



Pontificia Universidad  
**JAVERIANA**  
Cali

**Facultad de Ingeniería y Ciencias  
Secretaría de la Facultad**

**Acta de Correcciones al Proyecto de Grado  
Matemáticas Aplicadas**

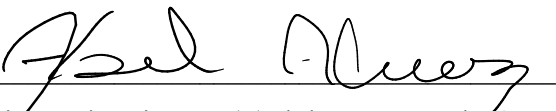
**Fecha:** 18 de Enero de 2022

**Autores:** Fabian David Navarro Cano.

**Nombre del Proyecto de Grado:** COMPORTAMIENTO EPIDEMIOLÓGICO COVID-19 USANDO EL MÉTODO DE DESCOMPOSICIÓN EN MODOS DINÁMICOS PARA LA CIUDAD DE SANTIAGO DE CALI.

**Director:** Abel Álvarez Bustos.

Como indica el artículo 2.27 de las Directrices de Trabajo de Grado, he verificado que los estudiantes indicados arriba han implementado todas las correcciones que los Jurados del Proyecto de Grado definieron que se efectuaran, como consta en el Acta de Calificación correspondiente.




---

Firma de Director(a) del Proyecto de Grado

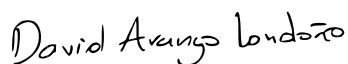
Nota de Aceptación

Aprobado por el Comité de Trabajo de Grado en cumplimiento de los requisitos exigidos por la Pontificia Universidad Javeriana para optar el título de Profesional en Matemáticas Aplicadas.

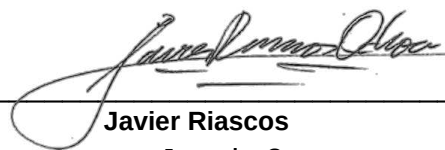
  
Hernán Camilo Rocha Niño  
Decano de la Facultad de Ingeniería y Ciencias

  
Diana Haidive Bueno Carreño  
Directora Carrera de Matemáticas Aplicadas

  
Abel Álvarez Bustos  
Director(a) Trabajo



David Arango Londoño  
Jurado 1



Javier Riascos  
Jurado 2

Santiago de Cali, 17 de diciembre de 2021.

Señores

**Pontificia Universidad Javeriana Cali.**

Dr. Diana Haidive Bueno Carreño.

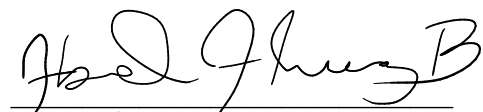
Directora de la carrera Matemáticas Aplicadas.

Cali.

Cordial Saludo.

Por medio de la presente me permito informarle que el estudiante de Matemáticas Aplicadas *Fabian David Navarro Cano*. (cod: 8926542) trabajo y finalizo bajo mi dirección, el proyecto de grado titulado **“Comportamiento epidemiológico COVID-19 usando el método de Descomposición en Modos Dinámicos para la ciudad de Santiago de Cali.”**, el cual considero se encuentra en consideraciones para ser sometido a evaluación.

Atentamente,



---

Dr. Abel Álvarez Bustos.

Director del proyecto de grado

Santiago de Cali, 17 de diciembre de 2021.

Señores

**Pontificia Universidad Javeriana Cali.**

Dr. Diana Haidive Bueno Carreño.

Directora de la carrera Matemáticas Aplicadas.

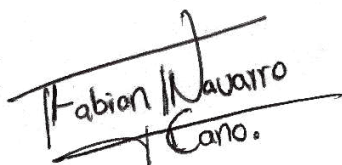
Cali.

Cordial Saludo.

Me permito presentar a su consideración el proyecto de grado titulado “**Comportamiento epidemiológico COVID-19 usando el método de Descomposición en Modos Dinámicos para la ciudad de Santiago de Cali.**” con el fin de cumplir con los requisitos exigidos por la Universidad para optar por el título de Profesional en Matemáticas Aplicadas.

Al firmar aquí, doy fe que entiendo y conozco las directrices para la presentación de trabajos de grado de la Facultad de Ingeniería aprobadas el 26 de Noviembre de 2009, donde se establecen los y norma para los desarrollos del anteproyecto y trabajo de grado.

Atentamente,

A handwritten signature in black ink that reads "Fabian David Navarro Cano". The signature is written in a cursive style and is positioned above a horizontal line.

---

*Fabian David Navarro Cano*

C.c. 1.144.201.933

Cod: 8926542

Pontificia Universidad Javeriana Cali  
Facultad de Ingeniería y Ciencias.  
Matemáticas Aplicadas.  
Proyecto de Grado.

**Comportamiento epidemiológico COVID-19 usando el  
método de Descomposición en Modos Dinámicos para  
la ciudad de Santiago de Cali.**

*Fabian David Navarro Cano.*

Director:  
Dr. Abel Álvarez Bustos.

Santiago de Cali, Colombia  
17 de diciembre de 2021



Pontificia Universidad  
**JAVERIANA**  
Cali

con Acreditación  
Institucional  
de Alta Calidad  
por 8 años



# Resumen

El SARS-CoV-2 es un virus que produce una enfermedad respiratoria conocida como coronavirus del 2019 (COVID-19). Este pertenece a la familia de los coronavirus, siendo un tipo de virus que infecta a seres humanos y algunos animales alrededor del mundo. De hecho, la infección por SARS-CoV-2 fue identificada por primera vez en personas expuestas a un mercado de mariscos en el sector de Wuhan, China, sobre el mes de noviembre del 2019, [22]. Desde entonces, ha sido una de las problemáticas de salud pública más grandes de la historia de la humanidad en los últimos siglos, debido a que este virus tiene una rápida tasa de transmisión y contagio. Es así como, la Organización Mundial de la Salud (OMS) emitió un estado de pandemia en cuestión de pocos meses, ocasionando grandes pérdidas económicas a nivel mundial. Dada la dimensión de esta problemática, se ha despertado el interés por parte de diferentes equipos científicos con el objetivo de identificar el comportamiento de la propagación, a fin de hallar una forma pertinente para su tratamiento.

En virtud de esta situación, en este proyecto de grado se presentará un método emergente impulsado por datos, denominado Descomposición en Modos Dinámicos (DMD), el cual permitirá conocer la dinámica epidemiológica de manera espacio-temporal sobre un periodo corto de tiempo [25, 15], utilizando los reportes generados dentro del área metropolitana de Santiago de Cali, Colombia, ver [4]. En este sentido, se emplearán las herramientas computacionales MATLAB y R para ejecutar las diferentes simulaciones que serán de utilidad para comprender el comportamiento de este virus; como de igual manera, se analizarán algunos conceptos complementarios y auxiliares que serán pertinentes para tener un adecuado manejo del método DMD, como son la Descomposición en Valores Singulares (SVD), el Análisis de Componentes Principales (PCA), el Análisis de Componentes Independientes (ICA) y finalmente el Análisis de Koopman.

**Palabras Clave:** Descomposición en Modos Dinámicos, Coronavirus, DMD, COVID-19, SARS-CoV-2.





# Abstract

The SARS-CoV-2 is a virus that produces a respiratory disease known as coronavirus 2019 (COVID-19). This belongs to the family of coronaviruses, being a type of virus that infects humans and some animals around the world. In fact, SARS-CoV-2 infection was identified for the first time in people exposed to a seafood market in the Wuhan sector, China, around November 2019, [22]. Since then, it has been one of the biggest public health problems in the history of mankind in recent centuries, since this virus has a rapid rate of transmission and contagion. Thus, the World Health Organization (WHO) issued a state of pandemic in a matter of a few months, causing great economic losses worldwide. Given the dimension of this problem, interest has been aroused by different scientific teams to identify the behavior of the spread, to find a relevant way for its treatment.

By virtue of this situation, in this degree project an emergent method driven by data will be presented, called Dynamic Mode Decomposition (DMD), which will allow to know the epidemiological dynamics in a spatio-temporal way over a short period of time [25, 15], using the reports generated within the metropolitan area of Santiago de Cali, Colombia, see [4]. In this sense, the computational tools MATLAB and R will be used to run the different simulations that will be useful to understand the behavior of this virus; Likewise, complementary, and auxiliary concepts that will be relevant to have an adequate handling of the DMD method will be analyzed, such as Singular Value Decomposition (SVD), Principal Component Analysis (PCA), Independent Component Analysis (ICA) and finally the Koopman Analysis.

**Keywords:** Dynamic Mode Decomposition , Coronavirus, DMD, COVID-19, SARS-CoV-2.



# Índice general

<b>1. Planteamiento del Proyecto</b>	<b>3</b>
1.1. Descripción . . . . .	3
1.2. Antecedentes . . . . .	3
<b>2. Desarrollo del Proyecto</b>	<b>7</b>
2.1. Marco Teórico . . . . .	7
2.1.1. Descomposición en valores singulares (SVD). . . . .	7
2.1.2. Análisis de Componentes Principales (PCA). . . . .	9
2.1.3. Análisis de Componentes Independientes (ICA) . . . . .	11
2.1.4. Descomposición en Modos Dinámicos (DMD) . . . . .	13
2.1.5. Análisis de Koopman . . . . .	19
<b>3. Resultados del Proyecto</b>	<b>25</b>
3.1. Covid19 . . . . .	25
3.1.1. Tratamiento de datos . . . . .	25
3.1.2. Interpretación básica de los datos . . . . .	29
3.1.3. Implementación del método DMD en el Covid19 . . . . .	31
3.1.4. Conclusiones . . . . .	43
<b>4. Conclusiones del Proyecto y Trabajos Futuros</b>	<b>45</b>
<b>A. Anexo: Tópicos teóricos</b>	<b>47</b>
A.1. Descomposición en Valores Singulares (SVD). . . . .	47
<b>Bibliografía</b>	<b>57</b>



# Índice de figuras

1.1.	Resultado de reconstruir un campo de flujos TLV desde un estado original [11] . . . .	4
1.2.	Energía espectral y espacial en el tiempo, para los casos o modos de DMD [36] . . . .	4
1.3.	Mapeo moto-sensorial, derivado de los estados o modos de DMD [5] . . . . .	5
1.4.	Resultado sobre la base de datos “Google Flu” por medio de DMD [25] . . . . .	6
2.1.	Descomposición en Valores Singulares $U\Sigma V^*$ de una matriz $A$ con orden $2 \times 2$ . . . .	8
2.2.	Explicación geométrica de PCA en sus dos componentes principales. . . . .	10
2.3.	Modos evolucionados para un módulo de turbulencia en una simulación con plasma. . . .	14
2.4.	Descripción esquemática de un flujo con el comportamiento de DMD [6]. . . . .	19
2.5.	Funcionamiento del Operador de Koopman . . . . .	22
2.6.	Generar sistemas dinámicos de un sistema complejo desconocido en DMD/Koopman. . . .	23
3.1.	Diagrama de flujo para pacientes Covid19 del Ministerio de salud de Colombia [2]. . . .	26
3.2.	Estructura de un <i>Snapshot</i> en la ciudad de Santiago de Cali . . . . .	28
3.3.	Estructura total de <i>Snapshots</i> en la ciudad de Santiago de Cali . . . . .	28
3.4.	Snapshots normalizados de los contagiados por Covid19 a través del tiempo. . . . .	29
3.5.	Estadísticas básicas para el Covid19 en Santiago de Cali. . . . .	30
3.6.	Diagrama de Cajas para el Covid19 en Santiago de Cali. . . . .	31
3.7.	Representación del método de reducción utilizando SVD para DMD. . . . .	33
3.8.	Valores singulares por el método de reducción SVD en el Covid19. . . . .	33
3.9.	Estudio de relevancias que ejecutan los componentes principales. . . . .	34
3.10.	Geometría de la descomposición espectral con sus autovalores $\Lambda$ . . . . .	35
3.11.	Geometría de la descomposición espectral aplicado al Covid19. . . . .	36
3.12.	Magnitudes de comunas que generan aportes de contagio por Covid19. . . . .	37
3.13.	Mapa geográfico con las comunas de mayor aporte de contagio por Covid19. . . . .	37
3.14.	Gráfica que describe las magnitudes de frecuencia para el Covid19. . . . .	38
3.15.	Magnitudes de comunas que generan aportes de frecuencia por Covid19. . . . .	39
3.16.	Mapa de los mayores aportes a la frecuencia de contagio por comuna en el Covid19. . . .	40
3.17.	Reconstrucción y proyección futura de los modos de DMD. . . . .	41
3.18.	Tabla de errores para el modelo DMD entre lo predicho y lo actual. . . . .	42
A.1.	Esquema de matrices, para SVD Completo y SVD económico.[6] . . . . .	54
A.2.	Esquema de matrices para SVD Completo y SVD truncado.[6] . . . . .	55



# Índice de cuadros

2.1. Definición del operador de koopman [15] . . . . .	20
2.2. Definición del método DMD. [15] . . . . .	22
2.3. Definición de la descomposición de los modos de Koopman y el método DMD [15] . . . . .	24
A.1. Definición para la Pseudo-Inversa de Moore Penrose [26] . . . . .	49





# Introducción

Los avances del ser humano se encuentran en interacción con el medio que los rodea, lo cual lleva a concebir diferentes estrategias para avanzar en ámbitos personales, sociales y económicos. Sin embargo, las acciones realizadas para alcanzar estos avances, de alguna manera acarrearán diversos problemas sociales y medioambientales, [19, 21]. De hecho, actualmente, una de las problemáticas en el entorno del ser humano ha ocasionado un virus que produce una grave enfermedad respiratoria a nivel mundial, llevando a cada país a una crisis de salud y escasez en ámbitos socioeconómicos de grandes proporciones.

Esta afectación mundial está dada por el virus SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) siendo este, un tipo de virus que produce una enfermedad respiratoria denominada coronavirus del 2019 o COVID-19; esta enfermedad zoonótica (enfermedad que se transmite entre animales y seres humanos) fue identificada por primera vez en algunas personas en la ciudad de Wuhan, China, dentro de un mercado de productos húmedos. Tal problema de salud tuvo su primer registro el día 17 de noviembre del 2019, lo que derivó en la actualidad (marzo del 2021) en un registro con casos confirmados de más de 125 millones de personas en todos los continentes. La gravedad de esta enfermedad fue tal que conllevó a la Organización Mundial de la Salud (OMS), a declarar el inicio de un estado de pandemia el día 11 de marzo del 2020, cuando presentó una tasa de mortalidad del 3% al 7% aproximadamente [37, 17, 9].

En este sentido, al declarar el estado de pandemia a raíz del virus, se presentó un incremento exponencial de contagio en diferentes sectores del mundo que obligó a sus gobernantes a decretar aislamientos preventivos y cierres de fronteras a fin de mitigar las tasas de contagio, no obstante, realizar este tipo de estrategias, generó efectos secundarios, llevando a la población a un estado de salud grave en conjunto con una gran parálisis de la economía mundial, la cual implicaría millonarias pérdidas monetarias y una crisis socioeconómica por cuestiones de la pandemia, [22]. Dada la dimensión de este problema, se despertó el interés por parte de diferentes equipos científicos de todo el mundo, apostando por una carrera en la identificación del comportamiento de la propagación del virus con el fin de hallar una forma efectiva de tratar esta enfermedad, ver [16].

En virtud de esta situación, en este proyecto de grado se presentará un método emergente impulsado por datos, denominado Descomposición en Modos Dinámicos (DMD), el cual permitirá conocer la dinámica epidemiológica sobre un corto periodo de tiempo [32], utilizando los reportes de Covid19 que fueron generados en la zona metropolitana de Santiago de Cali, ver [4]. En este orden de ideas, el proyecto de grado está desarrollado en **cuatro capítulos** empezando por el **Planteamiento del Proyecto** donde se tendrá una breve descripción del problema a abordar y que antecedentes se han estudiado para el método de interés, luego se tiene el **Desarrollo del Proyecto** donde se definirá el marco teórico con algunos conceptos complementarios y auxiliares que serán de utilidad para el método DMD, tales como la Descomposición en Valores Singulares (SVD), el Análisis de Componentes Principales (PCA), el Análisis de Componentes Independientes (ICA) y por último, el Análisis de Koopman. Seguido de esto, se tendrá los **Resultados del Proyecto** donde se llevará a cabo la teoría que fue estructurada en el capítulo anterior a partir de tres segmentos que son la “Recolección”, “Organización” y “Ejecución” de los datos, finalmente se llega a las **Conclusiones del proyecto** donde se centralizarán las perspectivas e interpretaciones de los resultados que fueron ejecutados bajo las herramientas computacionales MATLAB y R.



# Planteamiento del Proyecto

---

## 1.1. Descripción

El SARS-CoV-2 siendo una de las afectaciones de salud fuertes en la actualidad, ha generado que diversos equipos científicos a nivel mundial se encuentren analizando el comportamiento de propagación del virus Covid19, aunque en muchos casos de estos acercamientos se han realizado por modelos compartimentales, ver [7], puesto que estos modelos presentan ciertos inconvenientes al confrontar los datos obtenidos con la realidad, porque no estudian el virus de manera geográfica y por ende, no es una forma adecuada para planear políticas de tratamiento con este tipo de modelos.

Sin embargo, en virtud de esta situación, se espera que este proyecto de grado al emplear un método emergente impulsado por datos denominado Descomposición en Modos Dinámicos (DMD) el cual no es nuevo y donde se han tenido diferentes acercamientos en contextos variados, tal como se vera en la siguiente sección. Este pueda suplir la deficiencia que presentan los modelos que fueron previamente descritos, ya que este modelo de DMD se ejecuta de manera espacio-temporal captando los datos en un corto lapso de tiempo de los contagiados según la comuna, utilizando los registros emitidos por la secretaria de salud de Santiago de Cali, [4]

## 1.2. Antecedentes

Tal como se menciona en la sección anterior, la Descomposición en Modos Dinámicos es método reciente, de fácil aplicación y con diversos alcances; lo cual ha permitido que investigadores apliquen este proceso en diferentes contextos. Por ende, a continuación se describirán algunos artículos que han venido trabajando el método DMD, desde su creación en el año 2008 por Peter Schmid bajo procesos novedosos en sectores como: la industria de turbinas, mecánica de fluidos, neurociencia, epidemiología y otros campos de interés social.

- **Dynamic mode decomposition and reconstruction of tip leakage vortex in a mixed flow pump as turbine at pump mode [11]**

En este trabajo, publicado en el año 2020 por los autores Yadong Han y Lei Tan, presentan con éxito el tratamiento de reconstrucción gráfica para un campo de flujo dinámico. En él utilizan el método DMD centralizado en un análisis para el comportamiento de un vórtice de fuga, sobre una bomba de turbina mixta (TLV); obteniendo modos de descomposición a

partir de métodos complementarios como la Descomposición Ortogonal Adecuada (POD) que mejoran el rendimiento y pronóstico en sectores industriales.

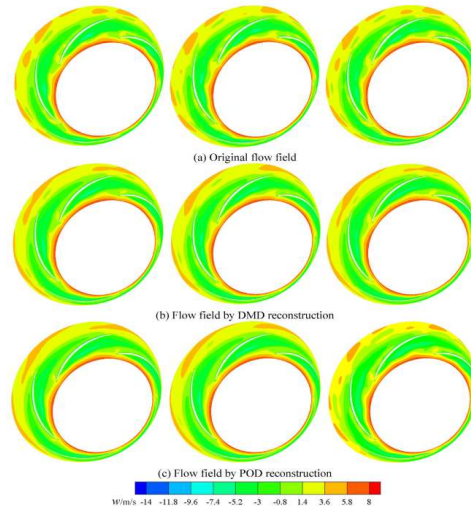


Figura 1.1: Resultado de reconstruir un campo de flujos TLV desde un estado original [11]

- **The identification of coherent structures using proper orthogonal decomposition and dynamic mode decomposition [36]**

En este otro trabajo el cual fue publicado en el año 2013 por 3 matemáticos chinos, presentan un estudio exhaustivo de la dinámica del agua como un fluido sobre tubos cilíndricos de diferentes tamaños, destacando los procesos de mejor respuesta, siendo estos los métodos DMD y POD, con la finalidad de verificar estructuras altamente armónicas y coherentes, dentro de todo el campo de fluidos.

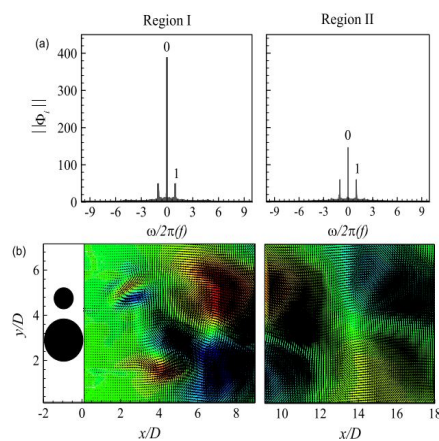


Figura 1.2: Energía espectral y espacial en el tiempo, para los casos o modos de DMD [36]

- **Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition** [5]

Por consiguiente, para este proyecto publicado en el año 2016 por diversos científicos especializados en el ámbito biológico y matemático, deciden acoplar el método DMD, para aproximarse a entender y visualizar los flujos grabados de las actividades neuronales de manera espacio temporal en las lecturas de diferentes electrodos, los cuales captan la interacción de las neuronas respecto a los diversos estímulos físicos y sensoriales.

