



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS MATEMÁTICAS

UN SISTEMA PARA LA RECONSTRUCCIÓN DE SUPERFICIES MEDIANTE
TÉCNICAS DE INTERPOLACIÓN DE DATOS DISPERSOS

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS

PRESENTA:
JORGE ZAVALETA SÁNCHEZ

DIRECTOR DE LA TESIS
DR. PABLO BARRERA SÁNCHEZ
FACULTAD DE CIENCIAS

MÉXICO, D. F. A 31 DE OCTUBRE DE 2013.

Prefacio

En este trabajo se está interesado en el problema de interpolación de datos dispersos, que en términos simples, se refiere al problema de ajustar una superficie con ciertas propiedades a través de muestras de datos con distribución no uniforme o aleatoria. Este tema es de gran importancia práctica dentro de diversos campos de la ciencia e ingeniería debido a que en ellos es necesario analizar y modelar diversos fenómenos físicos, experimentos, entre otros, a partir de datos obtenidos en posiciones irregulares dentro de un dominio de interés, por lo que requieren utilizar la interpolación para propagar la información sobre todas las posiciones de dicho dominio para ayudar con las tareas de análisis y modelación.

Esta variedad de aplicaciones en la que aparece el problema de interpolación de datos dispersos motivo el desarrollo de una herramienta computacional dentro de `Matlab`, el sistema *UNAMSI*, que además de facilitar su resolución, permite utilizar algunas técnicas con diferentes enfoques que posibilita al usuario usar su experiencia para elegir la solución que mejor se adecue a sus necesidades. Entre las técnicas empleadas están los métodos de distancia inversa, las funciones de base radial, B-splines multinivel y en particular kriging que es probablemente la más utilizada en la práctica. Además, dentro del sistema se ha creado un módulo para la construcción de mallas 3D a partir de la reconstrucción de superficies, ya que en numerosas aplicaciones es necesario generar una malla estructurada 3D que sea adecuada para la simulación numérica de algún fenómeno dentro de una región interés, como por ejemplo, la simulación numérica del flujo de hidrocarburos dentro de un yacimiento.

Aquí se presenta el sistema de reconstrucción de superficies mediante técnicas de interpolación dispersa y su aplicación para la generación de mallas 3D. Con este fin, se describirán primeramente los métodos usados para después mostrar el sistema y las ideas de la construcción de la malla utilizadas en el módulo de construcción. Dado que el enfoque de este trabajo es la aplicación de dichos métodos, se extiende la teoría lo necesario para cubrir las necesidades prácticas que se buscan, dando las referencias pertinentes sobre los temas que no llegaron a ser tratados a profundidad.

El presente escrito está estructurado de la siguiente manera:

- *Capítulo 1.* Se da una introducción y formulación del problema a tratar a modo de motivación del trabajo.
- *Capítulo 2.* A partir de este capítulo se comienzan a describir los métodos de interpo-

lación de datos dispersos usados, comenzando con los métodos de distancia inversa.

- *Capítulo 3.* Aquí se presentan algunos aspectos sobre la teoría de las funciones de base radial.
- *Capítulo 4.* Se describen las bases del método kriging que se distingue por su enfoque estadístico del problema de interpolación y su gran uso en la práctica.
- *Capítulo 5.* Se presenta el método de aproximación B-splines multinivel.
- *Capítulo 6.* Se describen las partes esenciales y el uso del sistema *Gaia SR*, además del módulo de construcción de mallas 3D.
- *Capítulo 7.* Se muestran las conclusiones y el trabajo a futuro.
- *Apéndice A.* Exhibe algunos resultados auxiliares de los métodos de distancia inversa y de funciones de base radial.
- *Apéndice B.* Es una breve revisión de conceptos básicos de la teoría de la probabilidad como complemento de los conceptos necesarios para kriging.

Índice general

Prefacio	III
1. Introducción	1
1.1. Creación de modelos de fenómenos físicos	2
1.1.1. Modelación de un yacimiento de hidrocarburos	3
1.2. Interpolación y aproximación de datos dispersos	4
2. Métodos de distancia inversa	7
2.1. Método de Shepard	7
2.2. Método de Shepard modificado	18
2.3. Aspectos prácticos	23
3. Funciones de base radial	29
3.1. Interpolación unidimensional	29
3.1.1. Interpolación lineal	29
3.1.2. Splines cúbicos	31
3.2. Interpolación multivariada	34
3.3. Matrices de interpolación no singulares	36
3.3.1. Funciones definidas positivas	36
3.3.2. Funciones condicionalmente definidas positivas	42
3.4. Aspectos prácticos	49
4. Métodos geoestadísticos	57
4.1. Preliminares	57
4.2. Algunos aspectos sobre el variograma	59
4.2.1. Variograma Experimental	60
4.2.2. Propiedades de la función de covarianza y del variograma teórico	61
4.3. Kriging	70
4.3.1. Kriging Simple	71
4.3.2. Kriging Ordinario	74
4.3.3. Kriging Universal	77
4.4. Aspectos prácticos	79

5. B-spline multinivel	87
5.1. Preliminares	87
5.2. Aproximación B-Spline	92
5.2.1. Idea Básica	92
5.2.2. Algoritmo	97
5.3. Aproximación B-Spline Multinivel	99
5.3.1. Algoritmo Básico	100
5.4. Algoritmo B-spline Multinivel con refinamiento	102
5.4.1. Refinamiento del spline bicúbico uniforme	103
5.4.2. Algoritmo con refinamiento B-spline	108
6. Sistema y aplicaciones.	111
6.1. Sistema <i>UNAMSI v1.0</i>	111
6.1.1. Abrir, crear o importar un conjunto de datos.	112
6.1.2. Guardar un conjunto de datos.	114
6.1.3. Herramientas y opciones de visualización y modificación de los datos.	115
6.1.4. Reconstrucción de superficies	118
6.1.5. Cálculo y ajuste de variogramas	123
6.2. Módulo de construcción de mallas 3D.	127
6.2.1. Descripción del módulo.	128
7. Conclusiones	133
Apéndices	
A. Resultados auxiliares	135
A.1. Algunos resultados sobre el interpolador $F_{S_0}^*$	135
A.2. Espacios de funciones y optimalidad de las funciones de base radial	137
A.2.1. Espacios de Hilbert con kernel reproductor y espacios nativos para funciones estrictamente definidas positivas	138
A.2.2. La optimalidad de la interpolación con funciones de base radial	140
B. Conceptos básicos de la teoría de la probabilidad.	143
B.1. Variables aleatorias	143
B.2. Vectores aleatorios.	146
B.3. Procesos estocásticos.	148
Bibliografía	153

Capítulo 1

Introducción

El tema principal de estudio en este trabajo es la interpolación de datos dispersos, que en términos simples, se refiere al problema de ajustar una superficie con ciertas propiedades a través de muestras de datos con distribución no uniforme o aleatoria. Este tema es de importancia práctica en muchos campos de la ciencia e ingeniería donde los datos son con frecuencia medidos o generados en posiciones dispersas e irregulares.

El objetivo de la interpolación es reconstruir, a partir de los datos disponibles, una función que permita propagar la información asociada a estos sobre todas las posiciones del dominio de interés y así conocer los valores en todas las posiciones que se requiera.

Existe una infinidad de fuentes de datos dispersos dentro de las que destacan, valores medidos de cantidades físicas, resultados experimentales y valores computacionales. Estos son encontrados en diversas aplicaciones científicas e ingeniería. Por ejemplo, mediciones no uniformes de cantidades físicas son tomadas en geología, meteorología, oceanografía, cartografía, y minería; datos dispersos experimentales son producidos en química, física, e ingeniería; y valores computacionales espaciados de manera no uniforme surgen en la salida de las soluciones de ecuaciones diferenciales parciales obtenidas mediante elemento finito, y varias aplicaciones en gráficos y visión computacional.

Estos campos requieren la interpolación de datos dispersos para determinar valores en posiciones arbitrarias y no solo en las que la información está disponible. Esto facilita varias operaciones útiles para la visualización de datos dispersos en varias dimensiones. Por ejemplo, en imágenes médicas, la interpolación de datos dispersos es esencial para construir una superficie cerrada de imágenes de órganos humanos o tumores dados por una tomografía axial computarizada (TAC) o una imagen por resonancia magnética (IMR). La visión computarizada utiliza interpolación de datos dispersos para realizar la reconstrucción de la superficie visual sobre mediciones dispersas obtenidas de la características extraídas de las imágenes o el movimiento. En el morphing de imágenes, la interpolación de datos dispersos es útil para derivar un mapeo suave desde la correspondencia de una propiedad de puntos entre un par de imágenes. En aplicaciones geológicas, la función de interpolación derivada facilita la graficación de un mapa de contornos, usados para la creación de mapas topográficos (figura 1.1a) o la creación de modelos digitales de elevación (figura 1.1b), que son representaciones tridimensionales de superficies terrestres.

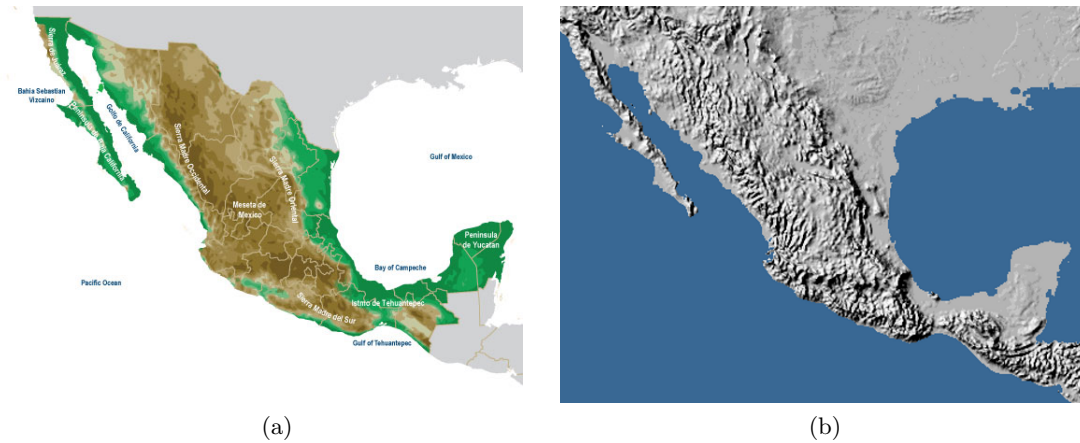


Figura 1.1: (a) Mapa topográfico y (b) modelo digital de elevación de México

Dentro de las aplicaciones geológicas donde estos mapas y modelos topográficos generados mediante interpolación son usados, podemos mencionar por ejemplo:

- Modelado de flujo de agua o movimientos de masas (por ejemplo, avalanchas y desplazamientos de tierra).
- Creación de modelos de fenómenos físicos.
- Análisis del terreno en la geomorfología y la geografía física.
- Sistemas de Información Geográfica (SIG).
- Ingeniería y diseño de infraestructura.
- Sistemas de Posicionamiento Global (GPS).

Algunas de estas aplicaciones son lo bastante generales como para contener una gran cantidad de problemas donde el uso de la interpolación de datos dispersos es de gran utilidad, es por ello que el trabajo se centra mayormente en la descripción de métodos y técnicas que son adecuados en el campo de las aplicaciones geológicas, aunque no están restringidas solo a este. La siguiente sección extiende una de las aplicaciones.

1.1. Creación de modelos de fenómenos físicos

En diversas aplicaciones y estudios de fenómenos físicos es necesario la modelación de un objeto geométrico 3D que represente adecuadamente la región de interés, por ejemplo:

- En la estimación de reservas es necesario la modelación de la estructura geomecánica donde se encuentran estas para poder realizar estudios de rentabilidad que permitan determinar si es factible o no la explotación de dichas reservas.
- Otro ejemplo es la explotación de acuíferos. Hoy en día el uso del agua es uno de los temas más relevantes, por lo que es importante modelar estos acuíferos tanto para estimar las reservas de agua que hay en el subsuelo así como para generar modelos que permitan la explotación racional y cuidado de este recurso.
- La explotación de hidrocarburos es importante ya que todavía de estos se obtiene una gran cantidad de los energéticos y materiales que se utilizan día con día.

En este último se centrará para mostrar el uso de la interpolación de datos dispersos para la creación de modelos físicos.

1.1.1. Modelación de un yacimiento de hidrocarburos

Un yacimiento de hidrocarburos es una acumulación natural de estos en el subsuelo, contenidos en rocas porosas o fracturadas. Es muy complicado y costoso obtener información directa de los yacimientos, por lo cual solo unos pocos pozos de exploración son construidos para mediciones de datos relevantes de los mismos. Para complementar esta información se utilizan técnicas indirectas, como los estudios de sísmica, para dar un panorama general de la estructura del yacimiento y así tomar decisiones sobre la forma en como se debe explotar.

Una herramienta que se ha vuelto indispensable en la industria petrolera sobre la producción y recuperación de hidrocarburos a lo largo de la vida productiva de un yacimiento, es la simulación numérica de flujos en yacimientos, la cual utiliza modelos matemáticos, técnicas numéricas y computacionales, que ayudan a simular varios escenarios y permiten así valorar un gran número de posibilidades de explotación.

Los modelos matemáticos que se utilizan en la simulación numérica de yacimientos, consisten en un conjunto de ecuaciones diferenciales parciales que describen los procesos de transferencia de masa dentro de un medio poroso, y que generalmente son resueltas mediante técnicas numéricas, por lo que es necesario generar un modelo geométrico discreto adecuado del yacimiento.

Esta discretización del yacimiento tiene que ser construida con la poca información precisa con que se cuenta, pero esta puede ser insuficiente, por lo cual la interpolación de datos dispersos toma un papel importante en este ámbito, ya que sirve tanto para construir el modelo geométrico 3D así como para la asignación de propiedades petrofísicas necesarias para la resolución de las ecuaciones diferenciales parciales que describen el flujo de los hidrocarburos dentro del yacimiento.

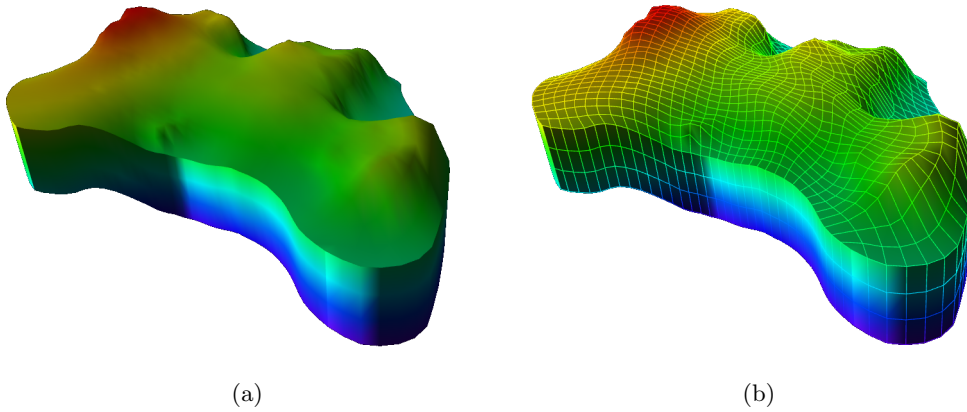


Figura 1.2: Discretización del yacimiento

Existen varias formas de hacer la discretización (figura 1.2), pero una de ellas y en la cual se enfatizará posteriormente, es la reconstrucción de superficies mediante técnicas de interpolación dispersa y la utilización de estas superficies para la construcción de mallas 3D mediante copias verticales que sirva como modelo geométrico 3D.

1.2. Interpolación y aproximación de datos dispersos

El problema principal que se quiere abordar es el de la interpolación de datos dispersos en dos dimensiones que puede ser formulado como sigue: Sea $\mathcal{P} = \{(x_i, y_i, z_i)\} \subset \mathbb{R}^3$ un conjunto de puntos dispersos en el espacio, donde $\mathbf{x}_i = (x_i, y_i)^\top \in \mathcal{D}$ para $\mathcal{D} \subset \mathbb{R}^2$ un dominio en el plano e $i = 1, \dots, N$. Resolver el problema consiste en encontrar una función escalar bivariada F , tal que $F(\mathbf{x}_i) = F(x_i, y_i) = z_i$ para todo i . En algunas ocasiones el problema de interpolación se puede reformular como el problema de reconstrucción de superficies, esto es, se supone que los valores de los datos z_i son obtenidos a partir de una función escalar desconocida $f : \mathcal{D} \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ al ser evaluada en un conjunto de posiciones arbitrarias $\mathbf{x}_i \in \mathcal{D}$, es decir, $f(\mathbf{x}_i) = z_i$, entonces se busca una función bivariada F , tal que $F(\mathbf{x}_i) = F(x_i, y_i) = z_i$ para todo i y F sea una aproximación a f .

Algunas veces se puede tener que los valores de los datos z_i comprendan cierto error, sobre todo cuando estos provienen de mediciones de alguna magnitud física, entonces en lugar de considerar el problema de interpolación se plantea el problema de aproximación, al tomar en cuenta que hay cierto error en los valores de los datos z_i . La única diferencia entre estos dos es que para el problema de aproximación la condición de interpolación se cambia por $F(\mathbf{x}_i) \approx z_i$. Debido a su importancia práctica este problema también es considerado.

Existen varios métodos para resolver este problema, aunque en el presente trabajo solo

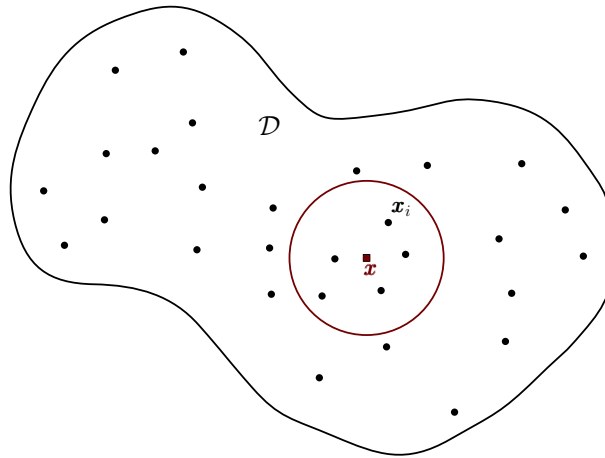


Figura 1.3: Influencia de los valores en las posiciones x_i sobre el nuevo valor en x

se describen algunos de ellos en los que se está interesado por su uso práctico. Un estudio general de varios de ellos es dado por Franke [7]. Todos los métodos están basados en la idea de que el valor desconocido en la posición x debe ser muy parecido a los valores de los datos z_i en las posiciones x_i que sean más cercanas a x . Con esto, se puede decir que solo los valores de los datos dentro de cierta vecindad o radio influyen en la determinación del valor en la posición x donde este se desconoce (véase figura 1.3). Los métodos comprendidos en el trabajo son:

- Métodos de distancia inversa (Shepard y Shepard modificado).
- Métodos de función de base radial.
- Métodos geoestadísticos.
- Método de aproximación B-spline multinivel.

Los métodos provienen de diferentes enfoques de la resolución del problema de interpolación, por lo cual es difícil clasificarlos, pero una forma general de hacerlo es dependiendo si utilizan todos los puntos de los datos o no. Cuando se utilizan todos los puntos se dice que el método es *global* de lo contrario se dice que es *local*. En general los métodos aquí expuestos son descritos en el caso bidimensional, aunque todos estos pueden ser extendidos al caso multidimensional.

El enfoque de este trabajo es la aplicación de estos métodos a algunos problemas prácticos, como por ejemplo los vistos en la sección previa, por lo que en los siguientes capítulos se extenderá la teoría lo necesario para cubrir las necesidades prácticas que se buscan para pasar después a su implementación y aplicación.

Capítulo 2

Métodos de distancia inversa

Los primeros métodos que se describirán son los de distancia inversa. Estos utilizan un promedio ponderado de los valores de los datos o de aproximaciones locales en las posiciones conocidas \mathbf{x}_i para determinar el valor en alguna posición desconocida \mathbf{x} de interés. Su nombre fue motivado por los pesos empleados en el promedio, debido a que son función de la inversa de la distancia en cada posición conocida. De esta manera, se crea una región de influencia (véase figura 1.3) donde solo los valores de los datos más cercanos a \mathbf{x} participan en el promedio, ya que sólo estos tienen un peso significativo que influye en el valor que se obtiene para \mathbf{x} .

La idea original del método de distancia inversa es debido a Shepard [18]. Todos los métodos de este tipo pueden ser vistos como generalizaciones del método de Shepard o modificaciones de estas generalizaciones (Franke [7], Spath [19]).

2.1. Método de Shepard

Dentro de los métodos de distancia inversa el más simple es el método de Shepard (Shepard [18]), el cual es un método global y está basado en las siguientes consideraciones (Spath [19]).

Para pesos dados $w_i(\mathbf{x}) > 0$ se busca el valor de S que mejor aproxime los valores de los datos z_i , $i = 1, \dots, N$, en el sentido de mínimos cuadrados ponderados, es decir, se quiere resolver el problema de minimización

$$\min_S \sum_{i=1}^N w_i(\mathbf{x})(z_i - S)^2. \quad (2.1)$$

Para encontrar el valor de S que minimice (2.1) se deriva esta con respecto a S , esto es

$$\frac{d}{dS} \left(\sum_{i=1}^N w_i(\mathbf{x})(z_i - S)^2 \right) = \sum_{i=1}^N \frac{d}{dS} (w_i(\mathbf{x})(z_i - S)^2) = -2 \sum_{i=1}^N w_i(\mathbf{x})(z_i - S),$$

e igualando a cero esta última expresión

$$\begin{aligned} \sum_{i=1}^N w_i(\mathbf{x})z_i - S \sum_{i=1}^N w_i(\mathbf{x}) &= 0 \\ S \sum_{i=1}^N w_i(\mathbf{x}) &= \sum_{i=1}^N w_i(\mathbf{x})z_i, \end{aligned}$$

se tiene que el valor de S que minimiza la expresión (2.1) esta dado por

$$S = \frac{\sum_{i=1}^N w_i(\mathbf{x})z_i}{\sum_{i=1}^N w_i(\mathbf{x})}.$$

El método de Shepard se obtiene ahora eligiendo los pesos $w_i(\mathbf{x})$ de tal manera que estos varíen como una potencia del inverso de la distancia euclidiana (Shepard [18]), donde la distancia es tomada de una posición arbitraria $\mathbf{x} = (x, y)^\top \in \mathbb{R}^2$ a \mathbf{x}_i , *i.e.*

$$w_i(\mathbf{x}) = \frac{1}{d_i^p} \quad (p > 0),$$

con

$$d_i = d_i(\mathbf{x}) = d(\mathbf{x}, \mathbf{x}_i) = \sqrt{(x - x_i)^2 + (y - y_i)^2}, \quad \text{para } i = 1, \dots, N.$$

Esta elección de los pesos $w_i(\mathbf{x})$ tiene relación con la idea de que el valor del nuevo punto \mathbf{x} debe estar mayormente influenciado por los valores más cercanos a éste que por los que se encuentran más alejados. Así, cuando $d_i \rightarrow \infty$ se tiene que $w_i \rightarrow 0$ para cualquier $p > 0$, consecuentemente, el valor en la posición \mathbf{x}_i no tiene mucha influencia sobre la determinación del valor en \mathbf{x} . El valor de p sirve para determinar que tan rápido o lento se pierde esta influencia además de jugar un papel importante en la función interpolante resultante, pero de ello se hablará con detalle más adelante.

Resumiendo, entre mayor sea la distancia de \mathbf{x} a \mathbf{x}_i , menor será la influencia de \mathbf{x}_i sobre los valores de la función interpolante de Shepard dada por,

$$F_{S_0}(\mathbf{x}) = \sum_{i=1}^N z_i v_i(\mathbf{x}), \quad (2.2)$$

donde

$$v_i(\mathbf{x}) = \frac{w_i(\mathbf{x})}{\sum_{k=1}^N w_k(\mathbf{x})} = \frac{d_i^{-p}}{\sum_{k=1}^N d_k^{-p}}, \quad \mathbf{x} \in \mathbb{R}^2. \quad (2.3)$$

La función v_i dada en (2.3) se puede reescribir como

$$v_i(\mathbf{x}) = \frac{1}{d_i^p} = \frac{\frac{1}{d_i^p} \prod_{j=1}^N d_j^p}{\sum_{k=1}^N \frac{1}{d_k^p}} = \frac{\prod_{\substack{j=1 \\ j \neq i}}^N d_j^p}{\left(\frac{1}{d_1^p} + \frac{1}{d_2^p} + \cdots + \frac{1}{d_N^p} \right) \prod_{j=1}^N d_j^p} = \frac{\prod_{\substack{j=1 \\ j \neq i}}^N d_j^p}{\prod_{\substack{j=1 \\ j \neq 1}}^N d_j^p + \prod_{\substack{j=1 \\ j \neq 2}}^N d_j^p + \cdots + \prod_{\substack{j=1 \\ j \neq N}}^N d_j^p},$$

lo que lleva a una forma numéricamente más estable de v_i dada por

$$v_i(\mathbf{x}) = \frac{\prod_{\substack{j=1 \\ j \neq i}}^N d_j^p}{\sum_{k=1}^N \prod_{\substack{j=1 \\ j \neq k}}^N d_j^p}. \quad (2.4)$$

Observación 2.1. Al decir que v_i en (2.4) es numéricamente más estable se refiere a que de esta forma el denominador no se hace cero, como se argumenta a continuación. Véase que v_i sólo depende de d_j para todo $j = 1, \dots, N$ y las distancias $\{d_j\}$ son cero sólo en \mathbf{x}_i si $i = j$, dado que las posiciones $\{\mathbf{x}_i\}$ se suponen distintas. Así, al evaluar v_i en \mathbf{x}_ℓ para $1 \leq \ell \leq N$, se tiene que el único producto distinto de cero es $\prod_{j \neq \ell} d_j^p$ ya que es el único que no contiene a $d_\ell^p = 0$ y por tanto el denominador en v_i es diferente de cero. De esta manera v_i está bien definida para toda $\mathbf{x} \in \mathbb{R}^2$.

Ahora se verificará que $F_{S_0}(\mathbf{x})$ dado en (2.2) es la función interpolante que se busca. Para comprobar esto, se verá primero las propiedades más importantes de las funciones de peso v_i , $i = 1, \dots, N$, que se resumen a continuación.

Resultado 2.1. Sean las posiciones denotados por $\mathbf{x}_j \in \mathbb{R}^2$ para $j = 1, \dots, N$, distintas entre sí y $v_i(\mathbf{x})$ definida por (2.4) para $i = 1, \dots, N$, entonces se cumplen las siguientes propiedades:

a) Las funciones de peso $v_i(\mathbf{x})$ son continuas.

$$b) v_i(\mathbf{x}_j) = \delta_{ij} = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases}.$$

c) $v_i(\mathbf{x}) \geq 0$.

$$d) \sum_{i=1}^N v_i(\mathbf{x}) = 1.$$

Demostración. De la ecuación (2.4), se puede ver que v_i es una función continua ya que d_i es continua y el denominador siempre es distinto de cero (véase la observación 2.1).

Para probar la propiedad del inciso b) considérese primero el caso $i = j$. Entonces $d_i = 0$ y por tanto $\prod_{\ell \neq i} d_\ell^p \neq 0$, también siendo el único producto diferente de cero en la suma de los productos del denominador ya que este es el único producto donde no se encuentra d_i , por lo tanto

$$v_i(\mathbf{x}_j) = \frac{\prod_{\substack{\ell=1 \\ \ell \neq i}}^N d_\ell^p}{\sum_{k=1}^N \prod_{\substack{\ell=1 \\ \ell \neq k}}^N d_\ell^p} = \frac{\prod_{\substack{\ell=1 \\ \ell \neq i}}^N d_\ell^p}{\prod_{\substack{\ell=1 \\ \ell \neq i}}^N d_\ell^p} = 1.$$

Ahora, cuando $i \neq j$, se tiene que $d_i \neq 0$ y $d_j = 0$, consecuentemente $\prod_{\ell \neq i} d_\ell^p = 0$, y como la suma de los productos en el denominador es distinta de cero ya que $\prod_{\ell \neq j} d_\ell^p \neq 0$, se cumple que

$$v_i(\mathbf{x}_j) = \frac{\prod_{\substack{\ell=1 \\ \ell \neq i}}^N d_\ell^p}{\sum_{k=1}^N \prod_{\substack{\ell=1 \\ \ell \neq k}}^N d_\ell^p} = \frac{0}{\prod_{\substack{\ell=1 \\ \ell \neq j}}^N d_\ell^p} = 0.$$

La propiedad del inciso c) se cumple ya que, como $d_j \geq 0$ para todo j , entonces $\prod_{j \neq i} d_j^p \geq 0$ y $\sum_k \prod_{j \neq k} d_j^p > 0$, por tanto $v_i(\mathbf{x}) \geq 0$.

Por último, la propiedad del inciso d) se cumple porque

$$\sum_{i=1}^N v_i(\mathbf{x}) = \sum_{i=1}^N \frac{\prod_{\substack{j=1 \\ j \neq i}}^N d_j^p}{\sum_{k=1}^N \prod_{\substack{j=1 \\ j \neq k}}^N d_j^p} = \frac{\sum_{i=1}^N \prod_{\substack{j=1 \\ j \neq i}}^N d_j^p}{\sum_{k=1}^N \prod_{\substack{j=1 \\ j \neq k}}^N d_j^p} = 1.$$

□

Resultado 2.2. La función $F_{S_0}(\mathbf{x})$ definida en la ecuación (2.2) satisface las siguientes propiedades:

a) Es continua y es una función interpolante, es decir, cumple con la condición de interpolación $F(\mathbf{x}_i) = z_i$ para toda i .

b) *Es acotada.*

Demostración. De la continuidad de las funciones de peso $v_i(\mathbf{x})$ (inciso a) resultado 2.1) y de la definición de $F_{S_0}(\mathbf{x})$ dada en (2.2) se tiene que F_{S_0} es continua. Además, de (2.2) y la propiedad b) en el resultado 2.1, se cumple que $F_{S_0}(\mathbf{x})$ es una función interpolante, esto es

$$F_{S_0}(\mathbf{x}_i) = \sum_{k=1}^N z_k v_k(\mathbf{x}_i) = \sum_{k=1}^N z_k \delta_{ki} = z_i, \quad i = 1, \dots, N.$$

Ahora se probará la propiedad del inciso b). Defínase

$$\alpha = \min_i z_i \quad \text{y} \quad \beta = \max_i z_i.$$

Se tiene que $\alpha \leq z_i$ para toda i , entonces utilizando las propiedades c) y d) en resultado 2.1 se cumple

$$F_{S_0}(\mathbf{x}) = \sum_{i=1}^N z_i v_i(\mathbf{x}) \geq \sum_{i=1}^N \alpha v_i(\mathbf{x}) = \alpha \sum_{i=1}^N v_i(\mathbf{x}) = \alpha.$$

Por otra parte, $z_i \leq \beta$ para toda i , entonces utilizando nuevamente las propiedades c) y d) en resultado 2.1 se obtiene

$$F_{S_0}(\mathbf{x}) = \sum_{i=1}^N z_i v_i(\mathbf{x}) \leq \sum_{i=1}^N \beta v_i(\mathbf{x}) = \beta \sum_{i=1}^N v_i(\mathbf{x}) = \beta.$$

Esto implica que la función interpolante F_{S_0} está acotada entre el valor mínimo y máximo de los valores $\{z_i\}$, *i.e.*

$$\min_i z_i \leq F_{S_0}(\mathbf{x}) \leq \max_i z_i.$$

□

Ahora se presentan algunos ejemplos para ilustrar los resultados anteriores y algunas ideas adicionales sobre la función interpolante de Shepard F_{S_0} (2.2).

