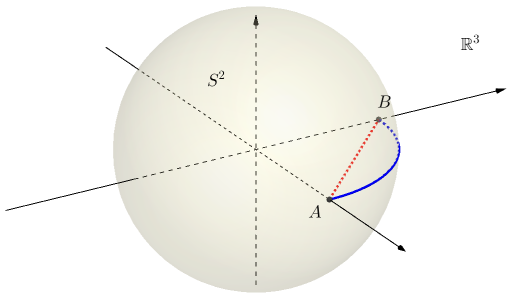
Así, dado un conjunto de datos $p_1,\ p_2, ...\ p_n$ con media μ en una variedad Riemanniana definimos la varianza de los datos como

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} d(\mu,\ p_i)^2 = \frac{1}{n}\sum_{i=1}^{n} ||Log_\mu(p_i)||^2.$$

3.2. SUBVARIEDAD GEODÉSICA

Lo siguiente será generalizar el concepto de subespacio lineal. Una geodésica es localmente el camino más corto entre dos puntos. Así la extensión de línea recta en el espacio Euclidiano la entenderemos como la curva geodésica en una variedad Riemanniana. No obstante, si N es una subvariedad de una variedad M, las geodésicas en N no necesariamente son geodésicas en M.

Figura 6. Geodésicas entre A y B en S^2 (curva azul) y $\mathbb{R}^3$ (línea roja).



Por ejemplo, la esfera S^2 es una subvariedad de $\mathbb{R}^3$ pero las geodésicas en S^2 serán segmentos de arco, mientras que las geodésicas en $\mathbb{R}^3$ son líneas rectas, esto nos supone definir una característica sobre una subvariedad.

Definición 9 *Una subvariedad Riemanniana N de una variedad Riemanniana M se dice geodésica en $x \in N$ si todas las geodésicas de N que pasan por x son geodésicas de M.*

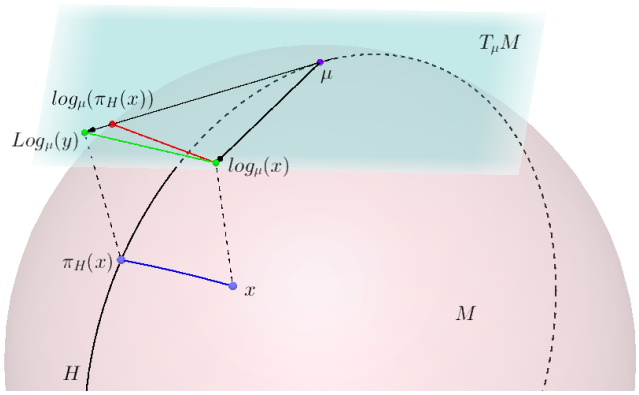
Subvariedades geodésicas en x preservan la distancia a x, esto supone una propiedad importante para el AGP ya que la varianza se define como la distancia al cuadrado de la media; de esta manera las subvariedades geodésicas serán la generalización de los subespacios lineales.

3.3. PROYECCIÓN

Sea M una variedad Riemanniana y H una subvariedad geodésica, la proyección de un punto $x \in M$ en H se define como el punto en H más cercano a x en la distancia Riemanniana. Se define entonces el operador $\pi_H : M \to H$ tal que

$$\begin{aligned} \pi_H(x) &= \arg\min_{y \in H} d(x,y)^2 \\ &= \arg\min_{y \in H} ||Log_x(y)||^2. \end{aligned}$$

Figura 7. Aproximación del operador $\pi_H(x)$. Representación de la proyección de un punto $x \in M$ sobre H en μ (punto azul), la proyección mapeada a $T_\mu M$ (punto verde) y la aproximación de la proyección sobre $T_\mu M$ (punto rojo).



La proyección se define mediante una minimización, dado que $PD(n)$ es una variedad Riemanniana que satisface el teorema de Hadamard-Cartan este variedad no posee puntos conjugados lo cual garantiza que la proyección sobre una subvariedad geodésica existe y es única.

Esta proyección sobre una subvariedad geodésica en μ puede ser aproximada li-

nealmente mediante el espacio tangente $T_\mu M$

$$\pi_H(x) \approx \arg\min_{y \in H} ||Log_\mu(x) - Log_\mu(y)||^2.$$

Observe que $Log_\mu(y)$ es simplemente un vector de $T_\mu H$ así, reescribiendo la anterior aproximación en términos de vectores en $T_\mu H$

$$Log_\mu(\pi_H(x)) \approx \arg\min_{v \in T_\mu H} ||Log_\mu(x) - v||^2.$$

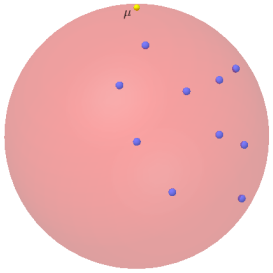
La anterior minimización es precisamente la proyeccón ortogonal de $Log_\mu(x)$ en un subespacio de $T_\mu M$ luego, si $v_1,\ v_2,\ ...,\ v_k$ es una base ortonormal de $T_\mu H$ entonces el operador π_H puede ser aproximado mediante la fórmula

$$Log_\mu(\pi_H(x)) \approx \sum_{i=1}^{k} \langle v_i, Log_\mu(x) \rangle_\mu. \tag{6}$$

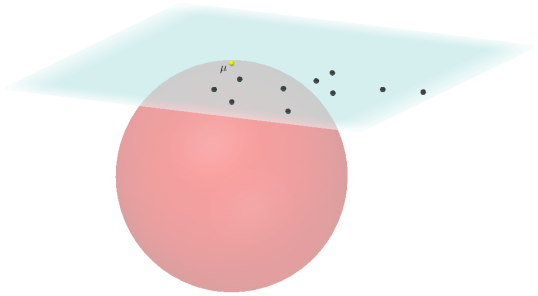
Lo anterior nos permite encontrar los subespacios geodésicos H de M que minimizan las distancias al cuadrado de los puntos en M con su proyección en H. Como en cualquier aproximación sobre el espacio tangente, el error en el operador proyección aumentará si los puntos están más lejos de su media.

Ejemplo 1 *Ilustraremos la aproximación de la proyección mediante la esfera unitaria S^2, considere 10 puntos sobre la esfera y un punto μ, queremos encontrar la subvariedad geodésica H en μ que minimice las distancias de los 10 puntos.*

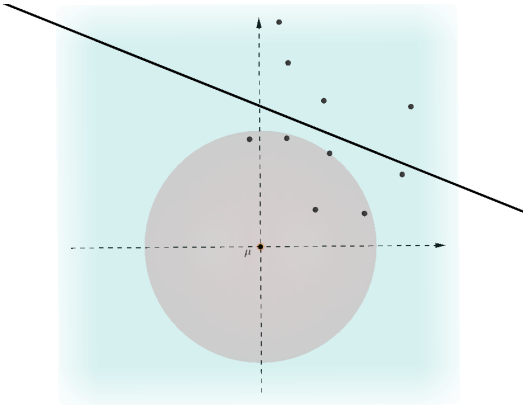
Figura 8. Ilustración de la aproximación en la esfera unitaria S^2.



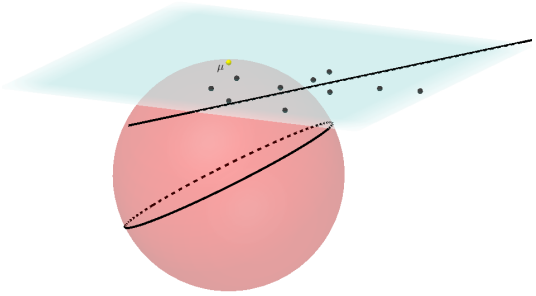
(a) 10 puntos sobre la esfera unitaria S^2 y un punto μ.



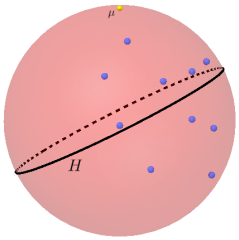
(b) Se define el espacio tangente en μ y se mapean los puntos al espacio mediante la función Log_μ.



(c) Sobre el espacio tangente T_μ se calcula el ajuste lineal de los puntos $Log_\mu(x_i)$.



(d) Subespacio lineal que devolvemos a la esfera unitaria S^2.



(e) Obteniendo la aproximación de la subvariedad geodésica H en S^2.

3.4. ANÁLISIS DE GEODÉSICAS PRINCIPALES

Sean $p_1,\ p_2\ ,...,\ p_N$ puntos en un variedad Riemanniana M y μ su media. El objetivo será, análogo al ACP, encontrar una secuencia de subvariedades geodésicas que maximicen la varianza de los datos proyectados, es decir, determinar vectores ortonormales $v_1, ..., v_k$ en $T_\mu M$ generadores de las subvariedades geodésicas que maximizan la varianza, de donde las geodésicas principales vendrán dadas por las imágenes de los vectores bajo la aplicación exponencial.

La primera dirección principal es aquella que maximiza la varianza proyectada sobre la correspondiente geodésica

$$v_1 = \arg\max_{\|v\|=1} \sum_{i=1}^{N} ||Log_\mu(\pi_H(p_i))||^2, \tag{7}$$

donde $H = Exp_\mu(gen(\{v\}))$.

La $k-$ésima dirección principal se define recursivamente como

$$v_k = \arg\max_{\|v\|=1} \sum_{i=1}^{N} ||Log_\mu(\pi_H(p_i))||^2, \tag{8}$$

donde $H = Exp_\mu(gen(\{v_1,\ \cdots,\ v_{k-1},\ v\}))$.

Aplicando (6) al operador de proyección $\pi_H(x)$ en (7) y (8) obtenemos

$$v_1 \approx \arg\max_{\|v\|=1} \sum_{i=1}^{N} \langle v, Log_\mu(p_i)\rangle^2, \tag{9}$$

$$v_k \approx \arg\max_{\|v\|=1} \sum_{i=1}^{N} \sum_{j=1}^{k-1} \langle v_j, Log_\mu(p_i)\rangle_\mu^2 + \langle v, Log_\mu(p_i)\rangle^2. \tag{10}$$

Si comparamos las anteriores aproximaciones con las ecuaciones (4) y (5), observamos que el anterior problema de maximización no es más que el ACP en $T_\mu M$

aplicado a los vectores $Log_{\mu}(p_i)$. Nos referiremos entonces a las ecuaciones (9) y (10) como PGA *linealizado* y, las ecuaciones (7) y (8) se denominarán AGP *exacto*[20].

