

UNIVERSIDAD DE SONORA



DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE FÍSICA

**Análisis de Componentes Principales de Espectros
Raman de Glóbulos Rojos con distintos niveles de
Radiación**

TESIS

QUE PARA OBTENER EL TÍTULO DE
LICENCIADO EN FÍSICA

PRESENTA:

FRANCISCO ALEJANDRO PEDROZA MONTERO

ENERO 2017

Universidad de Sonora

Repositorio Institucional UNISON



**"El saber de mis hijos
hará mi grandeza"**



Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess

UNIVERSIDAD DE SONORA

Resumen

Departamento de Física

Licenciatura en Física

Análisis de Componentes Principales de Espectros Raman de Glóbulos Rojos con distintos niveles de Radiación

by Francisco Alejandro PEDROZA MONTERO

La *Espectroscopia Raman* es una técnica usada para la observación de modos rotacionales y vibracionales, o bien, estados energéticos ligados a ciertas longitudes de onda en una molécula (o átomo), que permiten caracterizarla dando como resultado un espectro Raman particular para cada muestra. La necesidad que va más allá de la identificación de estos estados vibracionales mencionados va ligada a la interpretación adecuada de los datos y es por este motivo que se presenta el método de Análisis de Componentes Principales como auxiliar para la manipulación e interpretación de muchos espectros Raman.

El propósito principal de este escrito es mostrar el uso (y versatilidad) del método para la interpretación de muchos datos. Esta técnica reduce la dimensión del conjunto de datos y en consecuencia hace el manejo de conjuntos grandes de datos mucho más cómodo. El método descansa sobre el álgebra lineal y su uso es generalizado para cualquier conjunto de datos.

Índice general

Resumen	III
1. Introducción	1
1.1. Espectroscopia Raman	1
1.2. Sangre Irradiada	4
2. Marco Teórico	7
2.1. Objetivo	8
2.2. Cambio de Base	10
2.3. Ruido y Rotación	12
2.4. Redundancia	14
2.5. La Matriz de Covarianzas	15
2.5.1. Diagonalización de la Matriz de Covarianzas . . .	17
2.5.2. Solución del ACP con Descomposición en Vectores Propios	19
3. Resultados y Discusión	21
3.1. Tratamiento de Datos	22
3.2. Análisis de Componentes Principales	24
3.2.1. Muestra de Control	24
3.2.2. Muestra Irradiada a 30 Gy	26
3.2.3. Muestra Irradiada a 100 Gy	28
3.3. Comparación de espectros Raman y Componentes Principales	31
4. Conclusiones	33
A. Apéndice	41

DEDICADO A MI FAMILIA Y A MIS MAESTROS.

Capítulo 1

Introducción

La **Espectroscopía Óptica** es la medición de la interacción de la radiación electromagnética (luz, por ejemplo) con cierto material. I. Newton con sus experimentos relacionados con la óptica lo llevaron a demostrar que la luz blanca podía ser dispersada al pasar a través de un prisma y se podían observar los distintos colores en los que se descomponía. Los fenómenos que corresponden a dichas interacciones son la reflexión, refracción, dispersión, absorción y emisión. Estos tipos de interacción luego llegan a ser métodos básicos para analizar una muestra ya que es posible determinar de qué se compone ésta.

El espectro de radiación electromagnética varía desde ondas muy grandes, como lo son las ondas de radio, hasta muy pequeñas, como pueden ser los rayos X. Diversas técnicas espectroscópicas son usadas frecuentemente para caracterizar muestras: por ejemplo, muestras biológicas. Entre las técnicas más conocidas destacan la espectroscopía Infrarroja (**IR**), la espectroscopía Ultravioleta (**UV**), la espectroscopía Ultravioleta-Visible (**UV-Visible**) y la espectroscopía **Raman**. Estas técnicas son basadas en alguna de las interacciones ya mencionadas; por ejemplo, la espectroscopía Raman se basa en la dispersión de luz.

1.1. Espectroscopia Raman

Esta técnica recibe su nombre tras su descubridor (en 1928): Sir Chandrasekhara Venkata Raman, quien trabajó principalmente en el campo de la dispersión de luz (lo cual le consiguió el premio Nobel en física en 1930). En sus trabajos descubrió que cuando la luz atraviesa un material transparente, las ondas reflejadas por el material cambian su longitud de onda (λ). Este fenómeno, con posterioridad llamado Efecto Raman. Su descubrimiento se remonta a experimentos con líquidos (C.

V. Raman) y luego en cristales (G. Landsberg y L. Mandelstam). Este efecto fue predicho por un físico teórico austriaco (Adolf Smekal) en 1923.

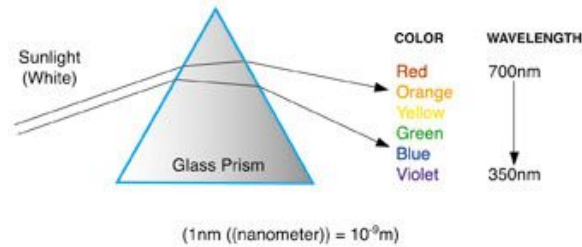


FIGURA 1.1: Dispersión de la luz blanca por un prisma.

El Efecto Raman es la aparición de líneas adicionales en el espectro de luz monocromática que ha sido dispersada por un material transparente. La energía (y por consiguiente la frecuencia y longitud de onda) de la luz dispersada es distinta debido a que la luz (incidente) imparte o quita energía vibracional y/o rotacional; las líneas espectrales de la luz dispersada tendrán una línea prominente correspondiente a la longitud de onda de la luz incidente y tendrá además líneas secundarias que serán longitudes de ondas alteradas.

Raman observó que no todas las longitudes de ondas dispersadas eran las mismas. Observó que también se tenían longitudes de onda adicionales, las cuales aisló con un filtro en la salida del haz de luz. Raman mostró (en más de 60 muestras) que se podían observar longitudes de ondas distintas. Alrededor de 1 fotón en 1 millón de fotones es dispersado. Las observaciones cualitativas (de colores) por si solas no son suficientes para poder caracterizar un material o sustancia. Se necesita medir de manera precisa la longitud de onda de la luz incidente y las longitudes de onda dispersadas; para esto Raman utilizó un espectrómetro de cuarzo para poder medir las longitudes de onda a la salida. Estos resultados fueron publicados en 1928 en el Indian Journal of Physics [2].

El Efecto Raman ocurre cuando radiación electromagnética incide sobre la materia e interactúa con los electrones y enlaces moleculares presentes en ella. Esta radiación dispersa fotones inelásticamente; sin embargo, esto es **independiente** de la onda incidente: esto es debido

a que la longitud de onda de la onda incidente sólo marca un origen para las mediciones relativas de los modos normales de vibración que provienen de la misma muestra; por lo que, se obtiene una señal Raman exactamente a la misma diferencia de números de onda $\Delta\omega$:

$$\Delta\omega = \frac{1}{\lambda_0} - \frac{1}{\lambda_1}. \quad (1.1)$$

Esta cantidad $\Delta\omega$ tiene unidades de longitud⁻¹ y se le denomina Raman Shift o bien desplazamiento Raman donde: λ_0 es la longitud de onda de excitación y λ_1 es la longitud de onda del espectro Raman.

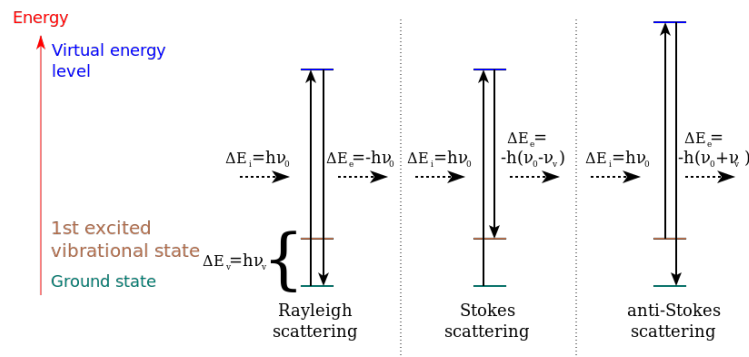


FIGURA 1.2: Diferentes casos para la dispersión de luz. En la dispersión de **Rayleigh** no hay un intercambio de energía; es decir, el átomo vuelve a su estado base.

El Raman es de tipo Stokes cuando la dispersión involucra la activación de un modo vibracional del material, y por tanto la luz pierde energía, y será Raman Anti-Stokes cuando la energía de un modo vibracional se suma a la energía del haz incidente. Habitualmente se trabaja en el rango de Raman Stokes y las unidades se expresan en cm^{-1} tomando el origen en la energía del haz incidente [26]. Por tanto la espectroscopia Raman nos da información sobre los modos vibracionales del material. Para que ocurra la dispersión Raman, la interacción de la radiación electromagnética con la molécula debe provocar un cambio en la polarizabilidad, es decir que mediante la aplicación de un campo de fuerza externo, los electrones pueden ser inducidos a moverse dentro de una molécula. Estos modos están ligados a los tipos de átomos, sus enlaces y a la simetría del material y por ello se trata de una técnica que nos proporciona información estructural [28].

1.2. Sangre Irradiada



FIGURA 1.3: Un eritrocito, también llamados glóbulos rojos.

El tejido sanguíneo (sangre) está compuesto por dos tipos de elementos: una parte sólida, compuesta por glóbulos blancos, rojos y plaquetas, y la parte líquida que es el plasma sanguíneo. Entre las principales funciones de este tejido, se encuentra el transporte de nutrientes y oxígeno. Los glóbulos rojos (GR) o mejor conocidos como eritrocitos son discos bicóncavos aproximadamente de $7.5 \mu m$ de diámetro y $2.5 \mu m$ de grosor. El uso de Espectroscopia Raman en glóbulos rojos ha demostrado ser una herramienta valiosa para monitorear la dinámica molecular de los estados de oxigenación (estado Tenso y Relajado) [3,5,10].

El procedimiento de irradiación de sangre no es algo nuevo: es un método estandarizado internacionalmente que inactiva a los linfocitos, la dosis de radiación establecida esta alrededor de 25-50 Gy con el fin de reducir el riesgo a la Enfermedad injerto contra huésped [6]. Los rayos gamma son un tipo de radiación electromagnética de alta frecuencia, siendo sus fotones de alta energía. Esta radiación es de tipo ionizante¹ con gran capacidad de penetración [7]. Los iones suelen ser altamente reactivos y se estabilizan mediante la captura de cargas o bien con la unión a otros átomos o moléculas. Al interaccionar la radiación gamma γ con la sangre se desencadena una serie de procesos bioquímicos en los que se generan radicales libres, los cuales inducen el deterioro temprano de los glóbulos rojos, funcional y estructuralmente. Estudios científicos reportan que al momento de irradiar la sangre se producen *especies reactivas de oxígeno* entre ellas están radicales hidroxilo OH^-

¹El proceso de Ionización sucede cuando se arranca a un electrón de un átomo o molécula. Se le llama Ion al átomo o molécula resultante.

y peróxido de hidrógeno H_2O_2 [23]. Para evaluar este efecto, la capacidad de transporte de oxígeno de los glóbulos rojos se aborda mediante el seguimiento de los cambios en el estado de oxigenación de la proteína de Hemoglobina [3,12].