Ejemplo 2.1. Considérese el conjunto de datos $\{\mathbf{x}_i\}$ dentro de la región $\mathcal{D} = [0, 1] \times [0, 1] \subset \mathbb{R}^2$ y sus correspondientes valores $\{z_i\}$, dados en la tabla 2.1, los cuales se utilizarán para analizar el comportamiento de F_{S_0} con diferentes valores de p .

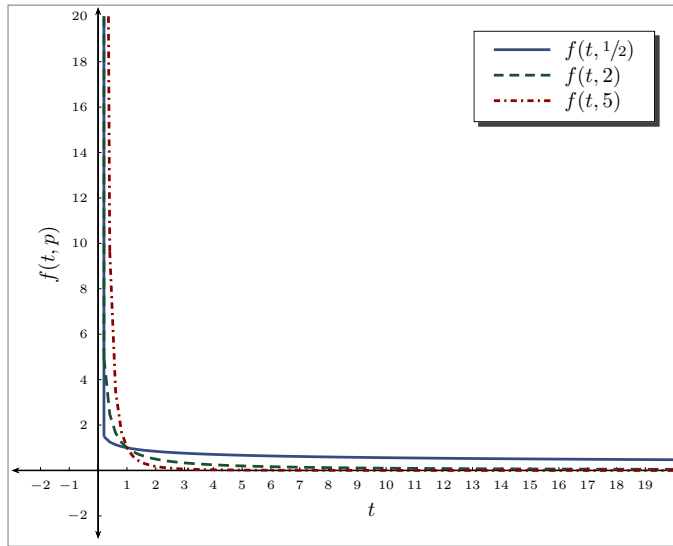
Pero antes, sea $f(t, p) : (0, \infty] \times (0, \infty] \rightarrow (0, \infty]$, definida como

$$f(t, p) = \frac{1}{t^p},$$

y obsérvese que $w_i(\mathbf{x}) = f(t, p)$ si $t = d_i$, de esta forma, estudiando la función $f(t, p)$ se puede examinar el comportamiento de los pesos $w_i(\mathbf{x})$ para diferentes valores de p . En particular se hará para $p = 1/2$, $p = 2$ y $p = 5$ (figura 2.1).

Posiciones	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	\mathbf{x}_4	\mathbf{x}_5
	$(0, 0)^\top$	$(1, 0)^\top$	$(0.5, 0.5)^\top$	$(0, 1)^\top$	$(1, 1)^\top$
Valores	z_1	z_2	z_3	z_4	z_5
	0	0	1	0	0

Tabla 2.1: Datos del ejemplo 2.1

Figura 2.1: Gráfica de la función $f(t, p)$ para algunos valores de p ($1/2, 2, 5$)

Para cualquier valor de p , cuando $t = 1$, $f(t, p) = 1$. Para $t > 1$, entre más grande es el valor de p , $f(t, p) \rightarrow 0$ de forma más rápida, como se muestra, por ejemplo, en la figura 2.1 para $p = 2, 5$. Consecuentemente, los pesos $w_i(\mathbf{x})$ son muy pequeños para puntos \mathbf{x} muy distantes de \mathbf{x}_i , por lo que no tendrán mucha influencia en el valor final de $F_{S_0}(\mathbf{x})$. Ahora, para cuando $t < 1$, sí los valores de p son pequeños, por ejemplo $p = 1/2$ (figura 2.1), sólo los puntos de los datos $\{\mathbf{x}_i\}$ que se encuentran muy cercanos a \mathbf{x} tienen una influencia significativa sobre $F_{S_0}(\mathbf{x})$, ya que solo estos cuentan con pesos considerables.

Así, el efecto que tiene la elección de algún valor de p sobre los pesos y por lo tanto sobre F_{S_0} , se observa en la figura 2.2. Para $p = 1/2$ la superficie obtenida presenta picos en cada uno de los puntos de los datos, dado que, solo los puntos \mathbf{x} muy cercanos a los puntos de los datos $\{\mathbf{x}_i\}$ comparten un valor parecido a ellos, de lo contrario, son un promedio de los valores $\{z_i\}$, como vemos en las figuras 2.2a-2.2b.

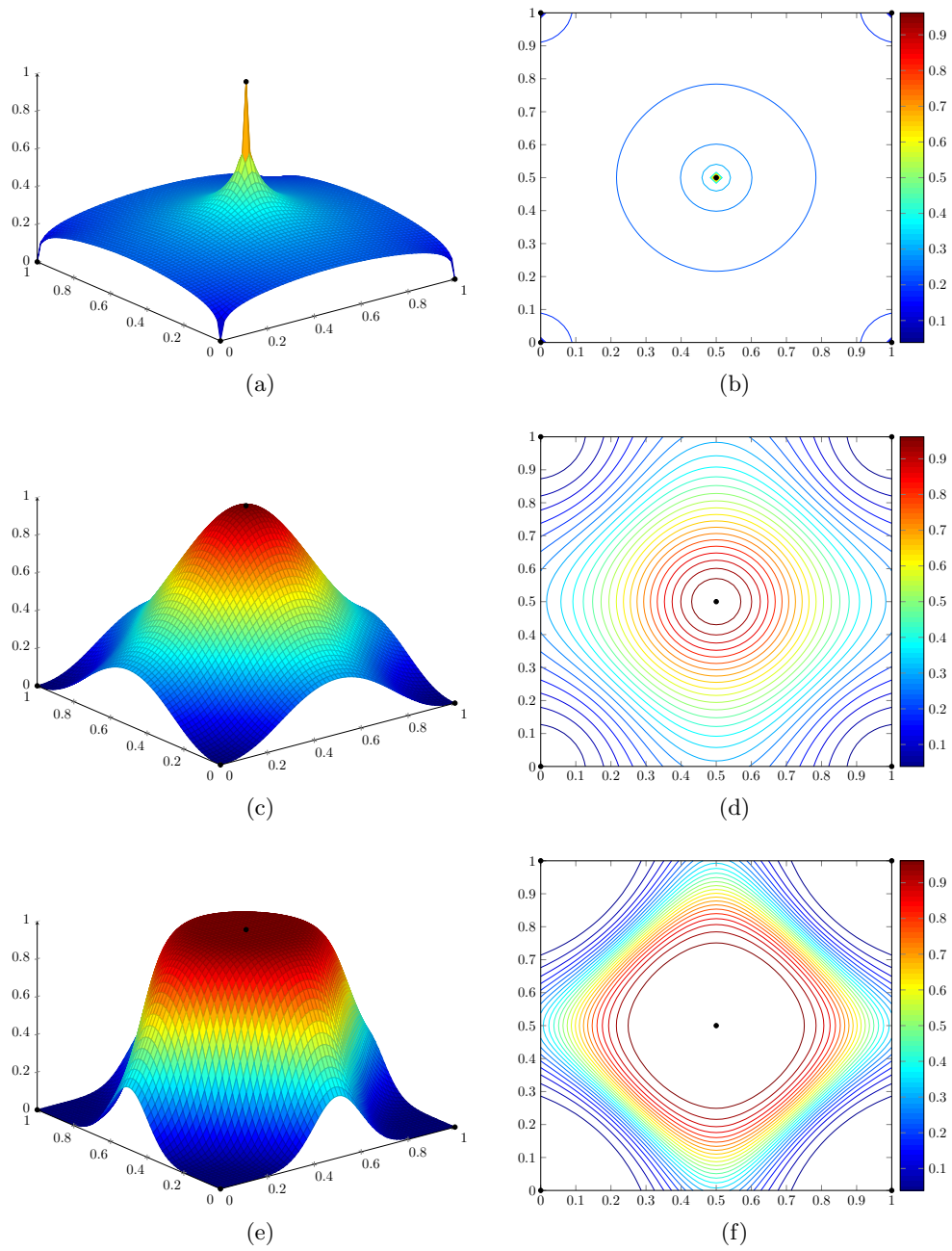


Figura 2.2: Superficies obtenidas mediante el interpolador F_{S_0} (2.2) para los datos de la tabla 2.1. (a) Para $p = 1/2$. (b) Curvas de nivel. (c) Para $p = 2$. (d) Curvas de nivel. (e) Para $p = 5$. (f) Curvas de nivel.

Para $p = 2$, ya no se presentan los picos como en el caso anterior e incluso en una vecindad de los puntos de los datos $\{\mathbf{x}_i\}$ los valores dados por F_{S_0} exhiben un comportamiento isotrópico (figuras 2.2c-2.2d). Para $p = 5$, se aprecia un aplanamiento alrededor de los puntos de los datos (figuras 2.2e-2.2f). Dicho efecto también se observa a partir de los pesos, porque, si un punto \mathbf{x} se encuentra más cercano a un punto de los datos que a otro, este es el que dominará en el valor de $F_{S_0}(\mathbf{x})$. El fenómeno del aplanamiento puede ser analizado calculando las derivadas de F_{S_0} en \mathbf{x}_i , una cuestión que se tratará posteriormente.

De esta forma, entre más grande sea la potencia p , mayor será el aplanamiento en una vecindad de los puntos de los datos $\{\mathbf{x}_i\}$, por esta razón, el valor de $p = 2$ es el más usado en la práctica, además de que simplifica el cálculo de los pesos, al evitar el cómputo de la raíz cuadrada. El aplanamiento, al parecer, una propiedad indeseable en algunas aplicaciones, puede estar bien situado en otras dentro de la demografía, ecología o al análisis de mercados, de acuerdo con Gordon & Wixom [9].

■

Ejemplo 2.2. La tabla 2.2 contiene valores de los datos $\{z_i\}$ y sus posiciones $\{\mathbf{x}_i\}$, para $i = 1, \dots, 5$, con los cuales se mostrará gráficamente el resultado 2.2.

Posiciones	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	\mathbf{x}_4	\mathbf{x}_5
	$(0, 0)^\top$	$(1, 1)^\top$	$(1.2, 0.2)^\top$	$(0, 0.5)^\top$	$(1, 0.5)^\top$
Valores	z_1	z_2	z_3	z_4	z_5
	5.5	1.5	3.5	2.5	2.5

Tabla 2.2: Datos del ejemplo 2.2

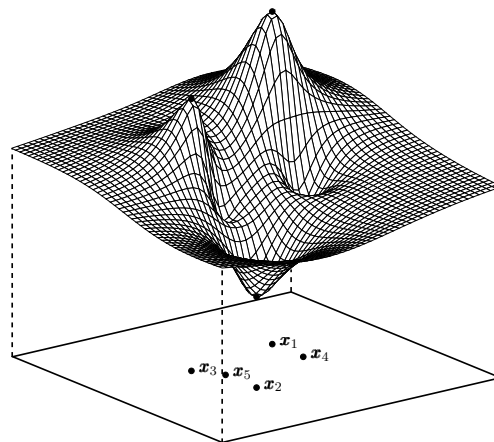


Figura 2.3: Superficie obtenida mediante Shepard para los datos de la tabla 2.2 con $p = 2$

La figura 2.3 muestra la superficie obtenida por la función interpolante de Shepard F_{S_0} para los datos de la tabla 2.2 con $p = 2$. Aquí se aprecia como F_{S_0} queda acotada por los valores máximo y mínimo de los $\{z_i\}$, de acuerdo con el inciso b) del resultado 2.2. Algo que también es importante notar, es que para las posiciones \mathbf{x} lejanas a las posiciones $\{\mathbf{x}_i\}$, el valor de la función F_{S_0} en \mathbf{x} es aproximadamente la media de los $\{z_i\}$. Para ver en detalle esta propiedad, refiérase al Corolario A.3 al final del apéndice A.1. ■

Observación 2.2. De la definición de las funciones $v_i(\mathbf{x})$ en (2.4), se puede ver que estas son diferenciables infinitamente casi donde sea excepto en las posiciones \mathbf{x}_i , donde todas las derivadas parciales hasta de orden $p - 1$ si p es entero y hasta $\lfloor p \rfloor$ si p no es entero, son cero. Para $p = 1$, las derivadas parciales laterales de $v_i(\mathbf{x})$ existen en \mathbf{x}_i , sin embargo, sus valores son casi siempre diferentes. Si $p < 1$ entonces no existe ninguna derivada parcial (Spath [19]). Así, $F_{S_0}(\mathbf{x})$ es diferenciable infinitamente casi donde quiera excepto en las posiciones \mathbf{x}_i , donde esta puede ser solo continua (*cf.* Teorema A.1 del apéndice A.1). Un análisis más detallado de las propiedades y del comportamiento del interpolador $F_{S_0}(\mathbf{x})$ es dado por Gordon & Wixom [9].

En particular, se tiene que,

$$\text{para } p > 1 \quad \Rightarrow \quad \left. \frac{\partial F_{S_0}}{\partial x} \right|_{\mathbf{x}_i} = \left. \frac{\partial F_{S_0}}{\partial y} \right|_{\mathbf{x}_i} = 0,$$

de donde se sigue que la superficie tiene la apariencia de ser plana en una vecindad de \mathbf{x}_i , $i = 1, \dots, N$ (véase la figura 2.2e para un ejemplo de esto). La causa de este fenómeno es que F_{S_0} envuelve solo la magnitud del vector de la diferencia entre \mathbf{x} y \mathbf{x}_i y no su dirección (Shepard [18], Spath [19]).

Con todo lo expuesto anteriormente, el método de Shepard descrito parece suficientemente simple y general para ser bastante atractivo, pero este tiene varios defectos o fallas (Shepard [18]), como ya se ha hecho notar, pero se resumirán a continuación.

- 1) Dado que es un método global, cuando el número de puntos es muy grande, el cálculo de $F_{S_0}(\mathbf{x})$ se convierte proporcionalmente grande. Eventualmente el método se volverá ineficiente o impráctico.
- 2) Solo las distancias de \mathbf{x} a las posiciones \mathbf{x}_i , y no la dirección son considerados.
- 3) Las derivadas direccionales obtenidas en cada punto de los datos \mathbf{x}_i representan una restricción arbitraria e indeseable en la superficie interpolada.
- 4) El error computacional se vuelve significativo en la vecindad de las posiciones \mathbf{x}_i dado que el término predominante resulta de la diferencia de dos números casi iguales.

Para solventar algunos de estos defectos, en [18] se pueden encontrar algunas ideas que el mismo Shepard propuso para cada uno de estos. También se han propuesto diferentes modificaciones del método de Shepard, varias referencias de estos métodos pueden ser encontradas

en los escritos de Franke [7] y Spath [19]. Dentro de estas modificaciones, se pueden encontrar algunas en las que se propone hacer el método de Shepard un método local (como en Franke & Nielson [8] y Renka [15]), cambiando las funciones de peso w_i y tomando algunas funciones conocidas, que sean aproximaciones locales de otras funciones alrededor de las posiciones \mathbf{x}_i , en lugar de tomar solo los valores $\{z_i\}$. Estas ideas llevan a lo que se conoce como el método de Shepard modificado, el cual será descrito en la siguiente sección. Pero, antes de pasar a esto, se verán algunas modificaciones que dan otra perspectiva mejor del comportamiento del interpolador F_{S_0} .

Una modificación sencilla es considerar diferentes potencias $p_i > 0$ ($i = 1, \dots, N$), en lugar de solo una potencia p , así, el interpolante F_{S_0} en (2.2) se puede reescribir como,

$$F_{S_0}^*(\mathbf{x}) = \sum_{i=1}^N z_i v_i^*(\mathbf{x}), \quad \text{con,} \quad v_i^*(\mathbf{x}) = \frac{\prod_{j=1, j \neq i}^N d_j^{p_j}}{\sum_{k=1}^N \prod_{j=1, j \neq k}^N d_j^{p_j}}. \quad (2.5)$$

No hay mucha diferencia entre los interpoladores $F_{S_0}(\mathbf{x})$ y $F_{S_0}^*(\mathbf{x})$, ya que F_{S_0} es un caso particular de $F_{S_0}^*$ al tomar todos los p_i iguales, pero los resultados que se obtienen al utilizar uno u otro son notables, tal y como se ejemplifica a continuación.

Ejemplo 2.3. Se comparará el uso de la función interpolante F_{S_0} y la función $F_{S_0}^*$ para algunos datos. Considérese los valores de los datos $\{z_i\}$ en las posiciones $\{\mathbf{x}_i\}$, $i = 1, \dots, 5$ y las potencias p_i dados en la Tabla 2.3.

Posiciones	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	\mathbf{x}_4	\mathbf{x}_5
	$(0, 0)^\top$	$(0, 1)^\top$	$(0.5, 0.5)^\top$	$(1, 0)^\top$	$(1, 1)^\top$
Valores	z_1	z_2	z_3	z_4	z_5
	1	5.5	4	3	7
Figura 2.4	p_1	p_2	p_3	p_4	p_5
(a)	2	2	2	2	2
(b)	10	1	5	2	3

Tabla 2.3: Datos del ejemplo Ejemplo 2.3

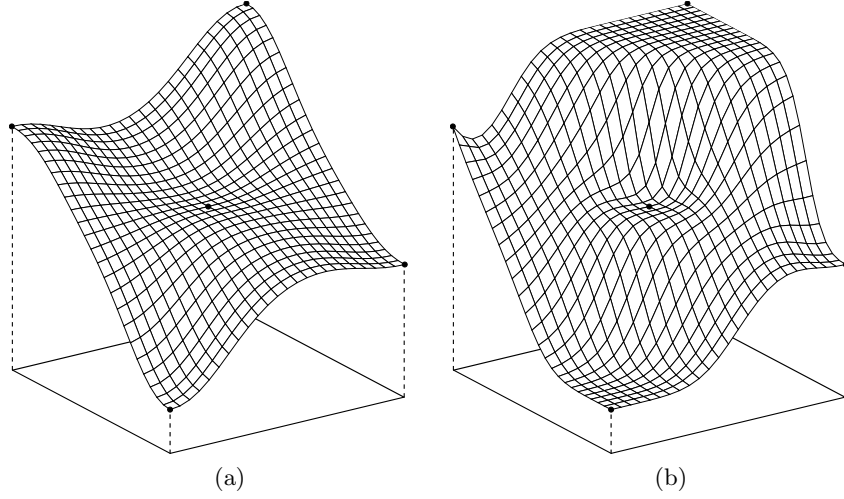


Figura 2.4: Superficies obtenidas de los datos en la tabla 2.3 por la función interpolante de Shepard $F_{S_0}^*$ para distintos valores p_i

La figura 2.4 muestra los resultados obtenidos. Para la figura 2.4a se usó el interpolador F_{S_0} ya que todos los $p_i = 2$, mientras que, para la figura 2.4b se utilizó el interpolador $F_{S_0}^*$ con diferentes potencias dadas en la tabla 2.3. Se puede observar un cambio radical al utilizar $F_{S_0}^*$ en lugar de F_{S_0} , debido a que, $F_{S_0}^*$ permite un control más preciso de su valor sobre los puntos de los datos $\{\mathbf{x}_i\}$, si se tienen en cuenta las propiedades que se analizaron para F_{S_0} en el ejemplo 2.1. ■

Para concluir esta parte, se tratará de forma breve el problema del aplanamiento alrededor de las posiciones \mathbf{x}_i . Este efecto se puede eliminar cuando se tienen valores aproximados para las dos primeras derivadas parciales

$$x'_i = \left. \frac{\partial F_{S_0}}{\partial x} \right|_{\mathbf{x}_i}, \quad y'_i = \left. \frac{\partial F_{S_0}}{\partial y} \right|_{\mathbf{x}_i}, \quad i = 1, \dots, n.$$

Entonces, la función de interpolación,

$$F_{S_1}(\mathbf{x}) = \sum_{i=1}^N [z_i + (x - x_i)x'_i + (y - y_i)y'_i] v_i(\mathbf{x}), \quad \mathbf{x} = (x, y)^T \in \mathbb{R}^2, \quad (2.6)$$

además de tener las mismas propiedades de F_{S_0} (Spath [19]), también toma los valores dados por las primeras derivadas parciales. Un ejemplo más específico del uso de estos interpoladores, se da en la última sección del capítulo, en la cual se compararán los métodos de interpolación de distancia inversa vistos.

2.2. Método de Shepard modificado

Una modificación al método de Shepard descrito en la sección anterior sería utilizar algunas funciones conocidas como aproximaciones locales alrededor de las posiciones $\{\mathbf{x}_i\}$ en lugar de solo los valores $\{z_i\}$ en F_{S_0} (2.2). En particular, Franke & Nielson [8] y Renka [15] utilizaron funciones cuadráticas bivariadas en conjunción con pesos que dan lugar a un método local conocido según Franke [7] como, método de Shepard cuadrático modificado. Aunque posteriormente Renka a utilizado aproximaciones de orden más alto (Renka [17]), en este trabajo se describirá la versión desarrollada por él mismo en [15].

Supóngase que existe una función escalar desconocida f con valores z_i en las posiciones o nodos \mathbf{x}_i para $i = 1, \dots, N$. Se define la función interpolante para el método de Shepard modificado como,

$$F_{SM}(\mathbf{x}) = \sum_{i=1}^N u_i(\mathbf{x}) Q_i(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^2, \quad (2.7)$$

con

$$u_i(\mathbf{x}) = \frac{w_i(\mathbf{x})}{\sum_{k=1}^N w_k(\mathbf{x})}, \quad (2.8)$$

donde w_i son pesos relativos y las funciones Q_i son aproximaciones locales a f en \mathbf{x}_i , conocidas dentro de este contexto como funciones nodales. Específicamente, Q_i es una función cuadrática bivariada definida como,

$$Q_i(\mathbf{x}) = c_{i_1}(x - x_i)^2 + c_{i_2}(x - x_i)(y - y_i) + c_{i_3}(y - y_i)^2 + c_{i_4}(x - x_i) + c_{i_5}(y - y_i) + z_i,$$

que claramente satisface,

$$Q_i(\mathbf{x}_i) = z_i, \quad (2.9)$$

y la cual se ajusta a los valores de f sobre un conjunto de nodos cercanos en un sentido de mínimos cuadrados ponderados. Los pesos relativos w_i son definidos por las funciones de distancia inversa, (propuesta por Franke & Little (véase [7], [8], para referencias)),

$$w_i(\mathbf{x}) = \left[\frac{(R_w - d_i)_+}{R_w d_i} \right]^2$$

donde

$$(R_w - d_i)_+ = \begin{cases} R_w - d_i, & \text{si } d_i < R_w \\ 0, & \text{si } d_i \geq R_w \end{cases}.$$

Aquí, d_i denota nuevamente la distancia euclidiana entre \mathbf{x} y \mathbf{x}_i , y R_w es un radio de influencia alrededor del nodo \mathbf{x}_i , por lo que, los datos en \mathbf{x}_i solo influyen valores interpolados en los puntos dentro de este radio. De esta manera el método se vuelve local al no involucrar todos los nodos \mathbf{x}_i en el cálculo del nuevo valor en la posición \mathbf{x} .

Las funciones u_i además de ser análogas a las funciones de peso v_i (dadas en (2.4) para la función interpolante de Shepard F_{S_0}), cumplen algunas de sus propiedades (resultado 2.1), como se muestra a continuación.

Resultado 2.3. *Sean las posiciones o nodos denotados por $\mathbf{x}_j = (x_j, y_j)^\top \in \mathbb{R}^2$ para $j = 1, \dots, N$, distintos entre sí y las funciones de peso $u_i(\mathbf{x})$ definidas por (2.8), $i = 1, \dots, N$, entonces para todo i , u_i satisface las siguientes propiedades:*

$$a) \quad u_i(\mathbf{x}_j) = \delta_{ij} = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases}.$$

$$b) \quad \sum_{i=1}^N u_i(\mathbf{x}) = 1.$$

$$c) \quad \frac{\partial u_i}{\partial x}(\mathbf{x}_j) = \frac{\partial u_i}{\partial y}(\mathbf{x}_j) = 0, \quad \forall j \in \{1, \dots, N\}.$$

Demostración. Sea $\mathbf{x} = (x, y)^\top \in \mathbb{R}^2$ y $R_w > 0$, entonces, solo para i tal que $R_w - d_i > 0$, se tiene que $w_i > 0$, de esta manera, considérese el conjunto de índices $\mathcal{I} = \{i \in \{1, \dots, N\} \mid R_w - d_i > 0\}$, así, la función de peso u_i puede reescribirse como,

$$u_i(\mathbf{x}) = \frac{w_i(\mathbf{x})}{\sum_{k \in \mathcal{I}} w_k(\mathbf{x})},$$

ya que para $k \notin \mathcal{I}$, $w_k = 0$. Luego, para $k \in \mathcal{I}$ sea

$$\eta_k = \frac{R_w d_k}{R_w - d_k},$$

que cumplen, $w_k = \eta_k^{-2}$, porque $(R_w - d_k)_+ = R_w - d_k > 0$, consecuentemente,

$$u_i(\mathbf{x}) = \frac{w_i(\mathbf{x})}{\sum_{k \in \mathcal{I}} w_k(\mathbf{x})} = \frac{\frac{1}{\eta_i^2}}{\sum_{k \in \mathcal{I}} \frac{1}{\eta_k^2}} = \frac{\prod_{\substack{j \in \mathcal{I} \\ j \neq i}} \eta_j^2}{\sum_{k \in \mathcal{I}} \prod_{\substack{j \in \mathcal{I} \\ j \neq k}} \eta_j^2}, \quad (2.10)$$

la cual es una expresión de u_i análoga a la de v_i dada en (2.4).

Con la ayuda de la expresión (2.10) para u_i , se puede proceder de forma análoga al resultado 2.1 a), para probar la propiedad dada en el inciso a). Primero, sea el caso en que $i = j$. Se tiene que $d_i = 0$ y por tanto $\eta_i = 0$. Como se evalúa u_i en \mathbf{x}_i , se tiene que \mathcal{I} está conformado por todos los índices k tal que $R_w - d_k(\mathbf{x}_i) > 0$, por lo tanto $i \in \mathcal{I}$ dado

que $R_w - d_i(\mathbf{x}_i) = R_w > 0$, de esta manera, el único producto diferente de cero en la suma es $\prod_{\ell \neq i} \eta_\ell^2$, obteniendo así,

$$u_i(\mathbf{x}_j) = \frac{\prod_{\substack{\ell \in \mathcal{I} \\ \ell \neq i}} \eta_\ell^2}{\sum_{k \in \mathcal{I}} \prod_{\substack{\ell \in \mathcal{I} \\ \ell \neq k}} \eta_\ell^2} = \frac{\prod_{\substack{\ell \in \mathcal{I} \\ \ell \neq i}} \eta_\ell^2}{\prod_{\substack{\ell \in \mathcal{I} \\ \ell \neq i}} \eta_\ell^2} = 1.$$

Ahora, cuando $i \neq j$, como se evalúa u_i en \mathbf{x}_j , se tiene que \mathcal{I} esta conformado por todos los índices k tal que $R_w - d_k(\mathbf{x}_j) > 0$. Así $j \in \mathcal{I}$ y $\eta_j = 0$, ya que $d_j = 0$. Luego, si $i \notin \mathcal{I}$, implica que $w_i(\mathbf{x}_j) = 0$ y por tanto $u_i(\mathbf{x}_j) = 0$. Si $i \in \mathcal{I}$ se tiene que el producto $\prod_{\ell \neq i} \eta_\ell^2 = 0$ ya que este incluye $\eta_j = 0$ y $\prod_{\ell \neq j} \eta_\ell^2 \neq 0$ dado que η_j no está en el producto y al menos $\eta_i \neq 0$ se encuentra en este. Así,

$$u_i(\mathbf{x}_j) = \frac{\prod_{\substack{\ell \in \mathcal{I} \\ \ell \neq i}} \eta_\ell^2}{\sum_{k \in \mathcal{I}} \prod_{\substack{\ell \in \mathcal{I} \\ \ell \neq k}} \eta_\ell^2} = \frac{0}{\prod_{\substack{\ell \in \mathcal{I} \\ \ell \neq j}} \eta_\ell^2} = 0.$$

La propiedad del inciso b) se cumple ya que,

$$\sum_{i=1}^N u_i(\mathbf{x}) = \sum_{i \in \mathcal{I}} \frac{\prod_{\substack{j \in \mathcal{I} \\ j \neq i}} \eta_j^2}{\sum_{k \in \mathcal{I}} \prod_{\substack{j \in \mathcal{I} \\ j \neq k}} \eta_j^2} = \frac{\sum_{i \in \mathcal{I}} \prod_{j \in \mathcal{I}} \eta_j^2}{\sum_{k \in \mathcal{I}} \prod_{j \in \mathcal{I}} \eta_j^2} = 1.$$

Por último, para $\mathbf{x} \in \mathbb{R}^2$ si $i \notin \mathcal{I}$, es decir, que $R_w - d_i = 0$, entonces $w_i(\mathbf{x}) = 0$ y por tanto $u_i \equiv 0$, implicando que $\partial u_i / \partial x = \partial u_i / \partial y = 0$. Si $i \in \mathcal{I}$, entonces $w_i(\mathbf{x}) \neq 0$, por lo cual, se puede utilizar la expresión de u_i dada en (2.10). Véase que, para $j \in \mathcal{I}$, se cumple que,

$$\frac{\partial}{\partial x} \eta_j^2 = 2 \frac{R_w^3 (x - x_j)}{(R_w - d_j)^3},$$

se hace cero en \mathbf{x}_j . Luego, para $k \in \mathcal{I}$, sea

$$B_k(\mathbf{x}) = \prod_{\substack{j \in \mathcal{I} \\ j \neq k}} \eta_j^2,$$

la cual cumple, $B_k(\mathbf{x}_j) = 0$ si $k \neq j$ y además se tiene que

$$\frac{\partial B_k}{\partial x}(\mathbf{x}) = \sum_{\substack{m \in \mathcal{I} \\ m \neq k}} \frac{\partial \eta_m^2}{\partial x} \prod_{\substack{\ell \in \mathcal{I} \\ \ell \neq m \\ \ell \neq k}} \eta_\ell^2.$$

Por tanto,

$$\frac{\partial B_k}{\partial x}(\mathbf{x}_j) = 0, \quad \text{si } k \neq j,$$

así, se puede ver que

$$\begin{aligned} \frac{\partial u_i}{\partial x}(\mathbf{x}_j) &= \frac{\left(\sum_{k \in \mathcal{I}} B_k(\mathbf{x}_j) \right) \cdot \frac{\partial B_i}{\partial x}(\mathbf{x}_j) - B_i(\mathbf{x}_j) \cdot \left(\sum_{k \in \mathcal{I}} \frac{\partial B_k}{\partial x}(\mathbf{x}_j) \right)}{\left(\sum_{k \in \mathcal{I}} B_k(\mathbf{x}_j) \right)^2} \\ &= \frac{B_i(\mathbf{x}_i) \cdot \frac{\partial B_i}{\partial x}(\mathbf{x}_i) - B_i(\mathbf{x}_i) \cdot \frac{\partial B_i}{\partial x}(\mathbf{x}_i)}{B_i^2(\mathbf{x}_i)} = 0. \end{aligned}$$

De forma análoga se tiene que $\frac{\partial u_i}{\partial y}(\mathbf{x}_j) = 0$, cumpliéndose así el inciso c). \square

Observación 2.3. De la prueba del inciso c) del resultado 2.3, se observa que la derivada de las funciones u_i existe para todo $\mathbf{x} \in \mathbb{R}^2$ y además se puede mostrar que esta es continua, lo cual implica que $F_{SM} \in C^1(\mathbb{R}^2)$.