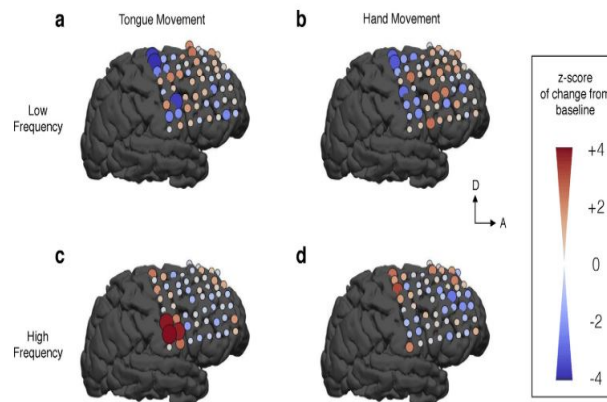


Figura 1.3: Mapeo moto-sensorial, derivado de los estados o modos de DMD [5]

- **Discovering dynamic patterns from infectious disease data using dynamic mode decomposition** [25]

Finalmente, para este trabajo realizado por Joshua L, Proctor y Philip A, Eckhoff, desarrollan una aplicación en el ámbito epidemiológico para verificar el comportamiento de la gripe en diferentes estados de EE. UU, utilizando una base datos denominada “Google Flu” elaborada en el año 2008.

Con este análisis se empleó el método DMD, donde se presentó la dinámica de la gripe y los patrones que se pueden obtener a partir de dicha información a fin de controlar esta enfermedad.

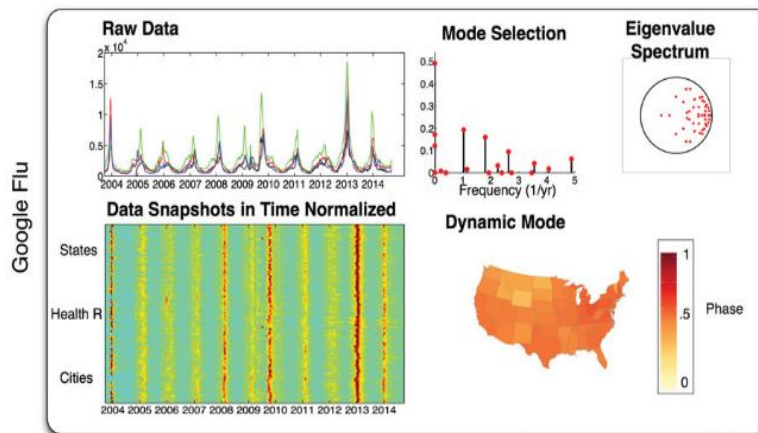


Figura 1.4: Resultado sobre la base de datos “Google Flu” por medio de DMD [25]

En definitiva, aunque el método de Descomposición en Modos Dinámicos sea algo reciente, se han elaborado hasta la actualidad proyectos de alta calidad, donde además de emplear el método DMD, también se utilizan otros procesos complementarios como son la Descomposición Ortogonal Adecuada, la Descomposición en Valores Singulares, el Análisis de Koopman y otras técnicas que posteriormente se irán abordando.

# Desarrollo del Proyecto

---

## 2.1. Marco Teórico

### 2.1.1. Descomposición en valores singulares (SVD).

Históricamente, el método de Descomposición en Valores singulares o *Singular Value Decomposition (SVD)*, se originó en el ámbito de las ciencias sociales a partir de pruebas inteligentes. En sus comienzos, los investigadores de este método analizaron las pruebas ejecutadas, donde evaluaron diferentes indicadores de inteligencia, para los cuales, hallaron una correlación entre lo verbal y lo espacial.

Debido a esto, los investigadores plantearon una hipótesis que mencionaba una medida de la inteligencia general, caracterizada con la letra “g”, elucidando la palabra *General Intelligence* o “Inteligencia General”. Actualmente esta es conocido como “I.Q.” (*Intelligence Quotient* o Coeficiente Intelectual). Por este motivo, los investigadores se dedicaron exclusivamente a analizar diferentes factores de valor que componían la inteligencia, a fin de extraer el factor más importante [18].

La SVD se conoce con diferentes nombres. En sus primeros momentos, debido a lo explicado anteriormente se le llamo “análisis factorial”, otros términos con el cual se identifico fue la “Descomposición de Componentes Principales” (PC) y el “análisis de función ortogonal empírica” (EOF). Todos estos son matemáticamente equivalentes, aunque la forma en que se trata dentro de la literatura es a menudo diferente. Hoy en día, la Descomposición en Valores Singulares se ha extendido a muchas ramas de la ciencia como: la psicología, sociología, climatología, astronomía. Tenido una gran utilidad en áreas como *machine learning* y estadísticas descriptivas como predictivas.

#### 2.1.1.1. Desarrollo del método Descomposición en Valores Singulares.

El método de descomposición en valores singulares es una factorización de matrices con entradas reales o complejas en tres matrices diferentes con propiedades relevantes que posteriormente se relacionarán. De esta forma, dada una matriz  $A$  se tiene:

$$\mathbf{A}_{n \times m} = \mathbf{U}_{n \times n} \mathbf{\Sigma}_{n \times m} \mathbf{V}_{m \times m}^*$$

donde:

1. **La matriz  $U$ :** es un arreglo matricial cuadrado de orden de  $n \times n$ , es conocida como “Matriz de vectores singulares a izquierda”. Esta matriz presenta ortogonalidad, es decir, que satisface la propiedad de  $U^*U = I$ , donde  $I$  es la matriz identidad.
2. **La matriz  $\Sigma$ :** es un arreglo matricial diagonal de orden  $n \times m$ , llamada “matriz de valores singulares”. Los valores de la diagonal  $[\sigma_1, \sigma_2, \dots, \sigma_r]$  con  $r \leq n$  son sus valores singulares y se encuentran ordenados de la forma  $\sigma_i \leq \sigma_j$  donde  $i > j$ .
3. **La matriz  $V$ :** Es un arreglo similar a la matriz  $U$  debido a que presenta ortogonalidad, aunque se conoce como “Matriz vectores singulares a derecha” con un orden de  $m \times m$ .

Se puede caracterizar el método SVD para una matriz  $A$  de tamaño  $n \times m$  con entradas reales o complejas como una transformación geométrica, donde una hiperesfera  $n \times n$  en un hiper elipsoide de  $m \times m$  se encuentra dentro un sistema de coordenadas; A modo de ejemplo, considere  $A$  como una matriz real de orden  $2 \times 2$ , donde mapea un círculo unitario a una elipse tal como se muestra en la figura 2.1. Siendo así, cuando se multiplica las coordenadas del círculo unitario por la matriz  $V^*$  esta genera una rotación de el mismo en el origen, donde luego, al multiplicar por la matriz  $\Sigma$  este círculo se transforma en una elipse, para finalmente al multiplicarse por la matriz  $U$  está rota la elipse en sus coordenadas originales. Por conclusión, note que el sistema de coordenadas  $[v_1, v_2]$  envía su vector  $v_1 \rightarrow \sigma_1 U_1$  y  $v_2 \rightarrow \sigma_2 U_2$ , donde  $[U_1, U_2]$  es otro sistema de coordenadas con  $\sigma_1, \sigma_2$  como números no negativos, ver [28].

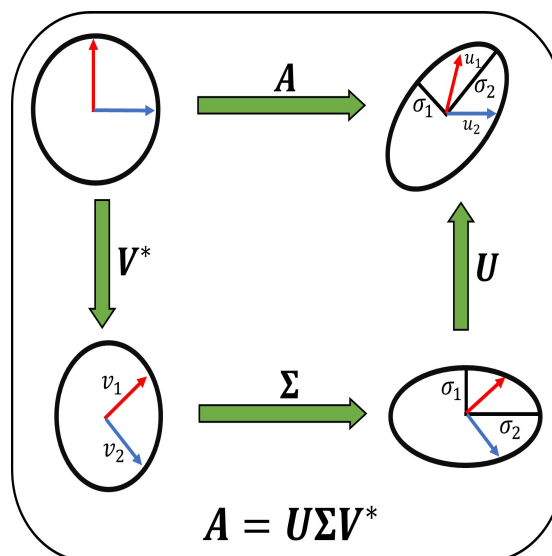


Figura 2.1: Descomposición en Valores Singulares  $U\Sigma V^*$  de una matriz  $A$  con orden  $2 \times 2$ .

**Arriba:** La acción de  $A$  indicando su efecto en el disco unitario  $D$  y sus dos vectores canónicos  $e_1$  y  $e_2$ .

**Izquierda:** La acción de  $V^*$  que genera una rotación de  $D$ ,  $e_1$  y  $e_2$ .

**Abajo:** La acción de  $\Sigma$  donde hace un escalado por los valores singulares  $\sigma_1$  y  $\sigma_2$ .

**Derecha:** La acción de  $U$  que genera otra rotación.



Finalmente, si deseas profundizar en este capítulo de SVD para entender adecuadamente el método de interés que es la Descomposición en Modos Dinámicos, ver el anexo A.1. En este anexo se tendrán diferentes secciones, donde se comprenderán cuales propiedades relevantes para SVD, como es el proceso para calcular este método y algunas variantes tales como: los métodos económicos, truncados y que relación existe entre estos.

### 2.1.2. Análisis de Componentes Principales (PCA).

En esta nueva sección, se pretende examinar la técnica de Análisis de Componentes Principales o *Principal Component Analysis (PCA)*, éste es uno de los procesos centrales más utilizados en el método de la Descomposición en Valores Singulares (SVD); Esta técnica, provee una metodología impulsada por datos con sistemas que representan información en altas dimensiones [6]. Fue inventada en el año de 1901 por Karl Pearson, como analogía a un teorema de la mecánica denominada “principal axis theorem” o teorema de ejes principales [23], tiempo después, se desarrolló de manera independiente en la década de 1930 por Harold Hotelling [12].

Frente a esto, PCA procesa los datos mediante la resta de la media para establecer una varianza con nuevas variables, la geometría resultante está determinada por los Componentes Principales (PCs) no correlacionados entre sí, pero si con las medidas *i.e.* PCA es una técnica para reducir conjuntos de datos grandes, preservando la mayor variabilidad de información [14].

#### 2.1.2.1. Interpretación geométrica del Análisis de Componentes Principales

Para entender el proceso de PCA, se visualizará los componentes principales desde un punto de vista geométrico y gráfico de acuerdo a la figura 2.2. Por tanto, supóngase un conjunto de observaciones para las que se dispone de dos variables ( $\mathbf{X}_1, \mathbf{X}_2$ ). El vector que define la primera componente principal ( $Z_1$ , línea de color roja en 2.2), sigue la dirección donde las observaciones varían más, es decir, la proyección de cada observación sobre esa dirección equivale a la primera componente para dicha observación (*principal component scores,  $z_{i1}$* ) [13].

Por otro lado, la segunda componente principal ( $Z_2$ , línea de color verde en 2.2) sigue una dirección alterna, tal que los datos muestran mayor varianza y no tienen correlación con la primera componente. Puesto que la condición de no correlación entre componentes principales equivale a decir que sus direcciones son perpendiculares/ortogonales.

#### 2.1.2.2. Funcionamiento del Análisis de Componentes Principales

La ejecución de esta técnica, se origina a partir de una matriz de información, denominada  $\mathbf{X}$  de orden  $n \times m$  tal que cada columna  $m$  representa un estado temporal del estudio en  $n$  entradas descriptivas de este estado, como se ve a continuación [6]:

$$\mathbf{X} = [I_1, I_2, \dots, I_m] \quad \text{donde } I_k = 1 \times n, \text{ con } k = [1, m].$$

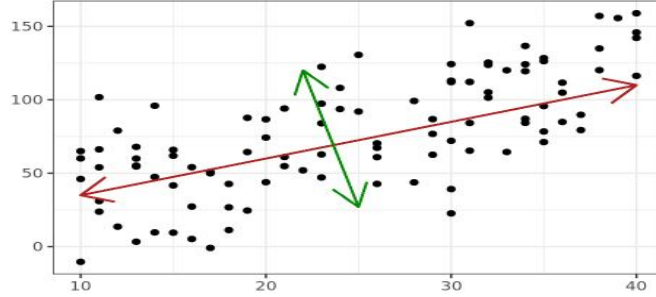


Figura 2.2: Explicación geométrica de PCA en sus dos componentes principales.

PCA de una **distribución normal multivariada** centrada en  $(24, 70)$ , donde los vectores (rojo y verde) muestran los autovectores de la matriz de correlación escalados mediante la raíz cuadrada del correspondiente autovalor, y desplazados para que su origen coincida con la media estadística.

Si se calcula  $\bar{\mathbf{x}}$  la media de todas las filas, se tendría:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n X_{ij},$$

siendo así, la media de la matriz esta dada como:

$$\bar{\mathbf{X}} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \bar{\mathbf{x}}.$$

Si se subtrae  $\bar{\mathbf{x}}$  de  $\mathbf{X}$ , resulta  $\mathbf{B}$  y por esta relación se tiene:

$$\mathbf{B} = \mathbf{X} - \bar{\mathbf{B}}.$$

Luego, la matriz de covarianza en filas de  $\mathbf{B}$  está dada por:

$$\mathbf{C} = \frac{1}{n-1} \mathbf{B}^* \mathbf{B}. \quad (2.1)$$

Justo aquí, se conoce el primer componente principal  $\mathbf{u}_1$ , que está dado como:

$$\mathbf{u}_1 = \underset{\|\mathbf{u}_1\|=1}{\operatorname{argmax}} \mathbf{u}_1^* \mathbf{B}^* \mathbf{B} \mathbf{u}_1,$$

donde el primer autovector de la matriz  $\mathbf{B}^* \mathbf{B}$  corresponde a sus autovalores más grandes y la matriz de covarianza (2.1), tiene una estrecha relación con el método SVD. El arreglo matricial de autovectores  $\mathbf{B}^* \mathbf{B}$  representa a la matriz  $\mathbf{U}$ , que son los componentes principales para las transformaciones de SVD, puesto que crean proyecciones de la forma  $\mathbf{Y} = \mathbf{U}^* \mathbf{X}$ .

### 2.1.2.3. Proporción de varianza en el Análisis de Componentes Principales

Uno de los cuestionamientos más frecuentes que surgen tras haber realizado el Análisis de Componentes Principales es: ¿Cuanta información del conjunto original se pierde al proyectar las observaciones en un espacio de menor dimensión? Para contestar esta pregunta se recurre a la proporción de varianza que explica el comportamiento de cada componente principal [13].

Asúmase que las variables se han normalizado para tener media cero, donde la varianza total en el conjunto de datos está definida como:

$$\sum_{j=1}^p \text{Var}(X_j) = \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n x_{ij}^2,$$

y la varianza explicada por la componente  $m$  es:

$$\frac{1}{n} \sum_{j=1}^n z_{im}^2 = \frac{1}{n} \sum_{i=1}^n \left( \sum_{j=1}^p \phi_{jm} x_{ij} \right)^2.$$

Teniendo que la proporción de varianza acumulada está explicada por:

$$\frac{\sum_{i=1}^n \left( \sum_{j=1}^p \phi_{jm} x_{ij} \right)^2}{\sum_{j=1}^p \sum_{i=1}^n x_{ij}^2}.$$

En conclusión, tanto la proporción de varianza como la varianza acumulada son dos valores útiles al momento de decidir el número de componentes principales a utilizar.

### 2.1.3. Análisis de Componentes Independientes (ICA)

En esta sesión, se estudiara una técnica similar denominada análisis de componentes independientes o *independent component analysis (ICA)*, centralizada en revelar factores ocultos que subyacen en conjunto de datos aleatorios, asumiendo que los subcomponentes no son gaussianos y estadísticamente sus valores son independientes. Fue introducida en la década de 1980 por J.Herault, C.Jutten y B.Ans [31].

El primer problema abordado por la técnica de ICA fue en 1982 dentro un entorno neurofisiológico, donde su objetivo era realizar un modelo simplificado de una contracción muscular a partir de señales cerebrales, de relevancia en el ámbito de las matemáticas, la cual con el pasar del tiempo fue acoplándose a otras áreas.

#### 2.1.3.1. Funcionamiento del Análisis de Componentes Independientes

Tomando una colección de vectores columna los cuales son estados temporales y conforman una matriz  $\mathbf{X}$  de orden de  $n \times m$ , esta se encuentra definida como [34]:

$$\mathbf{Y} = \mathbf{U}^* \mathbf{X}, \quad (2.2)$$

$\mathbf{U}$  es una matriz ortonormal de orden  $n \times n$  y sus componentes principales (columnas de la matriz  $\mathbf{U}$ ) son las direcciones de máxima variación de los datos. Los vectores de componentes principales son los autovectores de la matriz de covarianza  $C_x = \frac{1}{n} \mathbf{X} \mathbf{X}^*$ . Además, según la ecuación (2.2), se tiene que  $\mathbf{Y}$  es la matriz original de datos proyectada sobre el subespacio PCA definido por los autovalores. A partir de esta relación, se puede hacer una reducción de dimensionalidad en los datos, seleccionando un subconjunto de  $k$  autovectores de  $(\mathbb{R}^n \rightarrow \mathbb{R}^k)$ , siendo:

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{U}}^* \mathbf{X}.$$

De esta manera,  $\tilde{\mathbf{Y}}$  es una matriz de orden  $k \times m$  donde  $k$  cumpla  $k \ll n$  y la reducción de dimensión sea considerable. Cabe aclarar que la transformación PCA tiene una ligera perspectiva, considerando que  $\mathbf{U}$  es una de las matrices de descomposición SVD, teniendo:

$$\mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*, \quad (2.3)$$

Si,  $\mathbf{U}$  y  $\mathbf{V}$  son matrices con columnas ortonormales de orden  $n \times r$  y  $m \times r$ ,  $\mathbf{\Sigma}$  es una matriz diagonal no negativa de tamaño  $r \times r$ . Posibilita crear la representación de  $\mathbf{X}$  mediante la suma de  $r$  arreglos de rango unidad:

$$\mathbf{X} = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^*.$$

Lo anterior, implica que los valores singulares nulos sean ignorados, permitiendo realizar una aproximación de  $\mathbf{X}$  mediante los  $k$  autovectores de  $\mathbf{U}$  y  $\mathbf{V}$  de la forma:

$$\mathbf{X} \approx \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^* = \tilde{\mathbf{U}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{V}}^*.$$

Por otro lado, un vector aleatorio centrado en  $\mathbf{z}$  se dice que es *white* si sus elementos no se encuentran correlacionados y presentan varianza igual a la unidad, denotando que su matriz de covarianzas es igual a la matriz unidad  $\mathbf{I}$ . Como la operación de *whitening* puede realizarse mediante una decorrelación seguida de un escalado desde una matriz original, se tiene:

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{\Sigma}}^{-1} \tilde{\mathbf{U}}^* \mathbf{X} = \tilde{\mathbf{V}}^*, \quad (2.4)$$

en el desarrollo de ICA si se toma una matriz de observaciones aleatorias  $\mathbf{X}$ , se puede ver lo siguiente:

$$\mathbf{X} \approx \mathbf{A}\mathbf{S},$$

donde  $\mathbf{S}$  es la matriz que contiene los vectores aleatorios estadísticamente independientes y  $\mathbf{A}$  es la matriz mezcla, que se ve:

$$\mathbf{S} = \mathbf{W}\mathbf{X},$$

$\mathbf{W}$  es equivalente a la matriz pseudoinversa de  $\mathbf{A}$ , es decir,  $\mathbf{W} = \mathbf{A}^\dagger$  ( tabla A.1). En el modelo ICA, los vectores  $\mathbf{w}_i$  se obtienen en relación a las filas de la matriz  $\mathbf{S}$  de distribución no-gaussiana mutuamente decorrelacionada. Para conseguir esto, hacer *whitening* sobre los datos según la ecuación 2,4, luego, buscar una proyección ortogonal no-normal ( $\mathbf{R}$ ) de la siguiente manera:

$$\mathbf{S} = \mathbf{R}^* \tilde{\Sigma}^{-1} \tilde{\mathbf{U}}^* \mathbf{X} = \mathbf{R}^* \tilde{\mathbf{Z}},$$

respecto a esta ecuación, se puede decir que ICA es una operación de *whitening* seguida de una rotación, expresándose de la forma:

$$\mathbf{X} \approx \tilde{\mathbf{U}} \tilde{\Sigma} \tilde{\mathbf{V}}^* = \underbrace{\tilde{\mathbf{U}} \tilde{\Sigma}}_{\mathbf{A}} \underbrace{\tilde{\mathbf{R}}^* \tilde{\mathbf{V}}^*}_{\mathbf{S}} = \mathbf{A}\mathbf{S}.$$

Es importante enfatizar que en el modelo ICA ninguna de las matrices de transformación ( $\mathbf{A}$  o  $\mathbf{W}$ ) están restringidas a arreglos ortogonales [30].

#### 2.1.4. Descomposición en Modos Dinámicos (DMD)

Para desarrollar el método Descomposición en Modos Dinámicos o *Dynamic Mode Decomposition (DMD)*, científicos a nivel mundial han creado diferentes estrategias para conocer el comportamiento de las dinámicas sobre los campos de flujos; solo hasta el año 2008 se desarrolla el novedoso método DMD, con la capacidad de descomponer patrones gráficos obteniendo modos representativos de estos comportamientos dinámicos. El cual utiliza la combinación de la Descomposición Ortogonal Adecuada (POD) y la Descomposición de Valores singulares (SVD).

Para mayor información ver [29].

##### 2.1.4.1. Desarrollo de la Descomposición en Modos Dinámicos

El método de descomposición en modos dinámicos fue desarrollado en el año 2008, por el Phd. Peter Schmid, profesional en el área de la ingeniería aeroespacial, doctor en matemáticas y actual profesor de la universidad *Imperial College London, UK*. El crea este método a fin de contribuir en la comunidad de fluidos dinámicos y así identificar estructuras espacio-temporales, en datos de alta

dimensión, identificando mejores sistemas lineales para adquirir una aproximación al momento de predecir datos.

