

Tecnológico de Costa Rica
Escuela de Ingeniería Electrónica



**Optimización de la detección de cantidad, ancho y posición de
bandas en imágenes de geles de electroforesis**

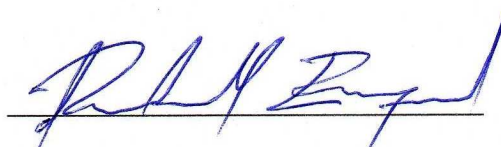
Informe de Proyecto de Graduación para optar por el título de
Ingeniero en Electrónica con el grado académico de Licenciatura

Randall José Esquivel Alvarado

Cartago, 20 de junio, 2011

Declaro que el presente Proyecto de Graduación ha sido realizado enteramente por mi persona, utilizando y aplicando literatura referente al tema e introduciendo conocimientos propios.

En los casos en que he utilizado bibliografía he procedido a indicar las fuentes mediante las respectivas citas bibliográficas. En consecuencia, asumo la responsabilidad total por el trabajo de graduación realizado y por el contenido del correspondiente informe final.



Randall José Esquivel Alvarado


Cartago, 20 de junio de 2011

Céd: 5-0357-0111

Instituto Tecnológico de Costa Rica
Escuela de Ingeniería Electrónica
Proyecto de Graduación
Tribunal Evaluador

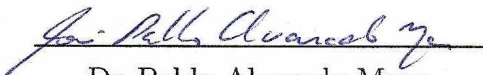
Proyecto de Graduación defendido ante el presente Tribunal Evaluador como requisito para optar por el título de Ingeniero en Electrónica con el grado académico de Licenciatura, del Instituto Tecnológico de Costa Rica.

Miembros del Tribunal


M.Sc. Eduardo Interiano Salguero
Profesor Lector




M.Sc. Arys Carrasquilla Batista
Profesora Lectora


Dr. Pablo Alvarado Moya
Profesor Asesor

Los miembros de este Tribunal dan fe de que el presente trabajo de graduación ha sido aprobado y cumple con las normas establecidas por la Escuela de Ingeniería Electrónica.

Cartago, 20 de junio de 2011

Resumen

En este documento se propone una mejora al proceso de detección de bandas en geles de electroforesis, utilizadas para la caracterización de patrones de ADN a nivel molecular.

La caracterización de bandas en geles de electroforesis puede presentar el inconveniente de que la resolución y contraste de la imagen no sean adecuados para la determinación de la posición y el ancho de las bandas, por lo que una herramienta computacional constituye una ayuda al profesional analista de los datos. El enfoque de este trabajo es la caracterización de las bandas ubicadas en carriles de electroforesis, de forma que el proceso pueda realizarse de manera automática, sin necesidad de la intervención del usuario para obtener aproximaciones iniciales.

El algoritmo parte de la imagen de un carril al cual se le realiza una estimación inicial de la desviación estándar utilizando el espacio de escalas y con la segunda derivada se calcula la posición inicial de las bandas en dicho carril. La evaluación iterativa de una función objetivo con estos parámetros iniciales permite mejorar la aproximación de la desviación estándar, con cuyo valor, se itera nuevamente utilizando el optimizador *Downhill Simplex* para mejorar los parámetros restantes, amplitud y posición, para cada banda.

Palabras clave: *Downhill Simplex*, espacio de escalas, función objetivo, minimización.

Abstract

This document proposes an improved band detection process in electrophoresis gel, used to characterize DNA patterns at molecular level.

The characterization of gel electrophoresis bands may have the disadvantage that the resolution and contrast of the image are not suitable for determining the position and width of the bands, so a computational tool is an aid to the professional analyst of that data. The focus of this work is the band characterization located in electrophoresis lanes, so that the process can be performed automatically, without user intervention to obtain initial approximations.

The algorithm starts with the image of a lane and computes an initial estimate of the standard deviation using the scale space and with the second derivative the different band location is calculated in that lane. The iterative evaluation of an objective function with these initial parameters allows the improvement in the approximation of the standard deviation, and with this value, the objective function is iterated again using the Downhill Simplex optimizer to improve the remaining parameters, amplitude and position for each band.

Keywords: *Downhill Simplex* least squares, minimization, objective function, scale space

...a mis queridos padres

Agradecimientos

Gracias a Dios por permitirme haber llegado al final de esta carrera y por darme fuerzas para continuar siempre.

Al Dr. Pablo Alvarado por haberme permitido realizar este trabajo, obtener nuevos conocimientos, así como por su ayuda y orientación durante la realización del mismo.

A mis padres Miguel y Flor, especialmente, por el apoyo incondicional que me han brindado siempre, el cual ha sido fundamental para completar esta etapa.

A mi tía Benilda y mi primo Jorge que me han apoyado en los momentos difíciles.

Al Departamento de Trabajo Social y Salud por la beca Mauricio Campos otorgada para el financiamiento de mis estudios.

Su apoyo se los agradezco profundamente.

Randall José Esquivel Alvarado

Cartago, 20 de Junio de 2011

Índice general

Índice de figuras	iii
Índice de tablas	v
1 Introducción	1
1.1 Electroforesis en Gel	1
1.2 Caracterización de bandas	3
1.3 Estrategia propuesta para la caracterización de bandas	4
1.4 Objetivos y estructura del documento	4
2 Marco Teórico	5
2.1 Forma de las bandas teóricas en carriles de electroforesis	5
2.1.1 Función Gaussiana	5
2.1.2 Función Lorentziana	6
2.1.3 Modelo seleccionado	6
2.2 Espacio de Escalas	7
2.2.1 Espacio de escalas unidimensional	8
2.3 Minimización lineal de funciones	10
2.3.1 Acorralamiento de mínimo utilizando la razón dorada	10
2.3.2 Interpolación parabólica y el método de Brent	11
2.3.3 Método de minimización lineal de funciones o <i>LineSearch</i>	12
2.4 Downhill Simplex	13
2.5 Filtros de Savitzky y Golay	15
3 Sistema para caracterización de bandas	19
3.1 Detección de desviación estándar mediante espacio de escalas	20
3.1.1 Análisis del espacio de escalas	20
3.1.2 Enventanado del espacio de escalas	23
3.1.3 Problemas en la determinación de la desviación estándar	24
3.2 Filtrado para extracción de fondo del carril	26
3.3 Mejora de la resolución utilizando derivadas	26
3.4 Modelado del carril como sistema lineal	28
3.5 Ajuste por minimización de una función objetivo	29
3.6 Ajuste de la posición de las bandas y segmentado del carril en ventanas	29

4	Resultados y Análisis	31
4.1	Espacio de escalas	31
4.1.1	Evaluación de carril con poco traslape entre bandas	33
4.1.2	Evaluación de carriles con traslape	33
4.2	Estimación de la posición y cantidad de bandas	36
4.2.1	Ubicación de bandas con la segunda derivada	36
4.2.2	Determinación de la cantidad de bandas en el carril	39
4.3	Determinación de la desviación estándar con Line Search y espacio de escalas	40
4.4	Mejora en ubicación de la posición utilizando Downhill Simplex	45
4.5	Evaluación de carriles reales	45
4.6	Rendimiento del algoritmo	46
5	Conclusiones	49
5.1	Conclusiones	49
5.2	Recomendaciones	50
	Bibliografía	51

Índice de figuras

1.1	Proceso de electroforesis.	2
1.2	Imagen de un gel de electroforesis.	2
1.3	Diagrama general del sistema.	3
1.4	Diagrama de bloques general del sistema propuesto.	4
2.1	Función gaussiana.	6
2.2	Función lorentziana.	7
2.3	Efecto del incremento de escala en dos dimensiones.	8
2.4	Esquema del espacio de escalas unidimensional.	9
2.5	Espacio de escalas de un impulso gaussiano.	9
2.6	Acorralamiento del mínimo de una función.	10
2.7	Minimización de una función por interpolación parabólica.	12
2.8	Cambios posibles para un simplex inicial.	14
2.9	Efecto de la aplicación de filtros de Savitzky y Golay.	17
3.1	Diagrama completo del sistema propuesto.	20
3.2	Normalización del máximo de la segunda derivada del espacio de escalas.	22
3.3	Espacio de escalas para una sola banda.	22
3.4	Enventanado del espacio de escalas.	23
3.5	Detección de la desviación estándar utilizando espacio de escalas.	24
3.6	Espacio de escalas de dos funciones de igual amplitud.	25
3.7	Efecto de distorsión en el espacio de escalas.	25
3.8	Carril real con fondo.	26
3.9	Derivadas de una función gaussiana.	27
3.10	Derivadas de una sumatoria de gaussianas con traslape severo.	28
3.11	Segmentación de un carril en ventanas.	30
4.1	Error en la determinación de la desviación estándar con espacio de escalas en función de la desviación estándar de las bandas.	33
4.2	Carril sintético con 10 bandas con $\sigma = 5$	33
4.3	Carril sintético de 9 bandas con traslape.	34
4.4	Espacio de escalas con traslape de bandas.	37
4.5	Carril teórico con poco traslape.	38
4.6	Carril teórico con mayor traslape.	38
4.7	Error en la determinación de la desviación estándar en carriles con una banda.	42

4.8	Error en la determinación de la posición utilizando Downhill Simplex.	46
4.9	Proceso de ajuste del carril mediante mínimos cuadrados.	47
4.10	Carril con una única banda utilizado para evaluar rendimiento.	47
4.11	Carril con 10 bandas utilizado para evaluar rendimiento.	48

Índice de tablas

4.1	Estimación de desviación estándar con espacio de escalas para carriles teóricos con una banda	32
4.2	Estimación de desviación estándar con espacio de escalas para un carril con 10 bandas con $\sigma = 5$	34
4.3	Máximos locales en un carril con 9 bandas traslapadas	35
4.4	Evaluación de la desviación estándar utilizando carriles sintéticos con 20 bandas traslapadas. Aproximación realizada con el espacio de escalas.	36
4.5	Posición de las bandas en un carril con poco traslape	37
4.6	Posición de las bandas en un carril sintético con 26 bandas	39
4.7	Evaluación de cantidad de bandas en carriles sintéticos. Cantidad de bandas observadas.	40
4.8	Evaluación de cantidad de bandas en carriles sintéticos. Cantidad de bandas obtenidas mediante la segunda derivada.	40
4.9	Estimación de desviación estándar utilizando la función objetivo para carriles teóricos con una banda	41
4.10	Aproximación de la desviación estándar por ajuste de función objetivo	42
4.11	Detección de bandas y desviación estándar en bandas traslapadas con espacio de escalas	43
4.12	Evaluación de la desviación estándar utilizando carriles sintéticos con 20 bandas.	44
4.13	Evaluación de la desviación estándar utilizando carriles sintéticos con 20 bandas traslapadas. Aproximación del espacio de escalas y minimización lineal.	44
4.14	Posición aproximada de las bandas en un carril sintético con 26 bandas utilizando Downhill Simplex	45
4.15	Evaluación del rendimiento del algoritmo anterior con algoritmos genéticos.	47
4.16	Evaluación del rendimiento del algoritmo propuesto en este trabajo.	48

Capítulo 1

Introducción

1.1 Electroforesis en Gel

Los geles de electroforesis son una herramienta de análisis cuyo principio básico es el movimiento controlado de partículas cargadas bajo la acción de un campo eléctrico [7]. Las moléculas en los ácidos nucleicos por lo general poseen carga negativa, por lo que al aplicar una diferencia de potencial entre los extremos del gel de poliacrilamida las moléculas tratarán de avanzar a través del gel desde el cátodo hasta el ánodo. La velocidad de migración de una molécula v es directamente proporcional al campo eléctrico E , a la carga neta de la molécula z e inversamente proporcional al coeficiente de fricción de f :

$$v = \frac{Ez}{f} \quad (1.1)$$

El coeficiente de fricción f es directamente proporcional a la masa y el tamaño de la molécula, y a la viscosidad del medio, en este caso el gel.

El proceso de electroforesis es realizado generalmente en geles debido a que estos sirven como una rejilla que facilita la separación de las moléculas. Las moléculas que son pequeñas en comparación con los poros del gel se mueven fácilmente a través de éste, mientras que las moléculas grandes tienen su movimiento limitado. La electroforesis se lleva a cabo en placas verticales de gel de poliacrilamida, donde se colocan las moléculas a estudiar en la parte superior, de forma que la dirección de movimiento sea de arriba hacia abajo. Las moléculas pequeñas se mueven rápidamente a través del gel, mientras que las moléculas grandes quedan, en general, cerca del punto de aplicación de la muestra.

En una placa de gel de electroforesis se aplica una serie de muestras que se analizan al mismo tiempo. La Figura 1.1 muestra el proceso de electroforesis descrito anteriormente, donde se aprecia la formación de diferentes acumulaciones de moléculas de similar carga y tamaño debido a las restricciones de movilidad dentro de la estructura del gel.

En la imagen de la Figura 1.2 se muestra el resultado de la electroforesis en un caso real. Los patrones verticales que se observan en la imagen del gel corresponden a los carriles, donde

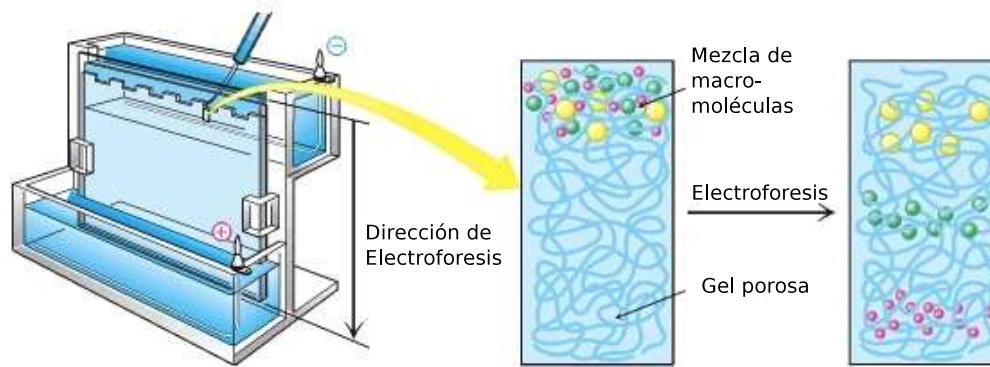


Figura 1.1: Proceso de electroforesis. Tomado de [7]

cada uno de estos representa la distribución de moléculas para una muestra en análisis. Las formaciones horizontales oscuras dentro de un carril reciben el nombre de bandas y corresponden a la acumulación de moléculas mostrada en el esquema de la Figura 1.1.

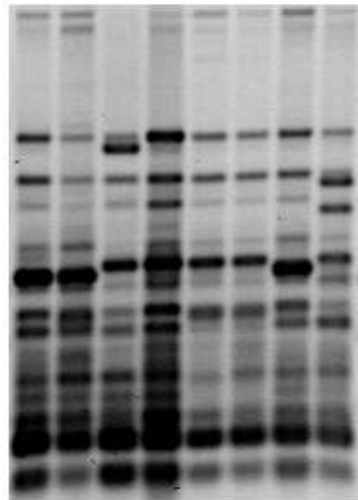


Figura 1.2: Imagen de un gel de electroforesis.

Uno de los usos de los gels de electroforesis es el análisis de compuestos de proteínas, y permite realizar un análisis cualitativo en una imagen por la forma en la que se desplazan las moléculas a través del gel [7].

La ubicación de las bandas dentro del carril se utiliza como una forma de comparar la composición molecular del ADN de dos organismos diferentes para determinar qué tan similares son genéticamente.

El proyecto que pretende obtener un sistema de procesamiento de imágenes de gels de electroforesis se denomina "Análisis por computador de imágenes de gels de electroforesis para la caracterización molecular de organismos" [6]. El desarrollo de este sistema se basa en la necesidad de mejorar la etapa de procesamiento de los datos para obtener mejores resultados en menor tiempo. En el ITCR se está avanzando en el desarrollo de la herramienta computacional que permita realizar este reconocimiento de las bandas presentes en cada muestra de ADN mediante el procesamiento de imágenes.

El diagrama general del proyecto se muestra en la Figura 1.3, con las diferentes etapas correspondientes al procesamiento de imágenes de las gels de electroforesis.

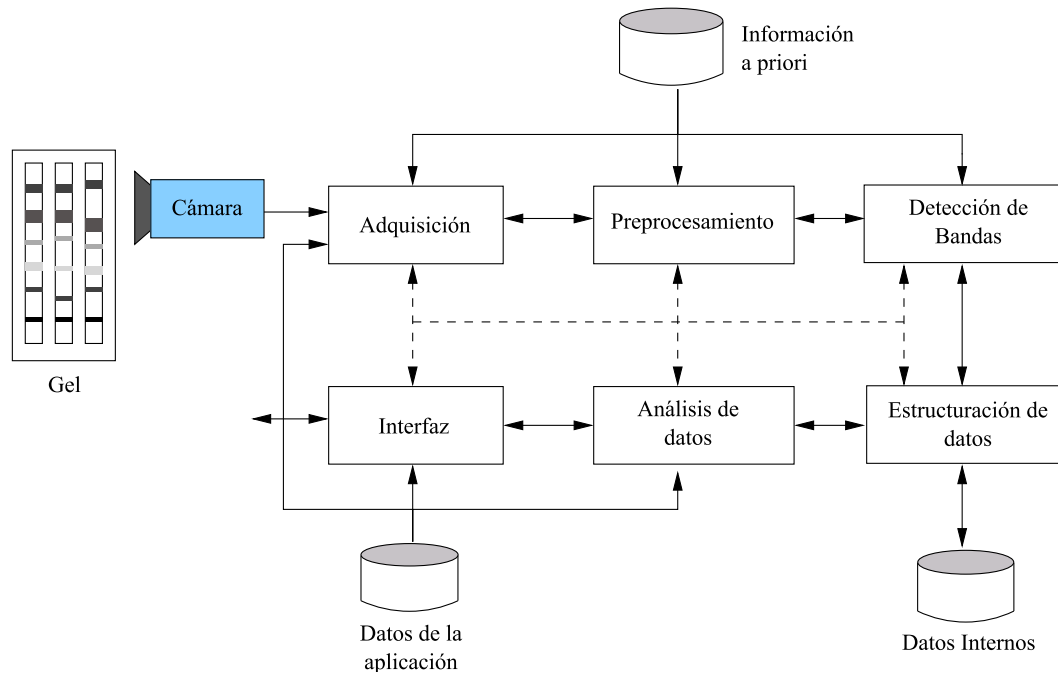


Figura 1.3: Diagrama general del sistema.