Resultado 2.4. Sean $Q_i(\mathbf{x})$ $\mathbf{x} \in \mathbb{R}^2$, funciones nodales que cumplen la ecuación (2.9), i.e., $Q_i(\mathbf{x}_i) = z_i$ para $i = 1, \dots, N$, entonces la función $F_{SM}(\mathbf{x})$ definida en la ecuación (2.7), satisface las siguientes propiedades:

a) Es una función interpolante, es decir, cumple con la condición de interpolación $F_{SM}(\mathbf{x}_i) = z_i$ para toda i .

b)

$$\frac{\partial F_{SM}}{\partial x}(\mathbf{x}_i) = \frac{\partial Q_i}{\partial x}(\mathbf{x}_i) \quad \text{y} \quad \frac{\partial F_{SM}}{\partial y}(\mathbf{x}_i) = \frac{\partial Q_i}{\partial y}(\mathbf{x}_i).$$

Demostración. De (2.9) y de la propiedad dada en el inciso a) del resultado 2.3 se cumple que,

$$F_{SM}(\mathbf{x}_i) = \sum_{k=1}^N u_k(\mathbf{x}_i) Q_k(\mathbf{x}_i) = \sum_{k=1}^N \delta_{ki} Q_k(\mathbf{x}_i) = Q_i(\mathbf{x}_i) = z_i.$$

Por otro lado, de las propiedades dadas en los incisos a) y c) del resultado 2.3, se obtiene que,

$$\frac{\partial F_{SM}}{\partial x}(\mathbf{x}_i) = \sum_{k=1}^N \left[\frac{\partial u_k}{\partial x}(\mathbf{x}_i) Q_k(\mathbf{x}_i) + u_k(\mathbf{x}_i) \frac{\partial Q_k}{\partial x}(\mathbf{x}_i) \right] = \frac{\partial Q_i}{\partial x}(\mathbf{x}_i).$$

De forma similar se tiene que $\frac{\partial F_{SM}}{\partial y}(\mathbf{x}_i) = \frac{\partial Q_i}{\partial y}(\mathbf{x}_i)$. \square

Al principio de la sección se menciono que las funciones nodales Q_i son funciones cuadráticas que cumplen la condición (2.9), *i.e.*, $Q_i(\mathbf{x}_i) = z_i$, además de que se ajustan a los valores de f sobre un conjunto de nodos cercanos en un sentido de mínimos cuadrados ponderados al ser estas una aproximación local de f en una vecindad de \mathbf{x}_i . Es por ello que las funciones nodales $Q_i(\mathbf{x})$, $\mathbf{x} = (x, y)^\top \in \mathbb{R}^2$, quedan dadas como,

$$Q_i(\mathbf{x}) = \bar{c}_{i1}(x - x_i)^2 + \bar{c}_{i2}(x - x_i)(y - y_i) + \bar{c}_{i3}(y - y_i)^2 + \bar{c}_{i4}(x - x_i) + \bar{c}_{i5}(y - y_i) + z_i,$$

donde los coeficientes \bar{c}_{ij} , $j = 1, \dots, 5$, son obtenidos al resolver el problema de minimización,

$$\min_{c_{i1}, c_{i2}, c_{i3}, c_{i4}, c_{i5}} \sum_{\substack{k=1 \\ k \neq i}}^N \hat{w}_k(\mathbf{x}_i) [Q_i(\mathbf{x}_k) - z_k]^2,$$

con

$$\hat{w}_k(\mathbf{x}) = \left[\frac{(R_q - d_k)_+}{R_q d_k} \right]^2,$$

donde R_q es un radio de influencia alrededor del nodo \mathbf{x}_k . Solo los nodos cuyos radios de influencia incluyen a \mathbf{x}_i tienen contribución no cero al ajuste de mínimos cuadrados ponderados, y así Q_i está localmente definida.

Antes de pasar a otro tema, un comentario final es pertinente. Existe una gran variedad de métodos de distancia inversa ya que estos se pueden obtener al considerar una familia de aproximaciones de distancia inversa de mínimos cuadrados ponderados, trabajo que fue realizado por McLain (Franke & Nielson [8]). La forma general de estos interpolantes esta dada por,

$$F(\mathbf{x}) = \sum_{k=1}^M \bar{a}_k(\mathbf{x}) \phi_k(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^2,$$

donde ϕ_k , $k = 1, \dots, M$, denota el conjunto de funciones base¹ y $\bar{a}_k(\mathbf{x})$ los parámetros elegidos que son solución (para un $\mathbf{x} = (x, y)^\top$ dado) del problema de minimización,

$$\min_{a_1, \dots, a_M} \sum_{i=1}^N w_i(\mathbf{x}) (z_i - F(\mathbf{x}))^2. \quad (2.11)$$

En este trabajo solo se ha considerado dos métodos de esta clase, los cuales son los más utilizados y que pueden ser obtenidos a partir de (2.11). El método de Shepard es un caso particular de (2.11) al tomar $F(\mathbf{x}) = S$, con S una constante y los pesos como función de una potencia del inverso de la distancia euclidiana. El método de Shepard modificado fue obtenido a partir de la familia de interpolantes de mínimos cuadrados por Franke & Nielson [8], solo que, a diferencia de los métodos de Shepard original y de McLain, este se obtiene al hacer una aproximación de los parámetros \bar{a}_k en lugar de encontrarlos directamente.

¹La elección de las funciones base ϕ_k consiste por lo general de monomios $x^m y^n$, de grado no muy grande, por lo que $F(\mathbf{x})$ se toma comúnmente como un polinomio de grado 2 ó 3.

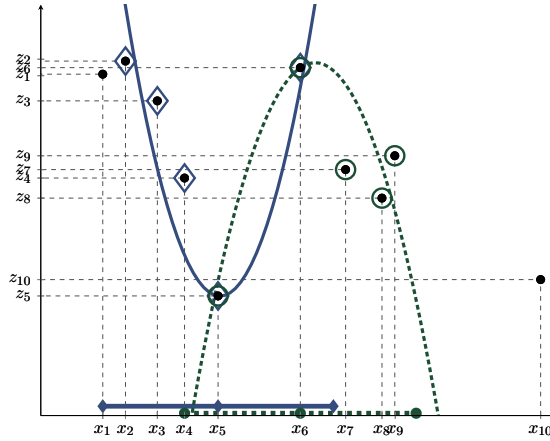


Figura 2.5: Ajuste de funciones cuadráticas alrededor de los nodos x_5 y x_6 utilizando 5 puntos (caso unidimensional). La línea azul $\blacklozenge-\blacklozenge$ denota el radio R_w alrededor del nodo x_5 y los nodos marcados con \blacklozenge son los que participan en el ajuste de la cuadrática azul. La línea verde $\bullet-\bullet$ denota el radio R_w alrededor del nodo x_6 y los nodos marcados con \circ son los que participan en el ajuste de la cuadrática verde.

2.3. Aspectos prácticos

Hasta el momento solamente se ha desarrollado la teoría sobre los métodos, pero ahora, se centrará en lo concerniente a la implementación de los mismos, sobre todo en detallar los aspectos prácticos del método de Shepard modificado.

Los métodos de Franke & Nielson [8] y Renka [15] son similares, la única diferencia consiste en definir los radios de influencia R_w y R_q . El método de Franke & Nielson usa radios fijos uniformes R_w y R_q , elegidos como resultado en el valor específico de N_w y N_q como promedios, donde N_q es el número de nodos \mathbf{x}_i que se incluyen en el ajuste de mínimos cuadrados de las funciones nodales y N_w es el número de funciones nodales que se incluyen en el cálculo del nuevo valor mediante el interpolador de Shepard modificado dado en (2.7). Con el enfoque de Renka se obtiene una mejor precisión al tomar los radios R_w y R_q variables, de manera tal que estos sean lo suficientemente grandes para incluir N_w y N_q nodos, respectivamente, para valores fijos de N_w y N_q . Por ejemplo, de forma más precisa, R_q varía con respecto al i -ésimo nodo, de modo tal, que R_q se toma como la distancia del nodo \mathbf{x}_i al nodo $(j + 1)$ -ésimo más cercano a este para $j + 1 > N_q$, así, solo se utilizan $j = N_q$ nodos, que son los más cercanos a \mathbf{x}_i . R_w se obtiene de manera similar. Los valores propuestos por Renka [15] para la mayoría de los casos son $N_q = 13$ y $N_w = 19$.

La figura 2.5 muestra, para el caso unidimensional, como se obtienen los radios R_q si se considera que $N_q = 4$. Para el caso del ajuste de la función cuadrática alrededor del nodo x_5 , R_q se toma como la distancia de x_5 a x_1 que es el quinto más cercano a este. De forma

análoga, cuando se ajusta una función cuadrática alrededor del nodo x_6 , R_q se toma como la distancia de x_6 a x_4 que es el quinto nodo más cercano a este, pero que ya no se considera dentro del ajuste. De forma similar se encuentra R_w .

Se darán a continuación los algoritmos para los métodos de Shepard y Shepard modificado. Para el método de Shepard (algoritmo 2.1), solo basta evaluar la función F_{S_0} (2.2), aunque hay que tener cierto cuidado en el cálculo de las funciones v_i , esto porque el cálculo de los productos puede causar un desbordamiento aritmético (*overflow*).

Algoritmo 2.1: Método de Shepard

Entrada: datos dispersos $\mathcal{P} = \{(x_i, y_i, z_i)\}$, posiciones de evaluación
 $\mathcal{Q} = \{(\hat{x}_\ell, \hat{y}_\ell)\}_{\ell=1, \dots, M}$, potencia p

Salida: valores en las posiciones de evaluación \hat{z}_ℓ

- 1 **Para** $\ell \leftarrow 1$ **hasta** M **hacer**
 - 2 $\hat{z}_\ell \leftarrow 0$
 - 3 **Para** $i \leftarrow 1$ **hasta** N **hacer** // Cálculo de $F_{S_0}(\hat{\mathbf{x}}_\ell)$ mediante (2.2)
 - 4 calcular $v_i(\hat{\mathbf{x}}_\ell)$ con (2.4) utilizando p dada
 - 5 sumar $z_i \cdot v_i(\hat{\mathbf{x}}_\ell)$ a \hat{z}_ℓ
-

Algoritmo 2.2: Método de Shepard modificado

Entrada: datos dispersos $\mathcal{P} = \{(x_i, y_i, z_i)\}$, posiciones de evaluación
 $\mathcal{Q} = \{(\hat{x}_\ell, \hat{y}_\ell)\}_{\ell=1, \dots, M}$, N_q , N_w

Salida: valores en las posiciones de evaluación \hat{z}_ℓ

- 1 **Para** $i \leftarrow 1$ **hasta** N **hacer** // Resolver el problema de los N_q y N_w vecinos más cercanos
 - 2 Para x_i , utilizar N_q , N_w , para determina R_q y R_w , respectivamente
 - 3 **Para** $i \leftarrow 1$ **hasta** N **hacer**
 - 4 Resolver el problema de mínimos cuadrados

$$\min_{c_{i_1}, c_{i_2}, c_{i_3}, c_{i_4}, c_{i_5}} \sum_{\substack{k=1 \\ k \neq i}}^N \hat{w}_k(\mathbf{x}_i) [Q_i(\mathbf{x}_k) - z_k]^2$$
 para obtener \bar{c}_{i_j} $j = 1, \dots, 5$
 - 5 **Para** $\ell \leftarrow 1$ **hasta** M **hacer**
 - 6 $\hat{z}_\ell \leftarrow 0$
 - 7 **Para** $i \leftarrow 1$ **hasta** N **hacer** // Cálculo de $F_{SM}(\hat{\mathbf{x}}_\ell)$ mediante (2.7)
 - 8 calcular $Q_i(\hat{\mathbf{x}}_\ell)$ con \bar{c}_{i_j} $j = 1, \dots, 5$
 - 9 calcular $u_i(\hat{\mathbf{x}}_\ell)$ con (2.8)
 - 10 sumar $Q_i(\hat{\mathbf{x}}_\ell) \cdot u_i(\hat{\mathbf{x}}_\ell)$ a \hat{z}_ℓ
-

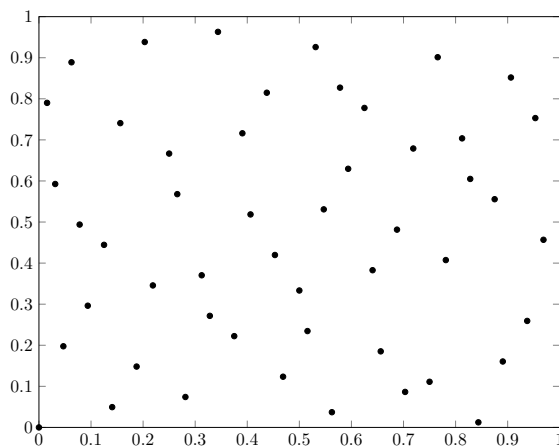
Para el método de Shepard modificado es necesario determinar los radios de influencia R_w y R_q para cada \mathbf{x}_i , por lo que el problema de la búsqueda de los N_w y N_q vecinos más cercanos surge dentro del método.

Definición 2.4. Dado $\mathbf{x} \in \mathcal{D}$, la búsqueda del vecino más cercano consiste en encontrar un punto en \mathcal{P} tal que tenga la mínima distancia a \mathbf{x} dentro de todos los puntos en \mathcal{P} . De forma más general, dado $\mathbf{x} \in \mathcal{D}$ y $\ell \in \mathbb{N}$, la búsqueda de los ℓ vecinos más cercanos, consiste en encontrar ℓ puntos en \mathcal{P} , diferentes entre sí, que tengan la mínima distancia a \mathbf{x} dentro de todos los puntos en \mathcal{P} .

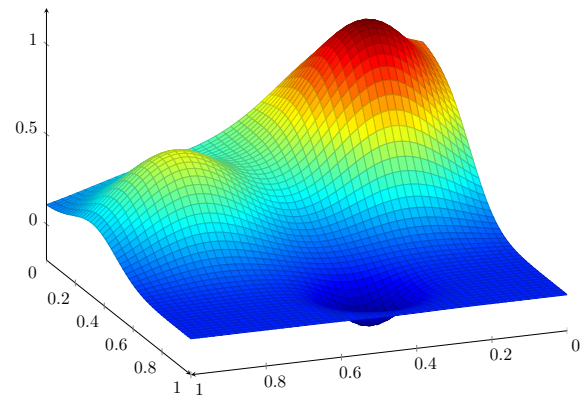
El problema de la búsqueda de los ℓ vecinos más cercanos presenta una gran cantidad de enfoques de solución, es por ello que el tema no será estudiado en el presente trabajo. Algunos de los métodos de solución se encuentran descritos en Wendland [22], aunque el que se utiliza fue dado por Renka en [15] y [16]. El método de Shepard modificado es descrito por el algoritmo 2.2. En el siguiente ejemplo, se estudiará el desempeño de las funciones interpolantes F_{S_0} , F_{S_1} y F_{SM} dados por (2.2), (2.6) y (2.7), respectivamente.

Ejemplo 2.4. Considérese una distribución aleatoria de 50 puntos, para los cuales se evalúa la función de prueba de Franke dada en [6], y la cual tiene la siguiente expresión,

$$f_1(x, y) = 0.75 \exp\left(-\frac{(9x-2)^2 + (9y-2)^2}{4}\right) + 0.75 \exp\left(-\frac{(9x+1)^2}{49} - \frac{9y+1}{10}\right) \\ + 0.5 \exp\left(-\frac{(9x-7)^2 + (9y-3)^2}{4}\right) - 0.2 \exp(-(9x-4)^2 - (9y-7)^2).$$



(a)



(b)

Figura 2.6: (a) Una distribución aleatoria de puntos. (b) Gráfica de la función de prueba f_1

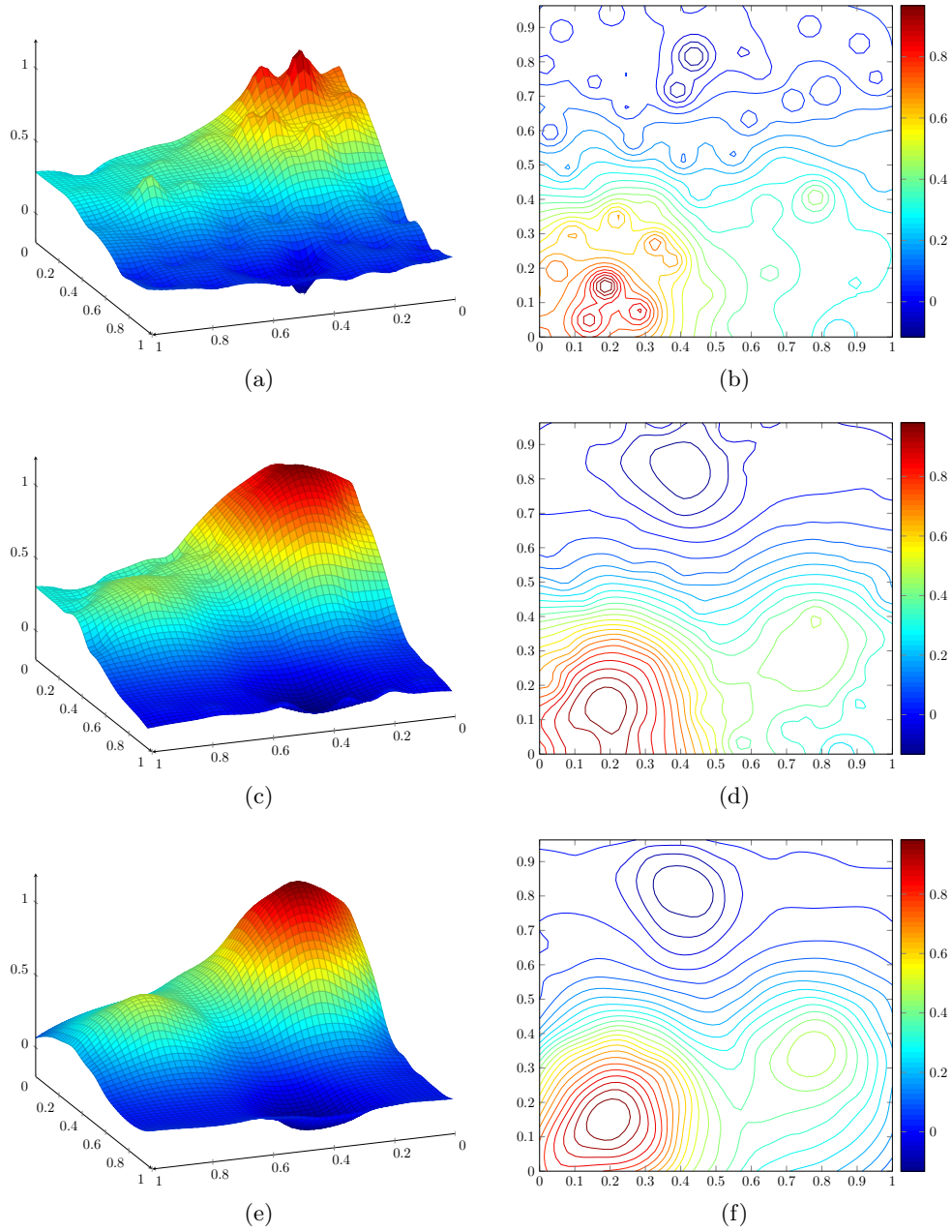


Figura 2.7: (a) Superficie obtenida mediante el interpolador F_{S_0} (2.2) para $p = 2$. (b) Curvas de nivel. (c) Superficie obtenida mediante el interpolador F_{S_1} (2.6) para $p = 2$ y aproximaciones de las derivadas direccionales. (d) Curvas de nivel. (e) Superficie obtenida mediante el interpolador F_{SM} (2.7) con $N_q = 13$ y $N_w = 19$. (f) Curvas de nivel.

En la figura 2.6, se muestra la función original junto con la distribución de las posiciones con las que fueron probadas las funciones interpolantes. En la figura 2.7 se puede ver los resultados obtenidos para los diferentes interpolantes. Se observa (figura 2.7e) que el interpolante F_{SM} dado en (2.1) tiene una mejor aproximación a la función de prueba (figura 2.6b), mientras que la figura 2.7a muestra algunas de las deficiencias del interpolador F_{S_0} dado en (2.2). La figura 2.7c muestra la superficie obtenida mediante el interpolador F_{S_1} dado en (2.6), el cual utiliza aproximaciones de las derivadas direccionales y que muestra una mejora con respecto a F_{S_0} , pero no tiene un desempeño tan bueno como F_{SM} . Tanto para F_{S_0} como para F_{S_1} se utilizó una potencia $p = 2$ y para F_{SM} se utilizó $N_q = 13$ y $N_w = 19$. Más adelante se hará un análisis detallado sobre la comparación de todos los métodos. ■

Capítulo 3

Funciones de base radial

Con el método de funciones de base radial se abordará el problema de aproximar una función escalar bivariada desconocida $f(\mathbf{x}) : \mathbb{R}^2 \rightarrow \mathbb{R}$ por una función conocida $s(\mathbf{x}) : \mathbb{R}^2 \rightarrow \mathbb{R}$, tal que para un conjunto arbitrario de posiciones $\{\mathbf{x}_i\}$, se cumpla que $f(\mathbf{x}_i) = s(\mathbf{x}_i) = z_i$, $i = 1, \dots, N$.

El método descrito en el presente capítulo, da múltiples opciones de s , debido a que se obtienen como una combinación lineal de una base de funciones conocidas como funciones de base radial. Las funciones s así obtenidas pueden ser consideradas como extensiones de los splines de una variable a varias variables. Varios de los resultados que se presentan en esta parte, están formulados para el caso multivariado, aunque como ya hemos mencionado, nuestro interés principal es la aplicación de estos en el caso de dos dimensiones.

Antes de definir el problema de interpolación mediante el uso de funciones de base radial, veremos la relación que tienen con los splines al estudiar el problema de interpolación lineal en una dimensión.

3.1. Interpolación unidimensional

3.1.1. Interpolación lineal

Sean los valores $\{f(x_i)\}$ dados, donde $i = 1, \dots, N$, $f : \mathbb{R} \rightarrow \mathbb{R}$ y la sucesión de puntos $\{x_i\}$ es estrictamente creciente. El spline lineal usual que interpola los datos está compuesto de los segmentos de recta

$$s(x) = \frac{(x_{k+1} - x)f(x_k) + (x - x_k)f(x_{k+1})}{x_{k+1} - x_k} \quad x_k \leq x \leq x_{k+1} \quad (3.1)$$

para $k \in \{1, \dots, N-1\}$. El siguiente resultado muestra una forma análoga de definir el spline lineal dado en (3.1), que interpola el conjunto de datos $\{(x_i, f(x_i))\}$.

Resultado 3.1. Sea el conjunto de datos $\{(x_i, f(x_i))\}$, $i = 1, \dots, N$ como se ha definido antes, y sea,

$$\widehat{s}(x) = \sum_{i=1}^N \lambda_i |x - x_i|, \quad x_1 \leq x \leq x_N, \quad (3.2)$$

donde los coeficientes $\{\lambda_i\}$ son definidos por las condiciones de interpolación,

$$\widehat{s}(x_i) = f(x_i). \quad (3.3)$$

Entonces $s(x) \equiv \widehat{s}(x)$ para $x_1 \leq x \leq x_N$, donde $s(x)$ es el spline lineal dado en (3.1).

Demostración. Sea $x \in [x_k, x_{k+1}]$ para $k = 1, \dots, N - 1$, luego, utilizando la definición de valor absoluto, podemos reescribir (3.2) como,

$$\begin{aligned} \widehat{s}(x) &= \sum_{i=1}^N \lambda_i |x - x_i| = \sum_{i=1}^k \lambda_i (x - x_i) + \sum_{i=k+1}^N \lambda_i (x_i - x) \\ &= x \left[\sum_{i=1}^k \lambda_i - \sum_{i=k+1}^N \lambda_i \right] - \sum_{i=1}^k \lambda_i x_i + \sum_{i=k+1}^N \lambda_i x_i. \end{aligned} \quad (3.4)$$

Por otra parte, para los extremos del intervalo, x_k y x_{k+1} , dado que \widehat{s} cumple las condiciones de interpolación (3.3), utilizando (3.4) se obtiene que,

$$\begin{aligned} f(x_{k+1}) &= x_{k+1} \left[\sum_{i=1}^k \lambda_i - \sum_{i=k+1}^N \lambda_i \right] - \sum_{i=1}^k \lambda_i x_i + \sum_{i=k+1}^N \lambda_i x_i, \\ f(x_k) &= x_k \left[\sum_{i=1}^k \lambda_i - \sum_{i=k+1}^N \lambda_i \right] - \sum_{i=1}^k \lambda_i x_i + \sum_{i=k+1}^N \lambda_i x_i, \end{aligned}$$

así, sustrayendo la segunda ecuación de la primera, tenemos,

$$\begin{aligned} f(x_{k+1}) - f(x_k) &= (x_{k+1} - x_k) \left[\sum_{i=1}^k \lambda_i - \sum_{i=k+1}^N \lambda_i \right], \\ \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} &= \sum_{i=1}^k \lambda_i - \sum_{i=k+1}^N \lambda_i. \end{aligned}$$

Además, utilizando esta última ecuación, se tiene también que,

$$f(x_k) - x_k \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} = - \sum_{i=1}^k \lambda_i x_i + \sum_{i=k+1}^N \lambda_i x_i.$$

De esta manera, se puede reescribir la ecuación (3.4) como,

$$\begin{aligned} \widehat{s}(x) &= x \left[\sum_{i=1}^k \lambda_i - \sum_{i=k+1}^N \lambda_i \right] - \sum_{i=1}^k \lambda_i x_i + \sum_{i=k+1}^N \lambda_i x_i \\ &= \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} (x - x_k) + f(x_k) \\ &= \frac{(x_{k+1} - x)f(x_k) + (x - x_k)f(x_{k+1})}{x_{k+1} - x_k} = s(x). \end{aligned}$$

Como esto se cumple para todos los intervalos donde están definidas las funciones, entonces $\widehat{s}(x) \equiv s(x)$ para $x_1 \leq x \leq x_N$. \square

3.1.2. Splines cúbicos

Para resolver el mismo problema, es posible utilizar otros polinomios de mayor grado en lugar de lineales en cada uno de los intervalos, es decir, que el spline restringido a cada uno de estos se encuentre en $\Pi_m(\mathbb{R})$, para $m > 1$, donde $\Pi_m(\mathbb{R}^d)$ denota al espacio de polinomios de d variables de grado absoluto a lo más m^1 .

En particular, se puede utilizar el conjunto de los splines cúbicos correspondientes a la sucesión estrictamente creciente de puntos $X = \{x_i \mid i = 1, \dots, N\}$ dado por

$$\mathcal{S}_3(X) = \left\{ s \in C^2[a, b] \mid s|_{[x_j, x_{j+1}]} \in \Pi_3(\mathbb{R}), 0 \leq j \leq N \right\},$$

donde $a = x_0 < x_1$, $b = x_{N+1} > x_N$. Este es el conjunto de todas las funciones dos veces diferenciables que coinciden con polinomios cúbicos sobre los intervalos dados por X . La dimensión del espacio $\mathcal{S}_3(X)$ es $N + 4$, por lo que las condiciones de interpolación (3.3) no son suficientes para garantizar la solución única del problema de interpolación. Existen varias formas de hacer que el problema de interpolación en una dimensión usando splines cúbicos tenga solución única y una de ellas es dada por el concepto de spline cúbico natural.

Defínase el conjunto de los splines cúbicos naturales como,

$$\mathcal{N}_{\mathcal{S}_3}(X) = \left\{ s \in \mathcal{S}_3(X) \mid s|_{[a, x_1]}, s|_{[x_N, b]} \in \Pi_1(\mathbb{R}) \right\},$$

el cual consiste en todos los splines cúbicos que son polinomios lineales en los intervalos exteriores $[a, x_1]$ y $[x_N, b]$. Se puede ver que un spline cúbico s es un spline cúbico natural si y sólo si $s''(x_1) = s^{(3)}(x_1) = 0$ y $s''(x_N) = s^{(3)}(x_N) = 0$. Con estas cuatro condiciones adicionales la dimensión de $\mathcal{N}_{\mathcal{S}_3}(X)$ es N y por lo tanto el problema de interpolación tiene solución única².

¹Con grado absoluto se refiere a que la suma de los exponentes de los monomios sea a lo más m .

²Los detalles de estos resultados pueden ser consultados en Powell [13] o para otras referencias en Wendland [22].

Observación 3.1. Una de las propiedades más importantes de los splines cúbicos naturales es la siguiente. Utilizando un enfoque variacional, el problema de interpolación se puede formular como la búsqueda de una función $g \in C^2$, que cumpla las condiciones de interpolación (3.3), $g(x_i) = f(x_i)$ y que minimice el funcional

$$I(g) = \int_{x_1}^{x_n} [g''(x)]^2 dx.$$

El spline cúbico natural único $s \in \mathcal{N}_{\mathcal{S}_3}(X)$ que cumple las condiciones de interpolación (3.3) es el que tiene norma mínima³, ya que, para cualquier g que cumpla las condiciones antes mencionadas, se tiene,

$$I(s) \leq I(g),$$

donde la igualdad sólo se cumple si $g(x) \equiv s(x)$.

