

# Métodos de Reconstrucción de Superficies Corneales mediante Funciones de Base Radial



Darío Ramos López  
Tutor: Andrei Martínez Finkelshtein

Almería, 13 de septiembre de 2011

Trabajo Final del Máster  
Interuniversitario en Matemáticas

Universidad de Almería



# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Preliminares</b>	<b>13</b>
2.1. Topografías corneales . . . . .	13
2.2. Polinomios de Zernike . . . . .	18
2.3. Funciones de Base Radial (RBF) . . . . .	23
2.4. RBF Anisótropas . . . . .	26
<b>3. Ajuste directo. Regularización de Tikhonov</b>	<b>29</b>
3.1. Pre-ajuste . . . . .	31
3.2. Ajuste directo por RBFs . . . . .	35
3.3. Regularización de Tikhonov . . . . .	39
3.3.1. Cross-Validación Generalizada (GCV)	45
3.3.2. SVD Compacta . . . . .	54
3.3.3. Uso eficiente de GCV . . . . .	57
3.3.4. L-Curve . . . . .	69
3.4. Resultados Experimentales . . . . .	72

<b>4. Ajuste Iterativo Adaptativo</b>	<b>85</b>
4.1. Pre-ajuste . . . . .	87
4.2. Procedimiento de ajuste . . . . .	88
4.3. Descripción del algoritmo iterativo . . . . .	92
4.4. Criterio de parada . . . . .	101
4.5. Resultados Experimentales . . . . .	106
<b>5. Conclusiones</b>	<b>121</b>
<b>Índice de figuras</b>	<b>131</b>
<b>Bibliografía</b>	<b>135</b>

# Capítulo 1

## Introducción

El objetivo de este trabajo es presentar dos nuevos métodos generales de reconstrucción de superficies. Ambos han sido verificados experimentalmente en superficies ópticas, como la de la córnea, mostrando buenas propiedades numéricas. Los dos métodos que se desarrollan a lo largo del trabajo están basados en las funciones de base radial (RBF) gaussianas, estándar en el primero de ellos (“ajuste directo”) y anisótropas en el segundo (“ajuste iterativo adaptativo”).

Este trabajo está motivado por la necesidad que hay ac-

tualmente de un modelado fiable y preciso de las superficies de la córnea, necesario tanto por motivos tecnológicos como clínicos.

Dada la importancia de la forma de la cara anterior (externa) de la córnea en la refracción del ojo [36] y la capacidad de corregir errores refractivos mediante ablación con láser de dicha superficie, es crucial un análisis detallado del frente de onda de cada topografía corneal concreta, como base para los clínicos para proporcionar un tratamiento personalizado. Se sabe que la cara anterior de la córnea proporciona normalmente la mayor parte de las aberraciones oculares en los ojos patológicos u operados [5].

El modelado corneal puede servir asimismo como herramienta para visualizar enfermedades o patologías corneales. El queratocono, por ejemplo, es una enfermedad ocular que distorsiona la forma de la córnea y como resultado, provoca

una pérdida significativa de visión (incluso total). Los pacientes sospechosos de padecer queratocono deben ser inspeccionados previamente a cualquier cirugía refractiva, porque tal tratamiento puede empeorar la forma corneal y conducir incluso al trasplante de córnea. Por tanto, el modelado de la córnea juega un papel esencial en el diagnóstico y observación del queratocono, para garantizar que un paciente es adecuado para el tratamiento e impedir casos de cirugía refractiva inapropiados [26]. Además, hay que tener en cuenta el importante rol de las herramientas fiables de visualización en la práctica clínica.

Las técnicas modernas de diseño y ajuste de lentes de contacto blandas pueden tener en consideración características del ojo del paciente, adaptado la superficie interna de la lente para que se adapte a la forma concreta de su córnea. Este método requiere, de nuevo, un detallado análisis de la

topografía de la cara anterior de la córnea.

Con la introducción de la videoqueratoscopía (visualización de la córnea) de alta velocidad [19, 30] en el estudio de la dinámica de la topografía de la superficie corneal [45] y la estabilidad de la capa lagrimal [21], el almacenamiento de la información ha comenzado a ser significativo, motivando otra importante aplicación del modelado de las superficies corneales: la compresión de datos [35].

La inmensa mayoría de los topógrafos corneales modernos recogen los datos (ya sean de elevación, curvatura, desplazamiento de anillas u otros) en un conjunto finito y discreto de puntos. Típicamente, estos puntos tienen una estructura o “casi-estructura”; por ejemplo, los dispositivos basados en la tecnología de discos de Plácido proporcionan elevaciones y curvaturas en las imágenes discretizadas de las anillas, cuya deformación es proporcional a la complejidad de la superfi-



cie (esto se explica algo más detalladamente en el Capítulo 2). En cualquier caso, los datos están contaminados de error, que procede de varias fuentes: el ruido natural del dispositivo, errores de medida y digitalización, errores de los algoritmos utilizados (como los que convierten el desplazamiento de anillas en alturas), errores de redondeo y otros. Por tanto, nos enfrentamos al problema del ajuste de la superficie real contaminada por ruido, con la menor cantidad de coeficientes o parámetros posible, para sus posteriores aplicaciones clínicas y tecnológicas.

La solución a este problema normalmente se clasifica en los llamados métodos zonales y modales. En los primeros, los métodos zonales, el dominio donde están los datos es dividido en subdominios más elementales (por ejemplo, triángulos) y la superficie se aproxima en cada uno de los subdominios con cierta independencia del resto. La herramienta más común

aquí son los splines, en particular, los B-splines [4, 16, 25], que son numéricamente estables y ampliamente utilizados en el Diseño Geométrico Asistido por Ordenador (en inglés, CAGD).

En los métodos modales de reconstrucción, en cambio, la aproximación a la superficie está dada por una combinación (casi siempre lineal) de funciones básicas definidas globalmente en todo el dominio de los datos, pudiendo depender éstas de un cierto número de parámetros. En este sentido, las decisiones más importantes a tomar son qué tipo de funciones utilizar, cuales serán sus parámetros y la cantidad necesaria de ellas para recuperar la información relevante de la superficie, tratando de evitar el sobre-ajuste (ya que en este caso, estaríamos ajustado el error de medición de los mismos).

Entre las ventajas de los métodos zonales están la flexibilidad y la precisión del ajuste, pero adolecen de la simplicidad

del enfoque modal, que proporciona expresiones funcionales válidas en todo el dominio, adecuadas para posibles cálculos futuros (como trazado de rayos u otros). Las técnicas zonales son además más exigentes computacionalmente, y codifican la forma final de la superficie en una cantidad grande de parámetros. Además, en algunos métodos modales las funciones básicas utilizadas son interpretables en términos físicos, ópticos o clínicos, pero esto no ocurre en el caso de los zonales.

Una base de funciones estándar para la reconstrucción modal, que se usa continuamente en oftalmología para representar el error de frente de onda ocular, son los polinomios de Zernike [24]. Los coeficientes de su expansión tienen interpretación en término de las aberraciones básicas de la visión, como son el desenfoque, astigmatismo, coma, trifolio o aberración esférica, junto con otras aberraciones de alto orden [11, 42].

Como procedimiento de ajuste, los polinomios de Zernike no se limitan al análisis de superficies de error de frente de onda, si no que se pueden utilizar también para representar otras superficies oculares, incluyendo la cara anterior de la córnea [9, 38]. Se ha sugerido que el análisis con polinomios de Zernike puede ser aplicable en el desarrollo de herramientas de diagnóstico mediante topografía corneal, utilizando los coeficientes de Zernike como datos de entrada para la clasificación corneal con redes neuronales [10, 37], reemplazando o complementando los índices corneales de clasificación que hay actualmente disponibles en muchos de los dispositivos de topografía.

Sin embargo, las limitaciones potenciales de este enfoque han sido discutidas en la literatura [24, 39]. Hay una creciente preocupación de que el método de ajuste por Zernikes podría ser impreciso en situaciones anómalas. Además, es muy difícil

evaluar a priori cuántos términos son necesarios para obtener una precisión adecuada en la reconstrucción por Zernikes, para una córnea dada [20]. Se sabe [39] que limitar el análisis de Zernike a órdenes bajos puede ser incorrecto a la hora de evaluar la severidad de los estados más avanzados del queratocono [36]. Esta información se necesita particularmente en el análisis discriminante de marcadores de la enfermedad, o cuando se eligen los datos de entrada numéricos para el software de diagnóstico basado en redes neuronales como la clasificación corneal y para graduar la gravedad de una enfermedad.

En este sentido, recientemente se han sugerido múltiples alternativas al ajuste modal por mínimos cuadrados con polinomios de Zernike. Algunas de ellas intentan combinar los enfoques modal y zonal para tratar de preservar las mejores cualidades de ambos enfoques [14], o bien usar métodos no

lineales [35].

La idea de obtener la precisión y flexibilidad de los métodos zonales dentro del marco del enfoque modal, por medio de las funciones de base radial, ha sido expuesta con anterioridad [27].

El Capítulo 2 contiene los preliminares matemáticos necesarios para desarrollar el núcleo de este trabajo (Capítulos 3 y 4). En él se dan unas pocas pinceladas sobre el funcionamiento de los topógrafos corneales. A continuación, se definen los polinomios de Zernike (la base estándar utilizada en problemas de óptica y visión), con fines comparativos, y se dan algunas nociones de las funciones de base radial (RBF), así como su generalización a RBF anisótropas.

En el “ajuste directo” (Capítulo 3) se combinan varias técnicas matemáticas previamente existentes, pero que no habían sido utilizadas conjuntamente. Como resultado de este

algoritmo de ajuste se ha obtenido la patente [28]. En resumen, se ajusta por mínimos cuadrados un conjunto fijo de funciones de base radial gaussianas (RBGF, de “radial basis gaussian function”) [15] a los datos. Debido a la inestabilidad numérica de la resolución de este tipo de problemas, cuando el número de funciones crece, es necesario utilizar alguna técnica de regularización. En este caso hemos escogido la regularización de Tikhonov, que depende de un nuevo parámetro que hay que estimar. Este parámetro se selecciona mediante el criterio estadístico de cross-validación generalizada (GCV, de “generalized cross-validation”), aunque también existen criterios como L-Curve y otros. El procedimiento GCV es en general muy costoso computacionalmente, pero veremos como la descomposición en valores singulares (SVD, de “singular value decomposition”) compacta permite utilizarlo con gran eficiencia en este caso.

En el caso del “ajuste iterativo adaptativo” (Capítulo 4), el procedimiento es original. Se trata de un algoritmo adaptativo y multi-escala, que intenta ajustar la superficie con el menor número posible de parámetros libres. Este enfoque ha sido publicado en su versión original en [29] y una versión refinada en [31]. Se trata de un algoritmo iterativo que trata de describir la superficie corneal por medio de una combinación lineal de funciones de base radial anisótropas, de tipo gaussiano (A-RBGF, de “anisotropic radial basis gaussian function”).



# Capítulo 2

## Preliminares

### 2.1. Topografías corneales

Aunque no es el objetivo prioritario de este trabajo, vamos a dar una breve descripción de cómo se toman los datos de la superficie de la córnea.

Un topógrafo corneal es un instrumento que obtiene in vivo una cierta cantidad de puntos de la superficie de la córnea de un paciente. En cada punto, es posible devolver el valor de la altimetría (elevación respecto de un plano), la curvatura de

la córnea en dicho punto, o el poder refractivo corneal (medido en dioptrías). Existen varios tipos de topógrafos corneales en el mercado: los que están basados en los discos de Plácido, los que utilizan la cámara Scheimpflug, aquellos basados en la tomografía óptica coherente (OCT) y otros, siendo el primer grupo el más común. Cada uno de ellos realmente mide una característica distinta (altimetría, curvatura, etc) y calcula el resto de ellas a partir de la medida inicial, utilizando ciertos algoritmos.

En un topógrafo de discos de Plácido (ver Figura 2.1, izquierda), su principal componente es un patrón iluminado de discos concéntricos o anillas. La luz que parte de las anillas es reflejada por la cara anterior (externa) de la superficie de la córnea, y el reflejo en ella de estos anillos es capturado por una cámara digital en el centro del cono. La imagen es digitalizada utilizando un número fijo de meridianos usando

algoritmos estándar de detección de aristas (en este caso las aristas están provocadas por la alternancia de color negro-blanco o negro-rojo en los discos de Plácido).

El número de discos de Plácido y de meridianos que se toman al digitalizar varía de unos aparatos a otros; el primer valor suele estar cercano a 20 y el segundo a 256. De esta forma, se obtienen algunos miles de puntos, en una estructura casi-concéntrica (los discos de Plácido son concéntricos, pero su reflejo en la córnea, no). La información primaria que mide uno de estos topógrafos son las posiciones (en 2D) de los puntos en las anillas, o equivalentemente la distancia de cada punto al centro de la córnea. Esta información es procesada por el software del topógrafo para dar los datos de altimetría, curvatura u otros, mediante el algoritmo arc-step [8] o sus variaciones. Debido a las características de este algoritmo, la información de la zona periférica es menos fiable

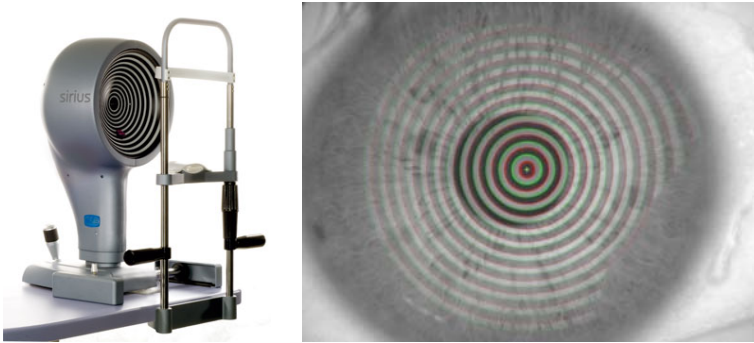


Figura 2.1: Fotografía de un topógrafo corneal basado en discos de Plácido (izquierda) y nube de puntos de la cara anterior de la córnea obtenida por él (derecha).

(más contaminada de ruido) que la parte central.

Los topógrafos basados en la cámara Scheimpflug son más recientes. Estos instrumentos utilizan una cámara de Scheimpflug giratoria, que captura hasta 50 imágenes de cortes transversales de la córnea en cada escaneo [34]. Estas imágenes son analizadas y procesadas por el software del instrumento, dando lugar a la información deseada de la superficie corneal: altimetría, curvatura, etc. Estos topógrafos tienen una ventaja sobre de los Plácido: además de la ca-

ra anterior (exterior) de la córnea, obtienen información de la cara posterior (interna), y por tanto permiten estudiar el grosor de la misma, lo cual tiene gran importancia desde el punto de vista clínico. Sin embargo, al parecer la fiabilidad de las medidas en la parte central de la córnea (la parte más importante desde el punto de vista óptico) es menor que en los topógrafos de Plácido, por lo que ahora están surgiendo topógrafos que combinan los dos sistemas: discos de Plácido y cámara Scheimpflug, para reunir las ventajas de ambas tecnologías.

En la gran mayoría de experimentos numéricos de este trabajo, se han utilizado datos procedentes del topógrafo CSO, basado en discos de Plácido, que toma un máximo de 24 anillas y hasta 256 puntos por anilla. En casi la totalidad de los casos, las medidas son incompletas. En la Figura 2.1 (derecha) se puede comprobar. En esa topografía se ve que hay

puntos en todas las anillas (lo cual rara vez ocurre, normalmente hay algunas anillas totalmente vacías), pero no todas están completas, a las más exteriores les faltan puntos. Este hecho, unido a la mayor presencia de ruido en la parte periférica, hay que tenerlo en cuenta a la hora de desarrollar algoritmos de reconstrucción.

## 2.2. Polinomios de Zernike

Los llamados polinomios de Zernike son una base ortogonal de funciones en el disco unidad. Expresados en coordenadas polares  $(\rho, \theta)$ , cada uno de ellos es producto de una parte angular (solamente depende de  $\theta$ ) y una parte radial (solamente depende de  $\rho$ ). La parte radial siempre es un polinomio en  $\rho$ , mientras que la parte angular es una función del tipo  $\cos(m\theta)$  o bien  $\sin(m\theta)$ , con  $m$  un número entero. Por tanto, en general no son realmente polinomios de dos

variables, aunque casi siempre se les llama polinomios.

Los polinomios de Zernike se agrupan según su orden radial, que es el grado máximo como polinomio de la parte radial. Se pueden definir mediante dos notaciones, la de índices simple y la de doble índice. Ambos son intercambiales con las fórmulas detalladas más abajo. En la notación de doble índice, diremos que el polinomio de Zernike  $(n, m)$  (de orden radial  $n$  y con frecuencia azimutal  $m$ , ambos enteros) es:

$$Z_m^n(\rho, \theta) = \begin{cases} \sqrt{n+1}R_n^0(\rho) & \text{si } m = 0 \\ \sqrt{2(n+1)}R_n^m(\rho) \cos(m\theta) & \text{si } m \text{ es par} \\ \sqrt{2(n+1)}R_n^m(\rho) \sin(-m\theta) & \text{si } m \text{ es impar} \end{cases}$$

con las restricciones de que  $n \geq 0$ ,  $|m| \leq n$  y que  $n - |m|$  sea un número par (alternativamente, fijado  $n \geq 0$ ,  $m$  admite valores desde  $-n$  a  $n$ , incrementados de 2 en 2:  $m \in \{-n, -n + 2, \dots, n - 2, n\}$ ).

El polinomio  $R_n^m(\rho)$  es la llamada parte radial y está dado por:

$$R_n^m(\rho) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s (n-s)!}{s!((n+|m|)/2-s)!((n-|m|)/2-s)!} \rho^{n-2s}$$

Los números a los que afectan los factoriales son siempre naturales positivos, debido a las restricciones que afectan a los valores que pueden tomar  $n$ ,  $m$  y  $s$ . Como vemos, el valor de la parte radial no depende realmente de  $m$ , sino de  $|m|$ , por lo que  $R_n^m(\rho) = R_n^{-m}(\rho)$  para cualesquiera  $n$  y  $m$ .

Desde el orden radial 0 hasta un orden  $n$  dado, hay un total de  $\frac{(n+1)(n+2)}{2}$  polinomios de Zernike.

En ocasiones es útil la nomenclatura de un único índice  $j$  para los polinomios de Zernike. La expresión para cambiar del sistema de doble índice  $(n, m)$  al de índice único  $j$  es:

$$j = \frac{n(n+2) + m}{2}$$



Y recíprocamente, para cambiar del índice único al doble:

$$n = \lceil \frac{-3 + \sqrt{9 + 8j}}{2} \rceil$$
$$m = 2j - n(n + 2)$$

siendo  $\lceil x \rceil$  el operador “techo” (el menor entero que es mayor o igual que  $x$ ).

La ortonormalidad de los polinomios de Zernike en el disco unidad se expresa formalmente como:

$$\int \int_{\mathbb{D}} Z_n^m(\rho, \theta) Z_r^s(\rho, \theta) d\rho d\theta = \delta_{n,r} \delta_{m,s}$$

Además, otra propiedad destacables de los polinomios de Zernike es la simetría rotacional: algunos de los polinomios de Zernike son invariantes bajo cualquier rotación (cuando  $m = 0$ ), mientras el resto lo son bajo rotaciones de ciertos ángulos concretos ( $2\pi/|m|$ , y sus múltiplos enteros, para  $m \neq 0$ ).