Una de las aplicaciones de la radiación gamma es la *esterilización* de tejidos, la cual se logra al reducir la reproducción de microorganismos y células. En relación a la sangre, se utiliza para la inactivación de linfocitos (células responsables de la respuesta del sistema inmune), y por lo tanto prevenir la enfermedad injerto contra huésped asociada, es una complicación poco frecuente, sin embargo cuando se presenta es letal se ha reportado una mortalidad entre el 90-100% [27]. Esta enfermedad es causada por los glóbulos blancos (linfocitos) al no ser reconocidos después de la transfusión.

En este trabajo se hará un análisis de espectros Raman en glóbulos rojos (sangre) con distintos niveles de radiación. Los niveles de radiación son distintos con el fin de poder observar que efecto tiene sobre la muestra de sangre. Debido al tamaño del conjunto de espectros Raman recopilados para este trabajo se sugiere el estudio de éstos con un análisis multivariable, con el fin de poder describir correctamente los tipos de muestras: *irradiadas* y *no irradiadas*. Esta idea se discute en el capítulo 2.2, dando lugar a la idea matemática detrás del método y es abordada con mayor extensión en el capítulo 2.5. Un beneficio del método de ACP es la reducción de las dimensiones del conjunto de espectros Raman, dando lugar a la discriminación de estructuras importantes en base a la varianza y covarianza.

Capítulo 2

Marco Teórico

En este capítulo se explicará los fundamentos del método de ACP, además de hacer una extensión de la matemática involucrada detrás de la solución mediante descomposición en vectores propios. Se aborda el tema con un ejemplo intuitivo de un sistema masa-resorte que pretende la abstracción a un conjunto de datos en general.

La matriz del conjunto de datos \mathbf{X} puede ser un conjunto de datos cualquiera; sin embargo, esta matriz toma particular importancia al momento de usar espectros Raman como el conjunto de datos. Con el método de ACP se logra extraer una cantidad de espectros *representativos* de cada muestra de sangre. El álgebra lineal sugiere un acomodo matricial del conjunto de espectros Raman, para así, realizar una serie de transformaciones que tienen en mente la extracción de la estructura Raman más representativa de todos aquellos contenidos en la matriz \mathbf{X} . Esto se logra con una transformación lineal, ecuación 2.2, un análisis de varianzas realizado en el capítulo 2.5 y su resolución final en el capítulo 2.5.2.

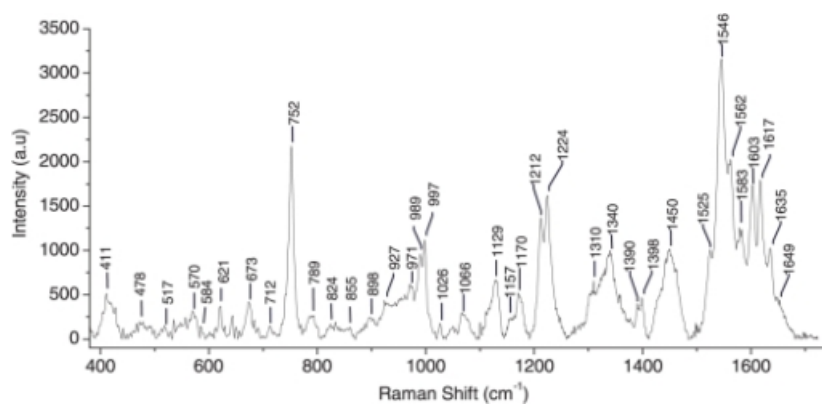


FIGURA 2.1: Espectro Raman de sangre (glóbulos rojos). Se denotan los picos más característicos [10].

En la figura 2.1 vemos que sólo hay una curva de Raman, la realidad no es así. Para la obtención de dicha curva se habrán realizado decenas de mediciones para poder observar que en efecto esos son los picos característicos y poder afirmar tener un espectro Raman adecuado para la muestra y no un espectro con **ruido**.

2.1. Objetivo

El propósito del ACP es identificar la base más significativa para re-expresar el conjunto de datos obtenidos. La esperanza es que esta nueva base remueva el ruido y revele la estructura oculta del experimento. Supongamos el siguiente ejemplo: se tiene un arreglo resorte-masa común y se pretende calcular la constante del resorte; para ello se medirán los desplazamientos con tres sensores colocados alrededor del resorte.

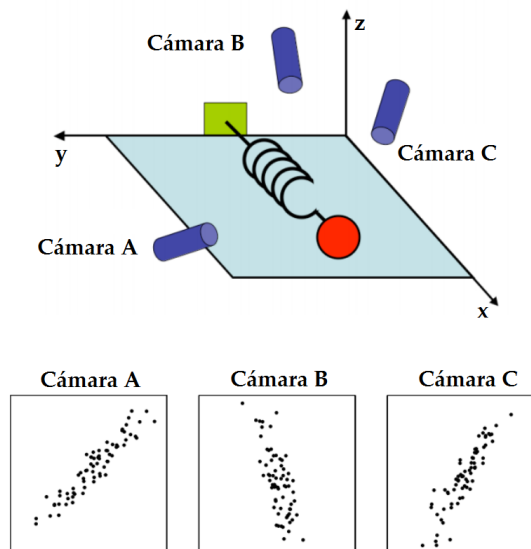


FIGURA 2.2: Es obvio que cada sensor dará información distinta de los demás; sin embargo, es el mismo experimento. [1]

El objetivo implícito de medir con tres cámaras el desplazamiento es realmente *determinar que la dinámica principal ocurre en el eje x* . En otras palabras, buscamos las dimensiones *importantes*, las redundantes y las pertenecientes al ruido. Con la definición más precisa de nuestro objetivo, también es importante *definir* nuestro conjunto de datos. Tratamos cada ensayo del experimento como una muestra individual de nuestro conjunto de datos completo: en el ejemplo del resorte podemos

expresar la posición que la cámara A tomó como (x_A, y_A) y de manera similar para las cámaras B y C. La representación matemática de un sólo ensayo de todas las cámaras (A, B y C) es representado como sigue:

$$\vec{X} = \begin{bmatrix} x_A \\ y_A \\ x_B \\ y_B \\ x_C \\ y_C \end{bmatrix}$$

Sólo para poner en perspectiva la cantidad de vectores que se obtendrían si se midiera la posición de la masa m por 10 minutos a 120 Hz, entonces se tendría la cantidad de $10 \times 60 \times 120 = 72000$ vectores [1].

Con este ejemplo concreto podemos abstraer la idea fundamental de hacer este arreglo con el vector de mediciones, cada vector \vec{X} tiene m *dimensiones* donde m es el número de mediciones hechas. Equivalentemente, podemos aseverar que cada medición es un vector que yace en un espacio vectorial m dimensional abarcado por alguna base ortonormal. Del álgebra lineal sabemos que todas las mediciones (o vectores) forman una combinación lineal de este conjunto de vectores base unitarios. Si regresamos al ejemplo del resorte y, además, solamente tomamos en cuenta la cámara A, ¿cuál sería la base ortonormal para (x_A, y_A) ? La respuesta puede ser tan obvia como $(1, 0)$, $(0, 1)$, pero ¿por qué no elegimos otra base como $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$, $(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$ o cualquier otra rotación arbitraria? La razón yace en la forma que se realiza el experimento. En el ejemplo del resorte, supongamos que la cámara A nos da la posición $(2, 2)$, eso se traslada a moverse 2 unidades hacia *la derecha* y 2 unidades hacia *arriba*, en consecuencia la base que describe mejor al experimento es la primera $(1, 0)$, $(0, 1)$, ya que ésta toma *pasos* unitarios en las direcciones canónicas. Esto puede generar problemas a la hora de la interpretación del experimento, ya que si no elegimos la base adecuada podemos terminar dando una descripción pobre del experimento. La pregunta que surge a partir de esto es ¿cómo expresar esta base con álgebra lineal? En el caso bidimensional, $(1, 0)$, $(0, 1)$ puede ser re-acomodado como vectores renglón, una matriz construida a partir de

estos vectores renglón es la matriz *identidad* I de 2×2 .

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Podemos generalizar esto para el caso m -dimensional construyendo la matriz identidad de $m \times m$ denotada por \mathbf{B} como

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_m \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \mathbf{I}, \quad (2.1)$$

donde cada renglón es un vector de la base ortonormal \mathbf{b}_i con m componentes. Podemos considerar esta base sencilla como un punto de partida. Todos nuestros datos han sido medidos en esta base y por consiguiente puede ser expresados como una combinación lineal del conjunto $\{\mathbf{b}_i\}$.

2.2. Cambio de Base

¿Hay alguna otra base, que sea una combinación lineal de la base original, que re-exprese mejor nuestro conjunto de datos? Note que se trata de *linealidad* entre bases, esta suposición simplifica la cantidad de posibles bases a elegir, ahora limitamos el método de ACP a una combinación lineal de los vectores base originales.

Sea \mathbf{X} la matriz de datos originales (lo que se mide), donde cada columna es una muestra individual (o momento en el tiempo) del conjunto de datos (e.g. \vec{X}). En el ejemplo del resorte: la matriz \mathbf{X} es de dimensiones $m \times n$ donde $m = 6$ (3 cámaras que miden dos coordenadas) y $n = 72000$ (el número de veces que se tomará una medición). Definimos la matriz \mathbf{Y} como otra matriz de $m \times n$ relacionada con \mathbf{X} bajo la transformación lineal \mathbf{P} .

$$\mathbf{PX} = \mathbf{Y} \quad (2.2)$$

La matriz \mathbf{X} es la matriz de datos originales, lo que se obtiene de medir

el experimento; sin embargo, \mathbf{Y} es la nueva representación de este conjunto de datos. Además, necesitamos definir las siguientes cantidades:

- \mathbf{p}_i son los renglones de la matriz \mathbf{P} .
- \mathbf{x}_i son las columnas de la matriz \mathbf{X} .
- \mathbf{y}_i son las columnas de la matriz \mathbf{Y} .

La ecuación 2.2 representa un cambio de base y puede tener varias interpretaciones:

1. \mathbf{P} es una matriz que tranforma a \mathbf{X} en \mathbf{Y} . Es decir, geométricamente \mathbf{P} es una matriz que alarga (o acorta) y rota a la matriz \mathbf{X} , esto es de nuevo, una transformación.
2. Los renglones de \mathbf{P} , $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m\}$, son una *nueva base* de vectores para expresar a \mathbf{X} .

Esta última puede no ser obvia. Si escribimos los productos explícitos de \mathbf{P} y \mathbf{X} :

$$\mathbf{P}\mathbf{X} = \begin{bmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_m \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_n \end{bmatrix} \quad (2.3)$$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{p}_1 \cdot \mathbf{x}_1 & \dots & \mathbf{p}_1 \cdot \mathbf{x}_n \\ \vdots & \ddots & \vdots \\ \mathbf{p}_m \cdot \mathbf{x}_1 & \dots & \mathbf{p}_m \cdot \mathbf{x}_n \end{bmatrix} \quad (2.4)$$

Si denotamos cada *columna* de \mathbf{Y} como \mathbf{y}_i se tiene:

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{p}_1 \cdot \mathbf{x}_i \\ \vdots \\ \mathbf{p}_m \cdot \mathbf{x}_i \end{bmatrix} \quad (2.5)$$

Se reconoce en la ecuación 2.5 que cada coeficiente de \mathbf{y}_i es un producto interno de \mathbf{x}_i con cada renglón correspondiente de \mathbf{P} . En otras palabras, el coeficiente j -ésimo de \mathbf{y}_i es una proyección en el renglón j -ésimo de \mathbf{P} , tenemos una nueva **base** para poder representar a \mathbf{X} ; los renglones de \mathbf{P} son un nuevo conjunto de vectores base para representar las columnas de \mathbf{X} . Si recordamos bien, cada columna de \mathbf{X} es una de las mediciones (o un momento en el tiempo en el cual se tomaron las mediciones) y lo representamos de otra forma en la ecuación 2.5

para poder trabajar con la nueva base. Asumiendo linealidad, el problema se reduce a encontrar el *cambio de base* adecuado. Los vectores renglón $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m\}$ en este tratamiento se vuelven **los componentes principales** de \mathbf{X} . Las preguntas que pudieran surgir de esta discusión podrían ser las siguientes:

- ¿Cuál es la mejor forma de re-expresar \mathbf{X} ?
- ¿Cuál es una buena elección para la matriz \mathbf{P} ?