Como se hizo antes para el spline lineal, es posible para el spline cúbico natural, encontrar una forma análoga que interpola el conjunto de datos X , como se mostrará a continuación. De acuerdo con Powell [13], el espacio $\mathcal{S}_3(X)$ tiene la base $(\cdot - x_i)_+^3$, $1 \leq i \leq N$, más la base canónica para $\Pi_3(\mathbb{R})$, de este modo, cada $s \in \mathcal{N}_{\mathcal{S}_3}(X)$ tiene una representación de la forma

$$s(x) = \sum_{i=1}^N \alpha_i (x - x_i)_+^3 + \sum_{j=0}^3 \beta_j x^j, \quad x \in [a, b].$$

Como s es lineal en los intervalos exteriores $[a, x_1]$ y $[x_N, b]$, es posible obtener información sobre los coeficientes $\{\alpha_i\}$ y $\{\beta_j\}$. Primero, si se toma el intervalo $[a, x_1]$, entonces $s(x) = \sum_{j=0}^3 \beta_j x^j$, porque $(x - x_i)_+ = 0$ para toda i , y como s debe ser un polinomio lineal en este intervalo, se tiene que $\beta_2 = \beta_3 = 0$. Así, $s(x)$ se reescribe como,

$$s(x) = \sum_{i=1}^N \alpha_i (x - x_i)_+^3 + \beta_0 + \beta_1 x, \quad x \in [a, b].$$

Para $x \in [x_N, b]$, se tiene que $(x - x_i)_+^3 = (x - x_i)^3$ para todo i , luego, expandiendo los términos al cubo, $s(x)$ en el intervalo $[x_N, b]$ esta dado por,

$$\begin{aligned} s(x) &= \sum_{i=1}^N \alpha_i \left(\sum_{k=0}^3 (-1)^k \binom{3}{k} x^{3-k} x_i^k \right) + \beta_0 + \beta_1 x \\ &= \sum_{k=0}^3 (-1)^k \binom{3}{k} \left(\sum_{i=1}^N \alpha_i x_i^k \right) x^{3-k} + \beta_0 + \beta_1 x, \quad x \in [x_N, b]. \end{aligned}$$

Dado que s es lineal en el intervalo $[x_N, b]$, esto implica que los términos cúbico y cuadrático deben ser cero, así, de la última expresión los coeficientes de s satisfacen (para $k = 0, 1$),

³Véase Powell [13], [14], Wendland [22]

$$\sum_{i=1}^N \alpha_i = \sum_{i=1}^N \alpha_i x_i = 0. \quad (3.5)$$

Con estos resultados, al igual como se procedió con el spline lineal, el siguiente resultado muestra una forma análoga de definir el spline cúbico natural que interpola los datos $\{(x_i, f(x_i))\}$.

Resultado 3.2. *Cada spline cúbico natural s puede ser representado en la forma*

$$s(x) = \sum_{i=1}^N \lambda_i \phi(|x - x_i|) + p(x), \quad a \leq x \leq b, \quad (3.6)$$

donde $\phi(r) = r^3$, $r \geq 0$, $p \in \Pi_1(\mathbb{R})$ y los coeficientes $\{\lambda_i\}$, $i=1, \dots, N$, tienen que satisfacer las condiciones (3.5).

Demostración. Como se ha visto, un spline cúbico natural tiene la representación

$$s(x) = \sum_{i=1}^N \alpha_i (x - x_i)_+^3 + \beta_0 + \beta_1 x, \quad x \in [a, b],$$

donde los coeficientes $\{\alpha_i\}$ cumplen las condiciones (3.5). Luego, utilizando la identidad $x_+^3 = (|x|^3 - x^3)/2$ y las condiciones (3.5), se tiene que,

$$\begin{aligned} s(x) &= \sum_{i=1}^N \frac{\alpha_i}{2} |x - x_i|^3 + \sum_{i=1}^N \frac{\alpha_i}{2} (x - x_i)^3 + \beta_0 + \beta_1 x \\ &= \sum_{i=1}^N \frac{\alpha_i}{2} |x - x_i|^3 + \sum_{k=0}^3 \frac{(-1)^k}{2} \binom{3}{k} \left(\sum_{i=1}^N \alpha_i x_i^k \right) x^{3-k} + \beta_0 + \beta_1 x \\ &= \sum_{i=1}^N \lambda_i |x - x_i|^3 + \tau_0 + \tau_1 x, \end{aligned}$$

con $\lambda_i = \frac{1}{2}\alpha_i$, para $1 \leq i \leq N$, y $\tau_0 = \beta_0 - \frac{1}{2} \sum \alpha_i x_i^3$, $\tau_1 = \beta_1 + \frac{3}{2} \sum \alpha_i x_i^2$. Luego, como los coeficientes $\{\alpha_i\}$ cumplen las condiciones (3.5), también las cumplen los coeficientes $\{\lambda_i\}$, por como están definidos. \square

Hasta el momento, se han dado representaciones análogas de los splines lineales y cúbicos naturales, que, a pesar de ser formas equivalentes a los planteamientos usuales para resolver el problema de interpolación en el caso univariado, estas sugieren diferentes métodos para la aproximación de funciones multivariadas, como veremos en la siguiente sección.

3.2. Interpolación multivariada

Ahora, considérese $\{\mathbf{x}_i\}$, $i = 1, \dots, N$, un conjunto de puntos en \mathbb{R}^d para el cual se quiere encontrar $s : \mathbb{R}^d \rightarrow \mathbb{R}$ que satisfaga las condiciones de interpolación análogas a (3.3),

$$s(\mathbf{x}_i) = f(\mathbf{x}_i), \quad (3.7)$$

para $\{f(\mathbf{x}_i) \mid f : \mathbb{R}^d \rightarrow \mathbb{R}\}$ dadas.

Una expresión análoga al interpolador lineal en una dimensión (3.2) (resultado 3.1) en este contexto, es que s tenga la forma,

$$s(\mathbf{x}) = \sum_{i=1}^N \lambda_i \|\mathbf{x} - \mathbf{x}_i\|, \quad \mathbf{x} \in \mathbb{R}^d, \quad (3.8)$$

donde $\|\cdot\| : \mathbb{R}^d \rightarrow \mathbb{R}$ es usualmente la norma euclidiana y la cual se tomara así durante todo el texto. Al definir s de esta manera, se tiene que s es una combinación lineal de las traslaciones de una función que es simétrica alrededor del origen.

Con esta formulación, el problema de interpolación multivariado, se reduce a encontrar solución al sistema de ecuaciones generado por las condiciones de interpolación (3.7), $A\boldsymbol{\lambda} = \mathbf{b}$, dado por,

$$\begin{pmatrix} \|\mathbf{x}_1 - \mathbf{x}_1\| & \|\mathbf{x}_1 - \mathbf{x}_2\| & \dots & \|\mathbf{x}_1 - \mathbf{x}_N\| \\ \|\mathbf{x}_2 - \mathbf{x}_1\| & \|\mathbf{x}_2 - \mathbf{x}_2\| & \dots & \|\mathbf{x}_2 - \mathbf{x}_N\| \\ \vdots & \vdots & \ddots & \vdots \\ \|\mathbf{x}_N - \mathbf{x}_1\| & \|\mathbf{x}_N - \mathbf{x}_2\| & \dots & \|\mathbf{x}_N - \mathbf{x}_N\| \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}_1) \\ f(\mathbf{x}_2) \\ \vdots \\ f(\mathbf{x}_N) \end{pmatrix},$$

el cual tiene solución, incluso única, si A es no singular. Para este caso A se conoce como matriz de distancias, y se puede mostrar que esta es no singular para cualquier conjunto de puntos distintos por pares en \mathbb{R}^d , por lo que el problema de interpolación multivariada está bien planteado bajo este esquema. No se probará directamente que A definida como la matriz de distancias es no singular⁴, aunque, más adelante se verá que este punto queda englobado en un contexto más general que derivaremos a partir de la generalización del spline cúbico natural al caso de interpolación multivariada.

Una generalización del spline cúbico natural (3.6) dado en el resultado 3.2 para el caso multivariado esta dada por,

$$s(\mathbf{x}) = \sum_{i=1}^N \lambda_i \phi(\|\mathbf{x} - \mathbf{x}_i\|) + p(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d, \quad (3.9)$$

⁴Se puede consultar para mayores referencias sobre este tema el libro de Fasshauer [5].

donde $\phi : [0, \infty) \rightarrow \mathbb{R}$ es una función fija univariada, $p \in \Pi_{m-1}(\mathbb{R}^d)$ es un polinomio en \mathbb{R}^d de grado bajo y los coeficientes $\{\lambda_i\}$ deben cumplir las condiciones análogas a (3.5),

$$\sum_{i=1}^N \lambda_i q(\mathbf{x}_i) = 0 \quad \text{para todo } q \in \Pi_{m-1}(\mathbb{R}^d). \quad (3.10)$$

El interpolante resultante es una combinación lineal de las traslaciones de una función radial $\Phi = \phi(\|\cdot\|)$ que es comparable a un polinomio de grado bajo. La función es llamada radial ya que esta es la composición de una función univariada con la norma euclídeana.

En muchos casos es posible que el término polinomial en (3.9) no sea utilizado, y consecuentemente las condiciones (3.10) tampoco sean necesarias, como es el caso de la matriz de distancias (3.8), para la cual $\Phi = \phi(\|\cdot\|) = \|\cdot\|$. De esta manera el problema de interpolación vuelve a recaer en la no singularidad de la matriz A , que ahora tiene las entradas $A_{ij} = \phi(\|\mathbf{x}_i - \mathbf{x}_j\|)$, entonces la pregunta que surge es:

¿Cómo elegir las funciones ϕ de tal manera que la matriz A sea no singular y así el problema de interpolación tenga solución?

Esta será la pregunta que responderemos en la siguiente sección, pero antes de finalizar esta parte queremos formalizar algunos conceptos que serán útiles posteriormente.

Definición 3.2. Una función $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$ es llamada radial, si existe una función univariada $\phi : [0, \infty) \rightarrow \mathbb{R}$ tal que

$$\Phi(\mathbf{x}) = \phi(\|\mathbf{x}\|),$$

donde $\|\cdot\|$ es la norma euclídeana en \mathbb{R}^d .

Observación 3.3. Para una función radial Φ y $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$, de acuerdo con la definición 3.2, se tiene que si

$$\|\mathbf{x}_1\| = \|\mathbf{x}_2\| \quad \Rightarrow \quad \Phi(\mathbf{x}_1) = \Phi(\mathbf{x}_2).$$

En otras palabras, el valor de Φ en cualquier punto a cierta distancia fija del origen (o cualquier otro punto fijo) es constante. Así, Φ es radialmente (o esféricamente) simétrica con respecto a su centro.

Observación 3.4. La función $s(\mathbf{x})$ dada por la ecuación (3.9) puede ser vista también como una combinación lineal de *funciones base* $\Phi_i(\mathbf{x}) = \phi(\|\mathbf{x} - \mathbf{x}_i\|)$ que son radiales, es en este contexto que surge el nombre de *funciones de base radial*.

Es oportuno hacer un comentario final sobre la relación entre las funciones de base radial y los splines. Se ha mostrado en la sección anterior que para el caso de intepolación en una dimensión la formulación con funciones de base radial y splines coinciden. Para el caso multivariado las funciones de base radial no es una generalización directa de los splines a dimensiones más grandes, pero se toma en cuenta la propiedad de norma mínima que cumple el spline cúbico natural (véase observación 3.1), ya que se busca que las funciones de base radial preserven esta propiedad en ciertos espacios de funciones. No se abundara sobre este tema aquí, pero algunos detalles se muestran en el apéndice A.2.

3.3. Matrices de interpolación no singulares

En la sección anterior se planteo el problema de interpolación de datos dispersos mediante funciones de base radial, cuya solución se reduce a que la matriz A con entradas $A_{ij} = \phi(\|\mathbf{x}_i - \mathbf{x}_j\|)$, $i, j = 1 \dots, N$ sea no singular. Por lo tanto, es necesario buscar para que tipos de funciones ϕ se garantiza que la matriz del sistema A sea no singular sobre cualquier conjunto $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ de puntos distintos por pares.

Un enfoque de solución, es buscar funciones que hagan a A una matriz definida positiva para así garantizar la no singularidad de A . Aquí se mostrarán los resultados teóricos bajo esta manera de abordar el problema, pero, dado que el objetivo del presente trabajo es la aplicación de los métodos, se omitirán varias de las pruebas de los resultados enunciados, debido a que son muy extensos y sería fácil desviarse del objetivo principal de este escrito. Varias de las pruebas y algunas referencias sobre ellas se encuentran en Fasshauer [5], Powell [14] y Wendland [22].

3.3.1. Funciones definidas positivas

Como se trabajara bajo el enfoque de matrices definidas positivas, es conveniente recordar como se definen.

Definición 3.5. Una matriz real simétrica A es llamada semi-definida positiva si su forma cuadrática asociada es no negativa, *i.e.*,

$$\sum_{i=1}^N \sum_{k=1}^N c_i c_k A_{ik} \geq 0 \quad (3.11)$$

para $\mathbf{c} = (c_1, \dots, c_N)^\top \in \mathbb{R}^N$. Si la forma cuadrática (3.11) es cero solo para $\mathbf{c} \equiv \mathbf{0}$, entonces A es llamada definida positiva.

Las matrices definidas positivas tienen la importante propiedad de que todos sus eigenvalores son positivos, por lo cual una matriz definida positiva es no singular. Así, si se tuviera funciones base $\Phi_i(\mathbf{x}) = \phi(\|\mathbf{x} - \mathbf{x}_i\|)$ en la que la función

$$s(\mathbf{x}) = \sum_{i=1}^N \lambda_i \Phi_i(\mathbf{x})$$

sujeta a las condiciones de interpolación $s(\mathbf{x}_i) = f(\mathbf{x}_i)$, genere una matriz del sistema definida positiva, entonces siempre se tendría un problema de interpolación bien planteado. Es por esto que, en analogía a la definición 3.5, se introduce el concepto de función definida positiva.

Definición 3.6. Una función continua de variable compleja $\Phi : \mathbb{R}^d \rightarrow \mathbb{C}$, es llamada definida positiva sobre \mathbb{R}^d si

$$\sum_{i=1}^N \sum_{k=1}^N c_i \bar{c}_k \Phi(\mathbf{x}_i - \mathbf{x}_k) \geq 0 \quad (3.12)$$

para cualquiera N puntos diferentes por pares $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbb{R}^d$, y $\mathbf{c} = (c_1, \dots, c_N)^\top \in \mathbb{C}^N$. La función Φ es llamada **estrictamente definida positiva** sobre \mathbb{R}^d si la forma cuadrática (3.12) es cero solo para $\mathbf{c} \equiv \mathbf{0}$. Aquí \bar{c} denota el conjugado del número complejo c .

Aunque se esta interesado en problemas con datos en los reales y coeficientes reales, una extensión de la notación para cubrir coeficientes complejos y funciones de variable compleja en la definición anterior será útil cuando se obtengan propiedades de funciones (estrictamente) definidas positivas.

Ejemplo 3.1. La función $\Phi(\mathbf{x}) = e^{i\mathbf{x}^\top \mathbf{y}}$ (i denota la unidad imaginaria y $\mathbf{x}^\top \mathbf{y}$ el producto interior estándar), para $\mathbf{y} \in \mathbb{R}^d$ fijo, es definida positiva sobre \mathbb{R}^d ya que la forma cuadrática (3.12) en la definición 3.6 se convierte en,

$$\begin{aligned} \sum_{j=1}^N \sum_{k=1}^N c_j \bar{c}_k \Phi(\mathbf{x}_j - \mathbf{x}_k) &= \sum_{j=1}^N \sum_{k=1}^N c_j \bar{c}_k e^{i(\mathbf{x}_j - \mathbf{x}_k)^\top \mathbf{y}} \\ &= \sum_{j=1}^N c_j e^{i\mathbf{x}_j^\top \mathbf{y}} \sum_{k=1}^N \bar{c}_k e^{-i\mathbf{x}_k^\top \mathbf{y}} \\ &= \left| \sum_{j=1}^N c_j e^{i\mathbf{x}_j^\top \mathbf{y}} \right|^2 \geq 0. \end{aligned}$$

■

Obsérvese que la función Φ no es necesariamente radial, una propiedad se necesitara y que se requerirá más adelante. Pero antes de restringirse a este tipo de funciones, se resumirá en el siguiente resultado algunas de las propiedades básicas sobre las funciones (estrictamente) definidas positivas.

Teorema 3.3. *Algunas propiedades básicas de las funciones definidas positivas son*

a) *Combinaciones lineales finitas no negativas de funciones definidas positivas son definidas positivas. Si Φ_1, \dots, Φ_N son definidas positivas sobre \mathbb{R}^d y $c_i \geq 0$, $i = 1, \dots, N$, entonces*

$$\Phi(\mathbf{x}) = \sum_{i=1}^N c_i \Phi_i(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d,$$

es también definida positiva. Más aún, si al menos una de las Φ_i es estrictamente definida positiva y el correspondiente $c_i > 0$, entonces Φ es estrictamente definida positiva.

b) $\Phi(\mathbf{0}) > 0$.

c) $\Phi(-\mathbf{x}) = \overline{\Phi(\mathbf{x})}$.

d) *Cualquier función definida positiva es acotada. De hecho,*

$$|\Phi(\mathbf{x})| \leq \Phi(\mathbf{0}).$$

e) Si Φ es definida positiva con $\Phi(\mathbf{0}) = 0$ entonces $\Phi \equiv 0$.

f) El producto de funciones (estrictamente) definidas positivas es (estrictamente) definido positivo.

La propiedad c) muestra que cualquier función (estrictamente) definida positiva de variable real, tiene que ser par, sin embargo, es posible también caracterizarlas usando solo coeficientes reales.

Teorema 3.4. *Una función continua de variable real Φ es definida positiva sobre \mathbb{R}^d si y sólo si es par y*

$$\sum_{i=1}^N \sum_{k=1}^N c_i c_k \Phi(\mathbf{x}_i - \mathbf{x}_k) \geq 0 \quad (3.13)$$

para cualquiera N puntos diferentes por pares $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbb{R}^d$, y $\mathbf{c} = (c_1, \dots, c_N)^\top \in \mathbb{R}^N$. La función Φ es estrictamente definida positiva sobre \mathbb{R}^d si la forma cuadrática (3.13) es cero solo para $\mathbf{c} \equiv \mathbf{0}$.

Ahora se presentaran algunos resultados sobre la caracterización integral de funciones definidas positivas. Estos fueron establecidos en los años 30 del siglo pasado por Bochner y Schoenberg. No obstante, se mencionaran las extensiones más recientes para funciones estrictamente definidas positivas que son esenciales para la aplicación de la teoría al problema de interpolación de datos dispersos. Uno de los resultados más importantes sobre funciones definidas positivas es su caracterización en términos de su transformada de Fourier establecida por Bochner.

Teorema 3.5 (Bochner). *Una función (de variable compleja) $\Phi \in C(\mathbb{R}^d)$ es definida positiva sobre \mathbb{R}^d si y sólo si es la transformada de Fourier de una medida de Borel⁵ no negativa finita μ sobre \mathbb{R}^d , i.e.*

$$\Phi(\mathbf{x}) = \widehat{\mu}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d}} \int_{\mathbb{R}^d} e^{-i\mathbf{x}^\top \mathbf{y}} d\mu(\mathbf{y}), \quad \mathbf{x} \in \mathbb{R}^d,$$

Demostración. Se probará solo una dirección del teorema que es importante para la aplicación al problema de interpolación de datos dispersos. Suponemos que Φ es la transformada de Fourier de una medida de Borel no negativa y mostraremos que Φ es definida positiva.

⁵ cfr. apéndice B en Fasshauer [5] para las definiciones de medida de Borel y transformada de Fourier.

Así,

$$\begin{aligned}
\sum_{j=1}^N \sum_{k=1}^N c_j \bar{c}_k \Phi(\mathbf{x}_j - \mathbf{x}_k) &= \frac{1}{\sqrt{(2\pi)^d}} \sum_{j=1}^N \sum_{k=1}^N c_j \bar{c}_k \left[\int_{\mathbb{R}^d} e^{-i(\mathbf{x}_j - \mathbf{x}_k)^\top \mathbf{y}} d\mu(\mathbf{y}) \right] \\
&= \frac{1}{\sqrt{(2\pi)^d}} \int_{\mathbb{R}^d} \left[\sum_{j=1}^N c_j e^{-i\mathbf{x}_j^\top \mathbf{y}} \sum_{k=1}^N \bar{c}_k e^{i\mathbf{x}_k^\top \mathbf{y}} \right] d\mu(\mathbf{y}) \\
&= \frac{1}{\sqrt{(2\pi)^d}} \int_{\mathbb{R}^d} \left| \sum_{j=1}^N c_j e^{-i\mathbf{x}_j^\top \mathbf{y}} \right|^2 d\mu(\mathbf{y}) \geq 0,
\end{aligned}$$

donde la última desigualdad se cumple por las condiciones impuestas sobre la medida μ . \square

Observación 3.7. Del teorema 3.5 se puede ver que la función $\Phi(\mathbf{x}) = e^{i\mathbf{x}^\top \mathbf{y}}$ del ejemplo 3.1 puede ser considerada como la función definida positiva fundamental ya que todas las funciones definidas positivas son obtenidas como combinaciones lineales (infinitas) de esta función. Mientras que la propiedad c) del teorema 3.3 implica que las combinaciones lineales de Φ van a ser nuevamente definidas positivas, la consecuencia más importante del teorema 3.5 es el hecho de que todas las funciones definidas positivas son generadas por Φ .

Con el objetivo de garantizar un problema de interpolación bien planteado se tiene que extender la caracterización de Bochner a funciones estrictamente definidas positivas. Esto servirá como un criterio para verificar cuando una función dada es estrictamente definida positiva.

Teorema 3.6. *Sea Φ una función continua en $L_1(\mathbb{R}^d)$. Φ es estrictamente definida positiva si y sólo si Φ es acotada y su transformada de Fourier es no negativa y no idénticamente igual a cero.*

Ahora se centrará en funciones radiales definidas positivas. Recuérdese que la definición 3.6 caracteriza las funciones (estrictamente) definidas positivas en términos de funciones multivariadas Φ , sin embargo, cuando estamos tratando con funciones radiales, *i.e.*, $\Phi(\mathbf{x}) = \phi(\|\mathbf{x}\|)$, entonces será conveniente referirse también como función radial definida positiva a la función univariada ϕ . Una consecuencia de esta convención en la notación es, el siguiente resultado.

Lema 3.7. *Si $\Phi = \phi(\|\cdot\|)$ es (estrictamente) definida positiva y radial sobre \mathbb{R}^d entonces Φ es incluso (estrictamente) definida positiva y radial sobre \mathbb{R}^σ para cualquier $\sigma \leq d$.*

Ahora se enuncia un resultado análogo al teorema 3.6 que da la caracterización en transformadas de Fourier de funciones radiales (estrictamente) definidas positivas sobre \mathbb{R}^d .

Teorema 3.8. *Una función continua $\phi : [0, \infty) \rightarrow \mathbb{R}$ tal que $r^{d-1}\phi(r) \in L_1([0, \infty))$ es estrictamente definida positiva y radial sobre \mathbb{R}^d si y sólo si la transformada de Fourier d -dimensional*

$$\mathcal{F}_d\phi(r) = \frac{1}{\sqrt{r^{d-2}}} \int_0^\infty \phi(t)t^{d/2} J_{(d-2)/2}(rt) dt$$

es no negativa y no idénticamente igual a cero. $J_{(d-2)/2}$ es la función de Bessel clásica de primer tipo de orden $(d-2)/2$.

Ya que el lema 3.7 establece que cualquier función que es (estrictamente) definida positiva y radial sobre \mathbb{R}^d es incluso (estrictamente) definida positiva y radial sobre \mathbb{R}^σ para cualquier $\sigma \leq d$, aquellas funciones las cuales son (estrictamente) definidas positivas y radiales sobre \mathbb{R}^d para toda d son de particular interés.

Teorema 3.9 (Schoenberg). *Una función continua $\phi : [0, \infty) \rightarrow \mathbb{R}$ es estrictamente definida positiva y radial sobre \mathbb{R}^d para todo d si y sólo si esta es de la forma*

$$\phi = \int_0^\infty e^{-r^2 t^2} d\mu(t),$$

donde μ es una medida de Borel no negativa sobre $[0, \infty)$ no concentrada en el origen.

Observación 3.8. Haciendo μ la medida de evaluación puntual en el teorema 3.9 se obtiene que la función gaussiana es estrictamente definida positiva y radial sobre \mathbb{R}^d para toda d .

A continuación, se darán algunos ejemplos de funciones radiales estrictamente definidas positivas que son obtenidas a partir de la teoría que se ha visto hasta ahora.

Ejemplo 3.2 (Funciones gaussianas). Se puede mostrar que la función gaussiana,

$$\Phi(\mathbf{x}) = e^{-\varepsilon^2 \|\mathbf{x}\|^2}, \quad \varepsilon > 0,$$

es estrictamente definida positiva (y radial) sobre \mathbb{R}^d para cualquier d . Esto es debido a que la transformada de Fourier de una función gaussiana es esencialmente gaussiana, es decir, esta dada por,

$$\widehat{\Phi}(\boldsymbol{\omega}) = \frac{1}{(\sqrt{2\varepsilon})^d} e^{-\frac{\|\boldsymbol{\omega}\|^2}{4\varepsilon^2}},$$

la cual es positiva independientemente de la dimensión del espacio d . Aquí ε se conoce como el parámetro de forma, y sirve para hacer las funciones base más planas, cuando ε tiene valores pequeños, o más localizadas (picudas) con respecto a su centro, en el caso contrario. La figura 3.1 muestra la gráfica de la función gaussiana para diferentes valores del parámetro de forma ε , donde se aprecia el efecto que tiene su valor sobre las funciones base.

De acuerdo con Fasshauer [5], las funciones gaussianas pueden ser vistas como el miembro fundamental de la familia de funciones que son estrictamente definidas positivas y radiales

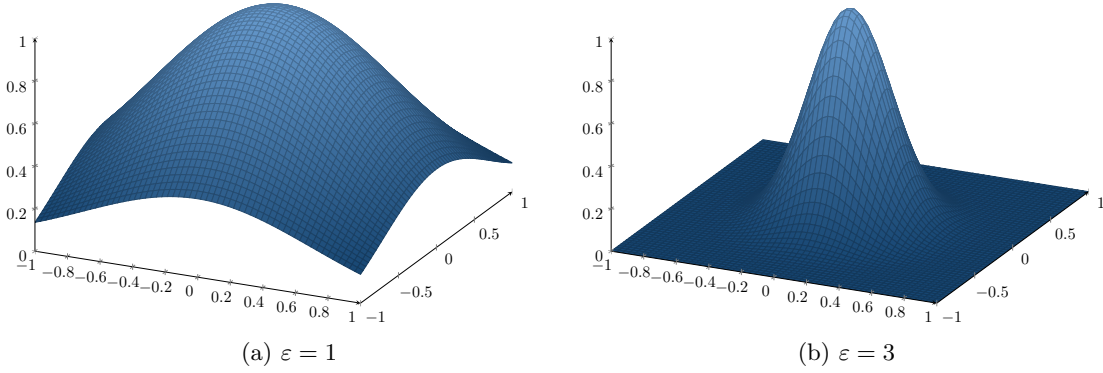


Figura 3.1: Función gaussiana centrada en el origen en \mathbb{R}^2 para diferentes valores del parámetro de forma ε

sobre \mathbb{R}^d para toda d , ya que a partir de estas, de la caracterización de Schoenberg, teorema 3.9 y de las propiedades de las funciones (estrictamente) definidas positivas (cfr. teorema 3.3) se pueden obtener muchas más. ■

Ejemplo 3.3 (Multicuadráticas inversas generalizadas). Las funciones multicuadráticas inversas generalizadas dadas por

$$\Phi(\mathbf{x}) = (1 + \|\mathbf{x}\|^2)^{-\beta}, \quad \beta > \frac{d}{2},$$

son estrictamente definidas positivas sobre \mathbb{R}^d para $d < 2\beta$, dado que su transformada de Fourier,

$$\widehat{\Phi}(\boldsymbol{\omega}) = \frac{K_{\beta-s/2}(\|\boldsymbol{\omega}\|)\|\boldsymbol{\omega}\|^{\beta-s/2}}{2^{\beta-1}\Gamma(\beta)}, \quad \beta > \frac{d}{2},$$

es una función radial positiva. K_ν es la función modificada de Bessel de segundo tipo de orden ν .

Es posible utilizar la teoría de funciones completamente monótonas para mostrar que solo se requiere que $\beta > 0$ (véase Fasshauer [5]), por lo tanto las funciones multicuadráticas inversas generalizadas son estrictamente definidas positivas sobre \mathbb{R}^d para cualquier d .

Para estas funciones, se utiliza también el parámetro de forma ε como se hizo para la función gaussiana al considerar la función múlticuadrática inversa generalizada como

$$\Phi(\mathbf{x}) = (1 + \varepsilon^2\|\mathbf{x}\|^2)^{-\beta}, \quad \varepsilon, \beta > 0,$$

que se evita poner en la definición inicial de estas funciones para obtener una forma sencilla de su transformada de Fourier, pero que se puede ver también que es estrictamente definida

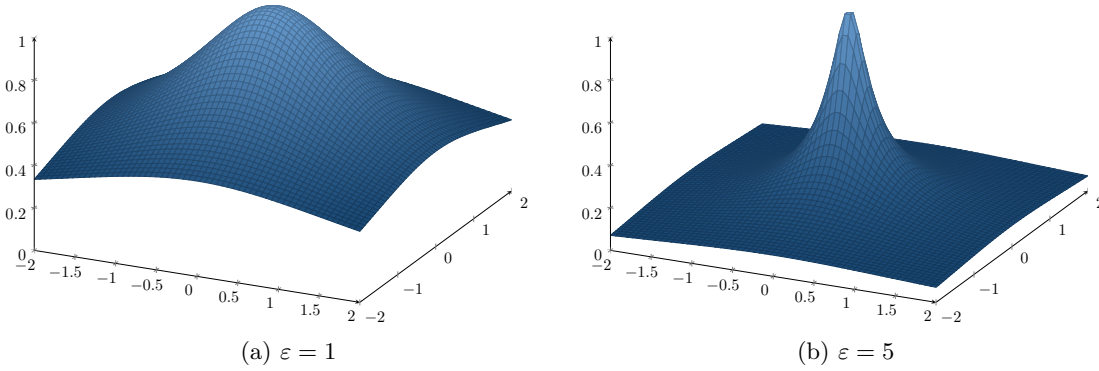


Figura 3.2: Función multicuadrática inversa ($\beta = 1/2$) centrada en el origen en \mathbb{R}^2

positiva.

La figura 3.2 muestra la gráfica de la función multicuadrática inversa para diferentes parámetros de forma ε y con $\beta = 1/2$. Para este valor de β , la función multicuadrática inversa generalizada esta relacionada con la función multicuadrática inversa propuesta por Hardy, dada por $\Phi(\mathbf{x}) = (c^2 + \varepsilon^2 \|\mathbf{x}\|^2)^{-1/2}$, al tomar $c^2 = 1/\varepsilon^2$ y escalar esta por $1/|\varepsilon|$. ■

3.3.2. Funciones condicionalmente definidas positivas

Ahora, se regresará al problema de interpolación que se obtuvo a partir de la generalización del spline cúbico natural en el caso multivariado cuando tomamos la función de aproximación $s(\mathbf{x})$ (véase ecuación (3.9)) como,

$$s(\mathbf{x}) = \sum_{i=1}^N \lambda_i \phi(\|\mathbf{x} - \mathbf{x}_i\|) + p(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d.$$

donde los coeficientes $\{\lambda_i\}$ deben cumplir las condiciones (3.10),

$$\sum_{i=1}^N \lambda_i q(\mathbf{x}_i) = 0 \quad \text{para todo } q \in \Pi_{m-1}(\mathbb{R}^d).$$

Es necesario encontrar funciones ϕ que den lugar a un problema de interpolación bien planteado al considerar ahora también la parte polinomial. El tipo de funciones que permiten hacer esto se conocen como funciones condicionalmente definidas positivas, pero antes de pasar a estudiarlas, daremos algunas definiciones previas que serán necesarias más adelante.