La naturaleza de DMD se fundamenta en los métodos de Descomposición Ortogonal Adecuada (POD) y el de Descomposición en Valores Singulares (SVD) [8]. Este contraste, da como resultado a una jerarquía de modos, basados en la correlación entre la energía y lo espacial, permitiendo que lo temporal predomine sobre la información. En conclusión, además de reducir las dimensiones de estos estados de funcionamiento, también, provee modos de evolución en el tiempo [6] (Véase la Figura 2.3).

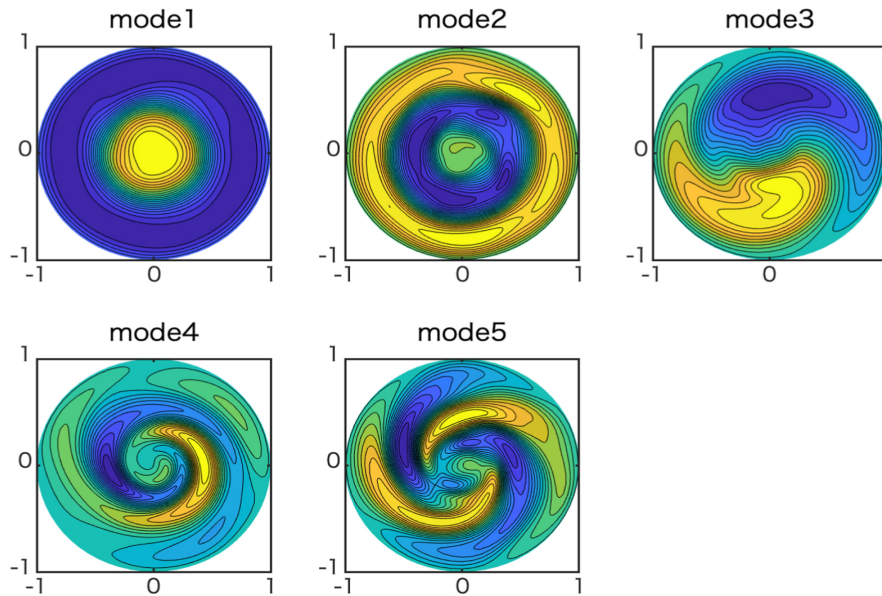


Figura 2.3: Modos evolucionados para un módulo de turbulencia en una simulación con plasma.

*Cinco autovectores dominantes que representan la evolución de una turbulencia en un espacio con patrones de dos dimensiones, donde se derivan de los resultados del análisis de DMD. Estos modos propios dominantes son típicamente estructuras fuertes no lineales, ver [27].*

#### 2.1.4.2. Funcionamiento de la Descomposición en Modos Dinámicos

Desde que surgió la idea del método DMD se han propuesto diferentes formas de representar su funcionamiento, debido a que la definición más técnica, se encuentra relacionada con la eficiencia y el buen condicionamiento del método Descomposición en Valores Singulares. Está se relaciona con una colección de datos desde un sistema dinámico de la forma:

$$\frac{dx}{dt} = f(x, t; \mu), \quad (2.5)$$

donde  $x(t) \in \mathbb{R}^n$  representa los estados del sistema dinámico en un tiempo  $t$  y  $\mu$  contiene los parámetros de este sistema bajo la representación dinámica  $f(\cdot)$ . Generalmente, estos sistemas dinámicos se representa por ecuaciones diferenciales ordinarias que a menudo son no lineales, donde los estados  $\mathbf{x}$  tienen dimensión  $n \gg 1$ , creando la discretización de una ecuación diferencial parcial en un numero de localizaciones espaciales discretas.

Por lo tanto, la ecuación 2.5 conduce a una correspondencia o flujo en tiempo discreto, en la que muestrea el sistema cada  $\Delta t$  en el tiempo como un subíndice  $x_k = x(k\Delta t)$ . En este sentido, al evolucionar este flujo para  $\Delta t$  en  $\mathbf{F}$  se tiene:

$$x_{k+1} = \mathbf{F}(x_k).$$

Con medidas del sistema, iguales a:

$$y_k = \mathbf{g}(x_k),$$

esta colección de medidas temporales  $t_k$  con  $k = 1, 2, \dots, m$ , en muchas aplicaciones son del modo  $y_k = x_k$ ; aunque algunas veces, la comunidad en este ámbito asimila estas medidas con expresiones implícitas  $G(x, t) = 0$ , con condiciones iniciales prescritas como  $x(t_0) = x_0$ .

Ahora bien, previamente se destacó que  $\mathbf{x}$  es un vector  $n$ -dimensional ( $n \gg 1$ ) que surge de la discretización de un sistema complejo, donde la ecuación que emplea y su condición inicial, especifican un problema de valor inicial bien planteado. Sin embargo, no es posible construir soluciones adecuadas a problemas no lineales evolucionados (2.5), por lo que se utilizan soluciones numéricas a fin de evolucionar un sistema a estados futuros.

Siendo así, en el marco de DMD donde la dinámica  $f(x, t; \mu)$  puede ser desconocida. Las mediciones de los datos del sistema solo se utilizan para aproximarse a la dinámica y predecir estados futuros, por ende, DMD construye una aproximación de un sistema dinámico lineal, de la siguiente manera:

$$\frac{dx}{dt} = \mathcal{A}x, \quad (2.6)$$

$\mathcal{A}$  se refiere a la matriz de dinámica en un tiempo continuo y si dada una condición inicial  $x(0)$  se tiene una solución conocida

$$x(t) = \sum_{k=1}^n \phi_k \exp(\omega_k t) b_k = \mathbf{\Phi} \exp(\Omega t) \mathbf{b},$$

donde  $\phi_k$  y  $\omega_k$  son autovectores y autovalores de la matriz  $\mathcal{A}$ , y los coeficientes  $b_k$  son las coordenadas  $x(0)$  en la base de autovectores.

Dada la dinámica lineal continua en (2.6), siempre es posible describir un sistema análogo de tiempo discreto muestreado cada  $\Delta t$  en el tiempo, de la forma:

$$x_{k+1} = \mathbf{A}x_k, \quad (2.7)$$

donde

$$\mathbf{A} = \exp(\mathcal{A}\Delta t).$$

La solución de este sistema es expresada en términos de autovalores  $\lambda_k$  y autovectores  $\phi_k$  que mapean una discretización en el tiempo para  $\mathbf{A}$ , siendo así:

$$x_k = \sum_{j=1}^r \phi_j \lambda_j^k b_j = \Phi \Lambda^k \mathbf{b}. \quad (2.8)$$

Por lo anterior,  $b$  son los coeficientes de la condición inicial  $x_1$  en la base de autovectores, por ende,  $x_1 = \Phi \mathbf{b}$ . El algoritmo DMD produce una descomposición en (2.8) de la matriz  $\mathbf{A}$  que se ajusta a las medidas en la trayectoria  $x_k$  para  $k = 1, 2, \dots, m$ ; teniendo una esencia de mínimos cuadrados de la forma:

$$\|x_{k+1} - \mathbf{A}x_k\|_2, \quad (2.9)$$

minimizando en todos los puntos  $k = 1, 2, \dots, m - 1$ . La optimización de esta aproximación se mantiene únicamente en las muestras construidas para la matriz  $\mathbf{A}$ , y su solución aproximada no solo es usada para realizar predicciones futuras sino también para descomponer la dinámica en varias escalas de tiempo, ya que se prescriben  $\lambda_k$  veces.

Para aproximarse a la minimización del error en (2.9), DMD primero colecciona pares de datos o *snapshots* de un sistema que va evolucionando a medida que transcurre el tiempo; estos pares de *snapshots* se denotan como  $\{\mathbf{x}(t_k), \mathbf{x}(t'_k)\}_{k=1}^m$ , donde  $t'_k = t_k + \Delta t$ , teniendo una variación  $\Delta t$  suficientemente pequeña, como para resolver altas frecuencias dinámicas. Permitiendo que todos los pares de datos se ajusten en dos matrices grandes  $\mathbf{X}$  y  $\mathbf{X}'$ , de la siguiente manera:

$$\mathbf{X} = \begin{bmatrix} | & | & \dots & | \\ \mathbf{x}(t_1) & \mathbf{x}(t_2) & \dots & \mathbf{x}(t_m) \\ | & | & & | \end{bmatrix},$$

$$\mathbf{X}' = \begin{bmatrix} | & | & \dots & | \\ \mathbf{x}(t'_1) & \mathbf{x}(t'_2) & \dots & \mathbf{x}(t'_m) \\ | & | & & | \end{bmatrix}.$$

Peter Schimd y C.W. Rowley originalmente formularon en asumir una muestra uniforme en el tiempo, tal que  $t_k = k\Delta t$  y  $t'_k = t_k + \Delta t = t_{k+1}$ . De ahí, al asumir esta uniformidad de las muestras



en el tiempo, se adopta la siguiente notación  $\mathbf{x}_k = \mathbf{x}(k\Delta t)$ . Por lo tanto, como el algoritmo DMD busca y lidera descomposiciones espectrales para obtener el mejor ajuste lineal de un operador  $\mathbf{A}$  en el tiempo, se obtiene lo siguiente:

$$\mathbf{X}' \approx \mathbf{A}\mathbf{X}.$$

De este modo,  $\mathbf{A}$  es el mejor operador que se ajusta a un sistema dinámico lineal que avanza las medidas de los *snapshots* al futuro, tal como en (2.7). Si se asume que las muestras presentan uniformidad en el tiempo, esto se convierte en:

$$\mathbf{X}_{k+1} \approx \mathbf{A}\mathbf{x}_k.$$

Matemáticamente,  $\mathbf{A}$  esta definido como:

$$\mathbf{A} = \arg \min_{\mathbf{A}} \|\mathbf{X}' - \mathbf{A}\mathbf{X}\|_F = \mathbf{X}'\mathbf{X}^\dagger, \quad (2.10)$$

donde  $\|\cdot\|_F$  es la norma de Frobenius y  $^\dagger$  denota la pseudo-inversa. Estos procesos, optimizan el algoritmo de DMD, para realizar regresiones sobre sus tiempos dinámicos, a fin de mejorar los modos del método y sus valores propios.

A partir de lo anterior, en teoría,  $\mathbf{A}$  contiene  $n^2$  elementos, lo cual hace que representar este operador o calcular una descomposición espectral del mismo, sea una cuestión intratable. En cambio, el algoritmo DMD aprovecha la reducción de dimensionalidad para calcular sus autovalores y autovectores sin requerir ninguna operación explícita usando  $\mathbf{A}$  directamente. En particular, la pseudo-inversa  $\mathbf{X}^\dagger$  en (2.10), es calculada por la descomposición en valores singulares de la matriz  $\mathbf{X}$ . Debido que esta matriz, normalmente presenta menos columnas que filas, es decir,  $m \ll n$  donde a lo máximo  $m$  serían sus valores singulares distintos de cero, haciendo que la matriz  $\mathbf{A}$  sea de rango  $m$ .

Por lo tanto, en lugar de calcular  $\mathbf{A}$  directamente, se calcula la proyección de  $\mathbf{A}$  sobre sus vectores singulares, creando una matriz  $\tilde{\mathbf{A}}$  de orden  $m \times m$  que permite obtener una aproximación a los modos de DMD. Finalmente, este algoritmo se da en los siguientes pasos:

**Paso 1:** Calcular el método de descomposición en valores singulares para  $\mathbf{X}$ :

$$\mathbf{X} \approx \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^\top,$$

donde  $\tilde{\mathbf{U}} \in \mathbb{C}^{n \times r}$ ,  $\tilde{\mathbf{\Sigma}} \in \mathbb{C}^{r \times r}$ ,  $\tilde{\mathbf{V}} \in \mathbb{C}^{m \times r}$ , y  $r \leq m$  denotando el rango exacto o aproximado de la matriz  $\mathbf{X}$ . En la práctica, elegir un rango aproximado  $r$  es uno de los pasos importantes para la reducción de dimensionalidad en el método DMD, adicionalmente, las columnas de la matriz  $\tilde{\mathbf{U}}$  y  $\tilde{\mathbf{V}}$  son conocidas como modos de la Descomposición Ortogonal Adecuada (POD), ya que satisface

respectivamente  $\tilde{U}^\top \tilde{U} = I$  y  $\tilde{V}^\top \tilde{V} = I$  dado que son matrices ortonormales.

**Paso 2:** De acuerdo a (2.10), la matriz  $\mathbf{A}$  se obtiene por la pseudo-inversa de  $\mathbf{X}$ :

$$\mathbf{A} = \mathbf{X}' \tilde{\mathbf{V}} \tilde{\Sigma}^{-1} \tilde{\mathbf{U}}^\top,$$

sin embargo, como solo interesan los autovalores y autovectores de la matriz  $\mathbf{A}$ , se realiza una proyección de esta matriz en los modos POD de  $\mathbf{U}$ , siendo de la siguiente manera:

$$\tilde{\mathbf{A}} = \tilde{\mathbf{U}}^\top \mathbf{A} \tilde{\mathbf{U}} = \tilde{\mathbf{U}}^\top \mathbf{X}' \tilde{\mathbf{V}} \tilde{\Sigma}^{-1}.$$

La clave hasta ahora, es que la matriz  $\tilde{\mathbf{A}}$  presenta los mismos autovalores distintos de cero que la matriz  $\mathbf{A}$  completa. Por lo tanto, es óptimo calcular la matriz reducida  $\tilde{\mathbf{A}}$  directamente, que su matriz completa. Por otro lado, la matriz reducida  $\tilde{\mathbf{A}}$  define un modelo lineal para el vector de coeficientes  $\tilde{\mathbf{x}}$  de POD, con la dinámica:

$$\tilde{\mathbf{x}}_{k+1} = \tilde{\mathbf{A}} \tilde{\mathbf{x}}_k,$$

note que  $\tilde{\mathbf{U}}$  proporciona un mapa para reconstruir el estado completo de  $\mathbf{x}$ , a partir de la reducción de  $\tilde{\mathbf{x}} : \mathbf{x} = \tilde{\mathbf{U}} \tilde{\mathbf{x}}$ .

**Paso 3:** Calcular la descomposición espectral de  $\tilde{\mathbf{A}}$ :

$$\tilde{\mathbf{A}} \mathbf{W} = \mathbf{W} \Lambda,$$

las entradas de la matriz diagonal  $\Lambda$  son los autovalores de DMD, además, corresponden a los autovalores de la matriz completa  $\mathbf{A}$ . las columnas de  $\mathbf{W}$  representan los autovectores del arreglo  $\tilde{\mathbf{A}}$ , donde proporciona una transformación de coordenadas que diagonaliza esta matriz y estas columnas, siendo así, pueden considerarse como combinaciones lineales del modo POD, con un patrón temporal único dado por  $\lambda$ .

**Paso 4:** Los modos de DMD  $\Phi$ , se construyen utilizando los autovectores  $\mathbf{W}$  del sistema reducido y la matriz de *snapshots* en el tiempo  $\mathbf{X}'$ , de acuerdo con:

$$\Phi = \mathbf{X}' \tilde{\mathbf{V}} \tilde{\Sigma}^{-1} \mathbf{W},$$

estos modos de DMD, sorprendentemente son autovectores de la matriz completa  $\mathbf{A}$ , en virtud de los autovalores  $\Lambda$ , como se muestra a continuación (más información ver [6]):

$$\begin{aligned} \Lambda \Phi &= (\mathbf{X}' \tilde{\mathbf{V}} \tilde{\Sigma}^{-1} \tilde{\mathbf{U}}^\top) \underbrace{(\mathbf{X}' \tilde{\mathbf{V}} \tilde{\Sigma}^{-1} \mathbf{W})}_{\tilde{\mathbf{A}}} \\ &= \mathbf{X}' \tilde{\mathbf{V}} \tilde{\Sigma}^{-1} \tilde{\mathbf{A}} \mathbf{W} \\ &= \mathbf{X}' \tilde{\mathbf{V}} \tilde{\Sigma}^{-1} \mathbf{W} \Lambda \\ &= \Phi \Lambda. \end{aligned}$$

Finalmente, este método DMD se puede ver gráficamente de la siguiente manera:

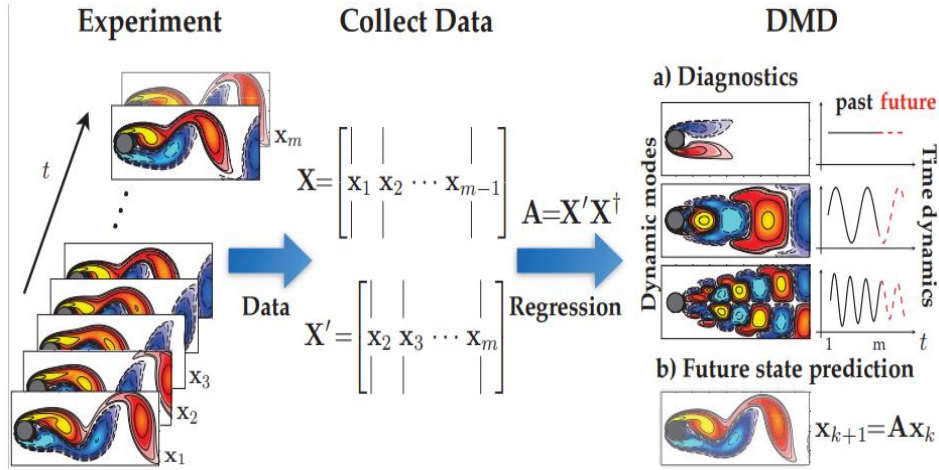


Figura 2.4: Descripción esquemática de un flujo con el comportamiento de DMD [6].

*Tener presente que el paso de regresión normalmente no construye  $A$ , sino que construye  $\tilde{A}$  donde evoluciona la dinámica de un subespacio en un bajo rango a través de POD. Debido que la descomposición propia de  $\tilde{A}$  es usada para aproximarse a la descomposición propia de la matriz  $A$  en una alta dimensión.*

### 2.1.5. Análisis de Koopman

Parte del interés que rodea al método DMD se debe a la fuerte conexión que existe con los sistemas dinámicos de la teoría espectral que tiene el operador Koopman. Este operador fue introducido en el año de 1931 por el matemático francés **Bernard Osgood Koopman** (1900-1981), como un operador lineal de dimensión infinita, que describe como las dimensiones de un sistema dinámico evolucionan a través de la dinámica no lineal. Debido que estas medidas son funciones que forman un espacio de Hilbert  $\mathcal{H}$ , por ende, el operador de Koopman es de dimensión infinita.

En el año 2009, el matemático estadounidense C.W.Rowley demostró que, en determinadas condiciones, el método DMD proporciona una aproximación de dimensión finita en autovalores y autovectores de un operador de Koopman con dimensión infinita (ver [15]). Por lo tanto, en esta sección se presentaran cuestionamientos teóricos, operativos y funcionales del operador de Koopman y su fuerte conexión con el método DMD.

#### 2.1.5.1. Funcionamiento del operador de Koopman

El trabajo original de B.O. Koopman en 1931, consideró los sistemas hamiltonianos y formulo el operador de Koopman en tiempo discreto; sin embargo, este operador se comenzó en tiempo continuo donde su derivada conlleva a un tiempo discreto asociado a esta formulación.

**Operador de Koopman:** Considere un sistema dinámico en tiempo continuo de la forma:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}),$$

donde  $\mathbf{x} \in \mathcal{M}$ , siendo este un estado suavizado  $n - dimensional$  bajo  $\mathcal{M}$ . El operador de Koopman  $\mathcal{K}$  es un operador lineal de dimensión infinita, tal que actúa sobre toda las funciones observables  $g : \mathcal{M} \rightarrow \mathbb{C}$  teniendo:

$$\frac{d}{dt} g(\mathbf{x}) = \mathcal{K} g(\mathbf{x}) = \nabla g(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}).$$

Cuadro 2.1: Definición del operador de koopman[15]

Por definición, el operador de Koopman actúa en espacios de Hilbert  $\mathcal{H}$  con escalares medibles de una función  $g$ , siendo en si, un operador de dimensión infinita. Por lo tanto, la transformación que se obtiene de la representación en el espacio de estados del sistema dinámica a la representación de Koopman, genera un cambio de la dinámica no lineal finita a una dinámica lineal de dimensión infinita [24, 6].