Estas preguntas explican en general *qué tipo de características quisiéramos que \mathbf{Y} exhibiera*. Evidentemente se requieren de más suposiciones, más allá de solamente linealidad para poder aterrizar en un resultado razonable. Se profundizará más acerca de esto más adelante en este trabajo.

2.3. Ruido y Rotación

El **ruido** en las mediciones es algo que no se puede quitar pero si se puede minimizar. Sin importar qué tan grande sea el conjunto de datos; si el ruido es *mucho*, no se podrá extraer la información relevante del conjunto de datos usando el método de ACP. No existe una escala absoluta para cuantificar el ruido; sin embargo, una forma de medición común es el coeficiente (por su nombre en inglés: Signal to Noise Ratio) SNR , que denota la razón que existe entre la varianza de la señal y la varianza del ruido,

$$SNR = \frac{\sigma_{señal}^2}{\sigma_{ruido}^2}. \quad (2.6)$$

De aquí debemos de hacer las siguientes observaciones:

1. Un SNR **alto** ($\gg 1$) indica una medición de alta precisión.
2. Un SNR **bajo** indica una medición pobre con mucho ruido.

Regresemos al ejemplo del resorte: sabemos que cada cámara grabará cierta dispersión de datos con un patrón lineal debido a que el resorte mismo se mueve (idealmente) en un solo eje. Mostrado en la figura 2.3.

Podemos observar como los datos tienden a ir en una recta; no obstante, hay ciertos datos que se *salen* de este movimiento lineal. Esto es el

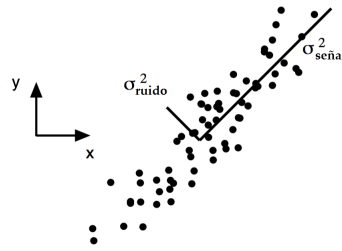


FIGURA 2.3: Ambos ejes: el de mayor varianza (el de la señal) y el eje perteneciente al ruido [1].

ruido, la varianza está señalada en ambas direcciones: la del ruido y la de la señal misma. La razón entre estas dos varianzas nos dice *que tan delgada es la nube de datos* o bien que tanta dispersión existe entre los datos agrupados. Esto engloba posibilidades como una línea delgada ($SNR \gg 1$), un círculo ($SNR = 1$) y otras. Si colocamos *buenas mediciones* (en el sentido de que sabemos que no poseen mucho ruido), cuantitativamente asumimos que las direcciones con las varianzas más grandes en nuestro espacio de mediciones son las que poseen la dinámica de interés. En la figura anterior la dirección con la mayor varianza no son los vectores canónicos de la cámara A : $\{\hat{x}_A = (1, 0), \hat{y}_A = (0, 1)\}$ sino la dirección a lo largo del eje mayor de la nube. En consecuencia, bajo la suposición de que la dinámica buscada se encuentra en aquellos ejes que contengan la mayor varianza y presumiblemente los de mayor SNR , son los ejes principales o bien los de interés.

Esta suposición sugiere que la base que buscamos *no es cualquier base*, no es la base canónica asociada a cada cámara ya que éstas no corresponden a las direcciones con la mayor varianza. Maximizando la varianza corresponde a encontrar la **rotación** adecuada de la base canónica. Cabe destacar que esto no sería necesario si pudieramos alinear desde un principio las cámaras; sin embargo, bajo el criterio de que es un experimento del cual no sabemos nada más allá de lo que observamos/medimos, se vuelve algo complicado. Así, rotando la base canónica en la dirección de la dinámica de la señal (la recta que se ajusta a lo largo de $\sigma_{señal}^2$ revelaría el movimiento del resorte; no obstante, ¿cómo generalizamos esta noción a un número arbitrario de dimensiones? Antes de poder dar respuesta a esta pregunta se necesita examinar el problema desde otra perspectiva.

2.4. Redundancia

En la figura 2.3 se puede notar un factor que puede llevar a malentender los datos, esto es la **redundancia**. El conflicto es particularmente evidente en el ejemplo del resorte ya que tenemos varias cámaras grabando *la misma dinámica* o vaya la misma información. Reexaminando esta figura podemos preguntarnos si realmente es necesario grabar 2 variables. La siguiente figura puede reflejar un rango de posibles gráficas hechas a base de mediciones arbitrarias r_1 y r_2 .

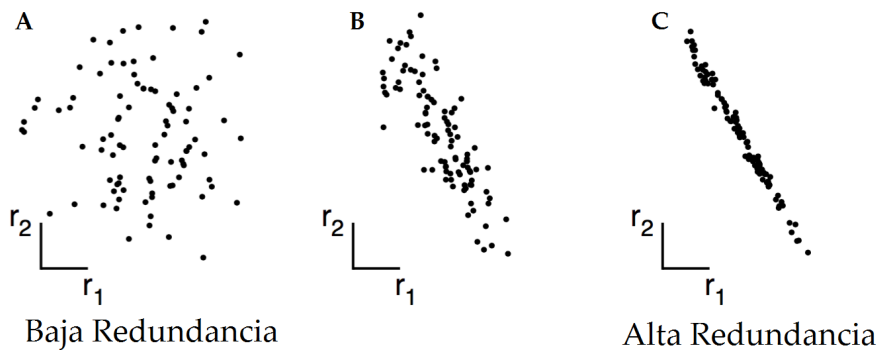


FIGURA 2.4: Se puede apreciar que hay casos extremos. Mucha redundancia intuye cierta predicción con solo tener una medición, sin embargo poca redundancia muestra mediciones pobres que no poseen una aparente relación.

En la figura 2.4A se muestran datos que *aparentemente* no están correlacionados ya que uno no puede predecir r_1 de r_2 . En el otro extremo, en la figura 2.4C se tienen datos altamente correlacionados y por último la figura 2.4B muestra un caso intermedio. Correlacionar los datos puede ser logrado de varias maneras:

- Una gráfica de (x_A, x_B) si las cámaras A y B están muy cerca una de otra.
- Una gráfica de (x_A, \tilde{x}_A) donde x_A está medido en metros y \tilde{x}_A está en centímetros o pulgadas.

Claramente la gráfica de la derecha sería más significativa si solamente midiéramos una variable (x_A ó \tilde{x}_A). El por qué es debido a que se puede calcular r_1 de r_2 o viceversa usando el mejor ajuste a los datos. Midiendo solamente una variable expresaría los datos más consistentemente y reduciría la cantidad de mediciones por hacer (en vez de usar 2

cámaras, se usaría solamente 1). Esta es la idea detrás de la reducción dimensional, poder tener un criterio para poder enfocarse en dinámicas apropiadas sin tener que hacer extensas mediciones.

2.5. La Matriz de Covarianzas

La matriz de covarianzas es una generalización de la varianza (o covarianza) a múltiples dimensiones. En su diagonal se tienen las *varianzas* sencillas, mientras que fuera de ésta se tienen las covarianzas respecto a los otros elementos.

En el caso bidimensional es sencillo identificar la redundancia ya que se puede encontrar el mejor ajuste lineal y juzgar la calidad del ajuste. Esta noción es particularmente útil; sin embargo, ¿cómo cuantificamos y generalizamos esta noción para dimensiones arbitrarias? Consideremos 2 conjuntos de mediciones con *media cero*

$$A = \{a_1, a_2, \dots, a_n\} \quad , \quad B = \{b_1, b_2, \dots, b_n\},$$

donde el subíndice denota el número de muestra. La varianza de A y B están individualmente definidos como,

$$\sigma_A^2 = \frac{1}{n} \sum_i a_i^2 \quad , \quad \sigma_B^2 = \frac{1}{n} \sum_i b_i^2.$$

La *covarianza* de A y B es una generalización inmediata

$$\text{covarianza de } A \text{ y } B \equiv \sigma_{AB}^2 = \frac{1}{n} \sum_i a_i b_i.$$

La covarianza mide el grado de relación lineal entre dos variables. La magnitud absoluta de la covarianza mide el grado de *redundancia*. Si A y B están **no** correlacionados (por ejemplo: la figura 2.4, panel de la izquierda) la covarianza $\sigma_{AB} = 0$. También es obvio que si $A = B \Rightarrow \sigma_{AB} = \sigma_{AA} \equiv \sigma_A$. Podemos equivalentemente convertir A y B en correspondientes vectores renglón:

$$\mathbf{a} = [a_1, a_2, \dots, a_n]$$

$$\mathbf{b} = [b_1, b_2, \dots, b_n]$$

de manera que podemos expresar la covarianza como cálculo del producto punto matricial:

$$\sigma_{\mathbf{ab}}^2 \equiv \frac{1}{n} \mathbf{ab}^T. \quad (2.7)$$

Finalmente, podemos generalizar de **dos** vectores renglón a un número arbitrario de dimensiones superiores. Renombremos los vectores renglón \mathbf{a} y \mathbf{b} como \mathbf{x}_1 y \mathbf{x}_2 , respectivamente. Ahora es sólo necesario considerar vectores renglón adicionales indexados apropiadamente $\mathbf{x}_3, \dots, \mathbf{x}_m$. Definimos una nueva matriz de $m \times n$ dimensiones llamada \mathbf{X}

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_m \end{bmatrix}.$$

Una interpretación de \mathbf{X} es la siguiente: Cada *renglón* de \mathbf{X} corresponde a todas las mediciones de un tipo particular, mientras que cada *columna* corresponde a todas mediciones de **un** solo tipo (ésto es \vec{X} en la sección (2.1). Ahora caemos en la definición de la **matriz de covarianzas**¹ $\mathbf{C}_\mathbf{X}$

$$\mathbf{C}_\mathbf{X} \equiv \frac{1}{n} \mathbf{X}\mathbf{X}^T. \quad (2.8)$$

El ij -ésimo elemento de $\mathbf{C}_\mathbf{X}$ es el producto punto de los vectores: i -ésimo tipo de medición y j -ésimo tipo de medición. Esta matriz posee las siguientes propiedades:

- $\mathbf{C}_\mathbf{X}$ es una matriz simétrica de $m \times m$.
- Los elementos diagonales de $\mathbf{C}_\mathbf{X}$ son la *varianza* de tipos de mediciones particulares.
- Los elementos que no están en la diagonal de $\mathbf{C}_\mathbf{X}$ son la *covarianza* entre tipos de mediciones.

$\mathbf{C}_\mathbf{X}$ captura la covarianza entre todos los posibles pares de mediciones. Los valores de ésta reflejan el ruido y la redundancia en las mediciones hechas. Podemos contruir la idea de lo que buscamos dentro de la matriz de covarianzas como sigue:

- En los elementos de la diagonal; por suposición, los valores grandes son de estructura importante, ya que representan datos (o mediciones) altamente correlacionadas.