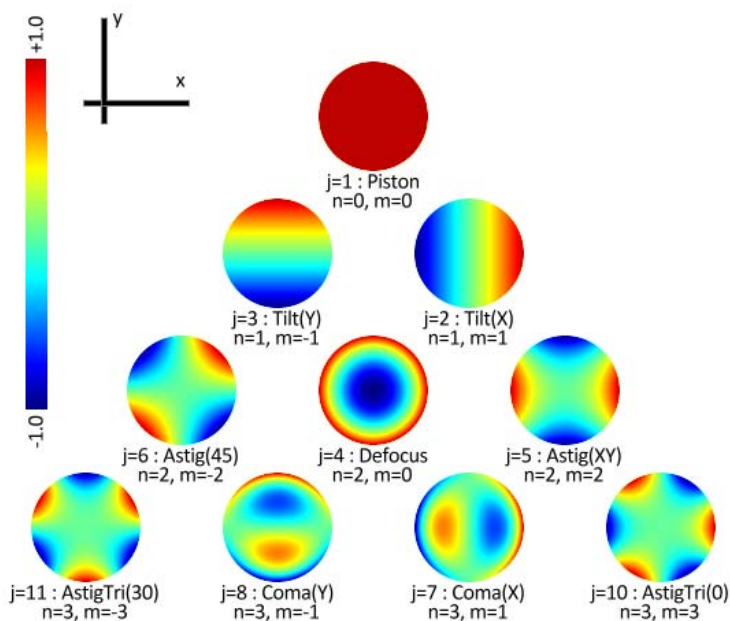


Figura 2.2: Representación de los primeros polinomios de Zernike en el disco unidad, numerados con índice simple ( $j$ ) y doble ( $n, m$ ), junto al nombre que se les suele dar.

### 2.3. Funciones de Base Radial (RBF)

Una función  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  se llama radial (centrada en el origen) si existe otra función  $\Phi : [0, +\infty) \rightarrow \mathbb{R}$  tal que

$$\phi(Q) = \Phi(\|Q\|^2), \quad \forall Q \in \mathbb{R}^d$$

Es decir, el valor de  $\phi$  sólo depende de la norma de su argumento. En otras palabras, las curvas de nivel para  $\phi$  son hipersferas de  $\mathbb{R}^d$  centradas en el origen.

A partir de las funciones radiales se definen las llamadas “funciones de base radial” (traducción del inglés “radial basis functions”, RBF). Una RBF es una combinación lineales de funciones radiales desplazadas, o sea, es del tipo:

$$g(Q) = \sum_{i=1}^n c_i \Phi(\|Q - Q_i\|^2)$$

donde se dice que  $Q_i$  es el “centro” de cada uno de los su-

mandos de la combinación lineal.

Algunos ejemplos de funciones radiales son:

- Gaussianas:

$$\phi(Q) = e^{-\alpha\|Q\|^2}, \quad \alpha > 0.$$

al valor  $\alpha$  se le suele llamar parámetro de forma.

- Multicuádricas inversas:

$$\phi(Q) = \left(c^2 + \|Q\|^2\right)^{-\beta}, \quad c > 0, \beta > 0.$$

Existen otros tipos de funciones radiales, y algunas generalizaciones como los thin-plate splines o las RBF de soporte compacto. Sin embargo, en este trabajo nos centraremos en las RBF gaussianas (o RBGF, como las llamaremos en adelante), y en su versión anisótropa (A-RBGF), que a continuación introduciremos.

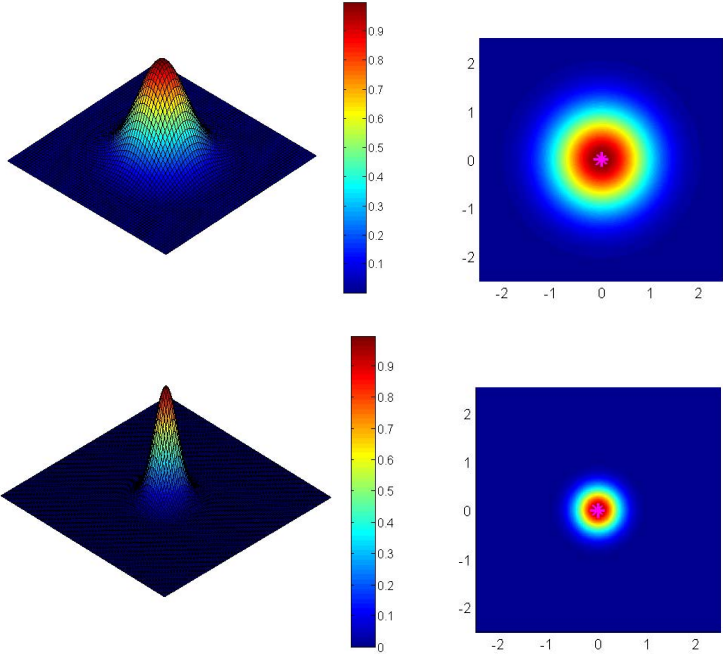


Figura 2.3: Representación gráfica de una RBF gaussiana, centrada en el origen, en perspectiva (izquierda) y en planta (derecha, se ha señalado también el centro), con parámetro de forma  $\alpha = 1$  (arriba) y  $\alpha = 5$  (abajo).

## 2.4. RBF Anisótropas

Además de las funciones radiales introducidas antes, es posible definir funciones radiales anisótropas, de la siguiente forma:

$$\phi(Q) = \Phi \left( \|Q\|_A^2 \right) = \Phi \left( Q^T A Q \right), \quad \forall Q \in \mathbb{R}^d$$

siendo  $A$  una matriz  $d \times d$  simétrica y definida positiva y el superíndice  $T$  denota la trasposición matricial. Para estas funciones anisótropas, las curvas de nivel ya no son hiperesferas, sino hiperelipsoides. El valor de  $\phi(Q)$  depende no solamente de la distancia de  $Q$  al origen (como en las RBF estándar), si no además, de la dirección en la que se encuentra  $Q$ .

Las RBF anisótropas han sido estudiadas previamente desde el punto de vista teórico para interpolación [6, 7]. Se puede ver la representación gráfica de una RBF gaussiana

anisótropa, con centro en el origen, y parámetro de forma

$$A = \begin{pmatrix} 1 & -1,1 \\ -1,1 & 3 \end{pmatrix} \text{ en la Figura 2.4.}$$

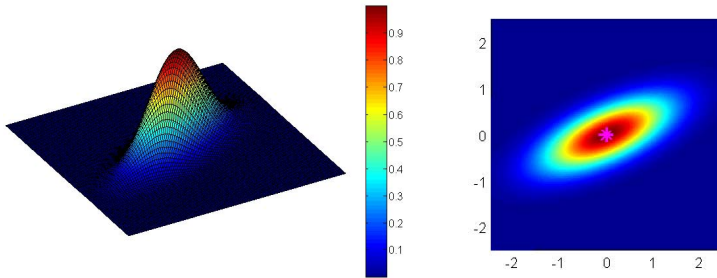


Figura 2.4: Representación gráfica de una RBF gaussiana anisótropa (A-RBGF), centrada en el origen, en perspectiva (izquierda) y en planta (derecha, se ha señalado también el centro), con parámetro de forma la matriz  $A$  anterior.





## Capítulo 3

# Ajuste directo. Regularización de Tikhonov

Cuando se reconstruye una córnea, es recomendable ajustar previamente una esfera a los datos, y posteriormente los datos “residuales” respecto de dicha esfera, con algún otro algoritmo, como el que se propone a continuación. Vamos pues a describir un algoritmo general para reconstrucción de superficies, que utiliza una base fija de funciones de base radial de tipo gaussiano (RBGF, del inglés “radial basis gaussian

function”). Dicha base dependerá de dos valores: la cantidad de funciones deseada  $N$ , y del parámetro de escala o de forma de las gaussianas  $\alpha$ . Se obtendrá la reconstrucción  $E$  como combinación lineal de las funciones base:

$$E(\mathbf{P}) = \sum_{j=1}^N c_j e^{-\alpha \|\mathbf{P} - \mathbf{Q}^{(j)}\|_2^2}, \quad \mathbf{P} = (x, y), \mathbf{Q}^{(j)} = (Q_x^{(j)}, Q_y^{(j)})$$

donde  $Q_i$  son los centros de las RBGF de la base.

Los coeficientes en la combinación lineal se calcularán por el método de mínimos cuadrados [41], a partir de los datos de entrada. Si el tamaño de la base utilizada en la reconstrucción es grande, surgen problemas numéricos debido a la mala condición numérica de la matriz del sistema. Para solventar este problema, se propone el uso de la regularización de Tikhonov [17], que depende de un parámetro  $\lambda$  llamado de regularización. Utilizar un valor correcto para este parámetro es crucial, y existen distintos criterios para elegirlo. Aquí ex-

plicamos dos de los más utilizados: el de L-Curve y el de cross validación generalizada (GCV). El criterio GCV es un método estadístico general, que se puede utilizar para escoger parámetros en un modelo cualquiera. Generalmente, la aplicación de GCV resulta costosa (en tiempo de cálculo), pero en el caso de la regularización de Tikhonov y gracias al uso de la descomposición en valores singulares (SVD) compacta de una matriz, se puede aplicar con gran eficiencia, permitiendo la selección del parámetro de regularización en tiempo real.

Finalmente, después de explicar los métodos y criterios mencionados, se muestran y se discuten algunos resultados experimentales del algoritmo.

### 3.1. Pre-ajuste

Los datos de entrada son una nube tridimensional  $(x_k, y_k, z_k)$ ,  $k = 1, \dots, N$ , que supondremos que corresponde o bien a da-

tos de elevación o curvatura  $z_k$  medidos por un topógrafo corneal en el nodo  $\mathbf{P}_k$  de la cara anterior de la córnea con coordenadas cartesianas  $(x_k, y_k)$ . Discutiremos el caso cuando  $z_k$  corresponde a la elevación. Teniendo en cuenta la forma global de la córnea, es un procedimiento estandar “aplanar” los datos ajustando primero la esfera de mejor ajuste [1]:

$$S(x, y) = z_0 + \sqrt{R^2 - (x - x_0)^2 - (y - y_0)^2}$$

donde  $R$  y  $(x_0, y_0, z_0)$  son su radio y las coordenadas cartesianas de su centro, respectivamente. A pesar de que la práctica común es ajustar por mínimos cuadrados estándar, se obtienen mejores resultados cuando se utiliza un ajuste de mínimos cuadrados ponderados, usando  $(1 + ||P_k||)^{-1}$  como el peso, de acuerdo con la distribución típica del error de medición de los topógrafos corneales [40].

Como resultado del paso anterior, los errores residuales

$\varepsilon_k = z_k - S(x_k, y_k)$  contienen información relevante de la superficie corneal además de ruido. Nuestro objetivo es ajustar esos residuos por una función  $E(x, y)$ , de tal forma que la expresión analítica para la altura de la córnea viene dada por

$$\text{Cornea}(x, y) = S(x, y) + E(x, y), \quad (3.1)$$

De esta forma,  $S$  contiene la forma esencial de la córnea, mientras que en  $E$  son capturadas las pequeñas irregularidades de la superficie. La función  $E$  está dada por una combinación lineal de  $n$  funciones de un diccionario dado,

$$E(x, y) = E_n(x, y) = \sum_{j=1}^n c_j h_j(x, y). \quad (3.2)$$

En un esquema ideal,  $n$  depende de los datos reales, y debe ser lo bastante grande para que toda la información relevante de las medidas de elevación esté modelada en  $E$ , pero no excesi-

vamente grande para evitar sobre-parametrizar el problema y ajustar el ruido. Para solventar las ya mencionadas dificultades de los polinomios de Zernike, usaremos como diccionario las funciones de base radial gaussianas, de la forma:

$$h(x, y) = \exp\left(-\alpha \|\mathbf{P} - \mathbf{Q}\|^2\right), \quad \mathbf{P} = (x, y)^T,$$

donde el superíndice  $T$  denota la trasposición matricial,

$\mathbf{Q} = (Q_x, Q_y)^T$  es un cierto punto del plano (llamado el “centro” de la función) y  $\alpha > 0$  un coeficiente libre, llamado el parámetro de forma.

Por tanto, buscamos una expresión de la forma

$$\begin{aligned} \text{Cornea}(x, y) = & S(x, y) + \\ & + \sum_{j=1}^n c_j \exp\left(-\alpha \|\mathbf{P} - \mathbf{Q}^{(j)}\|^2\right), \quad \mathbf{P} = (x, y)^T. \end{aligned}$$

Ajustaremos una combinación lineal de este tipo a los

datos, con los centros  $Q^{(j)}$ , el parámetro de forma  $\alpha$  y el número de funciones  $n$  fijados a priori.

## 3.2. Ajuste directo por RBFs

Basándonos en nuestros experimentos, hemos escogido colocar los centros  $Q$  distribuidos en una malla cuadrada regular, a pesar de que se pueden tomar muchas otras opciones igualmente válidas. Lo primero, supondremos que los datos han sido normalizados al disco unidad (se puede hacer de muchas formas, por ejemplo dividiendo por la norma del mayor), así que  $\|P\| \leq 1$ . Entonces, tomaremos un entero positivo  $n_0$  y  $n = n_0^2$  será el número de términos en la representación funcional 3.2 de la superficie.

Si definimos el valor  $a = 1/2$ , el cuadrado  $[-a, a]^2$  contiene al disco unidad, luego tiene todos los datos en su interior (estrictamente). Construimos la malla regular en dicho cua-

drado con  $n_0$  nodos equiespaciados en cada dimensión. Por tanto los  $n = n_0^2$  nodos de la malla bidimensional tienen las coordenadas cartesianas siguientes:

$$\left( -a + 2a \frac{r}{n_0 - 1}, -a + 2a \frac{s}{n_0 - 1} \right)$$

donde  $r, s = 0, 1, \dots, n_0 - 1$ . Los nodos de esta malla se pueden ordenar lexicográficamente según el par  $(r, s)$  y por tanto definir  $\mathbf{Q}_1 = (-a, -a)$ ,  $\mathbf{Q}_2 = (-a, -a + \frac{2a}{n_0-1})$ ,  $\dots$ ,  $\mathbf{Q}_{n_0} = (-a, a)$ ,  $\dots$ ,  $\mathbf{Q}_n = (a, a)$ . Así pues, una vez que se ha fijado el valor  $n_0$ , los  $n$  centros  $\mathbf{Q}$  están completamente determinados y se pueden numerar con el conjunto de índices  $j = 1, 2, \dots, n$ .

Para obtener la función que mejor ajusta a los datos corneales, una vez que los valores de  $n$  y  $\alpha$  se han fijado, los coeficientes de las RBGF se calculan por el método estándar



de mínimos cuadrados. Así que si llamamos

$$\mathbf{A} = \left( \exp \left( -\alpha^2 \left\| \mathbf{P}^{(i)} - \mathbf{Q}^{(j)} \right\|_2^2 \right) \right)_{(i,j)}$$

a la matriz de evaluaciones de las funciones base en los datos (usualmente llamada matriz de colocación, o en estadística, matriz de diseño), y  $\mathbf{b} = (z^{(i)})$  al vector de datos de elevación (o de curvatura) en los puntos  $\mathbf{P}$ , lo que queremos es resolver el sistema lineal de ecuaciones sobredeterminado

$$\mathbf{Ax} = \mathbf{b}$$

por medio de los mínimos cuadrados lineales. El vector solución  $\mathbf{x}$  satisface:

$$\mathbf{x} = \arg \min_{\mathbf{x}^*} \|\mathbf{Ax}^* - \mathbf{b}\|_2^2$$

y se puede obtener por métodos estándar, como la factori-

zación  $QR$  de la matriz  $\mathbf{A}$ , o la descomposición en valores singulares (ver [41]).

Los valores óptimos de  $n$  y  $\alpha$  se han estudiado en la práctica, basados en multitud de experimentos tanto con córneas reales como simuladas. Normalmente, los valores adecuados de  $n$  están entre 50 y 300; y para  $\alpha$  entre 1 y 20. Estos rangos de valores se discuten posteriormente. Como veremos, si no se utiliza la regularización, es conveniente elegir un valor pequeño de  $n$ , debido a la mala condición numérica del ajuste por mínimos cuadrados. Por otro lado, la elección de  $\alpha$  no parece tan crítica, obteniéndose buenos resultados para muchos valores dentro de un cierto rango.

Cuando  $n$  es relativamente grande (a partir de 150 o 200, en nuestros experimentos) o bien incluso con  $n$  más pequeño si  $\alpha$  es bastante bajo, la mala condición numérica de la matriz  $\mathbf{A}$  hace necesario el uso de algún tipo de “estabilización” o

“regularización”. Aquí utilizamos la llamada “regularización de Tikhonov”. Su objetivo es construir una solución  $E(x, y)$  con coeficientes pequeños, evitando así las grandes oscilaciones de la función.

### 3.3. Regularización de Tikhonov

El objetivo de este método de regularización es mejorar la condición numérica del ajuste por mínimos cuadrados discretos lineales de una matriz  $\mathbf{A}$  a uno vector de datos  $\mathbf{b}$ , es decir, queremos resolver por mínimos cuadrados el sistema de ecuaciones lineales sobredeterminado

$$\mathbf{Ax} = \mathbf{b}$$

Generalmente, lo que tendremos es un conjunto de  $n$  datos  $(\mathbf{t}, \mathbf{b})$ , con  $\mathbf{t} = (t_1, \dots, t_n)^T$ ,  $\mathbf{b} = (b_1, \dots, b_n)^T$  y un conjunto

de  $m$  funciones  $\varphi = (\varphi_1, \dots, \varphi_m)$  (con  $m < n$ ) con las que queremos hacer regresión lineal, o sea, queremos encontrar los coeficientes  $\mathbf{x} = (x_1, \dots, x_m)^T$  tales que

$$\sum_{j=1}^m x_j \varphi_j(t_i) = b_i, \text{ para } 1 \leq i \leq n$$

O bien, si llamamos  $\mathbf{A}$  a la matriz de evaluación de las funciones  $\varphi$  en los datos  $\mathbf{t}$ , es decir,  $\mathbf{A} = (\varphi_j(t_i))_{i,j}$ , la regresión lineal se escribe matricialmente como  $\mathbf{Ax} = \mathbf{b}$ . Normalmente no se podrá conseguir la igualdad, con lo que resolveremos el sistema por mínimos cuadrados para obtener una solución aproximada  $\mathbf{Ax} \approx \mathbf{b}$ .

En muchas ocasiones la matriz de ajuste  $\mathbf{A}$  está muy mal condicionada, con lo que pequeñas variaciones de los datos  $(\mathbf{t}, \mathbf{b})$  pueden producir grandes variaciones en los coeficientes obtenidos,  $\mathbf{x}$ . En estos casos, una forma de estabilizar la solución  $\mathbf{x}$  es utilizar la Regularización de Tikhonov, que se

explica a continuación.