La ventaja de tal compensación es que se puede resolver ecuaciones diferenciales lineales usando una representación espectral de la dinámica de fluidos. Por supuesto, una representación de dimensión infinita puede ser bastante problemática, pero en la práctica, se utiliza una suma de modos lo suficientemente grande y finita, tal que dé solución al espectral del operador Koopman. Ahora, cabe señalar que la definición del cuadro 2.1 se puede representar por una composición de observaciones con la evolución no lineal, de la siguiente manera:

$$\mathcal{K} g = g \circ \mathbf{f}.$$

En este sentido, el operador de Koopman también se puede definir para sistemas dinámicos en tiempo discreto, dado que son más general que los sistemas de tiempo continuo. De hecho, el sistema dinámico  $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x})$  inducirá un sistema dinámico de tiempo discreto por el mapa de flujo  $\mathbf{F}_t : \mathcal{M} \rightarrow \mathcal{M}$  bajo el mapa en su estado  $\mathbf{x}(t_0)$  que avanza al tiempo futuro  $\mathbf{x}(t_0 + t)$ :

$$\mathbf{F}_t(\mathbf{x}(t_0)) = \mathbf{x}(t_0 + t) = \mathbf{x}(t_0) + \int_{t_0}^{t_0+t} \mathbf{f}(\mathbf{x}(\tau)) d\tau,$$

esto impulsa a tener un sistema dinámico discreto, de la forma:

$$\mathbf{x}_{k+1} = \mathbf{F}_t(\mathbf{x}_k),$$

donde  $\mathbf{x}_k = \mathbf{x}(kt)$ , análogamente el operador de Koopman de tiempo discreto esta dado por  $\mathcal{K}_t$  donde  $\mathcal{K}_t g = g \circ \mathbf{F}_t$ . Por lo tanto, el operador de Koopman un sistema dinámico de tiempo continuo en la función  $g$  como:

$$\mathcal{K}_t g(\mathbf{x}_k) = g(\mathbf{F}_t(\mathbf{x}_k)) = g(\mathbf{x}_{k+1}).$$

Ahora, la descomposición espectral del operador de Koopman será fundamental para representar soluciones de un sistema dinámico de interés. Por lo tanto, considerar el problema de sus autovalores.

$$\mathcal{K} \varphi_k = \lambda_k \varphi_k,$$

las funciones  $\varphi_k(\mathbf{x})$  son funciones propias de Koopman, que definen un conjunto de coordenadas de medidas intrínsecas, sobre las cuales es posible avanzar estas medidas con un sistema dinámico lineal. Siendo así, se puede representar la evolución de la dinámica, utilizando una expansión para la solución de una función propia del operador de Koopman, como se ve a continuación:

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_p(\mathbf{x}) \end{bmatrix} = \sum_{k=1}^{\infty} \varphi_k(\mathbf{x}) \mathbf{V}_k,$$

donde  $\mathbf{V}_k$  es el  $k$ -ésimo modo de Koopman asociado a la  $k$ -ésima función propia de Koopman  $\varphi_k(x)$ . En la teoría original, Koopman consideraba los flujos hamiltonianos, es decir, la evolución de un sistema dinámico gobernado por ecuaciones de hamilton, donde describe mediante un mapeo particular un desarrollo que se asemeja al flujo de fluidos como preservadores de medidas, conllevando a que el operador de Koopman sea unitario. Por esta razón, las funciones propias son todas ortonormales y la ecuación anteriormente descrita se puede escribir explícitamente como:

$$\mathbf{g}(\mathbf{x}) = \sum_{k=1}^{\infty} \varphi_k(\mathbf{x}) \begin{bmatrix} \langle \varphi_k, g_1 \rangle \\ \langle \varphi_k, g_2 \rangle \\ \vdots \\ \langle \varphi_k, g_p \rangle \end{bmatrix} = \sum_{k=1}^{\infty} \varphi_k(\mathbf{x}) \mathbf{V}_k.$$

DMD se utiliza para aproximar los autovalores de Koopman  $\lambda_k$  y los modos  $\mathbf{V}_k$ . Donde la idea principal que se puede obtener de esta transformación, es que la dimensión finita de un sistema dinámico no lineal definido por  $\mathbf{f}$  y la dimensión infinita de un sistema dinámico lineal definido por  $\mathcal{K}$  de acuerdo al cuadro 2.1, es que son dos representaciones equivalentes del mismo comportamiento fundamental. Lo crítico de esto, es que para que la teoría de Koopman tenga éxito, es valioso tener la capacidad de enlazar las observaciones  $\mathbf{g}$  y la asociación de la expansión de los modos de Koopman en su evolución original definida por  $\mathbf{f}$ . Bajo condiciones adecuadas, esto se puede lograr de la siguiente manera:

$$\mathcal{K} \mathbf{g}(\mathbf{x}) = \mathcal{K} \sum_{k=1}^{\infty} \varphi_k(\mathbf{x}) \mathbf{V}_k = \sum_{k=1}^{\infty} \mathcal{K} \varphi_k(\mathbf{x}) \mathbf{V}_k = \sum_{k=1}^{\infty} \lambda_k \varphi_k(\mathbf{x}) \mathbf{V}_k.$$

Por lo tanto, las soluciones futuras se pueden calcular mediante una simple multiplicación con el autovalor de Koopman. Cabe resaltar que el operador de Koopman captura todo sobre el sistema dinámico no lineal  $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x})$ , y sus funciones propias definen un cambio no lineal de coordenadas en las que el sistema se vuelve lineal. De hecho, si se restringe las observaciones  $\mathbf{g}$  a un subespacio

invariante abarcado por funciones propias del operador de Koopman, induciría a un operador lineal  $\mathbf{K}$  de dimensión finita y que avanza en estos subespacios de funciones auto-observables, como aparece en la siguiente ilustración:

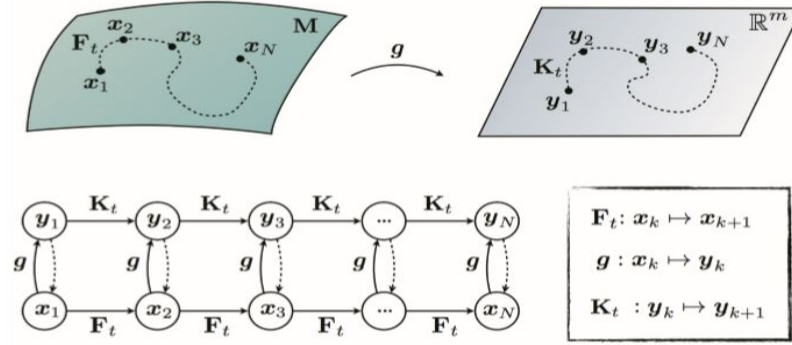


Figura 2.5: Funcionamiento del Operador de Koopman

*Esto es el Operador de Koopman restringido a un subespacio invariante de dimensión finita atravesado por funciones auto-observables [15]. Teniendo presente que la restricción de  $\mathcal{X}$  da como resultado el subespacio  $\mathbf{K}$ , que induce a un sistema lineal de dimensión finita en el espacio invariante.*

### 2.1.5.2. Conexión con el método de Descomposición en Modos Dinámicos

El algoritmo DMD determina los autovalores y modos de Koopman directamente desde datos en condiciones adecuadas. Específicamente, la elección de observaciones desempeñan un papel fundamental para el éxito del método de Koopman. La figura 2.6 muestra un acercamiento al método DMD de forma estándar y la contrasta con la descomposición de los modos de Koopman. Aunque antes de demostrar esta conexión, recordar la definición de DMD.

**DMD:** supóngase un sistema dinámico con dos conjuntos de datos

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_{m-1} \\ | & | & & | \end{bmatrix}, \quad \mathbf{X}' = \begin{bmatrix} | & | & & | \\ \mathbf{x}'_1 & \mathbf{x}'_2 & \dots & \mathbf{x}'_{m-1} \\ | & | & & | \end{bmatrix},$$

con  $\mathbf{x}_k$  como condición inicial y  $\mathbf{x}'_k$  como su salida después de una evolución en el tiempo  $\Delta t$ , considerando  $m - 1$  condiciones iniciales. los modos de DMD son autovectores de la forma

$$\mathbf{A}_x = \mathbf{X}' \mathbf{X}^\dagger,$$

Donde  $\dagger$  denota la Pseudoinversa de Moore-Penrose.

Cuadro 2.2: Definición del método DMD. [15]

Con esta definición, se puede considerar formalmente un conjunto de  $p$  observaciones, de la siguiente manera:

$$g_j : \mathcal{M} \rightarrow \mathbb{C}; \text{ donde } j = 1, 2, 3, \dots, p,$$

tal que  $g = [g_1 \ g_2 \ \dots \ g_p]^T$ , denotando un vector columna de observaciones. Ahora, con esto ya se construyen matrices de datos  $\mathbf{Y}$  y  $\mathbf{Y}'$  considerando un conjunto de condiciones iniciales  $\{x_1 x_2 \dots x_{m-1}\}$ . Las columnas de la matriz  $\mathbf{Y}$  estan dadas por  $\mathbf{y}_k = \mathbf{g}(\mathbf{x}_k)$ , mientras que las columnas de  $\mathbf{Y}'$  se dan evolucionando  $\frac{dx}{dt} = \mathbf{f}(\mathbf{x})$  hacia un tiempo futuro prescrito en un tiempo  $\Delta t$  y visualizándose como un vector salida, denotado por  $\mathbf{y}'_k = \mathbf{g}(\mathbf{x}'_k)$ . El algoritmo DMD resultante de esto es  $\mathbf{A}_Y = \mathbf{Y}'\mathbf{Y}^\dagger$ , generando así el requisito de la aproximación de Koopman. Finalmente, el procedimiento y la comparación del método DMD estándar se presentan en la siguiente ilustración:

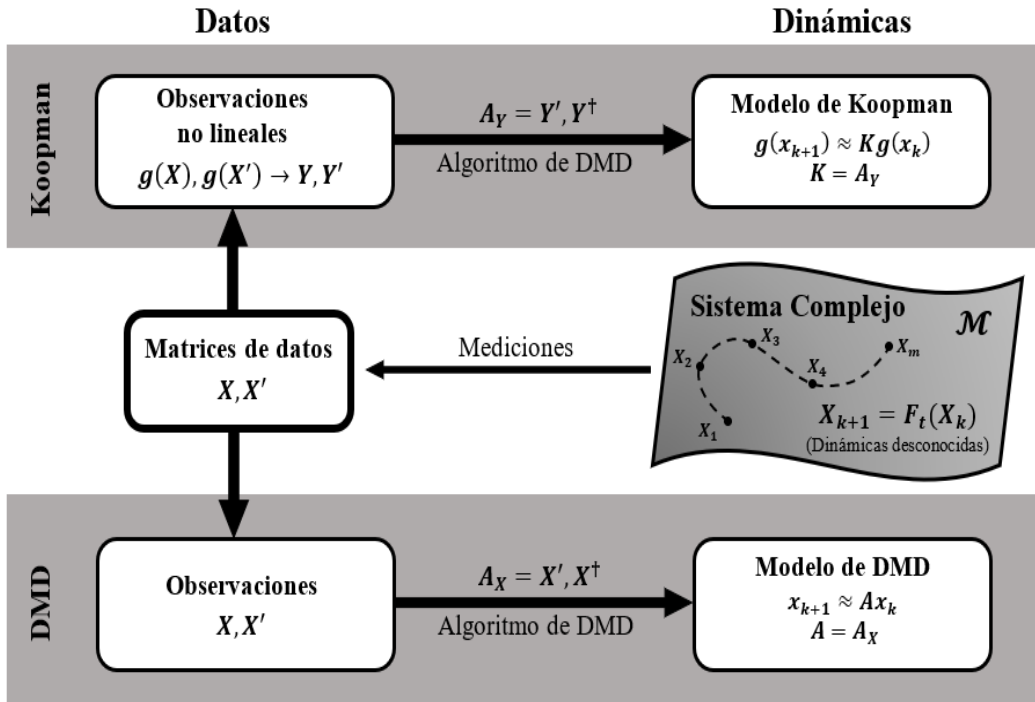


Figura 2.6: Generar sistemas dinámicos de un sistema complejo desconocido en DMD/Koopman.

En el método DMD, se toman medidas de los estados del sistema y se construye un modelo que mapea  $\mathbf{X}$  a  $\mathbf{X}'$ . En el análisis espectral de Koopman se enriquecen estas medidas con observaciones no lineales  $\mathbf{y} = \mathbf{g}(\mathbf{x})$  para proporcionar un mejor mapeo de  $\mathbf{Y}$  a  $\mathbf{Y}'$ , donde se aproxime a los mapas de Koopman de dimensión infinita. Las predicciones que se tienen de las observaciones futuras en este modelo de Koopman, pueden usarse para recuperar el estado futuro  $\mathbf{x}_{m+1}$ , siempre que la función  $\mathbf{g}$  sea inyectiva. En últimas, tanto DMD como Koopman son ecuaciones libres, en el sentido que no se centran en conocer  $\mathbf{F}_t$  [15].

Ahora bien, teniendo esta teoría e ilustración anteriormente descrita, se puede introducir el

siguiente teorema:

**Descomposición de los modos de Koopman y DMD:** Sea  $\varphi_k$  una función propia de  $\mathcal{X}$  con autovalores  $\lambda_k$  y suponiendo que  $\varphi_k \in \text{span}\{g_j\}$  se tiene:

$$\varphi_k(\mathbf{x}) = w_1 g_1(\mathbf{x}) + w_2 g_2(\mathbf{x}) + \dots + w_p g_p(\mathbf{x}) = \mathbf{W} \cdot \mathbf{g}$$

para algún  $\mathbf{W} = [w_1 w_2 \dots w_p]^T \in \mathbb{C}^p$ . Si  $\mathbf{W} \in R(\mathbf{Y})$  donde  $R$  es el rango, entonces  $\mathbf{W}$  sería un autovector izquierdo de  $\mathbf{A}_Y$  con autovalor  $\lambda_k$  donde  $\tilde{\mathbf{W}}^T \mathbf{A}_Y = \lambda_k \tilde{\mathbf{W}}^T$ .

Cuadro 2.3: Definición de la descomposición de los modos de Koopman y el método DMD [15]

Esto muestra, que los autovalores de Koopman sean los autovalores de DMD, proporcionando:

- I) El conjunto de observaciones sea lo suficientemente grande para que  $\varphi_k(\mathbf{x}) \in \text{span}\{g_j\}$ .
- II) Los datos sean lo suficientemente enriquecidos para que  $\mathbf{W} \in R(\bar{X})$

En conclusión, se considera que la elección de observaciones sea fundamental para conectar la teoría de DMD con el análisis espectral de Koopman. Si esto puede hacerse, simplemente se puede tomar *snapshots* de información como sistema dinámico no lineal finito y reparametrizarlo como un sistema lineal de dimensión infinita, siendo esto susceptible de una simple descomposición de función propia. Esta representación diagonaliza la dinámica y muestra que la evolución temporal de cada función propia corresponde a la multiplicación por su valor propio correspondiente.



# Resultados del Proyecto

---

## Desarrollo de la simulación para procesos diagnósticos

En el desarrollo de este proyecto, se emplearán dos lenguajes de programación denominados *MATLAB* y *R*, a fin de ejecutar y visualizar el método de Descomposición en Modos Dinámicos de acuerdo a los casos positivos de COVID-19 presentes en el área metropolitana de Santiago de Cali. Por lo tanto, se estructurarán tres fases de desarrollo: su primera fase será el **tratamiento de datos**, la segunda fase tendrá una **interpretación básica de los datos** y finalmente, se elaborará la **implementación del método DMD** respecto a cada enfermedad mencionada.

### 3.1. Covid19

#### 3.1.1. Tratamiento de datos

Para ejecutar el tratamiento de información del SARS-CoV-2 se empleó un conjunto de datos que fue emitido por la secretaria de salud de Santiago de Cali, para más información ver [4]. Este conjunto contiene el número de contagiados que fueron reportados en cada semana dentro de un periodo temporal de poco más de un año, comenzando desde la semana 10 del año 2020 (mitad de Marzo del 2020) hasta la semana 12 del año 2021 (finales de Marzo del 2021) contabilizando así un total de 56 semanas. Adicionalmente, la información fue recolectada a través de los reportes que generaban las prestadoras de salud de Santiago de Cali, de acuerdo a los lineamientos de detección y manejo de casos de esta enfermedad, ver [2].

Estos lineamientos contribuyeron a que las entidades prestadoras de servicios de salud confirmen situaciones de Covid19 de acuerdo a la información personal del paciente y la fecha en que se realizó los exámenes particulares de dicha enfermedad. De lo anterior, estos pacientes son organizados de acuerdo a los criterios del Ministerio de Salud Colombiana, donde seguidamente se le define una estructura de pasos ordenados para el manejo clínico de su proceso [1]. En síntesis, esta captura de casos positivos se muestra en el siguiente **flujograma**:

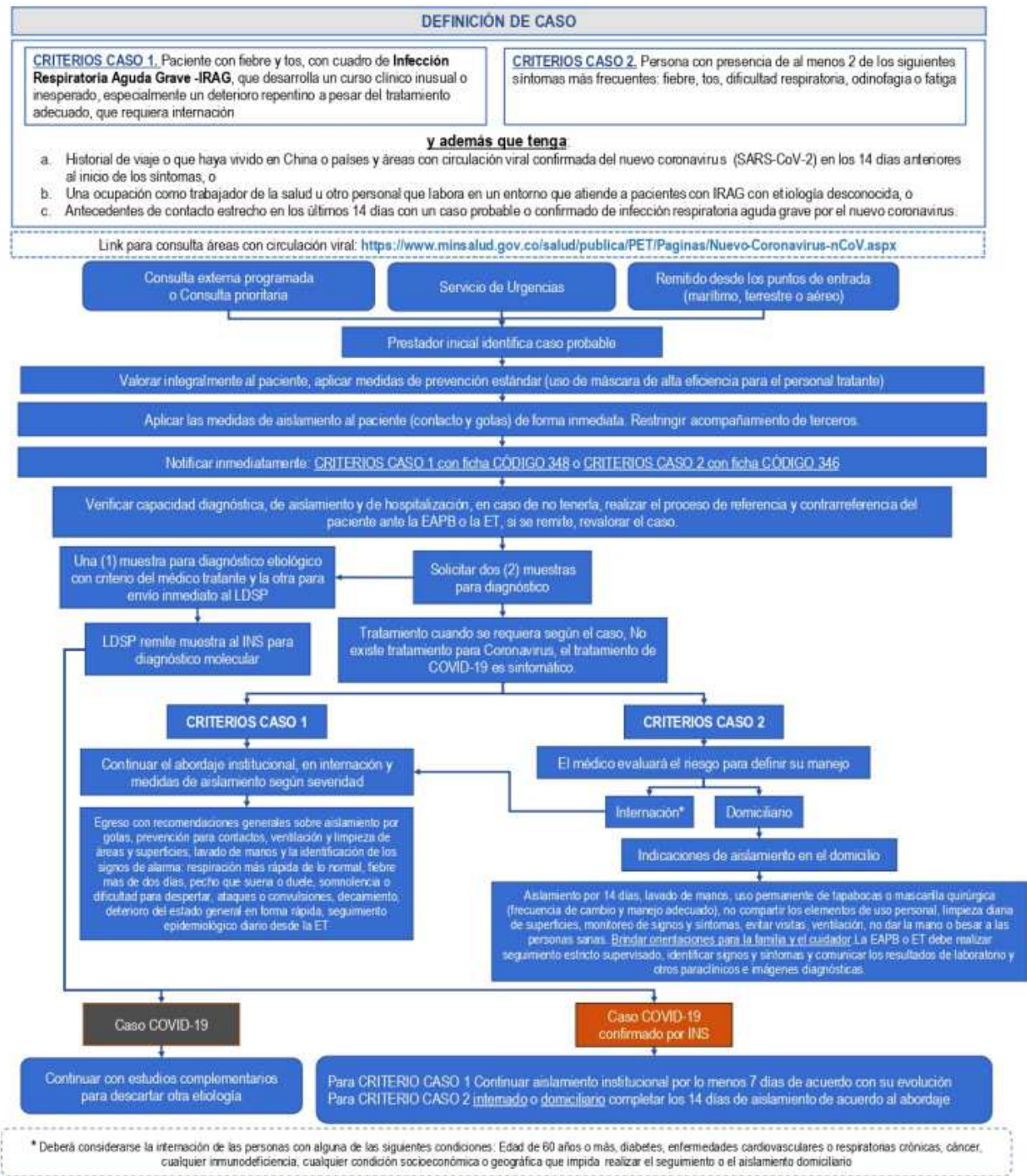


Figura 3.1: Diagrama de flujo para pacientes Covid19 del Ministerio de salud de Colombia [2].

Flujograma elaborado para la detección y manejo de casos por los prestadores de servicios de salud frente a la introducción del SARS-CoV-2 a Colombia [2]; tener presente, que cada paso de estos se adaptó al contexto del área metropolitana de Santiago de Cali.

Por lo tanto, conociendo como se confirma cada paciente positivo de Covid19, la secretaría de salud de Santiago de Cali decidió estructurar su conjunto de datos en relación a las siguientes variables:

- **Año** - Identifica los registros desde su origen en Cali, Colombia en el 2020 hasta 2021.
- **Semana** - Identifica los registros de la semana 10 del 2020 hasta la semana 12 del 2021.
- **Comunas** - Identifica cada una de las 22 comunas que comprende la ciudad.

Dicho conjunto refleja un arreglo matricial de orden  $58 \times 25$ , donde sus primeras dos columnas contienen los años y las semanas de casos registrados por esta enfermedad, luego desde la tercera columna hasta la penúltima columna se tienen las 22 comunas de la ciudad de Santiago de Cali y finalmente, se tiene una columna de totales que presenta la suma o total de contagiados que fueron hallados en cada comuna según su semana registrada. Por otro lado, en cuanto a las filas su primera fila representa las etiquetas de las variables que fueron anteriormente mencionadas y su última fila tiene la suma o total de contagiados que fueron hallados en cada semana según la comuna de la ciudad, debido que entre esta primera y última fila se tienen registradas las 56 semanas de contagio. Ahora bien, es válido aclarar que el cruce entre fila y columna (**Semana** y **Comuna**) representa un valor entero que denota la cantidad total de pacientes confirmados en esa respectiva semana, año y comuna según el procedimiento del ministerio de salud de la figura 3.1.

Luego, por cuestiones prácticas para posteriores pasos en la simulación de este proceso diagnóstico, este arreglo matricial se le realizaron dos modificaciones pertinentes:

1. **Eliminación de valores totales y columna de año:** Debido que el arreglo matricial contiene filas y columnas totales como una de años registrados, se decidió solo representar los valores que tienen pacientes contagiados que existe entre cada semana y comuna, a fin de tener una matriz de orden  $56 \times 22$ .
2. **Transposición del arreglo:** Por situación de orden en esta información, se decidió visualizar los datos de manera  $Comuna \times Semana$ , a fin de tener una matriz con un orden de  $22 \times 56$ .