¹En la práctica, la covarianza σ_{AB}^2 se calcula como $\frac{1}{n-1} \sum_i a_i b_i$. El ligero cambio en la constante de normalización surge de la teoría de la estimación, pero eso está más allá del enfoque tomado en este trabajo.

- En los elementos fuera de la diagonal, magnitudes grandes corresponden a ruido o alta redundancia.

Con estas ideas en mente, ¿es posible manipular \mathbf{C}_X ? Esta pregunta es de particular interés. Pretendamos por un momento que podemos manipular \mathbf{C}_X , ¿qué tipo de características serían las de apropiado interés para la mejor descripción del experimento? A esta nueva matriz le llamaremos (sugestivamente) \mathbf{C}_Y .

2.5.1. Diagonalización de la Matriz de Covarianzas

Podemos resumir nuestros objetivos como (1) minimizar la redundancia de datos (medida por la magnitud de la covarianza) y (2) maximizar la señal (medida por la varianza). ¿Cómo quedaría la matriz de covarianzas \mathbf{C}_Y bajo este enfoque?

- Todos los elementos fuera de la diagonal de \mathbf{C}_Y deben ser **cero**. Así que \mathbf{C}_Y sería una matriz *diagonal*. En otras palabras, \mathbf{C}_Y está *decorrelacionada*².
- Cada dimensión sucesiva de \mathbf{Y} debe estar ordenada de acuerdo con la varianza (de mayor a menor varianza).

Hay muchas maneras de diagonalizar a \mathbf{C}_Y . De manera simple, el método de ACP discutiblemente usa la manera más sencilla: el ACP asume que todos los vectores base $\{\mathbf{p}_1, \dots, \mathbf{p}_m\}$ son ortonormales, por ejemplo \mathbf{P} es una *matriz ortonormal*. ¿Por qué esta suposición es la más sencilla? Examinemos la figura 2.3: \mathbf{P} actúa como una **rotación** generalizada para alinear una base a lo largo del eje de máxima varianza. En dimensiones múltiples esto puede ser logrado con el siguiente algoritmo:

1. Seleccionamos una dirección normalizada en un espacio m -dimensional a lo largo de la cual la varianza de \mathbf{X} es maximizada. Guardamos este vector como \mathbf{p}_1 .
2. Buscamos otra dirección a lo largo de la cual la varianza de \mathbf{X} es maximizada; sin embargo, debido a la ortonormalidad de estos

²La *decorrelación* es un término general para cualquier proceso que reduce ruido o datos redundantes de un conjunto de señales y preserva los aspectos originales de la señal.

vectores, limitamos nuestra búsqueda a todos aquellos vectores ortogonales a los vectores previamente guardados. Guardamos este nuevo vector como \mathbf{p}_i .

3. Repetimos el procedimiento hasta tener m vectores seleccionados.

El conjunto resultante de \mathbf{p} 's son las **componentes principales**.

En principio este algoritmo funciona pero se debe reconocer que sucede debido a la suposición de ortonormalidad planteada. Suponer *ortonormalidad* tiene como beneficio verdadero el planteamiento de *soluciones analíticas* al problema. Hasta este punto, bajo la estipulación del orden según la varianza (mayor a menor), hemos desarrollado un método para juzgar las direcciones principales; esto es, las varianzas asociadas con cada dirección \mathbf{p}_i cuantifican que tan *principal* es esa dirección según el orden acordado con sus correspondientes varianzas. Podemos aterrizar la idea en los siguientes tres puntos:

I **Linealidad.**

La linealidad encuadra el problema como un cambio de base.

II **Varianzas grandes indican estructura importante.**

Esta suposición abarca conjuntos de datos con un alto SNR , por consiguiente, las componentes principales con una varianza alta indican estructuras importantes, mientras que aquellas componentes con una menor varianza corresponden al ruido y otros factores.

III **Las Componentes Principales son ortogonales.**

Suponer ortogonalidad entre las componentes principales hace que el problema general tenga solución con técnicas de descomposición provenientes del Álgebra Lineal.

Hasta ahora se ha descrito los aspectos que conforman el método de ACP; sin embargo, es necesario solucionar el problema con Álgebra Lineal, en particular, usando la Descomposición en Vectores Propios.

2.5.2. Solución del ACP con Descomposición en Vectores Propios

Los **Vectores Propios** o *Eigenectores*³ son aquellos que **no** cambian su dirección cuando una transformación lineal es aplicada a ellos. Esto puede ser escrito (de forma general) como:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \quad (2.9)$$

donde \mathbf{A} es la matriz de transformación, λ es el *valor propio* y \mathbf{v} es el *vector propio*. Esta ecuación tiene una similitud a la ecuación 2.2 y a partir de aquí construimos la primera solución algebraica para el ACP basada en una propiedad importante de la descomposición en vectores propios. El conjunto de datos es \mathbf{X} , una matriz de $m \times n$, donde m es el número de *tipos de medición* y n es el número de *muestras* de cada tipo de medición. El objetivo se resume como sigue: *Encontrar una matriz ortonormal \mathbf{P} en la ecuación $\mathbf{Y} = \mathbf{P}\mathbf{X}$ tal que $\mathbf{C}_\mathbf{Y} \equiv \frac{1}{n}\mathbf{Y}\mathbf{Y}^T$ es una matriz diagonal. Los renglones de \mathbf{P} son las componentes principales de \mathbf{X} . Se comienza por reescribir $\mathbf{C}_\mathbf{Y}$ en terminos de la variable desconocida.*

$$\begin{aligned} \mathbf{C}_\mathbf{Y} &= \frac{1}{n}\mathbf{Y}\mathbf{Y}^T \\ &= \frac{1}{n}(\mathbf{P}\mathbf{X})(\mathbf{P}\mathbf{X})^T \\ &= \frac{1}{n}\mathbf{P}\mathbf{X}\mathbf{X}^T\mathbf{P}^T \\ &= \mathbf{P}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^T\right)\mathbf{P}^T \\ \mathbf{C}_\mathbf{Y} &= \mathbf{P}\mathbf{C}_\mathbf{X}\mathbf{P}^T. \end{aligned} \quad (2.10)$$

Se identifica la matriz de covarianzas $\mathbf{C}_\mathbf{X}$.

El propósito es reconocer cualquier matriz \mathbf{A} es diagonalizada por una matriz ortonormal de sus vectores propios (Teoremas 3 y 4 en Apéndice A). Para una matriz simétrica, el Teorema 4 proporciona $\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^T$, donde \mathbf{D} es una matriz diagonal y \mathbf{E} es la matriz de los vectores propios de \mathbf{A} en un arreglo de columnas. Seleccionamos a la matriz \mathbf{P} en la ecuación 2.10, donde cada renglón \mathbf{p}_i es un vector propio

³También se les llaman *autovectores*, *vectores característicos*, entre otros.

de \mathbf{C}_X . Esto es de otra forma equivalente con

$$\mathbf{P} \equiv \mathbf{E}^T. \quad (2.11)$$

Con esta relación y el Teorema 1 podemos terminar por evaluar a \mathbf{C}_Y :

$$\begin{aligned} \mathbf{C}_Y &= \mathbf{P}\mathbf{C}_X\mathbf{P}^T \\ &= \mathbf{P}(\mathbf{E}^T\mathbf{D}\mathbf{E})\mathbf{P}^T \\ &= \mathbf{P}(\mathbf{P}^T\mathbf{D}\mathbf{P})\mathbf{P}^T \\ &= (\mathbf{P}\mathbf{P}^T)\mathbf{D}(\mathbf{P}\mathbf{P}^T) \\ \mathbf{C}_Y &= \mathbf{D}. \end{aligned} \quad (2.12)$$

Es evidente que la elección tomada en la ecuación 2.11 diagonaliza a \mathbf{C}_Y . Esto engloba el objetivo del ACP; podemos resumir todos los resultados del método en las matrices \mathbf{P} y \mathbf{C}_Y .

- Las componentes principales de \mathbf{X} son vectores propios de $\mathbf{C}_X = \frac{1}{n}\mathbf{X}\mathbf{X}^T$.
- El i -ésimo valor de \mathbf{C}_Y es la varianza de \mathbf{X} a lo largo del renglón \mathbf{P}_i .

En la práctica, realizar el método de ACP conlleva un análisis de datos previo; como puede ser, substraer el promedio de cada tipo de medición y calcular los vectores propios de \mathbf{C}_X . Esto se discutirá en el siguiente capítulo.

Capítulo 3

Resultados y Discusión

El propósito principal de este capítulo es mostrar el tratamiento de datos y las interpretaciones de los resultados obtenidos. Los espectros Raman llevan consigo imperfecciones que pueden provenir del mismo arreglo experimental con el que se trabaja y/o con la muestra en sí que se analiza. Como se mencionó en el capítulo 2.5.1,II: las estructuras que lleven varianzas grandes son las de particular interés ya que indican patrones que son similares a lo largo de muchos espectros en conjunción con un SNR alto, lo cual indica que poca parte de la señal Raman es debida al ruido.

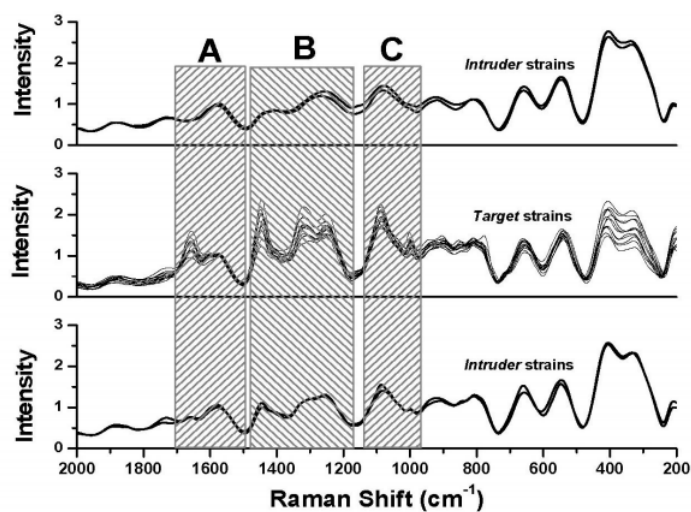


FIGURA 3.1: Identificación de bandas en espectros Raman [4].

Para el tratamiento de datos (espectros Raman) de este trabajo se uso el software comercial Eigenvector Solo+MIA [14]. Este software permite trabajar grandes conjuntos de datos e integra subrutinas en MATLAB [1]. La sangre usada para estos experimentos provino de voluntarios saludables, la cual luego fue irradiada con radiación gamma

(γ) usando una fuente de ^{60}Co (Nordion, Gammacell 220 Excel) con dosis de hasta 100 Gy a un ritmo de 0.5 Gy/min . Los espectros Raman de los glóbulos rojos se adquirieron con un microespectrómetro Raman (Horiba, LabRAM HR) en muestras preparadas como sigue: Se colocaron 10 ml de glóbulos rojos sobre un sustrato de silicio¹ revestido con una película delgada de poli-L-lisina y se sellaron con un cubreobjetos y esmalte de uñas. La excitación Raman fue proporcionada por la línea de 488 nm de un láser de iones de argón a potencia por debajo de 10 mW. Tanto la radiación incidente como la retro-dispersa fueron enfocadas y recogidas a través del mismo objetivo de 50 \times . Los experimentos Raman sobre GR se llevaron a cabo con 20s de tiempo de acumulación. Esto permitió obtener espectros Raman de buena calidad sin dañar la integridad de los GR. A estos espectros Raman se les sustrajo una línea de fondo.

