Singular Value Decomposition (SVD) y Dynamic Mode Decomposition (DMD)

Juan Diego Naranjo

May 20, 2025

1 SVD

1.1 ¿Qué es el SVD?

La descomposición en valores singulares (SVD) es una factorización de matrices en dos matrices unitarias y una diagonal.

Teorema 1.1: Descomposición SVD

Dada una matriz $A \in \mathcal{M}_{m,n}(\mathbb{R})$, existen matrices ortogonales $U \in \mathcal{M}_m(\mathbb{R})$ y $V \in \mathcal{M}_n(\mathbb{R})$ y una matriz diagonal $\Sigma \in \mathcal{M}_{m,n}(\mathbb{R})$ tales que

$$A = U\Sigma V^T$$

Sea $A \in \mathcal{M}_{m,n}(\mathbb{R})$ una matriz con rango $\operatorname{Rank}(A) = r$. Notamos que $A^T A \in \mathcal{M}_n(\mathbb{R})$ es una matriz simétrica y semidefinida positiva.

- $(A^T A)^T = A^T A$ (Es simétrica)

- Sean λ y $\mathbf{v}$ valores y vectores propios de $A^T A$. Luego,

  $$A^T A\mathbf{v} = \lambda\mathbf{v} \implies \mathbf{v}^T A^T A\mathbf{v} = \mathbf{v}^T(\lambda\mathbf{v}) = \lambda\mathbf{v}^T\mathbf{v} = \lambda\|\mathbf{v}\|^2$$

  Además,

  $$\mathbf{v}^T A^T A\mathbf{v} = (A\mathbf{v})^T(A\mathbf{v}) = \|A\mathbf{v}\|^2 \geq 0$$

  Entonces, $\lambda \geq 0$. Es decir, $A^T A$ es semidefinida positiva.

Teorema 1.2: Teorema Espectral (versión para matrices reales)

Si una matriz $A \in \mathcal{M}_n(\mathbb{R})$ es simétrica, entonces existe una base ortonormal de $\mathbb{R}^n$ formada a partir de los vectores propios de A, y los valores propios de A son reales.

Teniendo en cuenta el teorema espectral, existe una base ortonormal de $\mathbb{R}^n$ formada por los vectores propios de A^TA. Sea $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ dicha base ortonormal. Adicionalmente, sabemos que $\operatorname{Rank}(A^TA) = \operatorname{Rank}(A) = r$ es el número de valores propios no nulos de A^TA. Sean $\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$ los valores propios de A^TA ordenados de forma decreciente.

Por el teorema fundamental del álgebra lineal sabemos que $\dim N(A) = n-r$. Además, tenemos que los vectores propios $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ estan asociados a los valores propios nulos de A^TA; es decir, $A^TA\mathbf{v}_j = 0$ cuando $r < j \leq n$. Como $N(A) = N(A^TA)$, se sigue que $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ es una base ortonormal de $N(A)$.

Definimos los vectores $\mathbf{u}_i \in \mathbb{R}^m$ como

$$\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i \quad 1 \leq i \leq r$$

donde $\sigma_i = \sqrt{\lambda_i}$.

Verificamos que $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ es un conjunto ortonormal en $\mathbb{R}^m$:

- (Ortogonal)

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \left\langle \frac{1}{\sigma_i} A\mathbf{v}_i, \frac{1}{\sigma_j} A\mathbf{v}_j \right\rangle = \frac{1}{\sigma_i\sigma_j} \left\langle \mathbf{v}_i, A^TA\mathbf{v}_j \right\rangle = \frac{1}{\sigma_i\sigma_j} \langle \mathbf{v}_i, \lambda_j \mathbf{v}_j \rangle = \frac{\lambda_j}{\sigma_i\sigma_j} \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \frac{\lambda_j}{\sigma_i\sigma_j}(0) = 0$$