Finalmente, denominemos a este conjunto matricial de datos como  $X_f$  de orden  $22 \times 56$  donde se le puede crear subconjuntos matriciales, tal que estos subconjuntos pueden ser conformados por cada columna (semana) que contiene el total de contagiados en sus 22 comunas de Santiago de Cali. De este modo, los vectores columna generados anteriormente se les conoce dentro del método DMD como *Snapshots* y están esquematizados de la siguiente manera:

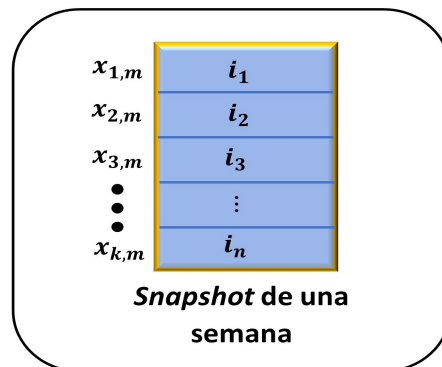


Figura 3.2: Estructura de un *Snapshot* en la ciudad de Santiago de Cali

Esquema matricial de **un vector columna** que representa la cantidad de contagiados por la enfermedad de Covid19 en una semana, donde  $i_n \in \mathbb{N}^0$  son el número de contagiados con  $n = 1, 2, \dots, 22$  ( $n=k$ ). Y donde cada uno de estos contagiados se encuentra asociado a una semana  $m$  en el rango de 1 a 56, según la comuna  $k$  que son 22 para la ciudad de interés en este trabajo, considerando así este vector como un *Snapshot*.

De la ilustración anterior, se tiene que el comportamiento matricial de un *Snapshot* está compuesto por la cantidad de contagiados que existe en una semana particular con sus 22 comunas, por ende, se puede decir que estos *Snapshots* reflejan un fragmento temporal de una semana para el Covid19 en la ciudad de Santiago de Cali, visualizándose esto en un paralelo de 56 vectores (14 meses) tal como se tiene a continuación:

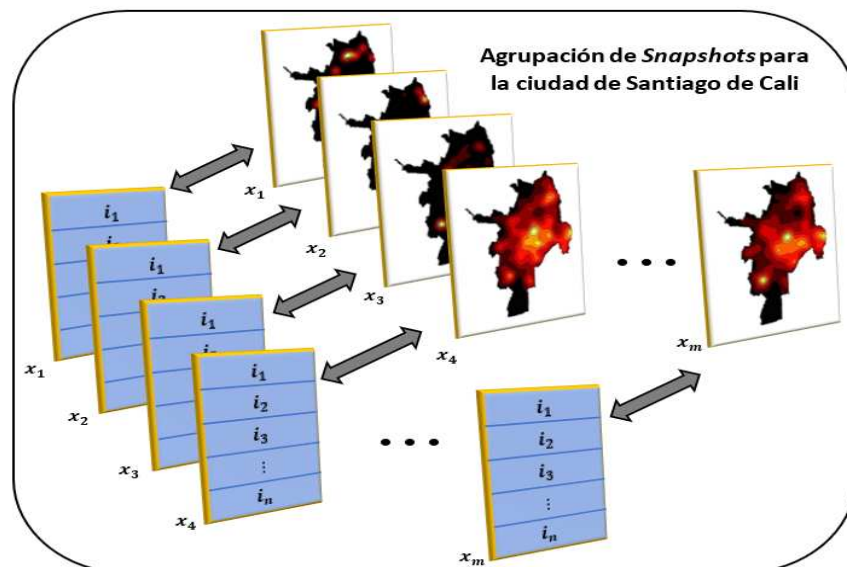


Figura 3.3: Estructura total de *Snapshots* en la ciudad de Santiago de Cali

Esquema con todos los arreglos matriciales y/o geográficos de los Snapshots, donde se representa en cada arreglo la cantidad de contagiados por la enfermedad de Covid19 en una semana, siendo  $i_n \in \mathbb{N}^0$  el número de contagiados con  $n = 1, 2, \dots, 22$  ( $n=k$ ). Y donde cada uno de estos contagiados se encuentra asociado a una semana  $m$  en el rango de 1 a 56, indicando incluso el avance de Snapshots  $x_n \in X_f$ , siendo  $X_f$  el arreglo matricial de orden  $22 \times 56$ .

Por último, teniendo identificado cada Snapshot del Covid19 que fue anteriormente presentado, se realizó un proceso de normalización en este conjunto de datos final con el objetivo de centralizar la información, lo cual contribuye a optimizar el proceso de diagnóstico con esta matriz final. Por tanto, estos snapshots (semanas) normalizados se visualizan como la matriz  $X_f$  dentro de un mapa de calor de la siguiente forma:

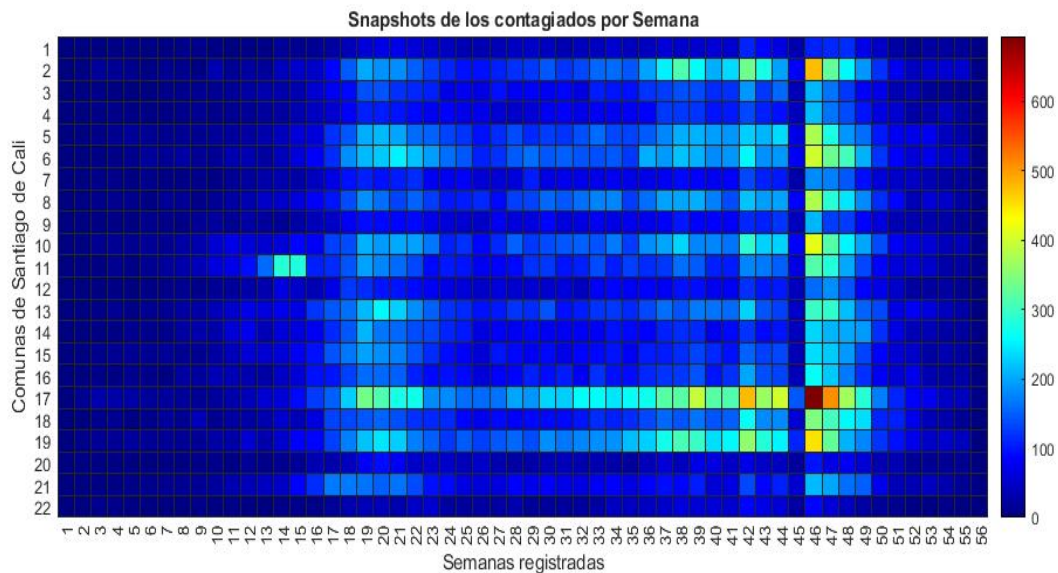


Figura 3.4: Snapshots normalizados de los contagiados por Covid19 a través del tiempo.

Mapa de calor que representa las franjas de contagio por Covid19 en las 56 semanas registradas según su comuna ( $x_n = 1, 2, \dots, 56$ ), donde a simple vista se reconocen dos picos de mediano y alto contagio, siendo estos aproximadamente entre la semana 19 y 20 para su primer pico y el segundo entre la semana 46 y 47.

### 3.1.2. Interpretación básica de los datos

Luego de haber examinado y estructurado de manera general este conjunto de datos de los pacientes contagiados por Covid19, en esta sección se optará por entender cuál es el comportamiento de dicha información recopilada en relación a planteamientos estadísticos básicos. De esta manera, si se ejecutan algunos análisis de medias, desviaciones, mínimos y máximos, se tendría lo siguiente:

Comuna	semanas	mean_Covid	std_Covid	min_Covid	max_Covid
1	56	37.857	30.763	0	114
2	56	123.48	103.38	1	474
3	56	72.643	54.291	0	209
4	56	63.821	47.187	0	216
5	56	112.63	83.244	0	375
6	56	126.02	93.019	0	400
7	56	61.679	44.005	0	179
8	56	111.2	81.487	0	378
9	56	57.357	42.1	0	209
10	56	124.29	88.064	0	416
11	56	107.5	74.205	0	316
12	56	60.107	42.91	0	182
13	56	109.54	74.36	0	297
14	56	83.125	57.798	1	231
15	56	85.964	59.254	2	236
16	56	83.982	58.679	0	263
17	56	189	150.09	2	693
18	56	98.875	78.447	0	344
19	56	141.45	105.64	0	450
20	56	34.054	26.028	0	99
21	56	75.464	54.668	0	213
22	56	26.375	23.489	0	88

Figura 3.5: Estadísticas básicas para el Covid19 en Santiago de Cali.

*Esta tabla presenta estadísticos básicos en relación a los datos capturados de la matriz  $X_f$ , ante esto, tener presente que la población total estudiada es de 111.238 pacientes positivos para la enfermedad de Covid19 en la ciudad de Santiago de Cali. Por lo tanto, dada esta información, a partir de la tabla se percibe que las comunas 17, 19 y 6 son los sectores de la ciudad con mayor promedio de contagio por Covid19, aunque de estos tres seleccionados se resalta la comuna 17, dado que tiene una mayor media de contagios respecto a otras comunas, teniendo en sí una representación del 9.51% de la población total, involucrando a que aproximadamente 189 personas se contagien por Covid19 cada semana, adicionalmente, esta comuna presenta una alta variabilidad de aproximadamente 150 persona por semana, es decir, indica la posibilidad que el número de contagiados fluctúe en aumento o disminución de manera considerable. Por otro lado, tomando nuevamente todas las comunas de la ciudad, la comuna 22 es el sector que presenta menos indicadores estadísticos, es decir, esta comuna tiene una representación del 1.33% de la población total con una variabilidad pequeña, denotando que aproximadamente cerca de 26 personas se contagian por Covid19 cada semana.*

Ahora bien, en relación a la tabla anterior y por cuestiones de percibir esta información de manera gráfica lo que fue anteriormente descrito, véase lo siguiente:

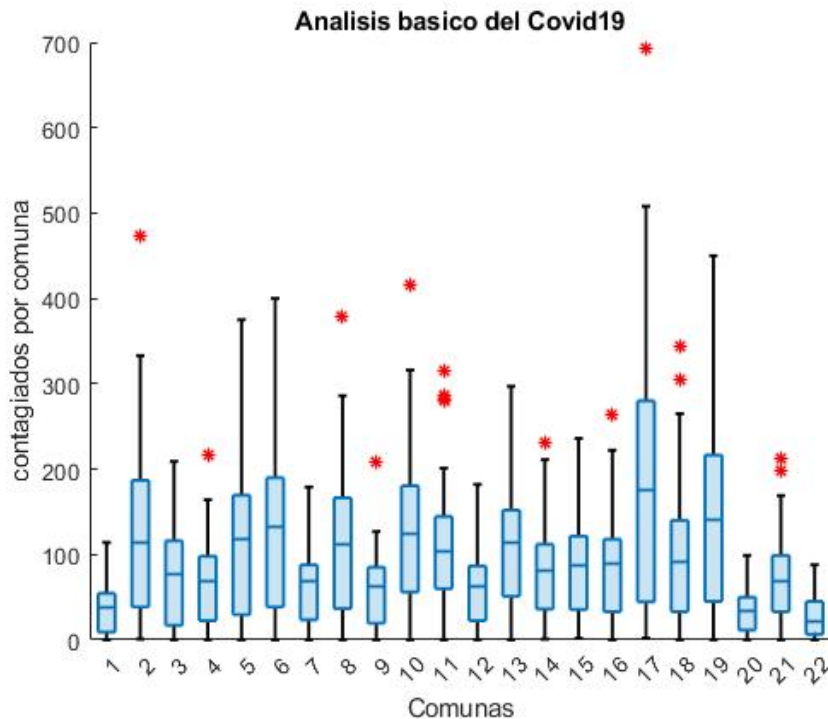


Figura 3.6: Diagrama de Cajas para el Covid19 en Santiago de Cali.

Gráfico que presenta las estadísticas de medias, varianzas y cuartiles de cada comuna según las 56 semanas de registros que fueron emitidas por la secretaria de salud de Santiago de Cali. Análogamente a la figura anterior, aquí se confirma visualmente la percepción de que las comunas 17, 19 y 6 son las tres comunas de mayor promedio de contagios por Covid19, adicional a esto, se puede percibir gráficamente que la comuna 17 tiene una mayor media y desviación estándar respecto a las otras comunas, conllevando a interpretar que esta comuna presente una alta dispersión de pacientes contagiados por Covid19 en las 56 semanas respecto a su media. Por otro lado, entre los cuartiles Q1 y Q3 presenta una concentración del 50% de casos por contagio Covid19, los cuales llevan a tener numerosos pacientes positivos en esta enfermedad que oscilan entre 45 personas y 217 personas cada semana aproximadamente.

### 3.1.3. Implementación del método DMD en el Covid19

El método DMD analiza la matriz  $X_f$  de la figura 3.4 tomando su información como una relación entre pares de datos de manera espacio temporal, es decir, consiste en una medición futura  $\mathbf{x}_{k+1}$  y una medición previa  $\mathbf{x}_k$  donde  $\mathbf{x} \in \mathbb{R}^n$ , organizándose cada uno de estos *Snapshots* o registros de acuerdo a la teoría explicada en el capítulo 2.1.4 con la siguiente expresión:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k, \quad (3.1)$$

de esto, se tiene que los datos captados de la figura 3.3, se expresan de la siguiente manera:

$$\begin{aligned} x_2 &= Ax_1 \\ x_3 &= Ax_2 \\ x_4 &= Ax_3 \\ &\vdots \\ x_{k+1} &= Ax_k \Rightarrow \mathbf{X}_{k+1} = \mathbf{A}\mathbf{X}_k, \end{aligned}$$

donde la matriz  $\mathbf{A} \in \mathbb{R}^{n \times n}$  se puede encontrar como la mejor aproximación en norma 2 aplicando la Pseudo-Inversa de Moore Penrose, tal como se menciona en el cuadro A.1, viéndose así:

$$\mathbf{X}_{k+1} (\mathbf{X}_k)^\dagger = \mathbf{A}.$$

Posteriormente, habiendo analizado el comportamiento de estos estados temporales anteriores, se pueden crear dos arreglos matriciales de las siguiente forma:

$$\mathbf{X} = \begin{bmatrix} | & | & | & \dots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \dots & \mathbf{x}_{m-1} \\ | & | & | & \dots & | \end{bmatrix} \quad \mathbf{X}' = \begin{bmatrix} | & | & | & \dots & | \\ \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \dots & \mathbf{x}_m \\ | & | & | & \dots & | \end{bmatrix},$$

recordar,  $m$  indica el avance de semanas o Snapshots lo cual permite definir una matriz  $\mathbf{X}$  compuesta desde el primer registro hasta el penúltimo registro de  $X_f$ , y de manera análoga, se puede definir la matriz  $\mathbf{X}'$  la cual está compuesta desde el segundo registro hasta su ultimo registro de  $X_f$ . Por lo tanto, el orden de estas matrices previamente explicadas serian  $\mathbf{X}, \mathbf{X}' \in \mathbb{R}^{n \times m-1}$ , adicionalmente, dado que  $\mathbf{A} \in \mathbb{R}^{n \times n}$  es necesario que  $n \gg m$  con el fin de que el  $\text{rang}(\mathbf{A})$  sea completo. De esta manera, las matrices  $\mathbf{X}$  y  $\mathbf{X}'$  se escogen particularmente para la implementación con un orden de  $22 \times 17$  datos relacionados con el Covid19 en la ciudad de Santiago de Cali.

Por lo tanto, utilizando recursivamente la ecuación 3.1 y teniendo presente lo anterior se obtiene el siguiente comportamiento:

$$\mathbf{X}' \approx \mathbf{A}\mathbf{X}. \quad (3.2)$$

Por otra parte, el manejo de información en esta etapa del modelado es crucial, por lo que simplificar los cálculos y el costo computacional que involucra la ejecución de estos arreglos matriciales, se convierte en una necesidad y ganancia para el modelo futuro. En este sentido, aplicando lo visto en la Sub-Sección 2.1.1 con relación a la reducción de elementos, se tendría el siguiente gráfico que presenta la identificación de características por medio SVD a fin de abordar estos sistemas:



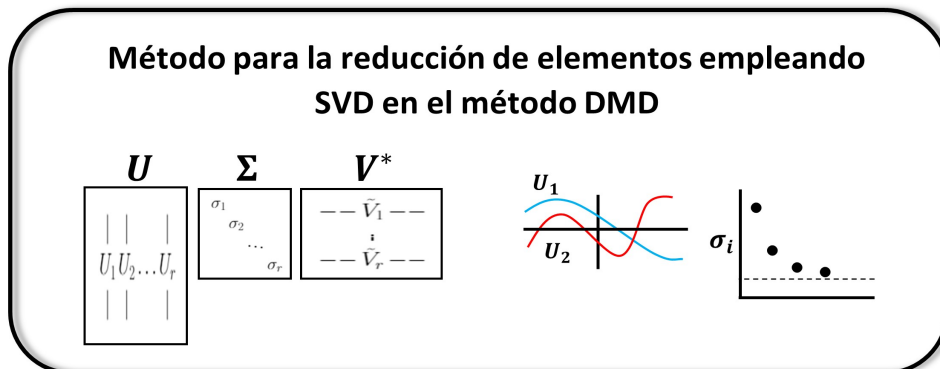


Figura 3.7: Representación del método de reducción utilizando SVD para DMD.

**Izquierda:** Esquema de tres matrices que presenta la descomposición del método SVD.

**Derecha:** Representación de los autovectores de rotación  $U$  y los elementos característicos de la matriz  $\Sigma$ .

De acuerdo con el esquema anterior y retomando con la ecuación 3.2, se aplica una descomposición de la matriz  $X$  siendo esta como  $X = U\Sigma V^*$  permitiendo que esta descomposición ejecutada caracterice los valores singulares ( $\sigma_i \in \Sigma$ ), a fin de ejecutar una reducción de estos elementos desde el método SVD. Adicional a esto, el proceso de SVD también es conocido como una acción análoga del método PCA (2.1.2) como proceso auxiliar a DMD. Por tanto, teniendo la descomposición y sus valores singulares, se contribuye al análisis y simplificación de estos datos por medio de truncamientos, de acuerdo al siguiente gráfico:

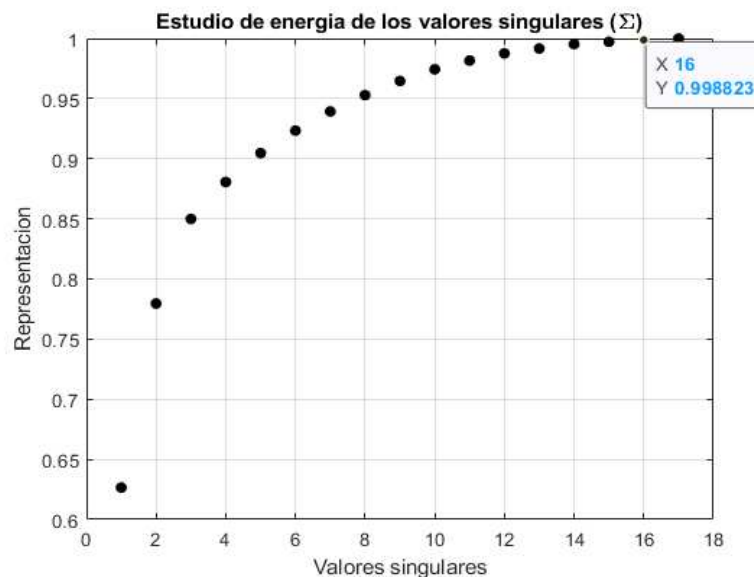


Figura 3.8: Valores singulares por el método de reducción SVD en el Covid19.

Gráfica que presenta todos los valores característicos o singulares  $\sigma_i$  con  $i = 1, 2, \dots, 22$  de manera ascendente, los cuales pertenecen a la matriz  $\Sigma$  de la descomposición SVD. Cada valor  $\sigma_i$  está presentando una relevancia como condición de representación porcentual en una matriz completa  $X$ .

La figura anterior, muestra un estudio de la energía del sistema donde se encuentran los valores singulares o representantes del comportamiento de la matriz total, estos valores, particularmente también son considerados como parte de los componentes principales del método PCA que son explicados bajo el método SVD. Por ende, estos valores enriquece y permite conocer con ciertos grados relevancia que influencia tiene cada comuna respecto a los contagios de la enfermedad Covid19. Siendo así, véase el siguiente análisis e interpretación de algunos componentes principales:

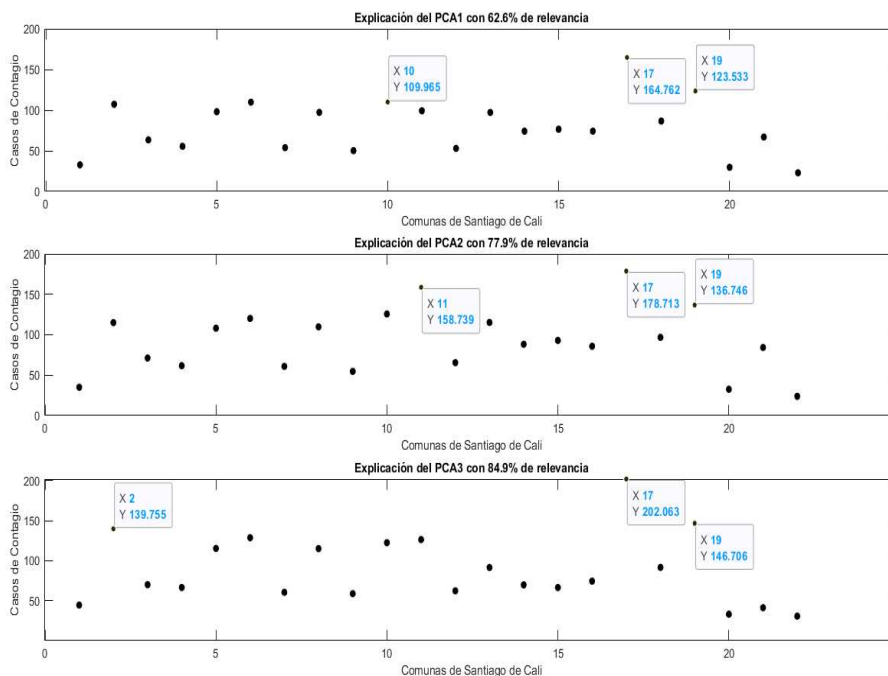


Figura 3.9: Estudio de relevancias que ejecutan los componentes principales.