3.1. Tratamiento de Datos

El conjunto de espectros Raman crudos (como los mostrados en la figura 3.3) llevan consigo ruido y diversos aspectos no deseables como son los spikes, discutido en las secciones 2.3 y 2.4. Es de interés corregir el conjunto de espectros mediante el uso de un baseline de cierto orden y una normalización al máximo.

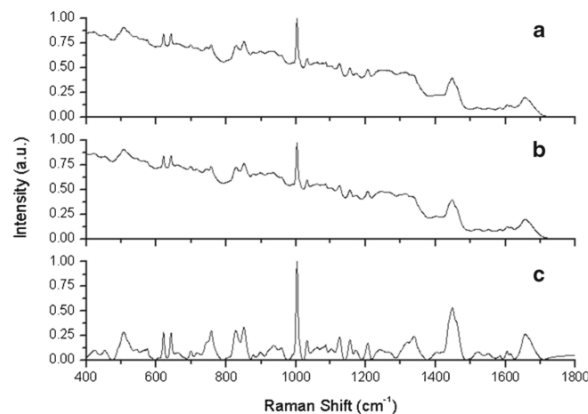


FIGURA 3.2:

a. Datos crudos. **b.** Datos filtrados. **c.** Datos corregidos con un baseline de cierto orden [13].

¹Debido al arreglo experimental, la respuesta Raman del silicio estará presente en todos los espectros Raman de GR.

En la figura 3.2a se muestran los datos originales, sin tratamiento alguno; en contraste con este trabajo, estos datos serían aquellos recolectados por el equipo LabRAM HR. En la figura 3.2b ya se han filtrado una cierta cantidad de datos en conjunción con la filtración de picos indeseados y para la parte 3.2c se tienen corregidos mediante un baseline de cierto orden. En el tratamiento de datos realizado en este trabajo se usó un **baseline de orden 12** para luego llevar una normalización al máximo absoluto.

Los espectros Raman de GR sin radiación sin tratamiento alguno salvo la sustracción de la línea de fondo se muestran en la siguiente figura:

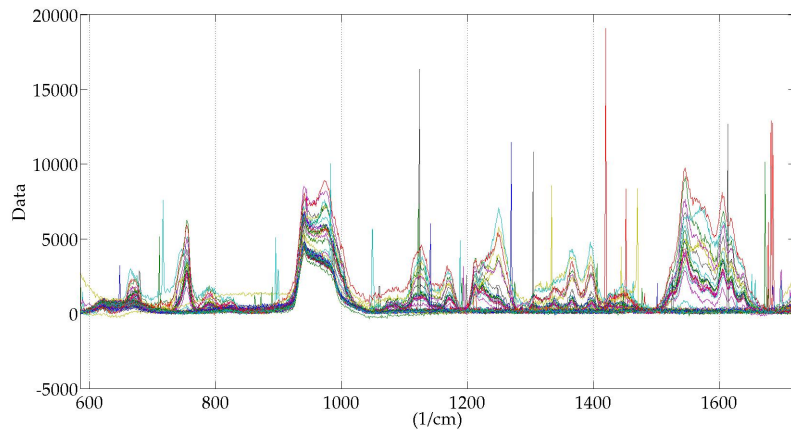


FIGURA 3.3: Espectros Raman crudos de GR sin radiación.

Se observa que hay *spikes*² y bandas no muy bien definidas, seña de ruido en la señal Raman. Al realizar los pasos mostrados en la figura 3.2 al conjunto de espectros se obtienen los espectros filtrados y normalizados. En la figura 3.4 se muestran los espectros *limpios* y normalizados a la unidad. Se pueden apreciar los cambios de pico a pico y los spikes han sido removidos. Este mismo método para tratar los datos es usado con todos los espectros Raman de este trabajo.

²Spikes: picos de alta intensidad que generalmente no son debidos a la muestra.

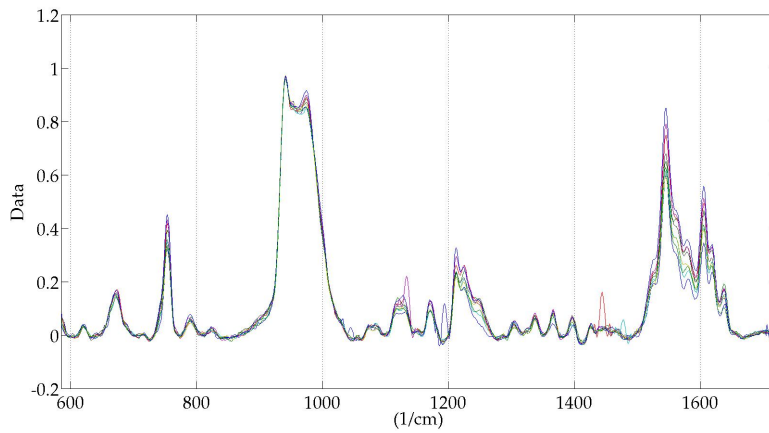


FIGURA 3.4: Tratamiento de datos de los espectros crudos de la figura 3.3.

La necesidad de tener una muestra de *control* surge como una manera de poder diferenciar cambios gráficamente en el espectro Raman de una muestra de sangre irradiada respecto a la no irradiada, este espectro de *control* se obtiene a partir de realizar el tratamiento de datos a la muestra de glóbulos rojos sin radiación.

3.2. Análisis de Componentes Principales

El ACP propone el replanteamiento de una base que mejor exprese el conjunto de espectros Raman con los que se comienza. El método despliega las componentes principales en forma de espectros Raman que son curvas que presumen una mejor representación de los espectros y reducen la cantidad de dimensiones. Las componentes principales sirven como auxiliar para dilucidar la estructura que tiene un conjunto de espectros.

3.2.1. Muestra de Control

La muestra de control es aquella cuya dosis de radiación es 0 Gy; en otras palabras, es sangre intacta. La sangre tiene diversos picos donde hay interacciones de interés [5,10]; sin embargo, el nivel de oxigenación de la sangre es uno de los factores de mayor impacto, ya que influye en el tiempo de almacenamiento y la viabilidad de linfocitos [3,31], como se mencionó en la sección 1.2. La gráfica siguiente muestra los espectros

Raman de control con el tratamiento de datos adecuado:

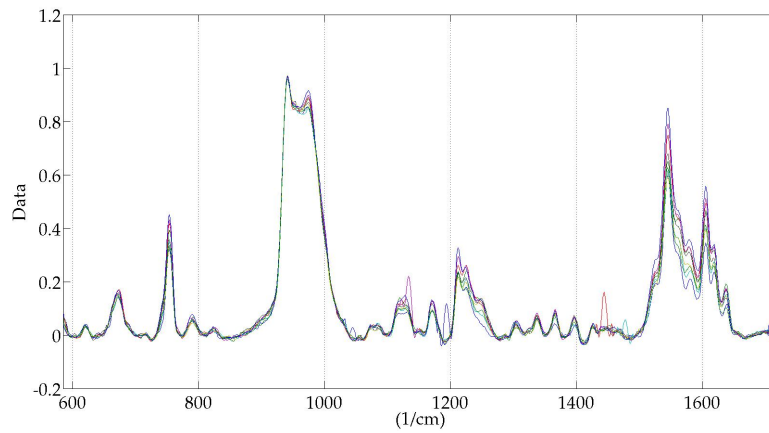


FIGURA 3.5: Muestra de control: Espectros Raman con una dosis de 0 Gy con tratamiento de datos.

Al analizar los espectros con el método de ACP se obtiene solamente una componente principal que se lleva el 99.47% de la varianza del conjunto de espectros. Esto es de cierta manera intuitivo debido a que todos estos espectros Raman tienen en común provenir de sangre *sana*. La única componente principal asociada a la muestra de control es la siguiente:

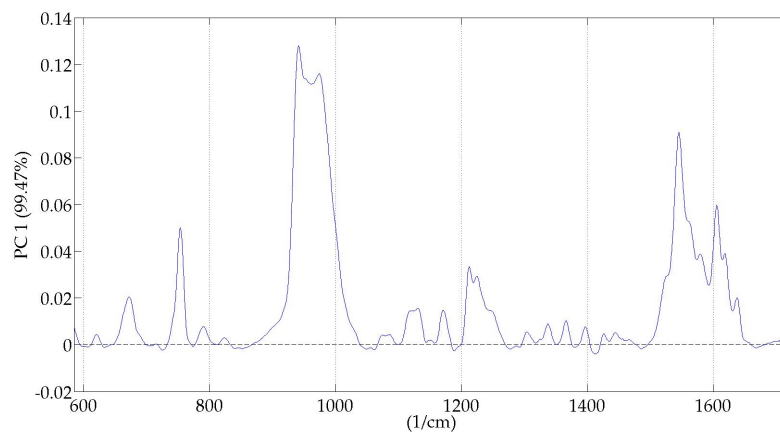


FIGURA 3.6: Componente Principal de la muestra de control.

Esto puede ser puesto en la perspectiva de la sección 2.4, donde vemos que casi todos los espectros tomados para el análisis son similares y por ende redundantes.

3.2.2. Muestra Irradiada a 30 Gy

El conjunto de espectros Raman obtenidos de la muestra irradiada con 30 Gy muestra distinciones ligeras respecto a la de control. Los espectros para la sangre irradiada a 30 Gy son los siguientes:

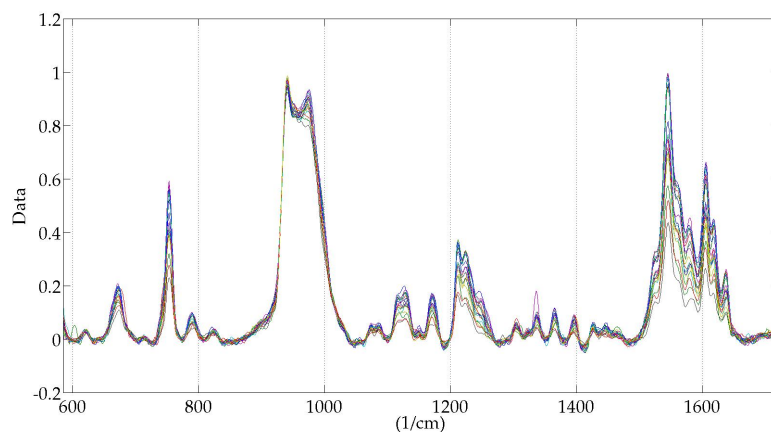


FIGURA 3.7: Espectros Raman con tratamiento de datos para la muestra con 30 Gy de dosis.