- (Normal)

$$\|\mathbf{u}_i\|^2 = \left\| \frac{1}{\sigma_i} A\mathbf{v}_i \right\|^2 = \left\langle \frac{1}{\sigma_i} A\mathbf{v}_i, \frac{1}{\sigma_i} A\mathbf{v}_i \right\rangle = \frac{1}{\sigma_i^2} \left\langle \mathbf{v}_i, A^TA\mathbf{v}_i \right\rangle = \frac{1}{\sigma_i^2} \langle \mathbf{v}_i, \lambda_i \mathbf{v}_i \rangle = \frac{\lambda_i}{\sigma_i^2} \langle \mathbf{v}_i, \mathbf{v}_i \rangle = \frac{\sigma_i^2}{\sigma_i^2} = 1$$

Si $r < m$ podemos encontrar los $m - r$ vectores restantes para formar una base ortonormal $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\} \cup \{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ de $\mathbb{R}^m$ (usando Gram-Schmidt por ejemplo).

Ahora bien, notamos que

$$A\mathbf{v}_i = \sigma_i \frac{1}{\sigma_i} A\mathbf{v}_i = \sigma_i \mathbf{u}_i$$

donde $A\mathbf{v}_i = 0$ para $r < i \leq n$.
Reescribiendo esto en forma matricial tenemos que

$$V = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \ldots & \mathbf{v}_n \\ | & & | \end{bmatrix} \in \mathcal{M}_n(\mathbb{R}) \quad U = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \ldots & \mathbf{u}_m \\ | & & | \end{bmatrix} \in \mathcal{M}_m(\mathbb{R})$$

son matrices ortogonales y

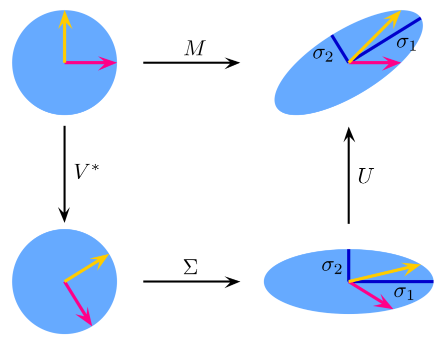
$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \in \mathcal{M}_{m,n}(\mathbb{R})$$

es una matriz diagonal tales que

$$AV = U\Sigma$$

Dado que V es una matriz ortogonal, concluimos que

$$A = U\Sigma V^T$$



$$M = U \cdot \Sigma \cdot V^*$$

Figure 1: Ilustración del efecto de una matriz M de 2×2 sobre el disco unitario y la base canónica de $\mathbb{R}^2$

1.2 Interpretación

Geométricamente, aplicar una matriz A a un vector $\mathbf{x} \in \mathbb{R}^n$ se puede entender como una secuencia de tres transformaciones:

1. Primero, $V^T\mathbf{x}$ reexpresa el vector $\mathbf{x}$ en la base ortonormal formada por los vectores propios de A^TA. Es un **cambio de base** en el dominio.

2. Luego, Σ actúa sobre el resultado, **escalando** las coordenadas a lo largo de direcciones específicas. Esta operación representa un **estiramiento o compresión** de cada componente por su valor singular σ_i.

3. Finalmente, U aplica una **rotación o reflexión** que alinea el resultado con la base ortonormal del codominio.

En resumen, la acción de A sobre cualquier vector se descompone en:

$$\text{rotación/reflexión} \longleftarrow \text{escalamiento} \longleftarrow \text{cambio de base (rotación).}$$

Desde un punto de vista geométrico, la SVD revela que toda transformación lineal con una representación matricial puede descomponerse en una rotación (o reflexión), seguida de un escalamiento, seguida de otra rotación (o reflexión). Esta interpretación permite entender mejor el comportamiento de una matriz: cómo estira, aplasta y gira el espacio.