En trabajos anteriores se han realizado avances en el área de preprocesamiento, en la cual se realizaron mejoras para la corrección de ruido [2], y en la corrección del efecto sonrisa [1], [4]; así como en el área de detección de bandas [21], [11].

1.2 Caracterización de bandas

En la detección de la posición y ancho de las bandas el primer trabajo realizado [21], consiste en la aproximación del carril mediante una sumatoria de bandas gaussianas, realizada utilizando como algoritmo optimizador el error determinado por la función de mínimos cuadrados, pero este algoritmo requiere como entrada la cantidad de bandas y su posición, por lo que depende de un usuario que estime y proporcione el valor de la desviación estándar de las bandas para realizar el ajuste en menor tiempo y además depende del usuario para estimar la cantidad de bandas en el carril.

En el segundo trabajo realizado [11] se introdujo el uso del espacio de escalas como una solución al problema, considerando que en un mismo proceso podría ser resuelta la estimación de la posición de las bandas, así como la desviación estándar, sin embargo, el algoritmo falla en presencia de bandas traslapadas, pero brinda una aproximación inicial de la desviación estándar, uso que se da a este método en el presente trabajo.

1.3 Estrategia propuesta para la caracterización de bandas

Para complementar los algoritmos desarrollados en trabajos anteriores, se propone en este trabajo el tratamiento de los carriles de geles de electroforesis para mejorar la detección de la cantidad de bandas, antes de aproximar la forma de las mismas.

La solución planteada consiste en encontrar la mayor cantidad de bandas considerando una función matemática que modela las bandas del carril, de forma que si una forma encontrada no se ajusta al modelo considerado pueda ser representado como dos ó mas funciones distintas que sí se ajusten al modelo. Una vez encontradas las diferentes bandas se calcula la posición de las mismas; con éstas posiciones conocidas se calcula el ancho de las bandas que aproxime con el menor error a los datos, como se muestra esquemáticamente en la Figura 1.4.

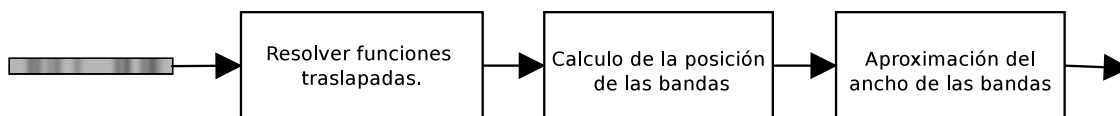


Figura 1.4: Diagrama de bloques general del sistema propuesto.

Este método se evalúa en este trabajo tanto para carriles teóricos como para carriles reales a los cuales se les extrae el fondo con la finalidad de realizar el ajuste de las bandas.

La solución se ha desarrollado programando en C++ sobre plataforma GNU/Linux utilizando la biblioteca de software LTI-Lib2 [8] para procesamiento digital de señales e imágenes.

1.4 Objetivos y estructura del documento

El objetivo de este trabajo consiste en realizar la caracterización unidimensional de las bandas en un carril de una imagen de gel de electroforesis, con lo cual se obtienen las variables amplitud, posición y desviación estándar para cada una de las bandas. Para ello se utiliza un algoritmo iterativo que realiza el ajuste del ancho de las bandas cuando se conoce su posición, hasta que el error sea mínimo, y se realiza una optimización al análisis del espacio de escalas para la ubicación de bandas, por lo tanto se incluye el análisis realizado al espacio de escalas, iniciado en [11]. La aproximación se realiza a su vez para la mejora en la determinación de la posición de las bandas obtenidas con el espacio de escalas.

La estructura de este documento es la siguiente: en el capítulo 2 se incluye la teoría necesaria para la implementación de la solución; el capítulo 3 explica detalladamente la implementación realizada; el capítulo 4 contiene los resultados obtenidos y la comparación entre los dos métodos analizados y el capítulo 5 está dedicado a las conclusiones y recomendaciones con respecto a estos métodos.

Capítulo 2

Marco Teórico

En este capítulo se presentan los conceptos en los que se basa el sistema propuesto para la caracterización de las bandas. A continuación se brinda información acerca de la forma de las bandas en carriles de electroforesis, el concepto de espacio de escalas y su implementación unidimensional para un vector de datos. Se incluye el modelado para el análisis matemático de un carril, así como la minimización de funciones objetivo por mínimos cuadrados para la aproximación iterativa de las bandas.

Además se resume la teoría para el filtrado y diferenciación de señales reales con ruido utilizando el método de Savitzky y Golay.

2.1 Forma de las bandas teóricas en carriles de electroforesis

En esta sección se explica el modelo de las distribuciones utilizada para la representación de las bandas en geles de electroforesis.

2.1.1 Función Gaussiana

La función gaussiana g es una función de área igual a uno, centrada en $x = \mu$. Es una función par con respecto a su centro, y su ancho se encuentra en función de la desviación estándar σ [5].

$$g(x, \sigma, \mu) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \quad (2.1)$$

La gráfica de la Figura 2.2 ilustra la variación de la función gaussiana para $\sigma = 1$, $\sigma = 2$ y $\sigma = 3$.

El ancho de una función gaussiana es infinito, pero tomando en cuenta que para una desvia-

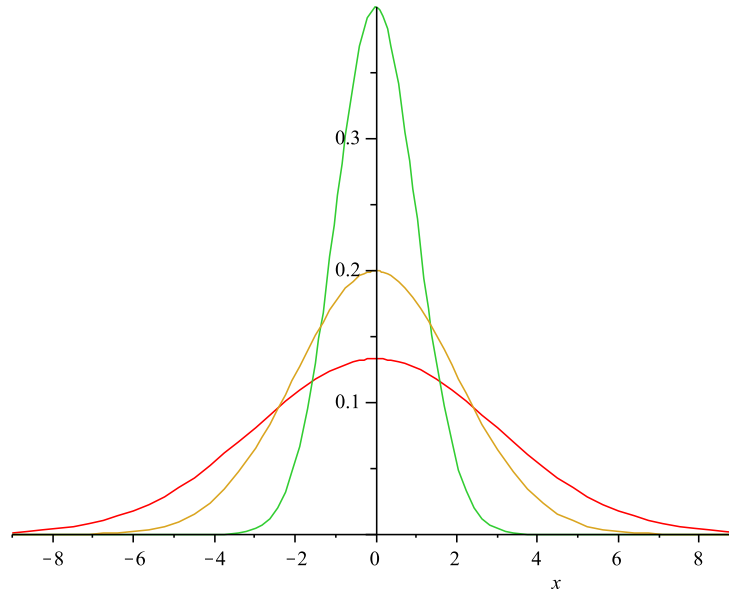


Figura 2.1: Función gaussiana. Función graficada con desviación estándar $\sigma = 1$ (verde), $\sigma = 2$ (amarillo), $\sigma = 3$ (rojo)

ción estándar σ , la integral de la función gaussiana desde -3σ hasta 3σ contiene aproximadamente un 99.73 % de los datos, por lo tanto se considera que el ancho de la función es de 6σ [19].

2.1.2 Función Lorentziana

La función lorentziana (2.2) corresponde a la distribución de probabilidad de Cauchy-Lorentz. Al igual que la función gaussiana se encuentra centrada en $x = \mu$ y su ancho se encuentra determinado por el parámetro Γ .

$$g(x, \Gamma, \mu) = \frac{\left(\frac{\Gamma}{2}\right)}{(x - \mu)^2 + \left(\frac{\Gamma}{2}\right)^2} \quad (2.2)$$

La Figura 2.2 muestra la función lorentziana para $\Gamma = 1$, $\Gamma = 2$ y $\Gamma = 3$.

El parámetro Γ corresponde al ancho de la distribución a la mitad de la altura. Esta distribución tiene la característica de que el 99 % de su área se encuentra contenido en un rango de $\frac{\Gamma}{\tan\left(\frac{\pi}{200}\right)}$ alrededor de su centro, esto es aproximadamente desde $-31,83\Gamma$ hasta $31,83\Gamma$.

2.1.3 Modelo seleccionado

Los estudios realizados en el ajuste de forma de las bandas en geles de electroforesis se basan comúnmente en la distribución gaussiana [3], y esto se justifica en la investigación realizada en [18], la cual tiene como resultado que las bandas en electroforesis capilar en el análisis de

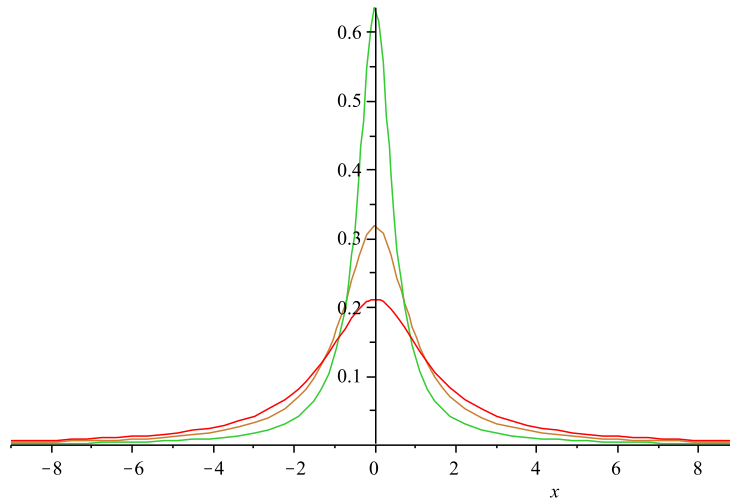


Figura 2.2: Función lorentziana. Función graficada con $\Gamma = 1$ (verde), $\Gamma = 2$ (amarillo), $\Gamma = 3$ (rojo)

ADN se aproximan a una distribución gaussiana, de donde se generaliza al caso de los geles con carriles más anchos.

En [3], se analiza el ajuste de bandas en geles de electroforesis utilizando la función lorentziana y se obtiene que este proporciona una mejor aproximación, partiendo de que las bandas analizadas son más anchas que como propone el modelo gaussiano, ya que este último decae rápidamente.

Sin embargo, según [10] la distribución lorentziana no es utilizada comúnmente para modelado de bandas en cromatografía, esto porque la cantidad de datos que deben tomarse en cuenta es mucho mayor dada la lenta caída de esta función. Además, si se realiza un análisis del espectro de esta distribución, la frecuencia de muestreo necesaria para obtener la información sin distorsión es considerablemente mayor que para el caso de la gaussiana, específicamente para obtener el 99,9% de la información las frecuencias máximas que deben considerarse son $\omega_{max} = 3,29/\sigma$ para el caso de la función gaussiana y $\omega_{max} = 8,67/\sigma$ para el caso del lorentziano, esto en el caso ideal de no tener traslape de bandas.

Por lo anterior se considera en este trabajo el uso de la distribución gaussiana como modelo para las bandas, puesto que implica realizar menos cálculos y considerar menos datos para cada banda, considerando teóricamente la distribución simétrica de estas para el análisis.

2.2 Espacio de Escalas

El espacio de escalas se refiere a una familia de funciones derivadas de una función inicial, a la cual representan en un nivel diferente de escala, donde t es considerado el parámetro de escala. Cada una de estas funciones $L(x; t)$ se obtiene de filtrados a la función original con escala en aumento [17]:

$$L(x; t) = \int_{-\infty}^{\infty} g(\xi, t) f(x - \xi) d\xi \quad (t > 0) \quad (2.3)$$

Cada uno de los niveles de escala resultante está relacionado a una cantidad de información, de forma al aumentar la escala, cada vez hay menos detalles en la imagen.

Según [15], los principios fundamentales del espacio de escalas se basan en que cada nivel de escala debe ser obtenido como resultado de la convolución lineal de un filtro pasabajas de ancho invariante con una imagen original real y que con cada incremento del espacio de escalas deben obtenerse señales con menos estructura.

En el caso de imágenes, por su naturaleza discreta, (2.3) se transforma en una sumatoria de componentes discretos ó píxeles de la imagen

$$L(x; t) = \sum_{n=-\infty}^{\infty} g(n, t) f(x - n) \quad (t > 0) \quad (2.4)$$

En el caso bidimensional, el filtro gaussiano aplicado debe considerar ambas coordenadas en la imagen, por lo que el filtro aplicado es

$$g(x, y, t) = \frac{e^{-\frac{(x)^2+(y)^2}{2t}}}{2\pi t} \quad (2.5)$$

donde $t = \sigma^2$ y σ es la desviación estándar del filtro.

La imagen de la Figura 2.3 muestra un ejemplo de aplicación a una imagen bidimensional, donde se aprecia el efecto de aumentar la escala, según el principio establecido en (2.5).



Figura 2.3: Efecto del incremento de escala en dos dimensiones. Ejemplo de imagen a escala t de 1, 16 y 256 respectivamente. (Tomado de [16]).

2.2.1 Espacio de escalas unidimensional

En el caso de una función unidimensional el espacio de escalas consiste en la realización de un filtrado sucesivo de forma que se aumente la escala aplicada consecutivamente hasta tener una señal cada vez con menos información. El espacio de escalas unidimensional puede considerarse como la aplicación sucesiva de un filtro gaussiano, también unidimensional

$$g(x, y, \sigma) = \frac{e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}}{\sqrt{2\pi}\sigma} \quad (2.6)$$

con desviación estándar σ cada vez mayor, tomando cada vez como base a la función original, a la cual le corresponde la escala cero. Esto se muestra a manera de esquema en la Figura 2.4.

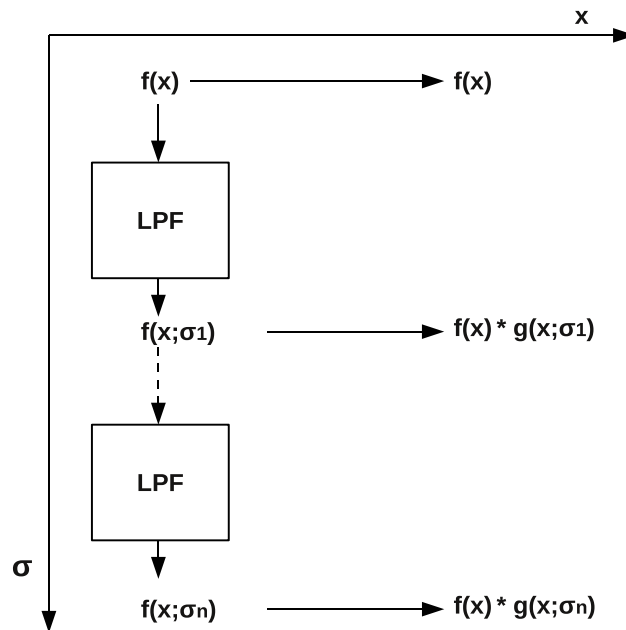


Figura 2.4: Esquema del espacio de escalas unidimensional. (Tomado de [11]).

La aplicación del esquema mostrado en la Figura 2.4 se muestra ejemplificado en la Figura 2.5, donde se genera el espacio de escalas en una dimensión a un impulso gaussiano con desviación estándar $\sigma = 9$.

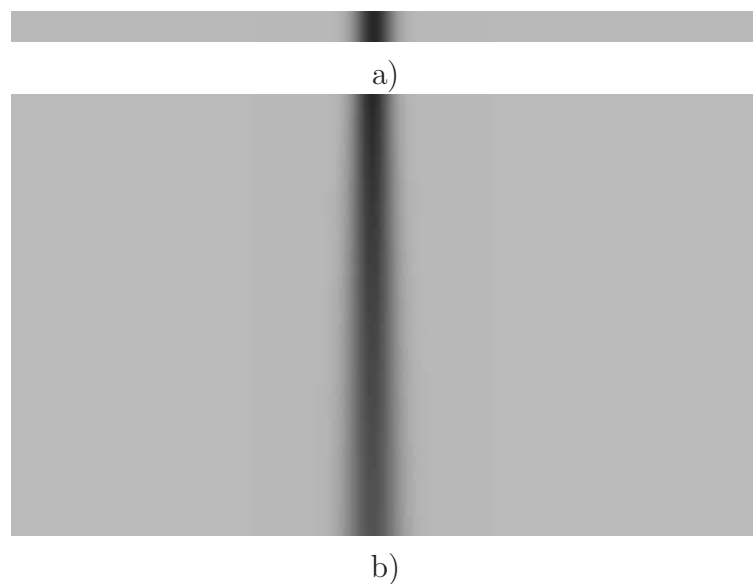


Figura 2.5: Espacio de escalas de un impulso gaussiano. a) Patrón de un impulso gaussiano. b) Espacio de escalas unidimensional aplicado a la fila central de la función en a).

2.3 Minimización lineal de funciones

En el caso de una función de con puntos máximos y mínimos, como la gráfica del error en un proceso, es necesario encontrar el punto en el cual la función alcanza el menor valor. Los algoritmos para la detección de mínimos se basan en poder encontrar estos puntos para una función teniendo en cuenta que debe realizarse con la mayor eficiencia posible.

Dada una función como la mostrada en la Figura 2.6 (ignorando los puntos identificados con números por ahora), la minimización se trata de encontrar el valor de la variable que hace mínimo el valor de la función, encontrarlo de forma rápida y con el menor esfuerzo computacional posible.

2.3.1 Acorralamiento de mínimo utilizando la razón dorada

La búsqueda de un mínimo en una función debe iniciar con la evaluación de dos abscisas, a y c , de forma que permita establecer un rango dependiendo del valor de la función evaluada en dichos puntos, suponiendo que el mínimo se encuentra en alguna parte entre esos dos puntos. Sean $f(a)$ y $f(c)$ los valores de la función evaluada en el intervalo, se buscará un punto b tal que $f(b) < f(a)$ y $f(b) < f(c)$ y se continúa cambiando dichos puntos descendiendo progresivamente al mínimo.