Definición 3.9. Un conjunto de puntos $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subset \mathbb{R}^d$ es llamado m -unisolvante si el polinomio de grado total a lo más m que interpola un conjunto de datos cero sobre X es el polinomio cero.

Esta definición garantiza la solución única de la interpolación para datos dados en un subconjunto de cardinalidad $M = \binom{m+d}{m}$ de los puntos $\mathbf{x}_1, \dots, \mathbf{x}_N$ por un polinomio de grado m . Aquí M es la dimensión del espacio lineal $\Pi_m(\mathbb{R}^d)$ de polinomios de grado total menor o igual que m en d variables.

Observación 3.10. Para referencias futuras se hace notar que la $(m-1)$ -unisolvencia de los puntos $\mathbf{x}_1, \dots, \mathbf{x}_N$ es equivalente al hecho de que la matriz P con

$$P_{i\ell} = p_\ell(\mathbf{x}_i) \quad i = 1, \dots, N, \ell = 1, \dots, M,$$

donde p_ℓ forma una base de $\Pi_m(\mathbb{R}^d)$, tiene rango completo (por columnas). Para $M = N$ esta es la matriz de interpolación polinomial.

En analogía a la sección anterior donde se trató la solución del problema de interpolación con el uso de funciones definidas positivas se introduce las funciones condicionalmente definidas positivas y las funciones estrictamente condicionalmente definidas positivas, para el problema de interpolación con funciones de base radial dado por la ecuación (3.9).

Definición 3.11. Una función continua de variable compleja Φ es llamada condicionalmente definida positiva de orden m sobre \mathbb{R}^d si

$$\sum_{i=1}^N \sum_{k=1}^N c_i \bar{c}_k \Phi(\mathbf{x}_i - \mathbf{x}_k) \geq 0 \quad (3.14)$$

para cualquier N puntos distintos por pares $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbb{R}^d$, y $\mathbf{c} = (c_1, \dots, c_N)^\top \in \mathbb{C}^N$, satisfaciendo

$$\sum_{i=1}^N c_i p(\mathbf{x}_i) = 0,$$

para cualquier polinomio de variable compleja p de grado a lo más $m-1$. La función Φ es llamada estrictamente condicionalmente definida positiva de orden m sobre \mathbb{R}^d si la forma cuadrática (3.14) es cero solo para $\mathbf{c} \equiv \mathbf{0}$.

Una observación inmediata es.

Lema 3.10. *Una función que es (estrictamente) condicionalmente definida positiva de orden m sobre \mathbb{R}^d es incluso (estrictamente) condicionalmente definida positiva de cualquier orden mayor. En particular, una función (estrictamente) definida positiva es siempre (estrictamente) condicionalmente definida positiva de cualquier orden.*

De la misma manera como se hizo para funciones definidas positivas, puede restringirse a funciones de variable real pares Φ y coeficientes reales.

Teorema 3.11. *Una función par continua de variable real Φ es llamada condicionalmente definida positiva de orden m sobre \mathbb{R}^d si*

$$\sum_{i=1}^N \sum_{k=1}^N c_i c_k \Phi(\mathbf{x}_i - \mathbf{x}_k) \geq 0 \quad (3.15)$$

para cualquier N puntos distintos por pares $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbb{R}^d$, y $\mathbf{c} = (c_1, \dots, c_N)^\top \in \mathbb{R}^N$, satisfaciendo

$$\sum_{i=1}^N c_i p(\mathbf{x}_i) = 0,$$

para cualquier polinomio de variable real p de grado a lo más $m-1$. La función Φ es llamada estrictamente condicionalmente definida positiva de orden m sobre \mathbb{R}^d si la forma cuadrática (3.15) es cero solo para $\mathbf{c} \equiv \mathbf{0}$.

A continuación se mostrará que el problema de interpolación dado por la ecuación (3.9) esta bien planteado cuando se usan funciones estrictamente condicionalmente definidas positivas.

Teorema 3.12. *Si la función par continua de variable real Φ es estrictamente condicionalmente positiva definida de orden m sobre \mathbb{R}^d y los puntos $\mathbf{x}_1, \dots, \mathbf{x}_N$ forman un conjunto $(m-1)$ -unisolvente, entonces el sistema de ecuaciones*

$$\begin{pmatrix} A & P \\ P^\top & O \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} \quad (3.16)$$

donde $A_{ik} = \Phi_k(\mathbf{x}_i) = \phi(\|\mathbf{x}_i - \mathbf{x}_k\|)$, $i, k = 1, \dots, N$, $P_{i\ell} = p_\ell(\mathbf{x}_i)$ los elementos de una base para $\Pi_m(\mathbb{R}^d)$, $\ell = 1, \dots, M$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)^\top$, $\mathbf{c} = (c_1, \dots, c_M)^\top$, $\mathbf{b} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_N))^\top$, $\mathbf{0}$ es el vector cero de tamaño M y O es una matriz de ceros de tamaño $M \times M$, tiene solución única.

Demostración. Supóngase que $(\boldsymbol{\lambda}, \mathbf{c})^\top$ es solución al sistema de ecuaciones lineales homogéneo, es decir, con $\mathbf{b} = \mathbf{0}$. Se mostrará que $(\boldsymbol{\lambda}, \mathbf{c})^\top = \mathbf{0}$ es la única solución posible.

La multiplicación por la izquierda del bloque superior por $\boldsymbol{\lambda}^\top$ da

$$\boldsymbol{\lambda}^\top A \boldsymbol{\lambda} + \boldsymbol{\lambda}^\top P \mathbf{c} = 0.$$

Por otra parte, del bloque de abajo de (3.16) se tiene que $P^\top \boldsymbol{\lambda} = \mathbf{0}$, lo cual implica que $\boldsymbol{\lambda}^\top P = \mathbf{0}^\top$ y por lo tanto

$$\boldsymbol{\lambda}^\top A \boldsymbol{\lambda} = 0.$$

Ya que la función Φ es estrictamente condicionalmente definida positiva de orden m por suposición, sabemos que la forma cuadrática de A (con coeficientes tal que $P^\top \boldsymbol{\lambda} = \mathbf{0}$), dada en la última ecuación es cero solo para $\boldsymbol{\lambda} = \mathbf{0}$, consecuentemente $\boldsymbol{\lambda} = \mathbf{0}$. La unisolvencia de los puntos de los datos, es decir, la independencia lineal de las columnas de P (véase observación 3.10) y el hecho de que $\boldsymbol{\lambda} = \mathbf{0}$ garantizan que $\mathbf{c} = \mathbf{0}$ a partir del bloque superior $A \boldsymbol{\lambda} + P \mathbf{c} = \mathbf{0}$ de (3.16). \square

Como antes se utiliza la caracterización mediante transformadas de Fourier para identificar las funciones que son estrictamente condicionalmente definidas positivas de orden m sobre \mathbb{R}^d , obtenida de la generalización de los resultados vistos para las funciones estrictamente definidas positivas.

Teorema 3.13. *Supóngase que la función de variable compleja $\Phi \in \mathcal{B}$ posee una transformada generalizada de Fourier $\widehat{\Phi}$ de orden m la cual es continua sobre $\mathbb{R}^d \setminus \{\mathbf{0}\}$. Entonces Φ es estrictamente condicionalmente definida positiva de orden m si y sólo $\widehat{\Phi}$ es no negativa y no se desvanece.*

El teorema anterior establece que las funciones estrictamente condicionalmente definidas positivas sobre \mathbb{R}^d son caracterizadas por el orden de la singularidad de sus transformadas de Fourier generalizadas en el origen, suponiendo que esta transformada es no negativa y no cero.

Ejemplo 3.4 (Multicuadráticas generalizadas). Las funciones multicuadráticas generalizadas,

$$\Phi(\mathbf{x}) = (1 + \|\mathbf{x}\|^2)^\beta, \quad \mathbf{x} \in \mathbb{R}^d, \beta \in \mathbb{R} \setminus \mathbb{N}_0,$$

tienen transformadas generalizadas de Fourier,

$$\widehat{\Phi}(\boldsymbol{\omega}) = \frac{2^{1+\beta}}{\Gamma(-\beta)} \|\boldsymbol{\omega}\|^{-\beta-d/2} K_{\beta+d/2}(\|\boldsymbol{\omega}\|), \quad \boldsymbol{\omega} \neq \mathbf{0},$$

de orden $m = \max(0, \lceil \beta \rceil)$, donde K_ν es nuevamente la función modificada de Bessel de segundo tipo de orden ν . Se necesita excluir los valores de los enteros para β ya que esto conduciría a polinomios de grado par (véase ejemplo 3.5).

Ya que las transformadas generalizadas de Fourier son positivas con una singularidad de orden m en el origen, el teorema 3.13 dice que las funciones,

$$\Phi(\mathbf{x}) = (-1)^{\lceil \beta \rceil} (1 + \|\mathbf{x}\|^2)^\beta, \quad \beta \in \mathbb{R} \setminus \mathbb{N}$$

son estrictamente condicionalmente definidas positivas de orden $m = \lceil \beta \rceil$. Para $\beta < 0$ se vuelve a las funciones inversas multicuadráticas generalizadas que son estrictamente condicionalmente definidas positivas de orden $m = 0$, *i.e.*, estrictamente definidas positivas.

De manera análoga al ejemplo 3.3, para las funciones multicuadráticas generalizadas es posible utilizar el parámetro de forma ε , de tal manera que estas se reescriben como $\Phi(\mathbf{x}) = (1 + \varepsilon^2 \|\mathbf{x}\|^2)^\beta$. La figura 3.3 muestra la función multicuadrática para $\beta = 1/2$ y diferentes parámetros de forma ε . Las funciones obtenidas son estrictamente definidas positivas de orden 1 y están relacionadas con las funciones propuestas por Hardy, por lo que, para este valor de β , llevan su nombre.

Los argumentos dados junto con el teorema 3.12 muestran que las funciones multicuadráticas de Hardy pueden ser utilizadas en la forma,

$$s(\mathbf{x}) = \sum_{i=1}^N \lambda_i \sqrt{1 + \varepsilon^2 \|\mathbf{x} - \mathbf{x}_i\|^2} + \tau \quad \mathbf{x} \in \mathbb{R}^d,$$

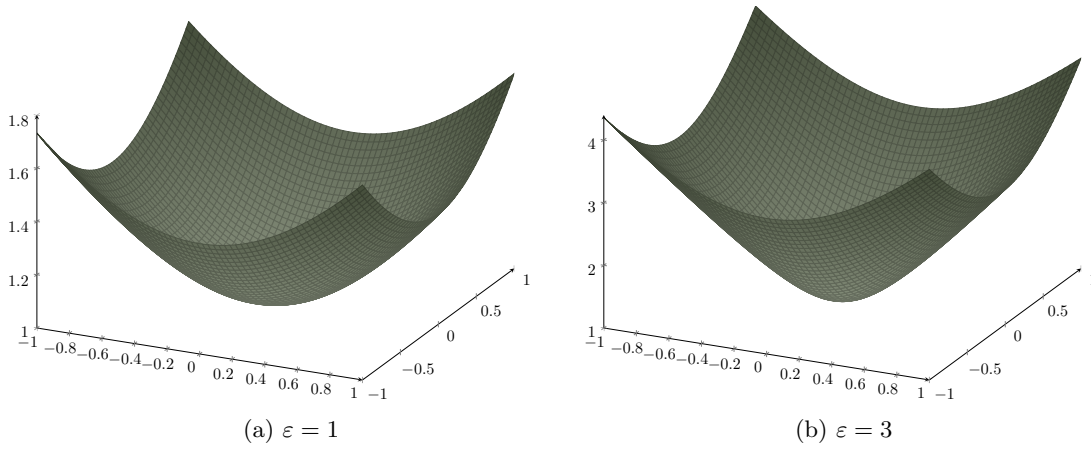


Figura 3.3: Función multicuadrática ($\beta = 1/2$) centrada en el origen en \mathbb{R}^2

junto con las condiciones,

$$\sum_{i=1}^N \lambda_i = 0,$$

para resolver el problema de interpolación multivariado. El uso del parámetro de forma no afecta el resultado dado por el teorema 3.12. ■

Ejemplo 3.5 (Potencias radiales). Las potencias radiales,

$$\Phi(\mathbf{x}) = \|\mathbf{x}\|^\beta, \quad \mathbf{x} \in \mathbb{R}^d, 0 < \beta \notin 2\mathbb{N}$$

tienen transformadas generalizadas de Fourier,

$$\widehat{\Phi}(\boldsymbol{\omega}) = \frac{2^{\beta+d/2} \Gamma\left(\frac{d+\beta}{2}\right)}{\Gamma(-\beta/2)} \|\boldsymbol{\omega}\|^{-\beta-d}, \quad \boldsymbol{\omega} \neq \mathbf{0},$$

de orden $m = \lceil \beta/2 \rceil$. Por lo tanto, las funciones

$$\Phi(\mathbf{x}) = (-1)^{\lceil \beta/2 \rceil} \|\mathbf{x}\|^\beta, \quad 0 < \beta \notin 2\mathbb{N},$$

son estrictamente condicionalmente definidas positivas de orden $\lceil \beta/2 \rceil$. La figura 3.4 muestra las potencias radiales lineal ($\beta = 1$, estrictamente condicionalmente definida positiva de orden 1) y cúbica ($\beta = 3$, estrictamente condicionalmente definida positiva de orden 2). De esta manera, las funciones base $\Phi(\mathbf{x}) = \|\mathbf{x}\|$, usadas para la matriz de distancia son estrictamente condicionalmente definidas positivas de orden 1, por lo que, de acuerdo con el teorema 3.12 estas debieron ser utilizadas junto con una constante como en el ejemplo anterior. El teorema 9.7 en Fasshauer [5] da una justificación para su uso como una matriz de distancias.

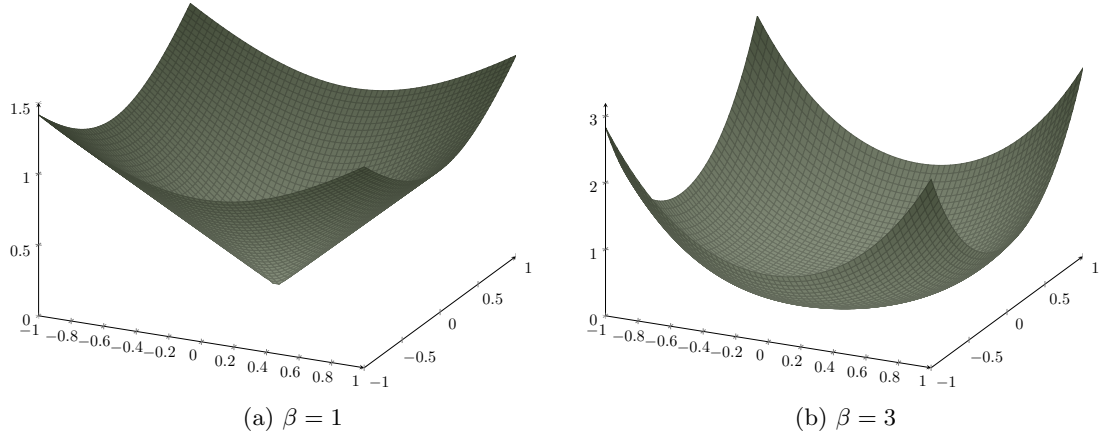


Figura 3.4: Potencias radiales para diferentes valores de β centrada en el origen en \mathbb{R}^2

Obsérvese que las potencias radiales no son afectadas por el escalamiento de sus argumentos, *i.e.*, el uso del parámetro de forma no tiene sentido aquí, es por ello que estas funciones se dice que son libres del parámetro de forma. Por último, el motivo por que se excluyen las potencias pares, es debido a que, una potencia par combinada con la raíz cuadrada en la definición de la norma euclidiana resulta en un polinomio, los cuales no pueden ser usados en el problema de interpolación para cualquier conjunto de posiciones dispersas.

■

Ejemplo 3.6 (Splines de placa delgada). En el ejemplo anterior se tiene que descartar las potencias pares, pero al multiplicar las potencias radiales por un término logarítmico se puede tener otras funciones que son definidas positivas para estas potencias.

Los splines de placa delgada de Duchon,

$$\Phi(\mathbf{x}) = \|\mathbf{x}\|^{2\beta} \log \|\mathbf{x}\| \quad \mathbf{x} \in \mathbb{R}^d, \beta \in \mathbb{N},$$

tienen transformadas generalizadas de Fourier,

$$\widehat{\Phi}(\boldsymbol{\omega}) = (-1)^{\beta+1} 2^{2\beta-1+d/2} \Gamma(\beta + d/2) \beta! \|\boldsymbol{\omega}\|^{-d-2\beta}$$

de orden $m = \beta + 1$. Por lo tanto, las funciones

$$\Phi(\mathbf{x}) = (-1)^{\beta+1} \|\mathbf{x}\|^{2\beta} \log \|\mathbf{x}\| \quad \beta \in \mathbb{N},$$

son estrictamente condicionalmente definidas positivas de orden $m = \beta + 1$.

La figura 3.5 muestra el spline placa delgada “clásico” ($\beta = 1$, estrictamente condicionalmente definida positiva de orden 2) y de orden 3 para $\beta = 2$. Observe que estas funciones

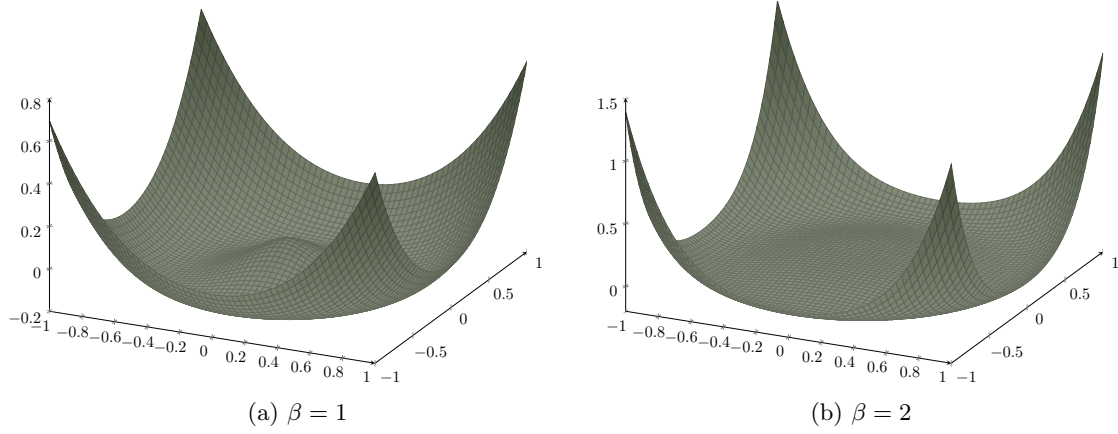


Figura 3.5: Spline de placa delgada para diferentes valores de β centrada en el origen en \mathbb{R}^2

no son monótonas además de que ambas tienen valores negativos.

Al igual que para las potencias radiales, el uso del parámetro de forma ε junto con los splines de placa delgada no tiene sentido (Fasshauer [5]). En particular, para la interpolación de datos dispersos en \mathbb{R}^2 , se puede usar,

$$s(\mathbf{x}) = \sum_{i=1}^N \lambda_i \|\mathbf{x} - \mathbf{x}_i\|^2 \log \|\mathbf{x} - \mathbf{x}_i\| + \tau_0 + \tau_1 x + \tau_2 y, \quad \mathbf{x} = (x, y)^T \in \mathbb{R}^2,$$

junto con las condiciones,

$$\sum_{i=1}^N \lambda_i = 0, \quad \sum_{i=1}^N \lambda_i x_i = 0, \quad \sum_{i=1}^N \lambda_i y_i = 0,$$

teniendo en cuenta que las posiciones no sean todas colineales. ■

3.4. Aspectos prácticos

A partir de las dos formas de definir la función de aproximación $s(\mathbf{x})$ se obtuvieron dos maneras diferentes de plantear el problema de interpolación mediante el uso de funciones radiales. En general se puede considerar solamente como el interpolador de base radial al que utiliza, adicionalmente a la base de funciones radiales, la parte polinomial, esto es, se toma como,

$$F_{RBF}(\mathbf{x}) = \sum_{i=1}^N \lambda_i \phi(\|\mathbf{x} - \mathbf{x}_i\|) + p(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d, p \in \Pi_{m-1}(\mathbb{R}^d),$$

donde los coeficientes $\{\lambda_i\}$ deben cumplir,

$$\sum_{i=1}^N \lambda_i q(\mathbf{x}_i) = 0 \quad \text{para todo } q \in \Pi_{m-1}(\mathbb{R}^d),$$

que junto con las condiciones de interpolación $F_{RBF}(\mathbf{x}_i) = f(\mathbf{x}_i) = z_i$, $i = 1, \dots, N$, reducen el problema de interpolación a encontrar la solución del sistema de ecuaciones $B\hat{\boldsymbol{\lambda}} = \hat{\mathbf{b}}$ donde

$$B = \begin{pmatrix} A & P \\ P^\top & O \end{pmatrix}, \quad \hat{\boldsymbol{\lambda}} = \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{c} \end{pmatrix} \quad \text{y} \quad \hat{\mathbf{b}} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix},$$

con $A_{ik} = \Phi_k(\mathbf{x}_i) = \phi(\|\mathbf{x}_i - \mathbf{x}_k\|)$, $i, k = 1, \dots, N$, $P_{i\ell} = p_\ell(\mathbf{x}_i)$, $\ell = 1, \dots, M$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)^\top$, $\mathbf{c} = (c_1, \dots, c_M)^\top$, $\mathbf{b} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_N))^\top$ (*cfr.* ecuación (3.16) para más detalles).

Cuando no se utiliza la parte polinomial, y por lo tanto, tampoco las condiciones adicionales sobre los coeficientes $\{\lambda_i\}$, se regresa al planteamiento en el cual solo se utiliza la base de funciones, quedando de esta manera como un caso particular donde el sistema de ecuaciones a resolver se reduce al sistema $A\boldsymbol{\lambda} = \mathbf{b}$.

Para garantizar la solución única de los sistemas y por consiguiente del problema de interpolación, además de las condiciones impuestas sobre los coeficientes λ_i , es necesario que la base de funciones sea elegida de manera particular. Por esta razón, se han mostrado los resultados teóricos que dan una gama diversa de funciones estrictamente definidas positivas y estrictamente condicionalmente definidas positivas que generan matrices A y B no singulares, que garantizan la solución única de los sistemas de ecuaciones (*cfr.* teorema 3.12).

En los ejemplos 3.2-3.6 se mostraron algunas de las familias de funciones más utilizadas y que dan una variedad extensa de funciones de base radial a elegir, pero las que se utilizarán son las siguientes:

- $\Phi(\mathbf{x}) = \phi(\|\mathbf{x}\|) = e^{-\varepsilon^2 \|\mathbf{x}\|^2}$ (Gaussianas),
- $\Phi(\mathbf{x}) = \phi(\|\mathbf{x}\|) = (1 + \varepsilon^2 \|\mathbf{x}\|^2)^{-1/2}$ (Multicuadráticas inversas),

- $\Phi(\mathbf{x}) = \phi(\|\mathbf{x}\|) = (1 + \varepsilon^2 \|\mathbf{x}\|^2)^{1/2}$ (Multicuadráticas),
- $\Phi(\mathbf{x}) = \phi(\|\mathbf{x}\|) = \|\mathbf{x}\|^\beta$, para $\beta = 1, 3$ (Potencias radiales),
- $\Phi(\mathbf{x}) = \phi(\|\mathbf{x}\|) = \|\mathbf{x}\|^2 \log \|\mathbf{x}\|$ (Splines de placa delgada),

donde nuevamente $\varepsilon > 0$ denota el parámetro de forma.

Con todo lo expuesto hasta el momento, el procedimiento para resolver el problema de interpolación mediante funciones de base radial consiste en construir alguno de los sistemas de ecuaciones anteriores (dependiendo de la función de base radial utilizada) y darle solución. De esta manera, en dependencia de los datos, se define de manera única la función interpolante F_{RBF} . El proceso para el caso de dos dimensiones queda descrito en el algoritmo 3.1, esto por las entradas del algoritmo, aunque se puede utilizar en el caso de mas dimensiones.

Algoritmo 3.1: Interpolación mediante funciones de base radial

Entrada: datos dispersos $\mathcal{P} = \{(x_i, y_i, z_i)\}$, posiciones de evaluación
 $\mathcal{Q} = \{(\hat{x}_j, \hat{y}_j)\}_{j=1, \dots, J}$, función de base radial ϕ , parámetro de forma ε
Salida: vector de tamaño J con los valores en las posiciones de evaluación $\hat{\mathbf{z}}$

```

1  $A \leftarrow O_{((N+M) \times (N+M))} // O_{((N+M) \times (N+M))}$  es la matriz de ceros
2 Para  $i \leftarrow 1$  hasta  $N$  hacer // Construcción del sistema  $B\hat{\lambda} = \hat{\mathbf{b}}$ 
3   Para  $k \leftarrow i$  hasta  $N$  hacer
4      $A_{i,k} \leftarrow \phi(\|\mathbf{x}_i - \mathbf{x}_k\|, \varepsilon)$ 
5      $A_{k,i} \leftarrow A_{i,k}$ 
6   Para  $\ell \leftarrow 1$  hasta  $M$  hacer
7      $A_{\ell+N,i} \leftarrow p_\ell(\mathbf{x}_i)$ 
8      $A_{i,\ell+N} \leftarrow A_{\ell+N,i}$ 
9    $\mathbf{b}_i \leftarrow z_i$ 
10 Resolver el sistema  $A\lambda = \mathbf{b}$ 
11 Para  $j \leftarrow 1$  hasta  $J$  hacer
12   Para  $i \leftarrow 1$  hasta  $N$  hacer
13      $C_{j,i} \leftarrow \phi(\|\hat{\mathbf{x}}_j - \mathbf{x}_i\|, \varepsilon)$ 
14   Para  $\ell \leftarrow 1$  hasta  $M$  hacer
15      $C_{j,\ell+N} \leftarrow p_\ell(\hat{\mathbf{x}}_j)$ 
16  $\hat{\mathbf{z}} \leftarrow C\lambda$ 

```

Observación 3.12. Algunas de las funciones de base radial que se utilizarán, como por ejemplo las potencias radiales y los splines de placa delgada, no requieren del parámetro de forma ε , por lo que debe de ser un argumento de entrada opcional del algoritmo 3.1. También, como se menciono anteriormente, la parte polinomial usada en el sistema es opcional, sobre todo para el caso en que se usan funciones estrictamente definidas positivas. En caso de

requerirse, los monomios utilizados por lo general en la práctica son a lo más de grado 2. Así, los que usualmente se utilizan para el caso de interpolación en dos dimensiones son:

$$p_1 = 1, p_2 = x, p_3 = y, p_4 = x^2, p_5 = y^2, p_6 = xy.$$

De esta manera, el interpolador de base radial para el caso de 2 dimensiones queda dada por,

$$F_{RBF}(\mathbf{x}) = \sum_{i=1}^N \lambda_i \phi(\|\mathbf{x} - \mathbf{x}_i\|) + \sum_{\ell}^M c_{\ell} p_{\ell}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^2,$$

donde los coeficientes $\{\lambda_i\}$ deben cumplir,

$$\sum_{i=1}^N \lambda_i p_{\ell}(\mathbf{x}_i) = 0, \quad \text{para } \ell = 1, \dots, M.$$

Para $M = 0$ no se utiliza parte polinomial, y para $M = 1$, $M = 3$ y $M = 6$ se utiliza la parte polinomial con monomios a lo más de grado 0, 1 y 2, respectivamente.

Finalmente, se exponen algunos ejemplos del uso de las funciones de base radial para resolver el problema de interpolación, además de mostrar como el uso del parámetro de forma afecta la solución entre otros aspectos de esta técnica.

Ejemplo 3.7. Considérese nuevamente, como en el ejemplo 2.4, la distribución aleatoria de 50 puntos, para los cuales se evalúa la función de prueba $f_1(\mathbf{x})$ de Franke dada en [6].