1.3 Aplicaciones del SVD: Mínimos Cuadrados, Pseudoinversa y SVD Truncado

1. Pseudoinversa

Definición 1.1: Pseudoinversa de Moore-Penrose

Dada una matriz $A \in \mathcal{M}_{m,n}(\mathbb{R})$ con descomposición SVD $A = U\Sigma V^T$, su pseudoinversa está dada por

$$A^\dagger = V\Sigma^\dagger U^T$$

donde $\Sigma^\dagger \in \mathcal{M}_{n,m}(\mathbb{R})$ es la matriz diagonal que resulta de tomar los recíprocos de los valores singulares no nulos σ_i en Σ y trasponiendo.

La pseudoinversa $A^\dagger$ de una matriz $A \in \mathcal{M}_{m,n}(\mathbb{R})$ es una solución al problema de mínimos cuadrados:

Dado un sistema de ecuaciones lineales

$$Ax = b$$

se cumple que

$$\|Ay - b\|_2 \leq \|Ax - b\|$$

para todo $x \in \mathbb{R}^n$, donde $y = A^\dagger b$

Este resultado se puede extender a sistemas de sistemas. Es decir, si $B \in \mathcal{M}_{m,p}(\mathbb{R})$ entonces

$$\|AY - B\|_F \leq \|AX - B\|_F$$

para todo $X \in \mathcal{M}_{n,p}$, donde $Y = A^\dagger B$ y $\| * \|_F$ es la norma Frobenius.

2. SVD Truncado y Aproximación de Bajo Rango

El SVD también permite aproximar una matriz A por una de rango bajo. Sea $A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ su descomposición completa.

Podemos truncar la suma en los primeros $k < r$ términos:

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

Entonces, A_k es una matriz de rango k que aproxima a A. El **teorema de Eckart–Young** establece que esta es la mejor aproximación posible en norma de Frobenius o norma 2:

$$\|A - A_k\|_2 = \sigma_{k+1}, \quad \|A - A_k\|_F^2 = \sum_{i=k+1}^{r} \sigma_i^2.$$

Esta propiedad hace que el SVD truncado sea muy útil en compresión, reducción de dimensionalidad y eliminación de ruido (por ejemplo, en PCA y procesamiento de imágenes).

2 DMD

2.1 ¿Qué es el DMD?

El modelado y control de sistemas complejos basado en datos es un campo importante en la ingeniería, biología y la física. Aunque hoy en día existe una disponibilidad sin precedentes de mediciones precisas provenientes de registros históricos, simulaciones numéricas y datos experimentales, los modelos siguen siendo difíciles de obtener. Esto se debe a que los sistemas de interés, tales como fluidos turbulentos, sistemas epidemiológicos, redes neuronales, mercados financieros o el clima, son sistemas dinámicos no lineales de alta dimensión que presentan grandes desafíos computacionales y de interpretabilidad. Sin embargo, muchos de estos sistemas evolucionan sobre un atractor de baja dimensión que puede describirse mediante estructuras espaciotemporales coherentes. La descomposición en modos dinámicos (DMD) es una técnica poderosa que permite descubrir estas estructuras de baja dimensión a partir de datos de alta dimensión.

El método DMD se originó en la comunidad de dinámica de fluidos como una técnica para descomponer flujos complejos en una representación simple basada en estructuras espaciotemporales coherentes. Schmid y Sesterhenn, fueron quienes definieron por primera vez el algoritmo DMD y demostraron su capacidad para extraer información valiosa a partir de datos de fluidos de alta dimensión. El creciente éxito de DMD se debe a que es un método basado en datos y libre de ecuaciones, capaz de descomponer con precisión un sistema complejo en estructuras espaciotemporales coherentes, lo que permite su uso en predicciones a corto plazo y en control de sistemas.

El método DMD se basa en la recolección de datos $\mathbf{x}_k$ que representan el estado de un sistema dinámico en diferentes momentos t_k, donde $k = 1, 2, \ldots, m$. Algorítmicamente, el DMD consiste en una regresión lineal de los datos sobre una dinámica local:

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

donde la matriz A se elige para minimizar la norma

$$\|\mathbf{x}_{k+1} - A\mathbf{x}_k\|_2$$

sobre $k = 1, 2, 3, \ldots, m - 1$.