Se puede observar que la radiación genera un cambio en la señal Raman del experimento; es decir, se observan distintas vibraciones que la sangre por sí sola no manifiesta [12]. En otras palabras, los estados vibracionales de glóbulos rojos sin radiación diferirán de aquellos con radiación; en consecuencia, se vuelven objeto de análisis; sin embargo, éste no es inmediato al comparar las gráficas. Al realizar el ACP se obtienen dos componentes principales que constan del 88.32 % y 8.70 %; dando un gran total del 97.02 % entre ambas. Las componentes principales son las siguientes:

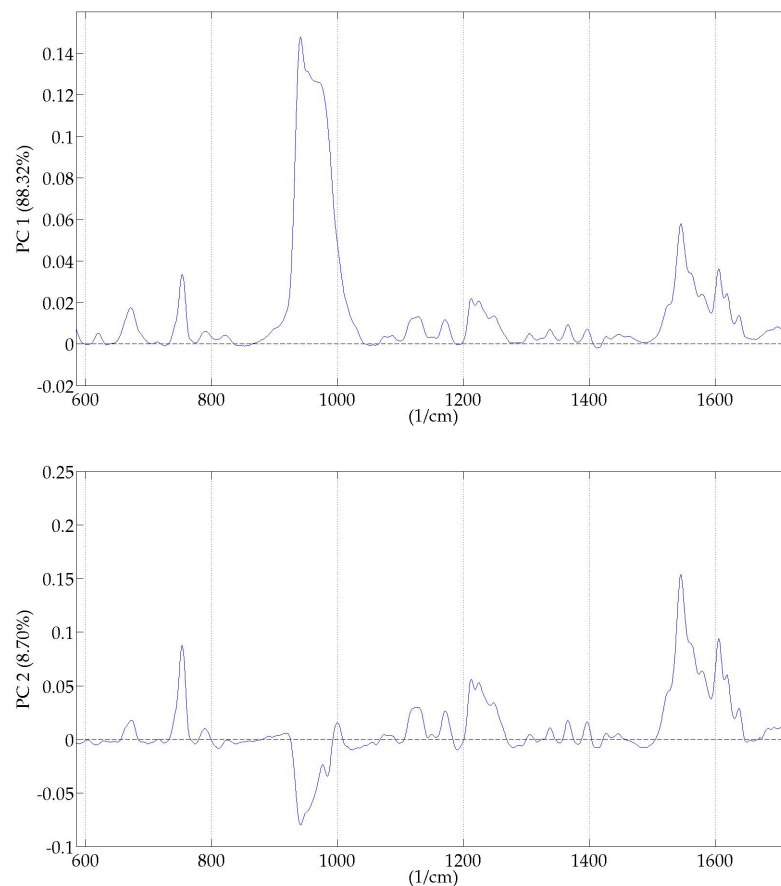


FIGURA 3.8: Componentes Principales: CP1 y CP2.

Al analizar las gráficas de la figura 3.8 se puede observar la componente principal adicional al caso de la sección 3.2.1; donde, la componente principal que se lleva la máxima varianza en este caso tiene la forma de un espectro Raman de sangre.

Como se ve en la figura 3.7 y la figura de control 3.6, se puede apreciar ciertas similitudes y diferencias en la banda de $1100\text{-}1400\text{ cm}^{-1}$, hay una reducción de picos en la señal Raman; sin embargo, en la banda comprendida entre los $950\text{-}1020\text{ cm}^{-1}$ no parece haber un cambio aparente. Esto es consecuencia de la radiación que ha llegado a los glóbulos rojos, la radiación puede dañar de diversas maneras a éstos; por ejemplo, la radiación inhibe la proliferación de linfocitos (alrededor de 25 Gy) y genera especies reactivas del oxígeno [37,38,39]. Mayores dosis pueden llegar a generar efectos dañinos y deterioros que llevan a menores tiempos de vida, por ende, menores tiempos de almacenamiento.

Estos efectos son debidos a la acción indirecta de la radiólisis³ en los glóbulos.

3.2.3. Muestra Irradiada a 100 Gy

Un caso extremo de dosis para la muestra es de 100 Gy; se ha mencionado que la radiación por sí sola genera especies reactivas de la molécula de oxígeno y de hidrógeno. Los espectros Raman obtenidos para las muestras con una dosis de radiación de 100 Gy son las siguientes:

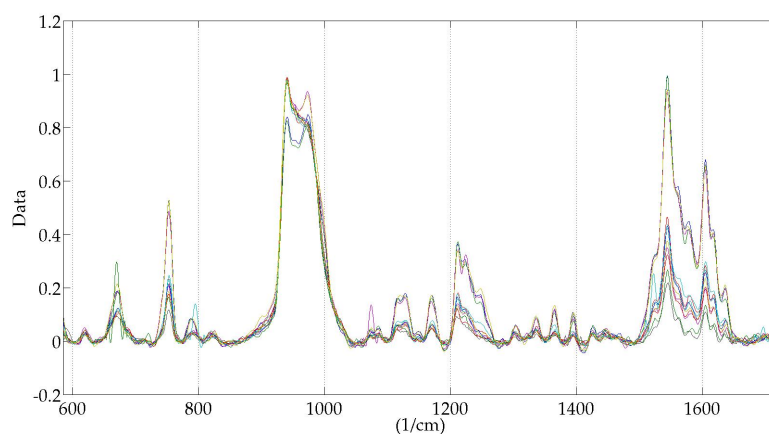


FIGURA 3.9: Espectros Raman con tratamiento de datos para glóbulos rojos con 100 Gy de dosis de radiación.

El ACP para este conjunto de datos da como resultado dos componentes principales que juntas consisten del 95.88 % de la varianza del conjunto de datos, la primera componente para este caso consiste en el 86.91 % de la varianza y la segunda componente principal consta del 8.97 %. Hasta este momento, se tiene una tendencia, entre mayor dosis de radiación pareciera que la segunda componente comienza a crecer significativamente; tomando en consideración que de la muestra de control sólo se desprendió una componente principal. Las componentes principales para esta muestra irradiada a 100 Gy son:

Estos espectros descritos por las componentes principales son espectros teóricos que logran reducir el tamaño de datos; en otras palabras,

³Radiólisis: descomposición molecular de una sustancia debido a la radiación ionizante.

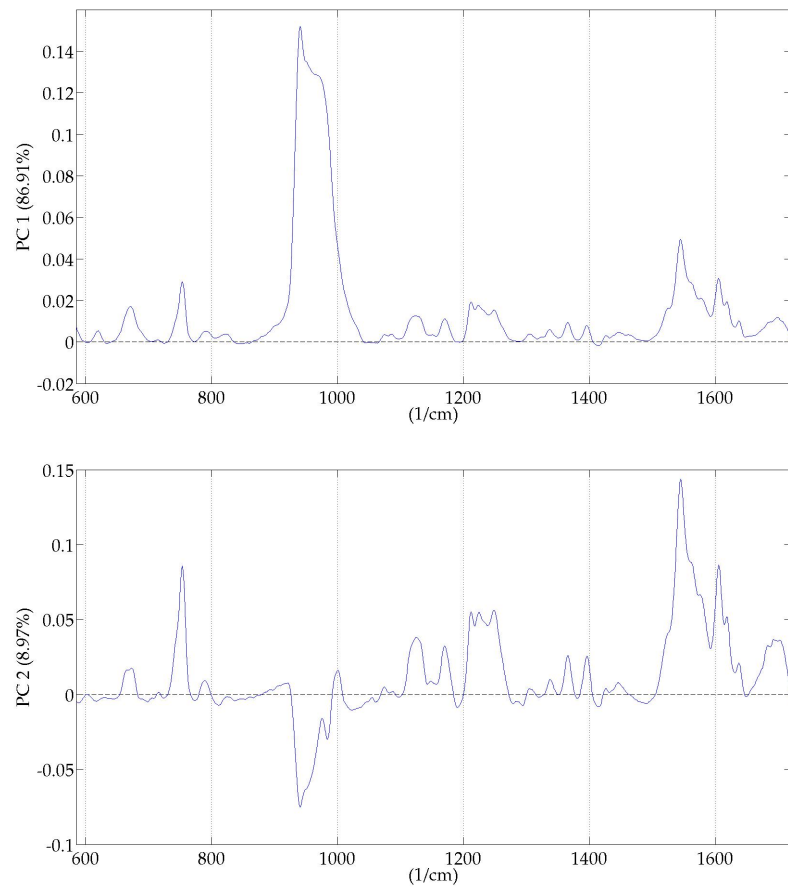
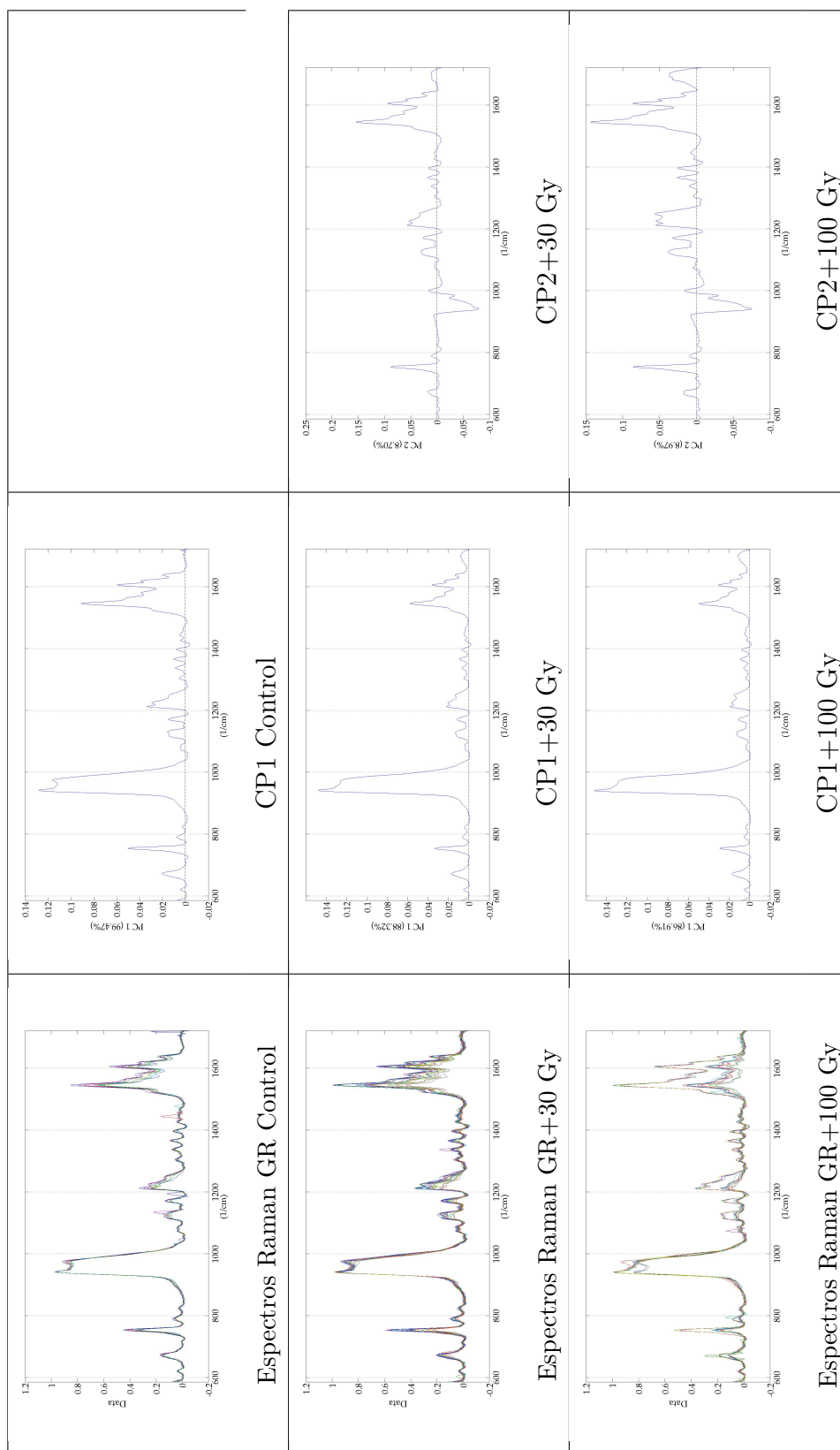


FIGURA 3.10: Primera y segunda componente principal para una muestra de glóbulos rojos con una dosis de 100 Gy de radiación.

redimensionaliza la cantidad de espectros Raman para describir la sangre y otras propiedades.