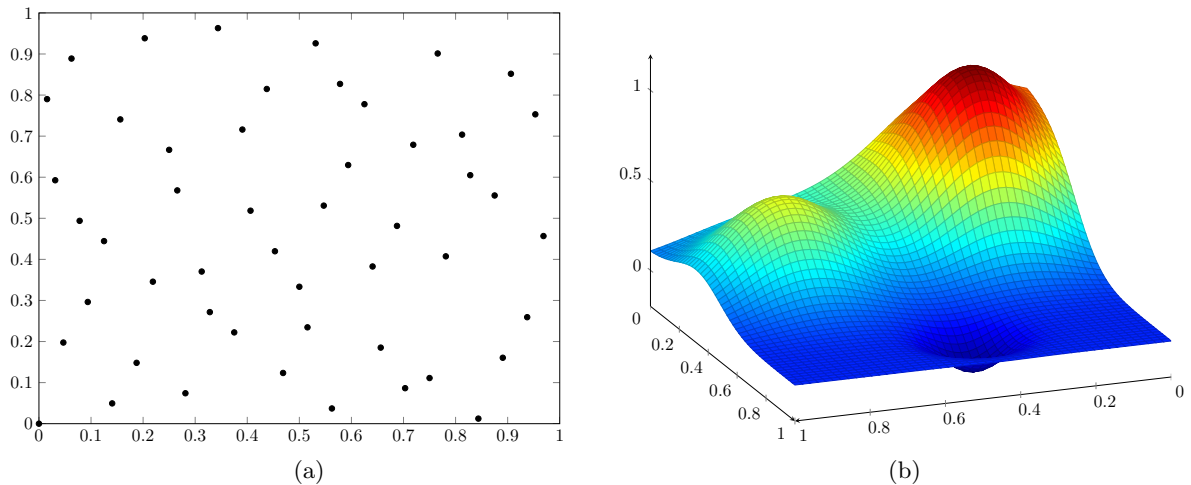


Figura 3.6: (a) Una distribución aleatoria de puntos. (b) Gráfica de la función de prueba f_1

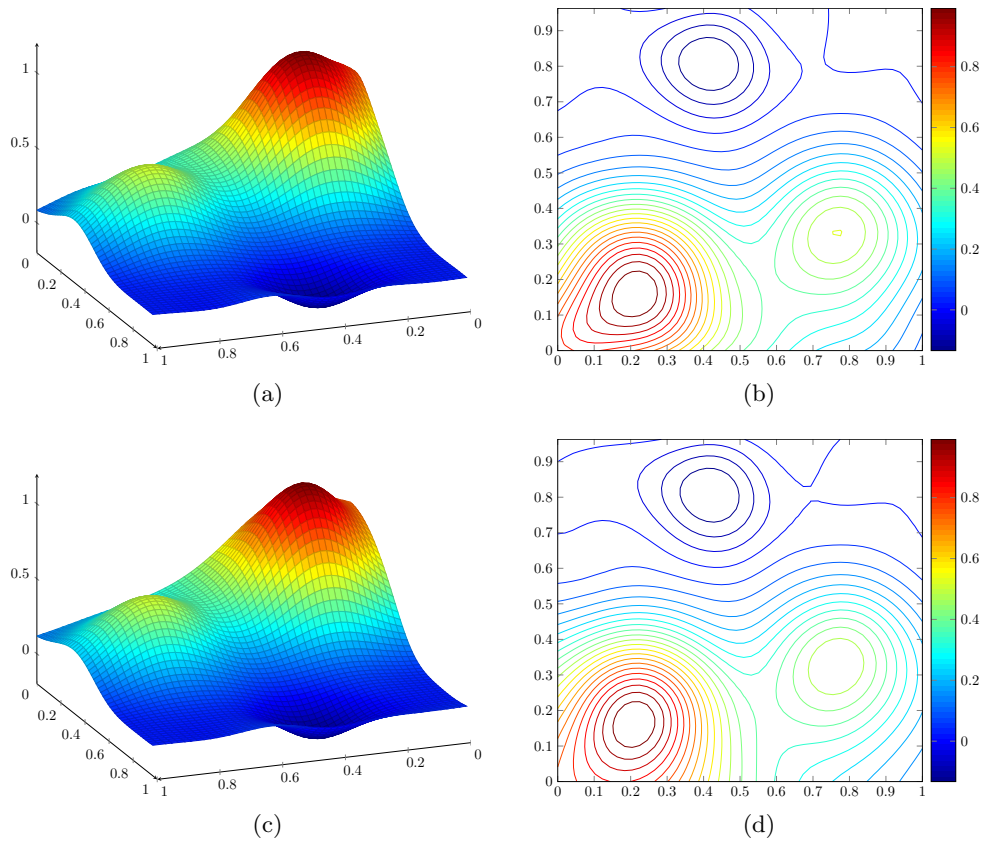


Figura 3.7: Superficies obtenidas mediante funciones de base radial para un parámetro de forma $\varepsilon = 5$. (a) Superficie obtenida mediante funciones gaussianas. (b) Curvas de nivel. (c) Superficie obtenida mediante funciones multicuadráticas. (d) Curvas de nivel.

La figura 3.7 muestra las superficies obtenidas mediante el uso de funciones radiales. La figura 3.7a muestra la superficie que se obtuvo utilizando funciones gaussianas ($\phi(\|\mathbf{x}\|) = e^{\varepsilon^2\|\mathbf{x}\|^2}$) mientras que la figura 3.7c muestra la superficie que se tiene al usar funciones multicuadráticas ($\phi(\|\mathbf{x}\|) = (1 + \varepsilon^2\|\mathbf{x}\|^2)^{1/2}$). Con cualquiera de los dos tipos de funciones se lograron buenos resultados de la aproximación de la función f_1 para el valor del parámetro de forma elegido que en ambos casos fue $\varepsilon = 5$. En este ejemplo el parámetro de forma fue elegido a prueba y error, de manera que visualmente la superficie interpolada fuera lo suficientemente parecida a la función original f_1 . ■

En el siguiente ejemplo se muestra como afecta a la solución del problema de interpolación la cantidad de datos disponibles y el valor del parámetro de forma que se elige.

Ejemplo 3.8. Se tomó nuevamente la función de prueba $f_1(\mathbf{x})$ de Franke y las funciones de base radial gaussianas. Para un valor fijo del parámetro de forma, $\varepsilon = 21.1$, se probaron algu-

k	N	error-RMS	error-max	RCOND
3	9	3.852831e-001	1.040605e+000	9.999922e-001
5	25	3.288661e-001	9.689059e-001	9.965305e-001
9	81	2.110352e-001	8.551264e-001	4.755635e-001
17	289	6.330603e-002	4.826697e-001	2.024096e-002
33	1089	6.355019e-003	8.584931e-002	4.615718e-006
65	4225	3.278815e-005	5.894380e-004	2.570946e-016

Tabla 3.1: Cantidades medidas para diferentes tamaños de distribuciones aleatorias

nas distribuciones aleatorias en la región $\mathcal{D} = [0, 1] \times [0, 1] \subset \mathbb{R}^2$, donde el tamaño del conjunto de datos esta dado por la progresión $\{(2^k + 1)^2 \mid k \in \mathbb{N}\} = \{9, 25, 81, 289, 1089, 4225, \dots\}$. El valor de ε es tomado así para tener funciones base más localizadas con respecto a las posiciones de los datos \mathbf{x}_i (véase el ejemplo 3.2 para más detalles).

Para evaluar el desempeño de la interpolación con funciones de base radial para los distintas distribuciones aleatorias, se midieron las siguientes cantidades:

- El error cuadrático medio (error-RMS; *root-mean-square error*), esto es,

$$\left(\frac{1}{J} \sum_{j=1}^J [F_{RBF}(\hat{\mathbf{x}}_j) - f(\hat{\mathbf{x}}_j)]^2 \right)^{1/2} = \frac{1}{\sqrt{J}} \|F_{RBF} - f\|_2.$$

donde $\hat{\mathbf{x}}_j$, $j = 1, \dots, J$ son las posiciones de evaluación.

- El error máximo (error-max),

$$\max_j |(F_{RBF}(\hat{\mathbf{x}}_j) - f(\hat{\mathbf{x}}_j))| = \|F_{RBF} - f\|_\infty.$$

- El estimador del recíproco del número de condición de la matriz de intepolación (RCOND). Este es calculado mediante la función `rcond` de `Matlab`.

En la tabla 3.8 se muestran las cantidades obtenidas para los diversos tamaños de distribuciones utilizados. Lo primero que se observa es que el error disminuye mientras más valores en las posiciones de los datos se tengan disponibles, esto sucede obviamente porque mientras más información se tenga de la función desconocida f mejor será la aproximación F_{RBF} . Por otro lado, mientras más aumenta el conjunto de datos, RCOND tiende a cero, es decir, la matriz de interpolación es mal condicionada, por lo que, si el conjunto de datos es muy grande la solución del sistema es poco confiable y por tanto la solución al problema de interpolación.

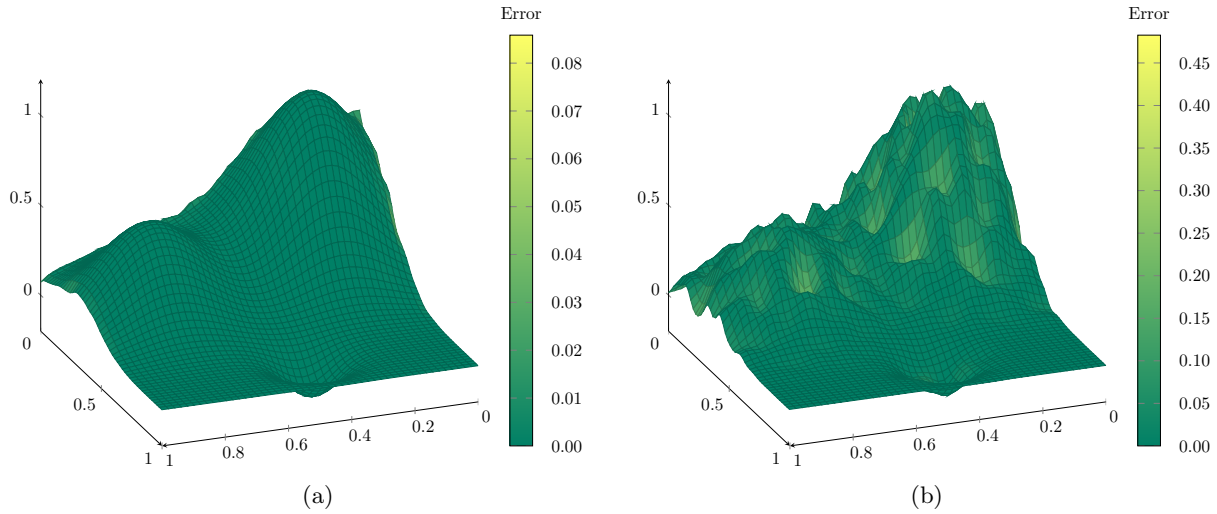


Figura 3.8: Interpolación mediante funciones de base radial para la función de Franke f_1 usando funciones gaussianas ($\varepsilon = 21.1$) (a) $N = 1089$. (b) $N = 289$.

Este es uno de los problemas a los cuales se enfrenta esta técnica al ser un método global, por lo cual, al igual que con el método de Shepard, existen modificaciones para hacerlo un método local, aunque con esto se pierde la ventaja de la rapidez del método al solamente resolver un único sistema de ecuaciones. Otra forma de abordar este problema es el de utilizar preconditionares para mejorar el número de condición de la matriz de interpolación. Para los tipos de problemas que se quieren tratar con el sistema, el enfoque del método de funciones radiales visto es suficiente. Para mayores referencias sobre las modificaciones al método véase Fasshauer [5] y Wendland [22].

La figura 3.8 muestra las superficies para el caso de $N = 1089$ (figura 3.8a) y $N = 289$ (figura 3.8b). En ambos casos el mapa de colores esta asociado con el error absoluto en cada uno de los nodos de la malla. Como se está utilizando un parámetro de forma que hace que las funciones base sean más centralizadas, cuando el tamaño de la distribución es pequeño, la superficie presenta varios picos y por tanto el error crece. Para el conjunto de datos más grande, los espacios donde no hay información se reducen, y por tanto el error también.

Para terminar esta parte se verá como se comportan los errores cuando se deja fijo el tamaño de la distribución de las posiciones aleatorias mientras se varia el valor del parámetro de forma ε cuando se utilizan funciones de base radial gaussianas. En la figura 3.9 se muestran los resultados. En todas las gráficas se puede apreciar una zona “errática” para valores pequeños del parámetro de forma (cercanos a cero), esto es debido a que para esos valores de ε , el sistema de ecuaciones es mal condicionado. También se aprecia que siempre hay un valor de ε para el cual ambos errores obtienen su valor mínimo. Este valor mínimo puede

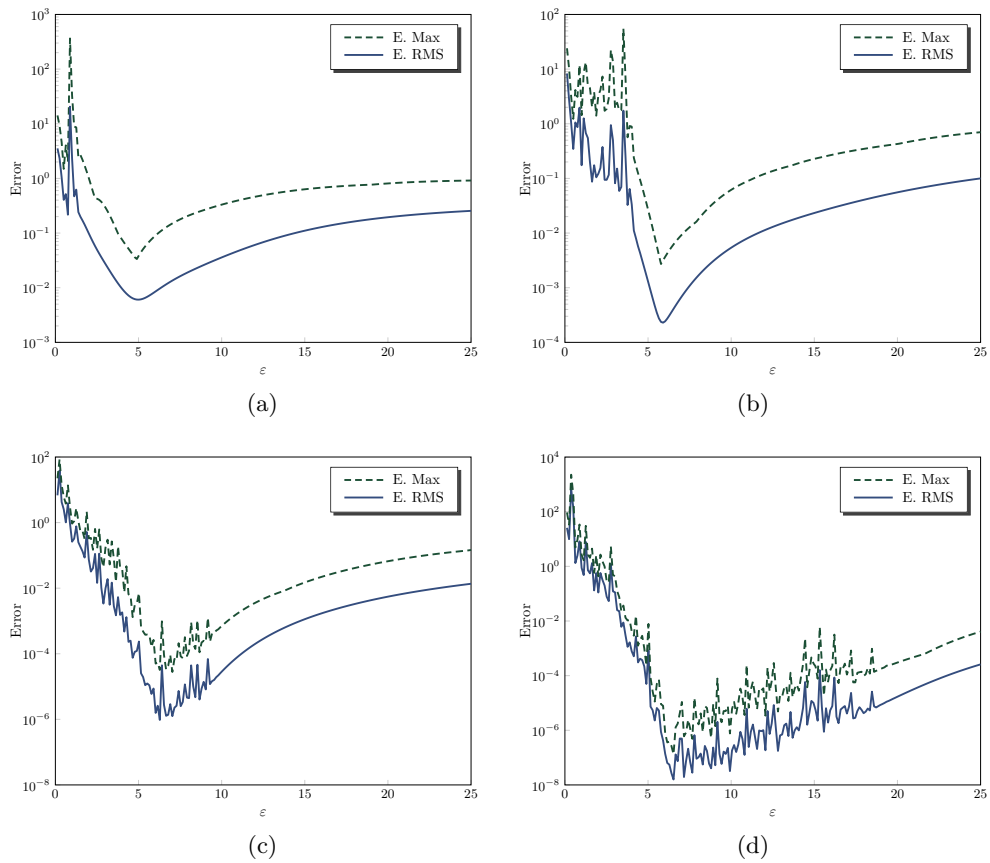


Figura 3.9: Error máximo (--) y RMS (—) obtenido para diferentes valores de forma ϵ y diferentes tamaños de distribuciones aleatorias. (a) $N = 81$. (b) $N = 289$. (c) $N = 1089$. (d) $N = 4225$.

ser alcanzado dentro de la zona errática (cuando el tamaño del conjunto de datos es grande, figuras 3.9c-3.9d) o fuera de ella (en el caso contrario, figuras 3.9a-3.9b). En general, estimar cual es el mejor parámetro de forma para cierta base de funciones no es un problema sencillo. Varias técnicas de como lograr esto son descritas en Fasshauer [5].



Capítulo 4

Métodos geoestadísticos

En varios campos de la ciencia y la industria se tienen conjuntos de datos dispersos obtenidos mediante mediciones de alguna variable de interés. Para el estudio de los datos medidos, que en particular siempre tienen asociados una posición particular en el espacio, se usan actualmente modelos matemáticos dentro de los cuales la geoestadística ha jugado un papel importante en las últimas décadas y ha probado su superioridad entre muchos tipos de estimación. Esto se debe a que, a diferencia de los métodos estadísticos clásicos, ofrecen una forma de describir la continuidad espacial, una característica esencial en la mayoría de los conjuntos de datos en ciencias de la tierra, y provee adaptaciones de técnicas clásicas de regresión para tomar ventaja de esta continuidad.

En particular, dentro de los métodos geoestadísticos, el método *kriging* es uno de los más utilizados ya que ha probado su efectividad en varios campos. Este método originalmente se creó en el campo de la minería cuando el ingeniero minero Danie Krige y el estadista Herbert Sichel se enfrentaron al problema de encontrar una buena estimación de la concentración de oro en las minas de Witwatersrand en Sudáfrica, razón por la cual varios conceptos están relacionados con términos en minería. Krige y Sichel solo dieron las bases de lo que sería el método, posteriormente el ingeniero francés Georges Matheron formalizó estas ideas que fueron la base de la geoestadística.

A partir del éxito que ha tenido el método sobre el campo de la minería este ha extendido su uso sobre otros campos de ciencias de la tierra como geología, geofísica, hidrología y en otras áreas como lo son las ciencias ambientales y la industria petrolera.

4.1. Preliminares

Considérese nuevamente que el conjunto de los datos $\{z_i\}$ es obtenido a partir de una función escalar desconocida $f : \mathcal{D} \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ al ser evaluada en las posiciones arbitrarias $\mathbf{x}_i = (x_i, y_i)^\top \in \mathcal{D}$, $i = 1, \dots, N$.

Hasta el momento se han dado algunas formas de aproximar la función f a partir de otra función F la cual tiene una forma predefinida. Ahora se usará un enfoque diferente conocido dentro de los métodos geoestadísticos como kriging. La principal diferencia con respecto a los

métodos antes descritos es que este comienza a partir de un modelo estadístico del problema de estudio en lugar de un modelo de la función de interpolación.

Es dentro de este enfoque estadístico que el problema de interpolación se puede ver ahora como un problema de estimación. Para plantear dicho problema dentro de este enfoque es necesario previamente definir algunos conceptos de la teoría de probabilidad. En esta parte solo se presentarán los conceptos necesarios para el desarrollo del tema, algunos otros están comprendidos dentro del apéndice B.

Definición 4.1. Una variable regionalizada se define como la función $z(\mathbf{x})$ para toda $\mathbf{x} \in \mathcal{D}$.

De acuerdo con la definición anterior, el conjunto de los datos $\{z_i\}$ puede ser visto como una colección de algunos valores de una variable regionalizada al tomar $z(\mathbf{x}_i) = z_i$. El término “regionalizado” se utiliza en el contexto de que cada valor del conjunto de los datos es dependiente de la posición en la región \mathcal{D} .

Luego, se supone que cada valor en el conjunto de los datos $z(\mathbf{x}_i)$ es una realización de una variable aleatoria $Z(\mathbf{x}_i)$, donde esta variable aleatoria puede ser diferente en cada posición \mathbf{x}_i . De esta manera, los valores de los datos son vistos como salidas de un mecanismo aleatorio. En general, se considera que para cada posición $\mathbf{x} \in \mathcal{D}$, cada $z(\mathbf{x})$ es una realización de una variable aleatoria $Z(\mathbf{x})$.

Definición 4.2. La familia de variables aleatorias

$$\{Z(\mathbf{x}) \mid \mathbf{x} \in \mathcal{D}\}$$

es llamada una función aleatoria.

En la literatura una función aleatoria es llamada proceso estocástico cuando \mathbf{x} varía en un espacio unidimensional (y puede ser interpretada como el tiempo) y un campo aleatorio cuando \mathbf{x} varía en un espacio de más de una dimensión (Chilès & Delfiner [3]). Una función aleatoria es descrita por sus distribuciones finito dimensionales que es el conjunto de todas las distribuciones multidimensionales de k -tuplas $(Z(\mathbf{x}_1), Z(\mathbf{x}_2), \dots, Z(\mathbf{x}_k))$, para todos los valores finitos de k y todas las configuraciones de los puntos $\mathbf{x}_1, \dots, \mathbf{x}_k$, que en algunas ocasiones se le denomina la distribución espacial de la función aleatoria.

La media de la función aleatoria es el valor esperado $m(\mathbf{x}) = E[Z(\mathbf{x})]$ de la variable aleatoria $Z(\mathbf{x})$ en la posición \mathbf{x} . La media también es conocida como *drift* de Z , específicamente cuando $m(\mathbf{x})$ varía con la posición. La covarianza, que se denotará por $\sigma(\mathbf{x}, \mathbf{y})$, es la covarianza de las variables aleatorias $Z(\mathbf{x})$ y $Z(\mathbf{y})$, esto es

$$\sigma(\mathbf{x}, \mathbf{y}) = E[(Z(\mathbf{x}) - m(\mathbf{x}))(Z(\mathbf{y}) - m(\mathbf{y}))].$$

En general, esta función depende de ambas posiciones \mathbf{x} y \mathbf{y} . Cuando $\mathbf{x} = \mathbf{y}$,

$$\sigma(\mathbf{x}, \mathbf{x}) = \text{Var}(Z(\mathbf{x})),$$

es la varianza de $Z(\mathbf{x})$. Naturalmente, en la teoría, estos momentos pueden no existir. Como es usual en teoría de la probabilidad la media está definida solo si $E[|Z(\mathbf{x})|] < \infty$. Si $E[Z(\mathbf{x})]^2$ es finita en cada punto, se dice que $Z(\mathbf{x})$ es una función aleatoria de segundo orden, esto es, esta tiene una varianza finita y la covarianza existe donde sea.

La mayoría de las veces es difícil conocer las distribuciones finito dimensionales de una función aleatoria, por lo que un caso particular de gran importancia práctica es cuando estas son invariantes bajo una traslación arbitraria de los puntos por un vector \mathbf{h} , si pasa esto, la función aleatoria es llamada estacionaria. Físicamente esto significa que el fenómeno es homogéneo en el espacio y, por decirlo así, se repite a sí mismo en todo el espacio.

Cuando la función aleatoria es estacionaria, sus momentos, si existen, son invariantes bajo traslaciones. Si solo se consideran la media y la covarianza, se tiene que para puntos $\mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D} \subset \mathbb{R}^2$,

$$\begin{aligned} E[Z(\mathbf{x})] &= m, \\ E[(Z(\mathbf{x}) - m)(Z(\mathbf{x} + \mathbf{h}) - m)] &= C(\mathbf{h}), \end{aligned} \quad (4.1)$$

es decir, la media es constante y la función de covarianza solo depende de la separación \mathbf{h} .

Definición 4.3. Una función aleatoria satisfaciendo las condiciones dadas en (4.1), es estacionaria de segundo orden o débilmente estacionaria.

Una hipótesis más débil es el suponer que para cada vector \mathbf{h} el incremento $Y_{\mathbf{h}}(\mathbf{x}) = Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})$ es una función aleatoria estacionaria en \mathbf{x} , es decir, la media de los incrementos es invariante para cualquier traslación de un vector \mathbf{h} dado entre el dominio, más aún, se supone cero cualquiera sea la posición de \mathbf{h} en el dominio y la varianza de los incrementos tiene un valor finito $2\gamma(\mathbf{h})$, que depende de la longitud y orientación del vector \mathbf{h} , pero no de su posición en el dominio. En resumen, para cualquier par de puntos $\mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D} \subset \mathbb{R}^2$, los incrementos deben cumplir las siguientes relaciones,

$$\begin{aligned} E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] &= 0, \\ E[(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}))^2] &= 2\gamma(\mathbf{h}), \end{aligned} \quad (4.2)$$

donde la función $\gamma(\mathbf{h})$ es llamada semi-variograma (o simplemente variograma), la herramienta básica para la interpretación estructural del fenómeno como también de la estimación y la cual se estudiará con mayor detenimiento en la siguiente sección.

Definición 4.4. Una función aleatoria $Z(\mathbf{x})$ cuyos incrementos satisfacen las condiciones dadas en (4.2), es una función aleatoria intrínseca.

4.2. Algunos aspectos sobre el variograma

La teoría de procesos estocásticos y funciones aleatorias ha sido usada por un tiempo relativamente largo para resolver problemas de interpolación y filtrado. Los métodos que

se describirán más adelante están basados en los dos primeros momentos de las funciones aleatorias, sin embargo, en el mundo real estos nunca son conocidos a priori y deben ser determinados. Una de las fortalezas de la geoestadística es el proponer una metodología para la identificación de estas características del modelo estocástico al igual que para la interpolación. El fenómeno espacial estudiado aquí es generalmente:

- Único, no reproducible.
- Definido en un dominio en 2 ó 3 dimensiones.
- Demasiado complejo para una precisa descripción determinista.
- Conocido para una muestra de puntos no uniforme.

Estos fenómenos son variables regionalizadas $\{z(\mathbf{x}) \mid \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d\}$, que se consideran realizaciones de funciones aleatorias.

Como se menciona anteriormente, una función aleatoria $\{Z(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^d\}$ es caracterizada por sus distribuciones finito dimensionales que determinan su distribución espacial. En general, estas distribuciones no pueden ser determinadas, excepto bajo una suposición de estacionariedad que introduce repetición en el espacio. Para los métodos lineales que se describirán más adelante, y que son los más ampliamente utilizados, es suficiente el conocer los momentos de segundo orden.

La herramienta principal será el variograma. Se distinguirán dos definiciones principales: 1) el variograma de funciones aleatorias, o variograma teórico, cuyo conocimiento es requerido para resolver el problema de interpolación, y 2) el variograma experimental o muestral, el cual es calculado a partir de los datos. El proceso será dividido en dos fases: 1) calcular el variograma experimental que mejor aproxime a los datos y 2) ajustar el modelo teórico a este variograma experimental.

4.2.1. Variograma Experimental

Se iniciará con los aspectos empíricos del análisis variográfico. Antes de proceder al cálculo del variograma experimental es necesario el realizar un análisis exploratorio de los datos (Chilès & Delfiner [3], Isaaks & Srivastava [11]), que comprende el cálculo de estadísticos usuales univariados y bivariados como por ejemplo, histogramas, diagramas de dispersión, entre otros, así como técnicas multivariadas como el análisis de conglomerados o análisis de componentes principales que pueden ser usados para reducir problemas multivariados a problemas univariados. Esto debe realizarse debido a que las herramientas geoestadísticas no reemplazan si no complementan las herramientas estadísticas clásicas, las cuales permiten detectar las inconsistencias existentes en los datos causantes de una mala interpretación del fenómeno de estudio y consecuentemente de una solución que no lo describa correctamente.

Se omitirá el estudio de estas técnicas estadísticas, ya que no son vitales en el desarrollo, pero éstas pueden ser consultadas por ejemplo, en Isaaks & Srivastava [11], donde son descritas con detalle sobre un caso muestra.

El variograma experimental $\hat{\gamma}(\mathbf{h})$, es un promedio de disimilitudes de los datos. Puede ser formado por una clase dada de vectores \mathcal{H} agrupando todos los $N_{\mathcal{H}}$ pares de puntos que pueden ser ligados o están separados por un vector \mathbf{h} perteneciente a \mathcal{H} . Tal clase \mathcal{H} agrupa vectores cuyas longitudes estén entre un intervalo especificado de longitudes y cuya orientación es la misma hasta una tolerancia dada sobre el ángulo θ que define su dirección con respecto a un eje de referencia. La disimilitud promedio con respecto a una clase de vectores \mathcal{H}_k es un valor del variograma experimental que se calcula mediante la siguiente fórmula

$$\hat{\gamma}(\mathcal{H}_k) = \frac{1}{2N_{\mathcal{H}_k}} \sum_{i=1}^{N_{\mathcal{H}_k}} (z(x_j) - z(x_i))^2 \quad \text{con } \mathbf{x}_j - \mathbf{x}_i = \mathbf{h} \in \mathcal{H}_k. \quad (4.3)$$

El variograma experimental es reemplazado por un variograma teórico esencialmente por la razón de que el modelo del variograma debe tener un significado físico. El reemplazo se realiza a través del ajuste de un modelo admisible de variograma teórico (esto porque el variograma teórico debe cumplir ciertas propiedades), donde el ajuste puede ser realizado de varias formas aunque lo más usual es un ajuste en el sentido de mínimos cuadrados (véase Chilès & Delfiner [3] para detalles sobre el ajuste).

4.2.2. Propiedades de la función de covarianza y del variograma teórico

Se consideran dos clases de funciones aleatorias; funciones aleatorias estacionarias y funciones aleatorias intrínsecas.

Covarianza de una función aleatoria estacionaria

Como se ha visto, una variable aleatoria estacionaria $Z(\mathbf{x})$ es caracterizada por su media

$$m = E[Z(\mathbf{x})],$$

y su función de covarianza

$$C(\mathbf{h}) = E[(Z(\mathbf{x}) - m)(Z(\mathbf{x} + \mathbf{h}) - m)].$$

La covarianza muestra como la correlación entre las variables evoluciona con respecto a la separación o intervalo \mathbf{h} . Esta función, por lo tanto, depende tanto de su longitud, la cual es la distancia entre \mathbf{x} y $\mathbf{x} + \mathbf{h}$ y su dirección. Cuando la covarianza depende solo de la distancia, se dice que es isótropa. La covarianza es una función par, y por la desigualdad de Schwarz está acotada por su valor en el origen (*i.e.* la varianza de la función aleatoria estacionaria), esto es

$$C(\mathbf{h}) = C(-\mathbf{h}), \quad |C(\mathbf{h})| \leq C(\mathbf{0}).$$

Más adelante, se mostrará que la covarianza es una función definida positiva. Ya que la función aleatoria tiene una varianza finita, está fluctúa alrededor de su media. Algunos fenómenos no muestran este comportamiento, ya que si se calcula la media y la varianza muestral sobre dominios cada vez más grandes, la media no se estabiliza, y la varianza muestral siempre crece. Esto motiva el siguiente modelo.

Variograma de una función aleatoria intrínseca

Una función aleatoria intrínseca es una función aleatoria cuyos incrementos son estacionarios de segundo orden. El variograma muestra como la disimilitud entre $Z(\mathbf{x})$ y $Z(\mathbf{x} + \mathbf{h})$ evoluciona con la separación \mathbf{h} . Al igual que la covarianza, es en general anisótropa. Obviamente el variograma es una función par no negativa valuada 0 en $\mathbf{h} = \mathbf{0}$, esto es

$$\gamma(\mathbf{h}) = \gamma(-\mathbf{h}), \quad \gamma(\mathbf{h}) \geq 0, \quad \gamma(\mathbf{0}) = 0.$$

Además, un variograma no puede ser cualquier función arbitraria, por lo cual, se mostrará más adelante que $-\gamma(\mathbf{h})$ debe ser una función condicionalmente definida positiva.

Dadas las dos clases de funciones aleatorias, se analizará porque el uso del variograma se preferencia sobre la función de covarianza. Una función aleatoria estacionaria es obviamente una función aleatoria intrínseca y por lo tanto tiene un variograma, implicando que el variograma tiene un grado de generalidad mayor que la función de covarianza. En el caso estacionario, el variograma esta ligado a la covarianza por la relación

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}),$$

por lo que el variograma de una función aleatoria estacionaria está acotado por $2C(\mathbf{0})$.

En síntesis, se puede enunciar dos razones para utilizar el variograma sobre la covarianza:

1. Ya que la clase de funciones aleatorias intrínsecas incluye a las funciones aleatorias estacionarias, el variograma es una herramienta más general que la covarianza.
2. La segunda razón es práctica esto porque el variograma, a diferencia de la covarianza, no requiere el conocimiento de la media, que por lo general, no es conocida y en tales condiciones tiene que ser estimada a partir de los datos, lo cual introduce un sesgo que produce interpretaciones erróneas del fenómeno en estudio.

Funciones definidas positivas y condicionalmente definidas positivas

La condición verdaderamente necesaria para que una función sea una covarianza o un variograma es que todos los cálculos de la varianza den un resultado no negativo.

En el caso estacionario, considérese cualquier combinación lineal $\sum_{i=0}^N \lambda_i Z(\mathbf{x}_i)$ de N términos. Es fácil ver que la esperanza de la combinación lineal es

$$E \left[\sum_{i=0}^N \lambda_i Z(\mathbf{x}_i) \right] = \sum_{i=0}^N \lambda_i m.$$

Su varianza, necesariamente positiva o cero, puede ser expresada con la covarianza como

$$\begin{aligned} \text{Var} \left(\sum_{i=0}^N \lambda_i Z(\mathbf{x}_i) \right) &= E \left[\left(\sum_{i=0}^N \lambda_i Z(\mathbf{x}_i) - \sum_{i=0}^N \lambda_i m \right)^2 \right] \\ &= E \left[\left(\sum_{i=0}^N \lambda_i (Z(\mathbf{x}_i) - m) \right)^2 \right] \\ &= E \left[\sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j (Z(\mathbf{x}_i) - m)(Z(\mathbf{x}_j) - m) \right] \\ &= \sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j E [(Z(\mathbf{x}_i) - m)(Z(\mathbf{x}_j) - m)] \\ &= \sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j C(\mathbf{x}_i - \mathbf{x}_j) \geq 0, \end{aligned}$$

por lo tanto, la función de covarianza $C(\mathbf{h})$ es una función definida positiva (refiérase al teorema 3.4 para más detalles sobre funciones definidas positivas).