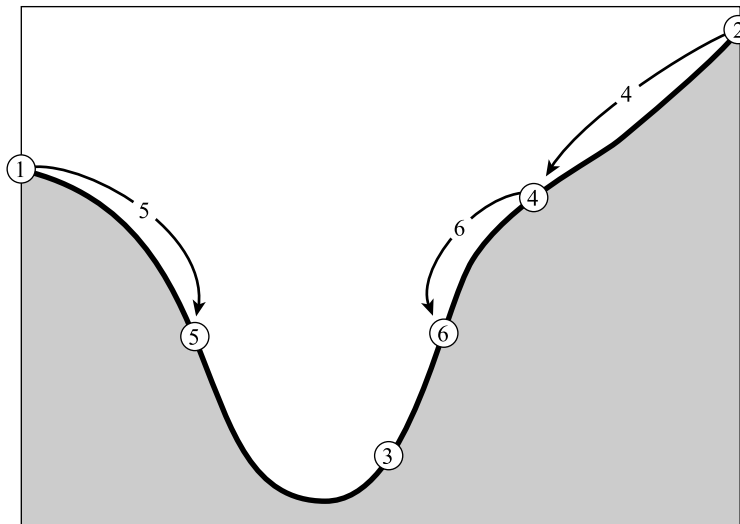


Figura 2.6: Acorralamiento del mínimo de una función. (Tomado de [12]).

La imagen de la Figura 2.6 muestra el acorralamiento sucesivo de una función. En un principio los valores del rango estimado son los puntos 1 y 2, el valor central en 3; con sucesivas evaluaciones el punto 2 es sustituido por el punto 4, el punto 1 por el 5 y por último el punto 4 por el 6.

La eficiencia del algoritmo depende de una estrategia adecuada de descenso dependiendo de los valores de la función. La razón dorada corresponde a un planteamiento de optimización del punto que debe ser evaluado dentro del intervalo entre a y b para la minimización del error. En [12] se detalla la deducción de la razón de oro, la cual se obtiene de

$$w^2 - 3w + 1 = 0 \implies w = \frac{3 - \sqrt{5}}{2} \approx 0,38197 \quad (2.7)$$

Este número representa la fracción de la distancia entre a y b que es óptima para la elección del punto siguiente b , de forma tal que se escoge b a una distancia 0,38197 veces la longitud del intervalo a partir de a , quedando a una distancia de $(1 - 0,38197) \approx 0,61803$ del punto b . Estos dos números son conocidos como la relación dorada.

De esta forma, por ejemplo, para una tripleta (a, b, c) se escoge el punto x entre b y c , manteniendo b como el punto mínimo, se realiza la evaluación de la función y en el caso $f(x) > f(b)$ la nueva tripleta es (a, b, x) y si $f(x) < f(b)$ la nueva tripleta será (b, x, c) .

Este aproximación del mínimo garantiza que el intervalo actual se reduce a uno 0,61803 veces menor en cada iteración.

2.3.2 Interpolación parabólica y el método de Brent

La interpolación parabólica se basa en la búsqueda de mínimos, asumiendo que la función a minimizar tiene una forma aproximadamente parabólica en las cercanías del mínimo.

Dada una tripleta de puntos (a, b, c) con el punto mínimo acotado entre a y b , se calcula la función parabólica que pasa por los tres puntos con la finalidad de encontrar el punto mínimo de la parábola. El punto mínimo de esa parábola se obtiene de

$$x = b - \frac{1}{2} \frac{(b-a)^2[f(b) - f(c)] - (b-c)^2[f(b) - f(a)]}{(b-a)[f(b) - f(c)] - (b-c)[f(b) - f(a)]} \quad (2.8)$$

Así el método consiste en la generación de parábolas que vayan aproximando al mínimo de la función como se muestra en la Figura 2.7.

Sin embargo, el método de interpolación parabólica por sí solo no debe ser utilizado como una estrategia de minimización, puesto que puede converger a mínimos parabólicos como a máximos.

El método de Brent es una optimización para la búsqueda lineal de puntos mínimos en una dimensión, el cual utiliza en forma eficiente el acorralamiento mediante la relación dorada en conjunto con la interpolación parabólica para minimizar la cantidad de evaluaciones de la función considerada.

Para una evaluación específica, este método hace uso de los seis puntos a, b, u, v, w y x , los cuales se definen como: a y b son los puntos iniciales que tienen acorralado el mínimo a encontrar; x , es el punto mínimo de todas las evaluaciones realizadas hasta el momento; w es el punto con la segunda menor evaluación de la función; v es el valor anterior de w y u es el último punto en el que la función fue evaluada.

El primer paso en la evaluación es realizar una interpolación parabólica entre los puntos x, v y w , y ésta para ser aceptable debe generar un mínimo en el intervalo (a, b) y el desplazamiento

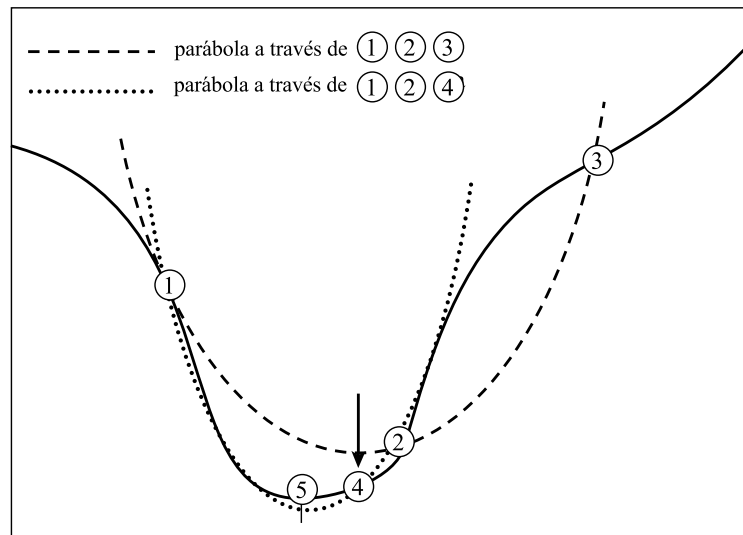


Figura 2.7: Minimización de una función por interpolación parabólica. 1, 2, 3 forman la primer parábola. Tras evaluar en el punto 4, éste reemplaza a 3 y se obtiene la nueva parábola con centro en 5. (Tomado de [12]).

del mínimo con respecto a x debe ser menor a la mitad del movimiento realizado en la iteración anterior, para asegurar la convergencia del mínimo. La interpolación parabólica es aceptable cuando el mínimo actual se encuentra lejos del real, por lo que el descenso se realiza más rápidamente que con el método de la razón dorada. Este último método se utiliza para como último paso para asegurar la convergencia al mínimo evitando los grandes *pasos* de la interpolación parabólica. Los puntos a y b son sustituidos en cada iteración de acuerdo con la ubicación actual del mínimo, reduciendo el intervalo de búsqueda.

El algoritmo se detiene cuando la diferencia entre a y b es menor a $2x\epsilon$, donde ϵ es la precisión establecida para el algoritmo y x el punto medio del intervalo (a, b) .

2.3.3 Método de minimización lineal de funciones o *LineSearch*

El método de minimización lineal puede ser aplicado en el espacio multidimensional con respecto a una de las variables. Para un punto inicial multidimensional P , avanzando en la dirección del vector n , se requiere calcular λ tal que la función f minimice su valor, hasta que el cambio en el valor del mínimo sea menor al valor de tolerancia aceptable ϵ . A continuación se enuncia el algoritmo seguido en esta minimización.

1. Dado el punto inicial P y dirección definida por n
2. Calcular λ que minimice la función $f(P + \lambda n)$
3. Reemplazar P por $P + \lambda n$
4. Reemplazar n por λn
5. Listo. Retorna P_{final} y $f(P_{final} + \lambda n)$

Este método es también conocido como *LineSearch*. Los pasos anteriores son realizados por el método de Brent para la minimización de la función, considerando la dirección establecida por el vector n .

2.4 Downhill Simplex

El algoritmo de *Downhill Simplex* es un algoritmo de optimización multidimensional de funciones, utilizado para encontrar los puntos mínimos de funciones que dependen de más de una variable. Este método fue introducido por Nelder y Mead [20].

El método es fundamentalmente geométrico. Un simplex es una figura geométrica de $N + 1$ dimensiones, donde N es la dimensión de la función a optimizar, por ejemplo en dos dimensiones el simplex es un triángulo, en un tres dimensiones es un tetraedro. En general, interesa es que el simplex sea no degenerado, es decir que encierre un volumen finito.

En optimización multidimensional de funciones no existe análogo al acorralamiento de mínimos como se realiza para el caso de funciones en una dimensión, por lo tanto lo mejor que puede realizarse es darle al algoritmo una aproximación inicial como primer punto a probar [12].

El algoritmo debe ser inicializado con $N + 1$ puntos que definan el simplex inicial, uno de los cuales debe ser el punto inicial P_0 . Los demás puntos P_i se obtienen utilizando la relación

$$P_i = P_0 + \Delta e_i \quad (2.9)$$

donde e_i son los vectores unitarios definidos desde P_0 hasta los otros puntos P_i , necesarios para formar el simplex. Δ es una constante de escala del problema de optimización.

Una vez que el algoritmo obtiene el valor inicial evalúa la función multidimensional en cada punto en busca del error más bajo. El algoritmo conserva los puntos del simplex con menor error y busca nuevas aproximaciones para el punto con mayor error. En esta búsqueda de nuevos puntos el simplex puede deformarse por cambios en los vértices en las formas presentadas en la Figura 2.8.

La optimización inicia ubicando el vértice que posee mayor error P_{may} , para luego calcular el centroide \bar{P} del simplex utilizando los restantes N vértices con la ecuación

$$\bar{P} = \frac{\sum_{i=1, i \neq may}^N P_i}{N} \quad (2.10)$$

Los cambios en el simplex mostrados en la Figura 2.8 se realizan utilizando la ecuación

$$\zeta = \bar{P} + \alpha(\bar{P} - P_{may}) \quad (2.11)$$

donde ζ es el punto multidimensional resultante y α es la constante que determina el tipo de movimiento a realizar según su magnitud y su signo. E_ζ representa el valor de ζ en la función de error.

En cada una de las iteraciones el algoritmo realiza una reflexión del punto con mayor error, pasando a través del centroide, para lo cual se establece ($\alpha = -1$) en (2.11). Si el error E_ζ es menor que el actual mínimo dentro del simplex, se realiza una expansión ($\alpha = -2$)

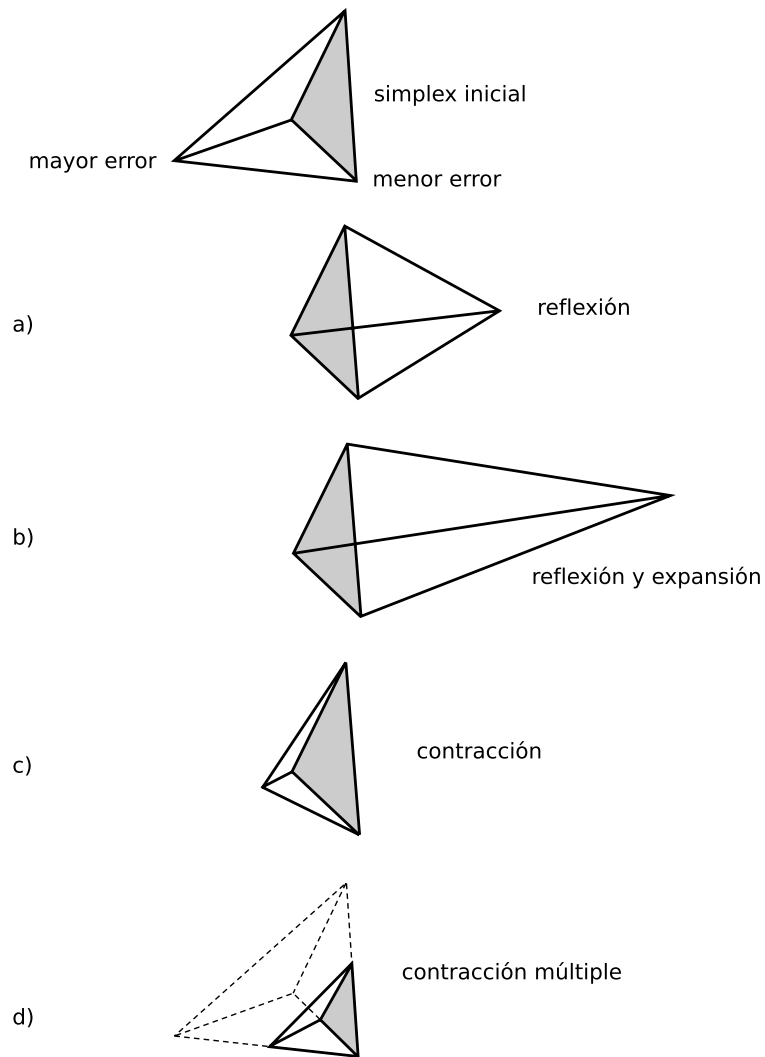


Figura 2.8: Cambios posibles para un simplex inicial (arriba). a) Reflexión alejándose del punto con mayor error, b) reflexión y expansión alejándose del punto con mayor error, c) contracción en una dimensión hacia el punto con menor error y d) una contracción de todas las dimensiones hacia el punto con menor error. (Tomado de [12]).

dando como resultado un nuevo punto ζ . Si $E_{\zeta} < E_{\varsigma}$ entonces el punto con mayor error es sustituido por el punto ζ , sino se utiliza el punto ς .

En el caso que $E_{min} < E_{\zeta} < E_{max}$, donde E_{min} y E_{max} representan el error mínimo y máximo del simplex actual, respectivamente, el punto con mayor error P_{may} es sustituido por ζ . Si el error E_{ζ} es mayor a E_{may} se busca un nuevo punto ζ realizando para ello una contracción en una dimensión ($\alpha = 0,5$) hacia el actual vértice con menor error, y en caso de que $E_{\zeta} < E_{max}$ se realiza la sustitución; sino la iteración se considera fallida y se realiza una contracción en todas las dimensiones en dirección al punto con el menor error.

2.5 Filtros de Savitzky y Golay

En el análisis de señales reales se utilizan métodos de mejora de la señal como filtros promediadores u otros filtros que ayuden a mejorar la señal estudiada. El método de filtrado de señales reales utilizado con mayor frecuencia en química analítica son los filtros propuestos por Savitzky y Golay [10].

El método se basa en el ajuste mediante mínimos cuadrados en una serie de datos de longitud $2m + 1$, donde el polinomio utilizado para el ajuste es evaluado generalmente en el centro de los $2m + 1$ puntos, siendo este entonces el valor aceptado como valor filtrado de la señal.

La idea de los filtros de Savitzky y Golay es la de aplicar a los datos un filtrado que permita preservar las características principales de la función como el ancho y la altura, mientras se realiza el proceso de mínimos cuadrados para atenuar el ruido.

El algoritmo de Savitzky y Golay permite obtener una función suavizada para un grupo de datos de $2m + 1$ puntos desde $i = -m$ a $i = m$, donde $i = 0$ es el punto donde la función es evaluada. Sin embargo los filtros de Savitzky y Golay a pesar de estar basados en una aproximación de mínimos cuadrados pueden ser calculados en forma fija y aplicarlos mediante convolución. La ecuación para los datos resultantes luego de la aplicación del filtro es

$$y_i = \sum_{k=-m}^m c_k x_{i+k} \quad (2.12)$$

donde c_i es el peso del filtro para la posición relativa i mientras que x_i corresponde al dato en la señal real. Inicialmente el algoritmo fue propuesto para ser resuelto mediante la resolución de un sistema lineal con $2m+1$ ecuaciones para obtener los coeficientes del filtro, sin embargo, el trabajo realizado por Gorry [13] hace una generalización utilizando polinomios de Gram para obtener los coeficientes para un filtro de cualquier ancho, incluyendo los coeficientes para las derivadas *suavizadas*.

La forma general del polinomio de Gram es la siguiente

$$f_n(t) = \sum_{k=0}^n b_k P_k^m(t) \quad (2.13)$$

donde $P_k^m(t)$ es el polinomio de Gram de orden k , con ancho de $2m + 1$ puntos, evaluado en el punto t . El polinomio de Gram está definido por

$$P_k^m(t) = \sum_{j=0}^k \frac{(-1)^{j+k} (j+k)^{(2j)} (m+t)^{(j)}}{(j!)^2 (2m)^{(j)}} \quad (2.14)$$

donde la función $(a)^{(b)} = (a)(a-1)\dots(a-b+1)$ y $(a)^{(0)} = 1$ es la función factorial generalizada. Sustituyendo (2.14) en (2.13) se obtiene la ecuación que representa el resultado de la

convolución con los datos, donde además se incluye la generalización para la derivada s .

$$f_n^s(t) = \sum_{j=-m}^m \sum_{k=0}^n \frac{(2k+1)(2m)^{(k)}}{(2m+k+1)^{(k+1)}} P_k^m(i) P_k^{m,s}(t) \quad (2.15)$$

donde los $P_k^{m,s}(t)$ se obtienen de su correspondiente derivada

$$P_k^{m,s}(t) = \left(\frac{\partial^s P_k^m(x)}{\partial x^s} \right)_{x=t} \quad (2.16)$$

Realizando la expansión de esta expresión se obtiene el polinomio P de orden n , derivada s ($s = 0$ cuando solamente se realiza suavizado de la función), evaluado en el rango de datos $-m \leq i \leq m$

$$P_k^m(i) = \frac{(2k+1)(2m)^{(k)}}{(2m+k+1)^{(k+1)}} [i P_{k-1}^m(i) + s P_{k-1}^{m,s-1}(i)] - \frac{(k-1)(2m+k)}{(2m+k+1)^{(k+1)}} P_{k-2}^{m,s}(i) \quad (2.17)$$

con lo que cada uno de los coeficientes $c_i^{t,s}$ se calcula con

$$c_i^{t,s} = \sum_{k=0}^n \frac{(2k+1)(2m)^{(k)}}{(2m+k+1)^{(k+1)}} P_k^m(i) P_k^{m,s}(t) \quad (2.18)$$

La ventaja de los filtros de Savitzky y Golay consiste en la menor deformación de la señal analizada, dado que los métodos comunes de filtrado por desplazamiento reducen el valor de la función en los puntos máximos, como se muestra en la Figura 2.9, donde se compara el filtrado utilizando un promediador simple con el filtro de Savitzky y Golay para 33 puntos de ancho, aplicados sobre una señal con ruido.