[20] Stefan SOMMER, Francois LAUZE y Mads NIELSEN. “The Differential of the Exponential Map, Jacobi Fields and Exact Principal Geodesic Analysis”. En: *Computing Research Repository - CORR* (2010).

4. ALGORITMOS ESTUDIADOS

En este trabajo se estudiaron además dos algoritmos propuestos para computar el AGP *linealizado*. Adicional a ello, mediante ejemplos con matrices simétricas definidas positivas 2×2 visualizaremos el proceso y comportamiento de los mismos.

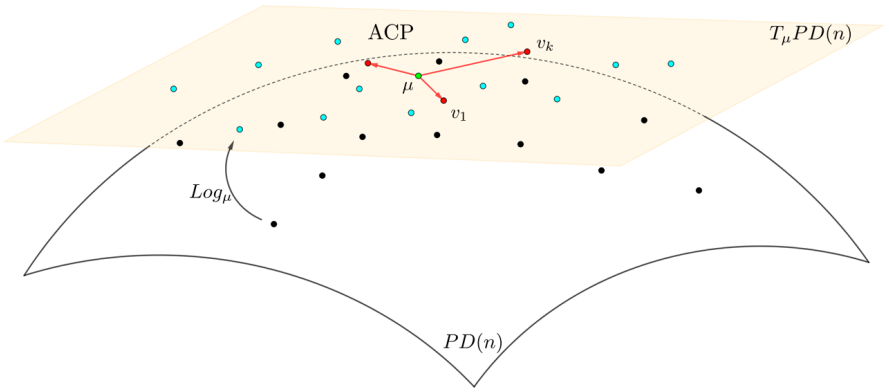
4.1. ALGORITMO DE FLETCHER

Siguiendo la anterior generalización del AGP, Thomas Fletcher y Sarang Joshi[4] proponen el siguiente algoritmo para calcular el AGP *linealizado* en el espacio $PD(n)$.

Algoritmo 1 AGP *linealizado*, método de Fletcher

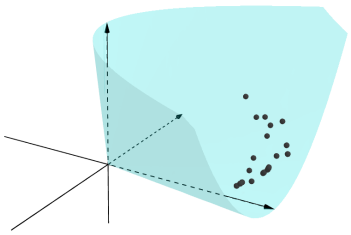
Entrada: $p_1,\ p_2,\ ...,\ p_n \in PD(n)$
Salida: direcciones principales $v_k \in T_\mu PD(n)$, varianzas $\lambda_k \in \mathbb{R}$
$\mu =$ media de $\{p_i\}_{i=1,2,...,n}$
$x_i = Log_\mu(p_i)$
$S = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$ (considerando x_i como un vector columna)
$\{v_k, \lambda_k\} =$ vectores/valores propios de S

Figura 9. Ilustración algoritmo de Fletcher. Dado un conjunto de puntos en $PD(n)$ se mapean al espacio tangente en la media $T_\mu PD(n)$ donde se aplica el ACP usual.

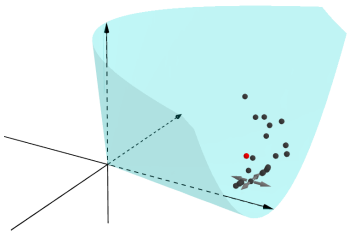


Ejemplo 2 *Con ayuda de la librería make_spd_matrix de Python la cual genera matrices simétricas definidas positivas de forma aleatoria, se crean* 20 *matrices* 2×2 *con las cuales ilustraremos el algoritmo de Fletcher en el espacio* $PD(2)$*.*

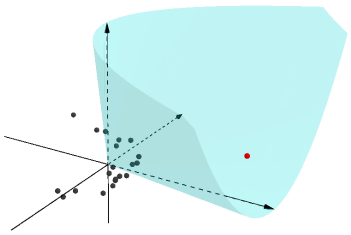
Figura 10. Ilustración algoritmo de Fletcher para un conjunto de matrices en el espacio $PD(2)$.



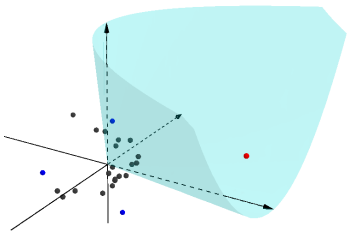
(a) 20 puntos en el Cono de matrices $PD(2)$.



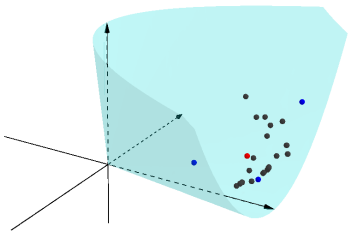
(b) Cálculo de la media del conjunto (punto rojo).



(c) Puntos llevados al espacio tangente en la media.



(d) Aplicación del ACP obteniendo las direcciones de varianza (puntos azules).



(e) Componentes devueltas al cono de matrices $PD(2)$ donde se visualiza el conjunto de puntos, su media y sus respectivas direcciones de varianza.

4.2. ALGORITMO DE YUCHEN XIE

Posteriormente, Yuchen Xie, Baba C. Vemuri, y Jeffrey Ho [21] estudian el trabajo presentado por Fletcher y Joshi, dando un aporte importante. Excepto en la identidad, el producto interno definido en la ecuación 3 no corresponde con el producto interno estándar Euclidiano el cual es necesario para el algoritmo usual del ACP. Por lo tanto, primero transformamos los datos al espacio tangente en la identidad, proceso que se logra a través de la siguiente transformación [22], si $v \in T_\mu PD(n)$

$$\begin{aligned} \phi_{\mu^{-1}} : T_\mu PD(n) &\longrightarrow T_I PD(n) \\ v &\longrightarrow \phi_{\mu^{-1}}(v) = \mu^{-\frac{1}{2}} v \mu^{-\frac{1}{2}}. \end{aligned}$$

Una vez los datos son llevados a $T_I PD(n)$, podemos aplicar el algoritmo usual del ACP para obtener las direcciones principales U_i con $i = 1, 2, \cdots, k$, las cuales devolvemos a $T_\mu PD(n)$ mediante la transformación inversa $\phi_\mu(u_i)$. En el siguiente algoritmo se presenta el proceso completo.

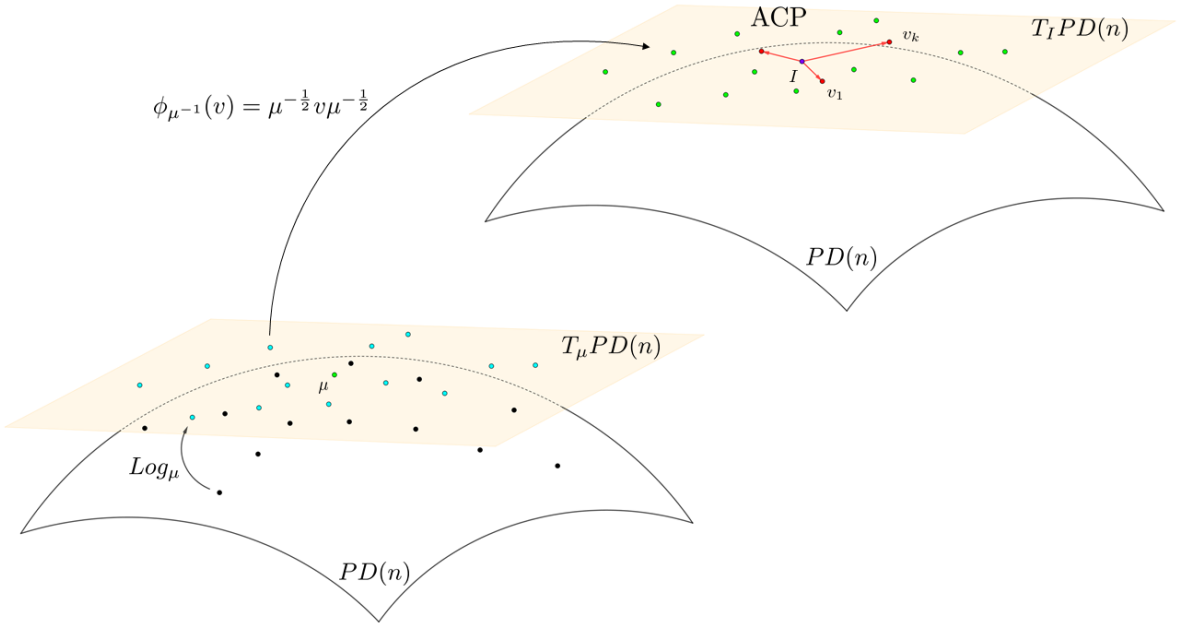
Algoritmo 2 AGP *linealizado*, método de Xie

Entrada: $p_1,\ p_2,\ ...,\ p_n \in PD(n)$
Salida: direcciones principales $v_k \in T_I PD(n)$, varianzas $\lambda_k \in \mathbb{R}$
$\mu =$ media de $\{p_i\}_{i=1,2,...,n}$
$y_i = Log_\mu(p_i)$
$x_i = \phi_{\mu^{-1}}(y_i)$
$S = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$ (considerando x_i como un vector columna)
$\{v_k, \lambda_k\} =$ vectores/valores propios de S

[21] Yuchen XIE, Baba C. VEMURI y Jeffrey HO. "Statistical Analysis of Tensor Fields". En: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010* (2010), págs. 682-689.

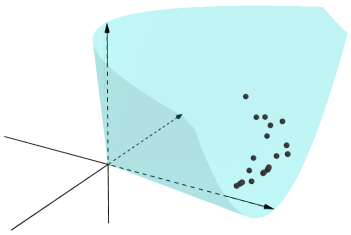
[22] Yair OR, Ben-Chen MIRELA y Talmon RONEN. "Parallel Transport on the Cone Manifold of SPD Matrices for Domain Adaptation". En: *IEEE Transactions on Signal Processing* 67 (2019), 1797–1811.

Figura 11. Ilustración algoritmo de Yuchen. Dado un conjunto de puntos en $PD(n)$ se llevan al espacio tangente en la media $T_\mu PD(n)$ y posteriormente se trasladan al espacio tangente en la media $T_I PD(n)$ donde se aplica el ACP usual.