Gráfica que muestra un estudio de los primeros 3 componentes principales, donde a simple vista se puede notar que después del 60% de relevancia o explicación del sistema  $X$ , se puede notar una alta influencia de contagio por parte de las comunas 17 y 19, teniendo un rango de contagio por paciente en aproximadamente 120 a 205 pacientes. De manera individual cada componente principal tiene lo siguiente:

- PCA1:** explica con una relevancia de 62,6% del total de los datos  $X$ . Donde la comuna de mayor influencia está dada por la comuna 17 teniendo un total aproximado de 165 pacientes cada semana bajo esa relevancia escogida, por otro lado, se identifica en el top 3 de influencia de contagio a la comuna 10 con un aproximado de 110 pacientes cada semana.

- **PCA2:** explica con una relevancia de 77,9% del total de los datos  $X$ . Se evidencia una prevaleciendo como mayor influencia de contagio por parte de la comuna 17 pero ahora con un total aproximado de 202 pacientes por semana dentro de esta relevancia, mientras que la comuna 19 a comparación del anterior componente principal, se baja dentro del top 3 para estar de último lugar y adicional a esto comienza a influir más en ese top 3 la comuna 11 con un aproximado de 159 pacientes semanalmente.
- **PCA3:** explica con una relevancia de 84,9% del total de los datos  $X$ . Y nuevamente se visualiza como la comuna 17 sigue teniendo la mayor influencia de los de contagio semanalmente, aunque por otro lado, para esta relevancia más alta dentro del tercer componente se evidencia que la comuna 2 tiene una influencia sobresaliente de alrededor 140 pacientes semanales.

De este modo, el método SVD permite explorar la influencia de la enfermedad por componentes tal como se vio anteriormente y de manera adicional, también este permite realizar reducciones o truncamientos de la matriz para tener explicaciones con cierto grado de relevancia, en este caso, para todo el modelo de DMD que sigue a continuación se tomara un truncamiento en relación a la gráfica 3.8 según el dato 16, debido que hasta este valor se retrata todo el sistema matricial  $X$  en un 99,88 %.

Luego, habiendo realizado este proceso de reducción, se toma la matriz  $A$  de la ecuación 3.2 dado que es el mejor ajuste para el sistema, entendiéndose esto desde la conexión con el cuadro 2.2 donde  $A = X'X^\dagger$  con el objetivo generar una proyección alterna en los modos POD de  $U$  como  $\tilde{A} = \tilde{U}^* A \tilde{U} = \tilde{U}^* X' \tilde{V} \tilde{\Sigma}^{-1}$ , siendo este arreglo matricial  $\tilde{A}$ , la forma adecuada para obtener la descomposición espectral geométrica vista como  $\tilde{A}W = W\Lambda$  y que se representa desde el siguiente esquema general:

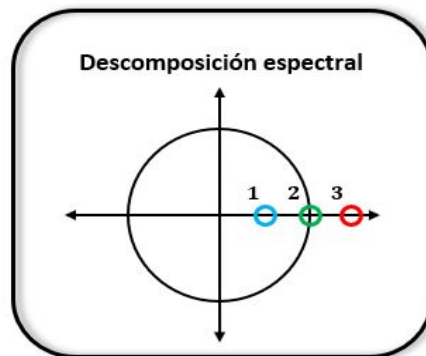


Figura 3.10: Geometría de la descomposición espectral con sus autovalores  $\Lambda$ .

*Ilustración de un círculo unitario en un plano complejo con tres autovalores teóricos  $\Lambda$ , de los cuales explican el comportamiento en magnitud de sus autovalores de contagio por comunas sobre tiempos futuros, dado que este trabajo se realiza en relación a un sistema diferencial lineal. Por lo tanto, En el caso del autovalor 1, se visualiza que la magnitud está por debajo del círculo unitario, lo cual indica que su autovalor tienda al origen y esto defina que la enfermedad para tiempos futuros pueda anularse, seguido de esto se tiene que el autovalor 2 en magnitud se encuentra exactamente sobre la frontera, lo cual indica que el autovalor*

oscile dado que es una multiplicación por la unidad y conlleve a que la enfermedad sea oscilante en tiempos grandes y perdure, por último, se tiene el autovalor 3 el cual en magnitud esta por fuera de la unidad, lo cual indica que el autovalor puede crecer rápidamente dada su solución diferencial lineal conllevando así, a que la enfermedad para este caso crezca y tome un papel problemático en el sector salud. Ahora bien, dada la gráfica e información anterior, ejecutando la descomposición espectral para los datos de Covid19 en la ciudad de Santiago de Cali, se tiene lo siguiente:

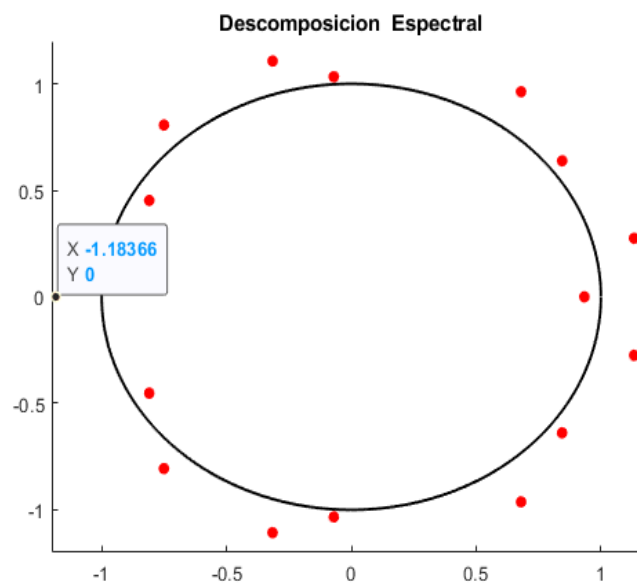


Figura 3.11: Geometría de la descomposición espectral aplicado al Covid19.

Este gráfico presenta la magnitud de sus autovalores donde la mayoría se encuentra por encima del círculo unitario, implicando por teoría que el comportamiento de la enfermedad Covid19 tenderá a crecer en tiempos futuros cercas para cada comuna de la ciudad. Aunque si bien, diagnosticar el comportamiento de estas magnitudes, conlleva a focalizarse en al menos uno de los autovalores de mayor magnitud dentro de este sistema, siendo este el autovalor resaltado que se encuentra un poco aislado de la frontera del círculo unitaria, indicando este una de las mayores magnitudes existente con un comportamiento propio de contagio entre comunas.

Por lo tanto, este comportamiento de autovalores y autovectores de  $\tilde{\mathbf{A}}$  que previamente fueron hallados, puede construir un nuevo arreglo matricial de la forma  $\Phi = \mathbf{X}'\mathbf{V}\Sigma^{-1}\mathbf{W}$ , tal que se denominan como “los modos dinámicos del método DMD” y donde este también puede ser un autovector de la forma  $\Phi = \mathbf{U}*\mathbf{W}$  donde se precisa como la dinámica de avance en relación a lo diagnostico en  $\tilde{\mathbf{A}}$  y  $\mathbf{A}$  dado que es el mejor ajuste dentro de esa ecuación diferencial ordinaria de la enfermedad Covid19.

Siendo así, ahora se analizará el aporte de contagio que ejerce cada comuna de la ciudad de Santiago de Cali, tomando como relación el autovalor de mayor magnitud que fue seleccionado en la gráfica 3.11, por ende, esto se visualiza de la siguiente manera:

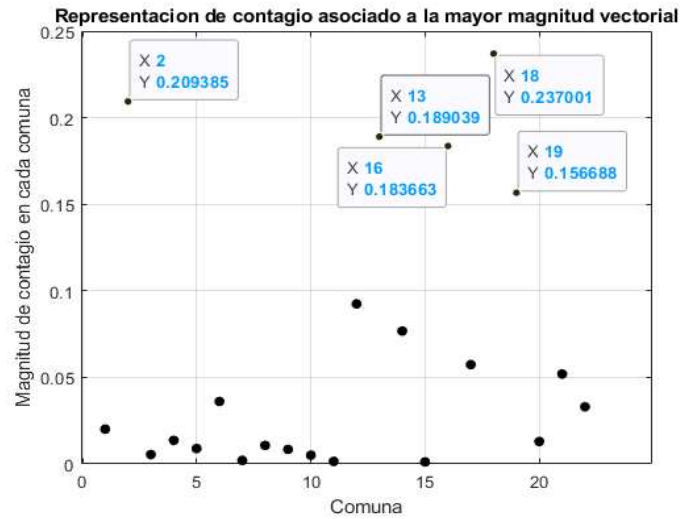


Figura 3.12: Magnitudes de comunas que generan aportes de contagio por Covid19.

*Este gráfico evidencia de manera diagnostica un alto aporte por magnitud de contagio Covid19 en cinco focos o comunas de la ciudad de Santiago de Cali, encabezadas por la comuna 18, seguida de la comuna 2, luego la comuna 13, luego la 16 y por último la comuna 19. Particularmente, es sorprendente notar algunas comparaciones de estos aportes por magnitud de contagio en esas 17 semanas seleccionadas con lo obtenido del análisis estadístico básico (figura 3.6), ya que por ejemplo la comuna 8 se evidencia como una comuna de mayor aporte o influencia de contagia en esas semanas anteriormente dichas, mientras que para las estadísticas básicas parece que esta comuna disminuye un poco su aporte de contagio en el tiempo. Por otro lado y en otro análisis, se evidencia que las comunas 2 y 19 han sido comunas destacadas desde comienzo de la pandemia y al parecer se mantienen dentro de ese rango de aporte a contagio por un buen tiempo, dado que sus valores de medias y variabilidad son altos respecto a las otras comunas.*

Dado lo anterior se visualizará de manera geográfica como aparece a continuación:

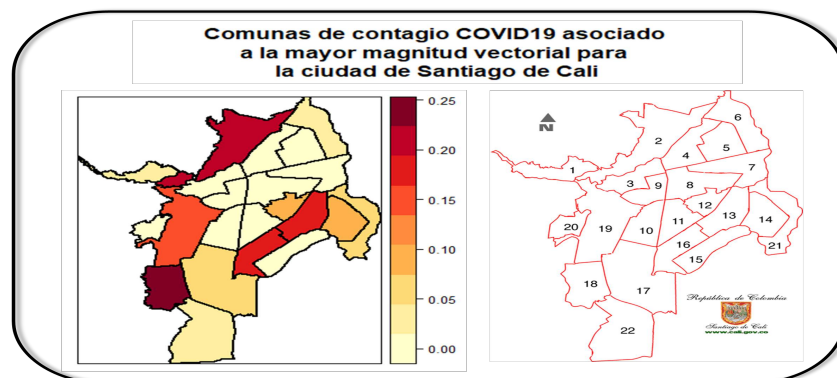


Figura 3.13: Mapa geográfico con las comunas de mayor aporte de contagio por Covid19.

Adicional al gráfico previo (3.12), este mapa presenta de manera diagnóstica los focos de contagio según las magnitudes halladas de cada comuna, por lo que la comuna 18 correspondiente al corredor sur de la ciudad es la comuna de mayor influencia en aporte de contagio Covid19, seguido de la comuna 2 en el corredor norte de la ciudad, mientras que en términos de aporte moderado-alto se tienen a las comunas 13 y 16 ubicadas en el sector centro-occidente de la ciudad. Por otro lado, se evidencia que las comunas en el sector Noroccidente y centro de la ciudad no intervienen mucho en aportes a contagio de esta enfermedad.

Ahora bien, habiendo realizado el comportamiento de los aportes en magnitud de contagio Covid19 por comunas, posiblemente surge la pregunta: “Dentro de esta enfermedad, ¿Cuáles son las comunas que tienen mayor frecuencia?” por tanto, en este nuevo desarrollo se conocerán los aportes de frecuencia que presenta la enfermedad de Covid19 según la ecuación que se muestra a continuación:

$$frecuencia_j = \frac{imag(\log(\lambda_j))}{2\pi\Delta t},$$

la ecuación, describe el número de repeticiones u oscilaciones de acuerdo a las magnitudes de los autovalores en  $\tilde{\mathbf{A}}$  ya que como se había mencionado antes, esta matriz tiene una relación con el método DMD en 2.2, debido que cada uno de estos valores presenta en su sistema una combinación lineal de datos espacio-temporales de la enfermedad Covid19. Por lo tanto, ejecutando este proceso de frecuencia para conocer las oscilaciones en el Covid19 de la ciudad de Santiago de Cali, se tendría lo siguiente:

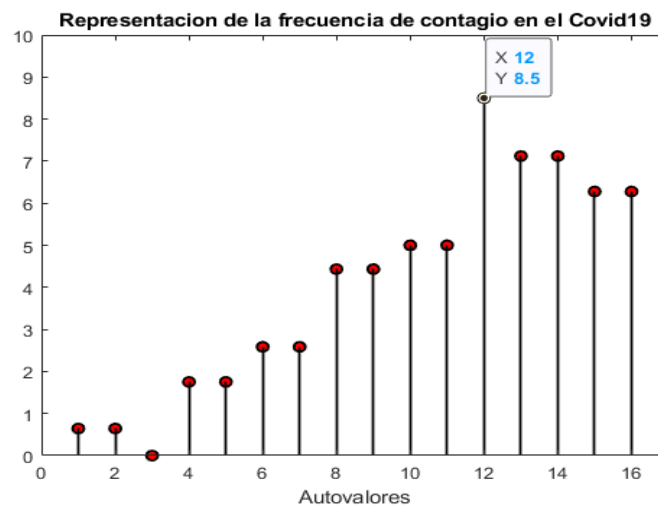


Figura 3.14: Gráfica que describe las magnitudes de frecuencia para el Covid19.

Esta gráfica se encuentra elaborada con un periodo de 17 semanas las cuales oscilan en relación a las magnitudes de los autovalores de la matriz  $\Lambda$ , con el fin de explicar cuáles de estas magnitudes aportan en mayor medida a la frecuencia de contagio Covid19 en la ciudad de Santiago de Cali. Siendo así, dado los autovalores

presentes en esta gráfica se tiene que el autovalor de mayor magnitud se encuentra en su entrada 12, lo cual conlleva a tener un patrón similar en comunas de las gráficas anteriores, pero ahora en relación a que estas comunas de la ciudad efectúan un mayor aporte a la frecuencia de contagio bajo este autovalor de mayor magnitud que fue seleccionado.

En este sentido, dada la figura previa y tomando de ella ese autovalor de mayor frecuencia, se tiene el siguiente gráfico de aportes a mayor frecuencia por comuna:

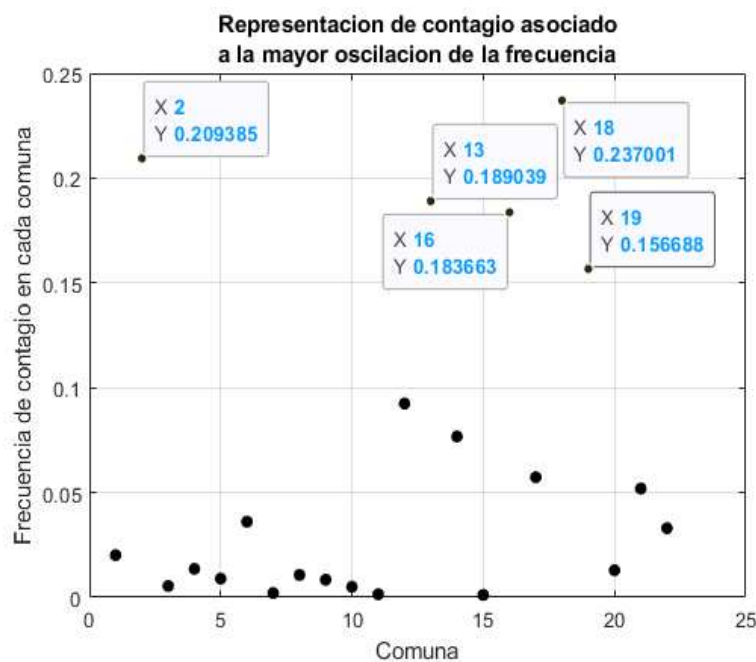


Figura 3.15: Magnitudes de comunas que generan aportes de frecuencia por Covid19.

Se puede notar en este gráfico de manera diagnóstica la existencia de varias comunas que aportan diferentes frecuencias de contagio en la enfermedad del Covid19, aunque de estos resultados se resaltan nuevamente los cinco focos anteriores de la ciudad de Santiago de Cali, encabezando estos aportes en primer lugar por la comuna 18, luego la comuna 2, seguida la comuna 13, después la comuna 16 y por último la comuna 19. En relación a este gráfico y su contexto de frecuencia, se puede inferir que tanto las comunas de aporte a magnitud de contagio son incluso también las frecuentadoras de la enfermedad Covid19, por lo que estas comunas son las precursoras de posibles problemas de contagio en el futuro cercano de acuerdo a sus comunas.

De este modo, habiendo obtenido estos aportes de frecuencia en contagio para la ciudad de Santiago de Cali, se visualiza lo anteriormente descrito pero de manera geográfica como aparece a continuación:

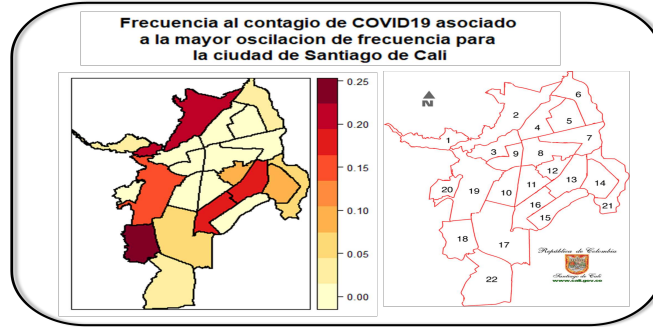


Figura 3.16: Mapa de los mayores aportes a la frecuencia de contagio por comuna en el Covid19.

Nuevamente, este mapa presenta de manera diagnóstica los aportes a frecuencia de contagio asociados a su autovalor de mayor magnitud, donde de manera análoga al caso de aportes en magnitud, se tiene que la comuna 18 correspondiente al corredor sur de la ciudad es la comunas de mayor aporte a frecuencia de contagio y magnitud, seguido de esto se tiene la comuna 2 ubicada en el corredor norte de la ciudad donde también presenta un aporte alto en frecuencia y magnitud. Por otro lado, en cuanto al aporte moderado-alto de frecuencias y magnitud de contagio se tiene las comunas 13 y 16 ubicadas en el sector centro-occidente. Por último, se evidencia nuevamente que en el Noroccidente y centro de la ciudad no hay mucha influencia por aporte a frecuencia de contagio.

Finalmente, habiendo realizado este análisis diagnóstico, donde se clarifico el comportamiento de la enfermedad Covid19 en la ciudad de Santiago de Cali y como esto contribuirá a lo que viene a continuación, se ha decidio ejecutar algunas estimaciones y predicciones de la enfermedad con el modelo de DMD. Pues bien, dado que el sistema completo presenta un orden de  $22 \times 56$  siendo este una Ecuación Diferencial Ordinaria no lineal, donde al dar solución se busca utilizar la definición del operador de Koopman de la figura 2,6 donde se solucionan estas predicciones realizando una conexión por medio de una ecuación lineal de infinita dimensión  $\mathbf{X}' \approx \mathbf{A}\mathbf{X}$  en su mejor ajuste matricial  $\mathbf{A}$  de orden  $n \times n$ .

Ahora bien, al generar un truncamiento particular de esta aproximación tal como se ha venido haciendo en la construcción diagnóstica, se puede tomar un valor  $n$  determinado si o solo si  $n \gg m$  haciendo que  $X' = f(x, \mu, t) \approx (X' \approx AX)$  y permitiendo que la matriz  $\mathbf{A}$  sea diagonalizable en la forma de  $A \sim PDP^{-1}$ . Por lo tanto, a partir de esto se pueden ver las siguientes aproximaciones lineales como solución de lo previamente definido:

$$\begin{aligned}
 X_2 &= AX_1 \Rightarrow X_2 = PDP^{-1}X_1 \\
 X_3 &= AX_2 \Rightarrow X_3 = (PDP^{-1})(PDP^{-1})X_1 = PD^2P^{-1}X_1 \\
 &\vdots \\
 X_{n+1} &= AX_n \Rightarrow X_{n+1} = PD^n \underbrace{P^{-1}X_1}_b = PD^n b,
 \end{aligned}$$



de lo anterior, se tiene que  $\mathbf{P}$  es la matriz de autovectores de  $\mathbf{A}$ ,  $\mathbf{D}$  es la matriz de valores diagonales y  $\mathbf{b} = \mathbf{P}^{-1}\mathbf{X}_1$  como la condición inicial particular que soluciona este sistema lineal. De este modo, partiendo ahora desde la descomposición espectral realizada en el sector diagnóstico y tomando el autovector de  $\mathbf{A}$  conocido como los modos dinámicos del método DMD, se tiene a  $\Phi = \mathbf{U}^*\mathbf{W}$  la cual permite crear una ecuación solución para reconstruir estas proyecciones de la manera  $\mathbf{X}_{n+1} = \Phi\Lambda^n\Phi^{-1}\mathbf{X}_1 \Rightarrow \Phi\Lambda^n\mathbf{b}$ , recordar,  $\mathbf{W}$  y  $\Lambda$  son los autovalores y autovectores que tienen en común las matrices  $\mathbf{A}$  y  $\hat{\mathbf{A}}$ .