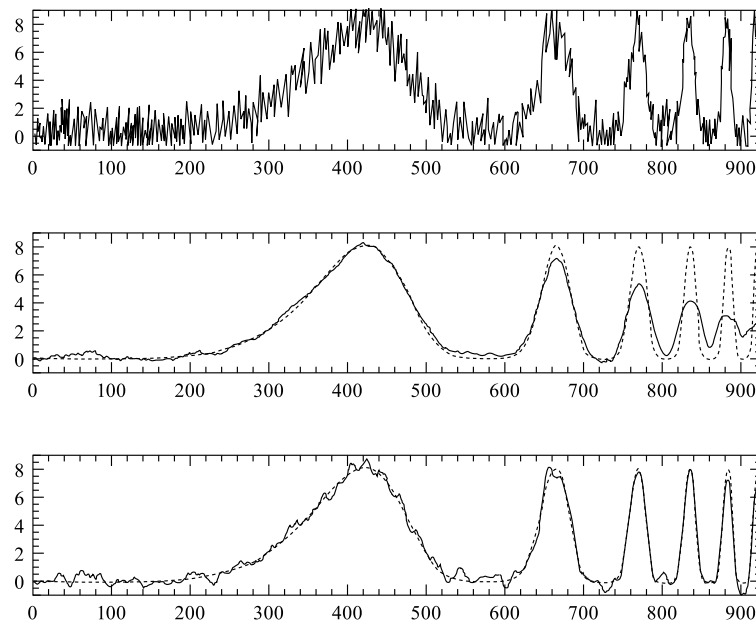


Figura 2.9: Efecto de la aplicación de filtros de Savitzky y Golay. Filtrado de la señal original con 6 picos de igual amplitud con ruido añadido (arriba). Para un filtro de 33 puntos de ancho se filtra la imagen con un promediador (centro) y con el filtro de Savitzky y Golay (abajo). La línea punteada corresponde a los datos sin ruido. (Tomado de [12]).

Capítulo 3

Sistema para caracterización de bandas

En este capítulo se presenta los detalles teóricos para la realización del sistema de caracterización de las bandas, lo cual incluye el filtrado de los datos de entrada para eliminar el ruido y fondo no deseado, así como el modelado del carril para lograr una caracterización óptima de los datos.

El método propuesto parte de la imagen de un carril real a la cual se le debe estimar la cantidad, posición y la desviación estándar de las bandas, para lo cual se consideran las bandas del mismo ancho. El esquema general del sistema propuesto se muestra en la Figura 3.1.

La primera aproximación de la desviación estándar se obtiene realizando el cálculo del espacio de escalas, el cual permite ubicar los puntos máximos correspondientes a cada una de las bandas y con ello determinar la desviación (σ) inicial. El valor de la desviación estándar obtenido es luego utilizado para realizar el filtrado y segunda derivada del carril dependiendo del valor de σ obtenido. Los valores máximos detectados con la segunda derivada permiten obtener la posición de las bandas, con lo cual puede realizarse una segunda aproximación de la desviación estándar ignorando los máximos cuya posición no corresponde a los máximos de la segunda derivada.

El valor de desviación estándar obtenido se utiliza en el siguiente paso de filtrado de la señal para eliminar el fondo y ruido indeseado, ya que este método necesita conocer el ancho de las bandas a optimizar para no deformar los datos. Cuando se obtienen el carril filtrado se hace una aproximación de σ para las posiciones encontradas con la segunda derivada utilizando minimización lineal para obtener el valor más probable de desviación estándar de las bandas. Este proceso se hace en forma iterativa hasta que la diferencia entre σ actual y el anterior sea menor a una tolerancia ϵ establecida de 0.001.

Finalmente el algoritmo realiza una optimización multidimensional de la amplitud y posición de las bandas con el valor de σ obtenido, para lo cual se utiliza el algoritmo *Downhill Simplex*.

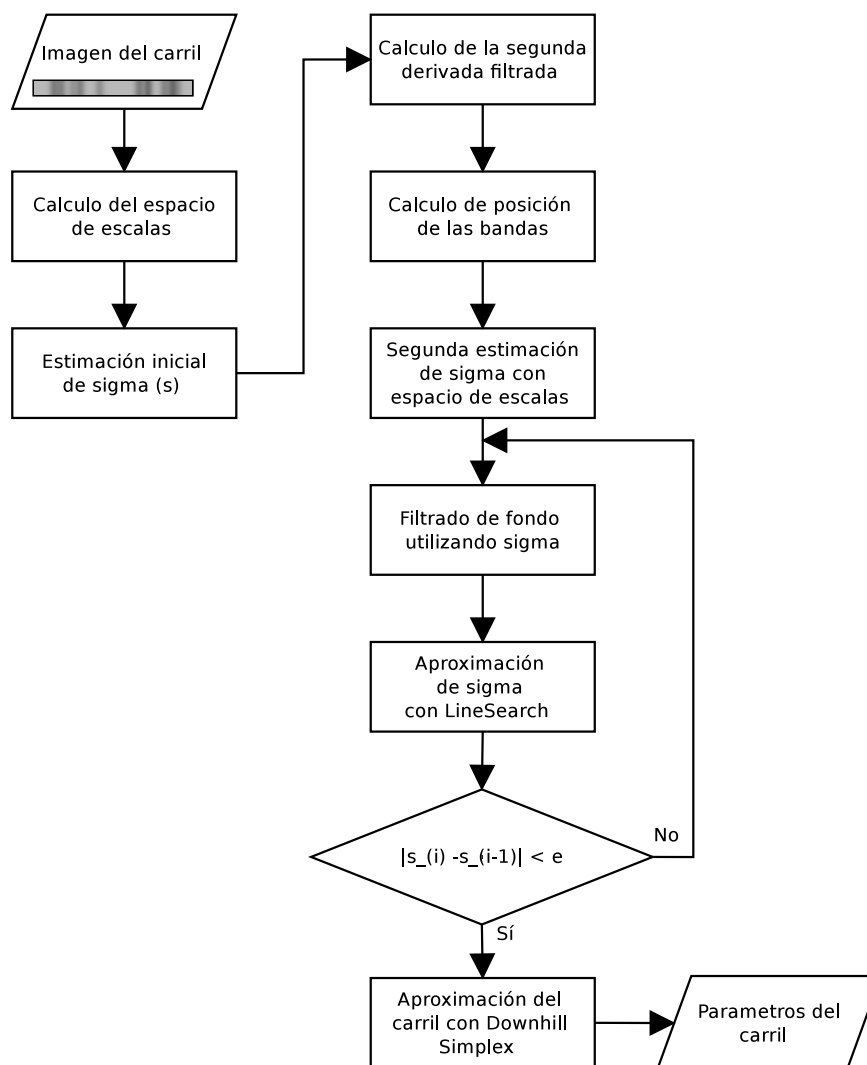


Figura 3.1: Diagrama completo del sistema propuesto.

3.1 Detección de desviación estándar mediante espacio de escalas

En esta sección se estudia el espacio de escalas para la caracterización de bandas en un carril de electroforesis. El análisis del espacio de escalas presentado corresponde al planteamiento realizado en el trabajo anterior en este sentido [11] con la variante de que se realiza el cálculo de la desviación estándar más probable en el espacio de escalas tomando en cuenta la cantidad de puntos máximos detectados en función de la escala utilizada.

3.1.1 Análisis del espacio de escalas

El espacio de escalas de un perfil unidimensional, como es el caso de la imagen que se analiza en el presente trabajo, se obtiene de convoluciones sucesivas con un kernel gaussiano discreto, cuya desviación estándar corresponde a la escala aplicada.

La expresión que describe la convolución de ambas funciones está dada por

$$L(x) = f(x) * g(x) \quad (3.1)$$

donde $f(x)$ representa el carril en estudio y $g(x, \sigma_f)$ corresponde al filtro aplicado con desviación estándar σ_f .

Considerando que los carriles de geles de electroforesis a analizar están compuestos por bandas con distribución gaussiana [18], la expresión en (3.1) consiste en convolucionar dos funciones gaussianas de diferente desviación estándar. de esta forma puede definirse la función f y la función g como

$$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x}{\sigma_b}\right)^2}}{\sigma_b\sqrt{2\pi}} \quad (3.2)$$

$$g(x) = \frac{e^{-\frac{1}{2}\left(\frac{x}{\sigma_f}\right)^2}}{\sigma_f\sqrt{2\pi}} \quad (3.3)$$

donde σ_b representa la desviación estándar de la banda en el carril de electroforesis y σ_f corresponde a la desviación estándar del filtro aplicado. En [11] se hace la deducción completa de las fórmulas utilizadas en esta sección. La convolución de las dos funciones anteriores genera una nueva función gaussiana $h(x)$

$$h(x) = \frac{e^{-\frac{1}{2}\left(\frac{x}{\sigma_r}\right)^2}}{\sigma_r\sqrt{2\pi}} \quad (3.4)$$

donde

$$\sigma_r = \sqrt{\sigma_b^2 + \sigma_f^2} \quad (3.5)$$

Al aplicar la segunda derivada a una banda con centro en $\mu = 0$, por simplicidad, se obtiene:

$$\frac{\partial^2 h(x)}{\partial^2 x} = \frac{(x - \sigma_r)(x + \sigma_r) e^{-\frac{1}{2}\left(\frac{x}{\sigma_r}\right)^2}}{\sqrt{2\pi} \sigma_r^5} \quad (3.6)$$

En el caso de una única banda, el máximo ocurre en su centro, esto es, en $x = \mu$, por lo que la expresión anterior puede simplificarse para obtener

$$\frac{\partial^2 h(x=0)}{\partial^2 x} = \frac{1}{\sqrt{2\pi} \sqrt{\sigma_b^2 + \sigma_f^2}^3} \quad (3.7)$$

Analizando esta ecuación se observa que el punto máximo depende del valor de la desviación estándar de la banda tanto como de la del filtro. La Figura 3.2 muestra el comportamiento

de esta función cuando se le normaliza multiplicando por la desviación estándar del filtro a diferentes potencias. Para la normalización con $\sigma_f^{1,5}$ se observa que el máximo del espacio de escalas coincide con la desviación estándar de la banda, por lo tanto éste es el mejor valor del cual se puede obtener la desviación de la banda, y por esta razón es el valor utilizado en el programa.

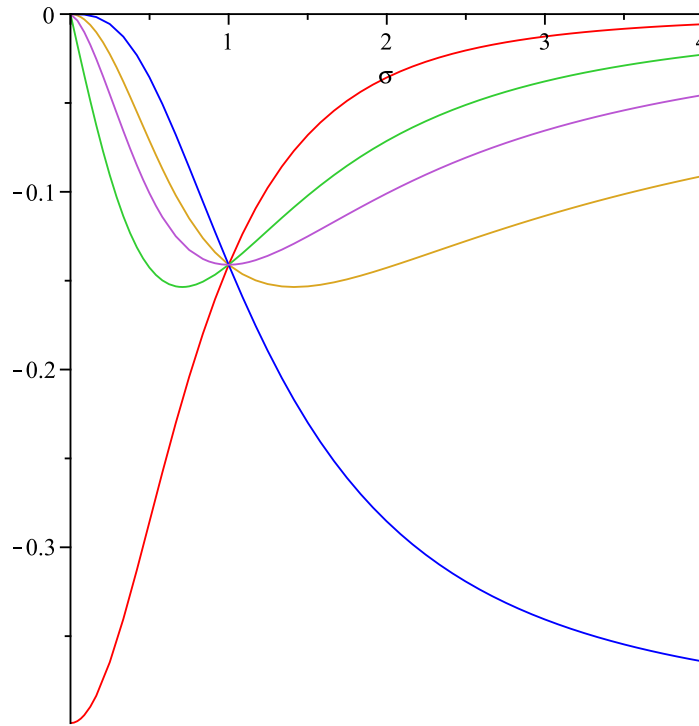


Figura 3.2: Normalización del máximo de de la segunda derivada del espacio de escalas. Normalización de (3.7) para valores de $n = 0$ (línea roja), $n = 1$ (línea verde), $n = 1, 5$ (línea morada), $n = 2$ (línea café) y $n = 3$ (línea azul) para un $\sigma_b = 1$. (Tomado de [11]).

La representación tridimensional del espacio de escalas se muestra en la Figura 3.3, en la cual se grafica la función (3.6) para un carril con una única banda.

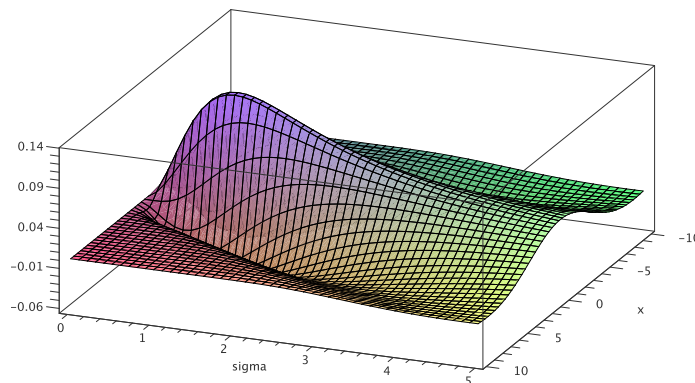


Figura 3.3: Espacio de escalas para una sola banda.

3.1.2 Enventanado del espacio de escalas

El espacio de escalas normalizado presenta una forma directa para la medición de la desviación estándar de las funciones gaussianas en un carril. En general las imágenes de geles de electroforesis presentan carriles con varias bandas, y partiendo de la consideración de que todas las bandas poseen la misma desviación estándar, puede encontrarse la región con la mayor cantidad de bandas de la imagen mediante una ventana.

El proceso de enventanado del espacio de escalas consiste en el cálculo del espacio de escalas tal como se muestra en la Figura 3.4. El espacio de escalas es recorrido en busca del punto máximo para cada una de las bandas, identificados como puntos blancos en la imagen.

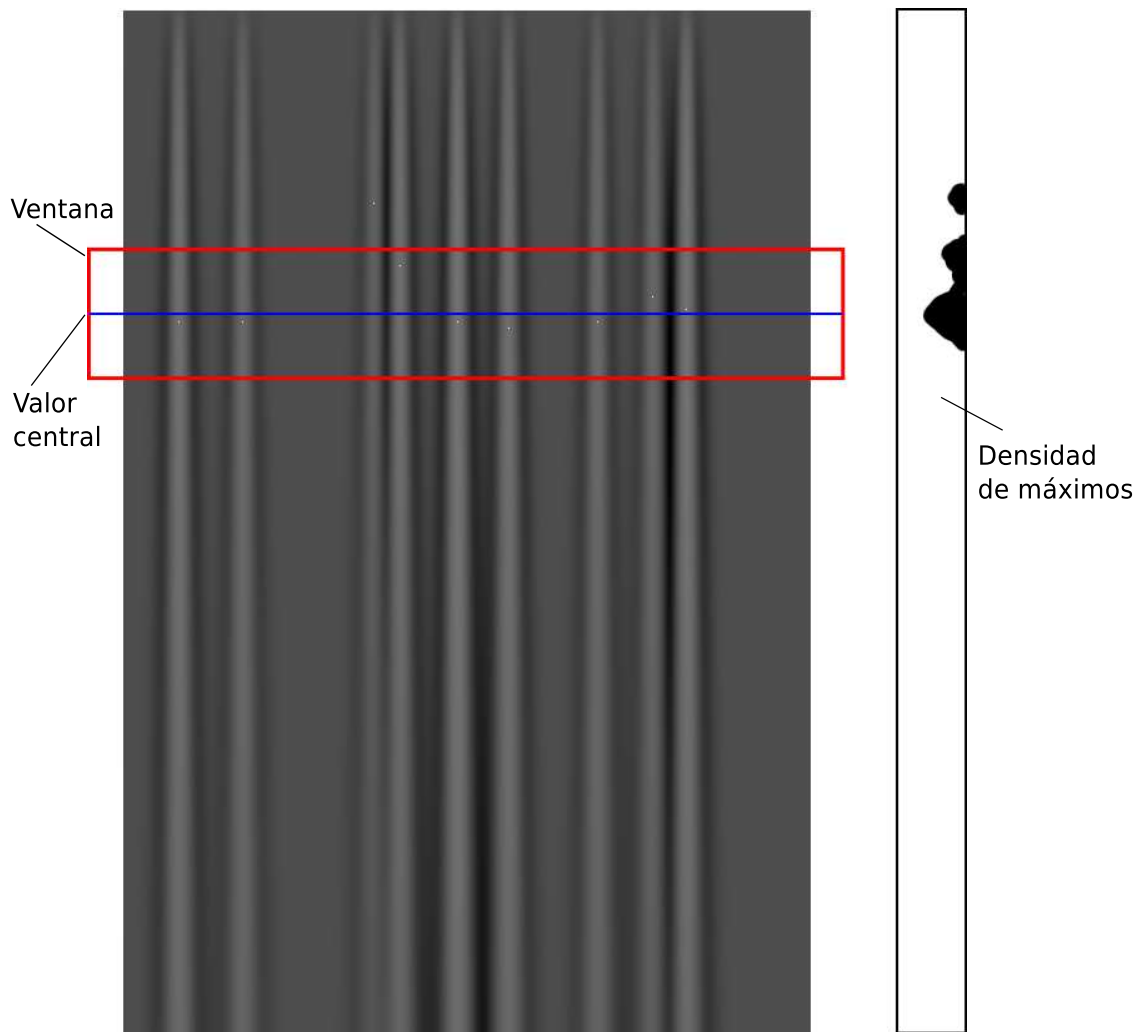


Figura 3.4: Enventanado del espacio de escalas.

Utilizando los valores máximos se genera un vector de densidad de puntos máximos con respecto a la desviación estándar realizando una convolución con una función rectangular y luego con una función semicircular para obtener una función suavizada que facilite la búsqueda del máximo, el cual corresponde al valor de desviación estándar más probable en el espacio de escalas. En la Figura 3.5 se ejemplifica este proceso de generación de una función de densidad de puntos máximos a partir de dos convoluciones, donde cada uno de

los filtros aplicados tiene un ancho de $\sigma = 1$, ó 100 píxeles que es la resolución utilizada en el vector de máximos con respecto a σ obtenidos del espacio de escalas.