3.3. Comparación de espectros Raman y Componentes Principales



La comparación desplegada en el arreglo de figuras anterior pretende mostrar que existen cambios en la señal Raman debido a la radiación. Debido a la complejidad sangre se tienen diversos picos de intensidades, los más prominentes son descritos en la siguiente figura:

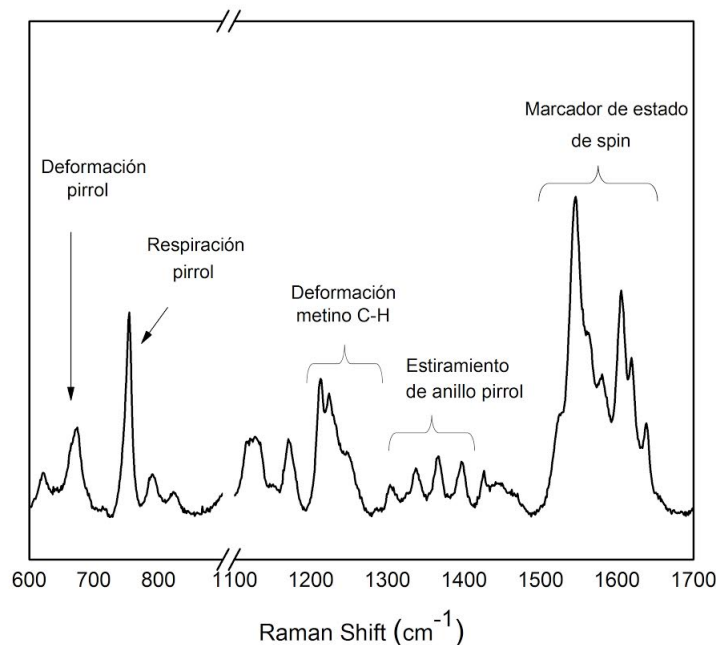


FIGURA 3.11: Picos observados en la señal Raman de una muestra de sangre [8].

Cada pico corresponde a un modo vibracional específico a alguna cadena presente en la sangre, los niveles de oxigenación se encuentran alrededor de los 1600 cm^{-1} , seguido de otro pico de desoxigenación. En las muestras irradiadas claramente se ve una disminución en la señal Raman en esta banda, lo que implica un nivel de oxigenación menor lo cual conlleva a un menor tiempo de almacenamiento.

Las componentes principales atribuidas a los espectros irradiados a 100 Gy muestran una señal muy débil alrededor de los 1300 cm^{-1} a 1400 cm^{-1} , también alrededor de 740 cm^{-1} se logra apreciar una atenuación en el pico en ese lugar. Este tipo de comportamiento se mantiene a través de todo el Raman y es consistente incluso con la muestra irradiada a 30 Gy ; se mantiene la atenuación de modos vibracionales que antes si estaban presentes.

Capítulo 4

Conclusiones

En este trabajo se observó que se pueden diferenciar muestras de sangre irradiadas respecto a una que no lo está utilizando la espectroscopía Raman y el análisis de componentes principales.

Se logró redimensionar el conjunto de espectros Raman, en el caso de la muestra de control bastó con una componente para describir cerca del 100% de los espectros originales. En las muestras irradiadas se observaron dos componentes principales: estas componentes mostraron cambios respecto a la de control y resultaron tener una estructura similar que se diferenciaba por tener una atenuación de picos. Un pico de interés particular es el correspondiente al estado de oxigenación de la sangre; mediante la comparación de las gráficas de componentes principales, se observa que la muestra irradiada con 100 Gy presenta los niveles más bajos de oxigenación. En general, el método de ACP logra desplegar la información relevante de un conjunto de datos y permite discriminar entre diversos tipos de muestras.

Bibliografía

- [1] Shlens, J. (2005). A Tutorial on Principal Component Analysis. *Measurement*, 51, 52. <https://doi.org/10.1.1.115.3503>
- [2] Raman, C. V., Krishnan, K. S. (1928). A New Type of Secondary Radiation. *Nature*, 121(3048), 501–502. <https://doi.org/10.1038/121501e0>
- [3] Santacruz-Gomez, K., Silva-Campa, E., Melendrez-Amavizca, R., Teran Arce, F., Mata-Haro, V., Landon, P. B., Lal, R. (2016). Carboxylated nanodiamonds inhibit γ -irradiation damage of human red blood cells. *Nanoscale*, 8(13), 7189–7196. <https://doi.org/10.1039/C5NR06789H>
- [4] Mobili, P., Londero, A., De Antoni, G., Gómez-Zavaglia, A., Araujo-Andrade, C., Ávila-Donoso, H., Frausto-Reyes, C. (2010). Multivariate analysis of Raman spectra applied to microbiology: Discrimination of microorganisms at the species level. *Revista Mexicana de Fisica*, 56(5), 378–385. [https://doi.org/10/2010;56\(5\):378-385](https://doi.org/10/2010;56(5):378-385)
- [5] Muro, C. K., Lednev, I. K. (2016). Identification of individual red blood cells by Raman microspectroscopy for forensic purposes: in search of a limit of detection. *Analytical and Bioanalytical Chemistry*, 1–7. <https://doi.org/10.1007/s00216-016-0002-2>
- [6] Beirana, A., Alcalá, D., Franco, A., (2000). Enfermedad injerto contra huésped. *Rev Cent Dermatol Pascua*, 9(2), 74–77. Retrieved from <http://www.medigraphic.com/pdfs/derma/cd-2000/cd002b.pdf>
- [7] Michael F. L'Annunziata (2003). *Handbook of Radioactivity Analysis*, Academic Press.
- [8] Luca, A.C., Pesce, G., Rusciano, G., Sasso, A. (2008). Raman Tweezers as a Diagnostic Tool of Hemoglobin-Related Blood Disorders. *Sensors*.

- [9] Bumbrah, G. S., Sharma, R. M. (2016). Raman spectroscopy – Basic principle, instrumentation and selected applications for the characterization of drugs of abuse. *Egyptian Journal of Forensic Sciences*, 6(3), 209–215. <https://doi.org/10.1016/j.ejfs.2015.06.001>
- [10] Bankapur, A., Zachariah, E., Chidangil, S., Valiathan, M., Mathur, D. (2010). Raman tweezers spectroscopy of live, single red and white blood cells. *PLoS ONE*, 5(4). <https://doi.org/10.1371/journal.pone.0010427>
- [11] P.J. de Groot, G.J. Postma, W.J. Melssen, L.M.C. Buydens, V. Deckert, R. Zenobi (2001). Application of principal component analysis to detect outliers and spectral deviations in near-field surface-enhanced Raman spectra.
- [12] Premasiri, W. R., Lee, J. C., Ziegler, L. D. (2012). Surface-enhanced Raman scattering of whole human blood, blood plasma, and red blood cells: cellular processes and bioanalytical sensing. *The Journal of Physical Chemistry. B*, 116(31), 9376–86. <https://doi.org/10.1021/jp304932g>
- [13] González-Solís, J. L., Martínez-Espinosa, J. C., Salgado-Román, J. M., Palomares-Anda, P. (2014). Monitoring of chemotherapy leukemia treatment using Raman spectroscopy and principal component analysis. *Lasers in Medical Science*, 29(3), 1241–1249. <https://doi.org/10.1007/s10103-013-1515-y>
- [14] Eigenvector Solo+MIA.
<http://www.eigenvector.com/software/solo+mia.htm>
- [15] Pieters, S., Vander Heyden, Y., Roger, J. M., D’Hondt, M., Hansen, L., Palagos, B., De Beer, T. (2013). Raman spectroscopy and multivariate analysis for the rapid discrimination between native-like and non-native states in freeze-dried protein formulations. *European Journal of Pharmaceutics and Biopharmaceutics*, 85(2), 263–271. <https://doi.org/10.1016/j.ejpb.2013.03.035>
- [16] Das, D.K., et al., (2013). Modulatory role of quercetin against gamma radiation-mediated biochemical and morphological alterations of red blood cells. *Int J Radiat Biol*. 89(6): p. 471-81.

- [17] Feng, S., Chen, R., Lin, J., Pan, J., Chen, G., Li, Y., Zeng Haishan, H. (2010). Nasopharyngeal cancer detection based on blood plasma surface-enhanced Raman spectroscopy and multivariate analysis. *Biosensors and Bioelectronics*, 25(11), 2414–2419. <https://doi.org/10.1016/j.bios.2010.03.033>
- [18] Huang, Z., Lin, J., Cao, G., Chen, X., Li, Y., Feng, S., Chen, R. (2014). Raman spectroscopy combined with multivariate analysis techniques as a potential tool for semen investigation, 9230(Pibm), 92301X. <https://doi.org/10.1117/12.2068836>
- [19] Chen, L., Wang, Y., Liu, N., Lin, D., Weng, C., Zhang, J., Feng, S. (2013). Near-infrared confocal micro-Raman spectroscopy combined with PCA–LDA multivariate analysis for detection of esophageal cancer. *Laser Physics*, 23(6), 65601. <https://doi.org/10.1088/1054-660X/23/6/065601>
- [20] Hollenhorst, J.. Agilent 101: An Introduction to Bio-Analytical Measurement. Agilent Technologies, Agilent Labs, 10. http://www.agilent.com/labs/features/2011_101_bio.html
- [21] Buckley, K., Kerns, J. G., Parker, A. W., Goodship, A. E., Matousek, P. (2014). Decomposition of in vivo spatially offset Raman spectroscopy data using multivariate analysis techniques. *Journal of Raman Spectroscopy*, 45(2), 188–192. <https://doi.org/10.1002/jrs.4434>
- [22] Shafer-Peltier, K. E., Haka, A. S., Fitzmaurice, M., Crowe, J., Myles, J., Dasari, R. R., Feld, M. S. (2002). Raman microspectroscopic model of human breast tissue: Implications for breast cancer diagnosis in vivo. *Journal of Raman Spectroscopy*, 33(7), 552–563. <https://doi.org/10.1002/jrs.877>
- [23] Hensley, K., et al., (2000). Reactive oxygen species, cell signaling, and cell injury. *Free Radical Biology and Medicine*. 28(10): p. 1456-1462.
- [24] Crawford, A., Silva, E., York, K., Li, C. (n.d.). Raman Spectroscopy: A Comprehensive Review. www.Academia.edu.