En la regresión lineal estándar, lo que se busca es minimizar en  $\mathbf{x}$  la norma euclídea del residuo del ajuste,  $\mathbf{Ax} - \mathbf{b}$ :

$$\min_x \|\mathbf{Ax} - \mathbf{b}\|_2$$

que es equivalente a minimizar su cuadrado:

$$\min_x \|\mathbf{Ax} - \mathbf{b}\|_2^2 \quad (3.3)$$

y se toma como solución  $\mathbf{x}$  el vector que da el mínimo de la expresión anterior.

Alternativamente, la Regularización de Tikhonov añade un término a la función objetivo de la minimización (3.3), que depende de un parámetro positivo  $\lambda$  y de la norma del vector solución  $\mathbf{x}$ :

$$\min_x \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2 \quad (3.4)$$

De esta forma, si el valor de  $\lambda$  es adecuado, se busca una solución próxima a la de mínimos cuadrados pero de norma lo más pequeña posible. Si  $\lambda = 0$ , estamos ante el ajuste estándar, sin regularizar. Por otro lado, si  $\lambda$  crece cada vez más, el vector de coeficientes  $\mathbf{x}$  se irá aproximando a cero, con lo que el ajuste será muy malo. Surge entonces el problema de qué valor asignar a ese parámetro  $\lambda$  y para esto se han propuesto distintos criterios en la literatura [17, 43]. Posteriormente explicaremos dos de los más utilizados: el de la función de cross validación generalizada (GCV) y el de la curva-L (o L-Curve).

El problema de Regularización de Tikhonov se puede reformular de las siguientes formas equivalentes a la expresión inicial (3.4):

$$\min_x \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2$$

$$\Downarrow$$

$$\min_x \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \|\lambda\mathbf{x}\|_2^2$$

$$\Downarrow$$

$$\min_x \|(\mathbf{Ax} - \mathbf{b}, \lambda\mathbf{x})\|_2^2$$

$$\Downarrow$$

$$\min_x \left\| \begin{pmatrix} \mathbf{A} \\ \lambda\mathbf{I}_m \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{b} \\ \mathbf{0}_m \end{pmatrix} \right\|_2^2$$

$$\Downarrow$$

$$\min_x \left\| \begin{pmatrix} \mathbf{A} \\ \lambda\mathbf{I}_m \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{b} \\ \mathbf{0}_m \end{pmatrix} \right\|_2$$

siendo  $\mathbf{I}_m$  la matriz identidad de orden  $m$  y  $\mathbf{0}_m$  el vector columna de  $m$  ceros. Para abreviar, en adelante pondremos

$$\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A} \\ \lambda \mathbf{I}_m \end{pmatrix} \quad \tilde{\mathbf{b}} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0}_m \end{pmatrix}$$

con lo que resulta que la Regularización de Tikhonov es de nuevo un problema de mínimos cuadrados lineales, esta vez formulado en términos de  $\tilde{\mathbf{A}}$  y  $\tilde{\mathbf{b}}$ :

$$\min_x \left\| \tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{b}} \right\|_2$$

En la definición de  $\tilde{\mathbf{A}}$  se aprecia mejor el efecto de la regularización. Si  $\mathbf{A}$  es de rango deficiente, y elegimos  $\lambda \neq 0$ , entonces la matriz ampliada  $\tilde{\mathbf{A}}$  tendrá rango completo. Aún así, si  $\lambda$  es demasiado pequeño, el número de condición de  $\tilde{\mathbf{A}}$  será todavía grande, y es por lo que se hace imprescindible un criterio adecuado para escoger el valor de  $\lambda$ . Primero vamos a



estudiar en detalle es la cross validación generalizada (GCV, del inglés “Generalized Cross-Validation”).

### 3.3.1. Cross-Validación Generalizada (GCV)

Ahora vamos a ver el método estadístico de la cross validación [17, 43]. Se trata de un método genérico para estimar un valor adecuado del parámetro en un modelo que estemos ajustando. Igual que en la sección previa, consideremos que tenemos un conjunto de datos  $(\mathbf{t}, \mathbf{b})$ , con  $\mathbf{t}$  no necesariamente unidimensional. Supongamos que a partir de los datos, estamos interesados en construir un modelo dependiente de un parámetro,  $f_\lambda$ , de forma que  $f_\lambda$  prediga lo mejor posible los valores de  $\mathbf{b}$  a partir de los valores de  $\mathbf{t}$  (se puede pensar en la regresión, por ejemplo).

Introduzcamos un poco de notación. Llamaremos  $\hat{f}_\lambda$  al modelo ajustado a nuestros datos  $(\mathbf{t}, \mathbf{b})$  completos. Ahora

supongamos que hemos eliminado el dato  $i$ -ésimo, de forma que quedan  $n - 1$  datos, y con ellos ajustamos otro modelo, al que llamaremos  $\hat{f}_\lambda^{-i}$ . Puesto que originalmente tenemos  $n$  datos, para cada valor de  $\lambda$  podemos construir  $n$  modelos del tipo  $\hat{f}_\lambda^{-i}$  distintos, uno por cada dato  $i$  que quitamos.

La idea de la Cross-Validación es obtener el modelo quitando el dato  $i$ -ésimo,  $\hat{f}_\lambda^{-i}$ , y ver en ese dato en cuánto se equivoca la predicción, es decir, el valor de la diferencia

$$\left( \hat{f}_\lambda^{-i}(t_i) - b_i \right)^2.$$

Si hacemos esto para todo  $1 \leq i \leq n$  y promediamos, tendremos una medida global de la bondad de ajuste del modelo con parámetro  $\lambda$  a los datos. Definimos así la función de Cross-Validación Ordinaria (OCV), dependiente de  $\lambda$ :

$$OCV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left( b_i - \hat{f}_\lambda^{-i}(t_i) \right)^2$$

Cuanto menor sea la función  $OCV(\lambda)$ , mejor será el ajuste del modelo  $\hat{f}_\lambda$  a los datos. Por tanto, un criterio adecuado para escoger el valor de  $\lambda$  es minimizar la función de Cross-Validación Ordinaria:

$$\min_{\lambda} OCV(\lambda) = \min_{\lambda} \left\{ \frac{1}{n} \sum_{i=1}^n \left( b_i - \hat{f}_\lambda^{-i}(t_i) \right)^2 \right\}$$

Sin embargo, por el momento no es un método eficiente: para cada valor de  $\lambda$  que fijemos, evaluar  $OCV(\lambda)$  requiere el ajuste de  $n$  modelos distintos,  $\hat{f}_\lambda^{-i}$ . Si  $n$  es relativamente grande, este criterio es imposible de usar en la práctica.

Afortunadamente, en muchas situaciones (incluyendo la regresión lineal, como la Regularización de Tikhonov) se puede tomar un atajo, que es el llamado 'Leave-One-Out Lemma'. Para enunciarlo conviene añadir algo más de notación. Primero, pondremos  $\mathbf{t}^{-i}$  al vector de  $n-1$  componentes que se obtiene de  $\mathbf{t}$  quitando la  $i$ -ésima, y lo análogo para  $\mathbf{b}$  será  $\mathbf{b}^{-i}$ .

Llamaremos  $\tilde{\mathbf{b}}_i$  al vector de  $n$  componentes obtenido a partir de  $\mathbf{b}$ , pero donde hemos sustituido la coordenada  $i$ -ésima, que era  $b_i$ , por el valor  $\hat{f}_\lambda^{-i}(t_i)$ , que es la predicción en  $t_i$  según el modelo ajustado con los datos  $(\mathbf{t}^{-i}, \mathbf{b}^{-i})$ . Así, podemos construir un nuevo modelo, con los  $n$  datos  $\tilde{\mathbf{b}}_i$ , a este modelo le llamaremos  $\tilde{f}_\lambda^{-i}$ . Se tiene entonces lo siguiente.

**Lema.** (*Leave-One-Out*)

*El modelo construido con los  $n - 1$  datos  $(\mathbf{t}^{-i}, \mathbf{b}^{-i})$ , evaluado en  $t_i$ , coincide con el modelo construido con los  $n$  datos  $(\mathbf{t}, \tilde{\mathbf{b}}_i)$ , evaluado en  $t_i$ . Con la notación anterior:*

$$\hat{f}_\lambda^{-i}(t_i) = \tilde{f}_\lambda^{-i}(t_i)$$

En muchos modelos, incluyendo a la regresión lineal o ajuste por mínimos cuadrados discretos, la estimación que se obtiene según el modelo,  $\hat{\mathbf{b}}$  está ligada con los datos originales  $\mathbf{b}$  por una matriz  $\mathbf{H} = \mathbf{H}(\lambda) = (h_{ij})_{i,j}$  que depende del

parámetro  $\lambda$  del modelo:

$$\hat{f}_\lambda(\mathbf{t}) = \hat{\mathbf{b}} = \mathbf{H}\mathbf{b}$$

Escribiendo esta expresión para cada fila, lo que se tiene es que, para cada  $1 \leq i \leq n$ :

$$\hat{f}_\lambda(t_i) = \sum_{j=1}^n h_{ij} b_j$$

Por otro lado, según el Lema de Leave-One-Out, y utilizando que, según lo dicho antes, la predicción para los datos  $\tilde{\mathbf{b}}_i$  viene dada por  $\mathbf{H}\tilde{\mathbf{b}}_i$ :

$$\hat{f}_\lambda^{-i}(t_i) = \tilde{f}_\lambda^{-i}(t_i) = \sum_{j=1}^n h_{ij} (\tilde{\mathbf{b}}_i)_j = \sum_{j=1, j \neq i}^n h_{ij} b_j + h_{ii} \hat{f}_\lambda^{-i}(t_i)$$

De aquí, restando de esta expresión la anterior, se cancelan

todos los términos de las sumas salvo cuando  $j = i$ :

$$\hat{f}_\lambda^{-i}(t_i) - \hat{f}_\lambda(t_i) = -h_{ii} \left( b_i - \hat{f}_\lambda^{-i}(t_i) \right)$$

y sumando y restando  $b_i$ , y agrupando:

$$\left( b_i - \hat{f}_\lambda(t_i) \right) - \left( b_i - \hat{f}_\lambda^{-i}(t_i) \right) = -h_{ii} \left( b_i - \hat{f}_\lambda^{-i}(t_i) \right)$$

$$b_i - \hat{f}_\lambda(t_i) = (1 - h_{ii}) \left( b_i - \hat{f}_\lambda^{-i}(t_i) \right)$$

de aquí, finalmente:

$$b_i - \hat{f}_\lambda^{-i}(t_i) = \frac{b_i - \hat{f}_\lambda(t_i)}{1 - h_{ii}}$$

La expresión anterior permite escribir la función de cross-validación ordinaria como:

$$OCV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left( b_i - \hat{f}_\lambda^{-i}(t_i) \right)^2 = \frac{1}{n} \sum_{i=1}^n \left( \frac{b_i - \hat{f}_\lambda(t_i)}{1 - h_{ii}} \right)^2$$

Con esta expresión hemos conseguido que la evaluación de  $OCV$  para cada valor de  $\lambda$  solo requiere ajustar un modelo,  $\hat{f}_\lambda$ , en lugar de los  $n$  modelos  $\hat{f}_\lambda^{-i}$  que era necesario ajustar antes, lo que significa una reducción muy importante del coste computacional. Como inconveniente, ahora es necesario conocer los valores diagonales  $h_{ii}$  de la matriz  $\mathbf{H} = \mathbf{H}(\lambda)$ .

Como variación de este método se propuso la Cross Validación Generalizada ( $GCV$ ), en el que se busca minimizar la misma función pero en la que se sustituyen todos los  $h_{ii}$  por la media de ellos,  $Tr(\mathbf{H})/n$ . Se supone que los mínimos de  $OCV$  y  $GCV$  tendrán valores próximos, y aunque aparentemente para conocer  $Tr(\mathbf{H})$  se necesita conocer los  $n$  valores  $h_{ii}$ , esto no es necesariamente cierto: en ocasiones se puede

calcular o estimar la traza sin conocer exactamente los valores diagonales. Se propone pues escoger el parámetro  $\lambda$  como el valor que minimiza la función *GCV*:

$$GCV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left( \frac{b_i - \hat{f}_\lambda(t_i)}{1 - \frac{Tr(\mathbf{H})}{n}} \right)^2$$

El denominador de la sumatoria se puede simplificar de la siguiente forma:

$$\begin{aligned} \left( 1 - \frac{Tr(\mathbf{H})}{n} \right)^2 &= \frac{1}{\sqrt{n}} (n - Tr(\mathbf{H}))^2 = \\ &= \frac{1}{\sqrt{n}} (Tr(\mathbf{I}_n) - Tr(\mathbf{H}))^2 = \\ &= \frac{1}{\sqrt{n}} (Tr(\mathbf{I}_n - \mathbf{H}))^2 \end{aligned}$$

con  $\mathbf{I}_n$  la matriz identidad de orden  $n$ . Puesto que este denominador no depende de  $i$ , se puede sacar de la sumatoria y



resulta que la expresión de  $GCV$  se simplifica a:

$$GCV(\lambda) = \frac{1}{n} \frac{\sqrt{n}}{Tr(\mathbf{I}_n - \mathbf{H})^2} \sum_{i=1}^n \left( b_i - \hat{f}_\lambda(t_i) \right)^2$$

Teniendo en cuenta que el único propósito de construir  $OCV$  y  $GCV$  es buscar el valor de  $\lambda$  que da el mínimo, a efectos prácticos se puede quitar el factor constante  $\frac{\sqrt{n}}{n}$  (o el  $\frac{1}{n}$  en el caso de  $OCV$ ), pues no afecta a la minimización, y podemos tomar  $GCV$  como:

$$GCV(\lambda) = \frac{\sum_{i=1}^n \left( b_i - \hat{f}_\lambda(t_i) \right)^2}{Tr(\mathbf{I}_n - \mathbf{H})^2}$$

La única restricción para la utilización de  $GCV$  en la práctica es que debe ser  $Tr(\mathbf{H}) = Tr(\mathbf{H}(\lambda)) \neq n$ , para que el denominador no se anule.

### 3.3.2. SVD Compacta

Antes de ver como se aplica la idea de la Cros-Validación Generalizada a la Regularización de Tikhonov, conviene recordar la Descomposición en Valores Singulares (SVD) de una matriz, así como su versión compacta (SVD compacta).

#### SVD estándar

Si  $\mathbf{A}_{n,m}$  es una matriz  $n \times m$  con entradas en  $K = \mathbb{R}$  ó  $K = \mathbb{C}$ , entonces admite la factorización:

$$\mathbf{A}_{n,m} = \mathbf{U}_{n,n} \mathbf{\Sigma}_{n,m} \mathbf{V}_{m,m}^T$$

donde el superíndice  $T$  denota la traspuesta o traspuesta conjugada, según quién sea el cuerpo  $K$ . Las matrices  $\mathbf{U}$  (orden  $n$ ) y  $\mathbf{V}$  (orden  $m$ ) son unitarias y  $\mathbf{\Sigma}$  es una matriz diagonal con entradas reales y positivas. Las entradas de  $\mathbf{\Sigma}$  generalmente se toman en orden descendente, para que  $\mathbf{\Sigma}$  esté de-

terminada de forma única, aunque  $\mathbf{U}$  y  $\mathbf{V}$  no lo están. Los valores diagonales de  $\mathbf{\Sigma}$  (reales y positivos) se llaman los valores singulares de  $\mathbf{A}$ . Finalmente, que  $\mathbf{U}$  y  $\mathbf{V}$  sean unitarias significa que:

$$\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_n$$

$$\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_m$$

### SVD compacta

La SVD compacta para una matriz  $\mathbf{A}_{n,m}$  se obtiene (teóricamente, no en la práctica) a partir de la SVD estándar, de la siguiente forma. La matriz  $\mathbf{\Sigma}$  de la SVD estándar es diagonal y en general es rectangular. Como los valores singulares están ordenados de forma decreciente, los últimos posiblemente son ceros. Si hay  $q$  valores singulares no nulos ( $q$  coincide con el rango de  $\mathbf{A}$ ), la SVD compacta (cuyas matrices denotaremos con el superíndice ( $C$ )) se obtiene a partir de la completa

quedándonos sólo con las  $q$  primeras columnas de  $\mathbf{U}$ ,  $\mathbf{V}$  y de  $\Sigma$  nos quedamos con las primeras  $q$  filas y  $q$  columnas. Es decir:

$$\mathbf{U}_{n,n} = \begin{pmatrix} \mathbf{U}_{n,q}^{(C)} & \mathbf{U}_{n,n-q}^{(resto)} \end{pmatrix}$$

$$\mathbf{V}_{m,m} = \begin{pmatrix} \mathbf{V}_{m,q}^{(C)} & \mathbf{V}_{m,m-q}^{(resto)} \end{pmatrix}$$

$$\Sigma_{n,m} = \begin{pmatrix} \Sigma_{q,q}^{(C)} & \mathbf{0}_{q,m-q} \\ \mathbf{0}_{n-q,q} & \mathbf{0}_{n-q,m-q} \end{pmatrix}$$

Entonces la SVD compacta de  $\mathbf{A}$  es:

$$\mathbf{A}_{n,m} = \mathbf{U}_{n,q}^{(C)} \Sigma_{q,q}^{(C)} (\mathbf{V}_{m,q}^{(C)})^T$$

y de esta forma, ahora

$$\Sigma^{(C)} = \text{diag}(\sigma_1, \dots, \sigma_q)$$

con los  $\sigma_j \neq 0$ , para  $1 \leq j \leq q$ . Las matrices  $\mathbf{U}^{(C)}$  y  $\mathbf{V}^{(C)}$  aho-

ra no son unitarias, pues hemos 'recortado' algunas columnas a cada una. Sin embargo, aún se mantienen las relaciones:

$$\left(\mathbf{U}^{(C)}\right)^T \mathbf{U}^{(C)} = \mathbf{I}_q$$

$$\left(\mathbf{V}^{(C)}\right)^T \mathbf{V}^{(C)} = \mathbf{I}_q$$

En muchas ocasiones esta descomposición compacta resulta más conveniente que la estándar, sobre todo porque si  $m \ll n$  (matrices con muchas más filas que columnas), se puede calcular mucho más rápido la SVD compacta que la estándar.

### 3.3.3. Uso eficiente de GCV

Anteriormente hemos visto cómo la función de Cross Validación Generalizada (*GCV*) se puede simplificar a través del lema 'Leave-One-Out', de forma que su evaluación resul-

ta más eficiente. Sin embargo, utilizarla de esa manera para la Regularización de Tikhonov aún resulta lento: para cada valor de  $\lambda$ , conocer  $GCV(\lambda)$  requiere resolver el ajuste por mínimos cuadrados lineales  $\tilde{\mathbf{A}}(\lambda)\mathbf{x} \approx \tilde{\mathbf{b}}$ . Vamos a ver cómo se puede utilizar la SVD compacta para utilizar la función  $GCV$  de forma muy eficiente.