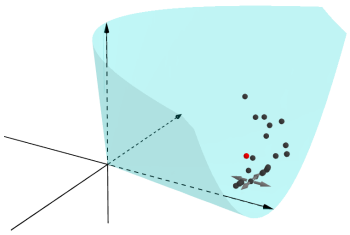


Ejemplo 3 *Con las* 20 *matrices simétricas definidas positivas generadas en el ejemplo 2 se ilustra ahora el algoritmo de Yuchen.*

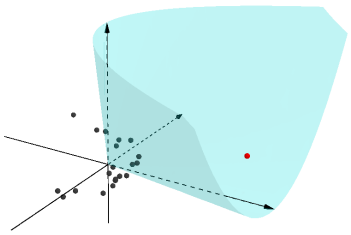
Figura 12. Ilustración algoritmo de Yuchen para un conjunto de matrices en el espacio $PD(2)$.



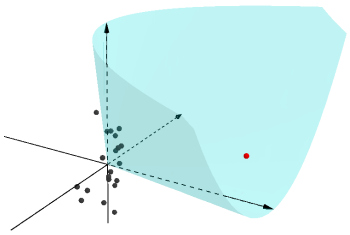
(a) 20 puntos en el Cono de matrices $PD(2)$.



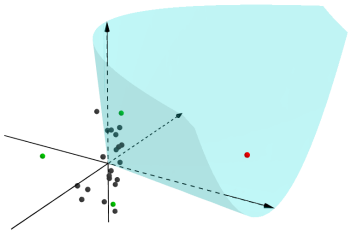
(b) Cálculo de la media del conjunto (punto rojo).



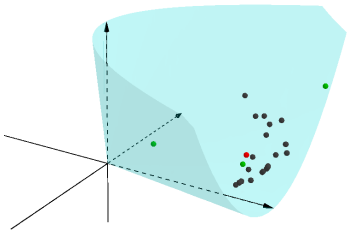
(c) Puntos llevados al espacio tangente en la media.



(d) Puntos mapeados al espacio tangente en la identidad.



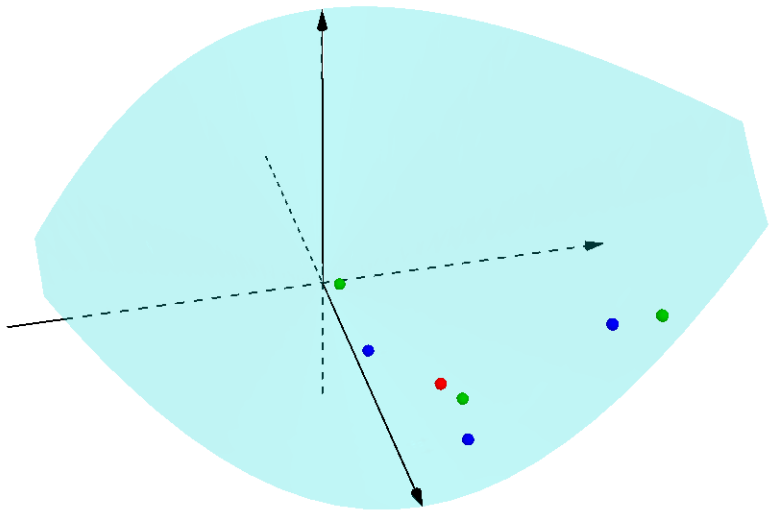
(e) Aplicación del ACP obteniendo las direcciones de varianza (puntos verdes).



(f) Componentes devueltas al cono de matrices $PD(2)$ donde se visualiza el conjunto de puntos, su media y sus respectivas direcciones de varianza.

Los dos algoritmos fueron aplicados al mismo conjunto de matrices, de esta manera tenemos una comparación entre las componentes obtenidas por cada algoritmo como se ilustra a continuación.

Figura 13. Comparación entre componentes obtenidas por los algoritmos. La media (punto rojo), las componentes dadas por el algoritmo de Fletcher (puntos azules) y las componentes dadas por el algoritmo de Yuchen (puntos verdes).



Dado que ambos procesos son aproximaciones se observan precisamente pequeñas diferencias en las direcciones dadas por cada algoritmo. En el capítulo 7, a partir de experimentos realizados se evidenciará cual de los dos algoritmos resulta más eficiente y cuyas direcciones dadas son una mejor aproximación de las direcciones de mayor varianza en el conjunto de matrices de covarianza.

5. MODELO PROPUESTO COMO DESCRIPTOR DE VIDEO

Iniciaremos este capítulo introduciendo el concepto de descriptor de covarianza y posteriormente presentando el modelo propuesto como descriptor para la clasificación de video.

5.1. DESCRIPTOR DE COVARIANZA POR CUADRO

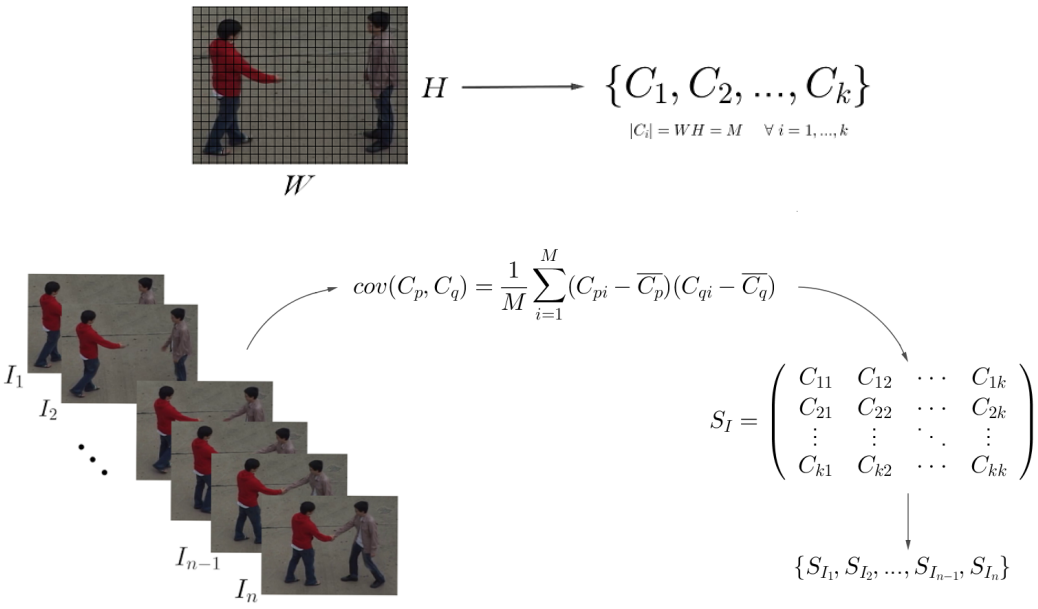
Dado un video podemos obtener una secuencia de imágenes $I_1, \ldots, I_n$ donde cada I_i tiene dimensión (en píxeles) de $H \times W$ correspondiente a su alto H y ancho W con $i = 1, 2, \ldots, n$ que puede representarse mediante un conjunto de k características $C_1, C_2, ..., C_k$ donde cada C_j posee dimensión $H \times W$; mediante una reordenación podemos obtener una expresión equivalente de tamaño $M = HW$, de esta forma cada C_j se puede definir como un vector columna que pertenece a $\mathbb{R}^M,\ j = 1, 2, \ldots, k$. Definiendo la matriz $X = [C_1, ..., C_k]$ con M mediciones y k variables, tendremos que el descriptor de covarianza S_i de I_i viene dado por la matriz de covarianza

$$S_i = \frac{1}{M} \sum_{i=1}^{M} (x_i - \mu)(x_i - \mu)^T$$

donde x_i es un vector columna que contiene las k variables sobre la medición i, $i = 1, 2, \ldots, M$ y μ representa la media de X esto es, μ es un vector columna que contiene la media de cada C_j con $j = 1, 2, \ldots, k$.

Si a cada imagen se le calcula su descriptor de covarianza podemos considerar el vídeo como un conjunto multivariante de gran dimensión donde los datos son precisamente matrices de covarianza $\{S_1, S_2, ..., S_n\}$.

Figura 14. Representación del descriptor de covarianza.



Ejemplo 4 *Considere la siguiente imagen de tamaño* 3×3 *sobre la cual queremos determinar* 3 *características: intensidad de color rojo, verde y azul (Valores que se encuentran en el intervalo* $[0, 255]$*) y posteriormente calcular su matriz de covarianza.*

Figura 15. Ejemplo del cálculo de la matriz de covarianza a una imagen de dimensión 3×3.



De donde obtenemos las siguientes matrices características que describen los valores de la intensidad de cada color en cada píxel.

Intensidad de color rojo $\begin{pmatrix} 226 & 235 & 54 \\ 223 & 207 & 41 \\ 200 & 182 & 63 \end{pmatrix}$ *Intensidad de color verde* $\begin{pmatrix} 123 & 255 & 239 \\ 80 & 228 & 209 \\ 103 & 192 & 183 \end{pmatrix}$ *Intensidad de color azul* $\begin{pmatrix} 54 & 47 & 45 \\ 61 & 52 & 49 \\ 68 & 92 & 63 \end{pmatrix}$

Reordenando cada matriz por filas podemos reescribirlas como vectores columna intensidad de rojo $(226,\ 235, 54,\ 223,\ 207,\ 41,\ 200,\ 182,\ 63)^T$*,*
intensidad de verde $(123,\ 255,\ 239,\ 80,\ 228,\ 209,\ 103,\ 192,\ 183)^T$*,*
intensidad de azul $(54,\ 47,\ 45,\ 61,\ 52,\ 49,\ 68,\ 92,\ 63)^T$*.*
Con lo cual la matriz de datos vendrá dada por

$$\begin{bmatrix} 226 & 123 & 54 \\ 235 & 255 & 47 \\ 54 & 239 & 45 \\ 223 & 80 & 61 \\ 207 & 228 & 52 \\ 41 & 209 & 49 \\ 200 & 103 & 68 \\ 182 & 192 & 92 \\ 63 & 183 & 63 \end{bmatrix},$$

donde las columnas representan las variables o características (de izquierda a derecha las intensidades rojo, verde y azul) y las filas las mediciones, así la media vendrá dada por $\mu = (159, 179.11, 59)^T$ *donde la componente* i *es la media de la columna* i*,* $i = 1, 2, 3$*. De esta manera si* $x_i = (x_{i1},\ x_{i2},\ x_{i3})^T$ *es un vector columna cuyas componentes representan la intensidad de color rojo, verde y azul respectivamente sobre la medición* i*, sustituyendo en la ecuación 1, la matriz de covarianza vendrá*

dada por

$$S = \frac{1}{9}\sum_{i=1}^{9} \begin{pmatrix} x_{i1} - 159 \\ x_{i2} - 179.11 \\ x_{i3} - 59 \end{pmatrix} \Big(x_{i1} - 159, \quad x_{i2} - 179.11, \quad x_{i3} - 59 \Big),$$

note por ejemplo que, si $i = 1$ *obtenemos la matriz*

$$\begin{pmatrix} 226 - 159 \\ 123 - 179.11 \\ 54 - 59 \end{pmatrix} \Big(226 - 159, \quad 123 - 179.11, \quad 54 - 59 \Big) = \begin{pmatrix} 4489 & -3759.37 & -335 \\ -3759.37 & 3148.33 & 280.55 \\ -335 & 280.55 & 25 \end{pmatrix}.$$