Las ventajas de este método son su simplicidad de implementación y el hecho de que requiere muy pocas suposiciones sobre el sistema subyacente. El costo computacional principal del algoritmo es realizar una descomposición en valores singulares (SVD) de la matriz de estados construida a partir de los datos $\mathbf{x}_k$.

2.2 Arquitectura

Consideramos el sistema dinámico

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

Donde $\mathbf{x}(t) \in \mathbb{R}^n$ es el vector que representa el estado del sistema en el tiempo t, μ son los parámetros del sistema y f es la dinámica del sistema.

Como ya se mencionó, la dimension del sistema es muy grande ($n >> 1$) y su dinámica f (posiblemente desconocida) es no-lineal. Para predecir estados futuros o caracterizar el sistema, intentamos simplificar el modelo. Para ello, buscamos aproximar localmente la dinámica del sistema con un sistema lineal de baja dimensionalidad

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, t; \mu) \quad \rightsquigarrow \quad \frac{d\mathbf{x}}{dt} = \mathcal{A}\mathbf{x}$$

Si tenemos la condición inicial $\mathbf{x}(0)$, la solución está dada por

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

,

donde $\boldsymbol{\phi}_k$ y ω_k son los vectores y valores propios de la matrix $\mathcal{A}$, y los coeficientes b_k son las coordenadas de $\mathbf{x}(0)$ en la base generada por los vectores propios. Así, el objetivo del DMD es encontrar la matriz A que mejor aproxime la solución.

Supongamos que tenemos a nuestra disposición $\mathbf{x}(t)$ en $t_1, ..., t_m$. Definimos las matrices de estado X y X':

$$X = \begin{bmatrix} | & & | \\ x_1 & ... & x_{m-1} \\ | & & | \end{bmatrix} \quad X' = \begin{bmatrix} | & & | \\ x_2 & ... & x_m \\ | & & | \end{bmatrix}$$

Buscamos una matriz A que nos lleve de un estado al siguiente con el menor error posible. Es decir,

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

tal que la norma

$$\|\mathbf{x}_{k+1} - A\mathbf{x}_k\|_2$$

sea mínima para todo $k = 1, 2 \ldots, m-1$.

Como ya se mencionó, la matriz A que mejor se ajusta al conjunto de datos está dada por

$$A = X'X^{\dagger}$$

donde $X^{\dagger}$ es la matriz psudoinversa de X. Esta solución minimiza el error

$$\|X' - AX\|_F$$

Teniendo en cuenta que la dimension de $A = X'X^{\dagger}$ es bastante grande, calcular los vectores y valores propios necesarios para calcular la pseudoinversa con SVD llega a ser muy costoso. Para remediar este inconveniente, reducimos la dimensionalidad de la matriz de aproximación usando descomposición SVD truncada.

2.3 Algoritmo DMD

Teniendo las matrices de estado X y X' tomamos $k \leq r = \text{Rank}(X)$.

1. (Descomposición SVD truncada)

$$X = U_k \Sigma_k^{\dagger} V_k^T$$

2. (Transformación lineal de semejanza)

$$\tilde{A} = U_k^T X' V_k \Sigma^{\dagger}$$

3. (Autovectores y autovalores)

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

4. (Modos)

$$\Phi = X' V \Sigma^{\dagger} W$$

2.4 Interpretación

Descomposición SVD truncada

La descomposición SVD truncada de X se puede entender dinámicamente como:

- U_k contiene las direcciones principales (modos espaciales),
- Σ_k contiene los valores singulares (energía de cada modo),
- V_k describe la evolución temporal de los modos.

Proyectamos los datos a un subespacio de dimensión reducida que captura la estructura esencial del sistema, eliminando ruido y redundancia.