- [25] Pichardo-Molina, J. L., Frausto-Reyes, C., Barbosa-García, O., Huerta-Franco, R., González-Trujillo, J. L., Ramírez-Alvarado, C. A., Medina-Gutiérrez, C. (2007). Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients. *Lasers in Medical Science*, 22(4), 229–236. <https://doi.org/10.1007/s10103-006-0432-8>
- [26] Solé, J., L. Bausa, and D. Jaque, (2005). An introduction to the optical spectroscopy of inorganic solids.
- [27] Kilic, S.S., S. Kavurt, and S.B. Adim, (2010). Transfusion-Associated Graft-Versus-Host Disease in Severe Combined Immunodeficiency. *Journal of Investigational Allergology and Clinical Immunology*.
- [28] Long, D.A., (1997). Raman spectroscopy. New York, p. 1-12.
- [29] Sikirzhyski, V., Virkler, K., Lednev, I. K. (2010). Discriminant analysis of Raman spectra for body fluid identification for forensic purposes. *Sensors*, 10(4), 2869–2884. <https://doi.org/10.3390/s100402869>
- [30] Bergholt, M. S., Duraipandian, S., Zheng, W., Huang, Z. (2013). Multivariate reference technique for quantitative analysis of fiberoptic tissue raman spectroscopy. *Analytical Chemistry*, 85(23), 11297–11303. <https://doi.org/10.1021/ac402059v>
- [31] Martínez Espinosa, J. C., González Solís, J. L., Frausto Reyes, C., Miranda Beltrán, M. L., Soria Fregoso, C., Medina Valtierra, J. (2009). Detection of Leukemia with Blood Samples Using Raman Spectroscopy and Multivariate Analysis. *AIP Conference Proceedings*, 99(2009), 99–104. <https://doi.org/10.1063/1.3175637>
- [32] Haydock, R. (2015). Multivariate Analysis of Raman Spectroscopy Data.
- [33] Gonzalez-Solis, J. L., Rodriguez-Lopez, J., Martinez-Espinosa, J. C., Frausto-Reyes, C., Jave-Suarez, L. F., Aguilar-Lemarroy, A. C., Longo, L. (2010). Detection of Cervical Cancer Analyzing Blood Samples with Raman Spectroscopy and Multivariate Analysis. *AIP Conference Proceedings*, 91(2010), 91–95. <https://doi.org/10.1063/1.3453792>

-
- [34] Notingher, I., Jell, G., Notingher, P. L., Bisson, I., Tsigkou, O., Polak, J. M., Hench, L. L. (2005). Multivariate analysis of Raman spectra for in vitro non-invasive studies of living cells. *Journal of Molecular Structure*, 744–747(SPEC. ISS.), 179–185. <https://doi.org/10.1016/j.molstruc.2004.12.046>
- [35] Madden, M. G., Ryder, A. G. (n.d.). *Machine Learning Methods for Quantitative Analysis of Raman Spectroscopy Data*.
- [36] Module, M. A. (n.d.). *LabSpec 6 Multivariate Analysis Module*. Ilin, Y., Kraft, M. L. (2014). Identifying the lineages of individual cells in cocultures by multivariate analysis of Raman spectra. *The Analyst*, 139(9), 2177–85. <https://doi.org/10.1039/c3an02156d>
- [37] Sanford, C. L., Mantooth, B. a, Jones, B. T. (2001). Determination of Ethanol in Alcohol Samples Using a Modular Raman Spectrometer. *Journal of Chemical Education*, 78(9), 1221. <https://doi.org/10.1021/ed078p1221>
- [38] Numata, Y., Iida, Y., Tanaka, H. (2011). Quantitative analysis of alcohol-water binary solutions using Raman spectroscopy. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 112(6), 1043–1049. <https://doi.org/10.1016/j.jqsrt.2011.01.005>
- [39] Giguère, P. A., Chen, H., Giguere, P. A. (1984). Hydrogen bonding in hydrogen peroxide and water. A Raman study of the liquid state. *Journal of Raman Spectroscopy*, 15(3), 199–204. <https://doi.org/10.1002/jrs.1250150313>

Apéndice A

Apéndice

Teorema 1 (La inversa de una matriz ortogonal es su transpuesta.). *Sea \mathbf{A} una matriz ortogonal de $m \times n$ donde \mathbf{a}_i es el i -ésimo vector columna. El ij -ésimo elemento de $\mathbf{A}^T \mathbf{A}$ es*

$$(\mathbf{A}^T \mathbf{A})_{ij} = \mathbf{a}_i^T \mathbf{a}_j = \begin{cases} 1 & \text{si } i=j \\ 0 & \text{otro caso.} \end{cases}$$

Entonces, como $\mathbf{A}^T \mathbf{A} = \mathbf{I}$, se sigue que $\mathbf{A}^{-1} = \mathbf{A}^T$.

Teorema 2 (Para cualquier matriz, $\mathbf{A}^T \mathbf{A}$ y $\mathbf{A} \mathbf{A}^T$ son simétricas). *Por definición de simetría, una matriz es simétrica si al transponerla obtenemos esta misma matriz. Comenzamos calculando la transpuesta de los productos matriciales $\mathbf{A}^T \mathbf{A}^T$ y $\mathbf{A} \mathbf{A}$:*

$$\begin{aligned} (\mathbf{A} \mathbf{A}^T)^T &= \mathbf{A}^{TT} \mathbf{A} = \mathbf{A} \mathbf{A}^T \\ (\mathbf{A}^T \mathbf{A})^T &= \mathbf{A}^T \mathbf{A}^{TT} = \mathbf{A}^T \mathbf{A}. \end{aligned}$$

Transponer el producto matricial $\mathbf{A}^T \mathbf{A}$ ó $\mathbf{A} \mathbf{A}^T$ da como resultado el mismo producto matricial original, entonces $\mathbf{A}^T \mathbf{A}$ y $\mathbf{A} \mathbf{A}^T$ son productos matriciales simétricos.

Teorema 3 (Una matriz \mathbf{A} es simétrica si es ortogonalmente diagonalizable.). *Que la matriz \mathbf{A} sea ortogonalmente diagonalizable significa que existe una matriz \mathbf{E} tal que podemos reescribir la matriz $\mathbf{A} = \mathbf{E} \mathbf{D} \mathbf{E}^T$ donde la matriz \mathbf{D} es diagonal (sus elementos son $d_{ij} = 0$ si $i \neq j \forall i, j \in \{1, 2, \dots, n\}$). La matriz \mathbf{E} es una matriz que diagonaliza a la matriz \mathbf{A} . Calculando \mathbf{A}^T :*

$$\mathbf{A}^T = (\mathbf{E} \mathbf{D} \mathbf{E}^T)^T = \mathbf{E}^{TT} \mathbf{D}^T \mathbf{E}^T = \mathbf{E} \mathbf{D} \mathbf{E}^T = \mathbf{A}.$$

Evidentemente, si \mathbf{A} es ortogonalmente diagonalizable, entonces también es simétrica.

Teorema 4 (Una matriz simétrica es diagonalizada por una matriz de sus vectores propios ortonormales). Sea \mathbf{A} de dimensiones $n \times n$, simétrica y asociada a sus vectores propios $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$. Sea ahora la matriz $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n]$ donde la i -ésima columna corresponde al i -ésimo vector propio \mathbf{e}_i . Este teorema afirma que existe una matriz diagonal \mathbf{D} tal que $\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^T$. Para esto comenzamos calculando los productos matriciales $\mathbf{A}\mathbf{E}$ y $\mathbf{E}\mathbf{D}$:

$$\begin{aligned}\mathbf{A}\mathbf{E} &= [\mathbf{A}\mathbf{e}_1, \mathbf{A}\mathbf{e}_2, \dots, \mathbf{A}\mathbf{e}_n], \\ \mathbf{E}\mathbf{D} &= [\lambda_1\mathbf{e}_1, \lambda_2\mathbf{e}_2, \dots, \lambda_n\mathbf{e}_n].\end{aligned}$$

Evidentemente, la transformación resultante del producto $\mathbf{A}\mathbf{e}_i$ es $\lambda_i\mathbf{e}_i$. Esto anterior es la definición de la ecuación de valores propios. Entonces se cumple que $\mathbf{A}\mathbf{E} = \mathbf{E}\mathbf{D}$. Un reacomodo proporciona la siguiente definición:

$$\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^{-1}. \quad (\text{A.1})$$

Ahora es necesario mostrar que si una matriz es simétrica siempre tendrá vectores ortogonales. Esto puede desarrollarse tomando en cuenta los vectores propios \mathbf{e}_i y \mathbf{e}_j como sigue,

$$\begin{aligned}\lambda_i\mathbf{e}_i \cdot \mathbf{e}_j &= (\lambda_i\mathbf{e}_i)^T \mathbf{e}_j \\ &= (\mathbf{A}\mathbf{e}_i)^T \mathbf{e}_j \\ &= \mathbf{e}_i^T \mathbf{A}^T \mathbf{e}_j \\ &= \mathbf{e}_i^T \mathbf{A} \mathbf{e}_j \\ &= \mathbf{e}_i^T (\lambda_j\mathbf{e}_j) \\ \lambda_i\mathbf{e}_i \cdot \mathbf{e}_j &= \lambda_j\mathbf{e}_i \cdot \mathbf{e}_j.\end{aligned} \quad (\text{A.2})$$

Esta última relación es equivalente a $(\lambda_i - \lambda_j)\mathbf{e}_i \cdot \mathbf{e}_j = 0$. Desde un principio se ha trabajado con la idea de que todos los vectores propios son únicos, por lo tanto, debe ser el caso de que $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ (ortogonalidad). Los vectores propios de \mathbf{A} son todos ortonormales (por elección). Esto tiene como consecuencia que la matriz \mathbf{E} sea ortonormal y por el Teorema 1, $\mathbf{E}^{-1} = \mathbf{E}^T$ y se puede reescribir la

ecuación A.1 como sigue

$$\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^T. \quad (\text{A.3})$$

Así, una matriz simétrica es diagonalizada por una matriz de sus vectores propios.

Teorema 5 (Para una matriz \mathbf{X} arbitraria de dimensiones $m \times n$, la matriz simétrica $\mathbf{X}^T\mathbf{X}$ tiene un conjunto de vectores propios ortonormales $\{\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \dots, \hat{\mathbf{v}}_n\}$ y un conjunto de valores propios asociados $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$. El conjunto de vectores $\{\mathbf{X}\hat{\mathbf{v}}_1, \mathbf{X}\hat{\mathbf{v}}_2, \dots, \mathbf{X}\hat{\mathbf{v}}_n\}$ forma una base ortogonal, donde cada vector $\mathbf{X}\hat{\mathbf{v}}_i$ tiene norma de $\sqrt{\lambda_i}$). Estas propiedades pueden obtenerse de la siguiente expresión:

$$\begin{aligned} (\mathbf{X}\hat{\mathbf{v}}_i) \cdot (\mathbf{X}\hat{\mathbf{v}}_j) &= (\mathbf{X}\hat{\mathbf{v}}_i)^T (\mathbf{X}\hat{\mathbf{v}}_j) \\ &= \hat{\mathbf{v}}_i^T \mathbf{X}^T \mathbf{X} \hat{\mathbf{v}}_j \\ &= \hat{\mathbf{v}}_i^T (\lambda_j \hat{\mathbf{v}}_j) \\ &= \lambda_j \hat{\mathbf{v}}_i \cdot \hat{\mathbf{v}}_j \\ (\mathbf{X}\hat{\mathbf{v}}_i) \cdot (\mathbf{X}\hat{\mathbf{v}}_j) &= \lambda_j \delta_{ij}. \end{aligned} \quad (\text{A.4})$$

La ecuación A.4 muestra una relación similar a la ecuación del Teorema 1; muestra que el conjunto de vectores propios de \mathbf{X} es ortogonal, resultando en una delta de Kronecker. Esto puede reescribirse como:

$$(\mathbf{X}\hat{\mathbf{v}}_i) \cdot (\mathbf{X}\hat{\mathbf{v}}_j) = \begin{cases} \lambda_j & \text{si } i=j \\ 0 & \text{otro caso.} \end{cases}$$

Esta ecuación establece que que dos vectores cualesquiera del conjunto son ortogonales. Una segunda propiedad surge como caso particular de la ecuación anterior, esto es:

$$\|\mathbf{X}\hat{\mathbf{v}}_i\|^2 = (\mathbf{X}\hat{\mathbf{v}}_i) \cdot (\mathbf{X}\hat{\mathbf{v}}_i) = \lambda_i. \quad (\text{A.5})$$