Tomando esta anterior ecuación solución del método de Descomposición en Modos Dinámicos, podríamos ver las siguientes proyecciones de contagio en la enfermedad del Covid19:

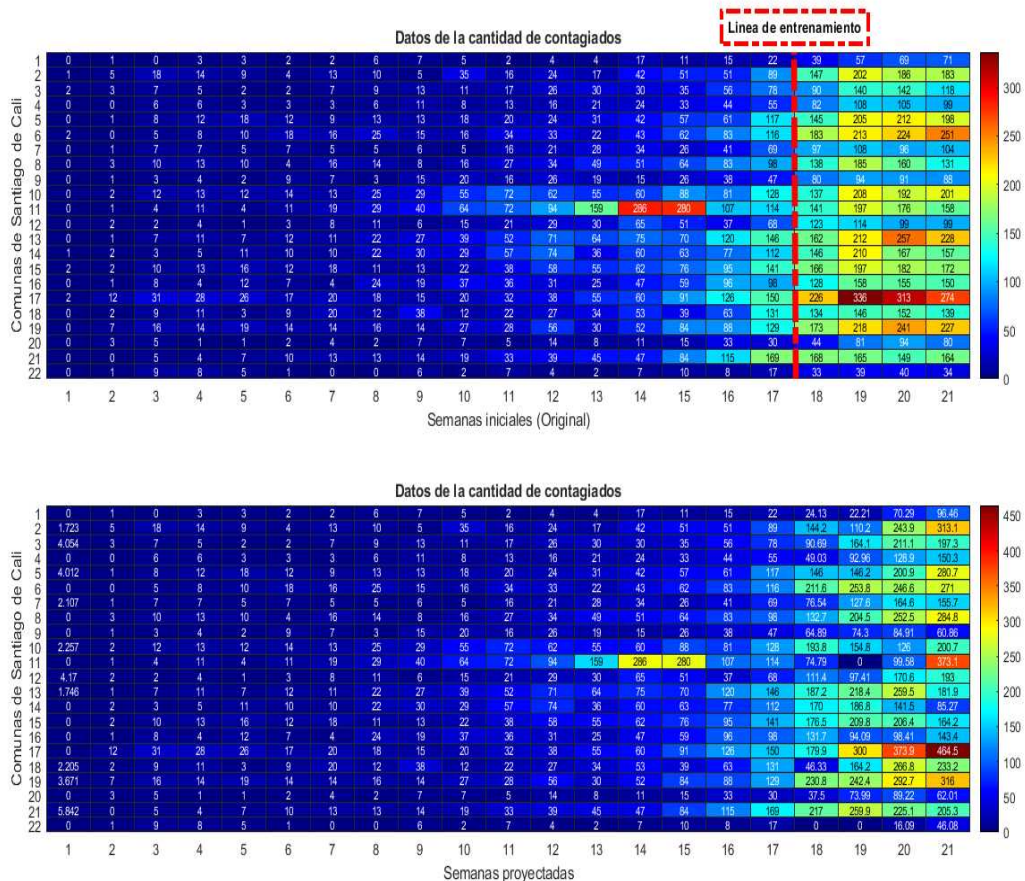


Figura 3.17: Reconstrucción y proyección futura de los modos de DMD.

En esta figura se presentan dos gráficas en mapas de calor, de esto, el primer mapa representa los valores originales que fueron registrados por la Secretaria de Salud de Santiago de Cali desde la semana 1 hasta la

semana 21, ahora bien, debido que en el proceso de diagnóstico se decidió trincar la matriz completa  $X$  de orden  $22 \times 56$  hasta el valor 17, esto conlleva a que esas semanas seleccionadas fueran los datos de entrada para el entrenamiento del modelo DMD y que a partir de estas se crearan los modos del método DMD para obtener así ciertas estimaciones, siendo ellas las resultantes del segundo mapa de calor.

De este modo, en el segundo mapa de calor se perciben predicciones bastante cercanas a los datos originales del primer mapa, siendo estos desde la semana 1 hasta la 17, después, se comienzan ahora a visualizar varios incrementos de contagio en cada comuna y es justo aquí donde la parte diagnóstica de la figura 3.11 cobra bastante sentido, ya que en esa figura de la descomposición espectral se tenían valores por fuera del círculo unitario los cuales indicaban mayores contagios en futuros cercanos y es que en cierta medida, se puede notar que las cifras se apegan un poco a los datos originales del primer mapa, adicional a esto, también se puede notar que las predicciones se encuentran advirtiendo y retratando el comportamiento del primer pico de contagio que se percibe en la figura 3.4 y tan solo con 17 semanas de entrenamiento.

Por último, es de gran valor verificar que tan aproximado se están haciendo las predicciones del método DMD respecto a los datos originales, por lo tanto, se analizó el comportamiento de estos desfases utilizando los procesos de Error Cuadrático Medio (MSE) y la Raíz del Error Cuadrático Medio (RMSE) tal como se ve a continuación:

Semana	MSE	RMSE
1	5.2856	2.299
2	2.5358e-27	5.0357e-14
3	2.393e-25	4.8919e-13
4	8.7216e-25	9.3389e-13
5	1.6118e-24	1.2696e-12
6	3.0265e-24	1.7397e-12
7	3.2615e-24	1.806e-12
8	2.6884e-24	1.6396e-12
9	1.8613e-24	1.3643e-12
10	1.4421e-24	1.2009e-12
11	2.9177e-24	1.7081e-12
12	1.0043e-23	3.169e-12
13	2.3337e-23	4.8308e-12
14	5.4404e-23	7.3759e-12
15	1.2063e-22	1.0983e-11
16	2.1146e-22	1.4542e-11
17	2.6043e-22	1.6138e-11
18	1299.3	36.046
19	3466.6	58.878
20	3107	55.74
21	8115.1	90.084

Figura 3.18: Tabla de errores para el modelo DMD entre lo predicho y lo actual.

*Esta tabla muestra que tanto desfase o error hubo en las 21 semanas que fueron utilizadas en el gráfico anterior de mapas de calor, mostrando así como los valores predichos y los originales en cuanto al MSE presentan desviaciones bastante pequeñas entre las semanas 2 al 17 como también unas varianzas muy mínimas dentro de misma semana según indica la columna de RMSE. Por otro lado, se notan algunas varianzas en los datos después de la semana 17 pero es debido a la información que se tomó dada la condición de que  $n \gg m$ .*

### 3.1.4. Conclusiones

Habiendo ejecutado lo diagnóstico y predictivo del método de Descomposición en Modos Dinámicos para el área metropolitana de Santiago de Cali en la enfermedad del Covid19, se llegó a las siguientes deducciones:

- 1) De acuerdo a la gráfica 3.12 con el mapa 3.13 y su otra gráfica análoga 3.15 con su mapa 3.16, se tiene que las comunas 18, 2, 13, 16 y 19 son las comunas que mayor aportan en magnitud y frecuencia de contagio para la enfermedad de Covid19, esto se debe ya que gran parte de las comunas anteriormente descritas se encuentran dentro de los 10 primeros sectores de mayor densidad poblacional según la alcaldía de Santiago de Cali (ver [10]) y adicional a esto, también algunas de estas comunas comprenden diferentes puntos de interés turísticos u ocio dentro de la ciudad, lo cual genera una afluencia considerable de personas en estas comunas de manera diaria. Por lo tanto, se propone crear una política pública que posicione entidades pertinentes en cada una de estas comunas con el fin de crear campañas de higiene y regularización de salud enfocadas al uso constante de implementos de bioseguridad y espacio de cuidado, permitiendo que estas campañas tengan una población más consciente que conlleve a reducir las cifras de contagio para estos sectores y que adicionalmente, esto genere ganancias para los centros hospitalarios ya que al haber una reducción de contagios habrá una mayor disponibilidad de recursos físicos como camas UCI evitando así posibles colapsos en los servicios de salud por los casos Covid19.
- 2) Habiendo construido el modelo de predicción para la gráfica 3.17 se dedujo que es necesario tener muy en cuenta el número de entradas que se establecen de "input" para entrenamiento el modelo ya que de ellas depende para la matriz  $A$  sea de rango completo y adicional a esto, también se debe tener cuidado con la selección de condición inicial, ya que esta es demasiado sensible a los cambios del sistema y puede presentar ciertas fallas al momento de hacer estimaciones futuras.
- 3) la cuantificación de los datos son necesario verificarlos si presentan una buena calidad de captura, ya que las malas prácticas de esta captación de información pueden afectar crucialmente la ejecución del modelo de predicciones.



# Conclusiones del Proyecto y Trabajos Futuros

---

En conclusión final, el método de Descomposición en Modos Dinámicos (DMD) puede pensarse como una combinación ideal de técnicas de reducción de dimensionalidad espacial. Puesto que las correlaciones de estos modos espaciales están asociadas con una frecuencia temporal, donde su ejecución, demuestra tasas de comportamientos en crecimiento o decadencia. Por lo tanto, el proceso de este método se centra simplemente en recopilar *Snapshots*  $\mathbf{x}_k$  de un sistema dinámico en un número de veces  $t_k$  donde  $k = 1, 2, 3, \dots, m$ . Siendo así, DMD es algorítmicamente una regresión lineal dinámica de datos locales  $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$ , donde  $\mathbf{A}$  se elige para minimizar  $\|\mathbf{x}_{k+1} - \mathbf{A}\mathbf{x}_k\|_2$  bajo  $k = 1, 2, \dots, m$  *Snapshots*.

Todo esto trayendo ventajas como la sencillez de ejecución de este método y las pocas suposiciones que se tienen que elaborar sobre su sistema, ya que el costo de este algoritmo es una Descomposición en Valores Singulares (SVD) de la matriz de *Snapshots*  $X$  que fue construida a partir de la matriz de datos  $X_f$ . Por tanto, se puede decir que DMD tiene varios usos e interpretaciones, siendo dos de ellos ya elaborado en este proyecto de los tres procesos que en síntesis se pueden ejecutar:

- I. **Diagnostico:** Este proceso particular fue el que se ejecutó en este proyecto de grado asociado a la enfermedad del Covid19. Puesto que este algoritmo extrae características espacio-temporales de sistemas de alta dimensión, donde históricamente desde sus inicios para el método DMD, se utilizó como una herramienta diagnostica, lo que permite tener resultados físicamente interpretables en términos de estructuras espaciales y sus respuestas temporales asociadas. Curiosamente, esta es quizás la función principal de DMD en muchas áreas de aplicación, puesto que la naturaleza del método permite el descubrimiento de sistemas complejos y análisis de flujos en relación a datos estructurados.
- II. **Estimación y predicción de datos futuros:** Este es un uso más sofisticado y desafiante que presenta el algoritmo DMD debido que se asocia con el uso de estructuras espacio-temporales que son dominantes en los datos, teniendo el objetivo de construir modelos dinámicos desde la información observada. En sí, esta es una labor difícil, especialmente porque DMD está limitada al construir el sistema dinámico lineal de mejor ajuste (mínimos cuadrados) al sistema dinámico no lineal que genera los datos. Así, a diferencia del objetivo diagnóstico, el objetivo de esta temática es anticiparse a un estado futuro del sistema del que no se han realizado mediciones. De este modo, lo que confunde la naturaleza regresiva del método DMD es el hecho de que la dinámica de los datos puede exhibir una dinámica multiescala tanto en el tiempo como en el espacio. Aunque, existen algunos modelos estratégicos del Big data como *intelligent sampling* y *updating the regression*, que permiten a DMD ser eficaz para generar este modelo de dinámica lineal útil, ya que esta ejecución de modelos han sido casos de éxito en muchas áreas de aplicación para la predicción de estados futuros.
- III. **Control:** Habilitar estrategias de control viables y sólidas directamente desde el muestreo de datos es el objetivo final y más desafiante del algoritmo DMD. Dado que se usa un modelo dinámico lineal para

predecir el futuro de un sistema dinámico no lineal, es razonable esperar que solo exista un tiempo limitado, donde una ventana de corta duración en el futuro pueda modelar los sistemas de diagnóstico y predicción al tiempo. Aunque la esperanza de esta ventana de predicción es que tenga una duración temporal lo suficientemente larga como para permitir una decisión de control, que sea capaz de influir en el estado futuro de un sistema. El algoritmo DMD en este caso permite un enfoque completamente basado en datos para controlar la teoría, proporcionando así un marco matemático convincente que de paso a controlar sistemas dinámicos complejos cuyas ecuaciones que gobiernan estos sistemas aún no se conocen o sencillamente son difíciles de modelar computacionalmente.

Finalmente, al utilizar el método DMD, siempre se debe tener en cuenta el uso previsto y el resultado esperado. Además, aunque parece bastante obvio, el éxito del algoritmo DMD también dependerá en gran medida de cuál de las tareas anteriores se esté intentando en conseguir. En el caso de este proyecto de grado, como se mencionó previamente, se optó por las temáticas diagnósticas y predictivas de las cuales se consiguieron panoramas con técnicas diferentes y novedosas a las que ya se ha venido empleando en relación a los estudios del Covid19, donde cabe aclarar que este desarrollo conlleva a diagnosticar de manera epidemiológica cuales son esos comportamientos de aporte a contagio y frecuencias que está generando la enfermedad, para así luego predecir posibles valores que pongan en riesgo a la población de Santiago de Cali.

Como sugerencia para posibles trabajos futuros, se recomienda ejecutar el mismo método diagnóstico y de predicción que se elaboraron en este proyecto, pero tomando datos a nivel nacional y no municipal como se hizo en este trabajo, esto debido a que la ciudad de Santiago de Cali bajo nuestra experiencia solo tiene 22 comunas, lo cual hace que se limite un poco esa indagación de datos debido a la condición  $n \gg m$ , por ende, trabajar a nivel nacional es más aconsejable ya que permitiría refinar la técnica de modelado dado que se pueden entrenar con cantidades grandes de datos en este modelo de DMD. Adicionalmente, también se sugiere de manera alcanzable continuar con la investigación de este proyecto a nivel municipal en cuanto a la temática predictiva para la condición inicial, ya que la sensibilidad de esta condición es tan alta, que cualquier perturbación por decimal hace que el modelo no pueda tener un funcionamiento correcto y lo cual no permitiría tener modelos tan acertados. Motivarse con estas sugerencias, hacen parte del proceso investigativo realizado y el empeño personal que se tuvo con este proyecto, puesto que esta puede ser la entrada a un nuevo campo de análisis epidemiológico y a una forma de monitorear y predecir no solo el Covid19 sino también otros tipos de enfermedades que se puedan presentar.

# Anexo: Tópicos teóricos

---

## A.1. Descomposición en Valores Singulares (SVD).

### Propiedades de la Descomposición en Valores Singulares.

Este método presenta cuatro propiedades fundamentales, en primera instancia asumir que el rango de una matriz  $A$  de orden  $m \times n$  es linealmente independiente por filas (o equivalentemente por columnas), siendo así, véase lo siguiente:

**Propiedad 1:** El rango de una matriz  $A = U\Sigma V^*$  es igual a los números que no son cero en las entradas de  $\Sigma$ .

**Demostración:** [35] Para comprobar esta propiedad, por hipótesis  $A = U\Sigma V^*$ , donde las definiciones anteriormente descritas, muestran que tanto  $U$  y  $V^*$  son matrices invertibles, es decir,  $UV^* = I$ . Concluyendo así, que  $A = \Sigma$ , por lo tanto, el  $Rang(A) = Rang(\Sigma)$ . ■

**Propiedad 2:** Si  $A$  es una matriz de tamaño  $n \times m$ , entonces  $|det(A)|$  de rango completo es igual a  $\sigma_1, \sigma_2, \dots, \sigma_n$ .

**Demostración:** [35] Por hipótesis  $A$  es un arreglo matricial de orden  $n \times m$ , por consiguiente:

$$det(A) = det(U\Sigma V^*),$$

siendo así, el  $det(U\Sigma V^*)$  se visualiza de la siguiente manera:

$$det(U\Sigma V^*) = det(U) det(\Sigma) det(V^*).$$

Aunque, por definición las matrices  $U$  y  $V$  son ortogonales, implicando que sus determinantes sean equivalentes a 1, Por lo tanto:

$$det(A) = det(\Sigma) \Rightarrow |det(A)| = \sigma_1, \sigma_2, \dots, \sigma_n.$$

Concluyendo así, que los determinantes de los productos son iguales a el producto de los integrales. ■

**Propiedad 3:** Si  $A$  es una matriz invertible de tamaño  $m \times m$ , entonces  $A^{-1} = V\Sigma^{-1}U^*$

**Demostración:** [35] Considere como punto de partida la **Propiedad 1**, teniendo presente que por hipótesis  $\Sigma$  es una matriz invertible, indicando lo siguiente:

$$\begin{aligned} \text{Rang}((A)^{-1}) &= \text{Rang}((U\Sigma V^*)^{-1}), \\ \text{Rang}((A)^{-1}) &= \text{Rang}((U)^{-1}(\Sigma)^{-1}(V^*)^{-1}), \end{aligned}$$

por definición  $U$  y  $V$  son ortogonales, empleando la propiedad de que  $U^* = U^{-1}$ , se podría decir que  $(U^*)^{-1} = (U^{-1})^*$ ; siendo esto una analogía para la matriz  $V$ , para obtener:

$$\begin{aligned} \text{Rang}((A)^{-1}) &= \text{Rang}((U)^*(\Sigma)^{-1}(V^*)^{-1}), \\ \text{Rang}((A)^{-1}) &= \text{Rang}((U)^*(\Sigma)^{-1}(V^{-1})^*), \end{aligned}$$

en conclusión, como  $V$  es ortogonal, es válido afirmar que la propiedad  $(V^{-1})^* = v$ , logra el resultado  $A^{-1} = V\Sigma^{-1}U^*$ . ■

**Propiedad 4:** Sea la matriz  $A$  de tamaño  $m \times n$ , donde  $A$  puede ser escrita como la suma del rango de matrices:

$$A = \sum_r^{i=1} \Sigma_i U_i V_i^*,$$

donde  $r$  es el rango de  $A$  y tanto  $U_i$  como  $V_i$  se encuentran en las  $i$ -ésimas columnas de  $U$  y  $V$ , siendo claramente de manera respectiva.

**Demostración:** [35]

$$\begin{aligned} A &= U\Sigma V^* = U \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} V^*, \\ &= U \left( \begin{bmatrix} \sigma_1 \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} + \begin{bmatrix} \sigma_2 \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} + \dots + \begin{bmatrix} \sigma_r \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} \right) V^*, \\ &= \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \dots + \sigma_r u_r v_r^*. \end{aligned}$$
■

**Nota:** Para la metodología SVD y el álgebra lineal, tener presente el concepto de inversa, puesto que para una matriz  $A$  de tamaño  $n \times n$  su rango coincide con el número de filas(o columnas), alcanzando propiedades como  $AA^{-1} = A^{-1}A = I$ , donde  $AA^{-1} = A^{-1}A$  deduciendo así que  $A^{-1}$  es la inversa “bilateral” de  $A$ , ver



Si  $A \in \mathbb{R}^{m \times n}$ , entonces existe una única matriz  $A^\dagger \in \mathbb{R}^{n \times m}$  donde satisface cuatro condiciones:

1.  $AA^\dagger A = A$ ,
2.  $A^\dagger AA^\dagger = A^\dagger$ ,
3.  $A^\dagger A = (A^\dagger A)^*$ ,
4.  $AA^\dagger = (AA^\dagger)^*$ ,

estas condiciones se denominan como **condiciones de Penrose**.

Cuadro A.1: Definición para la Pseudo-Inversa de Moore Penrose [26]

[26]. Si se llegara a obtener una matriz  $A$  que no presenta inversa, se puede realizar una generalización de la matriz, denominada **Pseudo-inversa de Moore Penrose**, definida de la siguiente manera:

En síntesis, la pseudo-inversa se ve reflejada en la siguiente expresión:

$$A^\dagger = (A^* A)^{-1} A^*.$$

### Proceso para calcular la Descomposición de Valores Singulares.

Conociendo las propiedades del método SVD, verificar teóricamente si una matriz  $A$  con tamaño  $m \times n$  siendo  $m \geq n$ , su factorización se presenta como:

$$A = U \Sigma V^*,$$

a modo descriptivo, sea  $U$  una matriz que presenta columnas ortogonales de orden  $m \times n$  luego, sea  $V$  una matriz ortogonal de orden  $n \times n$ , y por último, sea  $\Sigma$  una matriz diagonal de orden  $n \times n$ . Tal que, sus valores singulares  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ , con  $r \leq n$ .