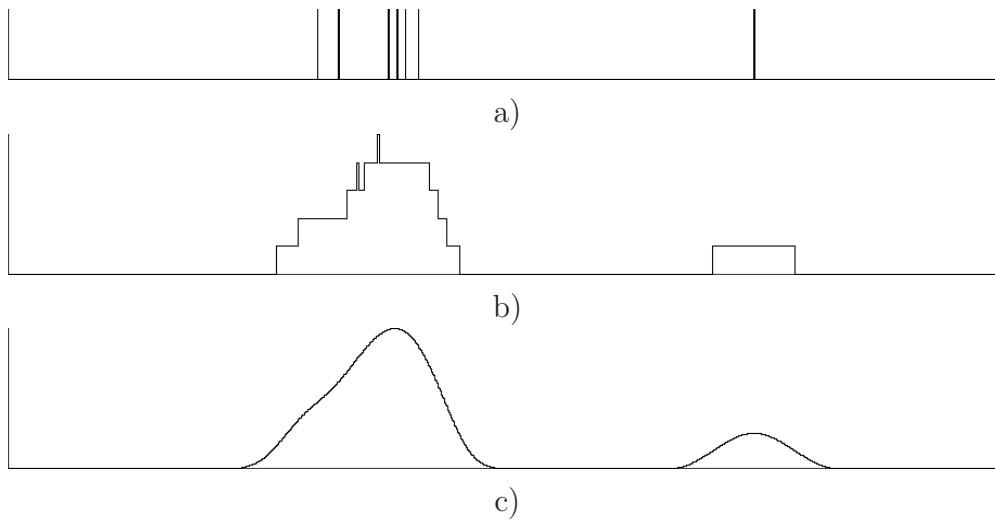


Figura 3.5: Detección de la desviación estándar utilizando espacio de escalas. a) Distribución máximos en el espacio de escalas con respecto a σ . b) convolución de a) con un impulso rectangular; c) convolución de b) con una función semicircular.

La aplicación de un filtro rectangular se realiza para evitar el decrecimiento de la función resultante de la convolución en la región entre máximos con poca separación al utilizar un filtro de suavizado de tipo gaussiano, semicircular ó alguna forma similar. Filtrando primero con una función rectangular es posible establecer como máximo del espacio de escalas un punto intermedio entre los dos máximos encontrados con poca separación, pero es necesario realizar un suavizado de la función que permita realizar la búsqueda del punto máximo, para lo cual en este trabajo se eligió una función semicircular. Sin embargo, para el filtrado de los datos se realiza primero la convolución entre estos dos filtros, para luego filtrar los datos en un único paso.

Una vez que se ha establecido el valor máximo se calcula una ventana centrada en ese punto, cuyo ancho es un parámetro definible por el usuario, de forma que se puedan tomar en cuenta las bandas cuyo máximo se encuentra dentro de esta ventana como los datos más representativos del carril, puesto que son los puntos máximos más cercanos al valor central.

3.1.3 Problemas en la determinación de la desviación estándar

El espacio de escalas permite ubicar de forma correcta la ubicación de los máximos para una banda, sin embargo, puesto que la segunda derivada posee lóbulos laterales de signo opuesto se afecta la respuesta de múltiples bandas adyacentes.

Por ejemplo en el caso de dos bandas de la misma amplitud y con $\sigma = 1$, cuyos máximos se encuentran separados por 3σ , el espacio de escalas obtenido es el de la Figura 3.6. Debido a la acción de la amplitud de los lóbulos, la ubicación de los máximos locales se ve afectada, donde 1 sería el valor máximo deseado.

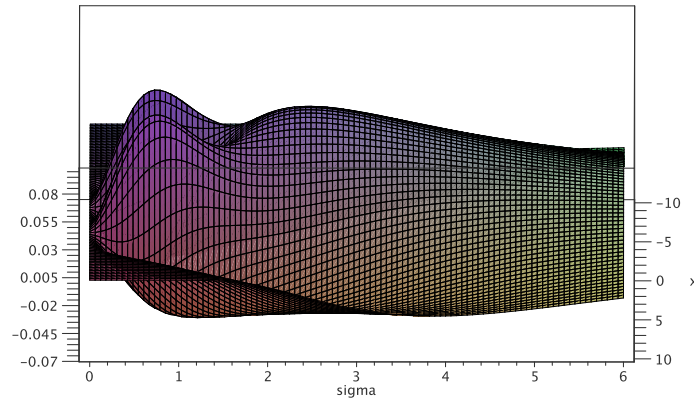


Figura 3.6: Espacio de escalas de dos funciones de igual amplitud, separadas por 3 sigma

La influencia de este fenómeno sobre otras bandas se incrementa cuando aumenta la cantidad de bandas juntas. El gráfico de la Figura 3.7 muestra el espacio de escalas para 6 bandas de $\sigma = 1$ de igual amplitud, separadas $2,5\sigma$.

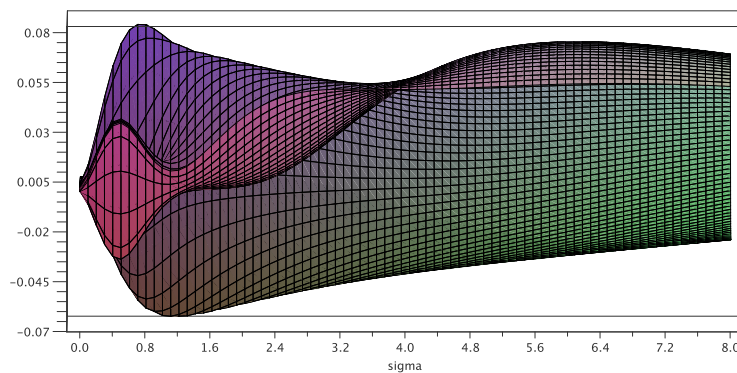


Figura 3.7: Efecto de distorsión en el espacio de escalas de seis funciones de igual amplitud, separadas por 2.5 sigma

La influencia de las otras bandas influye tanto en la posición del máximo como en la amplitud del espacio de escalas correspondiente a cada banda. En esta imagen es apreciable como en el caso de aglomeración de bandas, solamente las bandas laterales tienen su máximo ubicado cercano al valor teórico requerido, por lo tanto el valor de desviación estándar obtenido mediante el espacio de escalas es incorrecto en la presencia de bandas traslapadas.

Según [14] la desviación estándar de una sumatoria de funciones gaussianas no puede ser determinada exactamente por medio del espacio de escalas cuando existen dos ó más gaussianas traslapadas, sin embargo, puede utilizarse la desviación estándar de este método como un valor inicial para la aproximación de las funciones gaussianas mediante mínimos cuadrados.

3.2 Filtrado para extracción de fondo del carril

Las imágenes de carriles de geles de electroforesis suelen presentar imperfecciones que perjudican en el proceso de caracterización de las bandas. La presencia de fondo en los carriles dificulta la detección de bandas así como la distinción del ancho de las mismas, como se muestra en el carril real de la Figura 3.8.



Figura 3.8: Carril real con fondo.

La mejora se realiza mediante el filtrado de los carriles utilizando el algoritmo desarrollado en [6], en el cual se aplica un filtro de mínimos y luego un filtro de máximos a los datos.

Puesto que el ruido de fondo en la imagen tiene una varianza mayor a la de las bandas del carril se utiliza como parámetro para el tamaño del filtro aplicado un valor mayor al de la desviación estándar de las bandas. En este trabajo se utiliza un tamaño de 3 veces el ancho total de las bandas (18σ), de forma que el algoritmo remueva el ruido pero no reduzca el ancho de las bandas a encontrar. Por último este método realiza un suavizado de la imagen con el mismo tamaño de filtro.

3.3 Mejora de la resolución utilizando derivadas

La solución planteada se basa en el modelo de bandas descritas como funciones gaussianas de igual desviación estándar, de modo que el carril se encuentra formado por la sumatoria de estas funciones. En el caso ideal en el que las funciones se encuentran separadas podría realizarse una búsqueda de máximos con el fin de encontrar la posición de las bandas una vez eliminado el ruido.

Como se muestra en la Figura 3.9, en el punto máximo de una función gaussiana, la primera derivada pasa por cero cambiando su valor de positivo a negativo. La detección de este cambio es un indicio de que en este punto se encuentra un máximo.

Además la segunda derivada presenta un mínimo en la ubicación del máximo de la gaussiana de interés, por lo que el análisis de ambas derivadas es complementario en la búsqueda de los valores máximos en carriles formados por funciones gaussianas.

La detección de la ubicación de las bandas mediante el proceso de búsqueda de máximos no puede realizarse cuando se procesan imágenes en las cuales el traslape existente entre las bandas adyacentes es tan severo que es imposible distinguirlas, aún si la señal es carente de ruido, como se muestra en la Figura 3.10. En esta figura se muestra un carril teórico con tres funciones gaussianas con una separación de 2 veces la desviación estándar.

La Figura 3.10 muestra cómo al desaparecer los máximos locales en el carril, la primera derivada no cruza por cero en la ubicación correspondiente a los máximos de cada una de las bandas. Este hecho descarta el simple uso de la primera derivada como método para hallar

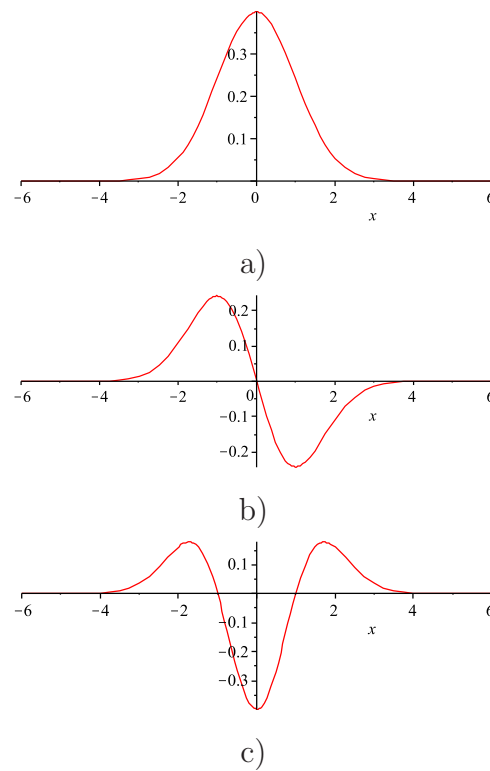


Figura 3.9: Derivadas de una función gaussiana. a) Función gaussiana, b) primera derivada, c) segunda derivada

los máximos.

Sin embargo, la segunda derivada sí conserva la generación de los mínimos en una posición próxima a la de los máximos de las tres gaussianas, por lo tanto esta es más confiable para la ubicación de la posición de las bandas en un carril con bandas traslapadas.

El cálculo de la segunda derivada se realiza en este trabajo una única vez utilizando el filtrado y suavizado de Savitzky y Golay en un único paso. Este se hace al inicio del programa para encontrar la posición de las bandas, considerando un ancho de filtro de 3σ de acuerdo con [9], donde se recomienda utilizar para el filtrado de funciones gaussianas un ancho del filtro no mayor a 0.9 veces el ancho de las bandas a la mitad de la altura (aproximadamente $2,35\sigma$ para una distribución gaussiana), esto para producir una distorsión de la altura de los picos no mayor al 1%, por lo tanto para el filtrado en el programa se utiliza un ancho de 2σ . Luego de la segunda derivada se aplica un filtrado para suavizado de los datos finales utilizando el suavizado de Savitzky y Golay. Tanto para la derivada como para el suavizado se utiliza un polinomio de Gram de segundo orden.

Sin embargo, para las bandas más angostas se utiliza un ancho mínimo del filtro de 7 píxeles, puesto que para anchos menores la extracción de ruido es deficiente en imágenes reales.

En este proceso de derivación se aplica también una umbralización de los máximos encontrados en la segunda derivada, de forma que sólo serán aceptables los que superen el promedio de la parte positiva de la segunda derivada.

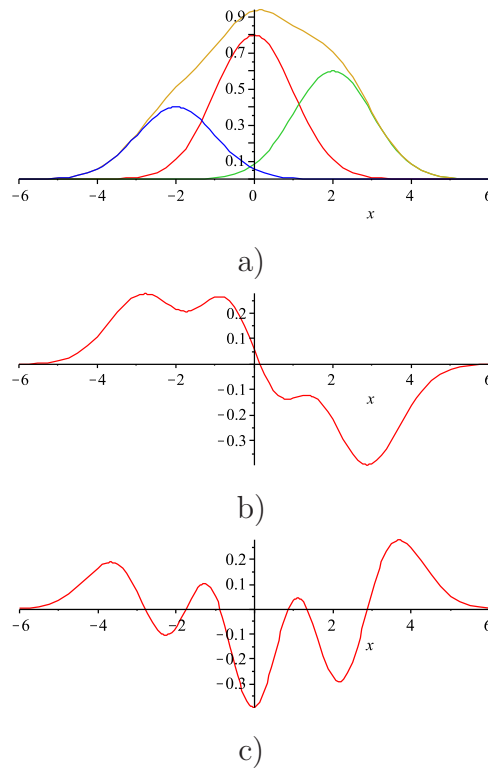


Figura 3.10: Derivadas de una sumatoria de gaussianas con traslape severo. a) Sumatoria de gaussianas con traslape severo, b) primera derivada, c) segunda derivada

3.4 Modelado del carril como sistema lineal

Las bandas presentes en un carril con traslape son identificables si se considera el análisis utilizando la segunda derivada, tal como se explicó en la sección anterior.

Conociendo las posiciones de cada una de las bandas se modela el carril completo ya sin fondo como una sumatoria de funciones gaussianas de diferentes amplitudes pero de igual desviación estándar, ubicadas en las posiciones μ_i encontradas con la segunda derivada.

De esta forma, el carril completo $l(x)$ puede expresarse como la siguiente serie con desviación estándar constante σ

$$l(x) = \sum_{i=1}^n A_i \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \quad (3.8)$$

Utilizando la expresión anterior puede construirse un sistema lineal con los valores del carril sin fondo $f(x)$, evaluado en cada una de los máximos encontrados μ_i como:

$$\begin{pmatrix} f(\mu_1) \\ f(\mu_2) \\ \vdots \\ f(\mu_n) \end{pmatrix} = \begin{pmatrix} g_1(\mu_1) & g_2(\mu_1) & \cdots & g_n(\mu_1) \\ g_1(\mu_2) & g_2(\mu_2) & \cdots & g_n(\mu_2) \\ \vdots & \vdots & \ddots & \vdots \\ g_1(\mu_n) & g_2(\mu_n) & \cdots & g_n(\mu_n) \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix} \quad (3.9)$$

En este sistema lineal se expresa el valor en un punto máximo como la suma de las componentes de cada una de las otras bandas, donde $g_i(x)$ se define como

$$g_i(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \quad (3.10)$$

Una vez que se define un valor de desviación estándar σ se puede realizar el cálculo de las amplitudes A_i correspondientes a cada una de las bandas para una desviación estándar determinada resolviendo el sistema lineal anterior.

3.5 Ajuste por minimización de una función objetivo

El carril de ajuste generado en la sección anterior $l(x)$ (3.8) se aproxima al carril real sin fondo definido por una función $f(x)$ cuando la desviación estándar utilizada para la aproximación es igual al valor teórico de las bandas, en cuyo caso el error sería el mínimo. La función de error cuadrático medio para un carril lineal consiste en evaluar la suma de las distancias al cuadrado para los diferentes puntos del carril x , para cada valor de σ , tal como se expresa en

$$E(\sigma) = \frac{1}{w} \sum_{x=1}^w \left(f(x) - \sum_{i=1}^n A_i \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \right)^2 \quad (3.11)$$

donde w corresponde al ancho en píxeles del carril en estudio. El proceso de aproximación del carril sintético al real puede realizarse por medio del cálculo del error para los diferentes valores de desviación estándar de las bandas en forma iterativa en el nuevo carril $l(x)$. El valor que hace mínima la función 3.11 es el valor más probable de σ de las bandas del carril real $f(x)$.

3.6 Ajuste de la posición de las bandas y segmentado del carril en ventanas

En este trabajo se utiliza la optimización de la posición de las bandas encontradas utilizando el algoritmo *Downhill Simplex*, el cual realiza una optimización multidimensional de un conjunto de los datos de posición y amplitud para un carril.

Puesto que la cantidad de iteraciones del algoritmo incrementa cuanto mayor sea la cantidad de dimensiones consideradas, se realiza la división del carril en ventanas, utilizando el método desarrollado en [21]. Este método divide el carril en ventanas traslapadas una distancia de 8σ , de forma que esa región pueda contener al menos una banda por completo. Esta segmentación se muestra de forma gráfica en la Figura 3.11.

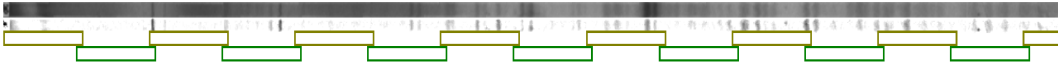


Figura 3.11: Segmentación de un carril en ventanas.(Tomado de [21])

Al realizar esta división y conocerse en forma aproximada la posición de las bandas, la cantidad de iteraciones necesarias para la convergencia del algoritmo se reduce al cálculo de vectores de menos dimensiones en cada ventana. Sin embargo, este método tiene como inconveniente que las bandas encontradas en las zonas de 8σ a partir los bordes de cada ventana estarán duplicadas en la ventana adyacente, por lo que debe establecerse una forma de fusión de bandas para estos casos en los datos finales.

Una vez que las bandas detectadas en todas las ventanas son obtenidas, éstas son incluidas en un único vector y ordenadas por la posición. Si dos bandas están separadas por una distancia menor a δ , que en este trabajo se considera como la desviación estándar aproximada, las bandas pueden ser fusionadas. Para esta fusión se debe considerar que si en una de las ventanas la banda no se encuentra completa, ubicada a menos de 3σ de uno de los extremos, se descarta y se toma la banda presente en la ventana adyacente porque está aproximada considerando más información. Si la banda está aproximada con suficiente información en ambas ventanas se utiliza la siguiente fórmula para realizar la fusión de las bandas:

$$\mu_t = \frac{A_i\mu_i + A_{i+1}\mu_{i+1}}{A_i + A_{i+1}} \quad (3.12)$$

donde A_i y μ_i representan la amplitud y posición de una banda; A_{i+1} y μ_{i+1} representan la otra banda a fusionar y μ_t es la posición de la banda resultante, donde la amplitud en este punto A_t , se obtiene de leer el valor del carril en esa posición.