En el caso estacionario cualquier combinación lineal finita tiene varianza finita. Esto no se cumple para el caso en que las funciones aleatorias tienen variogramas pero no covarianzas. En el caso intrínseco las únicas combinaciones lineales para las cuales se pueden calcular la varianza son combinaciones lineales de incrementos. Estas son llamadas combinaciones lineales admisibles y son caracterizadas por la condición

$$\sum_{i=0}^N \lambda_i = 0. \quad (4.4)$$

Es claro que todas las combinaciones lineales de incrementos satisfacen la ecuación (4.4), ya que cada incremento lo satisface. Recíprocamente, cualquier combinación $\sum_{i=0}^N \lambda_i Z(\mathbf{x}_i)$ satisfaciendo la ecuación (4.4), es igual a $\sum_{i=0}^N \lambda_i (Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}}))$ para cualquier elección del punto $\hat{\mathbf{x}}$, que es una combinación lineal de los incrementos.

Se trata ahora de expresar la varianza de una combinación lineal admisible en términos del variograma. Sea $\hat{\mathbf{x}} \in \mathcal{D}$, dado que se usa una combinación lineal admisible, esto es, que cumple la ecuación (4.4), se tiene que

$$\begin{aligned}
\text{Var} \left(\sum_{i=0}^N \lambda_i Z(\mathbf{x}_i) \right) &= \text{Var} \left(\sum_{i=0}^N \lambda_i Z(\mathbf{x}_i) - \sum_{i=0}^N \lambda_i Z(\hat{\mathbf{x}}) \right) \\
&= \text{Var} \left(\sum_{i=0}^N \lambda_i (Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}})) \right) \\
&= \text{E} \left[\left(\sum_{i=0}^N \lambda_i (Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}})) \right)^2 \right] \\
&= \text{E} \left[\sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j (Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}}))(Z(\mathbf{x}_j) - Z(\hat{\mathbf{x}})) \right] \\
&= \sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j \text{E} [(Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}}))(Z(\mathbf{x}_j) - Z(\hat{\mathbf{x}}))]
\end{aligned}$$

Por otro lado, se tiene que,

$$\begin{aligned}
(Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2 &= [(Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}})) - (Z(\mathbf{x}_j) - Z(\hat{\mathbf{x}}))]^2 \\
&= (Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}}))^2 - 2(Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}}))(Z(\mathbf{x}_j) - Z(\hat{\mathbf{x}})) + (Z(\mathbf{x}_j) - Z(\hat{\mathbf{x}}))^2,
\end{aligned}$$

y aplicando la esperanza a esta última expresión,

$$\begin{aligned}
\text{E} [(Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2] &= \text{E} [(Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}}))^2] \\
&\quad - 2\text{E} [(Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}}))(Z(\mathbf{x}_j) - Z(\hat{\mathbf{x}}))] + \text{E} [(Z(\mathbf{x}_j) - Z(\hat{\mathbf{x}}))^2].
\end{aligned}$$

Dado que se está suponiendo que la función aleatoria es intrínseca, se tiene que

$$2\gamma(\mathbf{x}_i - \hat{\mathbf{x}}) = \text{E} [(Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}}))^2] \quad \text{y} \quad 2\gamma(\mathbf{x}_i - \mathbf{x}_j) = \text{E} [(Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2],$$

por lo cual

$$\text{E} [(Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}}))(Z(\mathbf{x}_j) - Z(\hat{\mathbf{x}}))] = \gamma(\mathbf{x}_i - \hat{\mathbf{x}}) + \gamma(\mathbf{x}_j - \hat{\mathbf{x}}) - \gamma(\mathbf{x}_i - \mathbf{x}_j).$$

Utilizando esto, finalmente se ve que

$$\begin{aligned}
 \text{Var} \left(\sum_{i=0}^N \lambda_i Z(\mathbf{x}_i) \right) &= \sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j \mathbb{E} [(Z(\mathbf{x}_i) - Z(\hat{\mathbf{x}}))(Z(\mathbf{x}_j) - Z(\hat{\mathbf{x}}))] \\
 &= \sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j (\gamma(\mathbf{x}_i - \hat{\mathbf{x}}) + \gamma(\mathbf{x}_j - \hat{\mathbf{x}}) - \gamma(\mathbf{x}_i - \mathbf{x}_j)) \\
 &= \sum_{j=0}^N \lambda_j \left(\sum_{i=0}^N \lambda_i \gamma(\mathbf{x}_i - \hat{\mathbf{x}}) \right) \\
 &\quad + \sum_{j=0}^N \lambda_i \left(\sum_{j=0}^N \lambda_j \gamma(\mathbf{x}_j - \hat{\mathbf{x}}) \right) - \sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \\
 &= - \sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \geq 0,
 \end{aligned}$$

es decir, $-\gamma(\mathbf{h})$ es una función condicionalmente definida positiva (véase el teorema 3.11 para más detalles sobre funciones condicionalmente definidas positivas).

En resumen, el variograma teórico que está dada por,

$$\gamma(\mathbf{h}) = \frac{1}{2} \mathbb{E} [(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}))^2],$$

cuenta con las siguientes propiedades:

- El valor del variograma en el origen es cero por definición, $\gamma(\mathbf{0}) = 0$.
- Los valores del variograma son positivos $\gamma(\mathbf{h}) \geq 0$, y es una función par, $\gamma(-\mathbf{h}) = \gamma(\mathbf{h})$.
- El variograma crece más lento que $|\mathbf{h}|^2$, i.e., $\lim_{|\mathbf{h}| \rightarrow \infty} \frac{\gamma(\mathbf{h})}{|\mathbf{h}|^2} = 0$ (Chilès & Delfiner [3]).
- $-\gamma(\mathbf{h})$ es una función condicionalmente definida positiva. Esto significa que el uso del variograma $\gamma(\mathbf{h})$ para el cálculo de la varianza de una combinación lineal de $N + 1$ variables aleatorias estacionarias $Z(\mathbf{x}_i)$ debe ser positivo bajo la restricción de que la suma de los pesos sea cero, es decir

$$\text{Var} \left(\sum_{i=0}^N \lambda_i Z(\mathbf{x}_i) \right) = - \sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \geq 0, \quad \text{sí } \sum_{i=0}^N \lambda_i = 0,$$

por lo tanto, una matriz Γ con entradas $\Gamma_{i,j} = -\gamma(\mathbf{x}_i - \mathbf{x}_j)$ es condicionalmente definida positiva, i.e., $\boldsymbol{\lambda}^\top \Gamma \boldsymbol{\lambda} \geq 0$ para $\sum_{i=0}^N \lambda_i = 0$. Esto es necesario para garantizar la solución única del sistema de ecuaciones en kriging, como se verá más adelante.

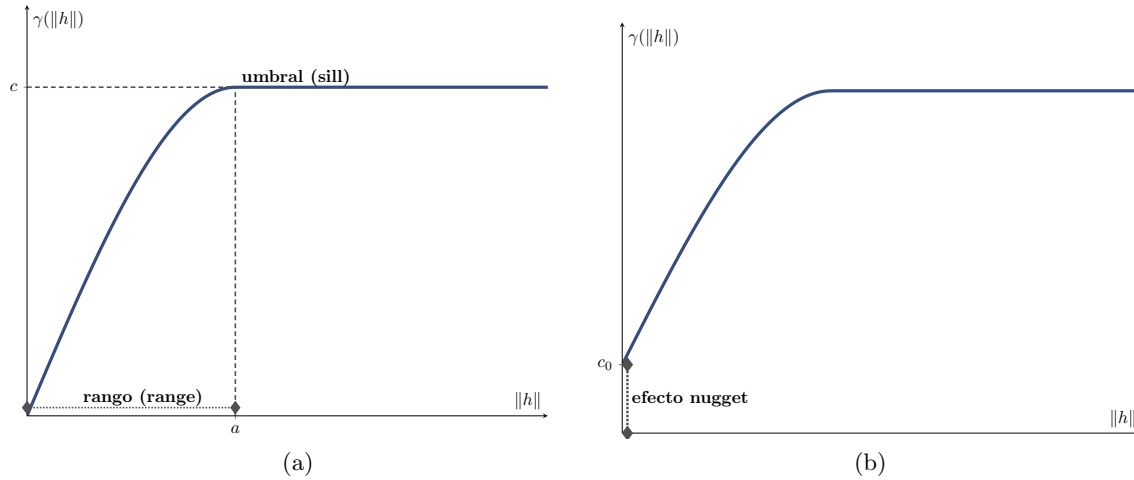


Figura 4.1: (a) Forma típica de un variograma que tiene un valor asintótico llamado su umbral a una distancia denominada como su rango. (b) Variograma con efecto nugget

El variograma presenta las siguientes características:

- Siempre empieza en $\mathbf{0}$ (para $\mathbf{h} = \mathbf{0}$, $Z(\mathbf{x} + \mathbf{h}) = Z(\mathbf{x})$). Este puede ser discontinuo justo después del origen, a lo que se denomina *efecto nugget* (figura 4.1b).
- Generalmente se incrementa con \mathbf{h} .
- Este crece hasta cierto nivel llamada “umbral” (*sill*) y después muestra un comportamiento asintótico a este valor. Alternativamente este puede seguir creciendo.

En la figura 4.1 se observa la forma típica de un variograma mostrando las características antes enunciadas. La tasa de crecimiento del variograma con la distancia indica cuan rápido la influencia de la muestra decae con la distancia. Después de que el variograma ha alcanzado su valor límite (su umbral) ya no hay más correlación entre las muestras. Esta distancia crítica, llamada el “rango” (*range*), da una definición más precisa de la noción de “zona de influencia” (figura 4.1a).

Lo anteriormente mencionado muestra el comportamiento del variograma para valores grandes de \mathbf{h} , pero es aún mas importante el analizar que pasa para valores pequeños de \mathbf{h} porque está relacionado a la continuidad y la regularidad espacial del fenómeno bajo estudio. Se consideran cuatro tipos diferentes de comportamiento cerca del origen (Armstrong [1], Chilès & Delfiner [3]).

1. **Cuadrático.** Esto indica que la variable regionalizada es continua y diferenciable.
2. **Lineal.** La variable regionalizada es continua pero no diferenciable.

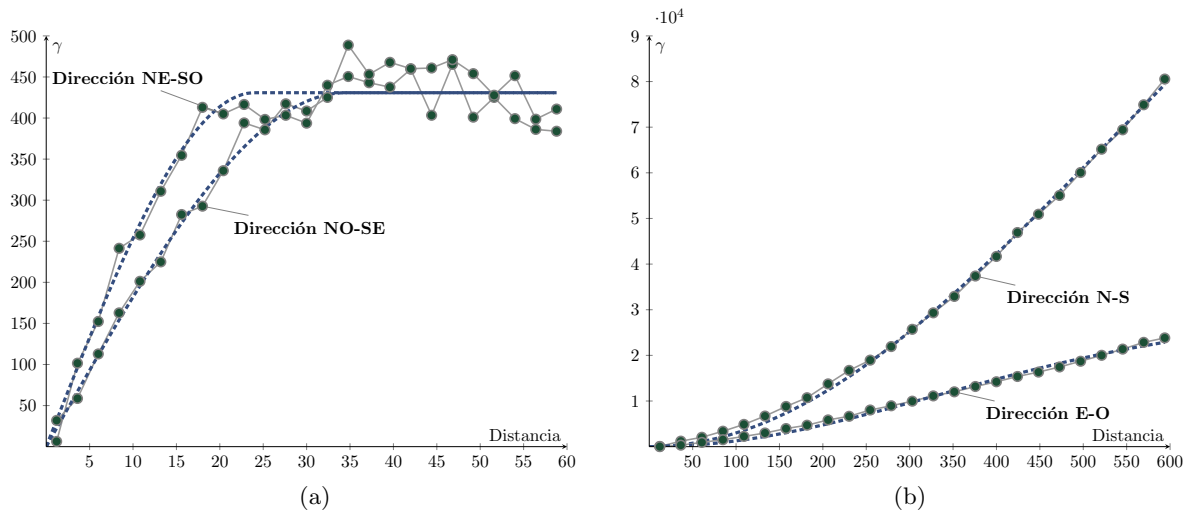


Figura 4.2: Ejemplos de anisotropía. (a) Anisotropía geométrica. (b) Anisotropía zonal.

3. **Discontinua en el origen.** Esto significa que la variable es altamente irregular en distancias cortas.
4. **Plana.** Las variables regionalizadas $z(\mathbf{x} + \mathbf{h})$ y $z(\mathbf{x})$ son no correlacionadas para todos los valores de \mathbf{h} no importando que tan cerca se encuentren.

Anisotropía

Otra característica importante ocurre cuando el variograma es calculado en varias direcciones, ya que su comportamiento observado en estas puede diferir significativamente. En tal situación, el variograma depende tanto de la dirección como de la distancia entre las posiciones, es decir, el variograma es anisótropo. Cuando el variograma solamente depende de la distancia se dice que es isótropo.

Dos casos usuales de anisotropía son los siguientes:

1. La anisotropía geométrica sucede cuando para variogramas calculados en diferentes direcciones se obtienen diferentes rangos y ambos alcanzan el mismo umbral. Un ejemplo se muestra en la figura 4.2a donde los variogramas son calculados en las direcciones noreste-suroeste (NE-SO) y noroeste-sureste (NO-SE).
2. La anisotropía zonal esta relacionada con la diferencia entre los umbrales de los variogramas calculados para diferentes direcciones. La figura 4.2b muestra la anisotropía zonal para otro conjunto de datos donde los variogramas son calculados en las direcciones norte-sur (N-S) y este-oeste (E-O).

Modelos Admisibles

Los modelos admisibles de variogramas son modelos que cumplen con las propiedades y características antes mencionadas del variograma teórico y estos pueden ser diferenciados por presentar alguno de los cuatro comportamientos cerca del origen antes enunciados.

Los más comúnmente usados, donde c denota el umbral y a el rango, son los siguientes (para otros modelos véase Armstrong [1], Isaaks & Srivastava [11], Wackernagel [21]):

- El modelo esférico es probablemente el modelo más comúnmente usado el cual tiene un comportamiento lineal en el origen.

$$\gamma(\mathbf{h}) = \begin{cases} c \left(\frac{3 \|\mathbf{h}\|}{2a} - \frac{1 \|\mathbf{h}\|^3}{2a^3} \right) & \|\mathbf{h}\| < a \\ c & \|\mathbf{h}\| \geq a \end{cases} .$$

- El modelo exponencial, al igual que el anterior, en el origen se comporta de manera lineal.

$$\gamma(\mathbf{h}) = c \left(1 - \exp \left(-\frac{\|\mathbf{h}\|}{a} \right) \right) .$$

- El modelo gaussiano en el origen se comporta de manera cuadrática, por lo cual representa un fenómeno extremadamente continuo.

$$\gamma(\mathbf{h}) = c \left(1 - \exp \left(-\frac{\|\mathbf{h}\|^2}{a^2} \right) \right) .$$

- El modelo de efecto nugget es un modelo que se utiliza junto con alguno de los otros modelos para cuando el fenómeno presenta algunas discontinuidades.

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \|\mathbf{h}\| = 0 \\ c_0 & \|\mathbf{h}\| > 0 \end{cases} .$$

- Los modelos de funciones potencia no presentan el comportamiento asintótico de los modelos anteriores, es decir, este no alcanza un umbral y esta dado por

$$\gamma(\mathbf{h}) = \hat{m} \|\mathbf{h}\|^\alpha \quad \text{con } 0 < \alpha < 2,$$

donde el parámetro \hat{m} denota la pendiente. Un caso especial que se utilizará frecuentemente es el modelo lineal $\gamma(\mathbf{h}) = \hat{m} \|\mathbf{h}\|$ (cuando $\alpha = 1$).

Se puede probar que una combinación lineal de modelos admisibles sigue siendo un modelo admisible (Chilès & Delfiner [3]). Lo más usual es el utilizar el modelo de efecto nugget junto con alguno de los otros modelos. De esta manera, si c_0 es el efecto nugget y lo combinamos con cualquiera de los otro modelos, $c + c_0$ es el umbral, donde c es el umbral para el modelo sin efecto nugget y a el rango para el modelo combinado.

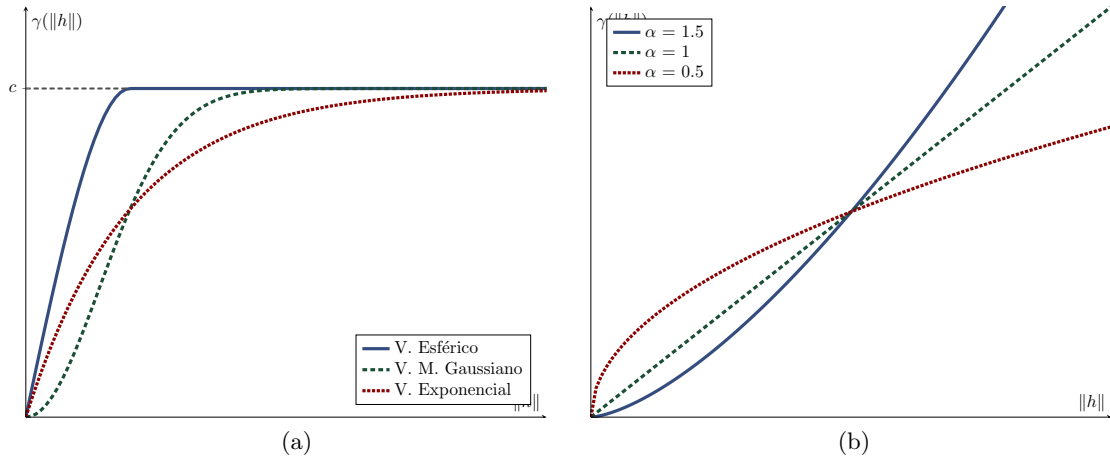


Figura 4.3: Gráfica de modelos admisibles. (a) Modelos acotados. (b) Tres modelos de funciones potencia con exponentes $\alpha = 0.5, 1$ y 1.5 .

La figura 4.3 muestra una gráfica de algunos modelos. Estos modelos son validos para el caso isótropo, por lo tanto, cuando en el análisis del variograma se encuentra algún tipo de anisotropía se tiene que construir un variograma omnidireccional que describa lo mejor posible la variabilidad de todas las direcciones de los datos en la región. Otra forma de tratar la anisotropía es el utilizar algunas transformaciones para volver al caso isótropo y describir esta en términos de los modelos vistos.

Por ejemplo, cuando se presenta anisotropía geométrica en dos dimensiones, al calcular el variograma en diferentes direcciones, se puede tomar el rango como función de la dirección, *i.e.* $a(\theta)$, donde θ denota la dirección y es el ángulo medido a partir del eje de las abscisas. La curva descrita por la función $a(\theta)$ es una elipse cuyo semieje mayor corresponde a la dirección con el rango más grande y el semieje menor es la dirección perpendicular a esta (véase figura 4.4a), es por ello que la anisotropía geométrica también se conoce como elíptica.

Para tratar la anisotropía geométrica se utiliza una transformación en dos pasos que la lleva a condiciones isótropas. El primer paso es alinear los ejes coordenados (x, y) con los ejes de la elipse de anisotropía. Específicamente, si $\hat{\theta}$ es la dirección donde se alcanza el mayor rango, se rota un ángulo $\hat{\theta}$ los ejes coordenados cartesianos (x, y) de manera tal que el eje de las ordenadas coincida con el eje menor de la elipse. Esto se logra mediante la matriz de rotación

$$R = \begin{pmatrix} \cos \hat{\theta} & \sin \hat{\theta} \\ -\sin \hat{\theta} & \cos \hat{\theta} \end{pmatrix},$$

para dar lugar al sistema de coordenadas (x', y') (figura 4.4b). El segundo paso consiste en elongar el eje menor de tal forma que su longitud sea igual a la del eje mayor de la elipse

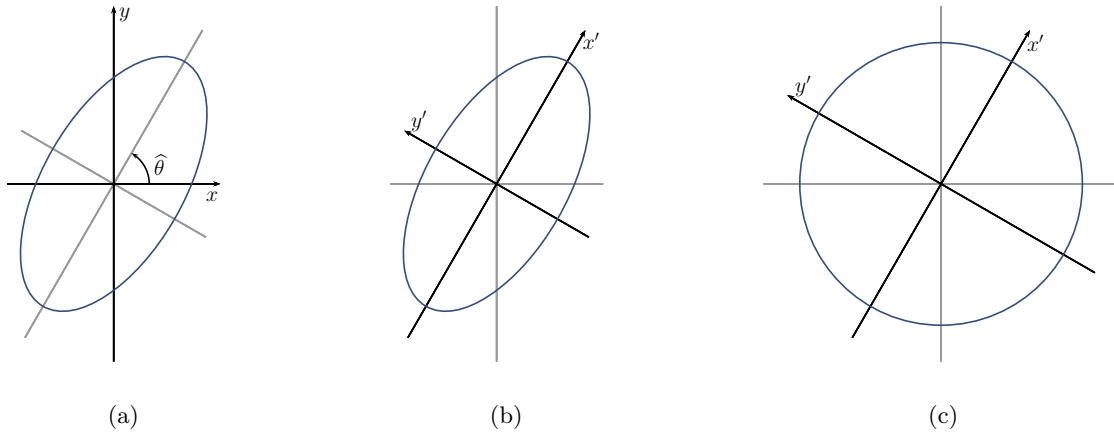


Figura 4.4: (a) Elipse de anisotropía. (b) Rotación de los ejes cartesianos a los ejes de la elipse. (c) Elongación del eje menor para volver a las condiciones isotrópicas

(figura 4.4c). Para ello se utiliza la matriz

$$E = \begin{pmatrix} 1 & 0 \\ 0 & 1/r_a \end{pmatrix},$$

donde $r_a = a_{\text{mín}}/a_{\text{máx}}$ es la razón de anisotropía, $a_{\text{máx}}$ y $a_{\text{mín}}$ son el rango más grande y más pequeño respectivamente. Así, el variograma anisótropo que era función de \mathbf{h} y θ es transformado a una función equivalente en condiciones isotrópicas, esto es

$$\gamma_a(\mathbf{h}) = \gamma(\Lambda\mathbf{h})$$

con $\Lambda = ER$, γ_a denota el variograma anisótropo y γ es alguno de los modelos admisible de variogramas isotrópicos.

4.3. Kriging

Un problema central en geoestadística es la reconstrucción de un fenómeno en un dominio sobre la base de valores observados en un número limitado de puntos. Matemáticamente, este problema puede ser considerado como un problema de interpolación. En el enfoque clásico, la función desconocida es aproximada por una función paramétrica cuya forma es postulada de antemano, tanto de forma implícita o explícita. Los parámetros son seleccionados de tal manera que optimicen algún criterio del mejor ajuste a los puntos de los datos, el cual puede ser de forma estadística o determinista. Una vez que la función de aproximación es determinada, solo basta con evaluarla donde quiera que sea necesario.

Tipo de kriging	Media	Prerequisito mínimo	Nombre del modelo
Kriging Simple (KS)	Constante, conocida	Covarianza	Estacionario
Kriging Ordinario (KO)	Constante, desconocida	Variograma	Intrínseco
Kriging Universal (KU)	Variante, desconocido	Variograma	Modelo KU

Tabla 4.1: Formas principales de kriging lineal

Un enfoque diferente para el desarrollo de este problema es conocido como kriging. La diferencia central es que kriging comienza como un modelo estadístico de la naturaleza del fenómeno en lugar de un modelo de la función de interpolación. Es por esto que el término estimación entra en este contexto, ya que este se utiliza para referirnos a la inferencia de un parámetro fijo pero desconocido del modelo.

Ahora, se centrará en resolver el problema de estimación al usar estimadores lineales. La teoría se desarrolla dentro del ámbito de un modelo de estacionariedad de segundo orden que solo envuelve la media $m(\mathbf{x})$ y la función de covarianza $\sigma(\mathbf{x}, \mathbf{y})$ o el variograma $\gamma(\mathbf{x}, \mathbf{y})$, que se suponen son conocidos. La tabla 4.1 resume las tres formas principales de kriging que trataremos y bajo que modelos se encuentran.

4.3.1. Kriging Simple

El primer estimador que se presenta es kriging simple (KS), el cual puede ser utilizado cuando la media de la función aleatoria es conocida, un caso poco común en la práctica. De acuerdo con Chilès & Delfiner [3], en la vida real, la media solamente puede ser conocida si hay repeticiones del fenómeno, como en los procesos espacio-temporales o cuando el número de datos se vuelve tan grande como para estimar la media casi a la perfección.

El problema que se quiere resolver es el estimar $Z(\mathbf{x}_0)$ en una nueva posición \mathbf{x}_0 a partir de las N observaciones $Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_N)$, usando el estimador

$$Z^*(\mathbf{x}) = \sum_{i=1}^N \lambda_i Z(\mathbf{x}_i) + \alpha,$$

donde los pesos de la combinación lineal y la constante α dependen de \mathbf{x} , esto es, $\lambda_i(\mathbf{x})$ ($i = 1, \dots, N$) y $\alpha(\mathbf{x})$. La constante α y los pesos λ_i son elegidos de tal manera que se minimice el error $Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)$, caracterizado por el error cuadrático medio $E[(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))^2]$, que puede ser escrito como

$$E[(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))^2] = \text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)) + (E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)])^2.$$

Si el estimador Z^* es insesgado, *i.e.* $E[Z^*(\mathbf{x}) - Z(\mathbf{x})] = 0$, dado que la varianza siempre es no negativa, el error cuadrático medio se minimiza al eliminar el sesgo, ya que de esta manera, este solo depende de la varianza. Por lo tanto, si queremos que $E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = 0$ esto implica

$$\begin{aligned} E[Z^*(\mathbf{x}_0)] - E[Z(\mathbf{x}_0)] &= 0 \\ \sum_{i=1}^N \lambda_i m(\mathbf{x}_i) + \alpha - m(\mathbf{x}_0) &= 0 \\ \alpha &= m(\mathbf{x}_0) - \sum_{i=1}^N \lambda_i m(\mathbf{x}_i), \end{aligned}$$

así, el estimador se reescribe como,

$$Z^*(\mathbf{x}_0) = m(\mathbf{x}_0) + \sum_{i=1}^N \lambda_i (Z(\mathbf{x}_i) - m(\mathbf{x}_i)).$$

Si se toma $\lambda_0 = -1$ se tiene que

$$Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0) = \sum_{i=1}^N \lambda_i (Z(\mathbf{x}_i) - m(\mathbf{x}_i)),$$

consecuentemente, el error cuadrático medio (que ahora coincide con la varianza) puede ser expresado en términos de la covarianza ($\sigma(\mathbf{x}, \mathbf{y})$)

$$\begin{aligned} E[(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))^2] &= E \left[\sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j (Z(\mathbf{x}_i) - m(\mathbf{x}_i)) (Z(\mathbf{x}_j) - m(\mathbf{x}_j)) \right] \\ &= \sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j E[(Z(\mathbf{x}_i) - m(\mathbf{x}_i)) (Z(\mathbf{x}_j) - m(\mathbf{x}_j))] \\ &= \sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j \sigma(\mathbf{x}_i, \mathbf{x}_j) \\ &= \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \sigma(\mathbf{x}_i, \mathbf{x}_j) - 2 \sum_{i=1}^N \lambda_i \sigma(\mathbf{x}_i, \mathbf{x}_0) + \sigma(\mathbf{x}_0, \mathbf{x}_0). \end{aligned}$$

Para obtener el mínimo de esta función cuadrática, se deriva parcialmente con respecto a los pesos y se iguala a cero, esto es

$$\frac{\partial}{\partial \lambda_i} E[(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))^2] = 2 \sum_{j=1}^N \lambda_j \sigma(\mathbf{x}_i, \mathbf{x}_j) - 2 \sigma(\mathbf{x}_i, \mathbf{x}_0) = 0,$$

para cada $i = 1, \dots, N$, lo cual da lugar al sistema de ecuaciones

$$\sum_{j=1}^N \lambda_j \sigma(\mathbf{x}_i, \mathbf{x}_j) = 2\sigma(\mathbf{x}_i, \mathbf{x}_0), \quad i = 1, \dots, N,$$

que se conoce como *sistema kriging simple*. En forma matricial puede ser reescrito como

$$\Sigma \boldsymbol{\lambda} = \boldsymbol{\sigma}_0$$

donde Σ es la matriz de covarianzas de tamaño $N \times N$, con entradas $\Sigma_{ij} = \sigma(\mathbf{x}_i, \mathbf{x}_j)$, $\boldsymbol{\sigma}_0$ un vector de tamaño N , donde la entrada i -ésima denota la covarianza entre $Z(\mathbf{x}_i)$ y $Z(\mathbf{x}_0)$ y $\boldsymbol{\lambda}$ es un vector de tamaño N con los pesos. El sistema de ecuaciones anterior tiene solución única si la matriz Σ es no singular. Este siempre es el caso si la función de covarianza $\sigma(\mathbf{x}, \mathbf{y})$ es estrictamente positiva definida y si todos los puntos de los datos son distintos.

La varianza de la estimación σ_{KS}^2 asociada con Z^* , llamada la varianza kriging, se obtiene al substituir la solución del sistema kriging simple en el error cuadrático medio. Esto es, si se considera que el sistema tiene solución única, entonces $\boldsymbol{\lambda} = \Sigma^{-1} \boldsymbol{\sigma}_0$ y si se reescribe el error cuadrático medio en forma matricial

$$(E[(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))])^2 = \text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)) = \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{\sigma}_0 + \sigma(\mathbf{x}_0, \mathbf{x}_0)$$

al substituir $\boldsymbol{\lambda}$ se tiene que

$$\begin{aligned} \sigma_{SK}^2 &= \text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)) \\ &= \boldsymbol{\lambda}^\top \Sigma \Sigma^{-1} \boldsymbol{\sigma}_0 - 2\boldsymbol{\lambda}^\top \boldsymbol{\sigma}_0 + \sigma(\mathbf{x}_0, \mathbf{x}_0) \\ &= \sigma(\mathbf{x}_0, \mathbf{x}_0) - \boldsymbol{\lambda}^\top \boldsymbol{\sigma}_0 \\ &= \sigma(\mathbf{x}_0, \mathbf{x}_0) - \sum_{i=1}^N \lambda_i \sigma(\mathbf{x}_i, \mathbf{x}_0). \end{aligned}$$

La varianza kriging, o mejor dicho su raíz cuadrada, la desviación estándar kriging, da una medida del error asociado con el estimador kriging.