Por lo tanto, al realizar el anterior proceso para cada i*, sumar las* 9 *matrices resultantes y dividir entre el número de mediciones, obtendremos la matriz de covarianza*

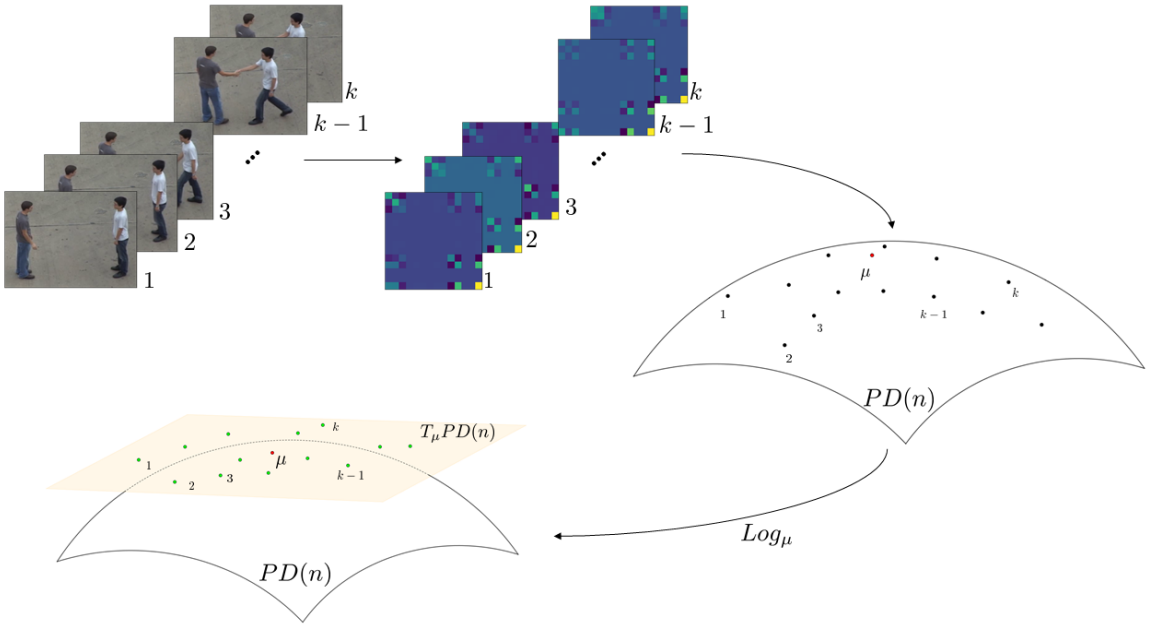
$$S = \begin{pmatrix} 5893.33 & -1666.77 & 215.44 \\ -1666.77 & 3508.32 & -283.55 \\ 215.44 & -283.55 & 189.33 \end{pmatrix}.$$

5.2. DESCRIPTOR DE VIDEO USANDO COMPONENTES PRINCIPALES

A continuación se presentan los modelos propuestos como descriptores para la clasifiación de video haciendo uso de los dos algoritmos estudiados.

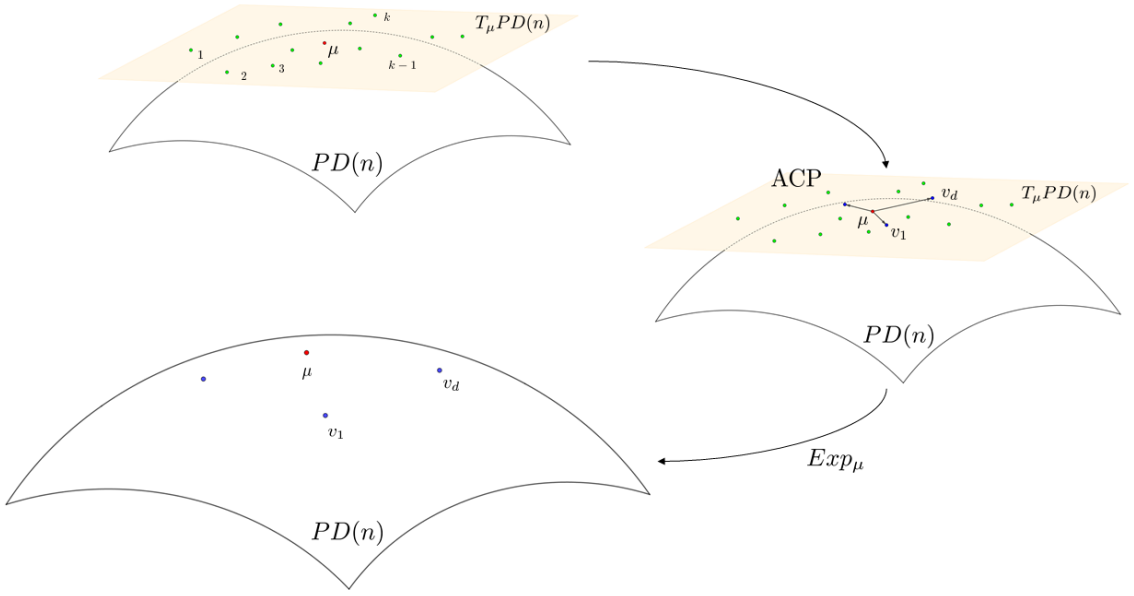
Dado un video secuenciado en k imágenes calculamos n características a cada imagen, las cuales agrupamos en una matriz de covarianza obteniendo una matriz de tamaño $n \times n$. Realizado el proceso para cada imagen se tendrán k matrices de covarianza de tamaño $n \times n$ las cuales regulamos y las llevamos al espacio $PD(n)$. Teniendo nuestros datos sobre $PD(n)$ calculamos la media μ presentada en [13], usamos la aplicación logaritmo Log_μ para llevar los puntos al espacio tangente a la media $T_\mu PD(n)$ y se procede a aplicar los diferentes algoritmos.

Figura 16. Modelo experimental, proceso inicial.



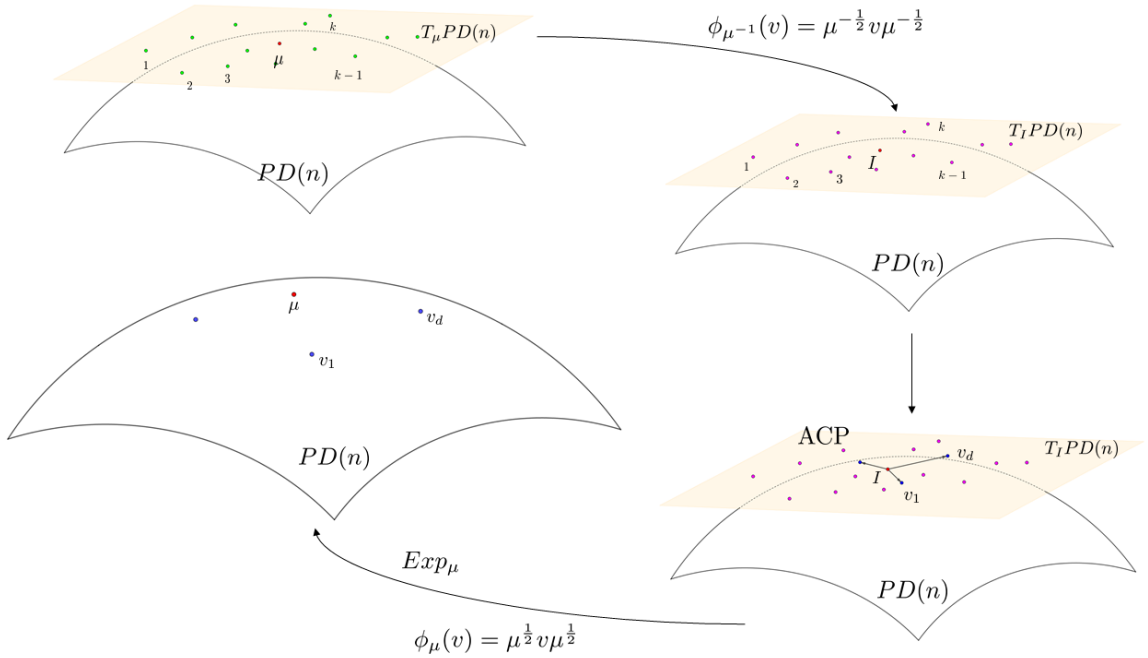
- **Algoritmo de Fletcher** Con nuestros puntos sobre el espacio tangente a la media $T_\mu PD(n)$ transformamos cada matriz en un vector de dimensión $\frac{n(n+1)}{2}$ y calculamos la matriz de covarianza S determinando los vectores $\{v_1, \ ..., \ v_d\}$ y valores propios respectivos, esto es, aplicar el proceso usual del ACP. Reconstruimos a partir de los vectores propios las matrices simétricas respectivas y utilizando la aplicación exponencial Exp_μ devolvemos los vectores propios a nuestra variedad $PD(n)$.

Figura 17. Modelo experimental del algoritmo de Fletcher.



- **Algoritmo de Yuchen** Con nuestros puntos sobre el espacio tangente a la media $T_\mu PD(n)$, trasladamos mediante la función $\phi_{\mu^{-1}}(v) = \mu^{-\frac{1}{2}} v \mu^{-\frac{1}{2}}$ los puntos de $T_\mu PD(n)$ al espacio tangente en la identidad $T_I PD(n)$ donde transformamos cada matriz en un vector de dimensión $\frac{n(n+1)}{2}$ y calculamos la matriz de covarianza S determinando los vectores $\{v_1, \ ..., \ v_d\}$ y valores propios respectivos. Nuevamente se aplica el proceso usual del ACP, reconstruimos a partir de los vectores propios las matrices simétricas respectivas y devolvemos nuestras componentes principales de nuevo a la variedad $PD(n)$ esto, mediante la aplicación primero, de la función inversa $\phi_\mu(v) = \mu^{\frac{1}{2}} v \mu^{\frac{1}{2}}$ que nos traslada las direcciones principales a $T_\mu PD(n)$ y posteriormente con la aplicación Exp_μ llevando nuestras direcciones a la variedad $PD(n)$.