Dado el rango de la matriz  $A$  donde  $r < n$ , entonces,  $\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_n = 0$ , sí la matriz  $\Sigma$  es de orden  $n \times n$  tendría forma de bloques, siendo el primer bloque en su diagonal principal denominado  $D$  de orden  $r \times r$  (considerando que probablemente  $r = n$ ). Los siguientes bloques son matrices nulas(ceros) denotados como  $\mathfrak{D}$ , tal que en el sector derecho a  $D$  presenta un orden de  $r \times n - r$ , luego, en la parte inferior  $D$  tiene un bloque de orden  $n - r \times r$  por último, se tiene un bloque diagonal de orden  $n - r \times n - r$ , de la siguiente forma:

$$\Sigma = \begin{pmatrix} D & \mathfrak{D} \\ \mathfrak{D} & \mathfrak{D} \end{pmatrix},$$

los bloques matriciales anteriores, se pueden describir como:

$$D = \begin{pmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{pmatrix}.$$

Ahora bien, es relevante hallar los factores de  $U$  y  $V^*$ , para hallar la matriz ortogonal  $V$  siendo ésta generada por una base ortogonal  $(V_1, V_2, \dots, V_n)$  de vectores en  $\mathbb{R}^n$ , tal que esté compuesta por autovectores de la matriz  $A^*A$  de orden  $n \times n$ , siendo:

$$V = [V_1, V_2, \dots, V_n],$$

Tal como se visualiza,  $V$  es una matriz ortogonal de orden  $n \times n$ .

Por otro lado, para determinar la matriz  $U$  de orden  $m \times n$ , observar que los vectores en  $\mathbb{R}^m$  se comportan como:

$$AV_1, AV_2, \dots, AV_n,$$

lo anterior es un conjunto de  $n$  vectores, que dependen de  $r$ , para ser  $r$  vectores ortogonales, obtenidos a partir de los autovectores  $A^*A$ ,  $v_i$  y  $v_j$  con  $i \neq j$ , al conocer que son ortogonales se cumple que:

$$v_j^* A^* A v_i = v_j^* \cdot \lambda_i v_i = 0,$$

de ahí  $Av_i$  es ortogonal a  $Av_j$ , en el caso de que sean no nulos  $v_i$  es ortogonal a  $v_j$ . Ahora bien, los valores singulares de  $\sigma_i = \|Av_i\|$ , con sus primeros  $r$  no nulos (con la posibilidad que  $r = n$ ). Por lo tanto, se normalizará los  $Av_i$  con  $i = 1, \dots, r$ , considerando:

$$u_i = \frac{Av_i}{\sigma_i},$$

para  $i = 1, \dots, r$ , corresponde un  $\sigma_i$  no nulo, garantizando  $[u_1, \dots, u_r]$  siendo ortonormales en  $\mathbb{R}^m$  bajo la premisa que  $r < n$  donde  $r + 1, \dots, n$ , dando como resultado:

$$v_j^* A^* A v_i = v_j^* \cdot \lambda_i v_i = 0,$$

si  $\lambda_i = 0$ , entonces  $\|Av_i\|^2 = 0$ , cumpliendo:  $Av_i = 0$ , por ende, son nulos para  $i = r + 1, 2, \dots, n$ . Posteriormente a esto, se abordaría el caso  $r < n$ .

En este orden de ideas, es importante extender el conjunto  $U = [U_1, U_2, \dots, U_r, U_{r+1} \dots U_n]$  para llegar a obtener  $n$  vectores ortonormales en  $\mathbb{R}^m$ .

En este punto, la factorización será compleja, ya que se desea obtener:  $A = U\Sigma V^*$ . como la matriz  $U$  está conformada por columnas ortonormales que son  $U = [U_1, \dots, U_r, |U_{r+1}, \dots U_n]$ . Si

$m > n$  la matriz  $U$  no es una base ortogonal de  $\mathbb{R}^{m \times m}$ .

Por tanto, verificar que las matrices  $V$ ,  $U$  y  $\Sigma$ , son de la siguiente manera:

$$A = U\Sigma V^*,$$

dado que  $V^* = V^{-1}$  por ser matrices ortogonales, se valida lo formulado, si se multiplica por  $V$  teniendo como posibilidad que  $r < n$ , se obtendría:

$$AV = U\Sigma,$$

desde este punto, se conoce que  $Av_i = \sigma_i u_i$  para  $i = 1, 2, \dots, r$ ; que a su vez:

$$\|Av_i\| = \sigma_i \geq 0; \quad i = r + 1, \dots, n.$$

Por lo tanto, si  $r < n$ , se tiene:

$$AV_i = 0; \quad i = r + 1, \dots, n,$$

generando lo siguiente:

$$AV = A[v_1, \dots, v_n] = [Av_1, \dots, Av_n].$$

En conclusión

$$AV = [Av_1, \dots, Av_r, 0 \dots 0] = [\sigma_1 u_1 \dots \sigma_r u_r, 0 \dots 0],$$

$$AV = (u_1 u_2 \dots u_n) \begin{pmatrix} \sigma_1 & \dots & 0 & \\ \vdots & \ddots & \vdots & O \\ 0 & \dots & \sigma_r & \\ & & O & O \end{pmatrix} = U\Sigma.$$

Finalmente, esta verificación teórica para la Descomposición de Valores Singulares es  $A = U\Sigma V^*$ , ver [33].

**Nota:** El método SVD Consiste en encontrar autovalores y autovectores de  $AA^*$  y  $A^*A$ . Ya que los autovectores  $A^*A$  construyen las columnas de  $V$  y los autovectores  $AA^*$  forman las columnas de  $U$ . Por último, los valores singulares  $\Sigma$  son raíces cuadradas de los valores propios  $AA^*$  y  $A^*A$ ; ubicados en la diagonal principal de la matriz  $\Sigma$  de manera descendente, ver [20].

Es conveniente observar un ejemplo a modo de afianzar la aplicación del método SVD:

Sea el vector de orden  $4 \times 2$ , el cual es:  $A = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ .

**Solución:** primero, para hallar la matriz  $U$  es necesario hacer  $AA^*$ , teniendo lo siguiente:

$$AA^* = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 4 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 20 & 14 & 0 & 0 \\ 14 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = W,$$

con esta matriz de tamaño  $4 \times 4$ , se procede a calcular los autovalores, partiendo de  $Wx = \lambda x \rightarrow (W - I\lambda)x = 0$  y observándose como:

$$\begin{bmatrix} 20 - \lambda & 14 & 0 & 0 \\ 14 & 10 - \lambda & 0 & 0 \\ 0 & 0 & -\lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{bmatrix} x = (W - I\lambda)x = 0,$$

por consiguiente, se obtendrían los siguientes autovalores:

1.  $\lambda_1 = 0$ ,
2.  $\lambda_2 = 0$ ,
3.  $\lambda_3 = 0,1339$ ,
4.  $\lambda_4 = 29,883$ .

Luego de hallar estos autovalores, se encontrarán los autovectores, siendo el primer caso con el valor  $\lambda = 0,1339$ , obteniendo:

$$\begin{bmatrix} 19,883 & 14 & 0 & 0 \\ 14 & 9,883 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

posteriormente, si se toma el valor de  $\lambda = 29,883$  y se realiza el proceso anterior, se obtiene:

$$\begin{bmatrix} -9,883 & 14 & 0 & 0 \\ 14 & -19,883 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

finalmente, realizando una combinación de las dos matrices halladas, se obtendrá la matriz  $U$ , de la siguiente manera:

$$U = \begin{bmatrix} 0,82 & -0,58 & 0 & 0 \\ 0,58 & 0,82 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Si se replica este proceso anterior se puede hallar la matriz  $V$  aplicando  $A^*A$ , obteniendo lo siguiente:

$$A^*A = \begin{bmatrix} 2 & 4 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

luego de ejecutar la expansión del producto  $A^*A$ , para hallar sus autovalores, se emplearán cada uno estos a fin de conseguir sus autovectores donde al realizar una combinación entre ellos, se definiría como:

$$V = \begin{bmatrix} 0,40 & -0,91 \\ 0,91 & 0,40 \end{bmatrix}.$$

Por último, se tiene en cuenta realizar las raíces cuadradas a los autovalores  $AA^*$  y  $A^*A$ , para obtener la matriz  $\Sigma$ , la cual se define como:

$$\Sigma = \begin{bmatrix} 5,47 & 0 \\ 0 & 0,37 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

### Método económico para la Descomposición de Valores Singulares

Dada una matriz  $A \in M_{m \times n}(\mathbb{C})$  con  $m > n$ , se formaliza lo siguiente:

$$A = U\Sigma V^*,$$

si  $n \geq m$ , conlleva que la matriz  $\Sigma$  presente más elementos ceros en su diagonal principal y por ende, se denota como  $\Sigma = \begin{bmatrix} \hat{\Sigma} \\ 0 \end{bmatrix}$ .

Lo anterior, posibilita representar la matriz  $A$ , usando método alterno denominado: Descomposición en Valores Singulares Económico (*SVD económico*), visualizándose de la siguiente forma [6]:

$$A = U\Sigma V^* = \begin{bmatrix} \hat{U} & \hat{U}^\perp \end{bmatrix} \cdot \begin{bmatrix} \hat{\Sigma} \\ 0 \end{bmatrix} \cdot V^* = \hat{U}\Sigma V^*,$$

según lo anterior, la columna  $U^\perp$  mide un vector en el espacio, el cual es complementario y ortogonal a  $\hat{U}$ . Posteriormente, los elementos de la diagonal están comprendidos  $\hat{\Sigma} \in \mathbb{C}^{m \times m}$ . Concluyendo: el rango de  $A$  es igual al número de valores singulares no negativos, como se puede ver a continuación:

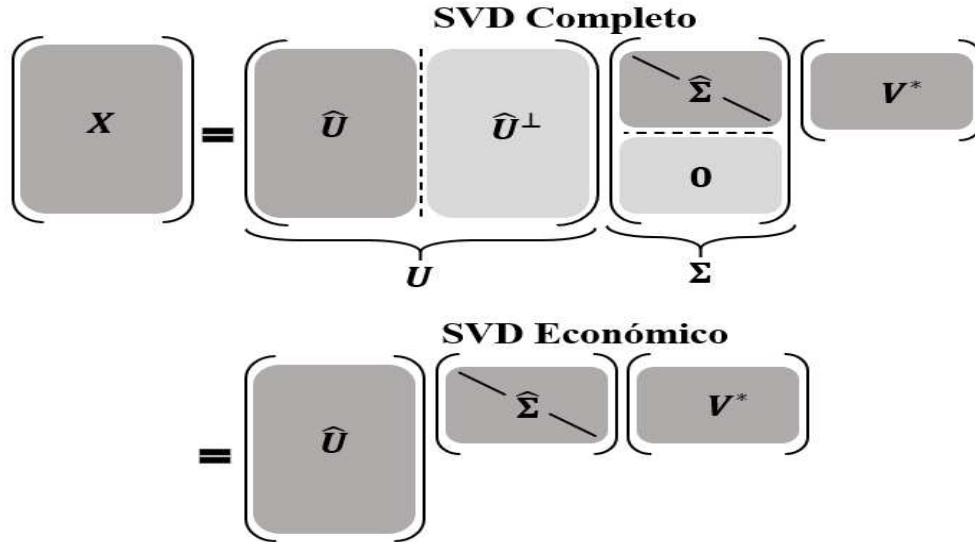


Figura A.1: Esquema de matrices, para SVD Completo y SVD económico.[6]

### Relación entre el método económico y el método truncado para SVD

Para entender de qué se trata esta relación, es necesario conocer el método de truncamiento, denotando:  $\tilde{U}$ ,  $\tilde{\Sigma}$  y  $\tilde{V}$  como matrices truncadas, donde solo ocurre esto si  $A$  no tiene un rango completo. Conllevando que para algunos valores singulares de  $\tilde{\Sigma}$  puedan ser cero y por lo tanto, este método llegue a ser exacto. Sin embargo, en valores de truncamiento donde  $r$  sea más pequeño que los valores singulares distintos de cero, se puede obtener que solo se aproxime a  $A$ , siendo de la siguiente manera:

$$A \approx \tilde{U} \tilde{\Sigma} \tilde{V}^*,$$

si se llegara a escoger un truncamiento específico donde los valores singulares sean diferentes de cero, se tendría que:  $A = \tilde{U} \tilde{\Sigma} \tilde{V}^*$  sería exacta. y su esquema es el siguiente:

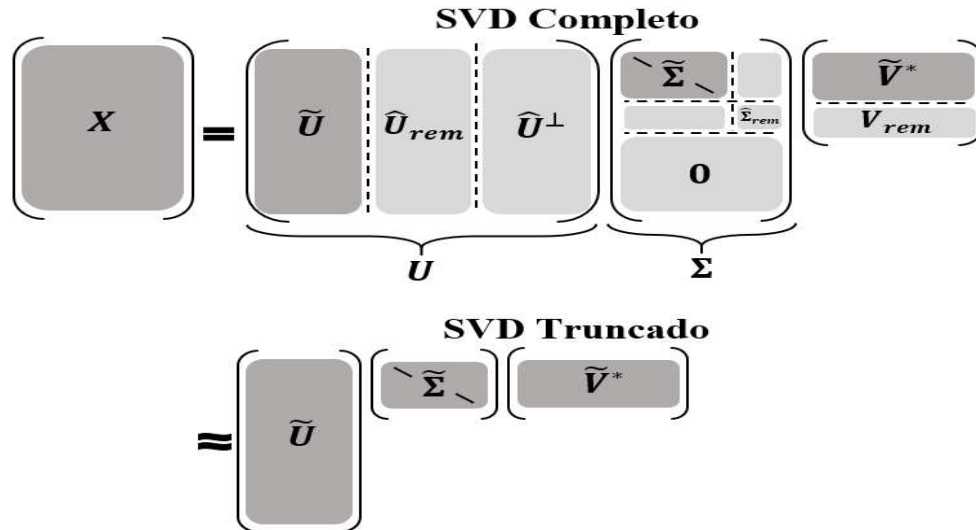


Figura A.2: Esquema de matrices para SVD Completo y SVD truncado.[6]

Es importante tener presente que una de las estrechas relaciones entre el método económico y truncado de SVD, es que reducen la dimensionalidad de una matriz original, obteniendo un mecanismo más eficiente.

### Norma de Frobenius

La norma de Frobenius o también conocida como la norma de Hilbert - Schmidt, está definida por una matriz  $A \in M_{m \times n}(\mathbb{C})$ , tal que:

$$\|A\|_F := (tr(A * A))^{1/2} = \left( \sum_{j=1}^m \sum_{k=1}^n |A_{j,k}|^2 \right)^{1/2},$$

cabe aclarar, que la variable  $tr$  es la función traza para las matrices.

**Requisitos:** dentro de esta norma, son relevantes los siguientes requisitos: que la norma esté en  $\mathbb{C}^n$  y  $\mathbb{R}^n$ , que tenga normas matriciales inducidas por normas vectoriales, notaciones para renglones y columnas de una matriz, desigualdad de Cauchy, definición formal del producto de matrices, propiedades de la multiplicación de matrices, producto interno en un espacio vectorial complejo y por ultimo propiedades de la traza de matrices, ver [3].





# Bibliografía

- [1] C. ABELLA BARRETO, GRUPO DE VIROLOGÍA, SUBDIRECCIÓN LABORATORIO NACIONAL DE REFERENCIA, AND DIRECCIÓN REDES EN SALUD PÚBLICA, *Guía para la vigilancia por laboratorio del virus de la influenza y otros virus respiratorios*, Instituto Nacional de Salud, (3027), pp. 1–26.
- [2] ALCALDIA DE SANTIAGO DE CALI, *Otros COVID-19*, 2020.
- [3] D. C. AQUINO AND E. MAXIMENKO, *Norma de Frobenius*, IPN— Instituto Politecnico Nacional, (2002), pp. 1–8.
- [4] D. ARANGO, *Evolución de los Casos de COVID-19 en Cali durante los primeros 2 meses de la Pandemia*, tech. rep., Secretaria de Salud, Santiago de Cali, Cali, Colombia, 2020.
- [5] B. W. BRUNTON, L. A. JOHNSON, J. G. OJEMANN, AND J. N. KUTZ, *Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition*, *Journal of Neuroscience Methods*, 258 (2016), pp. 1–15.
- [6] S. L. BRUNTON AND J. N. KUTZ, *Data-driven science and engineering: machine learning, dynamical systems, and control*, Cambridge University Press, 2019.
- [7] S. CARLOS ALBEROLA, S. GALLARDO BERMELL, AND S. MARTORELL ALSINA, *Modelos compartimentales para el cálculo de dosis interna*, *Modelling in Science Education and Learning*, 6 (2013), p. 19.
- [8] A. CHATTERJEE, *An introduction to the proper orthogonal decomposition*, *Current Science*, 78 (2000), pp. 808–817.
- [9] K. J. CLERKIN, J. A. FRIED, J. RAIKHELKAR, G. SAYER, J. M. GRIFFIN, A. MASOUMI, S. S. JAIN, D. BURKHOFF, D. KUMARAIHAH, L. R. RABBANI, A. SCHWARTZ, AND N. URIEL, *COVID-19 and Cardiovascular Disease*, 2020.
- [10] DEPARTAMENTO ADMINISTRATIVO DE PLANEACIÓN MUNICIPAL, *Documentos de Cali en Cifras*, 2020.
- [11] Y. HAN AND L. TAN, *Dynamic mode decomposition and reconstruction of tip leakage vortex in a mixed flow pump as turbine at pump mode*, *Renewable Energy*, 155 (2020), pp. 725–734.
- [12] H. HOTELLING, *Analysis of a complex of statistical variables into principal components*, *Journal of Educational Psychology*, 24 (1933), pp. 498–520.
- [13] A. R. JOAQUÍN, *Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE*, RStudio Pubs, (2017).

- 
- [14] I. T. JOLLIFE AND J. CADIMA, *Principal component analysis: A review and recent developments*, 2016.
- [15] J. N. KUTZ, S. L. BRUNTON, B. W. BRUNTON, AND J. L. PROCTOR, *Dynamic mode decomposition: data-driven modeling of complex systems*, SIAM, 2016.
- [16] N. LURIE, M. SAVILLE, R. HATCHETT, AND J. HALTON, *Developing covid-19 vaccines at pandemic speed*, 2020.
- [17] P. MEHTA, D. F. MCAULEY, M. BROWN, E. SANCHEZ, R. S. TATTERSALL, AND J. J. MANSON, *COVID-19: consider cytokine storm syndromes and immunosuppression*, 2020.
- [18] P. MILLS, *Singular Value Decomposition (SVD) Tutorial: Applications, Examples, Exercises*, 2017.
- [19] K. MILTON, *Ecologías: antropología, cultura y entorno*, 2007.
- [20] U. S. A. MIT CAMBRIDGE, *Singular Value Decomposition (SVD) tutorial*, 2002.
- [21] N. MURCIA, S. S. JAIMES, AND J. GÓMEZ, *La práctica social como expresión de humanidad*, *Cinta de moebio*, 57 (2016), pp. 257–274.
- [22] W. H. ORGANIZATION, *Coronavirus disease (COVID-19) Situation Report – 197 Data*, tech. rep., World Health Organization, 2020.
- [23] K. PEARSON, *LIII. On lines and planes of closest fit to systems of points in space*, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2 (1901), pp. 559–572.
- [24] J. L. PROCTOR, S. L. BRUNTON, AND J. NATHAN KUTZ, *Generalizing koopman theory to allow for inputs and control*, *SIAM Journal on Applied Dynamical Systems*, 17 (2018), pp. 909–930.
- [25] J. L. PROCTOR AND P. A. ECKHOFF, *Discovering dynamic patterns from infectious disease data using dynamic mode decomposition*, *International Health*, 11 (2019), pp. 83–92.
- [26] J. SALAS, A. TORRENTE, AND E. VILLASEÑOR, *Pseudoinversa y descomposición en valores singulares*, 2002.
- [27] M. SASAKI, Y. KAWACHI, R. O. DENDY, H. ARAKAWA, N. KASUYA, F. KIN, K. YAMASAKI, AND S. INAGAKI, *Using dynamical mode decomposition to extract the limit cycle dynamics of modulated turbulence in a plasma simulation*, *Plasma Physics and Controlled Fusion*, 61 (2019), p. 112001.
- [28] T. SAUER, *Numerical solution of stochastic differential equations in finance*, PEARSON, second ed., 2012.

- 
- [29] P. J. SCHMID, *Dynamic mode decomposition of numerical and experimental data*, Journal of Fluid Mechanics, 656 (2010), pp. 5–28.
- [30] J. SHLENS, *A Tutorial on Independent Component Analysis*, Google Research, 1 (2014), p. 13.
- [31] J. V. STONE, *Independent Component Analysis*, JOHNWILEY{\&} SONS, INC., New York / Chichester / Weinheim / Brisbane / Singapore / Toronto, 2018.
- [32] G. TISSOT, L. CORDIER, N. BENARD, AND B. R. NOACK, *Model reduction using Dynamic Mode Decomposition*, 2014.
- [33] UNLP— UNIVERSIDAD NACIONAL DE LA PLATA, *Descomposición en Valores singulares(SVD)*, 2010.
- [34] M. A. VICENTE, C. FERNÁNDEZ, A. GIL, AND L. PAYÁ, *Equivalencia entre ICA y PCA como métodos de extracción de características en reconocimiento visual basado en apariencia*, umh, 1 (2016), p. 6.
- [35] I. ZABALLA, *Capítulo 3 Valores singulares 3.1. Introducción*, tech. rep., Universidad del País Vasco / Euskal Herriko Unibertsitatea, Leioa, España, 2000.
- [36] Q. ZHANG, Y. LIU, AND S. WANG, *The identification of coherent structures using proper orthogonal decomposition and dynamic mode decomposition*, Journal of Fluids and Structures, 49 (2014), pp. 53–72.
- [37] Z. Y. ZU, M. D. JIANG, P. P. XU, W. CHEN, Q. Q. NI, G. M. LU, AND L. J. ZHANG, *Coronavirus Disease 2019 (COVID-19): A Perspective from China*, Radiology, (2020), p. 200490.