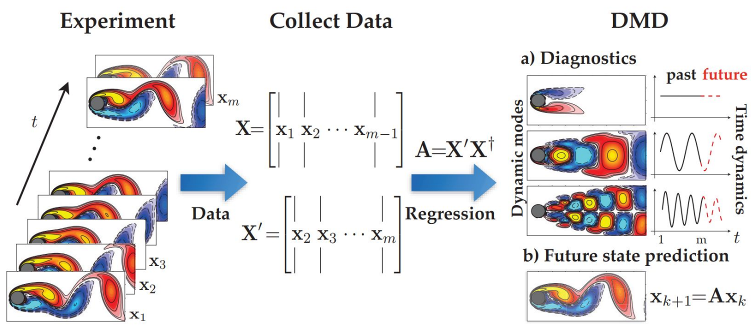


Figure 2: Diagrama del metodo DMD

Transformación lineal de semejanza

Construimos una matriz reducida $\tilde{A}$ que actúa en el subespacio generado por los modos dominantes. Esta matriz tiene una dimensión mucho menor que A, pero preserva su dinámica principal.

Autovalores y autovectores

Diagonalizamos $\tilde{A}$ para obtener:

- Λ: matriz diagonal con los autovalores (frecuencias y tasas de crecimiento/decaimiento),
- W: matriz de autovectores en el espacio reducido.

Estos representan las frecuencias dominantes y cómo evolucionan los modos principales en el tiempo.

Reconstrucción de los modos DMD

Reconstruimos los modos dinámicos Φ en el espacio original, donde cada columna de Φ es una forma espacial que evoluciona en el tiempo según su autovalor en Λ. Estas formas y sus dinámicas explican cómo evoluciona el sistema.

En resumen, el algoritmo DMD:

- Busca una matriz lineal que aproxime la evolución del sistema dinámico.

- Usa SVD para reducir la dimensión del problema y filtrar ruido.
- Extrae modos dominantes y sus dinámicas temporales.

2.5 Usos e Interpretaciones

El DMD tiene varios usos e interpretaciones. Generalmente, permite realizar tres tareas principales:

1. **Diagnóstico.**
   Originalmente, el algoritmo DMD se utilizó como herramienta de diagnóstico para caracterizar flujos de fluidos complejos. Extrae características espaciotemporales clave y de bajo rango de sistemas de alta dimensión, lo que permite obtener resultados físicamente interpretables en términos de estructuras espaciales y sus respuestas temporales asociadas. Este sigue siendo uno de los usos principales, ya que permite descubrir estructuras fundamentales en sistemas complejos, de forma similar al análisis POD (Proper Orthogonal Decomposition) en áreas como dinámica de fluidos, física de plasmas y modelado atmosférico.

2. **Estimación del estado y predicción futura.**
   Un uso más avanzado del DMD consiste en construir modelos dinámicos utilizando las estructuras espaciotemporales dominantes en los datos, con el objetivo de estimar el estado futuro del sistema en regiones donde no se han hecho mediciones. Aunque esto es complejo, debido a que DMD ajusta un sistema lineal a datos generados por un sistema no lineal, existen estrategias como el muestreo inteligente y la actualización del modelo que permiten construir modelos útiles. Este enfoque generativo se ha utilizado con éxito en diversas aplicaciones para la predicción futura de sistemas dinámicos.

3. **Control.**
   El objetivo más desafiante del DMD es desarrollar estrategias de control viables y robustas a partir del muestreo de datos. Dado que se utiliza un modelo lineal para predecir un sistema no lineal, la precisión de la predicción solo se mantiene durante una ventana de tiempo limitada. Sin embargo, si esta ventana es suficiente para tomar una decisión de control eficaz, es posible influir en el estado futuro del sistema. Este enfoque proporciona un marco matemático atractivo para controlar sistemas dinámicos complejos cuyos modelos no se conocen o son difíciles de obtener computacionalmente.

3 Bibliografía

- Horn, R.A. and Johnson, C.R. (2018) Matrix analysis. New York: Cambridge University Press.

- Johnston, N. (2021) Advanced linear and matrix algebra. Cham, Cham: Springer International Publishing Springer.

- Kutz, N.J. et al. (2016) Dynamic mode decomposition: Data-driven modeling of Complex Systems. Philadelphia (Pa.): Society for industrial and applied mathematics.