Figura 18. Modelo experimental del algoritmo de Yuchen.



En cualquier caso, obtendremos una aproximación de las direcciones de mayor varianza de nuestro conjunto de matrices de covarianza inicial, finalmente para efectos de clasificación, llevamos la media μ y nuestras direcciones $\{v_1, \ ..., \ v_d\}$ al espacio tangente a la identidad $T_I PD(n)$ donde el descriptor vendrá dado por la concatenación de la media junto a un número determinado de componentes, es decir,

$$\{\mu \ v_1 \ v_2 \ ... \ v_i\},$$

donde $i \leq d$.

De este modelo experimental surgen varios interrogantes como lo son, ¿Es conveniente utilizar como descriptor la media junto con las componentes?, ¿Es buena la forma en que se está clasificando?. Un pequeño estudio sobre estas preguntas pueden encontrarse en los Anexos (1) y (2) respectivamente.

6. CONFIGURACIÓN EXPERIMENTAL

A continuación se presenta la base de datos con el cual se experimenta y valida el modelo propuesto. También se exponen los parámetros bajo los cuales se implementaron los experimentos, una breve introducción a los métodos que permiten obtener características de secuencias de imágenes y los métodos de clasificación y validación utilizados.

6.1. BASE DE DATOS

El conjunto de datos utilizado es UT-Interaction [23] el cual contiene dos secuencias con seis clases de interacciones humano-humano tales como: dar la mano, señalar, abrazar, empujar, patear y golpear. Por cada clase hay un total de 10 videos cuya duración es de alrededor de 1 minuto. En los videos aparecen varios participantes con más de 15 condiciones de ropa diferente. Para nuestros experimentos consideramos solamente la primera secuencia que consta de videos grabados en un parqueadero con un fondo fijo de entre 120-150 imágenes de dimensión 480×320. Este conjunto de video presenta desafios significativos para su modelamiento y procesamiento automático, como por ejemplo, pequeños saltos de la cámara, variabilidad en las acciones, cambios de apariencia y cambios en el fondo.

---

[23] M. S. RYOO y J. K. AGGARWAL. *UT-Interaction Dataset, ICPR contest on Semantic Description of Human Activities (SDHA)*. 2010.

Figura 19. Primera secuencia del conjunto de datos UT-Interaction.



6.2. FLUJO ÓPTICO Y REDES NEURONALES CONVOLUCIONALES

Para describir cada cuadro en el video, se utilizaron dos estrategias diferentes: 1) utilizando descriptores de movimiento, obtenidos como base de representación el flujo óptico y 2) representando cada cuadro como un conjunto de respuestas a filtros locales aprendidos mediante una arquitectura convolucional. A continuación se describe brevemente los principios de las estrategias utilizadas para representar cada cuadro en el video.

El flujo óptico permite describir el movimiento local y aparente de un cuerpo a partir del análisis de una secuencia de imágenes o frames. Entre cuadros consecutivos del video se captura el desplazamiento de movimiento a nivel de píxel, conformando un campo vectorial de desplazamiento. La técnica más conocida para calcular el flujo óptico es el algoritmo de Lucas-Kanade. Al aplicar el flujo óptico sobre un video este quedará expresado como un campo vectorial de desplazamiento donde cada pixel tiene un vector de desplazamiento asociado. Esto permite determinar cinemáticas como lo son velocidad, aceleración, magnitudes entre otras. Características que permiten dar una representación de la acción que se esté ejecutando en el video. En la figura 20 se visualiza un ejemplo del flujo óptico calculado a una serie de imágenes donde los vectores de desplazamiento y las velocidades son representados mediante una gama de colores.

Las redes neuronales convolucionales (CNN por sus siglas en inglés) son un

tipo de red neuronal utilizadas generalmente para el análisis de datos de gran dimensión como imágenes o videos. Estos algoritmos han surgido dentro del marco de aprendizaje profundo que recientemente ha llamado la atención por los resultados obtenidos para representar y clasificar diferentes instancias en una amplia gama de ejemplos, como lo son: clasificación de imágenes, reconocimiento de voz, producción de texto, entre muchas otras. En términos específicos una red neuronal utiliza un conjunto de datos de entrenamiento para ajustar un conjunto de pesos que estan organizados jerárquicamente. Estos pesos, así como su disposición topológica, se denomina la representación aprendida. Esta representación se ajusta utilizando un algoritmo de gradiente descendiente y una estrategia de propagación hacia atrás.

Las redes convolucionales son redes neuronales diseñados para explotar representaciones de datos que están dispuestos matricialmente, donde las mayores correlaciones pueden explotarse localmente. Para lo anterior, la representación jerárquica inicia por aprender en sus primeros niveles un conjunto de kernels que realizan operaciones de convolución. Los resultados de operar estos kernels son activaciones, las cuales son sucesivamente convolucionadas con capas superiores de representación. Una CNN aprende características de bajo nivel en las capas iniciales como pueden ser líneas rectas, verticales, o ciertas curvas, seguidas de capas donde se identifican patrones más complejos e identificando caracteríticas de alto nivel. La función de una CNN puede entenderse como una canalización, con la diferencia clave en el aprendizaje automático de una jerarquía en las representaciones de características.

En este trabajo cada cuadro de video fue mapeado a una red convolucional propuesta en el estado del arte y pre-entrenada en un dominio de imágenes naturales. Por cada cuadro mapeado a la estructura convolucional se producen un conjunto de activaciones que resultan en las primeras capas de representación. Estas ac-

tivaciones son entonces tomadas como las características para hacer el posterior procesamiento con el modelo propuesto.

6.3. CLASIFICACIÓN

La clasificación de acciones mediante el descriptor propuesto se logra usando el algoritmo *Gaussian Naive Bayes classifier* el cual se fundamenta en el teorema de Bayes y la distribución normal (Gaussiana). Este algoritmo asume la independencia entre los valores (características), de esta manera se distribuyen los elementos de acuerdo a una distribución normal. A su vez resuelve la probabilidad mediante la fórmula de Bayes de pertenecer a ciertas clases establecidas. Para implementar este algoritmo es necesario que el descriptor pertenezca a un espacio Euclidiano, para ello nuevamente se reorganizan las componentes de nuestro descriptor (las cuales son matrices que pertenecen a $T_IPD(n)$) vectorizando la triangular superior de cada matriz, que a su vez son etiquetadas de acuerdo a la clase que pertenezca la acción.

En este sentido, cada video del conjunto de datos se describe mediante la concatenación de su matriz media de covarianza seguida de un número determinado d de componentes principales o matrices de covarianza que representan las direcciones de mayor varianza en el video $\{\mu\ v_1\ v_2\ ...\ v_d\}$ donde cada $v_i \in T_IPD(n)$. Elementos que vectorizamos de modo que obtenemos $\{\mu\ v_1'\ v_2'\ ...\ v_d'\}$ donde cada $v_i' \in \mathbb{R}^{\frac{n(n+1)}{2}}$ y son etiquetados de acuerdo a la clase que pertenezca el video. Con lo cual lo que ingresa al clasificador es un vector $X_i = \{\mu'\ v_1'\ v_2'\ ...\ v_i'\ e_i\}$ donde e_i representa la etiqueta. El clasificador asigna el vector $X_i \in \mathbb{R}^l$ donde $l = (d+1)\frac{n(n+1)}{2} + 1$ una de las seis clases $C_1,\ ...,\ C_6$ que representan las acciones. De esta manera, del teorema de Bayes se puede obtener la probabilidad de que dado $X_i \in \mathbb{R}^l$ este pertenezca a la clase C_k con $1 \leq k \leq 6$. Para ello se sabe que cada clase es mutuamente excluyente y se asume que los valores de cada clase se distribuyen de forma normal.

El clasificador asignará el vector X_i a la clase en la cual se haya obtenido el mayor valor de probabilidad entre las seis clases calculadas.

6.4. VALIDACIÓN EXPERIMENTAL

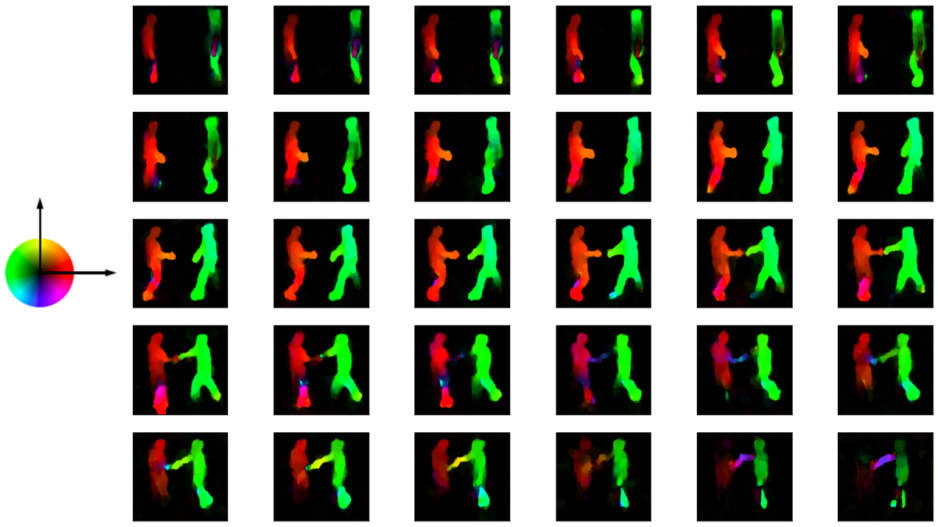
Para evaluar el proceso de clasificación se realiza una validación del modelo propuesto. En este trabajo se utiliza el método *K-fold cross validation* una técnica conocida para evaluar los resultados del análisis estadístico dado por el clasificador. Este método divide el conjunto de datos de forma aleatoria en K grupos donde $K-1$ grupos se utilizan para entrenar el modelo y el grupo faltante se emplea como validación. Este proceso se itera K veces de forma que en cada iteración se utiliza un grupo distinto obteniendo un procentaje de exactitud. El proceso genera K estimaciones que se promedian para dar un porcentaje de acierto final. Esta técnica resulta útil pues posee un alto nivel de precisión ya que se ejecuta sobre K combinaciones y garantiza la independencia entre los grupos de entrenamiento y validación.