Sea entonces  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$  la descomposición en valores singulares (SVD estándar) de la matriz original  $\mathbf{A}$  (antes de regularizar). Vamos a determinar explícitamente quién es la matriz  $\mathbf{H} = \mathbf{H}(\lambda)$  en este caso, utilizando la SVD compacta. La solución  $\mathbf{x}$  del problema regularizado es la que minimiza

$$\left\| \tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{b}} \right\|_2$$

y sabemos que es la que resuelve las ecuaciones normales, con

solución única puesto que  $\tilde{\mathbf{A}}$  es de rango máximo:

$$\tilde{\mathbf{A}}^T \tilde{\mathbf{A}} \mathbf{x} = \tilde{\mathbf{A}}^T \tilde{\mathbf{b}}$$

Recordando la definición de  $\tilde{\mathbf{A}}$  y  $\tilde{\mathbf{b}}$ , las ecuaciones normales se escriben como:

$$\begin{pmatrix} \mathbf{A}^T & \lambda \mathbf{I}_m \end{pmatrix} \begin{pmatrix} \mathbf{A} \\ \lambda \mathbf{I}_m \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{A}^T & \lambda \mathbf{I}_m \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{0}_m \end{pmatrix}$$

o bien, operando por bloques

$$(\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I}_m) \mathbf{x} = \mathbf{A}^T \mathbf{b}$$

Como  $\tilde{\mathbf{A}}$  es de rango completo,  $\tilde{\mathbf{A}}^T \tilde{\mathbf{A}} = \mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I}_m$  es inversible, con lo que

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I}_m)^{-1} \mathbf{A}^T \mathbf{b}$$

Utilizando la SVD (completa) de  $\mathbf{A}$  y las propiedades de las matrices  $\mathbf{U}$ ,  $\mathbf{\Sigma}$ ,  $\mathbf{V}$ , se puede operar de la siguiente forma:

$$\mathbf{A}_{n,m} = \mathbf{U}_{n,n} \mathbf{\Sigma}_{n,m} \mathbf{V}_{m,m}^T \quad \mathbf{A}^T = \mathbf{V} \mathbf{\Sigma}^T \mathbf{U}^T$$

$$\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{\Sigma}^T \mathbf{U}^T \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \mathbf{V} \mathbf{\Sigma}^T \mathbf{\Sigma} \mathbf{V}^T$$

$$\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I}_m = \mathbf{V} \mathbf{\Sigma}^T \mathbf{\Sigma} \mathbf{V}^T + \lambda^2 \mathbf{V} \mathbf{I}_m \mathbf{V}^T = \mathbf{V} (\mathbf{\Sigma}^T \mathbf{\Sigma} + \lambda^2 \mathbf{I}_m) \mathbf{V}^T$$

$$(\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I}_m)^{-1} = \mathbf{V} (\mathbf{\Sigma}^T \mathbf{\Sigma} + \lambda^2 \mathbf{I}_m)^{-1} \mathbf{V}^T$$

$$\begin{aligned} (\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I}_m)^{-1} \mathbf{A}^T &= \mathbf{V} (\mathbf{\Sigma}^T \mathbf{\Sigma} + \lambda^2 \mathbf{I}_m)^{-1} \mathbf{V}^T \mathbf{V} \mathbf{\Sigma} \mathbf{U}^T = \\ &= \mathbf{V} (\mathbf{\Sigma}^T \mathbf{\Sigma} + \lambda^2 \mathbf{I}_m)^{-1} \mathbf{\Sigma}^T \mathbf{U}^T \end{aligned}$$

Llamaremos  $\mathbf{D}$  a la matriz (de orden  $m \times n$ ) siguiente:

$$\mathbf{D} = (\mathbf{\Sigma}^T \mathbf{\Sigma} + \lambda^2 \mathbf{I}_m)^{-1} \mathbf{\Sigma}^T$$

que es diagonal (no cuadrada en general) y cuyos elementos



diagonales son

$$\frac{\sigma_i}{\sigma_i^2 + \lambda^2}, \quad \text{para } 1 \leq i \leq m,$$

y los  $\sigma_i$  son los valores singulares de  $\mathbf{A}$  (posiblemente, los últimos son nulos). Con esta notación resulta:

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I}_m)^{-1} \mathbf{A}^T \mathbf{b} = \mathbf{V} \mathbf{D} \mathbf{U}^T \mathbf{b}$$

y a partir de aquí, los valores estimados  $\hat{\mathbf{b}}$  según el ajuste son

$$\hat{\mathbf{b}} = \mathbf{A} \mathbf{x} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \mathbf{V} \mathbf{D} \mathbf{U}^T \mathbf{b} = \mathbf{U} \mathbf{\Sigma} \mathbf{D} \mathbf{U}^T \mathbf{b}$$

con lo que finalmente obtenemos que la matriz  $\mathbf{H}$ , que da los valores estimados  $\hat{\mathbf{b}}$  en función de los valores originales  $\mathbf{b}$  (es decir,  $\hat{\mathbf{b}} = \mathbf{H} \mathbf{b}$ ) es:

$$\mathbf{H} = \mathbf{H}(\lambda) = \mathbf{U} \mathbf{\Sigma} \mathbf{D} \mathbf{U}^T$$

y de los factores que aparecen en esta expresión, solamente  $\mathbf{D} = \mathbf{D}(\lambda)$  depende del parámetro  $\lambda$ .

Con esto ya podemos calcular  $Tr(\mathbf{I}_n - \mathbf{H})$ , que aparece en el denominador de la función  $GCV$ :

$$\begin{aligned} Tr(\mathbf{I}_n - \mathbf{H}) &= Tr(\mathbf{I}_n - \mathbf{U}\Sigma\mathbf{D}\mathbf{U}^T) = Tr(\mathbf{U}(\mathbf{I}_n - \Sigma\mathbf{D})\mathbf{U}^T) = \\ &= Tr(\mathbf{I}_n - \Sigma\mathbf{D}) \end{aligned}$$

ya que por ser  $\mathbf{U}$  una matriz unitaria (ortogonal), conjugar una matriz por  $\mathbf{U}$  no afecta a la traza.

Aquí se aprecia una ventaja de la  $GCV$  frente a  $OCV$ . Para conocer los valores  $h_{ii}$  de  $\mathbf{H}$  (necesarios para evaluar  $OCV$ ) tendríamos que conocer  $\mathbf{U}$  y calcular  $\mathbf{H}$  según lo anterior. Sin embargo, para evaluar  $GCV$  sólo necesitamos la traza de  $\mathbf{H}$ , y la podemos conocer solamente a partir de los valores singulares de  $\mathbf{A}$  y de  $\lambda$ , sin conocer  $\mathbf{U}$ .

Finalmente obtenemos el denominador de la función  $GCV$ :

$$\begin{aligned}
 Tr(\mathbf{I}_n - \mathbf{H}) &= Tr(\mathbf{I}_n - \mathbf{\Sigma D}) = Tr(\mathbf{I}_n) - Tr(\mathbf{\Sigma D}) = \\
 &= n - \sum_{i=1}^q \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} = n - \sum_{i=1}^q \left(1 - \frac{\lambda^2}{\sigma_i^2 + \lambda^2}\right) = \\
 &= n - q + \sum_{i=1}^q \frac{\lambda^2}{\sigma_i^2 + \lambda^2} \tag{3.5}
 \end{aligned}$$

siendo  $q \leq m$  el número de valores singulares no nulos (rango de  $\mathbf{A}$ ).

Para obtener el numerador de  $GCV$  de forma simplificada, aún hay queda algo de trabajo. Recordemos que si  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$  es la SVD completa de  $\mathbf{A}$ , entonces habíamos llegado a que

$$\mathbf{Ax} = \mathbf{U}\mathbf{\Sigma}\mathbf{D}\mathbf{U}^T\mathbf{b}$$

Pues bien, resulta que la matriz  $\mathbf{U}\mathbf{\Sigma}\mathbf{D}\mathbf{U}^T$  es la misma si  $\mathbf{U}$ ,  $\mathbf{\Sigma}$  son los factores procedentes de la SVD completa o lo son de

la SVD compacta, es decir:

$$\mathbf{U}\mathbf{\Sigma}\mathbf{D}\mathbf{U}^T = \mathbf{U}^{(C)}\mathbf{\Sigma}^{(C)}\mathbf{D}^{(C)} \left(\mathbf{U}^{(C)}\right)^T$$

Se puede comprobar fácilmente teniendo en cuenta cómo se obtiene la SVD compacta a partir de la SVD completa. Por lo tanto, para lo que sigue podemos utilizar la SVD compacta en lugar de la SVD completa, puesto que es en general más eficiente computacionalmente. Omitiremos también el superíndice  $(C)$  para las matrices de la factorización, aunque no hay que olvidar que serán las de la SVD compacta.

El numerador de la función  $GCV$  es en general

$$\sum_{i=1}^n \left(b_i - \hat{f}_\lambda(t_i)\right)^2$$

pero en este caso, puesto que  $\hat{f}_\lambda(\mathbf{t}) = \hat{\mathbf{b}} = \mathbf{A}\mathbf{x}$ , es posible

escribirlo matricialmente como

$$\sum_{i=1}^n (b_i - \hat{f}_\lambda(t_i))^2 = \|\mathbf{b} - \mathbf{Ax}\|_2^2$$

Sabíamos que

$$\hat{\mathbf{b}} = \mathbf{Ax} = \mathbf{U}\Sigma\mathbf{D}\mathbf{U}^T\mathbf{b}$$

si definimos una nueva matriz  $\mathbf{F} = \mathbf{I}_n - \Sigma\mathbf{D}$ , se puede escribir:

$$\mathbf{Ax} = \mathbf{U}(\mathbf{I}_n - \mathbf{F})\mathbf{U}^T\mathbf{b} = \mathbf{U}\mathbf{U}^T\mathbf{b} - \mathbf{U}\mathbf{F}\mathbf{U}^T\mathbf{b}$$

no hay que olvidar que ahora  $\mathbf{U}\mathbf{U}^T \neq \mathbf{I}$  pues  $\mathbf{U}$  es el factor de la SVD compacta (aunque sí es cierto que  $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ , lo utilizaremos después). Entonces:

$$\mathbf{b} - \mathbf{Ax} = \mathbf{b} - \mathbf{U}\mathbf{U}^T\mathbf{b} + \mathbf{U}\mathbf{F}\mathbf{U}^T\mathbf{b}$$

Además resulta que el vector  $\mathbf{b} - \mathbf{U}\mathbf{U}^T\mathbf{b}$  es ortogonal al vector  $\mathbf{U}\mathbf{F}\mathbf{U}^T\mathbf{b}$ , veamos que el producto escalar de ambos es cero:

$$\begin{aligned} (\mathbf{U}\mathbf{F}\mathbf{U}^T\mathbf{b})^T (\mathbf{b} - \mathbf{U}\mathbf{U}^T\mathbf{b}) &= (\mathbf{b}^T\mathbf{U}\mathbf{F}^T\mathbf{U}^T) (\mathbf{b} - \mathbf{U}\mathbf{U}^T\mathbf{b}) = \\ &= \mathbf{b}^T\mathbf{U}\mathbf{F}^T\mathbf{U}^T\mathbf{b} - \mathbf{b}^T\mathbf{U}\mathbf{F}^T\mathbf{U}^T\mathbf{U}\mathbf{U}^T\mathbf{b} = \\ &= \mathbf{b}^T\mathbf{U}\mathbf{F}^T\mathbf{U}^T\mathbf{b} - \mathbf{b}^T\mathbf{U}\mathbf{F}^T\mathbf{U}^T\mathbf{b} = 0 \end{aligned}$$

Como hemos visto que son ortogonales, resulta que

$$\begin{aligned} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 &= \|\mathbf{b} - \mathbf{U}\mathbf{U}^T\mathbf{b} + \mathbf{U}\mathbf{F}\mathbf{U}^T\mathbf{b}\|_2^2 = \\ &= \|\mathbf{b} - \mathbf{U}\mathbf{U}^T\mathbf{b}\|_2^2 + \|\mathbf{U}\mathbf{F}\mathbf{U}^T\mathbf{b}\|_2^2 \end{aligned} \quad (3.6)$$

El primer sumando en (3.6), que es independiente de  $\lambda$ ,

se puede expresar así:

$$\begin{aligned}
 \|\mathbf{b} - \mathbf{U}\mathbf{U}^T\mathbf{b}\|_2^2 &= \|(\mathbf{I} - \mathbf{U}\mathbf{U}^T)\mathbf{b}\|_2^2 = \\
 &= ((\mathbf{I} - \mathbf{U}\mathbf{U}^T)\mathbf{b})^T ((\mathbf{I} - \mathbf{U}\mathbf{U}^T)\mathbf{b}) = \\
 &= \mathbf{b}^T (\mathbf{I} - \mathbf{U}\mathbf{U}^T)^T (\mathbf{I} - \mathbf{U}\mathbf{U}^T)\mathbf{b} = \mathbf{b}^T (\mathbf{I} - \mathbf{U}\mathbf{U}^T)^2 \mathbf{b} = \\
 &= \mathbf{b}^T (\mathbf{I} - 2\mathbf{U}\mathbf{U}^T + \mathbf{U}\mathbf{U}^T\mathbf{U}\mathbf{U}^T)\mathbf{b} = \mathbf{b}^T (\mathbf{I} - 2\mathbf{U}\mathbf{U}^T + \mathbf{U}\mathbf{U}^T)\mathbf{b} = \\
 &= \mathbf{b}^T (\mathbf{I} + \mathbf{U}\mathbf{U}^T)\mathbf{b} = \mathbf{b}^T\mathbf{b} - \mathbf{b}^T\mathbf{U}\mathbf{U}^T\mathbf{b} = \mathbf{b}^T\mathbf{b} - (\mathbf{b}^T\mathbf{U})(\mathbf{U}^T\mathbf{b}) = \\
 &= \|\mathbf{b}\|_2^2 - \|\mathbf{U}^T\mathbf{b}\|_2^2 \tag{3.7}
 \end{aligned}$$

Para terminar, el segundo sumando en (3.6), que sí depende de  $\lambda$ , se puede calcular de esta forma:

$$\begin{aligned}
 \|\mathbf{U}\mathbf{F}\mathbf{U}^T\mathbf{b}\|_2^2 &= (\mathbf{U}\mathbf{F}\mathbf{U}^T\mathbf{b})^T (\mathbf{U}\mathbf{F}\mathbf{U}^T\mathbf{b}) = \\
 &= (\mathbf{b}^T\mathbf{U}\mathbf{F}\mathbf{U}^T) (\mathbf{U}\mathbf{F}\mathbf{U}^T\mathbf{b}) =
 \end{aligned}$$

$$= \mathbf{b}^T \mathbf{U} \mathbf{F}^2 \mathbf{U}^T \mathbf{b} = (\mathbf{b}^T \mathbf{U} \mathbf{F}) (\mathbf{F} \mathbf{U}^T \mathbf{b}) = \|\mathbf{F} \mathbf{U}^T \mathbf{b}\|_2^2 \quad (3.8)$$

Finalmente, juntando los cálculos previos (3.5, 3.7 y 3.8), obtenemos la expresión eficiente para evaluar la función *GCV* en el caso de la Regularización de Tikhonov:

$$GCV(\lambda) = \frac{\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2}{Tr(\mathbf{I} - \mathbf{H})^2} = \frac{R_0 + R_\lambda}{(n - q + S_\lambda)^2} \quad (3.9)$$

donde  $R_0 = \|\mathbf{b}\|_2^2 - \|\mathbf{U}^T \mathbf{b}\|_2^2$  es independiente de  $\lambda$ ,

$R_\lambda = \|\mathbf{F} \mathbf{U}^T \mathbf{b}\|_2^2$  depende de  $\lambda$  a través de

$$\mathbf{F} = \text{diag} \left( \frac{\lambda^2}{\sigma_1^2 + \lambda^2}, \dots, \frac{\lambda^2}{\sigma_q^2 + \lambda^2} \right)$$

donde  $\sigma_i$  son los  $q$  valores singulares no nulos de  $\mathbf{A}$ ) y

$S_\lambda = \sum_{i=1}^q \frac{\lambda^2}{\sigma_i^2 + \lambda^2}$  es precisamente la traza de  $\mathbf{F}$ .



De esta forma, en el contexto de la regularización de Tikhonov, la expresión (3.9) nos permite evaluar la función  $GCV$  para los valores del parámetro  $\lambda$  que queramos, solamente conociendo la descomposición en valores singulares (SVD) compacta de la matriz del sistema  $\mathbf{A}$ .

#### 3.3.4. L-Curve

Un criterio alternativo a la  $GCV$  para la selección del parámetro  $\lambda$  en la regularización de Tikhonov, es el llamado criterio de L-Curve (o de la curva-L) [17]. La curva-L es una herramienta para el análisis de problemas mal condicionados, que consiste en un gráfico (para todos los valores del parámetro de regularización) de la norma de la solución regularizada  $\|\mathbf{x}\|_2$  frente a la norma del vector de residuos  $\|\mathbf{Ax} - \mathbf{b}\|_2$ , en escala logarítmica para las dos variables.

Esta curva representa el compromiso entre los dos sumandos en la minimización (3.4). Queremos que sean pequeñas las dos, pero sabemos que al aumentar la primera disminuye la segunda, y viceversa.

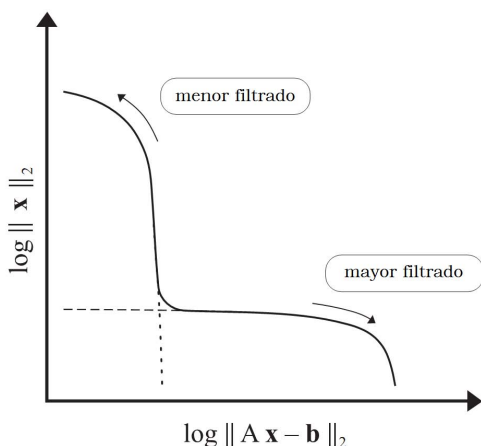


Figura 3.1: Forma característica de una curva-L.

Su representación gráfica normalmente tiene forma de L (ver Figura 3.1), de ahí el nombre de este criterio, con un vértice diferenciado que separa las partes vertical y horizontal de la curva. La parte superior (la más vertical) corresponde

a valores pequeños del parámetro  $\lambda$ , donde el filtrado o regularización es pequeño, y es posible disminuir mucho la norma del vector de coeficientes  $\mathbf{x}$  sin afectar demasiado la norma del residuo  $\mathbf{Ax} - \mathbf{b}$ . En la inferior (la más horizontal), la situación es justo la contraria, corresponde a valores grandes del parámetro de regularización, que provocan un aumento importante del error residual sin prácticamente disminuir la norma del vector de coeficientes. Por eso, el vértice de la curva es el punto óptimo para tener un equilibrio adecuado entre dichas normas.

En la práctica, para utilizar este criterio se calculan los valores de  $\|\mathbf{Ax} - \mathbf{b}\|_2$  y  $\|\mathbf{x}\|_2$  para un cierto conjunto de valores del parámetro  $\lambda$ , y se determina numéricamente dónde está el vértice de la curva y cual es su valor de  $\lambda$  asociado, que se toma como el óptimo para aplicar la regularización de Tikhonov.