Capítulo 4

Resultados y Análisis

En este capítulo se analizan los resultados obtenidos con la implementación del espacio de escalas, de forma que se discuten los resultados obtenidos y se analiza la aptitud de este método en la determinación de la posición y ancho de las bandas en carriles de geles de electroforesis.

Además en este capítulo se incluyen los resultados obtenidos con el método de detección de bandas mediante derivadas del carril, el cual permite aproximar las bandas mediante el ajuste de funciones gaussianas en las posiciones detectadas.

Los algoritmos de detección de bandas se evalúan para el caso de bandas teóricas con la finalidad de comprobar el comportamiento deseado, y se aplican luego a carriles de geles de electroforesis reales.

La ejecución del algoritmo se realiza en una computadora con procesador de doble núcleo AMD Athlon X2 4200+ a 2,20 GHz, con 2GB de memoria RAM.

4.1 Espacio de escalas

El espacio de escalas generado en este trabajo solamente toma en cuenta la distribución de los datos en una dimensión, por lo tanto el análisis realizado a cada carril simplemente toma una de las filas del mismo para el análisis.

El análisis de comportamiento del espacio de escalas se evalúa en esta sección en condiciones teóricas con carriles sintéticos generados como una sumatoria de funciones gaussianas, utilizando el algoritmo desarrollado en [21]. Con esto se analiza la caracterización que realiza el algoritmo considerando los casos en los cuales las bandas se encuentran aisladas, así como cuando se presenta traslape por alta cantidad de bandas en el carril.

La primera evaluación del algoritmo del espacio de escalas consiste en aplicar como entrada diferentes carriles sintéticos con una única banda sintética, con desviación que varía desde 0,5 hasta 10 en pasos de 0,5. En la Tabla 4.1 se muestran los resultados obtenidos para un conjunto de 10 carriles por cada valor de desviación estándar, junto con el error obtenido

con respecto a la desviación teórica y el coeficiente de variación de la medición, que toma en cuenta la desviación estándar (SD) y la media de los valores ($\bar{\sigma}$) del grupo de mediciones:

$$C.V. = \frac{SD}{\bar{\sigma}} 100\% \quad (4.1)$$

Tabla 4.1: Estimación de desviación estándar con espacio de escalas para carriles teóricos con una banda

Desviación teórica	Desviación promedio	C.V. (%)	Error (%)
0,5	1,066	8,967	113,200
1	1,409	2,469	40,900
1,5	1,745	0,906	16,333
2	2,167	2,393	8,350
2,5	2,650	0,000	6,000
3	3,146	1,228	4,867
3,5	3,603	0,614	2,943
4	4,100	0,771	2,500
4,5	4,583	0,698	1,844
5	5,065	0,666	1,300
5,5	5,566	0,227	1,200
6	6,063	0,473	1,050
6,5	6,576	0,425	1,169
7	7,094	0,503	1,343
7,5	7,542	0,232	0,560
8	8,032	0,286	0,400
8,5	8,545	0,425	0,529
9	9,062	0,530	0,689
9,5	9,533	0,284	0,347
10	10,06	0,297	0,550

Los datos obtenidos en esta tabla muestran que el algoritmo presenta un error en aumento cuanto más pequeña es la desviación estándar, llegando a ser de hasta 113% para $\sigma = 0,5$. Esto se debe a las pocas muestras que se tienen de la función gaussiana respectiva, insuficientes para caracterizar la distribución, porque como se observa en los datos el error disminuye significativamente al aumentar el valor de σ .

La Figura 4.1 muestra gráficamente el comportamiento de este error, excluyendo el error para $\sigma = 0,5$ debido a que el error correspondiente para este valor no permite apreciar claramente el comportamiento de la curva de error para los demás valores.

Para este cálculo que se realiza considerando solamente 1 banda por carril, la dispersión obtenida en los datos es muy baja como lo indica el coeficiente de variación, llegando a ser incluso 0 en uno de los casos.

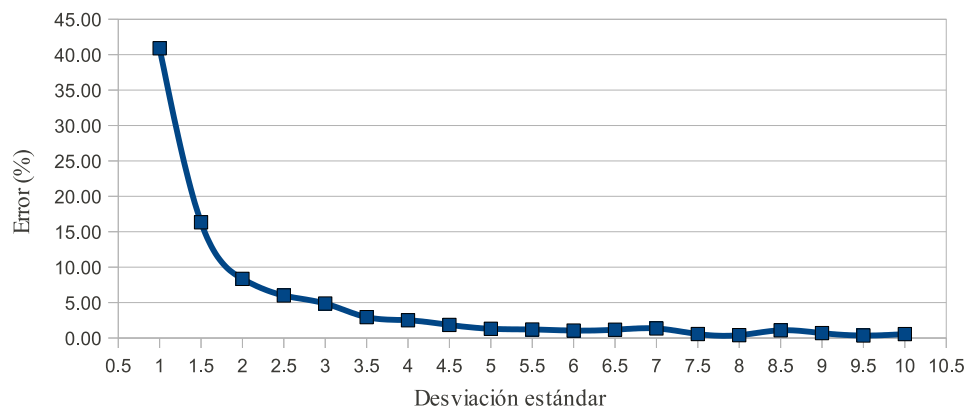


Figura 4.1: Error en la determinación de la desviación estándar con espacio de escalas en función de la desviación estándar de las bandas.

4.1.1 Evaluación de carril con poco traslape entre bandas

En esta sección se analiza el comportamiento del espacio de escalas con carriles teóricos con más de una banda en la determinación de la desviación estándar.

La primera evaluación del algoritmo se realiza en un carril con varias funciones gaussianas con traslape mínimo ó inexistente para algunas de las bandas. En la Tabla 4.2 se muestra el resultado de evaluar el espacio de escalas con un carril que contiene 10 bandas con $\sigma = 5$ para obtener el valor de la desviación estándar para cada una de las bandas del carril.

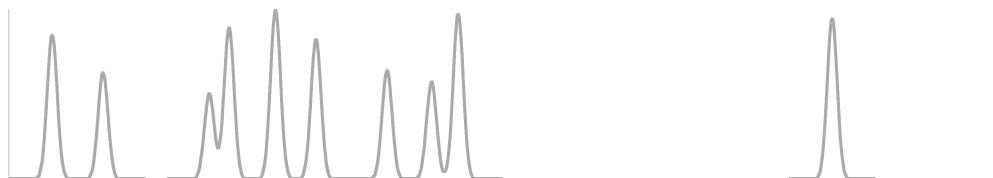


Figura 4.2: Carril sintético con 10 bandas con $\sigma = 5$.

En los datos obtenidos se observa es notable que las dos bandas con menor separación entre ellas poseen un error en su desviación de 8,13% y 19,75%, con una distancia de aproximadamente $4,2\sigma$ entre los máximos de cada una.

Sin embargo, cuando las bandas se encuentran completamente separadas, esto es, separadas por una distancia mayor a 6σ que corresponde a 30 píxeles para este carril, el error más alto es de 2,84%.

4.1.2 Evaluación de carriles con traslape

En el ejemplo de la Tabla 4.2 se realizó el inventariado del espacio de escalas, con el cálculo del vector de densidad de bandas con respecto a la desviación estándar con la finalidad de obtener el valor más probable, que resulta ser 4,98. A partir de este punto se aplica una ventana al espacio de escalas con $\pm 0,5\sigma$, esto es, desde 4,48 a 5,48. En este intervalo se

Tabla 4.2: Estimación de desviación estándar con espacio de escalas para un carril con 10 bandas con $\sigma = 5$

Posición	Desviación obtenida	Error %
44	5,06	1,19
95	5,06	1,19
201	4,01	19,75
222	4,59	8,13
268	5,06	1,19
309	5,11	2,18
380	5,06	1,19
424	4,87	2,84
451	4,96	0,80
826	5,06	1,19

encontraron 9 bandas, debido a que para la mayor cantidad de las bandas la desviación estándar sí fue estimada correctamente.

Para comprobar el corrimiento del valor de σ en el algoritmo implementado del espacio de escalas se hace necesaria la evaluación con un carril con traslape. En el siguiente ejemplo se utiliza un carril con 9 bandas severamente traslapadas con la finalidad de evidenciar el comportamiento descrito en la sección 3.1.3.

En el caso del carril de la Figura 4.3, se utilizan 9 bandas con traslape severo en el caso en que es apenas distinguible la presencia de dos bandas analizando la imagen de perfil. Las bandas utilizadas tienen $\sigma = 5$

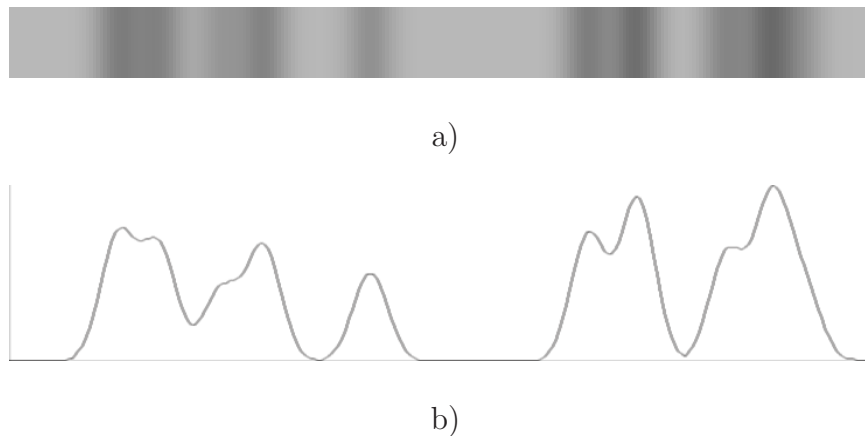


Figura 4.3: Carril sintético de 9 bandas con traslape. a) Perfil del carril, b) perfil del carril sin fondo

Los resultados de la estimación de la desviación estándar se muestran en la Tabla 4.3. En el caso de las bandas ubicadas en 30 y 218, los máximos locales desaparecen debido a la presencia del máximo global, puesto que la suma del espacio de escalas de cada una de las bandas

se une como se muestra en la Figura 3.6 en un solo máximo global, haciendo desaparecer el local para dicha banda y dando un valor de desviación de 8,51 y 9,54, respectivamente para estas dos bandas.

Tabla 4.3: Máximos locales en un carril con 9 bandas traslapadas

Posición	Desviación estándar	Error %
30	-	-
43	3,33	33,37
60	2,30	53,96
73	4,70	5,97
103	5,01	0,20
165	3,55	29,00
180	4,01	19,75
204	3,10	38,03
218	-	-

La aplicación del cálculo de la densidad de bandas en función de sigma da como resultado que la desviación estándar más probable dentro del carril es de 3,34, y aplicando la ventana de $\pm 0,5\sigma$, como se realizó en el ejemplo anterior, se encuentra que en el rango de 2,84 a 3,84 solamente se detectan 3 de las 9 bandas, por lo que no es confiable la estimación realizada. La estimación fue correcta en este ejemplo solamente para la banda ubicada en 103, para la que se obtuvo un error de 0,20 % en la estimación, pero se puede notar en el perfil del carril que ésta es la única banda que se encuentra aislada, sin distorsión por traslape.

En la Tabla 4.4 se muestran los resultados para la evaluación del algoritmo considerando múltiples carriles. Las pruebas realizadas consideran 10 carriles con 20 bandas para cada valor de desviación estándar en el rango 0,5 a 10, con paso de 0,5.

El valor de σ promedio mostrado en dicha tabla corresponde al promedio de los valores de desviación estándar más probables para cada carril, considerando los 10 carriles. Los carriles generados tienen la restricción de una separación mínima entre bandas de 2σ , de forma que todas las bandas puedan ser distinguidas y evitar bandas más anchas formadas por bandas cuya separación no permite distinguirlas.

Según estos datos obtenidos, el error en el cálculo de la desviación estándar llega a ser de 22% en el caso de $\sigma = 10$, cuyo error es de 0,55% para el caso de un banda y el coeficiente de variación es hasta 28,85% para $\sigma = 4,5$, por lo que los datos no pueden ser considerados estables ni precisos cuando hay traslape. Es por esta razón que no se utiliza por sí solo el espacio de escalas para la determinación de la desviación estándar.

Tabla 4.4: Evaluación de la desviación estándar utilizando carriles sintéticos con 20 bandas traslapadas. Aproximación realizada con el espacio de escalas.

σ teórico	σ Promedio	C.V. (%)	Error (%)	σ teórico	σ Promedio	C.V. (%)	Error (%)
0,5	0,999	2,185	99,800	5,5	4,361	4,494	20,709
1	1,246	5,272	24,600	6	4,978	7,506	17,033
1,5	1,541	4,965	2,733	6,5	5,719	9,904	12,015
2	1,863	2,994	6,850	7	5,577	7,558	20,329
2,5	2,326	5,015	6,960	7,5	6,853	6,592	8,627
3	2,679	2,607	10,700	8	6,952	8,600	13,100
3,5	3,037	6,050	13,229	8,5	7,803	8,748	8,200
4	3,436	7,124	14,100	9	8,052	9,639	10,533
4,5	4,483	28,851	0,378	9,5	8,513	9,663	10,389
5	4,294	5,974	14,120	10	7,778	7,319	22,220

4.2 Estimación de la posición y cantidad de bandas

En esta sección se presentan los resultados obtenidos utilizando la segunda derivada para la ubicación de las bandas dentro de un carril, en conjunto con la reducción de la función objetivo que caracteriza al carril como una suma de bandas gaussianas.

Esta es una solución complementaria al método del espacio de escalas, presentada en la sección anterior, debido a que no se obtuvo un resultado correcto para las situaciones donde se presenta traslape de bandas en el carril.

4.2.1 Ubicación de bandas con la segunda derivada

La localización de las bandas en un carril de electroforesis puede ser realizada utilizando la segunda derivada, debido a que esta aumenta los cambios en la señal [10], por lo cual facilita la tarea de búsqueda de máximos.

El aumento de la escala deja una señal cada vez con menos estructura, tal como se muestra en la Figura 4.4, donde se muestra cómo al aumentar la escala las bandas cercanas se unen en el espacio de escalas formando una sola banda y un único máximo, por lo que la segunda derivada provee la mejor resolución para encontrar las bandas, la cual corresponde a la fila inicial del espacio de escalas.

En [14] se destaca que para la caracterización de las bandas es más conveniente utilizar la más alta resolución del espacio de escalas, que corresponde a la segunda derivada del carril, porque al aumentar la escala los cruces por cero de la derivada y los máximos se desplazan más significativamente. En un caso sin traslape los cruces por cero de la segunda derivada se encuentran situados en $\pm\sigma$ a partir del punto central de la banda y la identificación de estos permitiría determinar la desviación estándar en forma directa. La derivada realizada

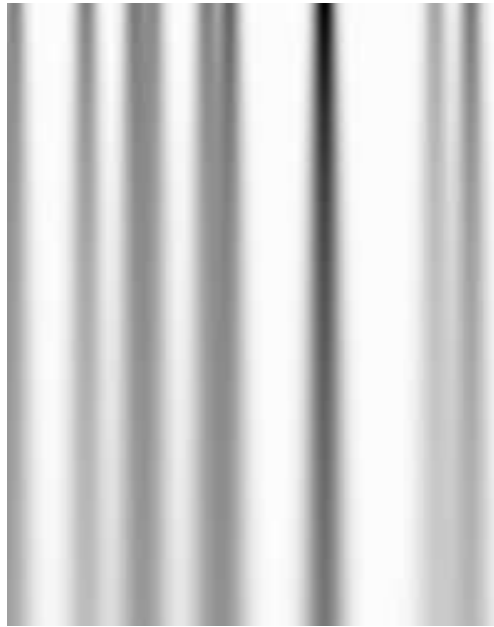


Figura 4.4: Espacio de escalas con traslape de bandas.

en este trabajo se hace utilizando el método de filtrado y derivación de Savitzky y Golay ya descrito en la sección 2.5 para encontrar directamente la segunda derivada.

En la sección 3.3 se muestra cómo al aplicar la segunda derivada a una función gaussiana; la función resultante presenta mínimos negativos en los puntos cercanos a los máximos de cada una de las bandas. Para la detección de máximos se utiliza entonces la segunda derivada invertida, la cual se obtiene de derivar el carril original, en el cual las bandas son oscuras en un fondo más claro, como se muestra en la Figura 4.5.

Aunque las posiciones correspondientes a los máximos de la segunda derivada y los cruces por cero no correspondan exactamente con los valores teóricos aun considerando funciones continuas, estos datos aproximan lo suficiente a las bandas como para estimar la cantidad de bandas en los datos considerados [14].

Este es el mismo carril utilizado en la sección anterior, con motivo de realizar la ubicación de las bandas cuando el traslape existente es mínimo. Las posiciones de los valores encontrados con respecto a los valores teóricos se encuentran en la Tabla 4.5.

Tabla 4.5: Posición de las bandas en un carril con poco traslape

Posición teórica	Posición obtenida	Posición teórica	Posición obtenida
44	44	309	309
95	95	380	380
202	202	425	425
222	222	451	451
268	268	826	826

En este caso de bandas separadas y con máximos perfectamente distinguibles la posición de-

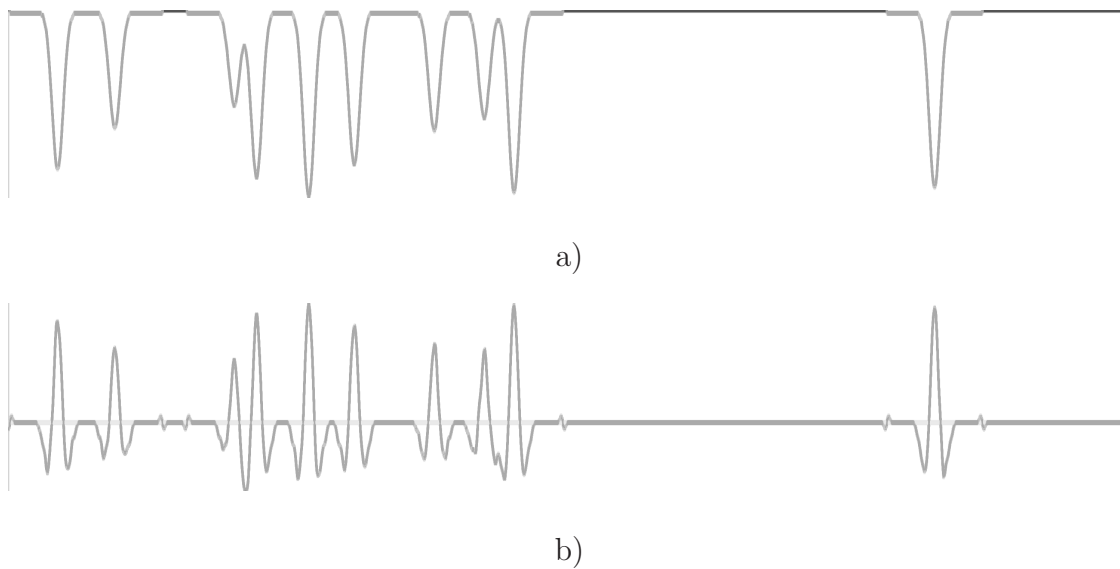


Figura 4.5: Carril teórico con poco traslape. a) Perfil del carril y b) su segunda derivada.

tectada coincide para todas las bandas con el valor teórico. Para evaluar el comportamiento de este método ante mayor nivel de traslape se utiliza un carril con 26 bandas con desviación estándar $\sigma = 5$, el cual se muestra en la Figura 4.6.