7. EVALUACIÓN Y RESULTADOS

Uno de los principales enfoques de esta investigación fue el estudio, análisis e implementación de representaciones basadas en descriptores de covarianza. En este caso los descriptores de video consideran la descomposición en componentes principales y su implementación en video permite una reducción dramática en el costo computacional, una representación eficiente que captura las principales características, discriminatorias entre clases, y su potencial uso en aplicaciones en línea. Para validar el descriptor de video propuesto, en este capítulo se utiliza la base de datos pública, académica mencionada en la sección 6.1, que contiene acciones relacionadas con aplicaciones de video vigilancia. Estos videos son procesados y se presenta una evaluación y los resultados obtenidos utilizando el descriptor propuesto.

Como representación inicial de los videos, y teniendo en cuenta la relevancia del movimiento para describir acciones, se calculó el flujo óptico correspondiente en cada secuencia del cual se utilizan 14 características cinemáticas:
Velocidad, tangente, aceleración y normal por componentes
$(Vx,\ Vy,\ Tx,\ Ty,\ a_x,\ a_y,\ N_x,\ N_y)$, magnitud de la velocidad, aceleración, tangente y normal $(||V||,\ ||a||,\ ||T||,\ ||N||)$, y aceleración tangente y normal $(A_T,\ A_N)$.

Figura 20. Ejemplo del flujo óptico calculado a un video de la secuencia 1, el color denota la dirección del flujo y la intensidad denota la longitud del vector del flujo.



El experimento inicial consistió en comparar los resultados obtenidos utilizando únicamente una media de Riemman con respecto a lo obtenido por la media junto con las componentes principales, calculadas con los dos algoritmos descritos anteriormente: el algoritmo de Fletcher y el algoritmo de Yuchen. Utilizando las características cinemáticas, calculadas del flujo óptico, se describió cada imagen en cada video con una matriz de covarianza de dimensión 14×14 y se procede de acuerdo al modelo experimental presentado. De lo anterior obtenemos los siguientes resultados.

Tabla 1. Resultados del experimento inicial.

	Media	Fletcher	Yuchen
Exactitud	66.95 %	67.79 %	68.84 %

Se compara el descriptor haciendo uso de los dos algoritmos con el descriptor de la media propuesto por Olmos[13] para observar el comportamiento al adjuntar componentes a la media. Vemos que la exactitud describiendo los datos con solo la media fue de $66.95\,\%$ aplicando el algoritmo de Fletcher obtenemos un porcentaje

del $67.79\,\%$ utilizando las primeras siete componentes principales, para el algoritmo de Yuchen se obtuvo un $68.84\,\%$ utilizando tan solo la primera componente principal. Los resultados del descriptor para los algoritmos de Fletcher y Yuchen se obtuvieron usando la media y adjuntando una por una componente principal de tal forma que se selecciona el mejor valor obtenido. De este primer experimento se obtienen resultados favorables pues se observa que los algoritmos representan una mejora en el valor de exactitud, el resultado obtenido por el algoritmo de Yuchen supone una mejora del $1.89\,\%$ respecto al valor alcanzado solo por la media donde solo fue necesario adjuntar la primera componente principal.
Adicional a ello se aplican los 4 métodos para determinar el número óptimo de componentes y observar el comportamiento entre este primer experimento y los resultados dados por los métodos presentados a continuación

- **Varianza acumulada** Para nuestro experimento tenemos que el número de componentes a utilizar para asegurar un $99\,\%$ de varianza son las 5 primeras componentes principales.

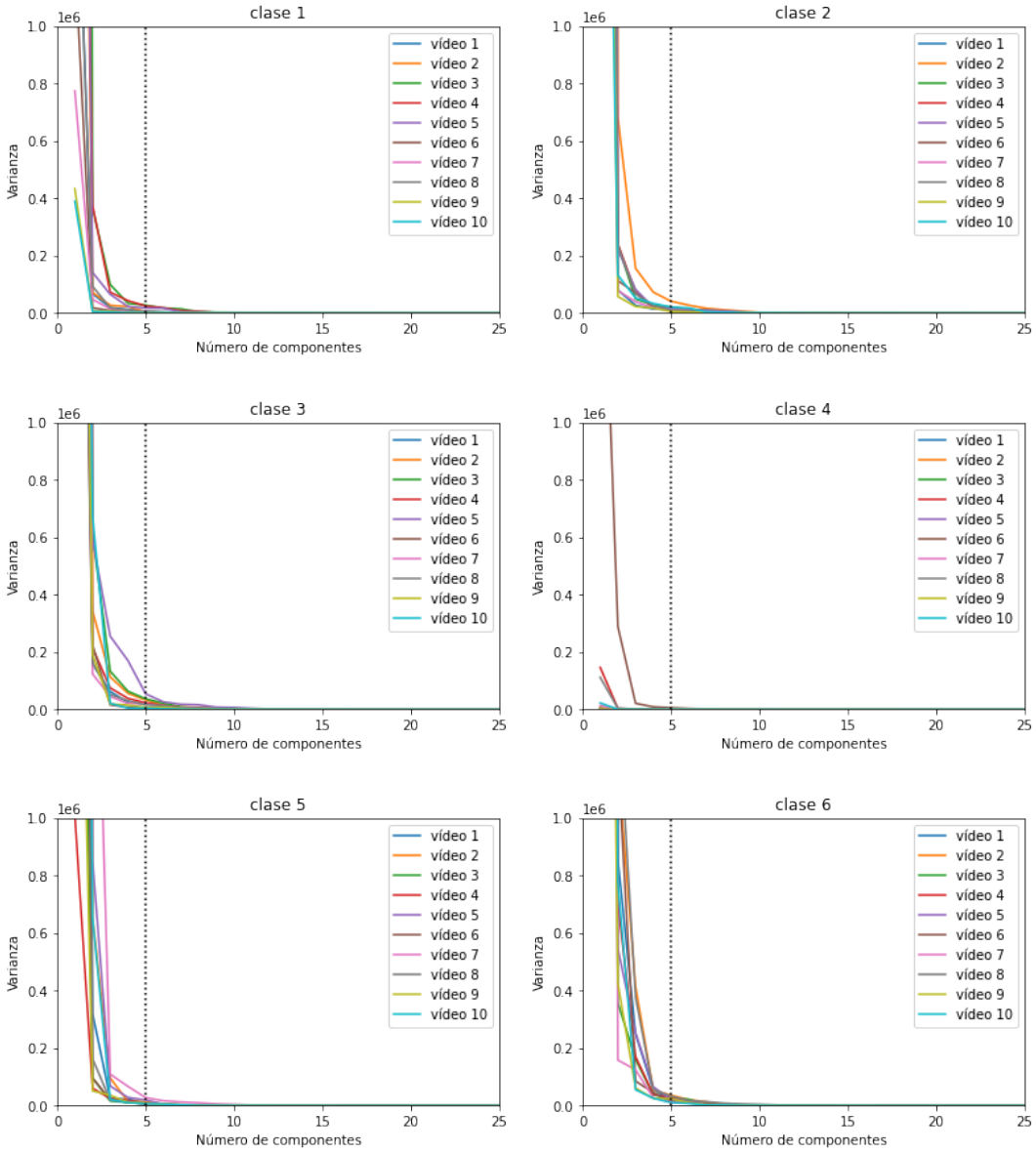
t^*	$90\,\%$	$95\,\%$	$99\,\%$
m	2	2	5

- **Regla de Kaiser** Al igual que el método de varianza acumulada obtenemos un valor de 5 componentes principales a usar.

	$\bar{l}$	$0.7\bar{l}$
m	5	6

- **Cattell** Graficando k vs l_k obtenemos

Figura 21. Gráficas número de componentes vs varianza respectiva para cada clase del conjunto de datos.

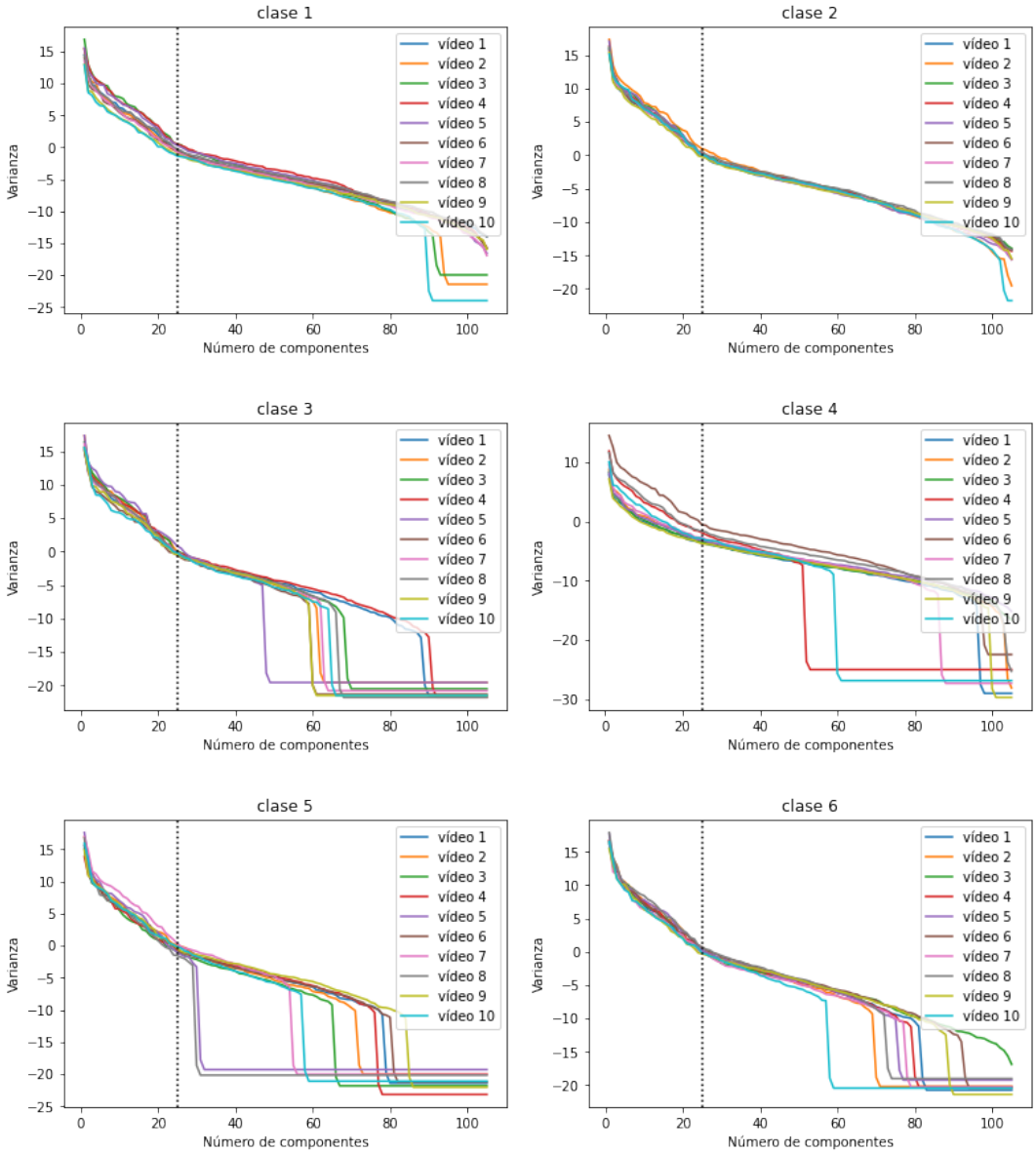


Por cada clase se graficaron los 10 videos correspondientes de tal forma que se tuviera una forma clara de comparar todas las varianzas en la clase, al observar las gráficas de cada clase podemos notar que cinco componentes principales es un buen valor para utilizar, es decir en cinco componentes se

obtiene de manera general para las seis clases el “codo” que nos indica el corte a hacer.