### 3.4. Resultados Experimentales

En esta sección, presentamos una comparativa de los resultados numéricos obtenidos por el algoritmo de ajuste directo. Todos los procedimientos han sido implementado en Matlab 7 y se ejecutan en plataformas estándar (PC o Mac con la configuración usual). Los datos altimétricos y de curvatura procedentes de córneas reales usados fueron tomados por los topógrafos corneales CM02 y Eye Top de la marca CSO, que en condiciones ideales digitaliza hasta 24 anillos con 256 puntos equiespaciados angularmente por anilla. El tiempo de ejecución del ajuste directo, incluidas las estimaciones del parámetro de regularización, es muy bajo, aproximadamente 1 segundo por superficie, aproximadamente igual que el ajuste por polinomios de Zernike.

Para comparar los resultados de distintos métodos de reconstrucción, un criterio es el de mínimo  $MSE$  (mean squa-

red error, o error cuadrático medio). Si para los datos  $(z_k)_{k=1,2,\dots,N}$ , el modelo predice los valores  $(\hat{z}_k)_{k=1,2,\dots,N}$ , entonces el  $MSE$  se define como:

$$MSE = \frac{1}{N} \sum_{k=1}^N (z_k - \hat{z}_k)^2$$

Por supuesto, este no es el único criterio, pero es uno de los más usados. Un  $MSE$  pequeño indica que el modelo ajusta bien a los datos de entrada. Normalmente, también conviene tener en cuenta la complejidad del modelo, pues a igualdad de  $MSE$  es preferible un modelo más sencillo. Además, es posible estudiar la distribución estadística de los residuos  $(\varepsilon_k) = (z_k - \hat{z}_k)$ , y analizar por ejemplo si se distribuyen de forma normal. En el caso de datos corneales, la distribución del error de medición es poco conocida, con lo que este tipo de análisis se vuelve más complicado.

La primera cuestión es el tipo RBF a utilizar. Se han rea-

lizado distintos experimentos con gaussianas, multicuádricas y de soporte compacto. Aunque los resultados de estos experimentos no se muestran aquí, comentaremos que los resultados con los tres tipos de funciones son comparables, siendo las gaussianas las funciones más sencillas de las tres, y por tanto las que hemos decidido utilizar para el ajuste directo.

Una vez que hemos decidido utilizar las RBGF, hay que elegir cómo se colocan los centros. En la Figura 3.2 se muestra un ejemplo del tipo de malla que se utiliza para colocar los centros de las RBGF en el ajuste, una rejilla cuadrada regular. Otro tipo de distribuciones, como la hexagonal, han sido analizadas, pero no parecen aportar ventajas con respecto a la malla cuadrada escogida en la reconstrucción. En la práctica, parece razonable utilizar mallas de entre 50 y 200 centros aproximadamente.

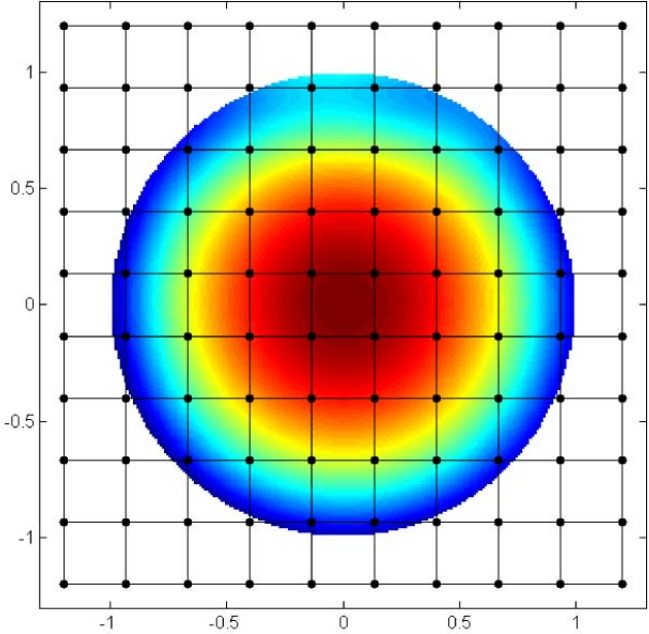


Figura 3.2: Malla cuadrada regular de 100 centros para el ajuste directo por RBFs y reconstrucción de una córnea el disco unidad.

Una vez decidida la colocación de los centros, la siguiente etapa es decidir el valor del parámetro de forma  $\alpha$  de las RBGF (en este algoritmo, es el mismo para todas). Basándonos en el  $MSE$  que se obtiene para distinta cantidad de funciones  $n$  en la reconstrucción, se concluye que un valor

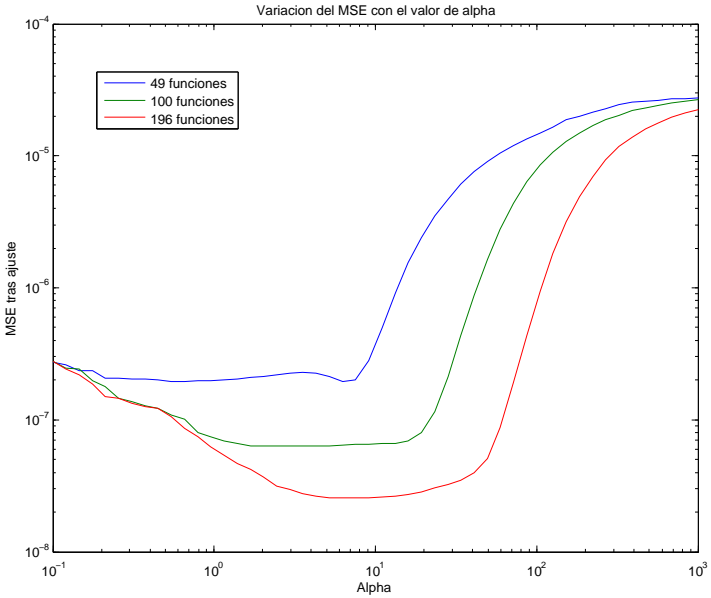


Figura 3.3: Variación del error cuadrático medio  $MSE$  según el valor del parámetro de forma  $\alpha$  de las RBGFs y según el tamaño de la base  $n$ .

cercano a 8 ó 10, aunque dentro de un rango suficientemente amplio, es donde se obtienen los mejores resultados para todos los valores de  $n$  (ver Figura 3.3).

Ahora, una vez fijados  $n$  y  $\alpha$ , estudiaremos como se comporta el problema del ajuste directo desde el punto de vista del número de condición.



La Figura 3.4 muestra el comportamiento del número de condición de la matriz de colocación del problema. Primero se observa cómo en el caso de no utilizar regularización, el número de condición de la matriz crece siempre con el tamaño de la base  $n$ . A partir de  $n = 300$ , la inestabilidad numérica es muy importante. Si utilizamos la regularización de Tikhonov, con el valor del parámetro  $\lambda$  dado por L-Curve, se observa cómo el número de condición de la matriz ampliada prácticamente es constante en torno a  $10^3$ , para cualquier tamaño  $n$  de la base .

En cambio, si utilizamos el criterio de GCV, el efecto de regularizar casi no se nota hasta  $n = 350$  aproximadamente; pero a partir de ahí de nuevo se estabiliza en un valor cercano a  $10^{14}$ . A pesar de que *GCV* consigue reducir el número de condición respecto a no usar la regularización, el valor en el que se estabiliza es excesivamente grande. Lo que sí consigue

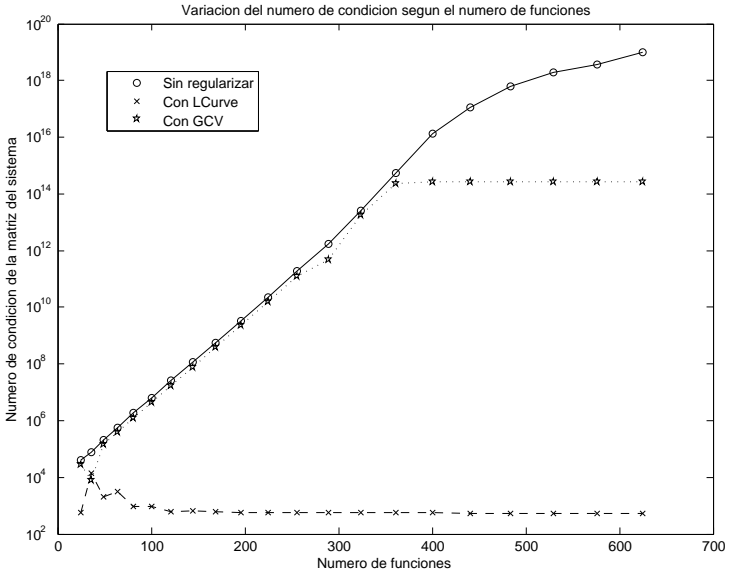


Figura 3.4: Número de condición del ajuste directo, sin regularización y con ella, usando los criterios L-Curve y GCV.

el criterio GCV es que la matriz del sistema siempre sea de rango maximal, propiedad que se pierde si no se usa regularización y el tamaño de la base es mayor a 400. Aparentemente, el criterio de GCV subestima el parámetro óptimo  $\lambda$ . Este hecho ya ha sido comentado en la literatura anteriormente (se puede consultar [44]).

Hemos visto que las consecuencias numéricas de elegir L-Curve o GCV son grandes. Las discrepancias entre los dos criterios para la selección del  $\lambda$  óptimo son importantes. En la Figura 3.5 se representan los valores de dicho parámetro que proporciona cada uno de los métodos, para distintos tamaños de la base  $n$ .

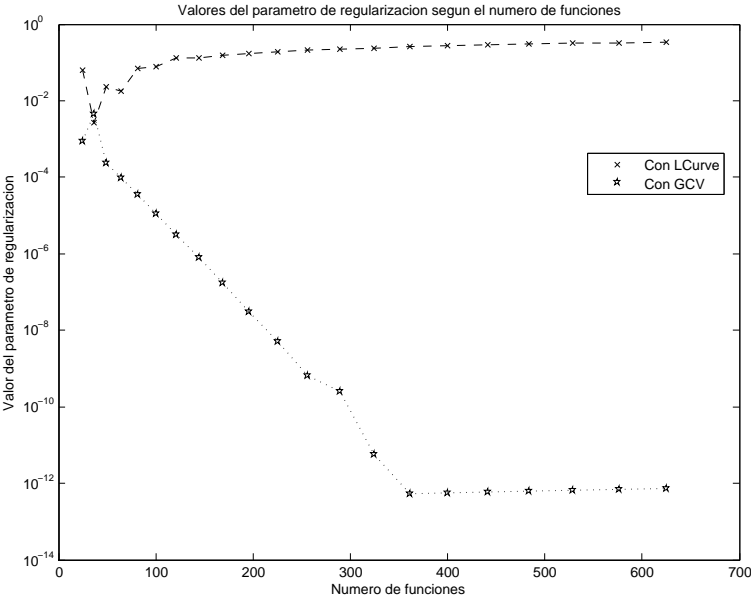


Figura 3.5: Valor de los parámetros de regularización dados según los criterios L-Curve y GCV.

Analicemos el comportamiento del error cuadrático medio  $MSE$  de la reconstrucción en tres casos: sin regularizar y con regularización, para los dos criterios estudiados (ver Figura 3.6).

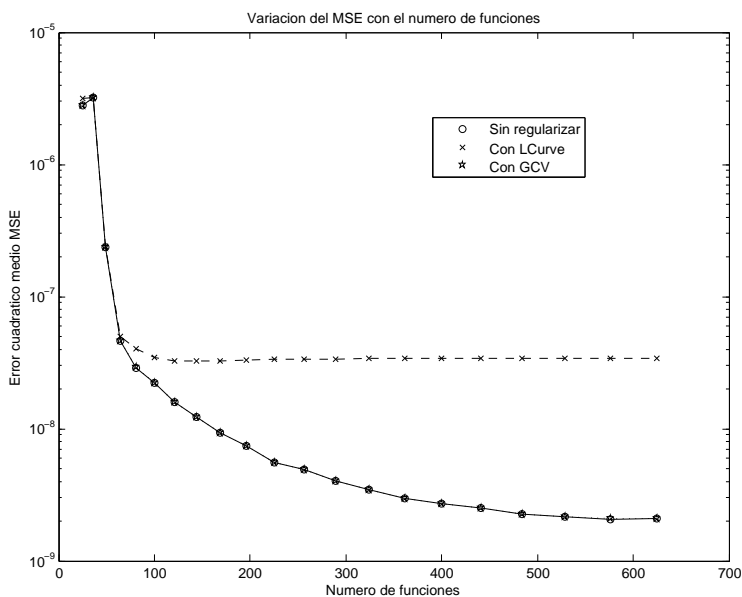


Figura 3.6: Variación del error cuadrático medio  $MSE$  segun el tamaño de la base  $n$ , utilizando la regularización con los dos criterios estudiados, y sin ella.

Observamos que, desde el punto de vista del  $MSE$  del ajuste, regularizar con  $GCV$  es casi equivalente a no regularizar, lo que es coherente con lo que muestra la Figura 3.4. En el caso de regularizar con L-Curve, es conveniente utilizar  $n$  próximo a 100, puesto que aunque crezca mucho el tamaño de la base, el  $MSE$  se mantiene constante.

Como conclusiones parciales, podemos decir que L-Curve es un criterio demasiado conservador. Tiende a proporcionar un número de condición bajo, lo cual no es malo, pero tampoco es necesario llegar a ese nivel. En cambio,  $GCV$  es más arriesgado: el número de condición de la matriz ampliada puede ser excesivamente alto. En consecuencia, tal vez un criterio combinado, que promedie de alguna forma los dos mencionados, proporcione los resultados óptimos en la práctica.

Finalmente, se muestran algunas córneas reconstruïdas con el ajuste directo. La primera, en la Figura 3.7, se mues-

tran las curvas de nivel de la reconstrucción total (esfera de mejor ajuste + ajuste por RBGF). En la Figura 3.8, se muestran otras dos reconstrucciones, esta vez solamente el sumando  $E$  en (3.1) (se omite la esfera de mejor ajuste). Este tipo de representación facilita la visualización, ya que la forma esférica puede encubrir los principales rasgos de una córnea.

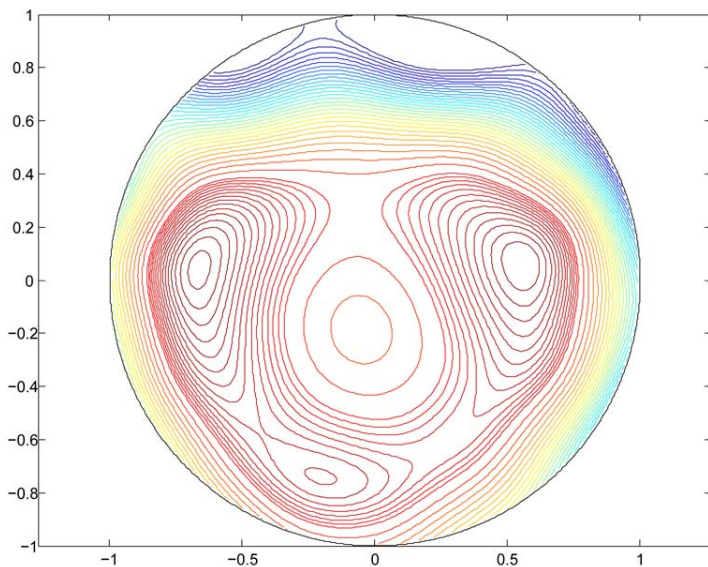


Figura 3.7: Curvas de nivel de la reconstrucción de una córnea mediante el ajuste directo.

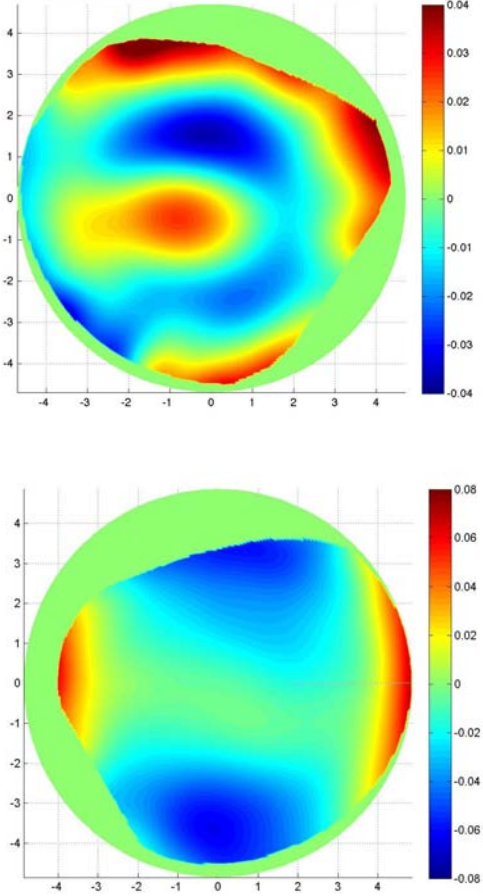


Figura 3.8: Reconstrucción con el ajuste directo de dos superficies corneales, previamente ajustadas por la esfera de mejor ajuste. Arriba: córnea operada mediante ablación láser. Abajo: córnea sana, sin cirugía.





## Capítulo 4

# Ajuste Iterativo Adaptativo

Vamos a desarrollar aquí un algoritmo original para la reconstrucción de superficies, utilizando las llamadas funciones de base radial gaussianas anisótropa (A-RBGF). Igual que en el caso del ajuste directo, es conveniente pre-ajustar los datos que suministra el topógrafo corneal por una esfera (cuando sean de altimetría) o por un plano (cuando sean de curvatura). A continuación, el residuo será ajustado por una combinación lineal de las A-RBGF, que serán construídas

dinámicamente a partir de los propios datos, y no fijas como en el caso anterior. Para que un algoritmo de este tipo funcione correctamente, los criterios para seleccionar los centros de las A-RBGF, sus parámetros de forma y coeficientes, deben escogerse con cuidado, lo cual será discutido a lo largo del capítulo.

Otra cuestión importante es decidir cuántas funciones de este tipo se utilizan para la reconstrucción. Se comentan algunos criterios de parada que pueden utilizarse, además de la opción de ajustar con un determinado número de funciones y aplicar un filtrado posterior. Finalmente, se muestra la potencia del método con algunos resultados experimentales, basados en nuestras simulaciones con ordenador, la mayoría de ellas con córneas reales.

## 4.1. Pre-ajuste

Consideraremos los mismos datos de entrada que para el ajuste directo, es decir, una nube tridimensional de puntos  $(x_k, y_k, z_k)$ ,  $k = 1, \dots, N$ . El pre-ajuste por la esfera, comentado en el Capítulo 2, es idéntico ahora. A los errores residuales, que ahora irán cambiando con cada iteración, serán ahora  $\varepsilon_k^{(1)} = z_k - S(x_k, y_k)$ , y el superíndice indica que estos son los residuos iniciales, antes de ejecutar el algoritmo iterativo que se va a describir.