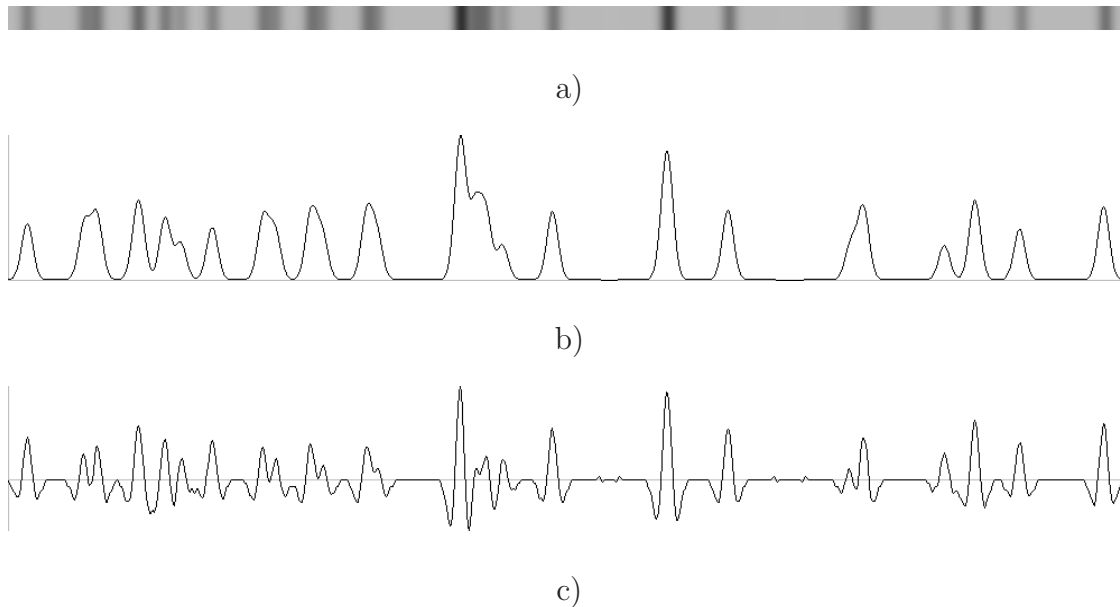


Figura 4.6: Carril teórico con mayor traslape. a) Perfil del carril, b) perfil invertido y c) su segunda derivada.

Esta imagen muestra el efecto de mejora de la resolución de los picos en el carril, lo cual permite distinguir aún aquellos que a la vista parecen como una única banda más ancha producto de la suma de dos bandas adyacentes.

Las posiciones de cada una de estas bandas se muestra en la Tabla 4.6, en la cual se muestran las posiciones con respecto a sus valores teóricos, así como el error en píxeles en que se incurre

en la determinación de la posición de cada una de estas bandas, que en el peor de los casos llega a ser de hasta 3 píxeles, esto es, 0,6 veces la desviación estándar de la banda.

Tabla 4.6: Posición de las bandas en un carril sintético con 26 bandas

Posición teórica	Posición obtenida	Error (píxeles)	Posición teórica	Posición obtenida	Error (píxeles)
17	17	0	403	403	0
70	67	3	416	418	2
78	79	1	423	426	3
116	116	0	440	441	1
140	140	0	485	485	0
154	155	1	587	587	0
182	182	0	642	642	0
228	227	1	753	750	3
237	239	2	762	763	1
271	269	2	835	835	0
279	281	2	862	862	0
321	320	1	902	902	0
327	329	2	977	977	0

4.2.2 Determinación de la cantidad de bandas en el carril

La segunda derivada permite distinguir entre bandas muy cercanas entre sí, tal como se ejemplificó en la sección anterior en el caso de carriles con traslape de bandas. Con el objetivo de analizar la determinación de la cantidad de las bandas que el algoritmo es capaz de detectar, se generan 10 carriles sintéticos para medir la cantidad de bandas detectadas para cada valor de desviación estándar empezando en 1 hasta 5,5 con paso de 0,5, cada uno con 20 bandas.

En estas mediciones interesa evaluar la detección de bandas en comparación con lo que un usuario a simple vista podría realizar, por lo tanto se genera la Tabla 4.7 con 100 mediciones donde se realizó un conteo de las bandas detectables a simple vista, y en la Tabla 4.7 se muestran los resultados de la detección de bandas utilizando la segunda derivada.

La cantidad de bandas en estos carriles es teóricamente de 20, sin embargo, el algoritmo generador de carriles sintéticos coloca algunas de las bandas en posiciones tan cercanas en algunos casos que ni siquiera con la segunda derivada se puede distinguir entre dos bandas adyacentes.

La cantidad de bandas detectadas mediante la segunda derivada mejora la resolución en varios de los casos analizados, puesto que para ninguno de los 100 carriles sintéticos analizados la cantidad de bandas estimadas a simple vista fue mayor a la cantidad estimada por el algoritmo. El aumento en la cantidad de bandas detectadas obedece a la mejora de resolución aplicada en los casos en que las bandas se traslapan y que a simple vista puede

Tabla 4.7: Evaluación de cantidad de bandas en carriles sintéticos. Cantidad de bandas observadas.

Medición σ	1	2	3	4	5	6	7	8	9	10
1	20	20	19	18	19	20	20	18	20	20
1.5	19	18	19	18	17	15	18	20	17	19
2	17	16	20	17	19	19	17	20	20	19
2.5	15	18	15	17	17	16	15	17	19	19
3	15	16	18	16	16	17	15	17	14	16
3.5	14	16	14	15	17	18	18	14	14	16
4	14	19	17	15	17	20	18	18	17	17
4.5	18	18	15	19	18	19	17	16	18	16
5	16	17	16	17	13	17	16	15	16	16
5.5	18	15	16	16	18	15	14	14	17	15

Tabla 4.8: Evaluación de cantidad de bandas en carriles sintéticos. Cantidad de bandas obtenidas mediante la segunda derivada.

Medición σ	1	2	3	4	5	6	7	8	9	10
1	20	20	19	18	19	20	20	18	20	20
1.5	19	18	19	18	17	16	18	20	17	19
2	17	16	20	17	19	19	17	20	20	19
2.5	16	18	15	17	18	16	15	19	19	19
3	16	17	18	17	16	17	16	18	16	16
3.5	15	16	14	15	17	18	19	15	15	16
4	15	20	17	16	18	20	19	18	17	18
4.5	18	19	15	20	18	19	17	16	18	16
5	17	18	16	18	13	17	16	15	17	16
5.5	19	15	16	18	19	16	15	15	17	16

parecer como una única banda más ancha en comparación con las demás, por lo tanto se considera satisfecho el criterio de identificación de más del 95% de las bandas del carril con respecto a la cantidad de bandas observadas, dado que es imposible determinar en este caso la totalidad de las 20 bandas a menos que no estén muy traslapadas.

4.3 Determinación de la desviación estándar con Line Search y espacio de escalas

La localización de las bandas por medio de la segunda derivada permite únicamente conocer la ubicación, sin embargo es necesario obtener el valor correspondiente a la desviación

estándar considerando todas las bandas encontradas. En este trabajo se asume que las bandas de los carriles poseen el mismo ancho.

Una vez conocidas las ubicaciones de las diferentes bandas, se realiza la optimización de la función objetivo, centrando las bandas en los máximos detectados, de forma que el error obtenido sea el mínimo al realizar el ajuste para diferentes valores de desviación estándar. El valor inicial de desviación estándar utilizado para aproximar estas bandas es el resultado de la aplicación del espacio de escalas.

El ajuste realiza la minimización del error en la aproximación mediante mínimos cuadrados utilizando el optimizador *LineSearch* en una dimensión. En la Tabla 4.9 se encuentran los resultados correspondientes a la aplicación de la optimización de la función objetivo considerando carriles sintéticos con una banda.

Tabla 4.9: Estimación de desviación estándar utilizando la función objetivo para carriles teóricos con una banda

σ teórico	σ promedio	C.V. (%)	Error (%) promedio	Error máximo 1 píxel (%)
0,5	0,578	16,239	15,597	33,333
1	1,043	2,661	4,270	16,667
1,5	1,548	1,983	3,230	11,111
2	2,041	0,981	2,055	8,333
2,5	2,537	0,574	1,471	6,667
3	3,034	0,478	1,128	5,556
3,5	3,535	0,430	0,995	4,762
4	4,029	0,384	0,736	4,167
4,5	4,535	0,411	0,779	3,704
5	5,045	0,432	0,895	3,333
5,5	5,537	0,355	0,667	3,030
6	6,036	0,531	0,596	2,778
6,5	6,564	0,456	0,987	2,564
7	7,037	0,534	0,522	2,381
7,5	7,540	0,465	0,528	2,222
8	8,030	0,640	0,374	2,083
8,5	8,516	0,315	0,183	1,961
9	9,027	0,378	0,303	1,852
9,5	9,530	0,233	0,314	1,754
10	10,051	0,635	0,506	1,667

Para cada valor de desviación estándar se realiza el cálculo utilizando 10 carriles sintéticos y se obtienen los valores medios de la desviación estándar y del error con respecto al valor teórico.

El error obtenido para las bandas con menor desviación estándar se reduce en comparación con el obtenido en la Tabla 4.1 utilizando solamente el espacio de escalas, tal como se

muestra en la Figura 4.7, donde destaca que para $\sigma = 0,5$ el error pasa de 113,20% a 15,597%. Además se incluye el error máximo permitido en la última columna, el cual es de 1 píxel para cada una de las bandas, donde se nota que para el método propuesto las bandas pueden ser aproximadas con error menor a este límite. Este error máximo está determinado por la siguiente ecuación

$$e_{max} = \frac{1}{6\sigma} 100\% \quad (4.2)$$

La Figura 4.7 muestra la mejora en la aproximación de la desviación estándar de las bandas utilizando el espacio de escalas en conjunto con el método de mínimos cuadrados en comparación con la utilización simple del espacio de escalas para carriles con una única banda. En este gráfico sí se incluye el error para las bandas con $\sigma = 0,5$ puesto que aunque para el espacio de escalas este error es de 113,2%, permite realizar la comparación con el error generado con el método propuesto para este ancho de la banda.

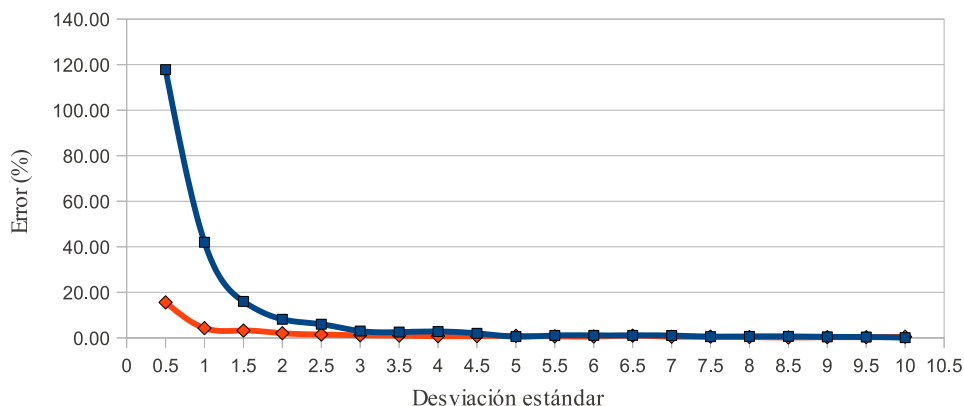


Figura 4.7: Error en la determinación de la desviación estándar en carriles con una banda. Error con espacio de escalas (□), error con espacio de escalas y mínimos cuadrados (◇).

Aplicando este algoritmo de aproximación a los carriles analizados en la sección anterior, se obtienen los resultados de la Tabla 4.10, donde se muestra la desviación estándar obtenida cuando se consideran todas las bandas del carril detectadas por la segunda derivada. En esta tabla se realiza una comparación entre los métodos de espacio de escalas y el método iterativo propuesto con espacio de escalas y lineSearch.

Tabla 4.10: Aproximación de la desviación estándar por ajuste de función objetivo

Caso	Carril	Desviación estándar (Espacio de escalas)	Error (%)	Desviación estándar (Espacio de escalas + LineSearch)	Error (%)
1	Figura 4.2	4,98	0,40	5,051	1,02
2	Figura 4.3	3,34	33,2	4,935	1,30
3	Figura 4.6	4,98	0,40	4,904	1,92

Los datos obtenidos en la tabla anterior muestran que la desviación estándar obtenida mediante el espacio de escalas puede ser más cercana al valor teórico que para el caso de la

aproximación con espacio de escalas y mínimos cuadrados. Sin embargo, el resultado del espacio de escalas incluye solamente a las bandas que no se encuentran traslapadas. Los valores con alta exactitud obtenidos mediante el espacio de escalas en los casos 1 y 2 de la Tabla 4.11 se deben a las bandas detectadas en la ventana de $\sigma = 1$ que se considera en cada caso, centrada en el punto más probable detectado y como existe una alta cantidad de bandas sin traslape estas dominan el valor global de la desviación estándar para el carril.

Tabla 4.11: Detección de bandas y desviación estándar en bandas traslapadas con espacio de escalas

Caso	Carril	Desviación estándar (Espacio de escalas)	Bandas Totales	Limites de la ventana	Bandas detectadas
1	Figura 4.2	4,98	10	4,48-5,48	9
2	Figura 4.3	3,34	9	2,84-3,84	3
3	Figura 4.6	4,98	26	4,48-5,48	9

Sin embargo, en el caso 2 solo existe una banda sin traslape por lo cual el valor dominante de desviación estándar se encuentra en 3,34, alejado del valor teórico donde además solamente se detectan 3 bandas en la ventana.

En los datos de la Tabla 4.10 los resultados obtenidos consideran todas las bandas pues la segunda derivada detectó la totalidad de estas en los tres casos, por lo que las aproximaciones de la desviación estándar toman en cuenta a todas las bandas del carril. Esto muestra que el método propuesto presenta menor error en la caracterización de las bandas aún cuando estas se encuentran severamente traslapadas.

En una evaluación más general del comportamiento de este método de detección de desviación estándar se generan carriles sintéticos con 20 bandas cada uno, para valores de desviación estándar de 1 a 10 con paso de 1, considerando 5 carriles por cada valor de desviación estándar. La Tabla 4.12 muestra los resultados.

En estos datos se observa que el error en la desviación estándar es más alto que para el caso de los carriles con banda única, lo cual se debe a que en los carriles sintéticos generados la posición de las bandas es aleatoria y cuando existen bandas traslapadas que no pueden ser distinguidas el algoritmo no aproxima correctamente la banda resultante aumentando el error. Este comportamiento aleatorio se indica en el coeficiente de variación de estos resultados, que es mucho más alto que en el caso de banda única.

Sin embargo, el algoritmo puede ser evaluado de manera más *justa* permitiendo que la aproximación pueda realizarse considerando todas las bandas del carril, de forma que se evite la aproximación de bandas más anchas a su valor real, como en el caso de los datos obtenidos en la Tabla 4.12, donde la segunda derivada ya no es capaz de distinguir entre bandas generadas con posición completamente aleatoria.

En la Tabla 4.13 se muestran los resultados para la evaluación del algoritmo considerando desviaciones estándar de 0,5 a 10, con paso de 0,5, para 10 carriles con 20 bandas para cada valor de desviación estándar. Los carriles generados tienen la restricción de una separación

Tabla 4.12: Evaluación de la desviación estándar utilizando carriles sintéticos con 20 bandas.

Medición σ	1	2	3	4	5	Promedio	CV (%)	Error (%)
1,0	1,091	1,065	1,107	1,194	1,056	1,103	4,987	10,260
2,0	2,135	2,198	2,069	2,115	2,056	2,115	2,684	5,730
3,0	3,045	3,064	2,958	2,838	2,877	2,956	3,372	1,453
4,0	4,395	3,919	4,069	3,846	4,184	4,083	5,350	2,065
5,0	5,252	5,015	5,316	5,174	5,129	5,177	2,234	3,544
6,0	6,067	6,505	6,183	6,273	5,793	6,164	4,258	2,737
7,0	6,984	7,054	7,474	6,968	7,421	7,180	3,438	2,574
8,0	8,057	8,058	8,31	8,551	7,941	8,183	3,003	2,293
9,0	9,143	8,738	9,323	9,052	7,981	8,847	5,976	1,696
10,0	9,788	11,283	10,893	10,582	9,472	10,404	7,275	4,036

mínima entre bandas de 2σ , de forma que todas las bandas puedan ser distinguidas.

Tabla 4.13: Evaluación de la desviación estándar utilizando carriles sintéticos con 20 bandas traslapadas. Aproximación del espacio de escalas y minimización lineal.