- **Farmer** Ahora con la variación k vs $log(l_k)$ obtenemos

Figura 22. Gráficas número de componentes vs logaritmo de la varianza respectiva para cada clase del conjunto de datos.



En este caso, un buen punto de corte visualmente será tomar 25 componentes principales ya que este valor cumple la condición descrita por Farmer para las seis clases.

Los primeros tres métodos coinciden en un valor cercano al 5 mientras que el cuarto da un valor mayor, recordemos que tanto el tercer como cuarto método pueden dar resultados subjetivos. De los resultados obtenidos en la Tabla 1 podemos ver una relación entre las componentes usadas y los valores de corte obtenidos por los métodos ya que para el algoritmo de Fletcher fueron necesarias 7 componentes mientras que para el algoritmo de Yuchen tan solo se usó la primera, valores menores o cercanos a 5.

Posteriormente, se exploran otras formas de obtener características; entre ellas, redes neuronales convolucionales. Se escoge la red convolucional MobileNetV2 [24] la cual tiene un tamaño de entrada de imagen de 224 por 224. MobileNetV2 tiene 53 capas de profundidad, sin embargo nos interesan las primeras capas ya que, se desean características simples como líneas y/o curvas, de esta manera, luego de observar y experimentar diferentes capas iniciales se selecciona la capa 4 la cual devuelve 32 características de las cuales se eligen 15 características más representativas.

---

[24] Mark SANDLER y col. "MobileNetV2: Inverted Residuals and Linear Bottlenecks". En: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018).

Figura 23. Ejemplo de las 15 características seleccionadas para una imagen con MobileNetV2.

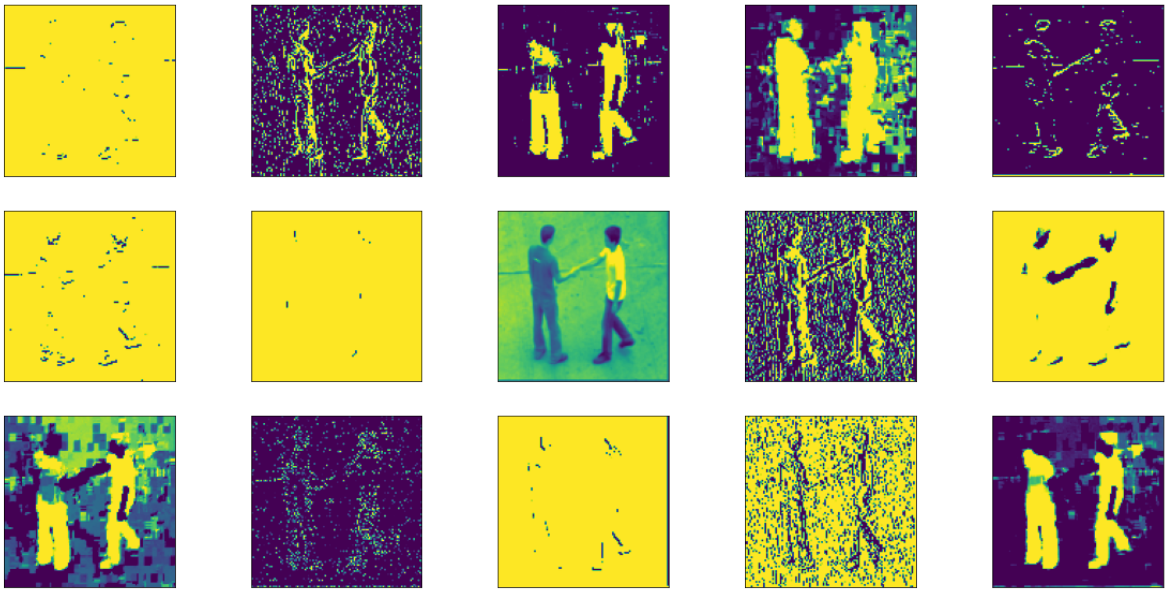


Tabla 2. Resultados del experimento con las características dadas por la MobileNetV2.

	Media	Fletcher	Yuchen
Exactitud	58.47 %	63.55 %	63.55 %

Para esta ocasión se obtuvo, usando solo la media una exactitud del $58.57\,\%$, usando el algortimo de Fletcher se obtiene $63.55\,\%$ andjuntando las primeras dos componentes principales, y con el algortimo de Yuchen se obtiene un porcentaje de $63.55\,\%$ adjuntando solo la primera componente principal. Nuevamente el algoritmo de Yuchen supone un aumento en este caso del $5.08\,\%$ sobre el valor alcanzado usando solo la media, es de destacar que este algoritmo mantiene la tendencia de sus resultados en ambos experimentos, de las Tablas 1 y 2 vemos como este algoritmo supera el valor de la exactitud usando en ambos casos tan solo la primera componente principal.

8. CONCLUSIONES

En el desarrollo de este trabajo se estudió las matrices de covarianza y el espacio generado por ellas, también se caracterizó este espacio como una variedad Riemanniana lo que permitió introducir conceptos necesarios para el cálculo de distancias y geodésicas. A su vez se realizó un estudio al proceso clásico del ACP mediante un enfoque geométrico y de optimización dando lugar a la extensión de nociones que nos permitieran extender este proceso particularmente al espacio de las matrices simétricas definidas positivas $PD(n)$. Con los conceptos gneralizados se procede finalmente a estudiar el proceso análogo al del ACP en un espacio no Euclidiano dando lugar al Ananálisis de Geodésicas Principales (AGP). No obstante, en el proceso de la generalización se presenta un problema difícil de optimización que hasta el momento sigue en estudio, se presenta entonces una aproximación mediante el espacio tangente con lo cual se obtiene una forma de computar el AGP de forma lineal. La anterior aproximación trae consigo errores pero hasta el momento es una buena forma de acercarnos a la noción del AGP. El estudio del espacio $PD(n)$ es motivado por su aplicación en el análisis de video, de esta manera se aborda el AGP con el propósito de proponer una forma de describir acciones en una secuencia de video teniendo en cuenta las direcciones de varianza en los datos proporcionados por estas secuencias. A partir de lo anterior y los resultados obtenidos en el presente trabajo de grado, es posible enunciar las siguientes conclusiones.

1. Se logra mostrar la forma en que el AGP puede ser computado mediante una aproximación en el espacio tangente.

2. Como se esperaba, en las Tablas 1 y 2 se observa un aumento en la exactitud obtenida, esto sustenta la idea que adjuntar componentes a la media ayuda a mejorar la descripción. De hecho los componentes principales mejoran el

modelamiento de la varianza intra clase, logrando un mejor comportamiento para la identificación automática de las acciones.

3. El estudio en el número de componentes no determina de buena forma este valor, sin embargo en las cuatro formas estudiadas se puede observar un valor estándar fijo que puede indicar un valor piso para el cual la exactitud disminuye.

4. En el estudio de los dos algoritmos se evidencia la importancia de conservar la métrica sobre el espacio, recordemos que para el algoritmo de Fletcher se calcula el ACP usual con la métrica Euclidiana en el espacio tangente, mientras que se trabaja con la métrica afín-invariante definida sobre el espacio tangente. Por otro lado el algoritmo de Yuchen traslada los puntos al espacio donde estas dos métricas coinciden. Efectivamente los resultados obtenidos son mayores o iguales con el algoritmo de Yuchen a su vez que hace uso de menos componentes que el algoritmo de Fletcher, adicionalmente se evidencia un mejor comportamiento en el algoritmo de Yuchen, observando el Anexo (1) vemos un comportamiento que tiende a disminuir cuando se usan más componentes, de igual forma en el Anexo (2) el algoritmo de Fletcher para cualquier versión estudiada fluctua entre los diferentes números de componentes usadas mientras que el algoritmo de Yuchen en ambas versiones mantiene un comportamiento que tiende a bajar a medida en que se adjuntan más componentes.

BIBLIOGRAFÍA

BOOTHBY, William M. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Second Edition. ACADEMIC PRESS, INC., 1986 (vid. págs. 19, 23, 25).

CARMO, Manfredo DO y Francis FLAHERTY. *Riemannian Geometry*. Birkhäuser, 1992 (vid. págs. 19-22, 24, 25).

FLETCHER, P. Thomas y Sarang JOSHI. "Riemannian Geometry for the Statistical Analysis of Diffusion Tensor Data". En: *National Alliance for Medical Image Computing (NAMIC)* 87 (feb. de 2007), págs. 250 -262 (vid. págs. 13, 14, 28, 44).

FRÉCHET, M. "Les éléments aléatoires de nature quelconque dans un espace distancié". En: *Annales de l'I. H. P* 10.4 (1948), págs. 215,310 (vid. págs. 28, 37).

GALLIER, Jean. *Notes on Differential Geometry and Lie Groups*. University of Pennsylvania, Philadelphia, USA, jun. de 2011 (vid. pág. 28).

HELGASON, Sigurdur. *Differential Geometry, Lie Groups, and Symmetric Spaces*. ACADEMIC PRESS, INC., 1978 (vid. págs. 19, 25, 28).

HOREV, Inbal, Florian YGER y Masashi SUGIYAMA. "Geometry-Aware Principal Component Analysis for Symmetric Positive Definite Matrices". En: *Asian Conference on Machine Learning*. Vol. 45. Proceedings of Machine Learning Research. Hong Kong, 2016, págs. 1-16 (vid. pág. 14).