Naturalmente, si no pre-ajusta por la esfera, es perfectamente posible tomar  $\varepsilon_k^{(1)} = z_k$  y utilizar el algoritmo igualmente. O bien, si se prefiere ajustar por un plano, simplemente hay que restar a los datos su media, y definir  $\varepsilon_k^{(1)} = z_k - \bar{z}$ . Para los datos corneales, es siempre conveniente preajustar por la esfera o por el plano, y entonces reconstruir los residuos.

## 4.2. Procedimiento de ajuste

Como resultado del paso anterior, los errores residuales  $\varepsilon_k^{(1)}$  contienen información relevante de la superficie corneal además de ruido. Nuestro objetivo es ajustar esos residuos por una función  $E(x, y)$ , de tal forma que la expresión analítica para la altura de la córnea viene dada por

$$\text{Cornea}(x, y) = S(x, y) + E(x, y), \quad (4.1)$$

De esta forma,  $S$  contiene la forma esencial de la córnea, mientras que en  $E$  son capturadas las pequeñas irregularidades de la superficie. La función  $E$  está dada por una combinación lineal de  $n$  funciones de un diccionario dado,

$$E(x, y) = E_n(x, y) = \sum_{j=1}^n c_j h_j(x, y). \quad (4.2)$$

En un esquema ideal,  $n$  depende de los datos reales, y debe ser lo bastante grande para que toda la información relevante de las medidas de elevación esté modelada en  $E$ , pero no excesivamente grande para evitar sobre-parametrizar el problema y ajustar el ruido. Para solventar las ya mencionadas dificultades de los polinomios de Zernike, usaremos como diccionario las funciones de base radial gaussianas, de la forma:

$$h(x, y) = \exp\left(-\|P - Q\|_A^2\right), \quad P = (x, y)^T,$$

donde el superíndice  $T$  denota la trasposición matricial,  $Q = (Q_x, Q_y)^T$  es un cierto punto del plano (llamado el “centro” de la función) y  $A$  es una matriz definida positiva en  $\mathbb{R}^{2 \times 2}$ . Para una tal matriz, la  $A$ -norma de un punto (vector columna) se

define como

$$\|\mathbf{P}\|_A = \sqrt{\mathbf{P}^T \mathbf{A} \mathbf{P}} = \sqrt{\alpha_x x^2 + \alpha_y y^2 + 2\alpha_{xy} xy},$$

para  $A = \begin{pmatrix} \alpha_x & \alpha_{xy} \\ \alpha_{xy} & \alpha_y \end{pmatrix}$  con  $\alpha_x > 0$  y  $\alpha_x \alpha_y > \alpha_{xy}^2$ .

En general, estas son funciones de base radial anisótropas, que resultan ser funciones de base radial estándar cuando los dos autovalores de  $A$  coinciden (en otras palabras, cuando  $A$  es un múltiplo positivo de la matriz identidad  $I_2$ ).

Por tanto, buscamos una expresión de la forma

$$\text{Cornea}(x, y) = S(x, y) + \sum_{j=1}^n c_j \exp\left(-\left\|\mathbf{P} - \mathbf{Q}^{(j)}\right\|_{A_j}^2\right), \quad \mathbf{P} = (x, y)^T.$$

Claramente, cualquier procedimiento de ajuste debe hacer una selección adecuada de todos los parámetros, a saber:

- centros  $Q^{(j)}$ ;
- matrices de forma (o parámetros de forma)  $A_j$ ;
- coeficientes (factores de escala)  $c_j$ ;
- número de términos  $n$  en la representación funcional.

Proponemos un algoritmo iterativo de reconstrucción, tal que en cada paso, se ajusta parcialmente el error residual por una RBGF anisotrópica (A-RBGF), y se calculan los nuevos residuos, que serán los datos de entrada para la siguiente iteración. Para mantener al máximo los grados de libertad de la reconstrucción, los centros, los parámetros (o matrices) de forma y los coeficientes serán seleccionados dinámicamente dependiendo de los datos residuales en cada etapa.

### 4.3. Descripción del algoritmo iterativo

Supongamos que  $E_{j-1}$  ya está calculado previamente (tomamos  $E_0 \equiv 0$ ). Los datos de entrada para la  $j$ -ésima iteración ( $j = 1, 2, \dots$ ) es la nube  $(x_k, y_k, \varepsilon_k^{(j)})$  de nodos  $\mathbf{P}_k = (x_k, y_k)^T$  y sus correspondientes residuos  $\varepsilon_k^{(j)}$ ,  $k = 1, 2, \dots, N$ ; recordamos que  $\varepsilon_k^{(1)} = z_k - S(x_k, y_k)$  son los primeros datos residuales tras el ajuste inicial por la esfera de mejor ajuste. Realizamos los siguientes pasos:

PASO 1: selección de los centros  $\mathbf{Q}^{(j)}$ .

El problema de selección del centro de una función de base radial se ha discutido en [23]. En este artículo, los centros se escogen como nodos  $\mathbf{P}_k$  de entre los propios datos, usando el criterio de maximizar la “cross-correlación”. Este método, a pesar de ser computacionalmente intensivo, se puede implementar para mejorar los resultados en el paso 1. Sin embargo, en la práctica hemos encontrado que el criterio (mucho más



simple) de escoger como el centro el nodo con mayor residuo (en valor absoluto) proporciona resultados igualmente satisfactorios, con un coste mucho menor; además, este criterio se correlaciona con la geometría de la A-RBGF, ya que ésta alcanza su máximo en el centro. Por tanto, escogemos:

$$\mathbf{Q}^{(j)} = (x_{k_0}, y_{k_0})^T \quad \text{donde} \quad k_0 = \arg \max_k |\varepsilon_k^{(j)}|$$

and denote

$$m^{(j)} := \varepsilon_{k_0}^{(j)}.$$

#### PASO 2: filtrado dinámico

Como se mencionaba antes, los datos altimétricos obtenidos de los dispositivos de medida como queratógrafos están contaminados por ruido. A pesar de que hay alguna información sobre la distribución estadística de esos errores, la información es aún muy limitada. Para afrontar este problema,

necesitamos filtrar aquellos datos que claramente corresponden a errores de medición y por tanto estropean la calidad de la reconstrucción. Esto se puede hacer previamente, antes de comenzar el procedimiento de ajuste, como en [2]. Pero aquí hemos escogido una alternativa más simple, que da buenos resultados: una vez que el centro  $Q^{(j)}$  se ha seleccionado, comprobamos el número,  $\ell_k$ , de nodos  $P_k$  que caen dentro del mayor disco posible centrado en  $Q^{(j)}$  que contiene solamente nodos cuyos residuos tienen el mismo signo que  $m^{(j)}$ . Es decir, consideramos los nodos más cercanos al centro con residuos del mismo signo. Si  $\ell_k < 20$ , consideramos que  $Q^{(j)}$  es un dato anómalo y no lo consideramos como un centro válido en esta iteración. Esto se puede hacer simplemente poniendo  $\varepsilon_{k_0}^{(j)} = 0$ , después de lo cual se vuelve al paso 1. De otro modo, avanzamos al siguiente paso.

Obviamente, este paso se puede ignorar si sabemos que los

datos son exactos o si el error que contienen es despreciable. Como ya se ha dicho, en el caso de superficies corneales, esto no es así, por lo que es necesario incluir esta etapa.

PASO 3: selección de los parámetros de forma

Primero determinamos los nodos de influencia  $\mathcal{P}_j(s)$ . Estos son el conjunto maximal de como máximo  $s$  nodos  $\mathbf{P}_k$  cercanos a  $\mathbf{Q}^{(j)}$  con residuos del mismo signo que  $m^{(j)}$ . Hay que observar que  $\mathcal{P}_j(s_1) \subset \mathcal{P}_j(s_2)$  si  $s_1 \leq s_2$ . Es conveniente paralelizar los cálculos siguientes para varios valores de  $s$ : hemos realizado experimentos numéricos usando los valores  $s = [s_{\text{mín}}, 100, 150, 200, 300]$ , donde  $s_{\text{mín}} = \text{mín}(\ell_k, 50)$ , y  $\ell_k$  fue definido en el Paso 2.

Las condiciones de interpolación

$$\varepsilon_{k_0}^{(j)} h_j(x_k, y_k) = \varepsilon_k^{(j)}, \quad k \in \mathcal{P}_j(s),$$

son equivalentes al sistema lineal de ecuaciones sobredeter-

minado

$$\begin{aligned} & \alpha_x (x_k - x_{k_0})^2 + 2\alpha_{xy} (x_k - x_{k_0})(y_k - y_{k_0}) \\ & + \alpha_y (y_k - y_{k_0})^2 = \log \left( \frac{\varepsilon_{k_0}^{(j)}}{\varepsilon_k^{(j)}} \right), \quad k \in \mathcal{P}_j(s) \end{aligned} \quad (4.3)$$

cuyas incógnitas son las 3 entradas desconocidas de la matriz de forma

$$A_j = \begin{pmatrix} \alpha_x & \alpha_{xy} \\ \alpha_{xy} & \alpha_y \end{pmatrix}.$$

Este problema se resuelve en el sentido de mínimos cuadrados ponderados (MCP), donde la  $k$ -ésima ecuación se multiplica por  $\eta_k := (1 + \left\| \mathbf{P}_k - \mathbf{Q}^{(j)} \right\|^2)^{-1}$  para considerar la mayor influencia de los nodos próximos al centro  $\mathbf{Q}^{(j)}$  en la matriz de forma  $A_j$ . Esta solución se obtiene por métodos estándar, como la factorización  $QR$  de la matriz de colocación correspondiente al sistema (4.3), o la descomposición en valores singulares. Se puede consultar por ejemplo [41].

Hay que remarcar que debido a la selección del centro actual  $Q^{(j)}$ ,

$$t_k := \log \left( \frac{\varepsilon_{k_0}^{(j)}}{\varepsilon_k^{(j)}} \right) \geq 0, \quad k \in \mathcal{P}_j(s).$$

Sin embargo, esta condición no garantiza que la matriz solución  $A_j$  of (4.3) en el sentido de mínimos cuadrados ponderados descrita arriba sea definida positiva. Esto normalmente ocurre en la periferia de la envolvente convexa de los nodos, donde la ausencia de datos en alguna dirección puede propiciar una matriz  $A_j$  no definida positiva. A pesar de que la correspondiente función  $h_j$  ajustaría los datos correctamente de forma local, no es válida, ya que crecería exponencialmente en la dirección del autovalor negativo de  $A_j$ , destruyendo así el ajuste previo lejos del centro de la función.

Para solucionar este problema, examinamos la solución  $A_j$  del sistema (4.3): si no es definida positiva, interpreta-

mos que hay una carencia de datos en un entorno de  $\mathbf{Q}^{(j)}$  y forzamos a la función  $h_j$  a ser una función de base radial isométrica (estándar):  $A_j = \alpha I_2$ . De esta forma, el sistema (4.3) se reduce a

$$\alpha \left\| \mathbf{P}_k - \mathbf{Q}^{(j)} \right\|^2 = t_k, \quad k \in \mathcal{P}_j(s) \quad (4.4)$$

donde la solución en el sentido de MCP es

$$\alpha = \sum_{k \in \mathcal{P}_j(s)} \theta_k t_k,$$

with

$$\theta_k = \frac{\eta_k^2 \left\| \mathbf{P}_k - \mathbf{Q}^{(j)} \right\|^2}{\sum_{t \in \mathcal{P}_j(s)} \eta_t^2 \left\| \mathbf{P}_t - \mathbf{Q}^{(j)} \right\|^4}$$

y  $\eta_k = \left( 1 + \left\| \mathbf{P}_k - \mathbf{Q}^{(j)} \right\|^2 \right)^{-1}$ . Observar que en este caso  $\alpha$

es positivo por construcción, y definimos

$$h_j(x, y) = \exp\left(-\alpha \left\| \mathbf{P} - \mathbf{Q}^{(j)} \right\|^2\right), \quad \mathbf{P} = (x, y)^T.$$

PASO 4: selección del factor de escala.

Podemos calcular el coeficiente  $c_j$  de

$$c_j h_j(x_k, y_k) = \varepsilon_k^{(j)}, \quad k \in \mathcal{P}_j(s),$$

usando MCP con los mismos pesos  $\eta_k$ :

$$c_j = \sum_{k \in \mathcal{P}_j(s)} \gamma_k \varepsilon_k^{(j)},$$

with

$$\gamma_k = \frac{\eta_k^2 h_j(x_k, y_k)}{\sum_{t \in \mathcal{P}_j(s)} \eta_t^2 h_j(x_t, y_t)^2}.$$

Sin embargo, hay que anotar que los experimentos numéricos revelan que la condición de interpolación  $c_k = m^{(j)}$  da

resultados que no son peores, siendo bastante más simple.

PASO 4: cálculo de los nuevos residuos.

Con los valores de  $c_j$  y  $A_j$  recién calculados, obtenemos

$$\varepsilon_k^{(j+1)} = \varepsilon_k^{(j)} - c_j h_j(x_k, y_k).$$

Como se mencionó antes, todos los cálculos se han realizado de forma paralela para distintos valores de  $s$  (típicamente, de 3 a 5 valores, entre 50 and 300), y por tanto, distintos conjuntos anidados de nodos de influencia  $\mathcal{P}_j(s)$  se han considerado simultáneamente. Ahora, mantenemos como válido el valor de  $s$  (y sus correspondiente valores de  $c_j$  y  $A_j$ ) que proporciona la menor norma del vector de residuos ( $\varepsilon_k^{(j+1)}$ ), y desechamos el resto de valores. Observar que una vez hecho esto, el valor de  $s$  ya no tiene influencia ninguna en la reconstrucción, solamente sus valores  $c_j$  y  $A_j$  asociados. Como resultado, encontramos la nueva aproximación  $E_j = E_{j-1} + c_j h_j$ .



Como paso final, comprobamos el criterio de parada, que va a ser discutido a continuación. Si este criterio no se verifica, incrementamos el contador de iteración  $j$  en 1 y volvemos al Paso 1.

Como resultado, tras la ejecución del ciclo iterativo, se obtiene una expresión analítica para la superficie del tipo

$$E(\mathbf{P}) = \sum_{j=1}^n c_j \exp\left(-\left\|\mathbf{P} - \mathbf{Q}^{(j)}\right\|_{A_j}^2\right), \quad \mathbf{P} = (x, y) \quad (4.5)$$

#### 4.4. Criterio de parada

En teoría, el algoritmo se ejecuta indefinidamente, dando una función interpolante, y en consecuencia, un vector residual nulo. En la vida real, con datos contaminados de ruido es muy importante el problema de seleccionar el orden del modelo: queremos capturar toda la información relevante, pero con el menor número posible de funciones, sin sobre-parametrizar

el modelo y sin ajustar el ruido. Muchas soluciones a este problema se han descrito en la literatura. Por ejemplo, la elección del número de polinomios de Zernike para la reconstrucción modal de datos altimétricos se ha discutido en [3, 22].

Los métodos estadísticos de selección del número  $n$  apropiado en (4.2) normalmente hacen hipótesis sobre el ruido. Sin embargo, apenas hay información a priori sobre cotas de ruido de medición o sobre cómo se distribuye. De acuerdo con [40], los errores no se pueden considerar como variables aleatorias i.i.d., a pesar de que la hipótesis de que están normalmente distribuidos (con varianza proporcional al cuadrado de la distancia del nodo al centro de la córnea) es aparentemente razonable. Estos criterios suelen ser computacionalmente intensivos [3, 18].

Métodos menos costosos utilizan criterios de teoría de información, como el Criterio de Información de Akaike (AIC),

o el Criterio de Detección Eficiente (EDC) [33], que estudia la evolución de

$$EDC_j(p) = N \log(MSE_j) + j(N \log N)^p, \quad 0 < p < 1,$$

con

$$MSE_j = \frac{\rho^2}{N} \sum_{k=1}^N \left( \varepsilon_k^{(j)} \right)^2, \quad \rho = \max_k \|P_k\|. \quad (4.6)$$

El valor de  $p$  normalmente se escoge experimentalmente.

Sin embargo, se puede ganar información analizando directamente el comportamiento de los errores normalizados definidos en (4.6). Típicamente, esos errores empiezan a decaer con tasa exponencial y orden de convergencia medio mayor que 1. Después de un número de iteraciones, se observa una estabilización de esta tasa de decaimiento, que se hace lineal; esto normalmente pasa cuando los valores de  $MSE_j$

están entre  $10^{-3}$  y  $10^{-4} \mu\text{m}^2$ . Basado en la experiencia, hemos utilizado con éxito el siguiente criterio de parada: permitimos que el algoritmo se ejecute hasta 50 iteraciones (esto es prácticamente instantáneo) y calculamos las pendientes ponderadas

$$\delta MSE_j = \frac{\log(MSE_{j+1}/MSE_j)}{j}.$$

La sucesión  $\delta MSE_j$ , a pesar de ser oscilatoria, es negativa y tiende a cero, así que buscamos la última iteración  $1 \leq J_1 \leq 50$  en la que  $\delta MSE_{J_1} \geq -0,02$ . Si  $MSE_{J_1} < 10^{-3} \mu\text{m}^2$ , fijamos  $J_1 + 1$  como la iteración de parada. Si no, buscamos la última iteración  $1 \leq J_2 \leq 50$  en la que  $\delta MSE_{J_2} \geq -0,01$ , y paramos el algoritmo en la  $(J_2 + 1)$ -ésima iteración.

Una alternativa al criterio de parada, descrita en [31], es aplicar el algoritmo con un número fijo de iteraciones y posteriormente filtrar aquellas funciones que consideramos innecesarias o perjudiciales para la reconstrucción. La representa-

ción funcional (4.5) permite un análisis multi-resolución, en el cual las bandas de valores  $\text{Vol}_j := c_j / \sqrt{\det A_j}$  (un valor proporcional al volumen total bajo la  $j$ -ésima función) puede especificar diferentes niveles de resolución. En particular, se puede considerar que las A-RBGF con valores de  $\text{Vol}_j$  pequeños corresponden al ruido.

Esto motiva el enfoque de filtrado dado en [12, 13] para la eliminación del ruido, poniendo a cero todos los coeficientes  $c_j$  para los que la condición  $\text{Vol}_j < \varepsilon$  se cumpla. La selección del parámetro  $\varepsilon > 0$  juega obviamente un papel crucial, y su valor debe estar basado o bien en la distribución estadística del ruido (poco conocida, en el caso de datos corneales) o en experimentos previos.

## 4.5. Resultados Experimentales

En esta sección, presentamos una comparativa de los resultados numéricos obtenidos por nuestro método aplicado tanto a córneas simuladas como sintéticas. Todos los procedimientos han sido implementado en Matlab 7 y se ejecutan en plataformas estándar (PC o Mac con la configuración usual). Los datos altimétricos y de curvatura procedentes de córneas reales usados fueron tomados por los topógrafos corneales CM02 y Eye Top de la marca CSO, que en condiciones ideales digitaliza hasta 24 anillos con 256 puntos equiespaciados angularmente por anilla.

Ya que el procedimiento esta pensado para reconstrucción en tiempo real de los datos corneales, cualesquiera métodos computacionalmente intensivos deben ser descartados. Sin embargo, en nuestra implementación en Matlab, el tiempo de ejecución es muy reducido, aproximadamente 1 segundo

por superficie.