σ teórico	σ Promedio	C.V. (%)	Error (%)	σ teórico	σ Promedio	C.V. (%)	Error (%)
0,5	0,645	3,277	20,137	5,5	5,459	0,907	1,973
1	1,149	3,490	8,052	6	5,974	0,774	0,160
1,5	1,526	3,369	3,473	6,5	6,440	0,791	0,226
2	2,018	1,426	0,310	7	6,967	1,032	0,350
2,5	2,515	0,883	2,576	7,5	7,477	0,384	0,567
3	3,003	0,583	0,042	8	7,990	0,473	0,518
3,5	3,523	0,298	0,611	8,5	8,480	0,436	0,350
4	3,978	0,767	0,484	9	8,986	0,467	0,309
4,5	4,484	1,432	0,330	9,5	9,476	0,549	0,376
5	4,974	0,769	0,419	10	9,921	0,490	0,945

Esta tabla permite comparar con los resultados mostrados en la Tabla 4.4 para el espacio de escala. La desviación estándar está dentro de los límites aceptables en este caso, puesto que todos los errores obtenidos son menores a un píxel, y los coeficientes de variación son menores a 3.277%, lo cual indica que los datos son precisos y estables utilizando este algoritmo, aún si las bandas se encuentran traslapadas.

4.4 Mejora en ubicación de la posición utilizando Downhill Simplex

La ubicación de los centros de las bandas utilizando la segunda derivada tiene el inconveniente que cuando el ancho de las bandas aumenta y el traslape es severo, la posición de los picos detectados se desvían del valor real como se ejemplificó en la sección 4.2.1.

Por esta razón se utiliza el valor obtenido de la desviación estándar mediante *LineSearch* y las posiciones calculadas con la segunda derivada para mejorar la aproximación de la posición de las bandas en el carril. Ejemplificando el trabajo del algoritmo se evalúa de nuevo para el carril de la Figura 4.6.

En esta evaluación se eligió de nuevo el mismo carril utilizado anteriormente por el traslape severo existente en ese carril como caso especial. En la Figura 4.8 se muestra el error obtenido en píxeles para las bandas del carril, que como se muestra no excede el límite máximo considerado en este trabajo que es de 2 píxeles.

Tabla 4.14: Posición aproximada de las bandas en un carril sintético con 26 bandas utilizando Downhill Simplex

Posición teórica	Posición obtenida	Posición teórica	Posición obtenida
17	16,77	403	402,57
70	68,67	416	415,52
78	78,87	423	424,17
116	115,67	440	441,54
140	140,25	485	485,30
154	154,14	587	587,55
182	181,73	642	641,84
228	227,21	753	752,26
237	237,76	762	762,87
271	269,97	835	834,74
279	279,94	862	862,26
321	321,45	902	902,08
327	327,88	977	977,03

4.5 Evaluación de carriles reales

El método de *LineSearch* necesita que las evaluaciones de la función objetivo tomen como base un carril sin distorsiones donde solamente se encuentren presentes las bandas, por lo que es necesario remover el fondo no deseado de los carriles reales.

Antes de utilizar el ajuste de mínimos cuadrados se remueve el fondo de los carriles de

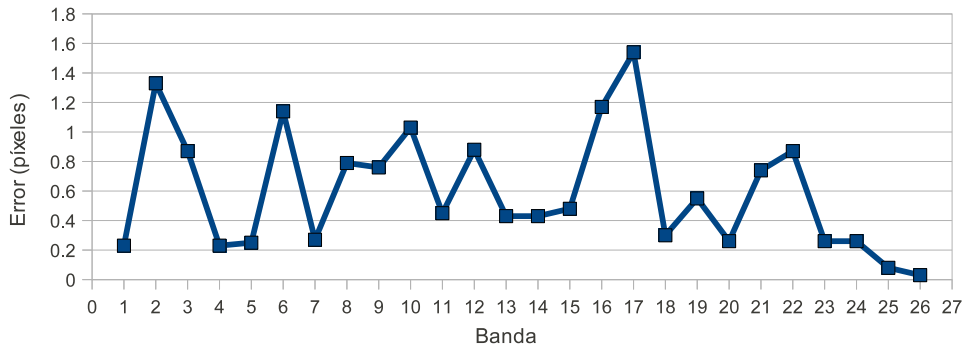


Figura 4.8: Error en la determinación de la posición utilizando Downhill Simplex.

acuerdo con la desviación estándar calculada mediante el espacio de escalas, por lo que antes de realizar el ajuste de la función objetivo se aplica un filtrado al carril real, para utilizarlo luego como entrada al programa. El método de filtrado de fondo utilizado es el desarrollado en [6].

La Figura 4.9 presenta el proceso de ajuste mediante mínimos cuadrados. Para un carril real a) que presenta distorsión por fondo indeseado, cuya imagen de perfil es b) donde se aprecia las regiones que no son banda. A este perfil se le realiza la extracción del fondo con lo cual se obtiene el nuevo perfil filtrado c). Finalmente el ajuste de mínimos cuadrados se realiza con un vector simulado d) con bandas es las posiciones encontradas mediante la segunda derivada, de forma que se minimice la función objetivo con c) como referencia.

Puesto que el valor de la desviación estándar obtenida con el espacio de escalas puede estar alejada del valor real, se realiza un proceso iterativo utilizando el σ calculado en la iteración anterior del *LineSearch* para filtrar nuevamente el fondo de carril y aproximarlos por sumatoria de gaussianas hasta que se establezca el valor de la desviación estándar obtenida.

Así el algoritmo propuesto en este trabajo tiene como resultado para el carril de la Figura 4.9 la detección de 50 bandas y una desviación estándar de 1.813 para el mejor ajuste.

4.6 Rendimiento del algoritmo

En el trabajo realizado se utiliza la mejora de la posición de las bandas utilizando el método de *Downhill Simplex*, por lo cual se evalúa el rendimiento del algoritmo implementado con relación al algoritmo utilizado en [21], el cual utiliza además algoritmos genéticos para la estimación de las bandas.

Esta evaluación consta del estudio de tres carriles que permitan comparar el comportamiento de ambos algoritmos con carriles con distinta aglomeración de bandas. Se utiliza un carril con una única banda el cual se muestra en la Figura 4.10, un carril con 10 bandas mostrado en la Figura 4.11 y un carril real con la mayor aglomeración de bandas observada, para lo cual se utiliza el carril de la Figura 4.9 de la sección anterior.

En la evaluación se utilizaron los mismos tres carriles para evaluar ambos algoritmos. Puesto

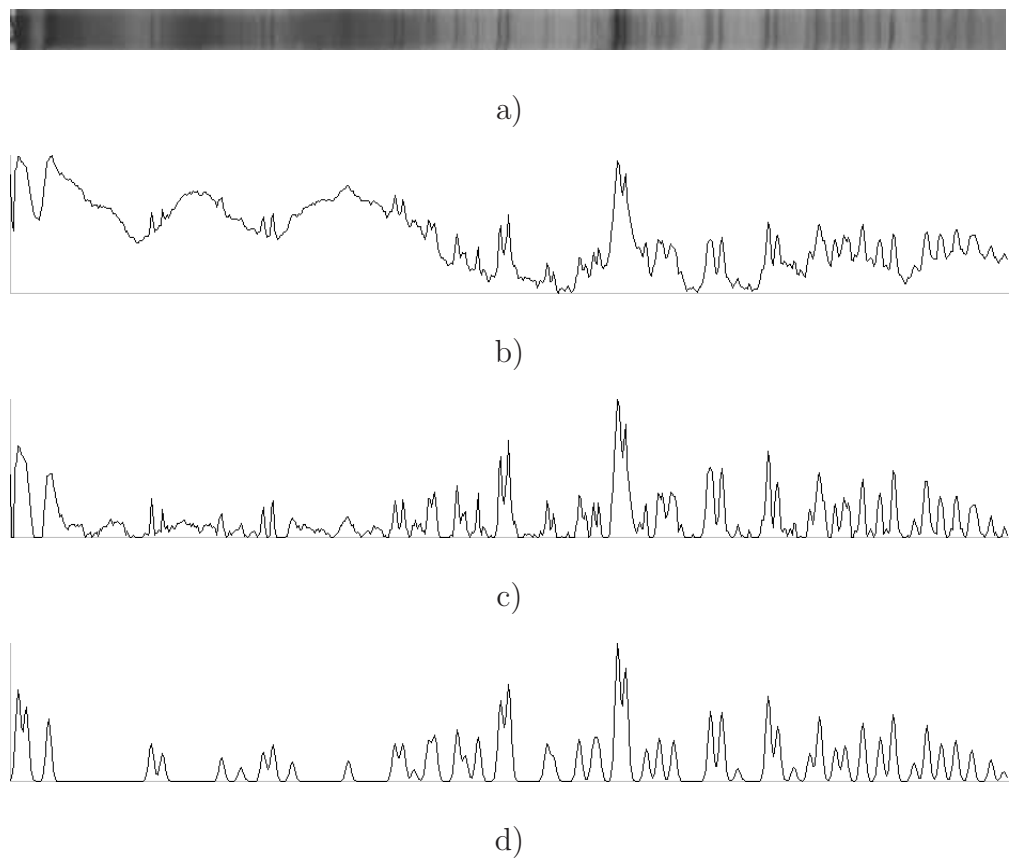


Figura 4.9: Proceso de ajuste del carril mediante mínimos cuadrados. a) Carril real, b) perfil, c) perfil sin fondo, d) perfil simulado de mejor ajuste.



Figura 4.10: Carril con una única banda utilizado para evaluar rendimiento.

que en el presente trabajo se utiliza el *Downhill Simplex* para optimizar la posición de las bandas y la división del carril en ventanas, se evalúan ambos algoritmos para ventanas de igual ancho. En el caso del método que utiliza algoritmos genéticos se da como valor de entrada la cantidad de bandas a considerar en cada ventana, como se muestra en la Tabla 4.15.

Tabla 4.15: Evaluación del rendimiento del algoritmo de [21] con *Downhill Simplex* y algoritmos genéticos.

Caso	Carril	Ancho del carril (píxeles)	Ancho de la ventana (píxeles)	Iteraciones	Tiempo (s)	Bandas por ventana
1	Figura 4.10	1000	100	417203	4,43	1
2	Figura 4.11	1182	60	6255996	102,45	2
3	Figura 4.9	623	60	13003209	214,93	3

Las iteraciones mostradas en estos resultados corresponden a la cantidad de veces que la función objetivo es evaluada por el algoritmo *Downhill Simplex*, considerando todas las



Figura 4.11: Carril con 10 bandas utilizado para evaluar rendimiento.

ventanas para un carril. En la tabla 4.16 los datos mostrados corresponden a los carriles evaluados con el método propuesto en este trabajo.

Tabla 4.16: Evaluación del rendimiento del algoritmo propuesto en este trabajo.

Caso	Carril	Ancho del carril (píxeles)	Ancho de la ventana (píxeles)	Iteraciones	Tiempo (s)
1	Figura 4.10	1000	100	4385	1,75
2	Figura 4.11	1182	60	29507	2,55
3	Figura 4.9	623	60	61572	5,25

Cabe resaltar que en el caso del primer algoritmo la desviación estándar de las bandas es un parámetro de entrada, por lo que el tiempo utilizado en cálculos es solamente para la ubicación de las bandas. Los tiempos obtenidos en la Tabla 4.16 para el algoritmo propuesto consideran tanto el cálculo de la desviación estándar como la mejora en la aproximación de la posición.

La disminución del tiempo de cálculo al estimar la posición inicial de las bandas con la segunda derivada permite disminuir el tiempo de cálculo en 40,8 veces para el carril del caso No. 3. y 40,2 veces para el caso 2, cuando el algoritmo debe considerar más de una banda en el carril. Para el caso No. 1 la mejora en tiempo de cálculo no es tan alta (2,53 veces), pero esto se debe a que como solamente una banda se encuentra en el carril, es una ó máximo dos ventanas las analizadas; las demás no tienen datos, por lo que son descartadas.

Capítulo 5

Conclusiones

5.1 Conclusiones

El espacio de escalas no permite conocer exactamente la posición de las bandas en un carril, por lo que es necesario utilizar una estrategia de aproximación para ubicarlos en los puntos correspondientes al máximo de la banda, lo cual se logró en el este trabajo al implementar un optimización multidimensional de los datos iniciales.

El ajuste que utiliza minimización lineal en conjunto con la segunda derivada para posición de las bandas permite aproximar mejor el valor de la desviación estándar, ya que permite tomar en cuenta las bandas del carril que no serían distinguibles por el traslape entre ellas. Se comprobó que cuando pueden detectarse las bandas presentes en el carril la desviación estándar será aproximada correctamente con error menor a 1 píxel permitido aún cuando existe traslape.

La detección de bandas con el espacio de escalas debe realizarse en la mayor resolución posible para evitar la pérdida de máximos por fusión con otras bandas cercanas al aumentar la escala.

Para obtener un sistema automatizado es necesario utilizar un método que brinde una aproximación inicial de la desviación estándar, como es en este caso el espacio de escalas, porque se necesita un valor cercano al real para realizar el filtrado inicial antes de realizar las aproximaciones de la función objetivo con el carril filtrado, y así corregir el valor de la desviación estándar.

La detección de la posición inicial de las bandas permite disminuir considerablemente el tiempo necesario para la caracterización de las bandas, en especial cuando las ventanas analizadas contienen varias bandas.

5.2 Recomendaciones

La aproximación del error disminuye cuando se utiliza minimización lineal de la función de error, pero se recomienda tener en cuenta la tendencia estadística del error en la aproximación de las bandas en futuros trabajos.

Aunque la segunda derivada es un método que efectivamente permite mejorar la resolución de los datos para bandas con traslape, depende de filtrado de ruido para operar apropiadamente, por lo que se recomienda para la mejora de esta resolución la implementación de algún otro algoritmo menos sensible al ruido en las imágenes de los carriles.

Bibliografía

- [1] José Antonio Aguilar. Detección y corrección del efecto sonrisa en imágenes de geles de electroforesis utilizando modelos activos de forma acoplados. Tesis de licenciatura, Escuela de Ingeniería Electrónica, ITCR, Cartago, Costa Rica, Abril 2010.
- [2] Miguel Angel Aguilar. Desarrollo de una etapa de preprocesamiento para reducción de ruido, mejoramiento de contraste y de nitidez en imágenes digitalizadas de geles de electroforesis. Tesis de licenciatura, Escuela de Ingeniería Electrónica, ITCR, Cartago, Costa Rica, 2007.
- [3] D. F. Allen, J. S. Bashkin, Hong Guo, S. E. Shadle, W. K. Pogozelski, and T. D. Tullius. Quantitative analysis of electrophoresis data: novel curve fitting methodology and its application to the determination of a protein–dna binding constant. *Nucleic Acids Research*, Vol25, No.4, 1997.
- [4] Pedro Alpízar. Optimización de la corrección del efecto sonrisa en imágenes de geles de electroforesis. Tesis de licenciatura, Escuela de Ingeniería Electrónica, ITCR, Cartago, Costa Rica, Diciembre 2010.
- [5] P. Alvarado. Modelos de sistemas. Notas de clase, 2008. URL <http://www.ie.itcr.ac.cr/palvarado/Modelos/modelos.pdf>.
- [6] P. Alvarado, A. Salazar, O. Murillo, F. Rojas, and J. Peraza. Análisis por computador de imágenes de geles de electroforesis para la caracterización molecular de organismos. Propuesta de proyecto de investigación, Escuela de Ingeniería Electrónica, ITCR, Cartago, Costa Rica, 2009.
- [7] J. M. Berg, J. L. Tymoczko, and L. Stryer. *Biochemistry*. W. H. Freeman and Company, 5th edition, 2002. URL <http://www.ncbi.nlm.nih.gov/books/NBK21154/>.
- [8] RWTH-Aachen Chair of Technical Computer Science. Lti-lib2 [online]. 2011 [visitado el 20 de Junio de 2011]. URL <http://www.ie.itcr.ac.cr/palvarado/ltilib-2/homepage/>.
- [9] C. G. Enke and T. A. Nieman. Signal to noise ratio enhancement by least-squares polynomial smoothing. In *Analytical Chemistry*. 1976.
- [10] A. Felinger. *Data Analysis and Signal Processing in Chromatography*. Elsevier, 1998.

- [11] Edison Fernández. Estimación de posición, cantidad y ancho de bandas en carriles de imágenes de electroforesis en gel utilizando espacios de escala. Tesis de licenciatura, Escuela de Ingeniería Electrónica, ITCR, Cartago, Costa Rica, Noviembre 2010.
- [12] B. P. Flannery, W. H. Press, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press, 3rd edition, 2007. URL <http://apps.nrbook.com/openaccess/index.html>.
- [13] P. Gorry. Curve fitting by a sum of general least-squares smoothing and differentiation by the convolution (savitzky-golay) method. *American Chemical Society, Analytical Chemistry, Vol. 62, No. 6*, 1990.
- [14] A. Goshtasby and W. D. O'Neill. Curve fitting by a sum of gaussians. *Graphical Models and Image Processing*, 1994. URL <http://www.engineering.wright.edu/~agoshtas/GMIP94.pdf>.
- [15] Tony Lindeberg. Scale space for discrete signals. Technical report, Computer Vision and Associative Pattern Processing Laboratory (CVAP), Royal Institute of Technology, S-100 44 Stockholm, Sweden.
- [16] Tony Lindeberg. Automatic scale selection as a pre-processing stage for interpreting the visual world. Technical report, Department of Numerical Analysis and Computing Science KTH, Royal Institute of Technology, S-100 44 Stockholm, Sweden, 1999.
- [17] Tony Lindeberg and Lars Bretzner. Real-time scale selection in hybrid multi-scale representations. Technical report, Computational Vision and Active Perception Laboratory (CVAP), Department of Numerical Analysis and Computer Science (KTH), Royal Institute of Technology, S-100 44 Stockholm, Sweden, 2003.
- [18] John A. Luckey, Tracy B. Norris, and Lloyd M. Smith. Analysis of resolution in dna sequencing by capillary gel electrophoresis. *The Journal of Physical Chemistry, Vol. 97, No. 12.*, 1993.
- [19] Douglas C. Montgomery and George C. Runger. *Probabilidad y Estadística Aplicadas a la Ingeniería*. Limusa Wiley, 2nd edition, 2005.
- [20] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal, vol. 7, pp. 308–313*, 1965.
- [21] David Soto. Detección automática de bandas en imágenes de geles de electroforesis por medio de la optimización de una función objetivo. Tesis de licenciatura, Escuela de Ingeniería Electrónica, ITCR, Cartago, Costa Rica, Junio 2010.