JOLLIFFE, I.T. *Principal Component Analysis (2nd Ed.)* Springer series in statistics, 2002 (vid. págs. 12, 32).

JOST, Jürgen. *Riemannian Geometry and Geometric Analysis*. Seventh Edition. Springer, 2017 (vid. págs. 19, 23, 24).

KARCHER, H. “Riemannian center of mass and mollifier smoothing”. En: *Communications on Pure and Applied Mathematics* 30 (1977) (vid. pág. 29).

LANG, Serge. *Fundamentals of Differential Geometry*. Springer, 1999 (vid. págs. 24, 26).

LEE, John M. *Introduction to Smooth Manifolds*. Second Edition. Springer, 2012 (vid. pág. 19).

MINH, Hà Quang y Vittorio MURINO. *Algorithmic Advances in Riemannian Geometry and Applications For Machine Learning, Computer Vision, Statistics, and Optimization*. Springer International, 2016 (vid. pág. 12).

— *Covariances in Computer Vision and Machine Learning*. Morgan & Claypool Publishers, 2018 (vid. pág. 13).

MOAKHER, Maher y Mourad ZÉRAÏ. “The Riemannian Geometry of the Space of Positive-Definite Matrices and Its Application to the Regularization of Positive-Definite Matrix-Valued Data.” En: *J Math Imaging Vis 40,* (2011) (vid. pág. 27).

OLMOS, Juan. *Cálculo de medias gemoétricas en el cono de las matrices simétricas semidefinidas positivas*. Bucaramanga, Colombia, 2020 (vid. págs. 26, 29, 36, 54, 64, 75).

OR, Yair, Ben-Chen MIRELA y Talmon RONEN. “Parallel Transport on the Cone Manifold of SPD Matrices for Domain Adaptation”. En: *IEEE Transactions on Signal Processing* 67 (2019), 1797–1811 (vid. pág. 47).

PENNEC, Xavier, Pierre FILLARD y Nicholas AYACHE. “A Riemannian Framework for Tensor Computing”. En: *RR-5255, INRIA* (2004), pág. 34 (vid. pág. 26).

REA, Alethea y William REA. “How Many Components should be Retained from a Multivariate Time Series PCA?” En: (oct. de 2016) (vid. pág. 35).

RYOO, M. S. y J. K. AGGARWAL. *UT-Interaction Dataset, ICPR contest on Semantic Description of Human Activities (SDHA)*. 2010 (vid. pág. 58).

SANDLER, Mark y col. “MobileNetV2: Inverted Residuals and Linear Bottlenecks”. En: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018) (vid. pág. 68).

SOMMER, Stefan, Francois LAUZE y Mads NIELSEN. “The Differential of the Exponential Map, Jacobi Fields and Exact Principal Geodesic Analysis”. En: *Computing Research Repository - CORR* (2010) (vid. pág. 43).

SÁNCHEZ, César. *Análisis de componentes principales*. Máster en Técnicas Estadísticas. Universidade de Santiago de Compostela, 2008-2009 (vid. págs. 13, 34, 35).

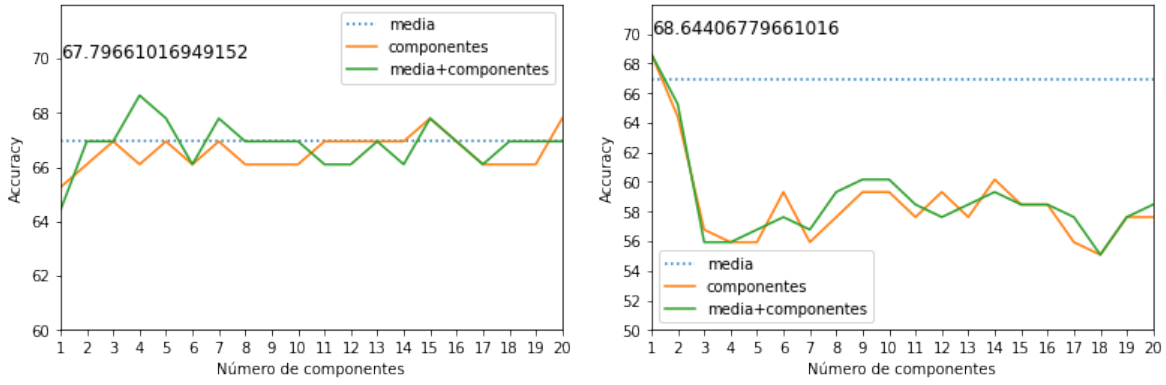
XIE, Yuchen, Baba C. VEMURI y Jeffrey HO. “Statistical Analysis of Tensor Fields”. En: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010* (2010), págs. 682-689 (vid. pág. 47).

ANEXOS

Anexo A. ¿Es conveniente adjuntar componentes a la media para la descripción?

Olmos en su trabajo [13] usa como descriptor la media. Se plantea el hecho que si a la media se adjuntan las direcciones de mayor varianza se puede obtener una mejor clasificación ya que estamos agregando más información sobre el conjunto de datos al descriptor. No obstante el hecho de usar solamente las componentes es decir, no adjuntar la media, podría ser también una buena forma por lo que se realizó un experimento comparando los resultados entre estas dos formas para describir.

Figura 24. Comparación entre los descriptores formados, en la figura de la izquierda aplicando el algoritmo de Fletcher y en la de la derecha usando el algoritmo de Yuchen.



El valor numérico presentado es el mejor valor obtenido en los experimentos realizados donde se usa la concatenación de la media junto a un número determinado de componentes. Se observa que para el algoritmo de Fletcher no hay un patrón que determine diferencias, incluso es inestable en el sentido que realizado esta vez el experimento se tiene que con 4 componentes hay una mejor exacitud de la obtenida en la Tabla 1. Para el algoritmo de Yuchen si hay un comportamiento notable, los

resultados del Capítulo 7 y de esta ocasión concuerdan, a su vez se observa una disminución en la exactitud al aumentar componentes, adicionalmente se observa que concatenar la media junto a un número de componentes dará un resultado levemente superior al de solo usar las componentes.

Anexo B. ¿Es buena la forma en que se está clasificando?

Nos referimos con esta pregunta al espacio del cual se toman los datos para la clasificación.

Para el algoritmo de Fletcher se plantean tres formas de clasificar, dados los puntos sobre $PD(n)$ se mapean al espacio tangente a la media $T_\mu PD(n)$ donde se aplica el ACP usual con lo cual obtenemos las direcciones de mayor varianza.

- El primer espacio a considerar es el espacio $T_\mu PD(n)$, una vez obtenidas las componentes se crea el descriptor con el cual se clasifica, esta primera versión la llamaremos Fletcher V1.

- El siguiente espacio a considerar es el espacio $PD(n)$, una vez obtenidas las componentes en $T_\mu PD(n)$ mapeamos de regreso tales componentes a $PD(n)$ con las cuales se crea el descriptor para la clasificación, nos referiremos a esta segunda versión como Fletcher V2.

- La tercera versión será llevar nuestras componentes de $PD(n)$ al espacio tangente en la identidad $T_I PD(n)$ con las cuales se crea el descriptor y se clasifica, esta versión la llamaremos Fletcher V3.

Para el algoritmo de Yuchen se sigue de igual forma, dados los puntos sobre $PD(n)$ se mapean al espacio tangente a la media $T_\mu PD(n)$ y posteriormente son llevados al espacio tangente en la identidad $T_I PD(n)$ donde se aplica el ACP usual, con lo cual obtenemos las direcciones de mayor varianza, para esta ocasión consideramos solo el espacio $T_I PD(n)$ para la clasificación.

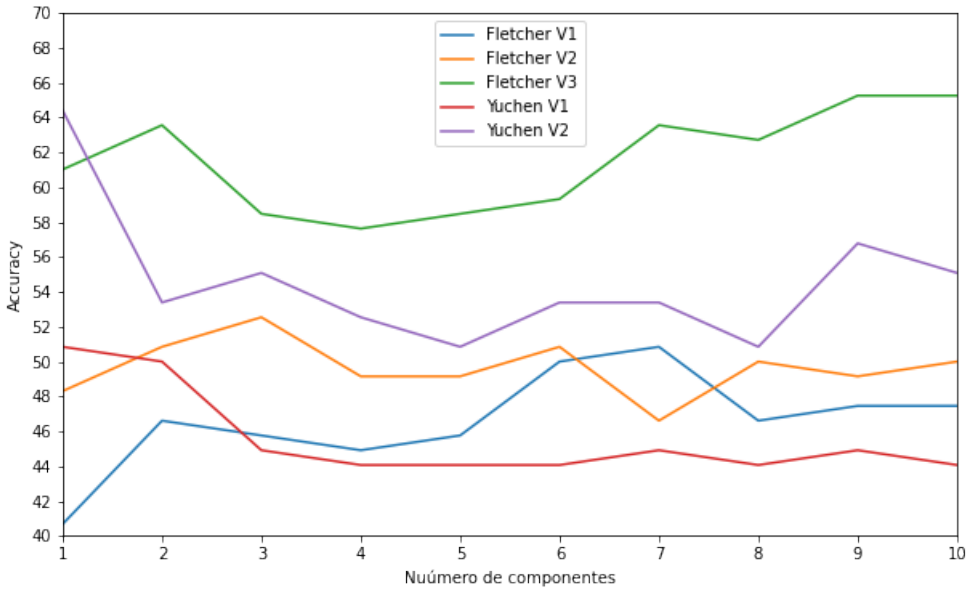
- La primera versión será crear el descriptor con las componentes en $T_I PD(n)$ y clasificar sobre este espacio, para esta versión nos referiremos como Yuchen V1.

- La segunda versión será, obtenidas las componentes en $T_I PD(n)$ mapearlas a $T_\mu PD(n)$, posteriormente llevarlas a $PD(n)$ y finalmente mapearlas de nuevo a

$T_I PD(n)$, donde se crea el descriptor y se clasifica, llamaremos a esta versión Yuchen V2.

Los resultados de los experimentos realizados a las cinco formas de clasificar se resumen en la siguiente figura.

Figura 25. Resultados para las cinco diferentes formas de clasificar, donde se grafica número de componentes versus la exactitud alcanzada.



Las formas que se tuvieron en cuenta para los resultados expuestos en el Capítulo 7, para el algoritmo de Fletcher y Yuchen fueron Fletcher V3 y Yuchen V2 respectivamente, ambas formas son las de más computo, esto motiva la idea que estas formas pueden acarrear consigo mayor error númerico, sin embargo de la Figura 25 vemos que estas dos formas son las de mejor exactitud.