Primero mostramos la potencia de la metodología propuesta usando algunos modelos de superficies elementales, comenzando por el ejemplo mas sencillo: una superficie que es la combinación lineal de tres exponenciales,

$$\text{Cornea}(x, y) = \sum_{j=1}^3 c_j \exp\left(-\|P - Q_j\|_{A_j}^2\right), \quad P = (x, y)^T,$$

con  $Q_1 = (0, 3, -0, 4)^T$ ,  $Q_2 = (0, 7, 0, 1)^T$ ,  $Q_3 = (-0, 4, 0, 3)^T$ ,

$c_1 = 0, 02$ ,  $c_2 = -0, 015$ ,  $c_3 = 0, 02$ , y

$$A_1 = \begin{pmatrix} 10 & -7 \\ -7 & 20 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 15 & 3,5 \\ 3,5 & 10 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 20 & -6 \\ -6 & 12 \end{pmatrix}.$$

Esta superficie obviamente está exagerada de lo que seria el residuo normal (respecto a la esfera) de una cornea normal, pero se ha considerado importante incluirlo con fines

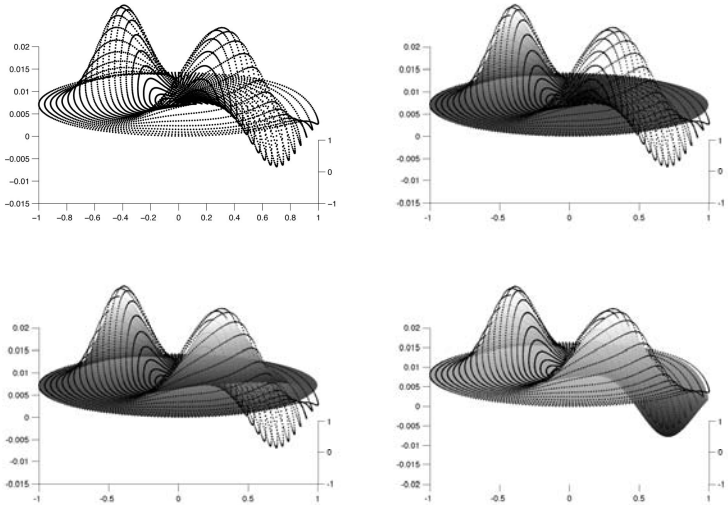


Figura 4.1: Ejemplo del rendimiento del algoritmo iterativo en el caso sencillo de datos muestreados de una superficie dada por tres gaussianas (arriba a la izquierda): la primera iteración (arriba a la derecha) localiza el “pico” más alto, mientras que las dos siguientes (abajo) capturan el resto de la superficie original.

ilustrativos. El resultado de las tres primeras iteraciones del algoritmo se pueden ver en la Figura 4.1.

Ahora se añade un ruido blanco de medición (ruido Gaussiano con media cero) a los datos de la superficie anterior. Concretamente, con varianza del 10% de la altura máxima



de la superficie. En la mayoría de aplicaciones en óptica esto correspondería a un nivel de ruido tremendamente grande, teniéndose prácticamente siempre niveles menores. Conviene remarcar que para el algoritmo propuesto no es necesario conocer la distribución estadística del ruido. Como resultado de la ejecución, los tres centros de las Gaussianas fueron estimados correctamente, a pesar de que los autovalores de las matrices de forma  $A_j$  fueron ligeramente sobreestimados (ver la Figura 4.2); esta sobreestimación se corrige en las iteraciones siguientes del algoritmo, añadiendo algunas otras A-RBGF con centros concentrados alrededor de los tres centros de la superficie original.

El error global, medido por  $MSE_j$ , tiende a decrecer monótonamente con tasa exponencial. El comportamiento típico para la reconstrucción de datos altimétricos tanto para corneales reales y sintéticas se puede observar en la Figu-

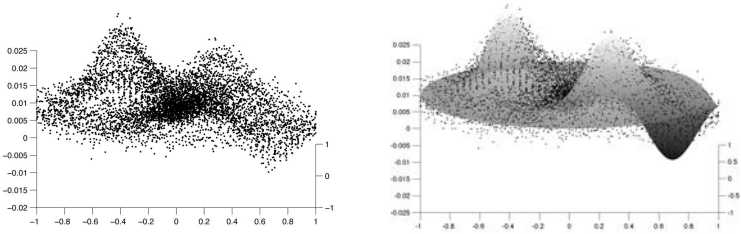


Figura 4.2: Rendimiento del algoritmo con los mismos datos de la Figura 4.1, contaminados con ruido blanco del 10% de la altura máxima. Se muestran los datos con ruido (arriba a la izquierda)

ra 4.3. En particular, en todos los casos se observa el comportamiento antes mencionado de la transición del decaimiento supra-exponencial al lineal, el cual se usa como criterio de parada. Para comparar, se han reconstruido los mismos datos con polinomios de Zernike utilizando el método estándar de mínimos cuadrados. El  $MSE_j$  para la reconstrucción con Zernikes se muestra en los mismos ejes coordenados, donde  $j$  indica el número total de polinomios de Zernike utilizados.

Recordamos que

$$j = \frac{(m+1)(m+2)}{2},$$

donde  $m$  es el máximo orden radial utilizado.

Observamos dos fenómenos destacables en esas gráficas. Primero, los polinomios de Zernike rápidamente capturan la forma global de la córnea, lo cual se expresa en la rápida caída inicial del error. Sin embargo, los pequeños detalles de la superficie (como las áreas localizadas con mayor pendiente) son mucho más difícil de ajustar con esta herramienta. Ello explica la clara saturación en el comportamiento del error con Zernikes después de una cierta cantidad de polinomios (típicamente, entre 21 y 36 polinomios, que corresponden a un orden radial  $5 \leq m \leq 7$ ). Esto no ocurre en el caso de la reconstrucción con A-RBGF, cuyo carácter adaptativo y multi-escala permite que se ajusten correctamente a los datos

en cada iteración.

Por otro lado, los experimentos numéricos muestran otro fenómeno interesante relacionado con el criterio de parada que se ha propuesto: en la mayoría de situaciones el momento de parada se corresponde a un número de iteraciones  $j$  para el cual  $MSE_j$  es aproximadamente igual para los dos métodos, el ajuste por A-RBGF y por polinomios de Zernike.

A pesar de lo que puedan ilustrar estas gráficas, el error global no es la mejor forma de comparar los dos métodos de reconstrucción. Recordamos que el enfoque modal con polinomios de Zernike se plantea precisamente para lograr el menor  $MSE_j$  posible para cada  $j$ ; mientras que el algoritmo iterativo propuesto aquí tiene otro objetivo totalmente distinto: localizar la parte más “abultada” de la superficie, modelarla por una A-RBGF de parámetros adecuados e incorporarla a la expresión analítica (4.1).

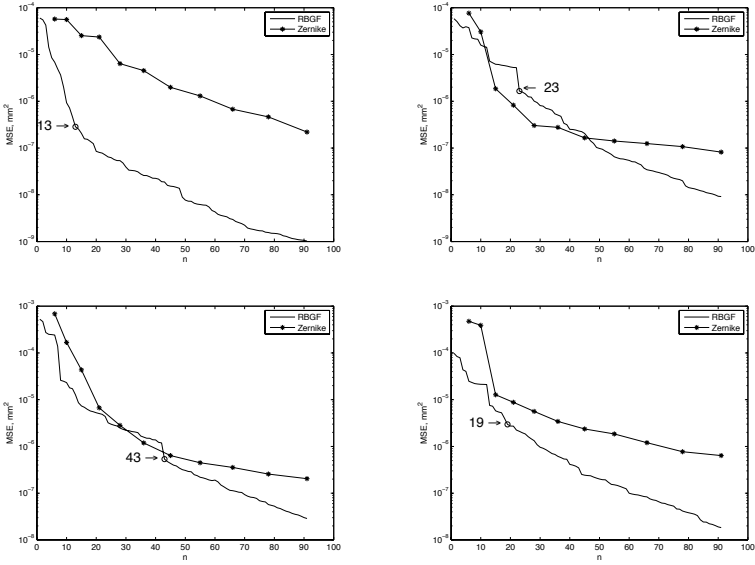


Figura 4.3: Comparativa de  $MSE_j$  para reconstrucción de datos altimétricos por polinomios de Zernike y con el ajuste por A-RBGF. Arriba a la izquierda: córnea simulada, dada como combinación lineal de 10 funciones gaussianas. Arriba a la derecha: córnea normal (sana). Abajo: dos córneas con queratocono. La flecha señala el momento de parada segun el criterio descrito antes.

Esto se puede ilustrar ajustando una córnea sintética con una cicatriz simulada, usada anteriormente en [27]. Su mapa de curvas de nivel está representado en la Figura 4.4, arriba a la izquierda. Arriba a la derecha se muestran las curvas de nivel de la superficie reconstruida con el algoritmo adaptativo descrito aquí, usando 20 funciones. Abajo, los otros dos gráficos corresponden a las curvas de nivel de la misma superficie reconstruida con polinomios de Zernike, concretamente la de la izquierda con 36 (orden radial 7) y la de la derecha con 136 (orden radial 15).

La situación se vuelve más evidente si comparamos los errores residuales sobre la octava anilla para ambos métodos (Figura 4.5, izquierda): mientras que los polinomios de Zernike funcionan perfectamente en las partes suaves de la superficie, tienen dificultades para adaptarse en las zonas de pendiente más cambiante, mientras que el algoritmo basado

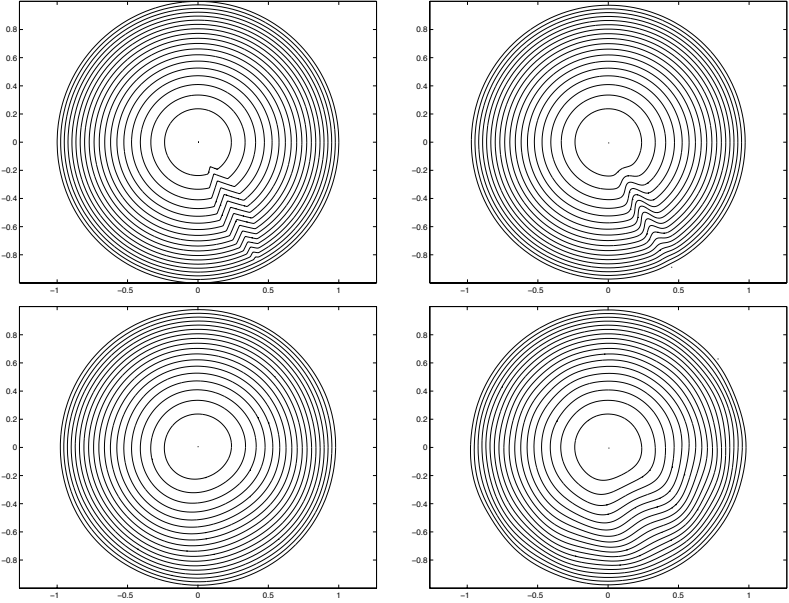


Figura 4.4: Comparación de la reconstrucción de una córnea simulada con una “cicatriz” (arriba a la izquierda) con 20 iteraciones del algoritmo adaptativo (arriba a la derecha); y con 36 y 136 polinomios de Zernike (abajo, izquierda y derecha resp.)

en A-RGBF utiliza su carácter adaptativo y multi-escala para ajustar la superficie casi perfectamente a lo largo de toda la anilla después de pocas iteraciones. Para el ajuste con polinomios de Zernike, al *MSE* le cuesta bastante avanzar, como se ilustra en la Figura 4.5, derecha.

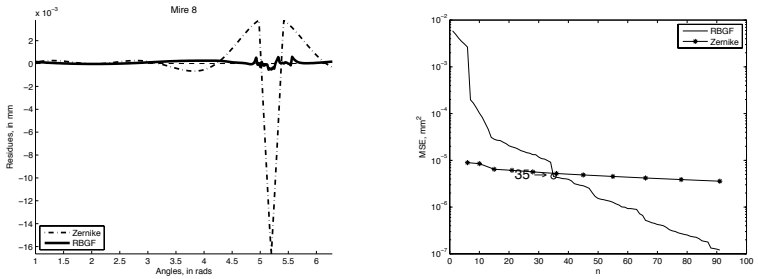


Figura 4.5: Izquierda: errores residuales en la anilla 8 para una la córnea simulada con cicatriz, reconstruida con 20 iteraciones del algoritmo adaptativo (línea gruesa) y con 36 polinomios de Zernike (línea discontinua). Derecha:  $MSE_j$  para la recontrucción de la misma córnea con polinomios de Zerinke y con el ajuste por A-RBGF.

Otra indicación del comportamiento consistente del algoritmo iterativo propuesto es la evolución de los parámetros calculados dinámicamente en cada iteración. A pesar de que los autovalores de la matriz de forma  $A_j$  tienden a crecer y pueden llegar a ser bastante grandes (cuando se ajusta una zona pequeña muy “abrupta”), su cociente (llamado el número de condición espectral de  $A_j$ ) se mantiene acotado, con algunas excepciones; recordamos que el número de condición 1



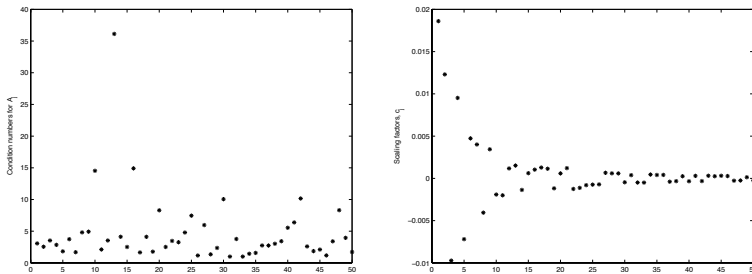


Figura 4.6: Evolución típica del número de condición de la matriz  $A_j$  (parámetro de forma de la A-RBGF) (izquierda) y factores de escala  $c_j$  (derecha).

corresponde a una función de base radial estándar (isotrópica). Por otro lado, los factores de escala  $c_j$  decrecen de forma estable, de acuerdo con la reducción gradual de los errores residuales (Figura 4.6).

Observando el momento de parada, los experimentos realizados con córneas reales y sintéticas muestran que el número razonable de iteraciones está entre 20 y 40; no hay una clara correlación entre el número de iteraciones y el estado clínico (normal, enferma, etc) de la córnea, como muestra la Figura 4.3. Por este motivo, hemos considerado apropiado realizar

50 iteraciones (aprovechando la velocidad del algoritmo) para a posteriori decidir el número adecuado de funciones  $n$  en la expresión funcional (4.2).

Sin embargo, parece hacer más correlación con la posición y el agrupamiento de los centros  $Q^{(j)}$  de las A-RBGF: para corneas normales, los centros normalmente se colocan en la parte más periférica, donde hay las oscilaciones más importantes, mientras que para corneas afectadas de queratocono observamos cómo algunos centros coinciden con la deformación (el llamado “cono”) ya en las primeras iteraciones (Figura 4.7).

El algoritmo iterativo descrito es adecuado para la reconstrucción de cualquier superficie a partir de un conjunto discreto de puntos de ella. En particular, se puede reconstruir un mapa de poder corneal, o el frente de onda (ver Figura 4.9). Teniendo en cuenta la forma típica de tales superficies, es

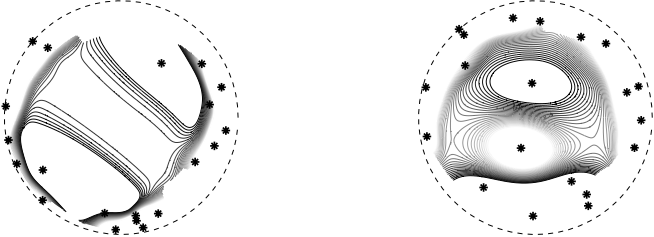


Figura 4.7: Posiciones de los centros de las primeras 20 A-RBGF para una córnea normal (izquierda) y una con queratocono (derecha). Las líneas son las curvas de nivel de la córnea reconstruída, una vez eliminada la esfera de mejor ajuste.

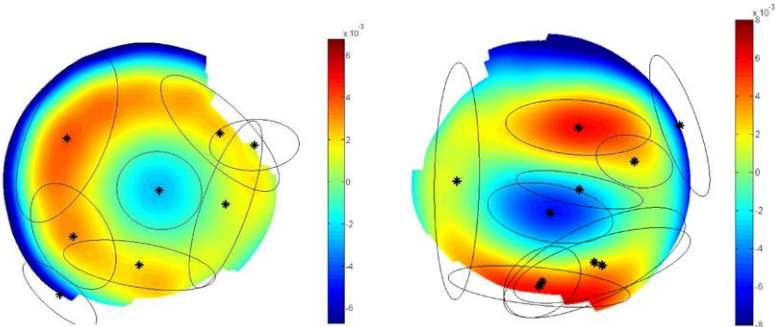


Figura 4.8: Córneas reconstruídas con el algoritmo basado en A-RBGF: córnea normal (izquierda) y queratocónica (derecha). Se muestran además los centros de las A-RBGF y curvas de nivel de cada una, que son elipes o circunferencias.

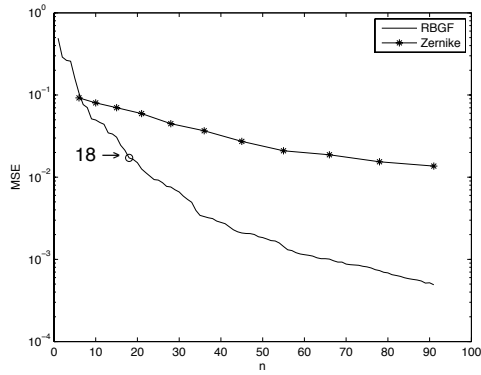


Figura 4.9: Evolución de  $MSE_j$  para datos de poder refractivo corneal reconstruidos por polinomios de Zernike y por el ajuste por A-RBGF.

conveniente saltarse el paso de ajuste previo por la esfera de mejor ajuste, haciendo  $S \equiv 0$  in (4.1). O alternativamente, ajustar un plano a los datos y calcular los residuos respecto de él (esto equivale a restar a los datos su propia media).

## Capítulo 5

# Conclusiones

En este trabajo, desarrollamos un método adaptativo de ajuste para datos corneales, combinando la simplicidad del enfoque modal con las ventajas de una reconstrucción zonal. Consiste en un ajuste preliminar (no es imprescindible) de los datos con alguna función global (una esfera en el caso de datos altimétricos, o bien un plano para datos de curvatura, etc) y un procedimiento iterativo que añade términos a la representación funcional de la superficie. Cada término consiste en una función de base radial anisotrópica, de tipo gaussiano,

con una escala adecuada. Sus parámetros se calculan dinámicamente y permiten ajustar los datos en cada iteración independientemente de la escala. El método también incluye un procedimiento de filtrado para descartar los datos anómalos (aquellos que corresponden claramente a errores de medición) y un criterio de parada que permite elegir el número final de términos en la representación analítica de la superficie, de acuerdo con la evolución del error residual.

La implementación numérica de este algoritmo en un ordenador personal estándar es muy rápida (el tiempo de ejecución esta por debajo de 2 segundos). Los resultados experimentales nos permiten obtener las siguientes conclusiones:

- el ajuste por mínimos cuadrados de una combinación lineal de polinomios de Zernike de orden radial bajo (entre 4 y 6, como máximo) es adecuado para los casos de córneas normales. Se puede usar además para

obtener las principales características de la forma de la superficie. Sin embargo, para córneas altamente aberradas, el procedimiento basado en Zernike se satura relativamente pronto, y necesitamos un número exageradamente grande de términos para alcanzar el nivel de precisión requerido en las regiones de mayor pendiente, lo que conlleva un sobre-parametrizado del modelo y necesariamente ajustar gran parte del ruido de la medida. Finalmente, la complejidad de cada término en la representación funcional aumenta con el índice.

- en contraste, el método iterativo presentado aquí muestra una caída exponencial constante del error de ajuste, independientemente de la complejidad de la córnea. Su tasa real está influida básicamente por la distribución de los residuos: la rápida caída en las primeras iteraciones, cuando se reconstruyen las partes más salientes,

es seguida por una caída lineal estable, cuando esencialmente lo que se ajusta es el ruido. Esto se puede utilizar como criterio de parada para el procedimiento iterativo. De esta forma, se utiliza la mínima cantidad posible de funciones en la representación analítica de cada córnea.

- como alternativa al criterio de parada, es posible utilizar un filtrado posterior de las funciones base, eliminando aquellas con volumen muy pequeño, que son las que producen las mayores oscilaciones en la reconstrucción y suponemos que corresponden al ruido. De esta forma, la superficie final queda suavizada sin apenas alterarla, y con un número pequeño de términos en su representación funcional.
  
- a diferencia de lo que ocurre con los polinomios de Zernike, la complejidad de cada término en la representa-



ción funcional con las A-RBGF es la misma, pero sus parámetros cambian para adaptarse a la escala actual. Esta escala está determinada solamente por los errores residuales y no por el número de iteración.

- debido al carácter localizado de las A-RBGF, la ubicación y agrupamiento de sus centros, así como el tamaño de sus parámetros de forma, proporciona una información espacial adicional sobre las regiones de mayor irregularidad. Estas ideas han sido usadas en la elaboración de nuevos índices corneales de irregularidad para detección de enfermedades de la córnea, y están actualmente en estudio [32].
- la sensibilidad de los coeficientes de Zernike al descentrado de los datos con respecto al ápex de la córnea se ha notificado en la literatura repetidamente. Debido al carácter local inherente al método de ajuste por A-

RBGF, este método es mucho más estable con respecto al desalineamiento.

- la reconstrucción basada en Zernikes es también muy sensible a anillas con datos incompletos. En la práctica clínica, ésto está usualmente provocado por obstrucción de pestañas o la falta de fluido en la capa lagrimal. Aún así, una buena parte de los datos de cada anilla incompleta está disponible, y es utilizada para la reconstrucción iterativa mediante A-RBGF, siendo muchos sensible a esta información faltante que el ajuste por Zernike.
- para el ajuste directo, la base está fija y es independiente del conjunto de datos. Esto permite precalcular y almacenar los valores de las derivadas y coeficientes de cuadratura de las funciones de la base, permitiendo la obtención de forma muy rápida y precisa de gradientes,

vectores normales y curvaturas de la superficie. Todos estos parámetros tienen gran importancia en la óptica y la visión.

- un inconveniente del algoritmo de A-RBGF es que la evaluación de la superficie reconstruida no es tan rápida como en el caso del ajuste directo, ya que en esta ocasión la base no es fija, si no que depende de cada superficie en concreto. Además no es posible precalcular las derivadas direccionales u otros parámetros, aunque debido a la expresión analítica sencilla de una A-RBGF, sí es posible obtener fórmulas cerradas con más o menos facilidad. A pesar de todo, en la práctica es posible utilizar el algoritmo con prácticamente la misma eficiencia que el ajuste directo u otros métodos existentes, como los polinomios de Zernike.
- el ajuste directo presenta fundamentalmente un incon-

veniente: la sobreparametrización. Es posible que en ocasiones, muchas de las funciones de la representación funcional sean innecesarias para la reconstrucción de una córnea en concreto.

- además de los algoritmos de reconstrucción, se ha estudiado el procedimiento general de la regularización de Tikhonov, aplicable a cualquier sistema lineal mal condicionado, así como dos de los principales criterios que se utilizan para estimar el valor apropiado del parámetro de regularización, L-Curva y GCV. De paso, también se ha visto cómo la descomposición en valores singulares (SVD) compacta permite reducir extraordinariamente la dificultad a la hora de evaluar la función de cross validación generalizada, que de otra manera tendría un coste computacional muy elevado y no serviría de nada en la práctica.

Los algoritmos propuesto aquí para el modelado de la córnea proporcionan un método para obtener una descripción matemática compacta de la topografía o el mapa de curvaturas de la córnea. Toda la información está codificada en un pequeño número de parámetros: radio y centro de la esfera de mejor ajuste, además de las posiciones de los centros, parámetros de escala y factores de escala para cada una de las funciones base. En el caso del ajuste directo, el número de parámetros es aún menor, ya que los centros se colocan de forma regular y el parámetro de escala es el mismo para todas las funciones base. Los dos métodos han sido validados experimentalmente, dando unos resultados comparables o mejores que los procedimientos usuales existentes.

Esta descripción puede ser utilizada para la visualización global de la córnea o algunas de sus partes, capturando detalles de la superficie que escapan a los procedimientos

estándar. Puede servir además como información de entrada para remuestreo o para el cálculo de otro tipo de cuestiones importantes en óptica, como el trazado de rayos, frente de onda, integración numérica y otros. Por ejemplo, los coeficientes de Zernike constituyen un método común de descripción de la córnea en la práctica clínica. La representación en términos de polinomios de Zernike para la representación obtenida mediante funciones de base radial, ya sean las estándar o las anisótropas, puede ser calculada sin dificultad mediante los procedimientos usuales de cuadratura.

# Índice de figuras

2.1. Fotografía de un topógrafo corneal basado en discos de Plácido (izquierda) y nube de puntos de la cara anterior de la córnea obtenida por él (derecha). . . . .	16
2.2. Representación de los primeros polinomios de Zernike en el disco unidad, numerados con índice simple (j) y doble (n,m), junto al nombre que se les suele dar. . . . .	22
2.3. Representación gráfica de una RBF gaussiana, centrada en el origen, en perspectiva (izquierda) y en planta (derecha, se ha señalado también el centro), con parámetro de forma $\alpha = 1$ (arriba) y $\alpha = 5$ (abajo). . . . .	25
2.4. Representación gráfica de una RBF gaussiana anisótropa (A-RBGF), centrada en el origen, en perspectiva (izquierda) y en planta (derecha, se ha señalado también el centro), con parámetro de forma la matriz $A$ anterior. . .	27
3.1. Forma característica de una curva-L. . . . .	70

3.2.	Malla cuadrada regular de 100 centros para el ajuste directo por RBFs y reconstrucción de una córnea el disco unidad. . . . .	75
3.3.	Variación del error cuadrático medio $MSE$ según el valor del parámetro de forma $\alpha$ de las RBGFs y según el tamaño de la base $n$ . . . .	76
3.4.	Número de condición del ajuste directo, sin regularización y con ella, usando los criterios L-Curve y GCV. . . . .	78
3.5.	Valor de los parámetros de regularización dados según los criterios L-Curve y GCV. . . .	79
3.6.	Variación del error cuadrático medio $MSE$ según el tamaño de la base $n$ , utilizando la regularización con los dos criterios estudiados, y sin ella. . . . .	80
3.7.	Curvas de nivel de la reconstrucción de una córnea mediante el ajuste directo. . . . .	82
3.8.	Reconstrucción con el ajuste directo de dos superficies corneales, previamente ajustadas por la esfera de mejor ajuste. Arriba: córnea operada mediante ablación láser. Abajo: córnea sana, sin cirugía. . . . .	83
4.1.	Ejemplo del rendimiento del algoritmo iterativo en el caso sencillo de datos muestreados de una superficie dada por tres gaussianas (arriba a la izquierda): la primera iteración (arriba a la derecha) localiza el “pico” más alto, mientras que las dos siguientes (abajo) capturan el resto de la superficie original. . . . .	108



- 4.2. Rendimiento del algoritmo con los mismos datos de la Figura 4.1, contaminados con ruido blanco del 10 % de la altura máxima. Se muestran los datos con ruido (arriba a la izquierda) 110
  
- 4.3. Comparativa de  $MSE_j$  para reconstrucción de datos altimétricos por polinomios de Zernike y con el ajuste por A-RBGF. Arriba a la izquierda: córnea simulada, dada como combinación lineal de 10 funciones gaussianas. Arriba a la derecha: córnea normal (sana). Abajo: dos córneas con queratocono. La flecha señala el momento de parada segun el criterio descrito antes. . . . . 113
  
- 4.4. Comparación de la reconstrucción de una córnea simulada con una “cicatriz” (arriba a la izquierda) con 20 iteraciones del algoritmo adaptativo (arriba a la derecha); y con 36 y 136 polinomios de Zernike (abajo, izquierda y derecha resp.) . . . . . 115
  
- 4.5. Izquierda: errores residuales en la anilla 8 para una la córnea simulada con cicatriz, reconstruída con 20 iteraciones del algoritmo adaptativo (línea gruesa) y con 36 polinomios de Zernike (línea discontinua). Derecha:  $MSE_j$  para la reconstrucción de la misma córnea con polinomios de Zerinke y con el ajuste por A-RBGF. 116
  
- 4.6. Evolución típica del número de condición de la matriz  $A_j$  (parámetro de forma de la A-RBGF) (izquierda) y factores de scala  $c_j$  (derecha). . . . . 117

- 4.7. Posiciones de los centros de las primeras 20 A-RBGF para una córnea normal (izquierda) y una con queratocono (derecha). Las líneas son las curvas de nivel de la córnea reconstruída, una vez eliminada la esfera de mejor ajuste. . . 119
- 4.8. Córneas reconstruídas con el algoritmo basado en A-RBGF: córnea normal (izquierda) y queratocónica (derecha). Se muestran además los centros de las A-RBGF y curvas de nivel de cada una, que son elipes o circunferencias. 119
- 4.9. Evolución de  $MSE_j$  para datos de poder refractivo corneal reconstruídos por polinomios de Zernike y por el ajuste por A-RBGF. . . . 120

# Bibliografía

- [1] S.J. Ahn, W. Rauh, and H.J. Warnecke. Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola and parabola. *Pattern Recognition*, 34 (12):2283–2303, 2001.
- [2] W. Alkhaldi, D.R. Iskander, A.M. Zoubir, and M.J. Collins. Enhancing the standard operating range of a placido disk videokeratoscope for corneal surface estimation. *IEEE Transactions on Biomedical Engineering*, 56 (3):800–809, 2009.
- [3] D. Alonso-Caneiro, D.R. Iskander, and M.J. Collins. Estimating corneal surface topography in videokeratoscopy in the presence of strong signal interference. *IEEE Transactions on Biomedical Engineering*, 55 (10):2381–2387, 2008.
- [4] M. Ares and S. Royo. Comparison of cubic b-spline and zernike-fitting techniques in complex wavefront reconstruction. *Applied Optics*, 45:6945–6964, 2006.
- [5] P. Artal, E. Berrio, A. Guirao, and P. Piers. Contribution of the cornea and internal surfaces to the change

- of ocular aberrations with age. *Journal of the Optical Society of America*, 19:137–143, 2002.
- [6] R. Beatson, O. Davydov, and J. Levesley. Error bounds for anisotropic rbf interpolation. *Journal of Approximation Theory*, 162 (3):512–527, 2010.
- [7] M. Bozzini, L. Lenarduzzi, M. Rossini, and R. Shaback. Interpolation by basis functions of different scales and shapes. *Calcolo*, 41 (2):77–87, 2004.
- [8] C. Campbell. Reconstruction of the corneal shape with the mastervue corneal topography system. *Optometry and Vision Science*, 74:899–905, 1997.
- [9] L.A. Carvalho. Accuracy of zernike polynomials in characterizing optical aberrations and the corneal surface of the eye. *Investigative Ophthalmology and Visual Science*, 46 (6):1915–1926, 2005.
- [10] L.A. Carvalho. Preliminary results of neural networks and zernike polynomials for classification of videokeratography maps. *Optometry and Vision Science*, 82 (2):151–158, 2005.
- [11] G. Conforti. Zernike aberration coefficients from seidel and higher-order power-series coefficients. *Optics Letters*, 8 (7):407–408, 1983.
- [12] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society: Series B*, 57 (2):301–369, 1995.
- [13] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Density estimation by wavelent thresholding. *Annals of Statistics*, 24 (2):508–539, 1996.

- 
- [14] J. Espinosa, D. Mas, J. Pérez, and C. Illueca. Optical surface reconstruction technique through combination of zonal and modal fitting. *Journal of Biomedical Optics*, 15 (2):026022, 2010.
- [15] G.E. Fasshauer. *Meshfree approximation methods with Matlab*. World Scientific Publishing, Singapore, 2007.
- [16] M.A. Halstead, B. Barsky, S.A. Klein, and R.B. Mandell. A spline surface algorithm for reconstruction of corneal topography from a videokeratographic reflection pattern. *Optometry and Vision Science*, 72 (11):821–827, 1995.
- [17] C. Hansen. *Rank-deficient and discrete ill-posed problems: Numerical aspects of linear inversion*. SIAM, Philadelphia, 1998.
- [18] D.R. Iskander, W. Alkhaldi, and A.M. Zoubir. On the computer intensive methods in model selection. In *33th IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008.
- [19] D.R. Iskander and M.J. Collins. Applications of high-speed videokeratoscopy. *Clinical and Experimental Optometry*, 88 (4):223–231, 2005.
- [20] D.R. Iskander, M.J. Collins, and B. Davis. Optimal modeling of corneal surfaces with zernike polynomials. *IEEE Transactions on Biomedical Engineering*, 48 (1):87–95, 2001.
- [21] D.R. Iskander, M.J. Collins, and B. Davis. Evaluating tear film stability in the human eye with high-speed videokeratoscopy. *IEEE Transactions on Biomedical Engineering*, 52 (11):1939–1949, 2005.

- [22] D.R. Iskander, M.R. Morelande, M.J. Collins, and B. Davis. Modeling of corneal surfaces with radial polynomials. *IEEE Transactions on Biomedical Engineering*, 56 (3):800–809, 2009.
- [23] A.A. Jamshidi and M.J. Kirby. Towards a black box algorithm for nonlinear function approximation over high-dimensional domains. *SIAM Journal on Scientific Computing*, 29 (3):941–963, 2007.
- [24] S.D. Klyce, M.D. Karon, and M.K. Smolek. Advantages and disadvantages of the zernike expansion for representing wave aberration of the normal aberrated eye. *Journal of Refractive Surgery*, 20 (5):S527–S541, 2004.
- [25] X. Liu and Y. Gao. B-spline based wavefront reconstruction for lateral shearing interferometric measurement of engineering surfaces. *Advances in Abrasive Technology*, 238-239:169–174, 2003.
- [26] N. Maeda, S.D. Klyce, and M.K. Smolek. Comparison of methods for detecting keratoconus using videokeratography. *Archives of Ophthalmology*, 113 (7):870–874, 1995.
- [27] A. Martínez-Finkelshtein, A.M. Delgado, G.M. Castro-Luna, A. Zarzo, and J.L. Alió. Comparative analysis of some modal reconstruction methods of the shape of the cornea from corneal elevation data. *Investigative Ophthalmology and Visual Science*, 50 (12):5639–5645, 2009.
- [28] A. Martínez-Finkelshtein, D. Ramos-López, G.M. Castro-Luna, and J.L. Alió. Procedimiento de reconstrucción de la topografía corneal a partir de datos alimétricos o de curvatura, 2010.

- [29] A. Martínez-Finkelshtein, D. Ramos-López, G.M. Castro-Luna, and J.L. Alió. Adaptive corneal modeling from keratometric data. *Investigative Ophthalmology and Visual Science*, 52 (8):4963–4970, 2011.
- [30] J. Nemeth, B. Erdelyi, B. Csakany, P. Gaspar, A. Soumelidis, F. Kahlesz, and Z. Lang. High-speed videotopographic measurement of tear film build-up time. *Investigative Ophthalmology and Visual Science*, 43 (6):1783–1790, 2002.
- [31] D. Ramos-López and A. Martínez-Finkelshtein. Multiscale modeling by anisotropic gaussian functions with applications to the corneal topography. In *CMMSE 2011: Proceedings of the 11th International Conference on Mathematical Methods in Science and Engineering, Alicante 26-30 June 2011: ISBN: 978-84-614-6167-7*, 2011.
- [32] D. Ramos-López, A. Martínez-Finkelshtein, G.M. Castro-Luna, D. Piñero, and J.L. Alió. Placido-based indices of corneal irregularity. *Optometry and Vision Science*, 88 (10), 2011.
- [33] C.R. Rao and Y. Wu. A strongly consistent procedure for model selection in a regression problem. *Biometrika*, 76 (2):369–374, 1989.
- [34] S.A. Read, M.J. Collins, D.R. Iskander, and B.A. Davis. Corneal topography with scheinplflug imaging and videokeratography : comparative study of normal eyes. *Journal of Cataract and Refractive Surgery*, 35(6):1072–1081, 2009.
- [35] M. Schneider, D.R. Iskander, and M.J. Collins. Modeling corneal surfaces with rational functions for high-speed

- videokeratoscopy data compression. *IEEE Transactions on Biomedical Engineering*, 56 (2):493–499, 2009.
- [36] J.T. Schwiegerling and J.E. Greivenkamp. Keratoconus detection based on videokerastoscopic height data. *Investigative Ophthalmology and Visual Science*, 73:721–728, 1989.
- [37] M.K. Smolek and S.D. Klyce. Current keratoconus detection methods compared with a neural network approach. *Investigative Ophthalmology and Visual Science*, 38 (11):2290–2299, 1997.
- [38] M.K. Smolek and S.D. Klyce. Zernike polynomial fitting fails to represent all visually significant corneal aberrations. *Investigative Ophthalmology and Visual Science*, 44 (11):4676–4681, 2003.
- [39] M.K. Smolek and S.D. Klyce. Goodness-of-prediction of zernike polynomial fitting to corneal surfaces. *Journal of Cataract and Refractive Surgery*, 31 (12):2350–2355, 2005.
- [40] W. Tang, M.J. Collins, L.G. Carney, and B. Davis. The accuracy and precision performance of four videokeratoscopes in measuring test surfaces. *Optometry and Vision Science*, 77 (9):483–491, 2000.
- [41] L.N. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM, 1997.
- [42] R.K. Tyson. Conversion of zernike aberration coefficients to seidel and higher-order power-series aberration coefficients. *Optics Letters*, 7 (6):262–264, 1982.



- [43] G. Wahba. *Spline models for observational data*. SIAM, Philadelphia, 1990.
- [44] Y. Wang. Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, 93 (441):341–348, 1998.
- [45] M. Zhu, M.J. Collins, and D.R. Iskander. Dynamics of ocular surface topography. *Eye*, 21:624–632, 2007.