Kriging simple es un interpolador exacto en el sentido que si \mathbf{x}_0 es idéntico con una posición de los datos, entonces el valor estimado es idéntico con el valor del dato en ese punto, *i.e.*

$$Z^*(\mathbf{x}_0) = Z(\mathbf{x}_i), \quad \text{si } \mathbf{x}_0 = \mathbf{x}_i.$$

Esto puede ser visto fácilmente. Cuando \mathbf{x}_0 es una de las posiciones \mathbf{x}_i , el lado derecho del sistema kriging simple es igual a una columna de la matriz del lado izquierdo. Un vector de pesos $\boldsymbol{\lambda}$ con un peso para esa columna igual a uno y todos los otros pesos igual a cero es la solución del sistema. Como la matriz del lado izquierdo es no singular, la solución es única.

4.3.2. Kriging Ordinario

En la mayoría de las situaciones prácticas la media $m(\mathbf{x})$ no es conocida. Bajo una suposición paramétrica sobre la función de la media, kriging da una solución óptima que envuelve solo un paso de la estimación. El caso más simple es cuando la media es constante $m(\mathbf{x}) = \mu$, que lleva al método conocido como kriging ordinario. Este fue desarrollado por Matheron el siglo pasado a principios de los sesentas y juega un rol especial ya que es compatible con un modelo estacionario, solo usa el variograma, y es de hecho la forma de kriging más usada.

Sea nuevamente el problema de estimar el valor desconocido $Z(\mathbf{x}_0)$ en una nueva posición \mathbf{x}_0 usando los valores de los datos $Z(\mathbf{x}_i)$ y tomándose nuevamente el estimador lineal

$$Z^*(\mathbf{x}) = \sum_{i=1}^N \lambda_i Z(\mathbf{x}_i) + \alpha.$$

De nueva cuenta, si se considera el error cuadrático medio y se supone que la media desconocida es constante, este se puede escribir como

$$\begin{aligned} E[(Z^*(\mathbf{x}_0) - Z(\mathbf{x}))^2] &= \text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x})) + (E[Z^*(\mathbf{x}_0) - Z(\mathbf{x})])^2 \\ &= \text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x})) + \left(\sum_{i=1}^N \lambda_i \mu + \alpha - \mu \right)^2 \\ &= \text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x})) + \left[\left(\sum_{i=1}^N \lambda_i - 1 \right) \mu + \alpha \right]^2. \end{aligned}$$

Solo el sesgo en el lado derecho de la última ecuación involucra a α , pero esta vez no se puede minimizar sin el conocimiento de μ . Una solución intuitiva sería reemplazar μ por un estimador $\hat{\mu}$ y resolver para α , pero este estimador necesariamente tendría que depender de los datos haciendo que α ya no fuera constante. La única solución que queda es hacer $\alpha = 0$ y si además se quiere que el sesgo sea cero, *i.e.* $E[Z^*(\mathbf{x}_0) - Z(\mathbf{x})] = 0$, esto implica que

$$\alpha = \left(\sum_{i=1}^N \lambda_i - 1 \right) \mu = 0,$$

imponiendo la condición que los pesos sumen 1, esto es $\sum_i \lambda_i = 1$, así el sesgo es cero cualquiera que sea la media constante desconocida μ .

La necesidad de hacer $\alpha = 0$ e imponer las condiciones sobre los pesos λ_i para eliminar el sesgo se hace en general cuando la media $m(\mathbf{x})$ no es conocida. De esta manera, solo se utilizan estimadores de la forma

$$Z^*(\mathbf{x}) = \sum_{i=1}^N \lambda_i Z(\mathbf{x}_i), \quad \text{sujeto a } \sum_{i=1}^N \lambda_i = 1. \quad (4.5)$$

Ahora, se formula el problema que se quiere resolver de la siguiente manera. Se desea estimar el valor desconocido $Z(\mathbf{x}_0)$ en una nueva posición \mathbf{x}_0 a partir de las N observaciones $\{Z(\mathbf{x}_i)\}$, tomadas en las posiciones $\{\mathbf{x}_i\}$, usando el estimador $Z^*(\mathbf{x}_0)$ definido en (4.5). Para este caso, se supone que los datos son parte de una realización de una función aleatoria intrínseca $Z(\mathbf{x})$ (*i.e.* cumplen las hipótesis intrínsecas) con variograma $\gamma(\mathbf{h})$, el cual se supone conocido.

Se busca que el estimador $Z^*(\mathbf{x}_0)$ sea insesgado y de mínima varianza. La propiedad de que el estimador sea insesgado es garantizada con la suma unitaria de los pesos, esto es

$$E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = E\left[\sum_{i=1}^N \lambda_i Z(\mathbf{x}_i) - Z(\mathbf{x}_0) \cdot \sum_{i=1}^N \lambda_i\right] = \sum_{i=1}^N \lambda_i (m(\mathbf{x}_i) - m(\mathbf{x}_0)) = 0,$$

ya que, $m(\mathbf{x}_i) - m(\mathbf{x}_0) = \mu - \mu = 0$, al ser la media constante.

La varianza de la estimación $\sigma_{KO}^2 = \text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))$ es la varianza de la combinación lineal

$$Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0) = \sum_{i=1}^N \lambda_i Z(\mathbf{x}_i) - 1 \cdot Z(\mathbf{x}_0) = \sum_{i=0}^N \lambda_i Z(\mathbf{x}_i),$$

con un peso λ_0 igual a -1 y $\sum_{i=0}^N \lambda_i = 0$. Así, la condición que los pesos $\{\lambda_i\}$ ($i = 1, \dots, N$) sumaran uno, incluso implica que el uso de el variograma para el cálculo de la varianza de la estimación del error es factible. Por lo tanto, la varianza de la estimación está dada nuevamente en términos del error cuadrático medio, esto es

$$\begin{aligned} \text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)) &= E[(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))^2] \\ &= E\left[\left(\sum_{i=1}^N \lambda_i Z(\mathbf{x}_i) - Z(\mathbf{x}_0)\right)^2\right] \\ &= \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j E[(Z(\mathbf{x}_i) - Z(\mathbf{x}_0))(Z(\mathbf{x}_j) - Z(\mathbf{x}_0))] \\ &= \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j (\gamma(\mathbf{x}_i - \mathbf{x}_0) + \gamma(\mathbf{x}_j - \mathbf{x}_0) - \gamma(\mathbf{x}_i - \mathbf{x}_j)) \\ &= 2 \sum_{i=1}^N \lambda_i \gamma(\mathbf{x}_i - \mathbf{x}_0) - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j). \end{aligned}$$

Dado que se tiene la restricción de los pesos, para minimizar la varianza se utiliza un multiplicador de Lagrange ρ , para considerar la función

$$Q = \text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)) - 2\rho \left(\sum_{j=1}^N \lambda_j - 1\right),$$

de esta forma se puede determinar el mínimo de la varianza sin restricciones equiparando las derivadas parciales de Q a cero, esto es

$$\begin{aligned}\frac{\partial Q}{\partial \lambda_i} &= 2\gamma(\mathbf{x}_i - \mathbf{x}_0) - 2 \sum_{j=1}^N \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) - 2\rho = 0 \\ \frac{\partial Q}{\partial \rho} &= -2 \left(\sum_{j=1}^N \lambda_j - 1 \right) = 0\end{aligned}$$

para cada $i = 1, \dots, N$, lo cual da lugar al sistema de ecuaciones

$$\begin{aligned}\sum_{j=1}^N \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) + \rho &= \gamma(\mathbf{x}_i - \mathbf{x}_0), \quad \text{para } i = 1, \dots, N \\ \sum_{j=1}^N \lambda_j &= 1\end{aligned}$$

que se conoce como *sistema kriging ordinario*. En forma matricial puede ser expresado como,

$$\begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_1 - \mathbf{x}_N) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{x}_N - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_N - \mathbf{x}_N) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \rho \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_N - \mathbf{x}_0) \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} \Gamma & P \\ P^\top & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \rho \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma}_0 \\ 1 \end{pmatrix}$$

donde Γ es la matriz de tamaño $N \times N$, con entradas $\Gamma_{ij} = \gamma(\mathbf{x}_i - \mathbf{x}_j)$, que describe las disimilitudes entre las posiciones de los datos, $\boldsymbol{\gamma}_0$ es un vector de tamaño N , donde la entrada i -ésima denota las disimilitudes entre cada posición de los datos y la posición donde se quiere estimar el nuevo valor \mathbf{x}_0 dada por el variograma, $\boldsymbol{\lambda}$ es un vector de tamaño N con los pesos y P es un vector de tamaño N con todas las entradas 1. El sistema tiene solución única si Γ es estrictamente condicionalmente negativa definida, lo cual se cumple usualmente si se utiliza un modelo válido de variograma.

La varianza de la estimación σ_{KO}^2 se obtiene al premultiplicar las primeras N ecuaciones del sistema kriging ordinario por λ_i , luego, sumando sobre i y usando las condiciones de los pesos se tiene que

$$- \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) = \rho - \sum_{i=1}^N \lambda_i \gamma(\mathbf{x}_i - \mathbf{x}_0),$$

que al sustituirse en la expresión de la varianza obtenida antes

$$\begin{aligned}\sigma_{KO}^2 &= \text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)) \\ &= 2 \sum_{i=1}^N \lambda_i \gamma(\mathbf{x}_i - \mathbf{x}_0) - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \\ &= \rho + \sum_{i=1}^N \lambda_i \gamma(\mathbf{x}_i - \mathbf{x}_0).\end{aligned}$$

Al igual que para kriging simple, kriging ordinario es un interpolador exacto en el sentido que si \mathbf{x}_0 es idéntico con una posición de los datos entonces el valor estimado es idéntico con el valor del dato en ese punto, esto es

$$Z^*(\mathbf{x}_0) = Z(\mathbf{x}_i), \quad \text{si } \mathbf{x}_0 = \mathbf{x}_i.$$

4.3.3. Kriging Universal

El modelo general, el cual Matheron llamo modelo kriging universal por razones que se explicarán después, supone que la media puede ser escrita como una expansión finita

$$m(\mathbf{x}) = \sum_{k=1}^M \omega_k p_k(\mathbf{x}),$$

donde $p_k(\mathbf{x})$ son funciones base conocidas y ω_k son coeficientes fijos pero desconocidos. Usualmente la primera función base ($k = 1$) es la función constante idénticamente igual a 1, la cual garantiza que el caso de media constante está incluido en el modelo. Las otras funciones son usualmente monomios de grado bajo en las componentes de \mathbf{x} (en la práctica el grado no excede dos, véase observación 3.12). Nótese que la expresión para la media puede ser considerada como una aproximación local a $m(\mathbf{x})$, esto es, los coeficientes ω_k pueden variar en el espacio pero lo suficientemente lento para ser considerados constantes entre las vecindades de estimación.

Desde un punto de vista físico el modelo kriging universal es la descomposición de la variable $Z(\mathbf{x})$ en la suma,

$$Z(\mathbf{x}) = m(\mathbf{x}) + Y(\mathbf{x})$$

de una función determinista suave $m(\mathbf{x})$, que describe el aspecto sistemático del fenómeno, y de una función aleatoria de media cero $Y(\mathbf{x})$, llamado el residual, que captura sus fluctuaciones erráticas.

Se quiere estimar $Z(\mathbf{x}_0)$ usando nuevamente el estimador lineal $Z^* = \sum_i \lambda_i Z(\mathbf{x}_i)$ buscando minimizar el error cuadrático medio dado el modelo lineal para la función de la media

$$m(\mathbf{x}) = \sum_{k=1}^M \omega_k p_k(\mathbf{x}).$$

El sesgo puede ser expandido como

$$E [Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = \sum_{i=1}^N \lambda_i \sum_{k=1}^M \omega_k p_k(\mathbf{x}_i) - \sum_{k=1}^M \omega_k p_k(\mathbf{x}_0),$$

luego, intercambiando el orden de las sumas sobre i y k , se obtiene

$$E [Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = \sum_{k=1}^M \omega_k \left(\sum_{i=1}^N \lambda_i (p_k(\mathbf{x}_i) - p_k(\mathbf{x}_0)) \right).$$

Con el objetivo de minimizar el error cuadrático medio $E [(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))^2]$, se quiere hacer el sesgo $E [Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)]^2 = 0$ donde quiera que las incógnitas ω_k aparezcan, lo cual implica eliminar los coeficientes en la última expresión. Esto conduce al conjunto de M condiciones

$$\sum_{i=1}^N \lambda_i p_k(\mathbf{x}_i) = p_k(\mathbf{x}_0), \quad k = 1, \dots, M, \quad (4.6)$$

que Matheron llamo las condiciones de universalidad y por consiguiente el nombre de kriging universal. Estas expresan que el estimador Z^* es insesgado para todos los valores de ω_k . Sujeto a estas condiciones el error cuadrático medio es igual a la varianza del error $Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)$ como en el caso de kriging ordinario, por lo cual, para este también la varianza queda en términos del variograma, esto es

$$\text{Var}(Z^*(x_0) - Z(x_0)) = 2 \sum_{i=1}^N \lambda_i \gamma(\mathbf{x}_i - \mathbf{x}_0) - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j).$$

El problema puede ser reformulado como sigue: Encontrar N pesos λ_i minimizando $\text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))$ sujeto a las restricciones lineales (4.6). Nuevamente se aplica el método de multiplicadores de Lagrange, para ello se toma la función

$$Q = \text{Var}(Z^*(x_0) - Z(x_0)) - 2 \sum_{k=1}^M \rho_k \left(\sum_{i=1}^N \lambda_i p_k(\mathbf{x}_i) - p_k(\mathbf{x}_0) \right),$$

donde ρ_k son M incógnitas adicionales (los multiplicadores de Lagrange), y se determina el mínimo sin restricciones de Q equiparando las parciales de Q a cero, esto es

$$\begin{aligned} \frac{\partial Q}{\partial \lambda_i} &= 2\gamma(\mathbf{x}_i - \mathbf{x}_0) - 2 \sum_{j=1}^N \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) - 2 \sum_{k=1}^M \rho_k p_k(\mathbf{x}_i) = 0, \\ \frac{\partial Q}{\partial \rho_k} &= -2 \left(\sum_{i=1}^N \lambda_i p_k(\mathbf{x}_i) - p_k(\mathbf{x}_0) \right) = 0. \end{aligned}$$

Que el valor extremo es en efecto un mínimo es de nuevo garantizado por la convexidad de $\text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))$ como función de λ_i . Esto nos lleva al siguiente conjunto de $N + M$

ecuaciones lineales con $N + M$ incógnitas que da lugar al *sistema kriging universal*

$$\begin{aligned} \sum_{j=1}^N \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) + \sum_{k=1}^M \rho_k p_k(\mathbf{x}_i) &= \gamma(\mathbf{x}_i - \mathbf{x}_0), \quad i = 1, \dots, N \\ \sum_{i=1}^N \lambda_i p_k(\mathbf{x}_i) &= p_k(\mathbf{x}_0), \quad k = 1, \dots, M. \end{aligned}$$

En forma matricial el sistema se expresa como,

$$\begin{pmatrix} \Gamma & P \\ P^\top & O \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\rho} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma}_0 \\ \mathbf{p}_0 \end{pmatrix},$$

donde Γ y $\boldsymbol{\gamma}_0$ son como antes, O es una matriz de ceros de tamaño $M \times M$, P es una matriz de tamaño $N \times M$ con las evaluaciones en los monomios, $\boldsymbol{\rho}$ los multiplicadores de Lagrange y \mathbf{p}_0 un vector de tamaño M con las evaluaciones de las funciones base en \mathbf{x}_0 . El sistema tiene solución única si $-\Gamma$ es estrictamente condicionalmente definida positiva y P de rango completo. La condición sobre Γ se cumple al usar un modelo admisible de variograma (véase subsección 4.2.2).

También para este caso se puede dar la varianza de la estimación,

$$\sigma_{KU}^2 = \text{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)) = \sum_{k=1}^M \rho_k p_k(\mathbf{x}_0) + \sum_{i=1}^N \lambda_i \gamma(\mathbf{x}_i - \mathbf{x}_0),$$

obtenida al proceder de manera análoga a como se hizo con kriging ordinario.

4.4. Aspectos prácticos

Hasta el momento se ha estudiado la teoría detrás del variograma y kriging. Aunque cada una de las formas del método kriging tienen ciertas consideraciones que los diferencian, ambos siguen la misma idea por lo que es posible dar una regla general para usarlos. Sin embargo, por el aspecto práctico de este trabajo se centrará en la implementación de las formas con media desconocida, kriging ordinario y universal, describiendo los algoritmos para este último ya que kriging ordinario, aunque es más utilizado, es un caso particular de kriging universal. Para aplicar alguna de las dos formas es necesario conocer el variograma de antemano el cual se estima al ajustar un modelo admisible al variograma experimental calculado a partir de los datos, como se mencionó en la subsección 4.2.1.

A *grosso* modo los pasos a seguir son los siguientes:

- 1) El análisis de los datos mediante el cálculo de estadísticos univariados y multivariados, para identificar tendencias, puntos atípicos, etc.
- 2) Cálculo del variograma experimental a partir de los datos.

- 3) Ajuste de un modelo de variograma admisible al variograma experimental.
- 4) Estimación de los nuevos valores utilizando el estimador lineal $Z^* = \sum_i \lambda_i Z(\mathbf{x}_i)$ y un modelo lineal de la función de la media $m(\mathbf{x})$ (constante para el caso de kriging ordinario).

El algoritmo 4.1 bosqueja el procedimiento seguido para realizar la interpolación de datos dispersos mediante kriging universal una vez conocido el variograma. El algoritmo 4.1 es muy parecido al algoritmo 3.1 para interpolación mediante funciones de base radial, con la diferencia que para este por cada nueva posición que se desea evaluar es necesario construir el lado derecho del sistema kriging y resolver el sistema para conocer la estimación correspondiente.

Algoritmo 4.1: Interpolación mediante kriging universal

Entrada: datos dispersos $\mathcal{P} = \{(x_i, y_i, z_i)\}$, posiciones de evaluación $\mathcal{Q} = \{(\hat{x}_j, \hat{y}_j)\}_{j=1, \dots, J}$, modelo admisible de variograma γ , parámetros del modelo de variograma ν

Salida: vector de tamaño J con los valores en las posiciones de evaluación \hat{z}

- 1 $A \leftarrow O_{((N+M) \times (N+M))}$ // $O_{((N+M) \times (N+M))}$ es la matriz de ceros
- 2 **Para** $i \leftarrow 1$ **hasta** N **hacer** // Construcción del sistema kriging universal
 - 3 **Para** $k \leftarrow i + 1$ **hasta** N **hacer**
 - 4 $A_{i,k} \leftarrow \gamma(\|\mathbf{x}_i - \mathbf{x}_k\|, \nu)$
 - 5 $A_{k,i} \leftarrow A_{i,k}$
 - 6 **Para** $\ell \leftarrow 1$ **hasta** M **hacer**
 - 7 $A_{\ell+N,i} \leftarrow p_\ell(\mathbf{x}_i)$
 - 8 $A_{i,\ell+N} \leftarrow A_{\ell+N,i}$
- 9 **Para** $j \leftarrow 1$ **hasta** J **hacer**
 - 10 **Para** $i \leftarrow 1$ **hasta** N **hacer**
 - 11 $b_i \leftarrow \gamma(\|\hat{\mathbf{x}}_j - \mathbf{x}_i\|, \nu)$
 - 12 **Para** $\ell \leftarrow 1$ **hasta** M **hacer**
 - 13 $b_{\ell+N} \leftarrow p_\ell(\hat{\mathbf{x}}_j)$
 - 14 Resolver el sistema $A\omega = \mathbf{b}$ // $\omega^\top = (\lambda^\top \ \rho^\top)$
 - 15 $\hat{z}_j \leftarrow \lambda^\top \mathbf{z}$ // Aquí \mathbf{z} denota el vector de tamaño N con entradas z_i

Observación 4.5. Algunas observaciones sobre el algoritmo 4.1 son pertinentes.

- 1) Los parámetros de los modelos varían con respecto a cual de ellos es utilizado, pero en general los que se consideran son el umbral (*sill*) y el rango (*range*), la pendiente o el exponente en los modelos de funciones potencia y el efecto *nugget* cuando se combina este con alguno de los otros modelos.
- 2) Cuando hay presencia de anisotropía geométrica se utiliza además el ángulo y la razón de anisotropía para pasar al caso isótropo, como se mostró al final de la subsección 4.2.1.

- 3) El lado izquierdo del sistema kriging (matriz A en el algoritmo) solo se construye una vez, por lo que es posible utilizar la descomposición LU para resolver el sistema de la línea 14 y así optimizar el algoritmo.
- 4) También es importante notar que la parte polinomial puede cambiar, ya que esta puede ser constante (kriging ordinario), lineal o a lo más cuadrática, ya que usualmente en la práctica los monomios utilizados son a lo más de orden 2. De este modo, como en la observación 3.12, los que se utilizarán en el caso de interpolación en dos dimensiones son:

$$p_1 = 1, p_2 = x, p_3 = y, p_4 = x^2, p_5 = y^2, p_6 = xy.$$

Ejemplo 4.1. Considérese nuevamente, como en el ejemplo 2.4, la distribución aleatoria de 50 puntos, para los cuales se evalúa la función de prueba $f_1(\mathbf{x})$ de Franke dada en [6].

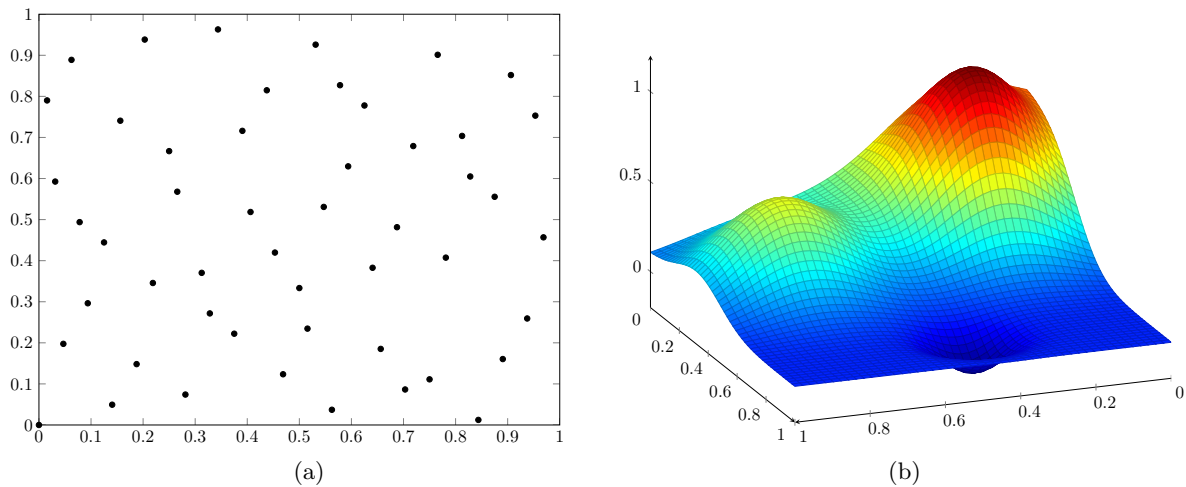


Figura 4.5: (a) Una distribución aleatoria de puntos. (b) Gráfica de la función de prueba f_1

Con los datos generados al evaluar la función de Franke en las posiciones sobre la región $\mathcal{D} = [0, 1] \times [0, 1]$ de la distribución aleatoria mostrada en la figura 4.5a, se tratará de seguir el procedimiento descrito al principio de la sección, profundizando en los detalles del cálculo del variograma experimental, una parte importante debido a que a partir de este se ajusta el modelo admisible que nos permite utilizar kriging para la interpolación de los datos dispersos.

El primer paso es realizar el análisis exploratorio de los datos. Dado que los datos son generados a partir de la función de prueba, para el ejemplo se omitirá dicho análisis ya que la información que se obtiene no será relevante para el desarrollo del mismo. En la subsección 4.2.1 se dieron las referencias pertinentes sobre el tema.

El segundo paso consiste en el cálculo del variograma experimental utilizando la fórmula dada en la ecuación 4.3. Como se consideran datos dispersos, pocos datos estarán ligados por

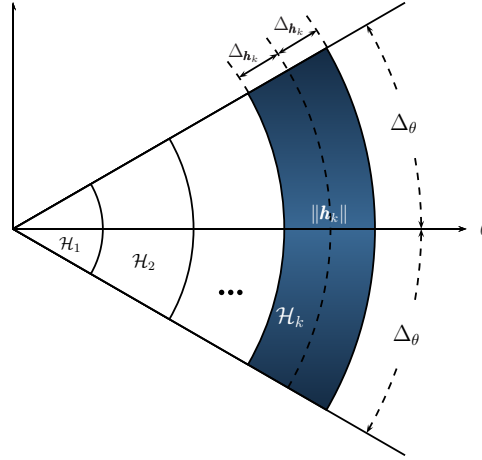


Figura 4.6: Esquema para determinar las clases en el cálculo del variograma experimental

un vector específico \mathbf{h} de la misma magnitud y dirección, es por ello que se deben tomar en cuenta una tolerancia tanto para la magnitud como para el ángulo que define su dirección¹, de tal manera que la k -ésima clase \mathcal{H}_k está formada por todos los vectores \mathbf{h} que cumplen $\|\mathbf{h}_k\| - \Delta_{\mathbf{h}_k} \leq \|\mathbf{h}\| < \|\mathbf{h}_k\| + \Delta_{\mathbf{h}_k}$, además de que el ángulo φ que define su dirección esté acotado entre $\theta - \Delta_\theta \leq \varphi < \theta + \Delta_\theta$, donde \mathbf{h}_k es el representante de la clase, θ es una dirección prefijada y $\Delta_{\mathbf{h}_k}, \Delta_\theta$ las tolerancias respectivas (véase figura 4.6).

En la práctica, el ángulo θ considerado anteriormente define la dirección o las direcciones en las que se quiere calcular el variograma experimental, principalmente para examinar si existe algún tipo de anisotropía. No hay una regla de cuantas direcciones deben ser consideradas ya que esto depende del fenómeno bajo estudio, pero usualmente se toman cuatro, las relacionadas con los puntos cardinales, este-oeste (0°), noreste-suroeste (45°), norte-sur (90°) y noroeste-sureste (135°), con una tolerancia Δ_θ de 22.5° (véase figura 4.2). Notesé que no es necesario considerar otras direcciones ya que como hemos visto el variograma es una función par. Con respecto al rango máximo de las distancias consideradas, es decir, la magnitud del vector de separación \mathbf{h} , se toma por lo general, a lo más la mitad de la extensión de los datos, ya que solo nos interesa cuantificar la correlación entre los valores de las posiciones cercanas entre sí. El número de clases considerado dependen nuevamente del fenómeno bajo estudio al igual si se ocupan tolerancias $\Delta_{\mathbf{h}_k}$ uniformes o no. Así, una vez fijando las direcciones θ_ℓ y estableciendo para estas las clases, la fórmula dada en la ecuación 4.3 se utiliza como

$$\hat{\gamma}_{\theta_\ell}(\mathcal{H}_k) = \frac{1}{2N_{\mathcal{H}_k}} \sum_{i=1}^{N_{\mathcal{H}_k}} (z_j - z_i)^2, \quad \text{con } \mathbf{x}_j - \mathbf{x}_i = \mathbf{h} \in \mathcal{H}_k,$$

la cual mide la disimilitud promedio con respecto a la clase \mathcal{H}_k en la dirección θ_ℓ . La gráfica

¹Los ángulos son medidos con respecto al eje de las abscisas.

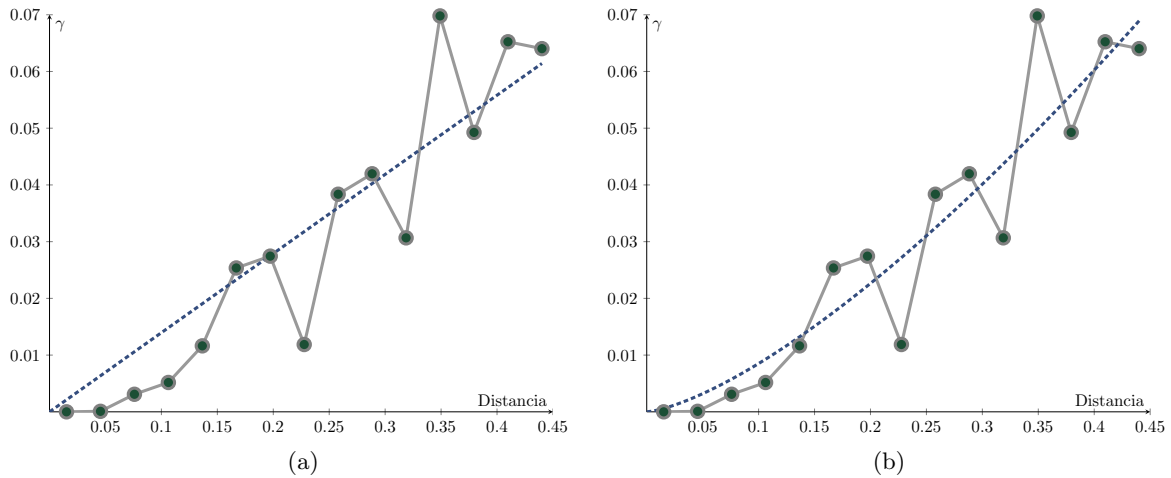


Figura 4.7: Variograma experimental para los datos del ejemplo 4.1 (a) Modelo lineal ajustado con pendiente $\hat{m} = 0.13953$. (b) Modelo de función potencia ajustado con exponente $\alpha = 1.4115$ y pendiente $\hat{m} = 0.21954$

del variograma experimental obtenida para este ejemplo se muestra en la figura 4.7 al considerar un variograma omnidireccional² y 15 clases.

Posteriormente, en el tercer paso, se tiene que ajustar un modelo admisible al variograma experimental. No es necesario que el ajuste sea perfecto, solo basta que el modelo ajustado respete las características esenciales exhibidas por el variograma experimental, que como se menciono anteriormente son el comportamiento cerca del origen, el rango, etc. En el ejemplo se han ajustado dos modelos de funciones potencia como se aprecia en la figura 4.7.

El último paso consiste en estimar los nuevos valores mediante kriging utilizando el algoritmo 4.1. Obsérvese que por cada nueva estimación es necesario resolver un sistema de ecuaciones. Por ejemplo, para cada superficie mostrada en la figura 4.8 fue necesario resolver 2500 sistemas kriging. La rapidez con que estos sistemas se resuelven depende de su tamaño, el cual esta ligado con el número de datos dispersos considerados debido a que, este es un método global, es por ello que se toman todos los datos en la construcción del sistema. Luego, entre más datos se tengan los sistemas serán más grandes y el algoritmo 4.1 será más lento. Este problema se puede abordar tratando de hacer una versión local del método, de forma análoga al método de Shepard modificado al considerar vecindades alrededor de cada punto a estimar.

²Es el variograma experimental para el caso isótropo, ya que solo se toma una dirección de calculo θ entre 0° y 179° con una tolerancia $\Delta_\theta = 90^\circ$, de tal modo que todos los vectores son considerados, no importando ya su dirección.

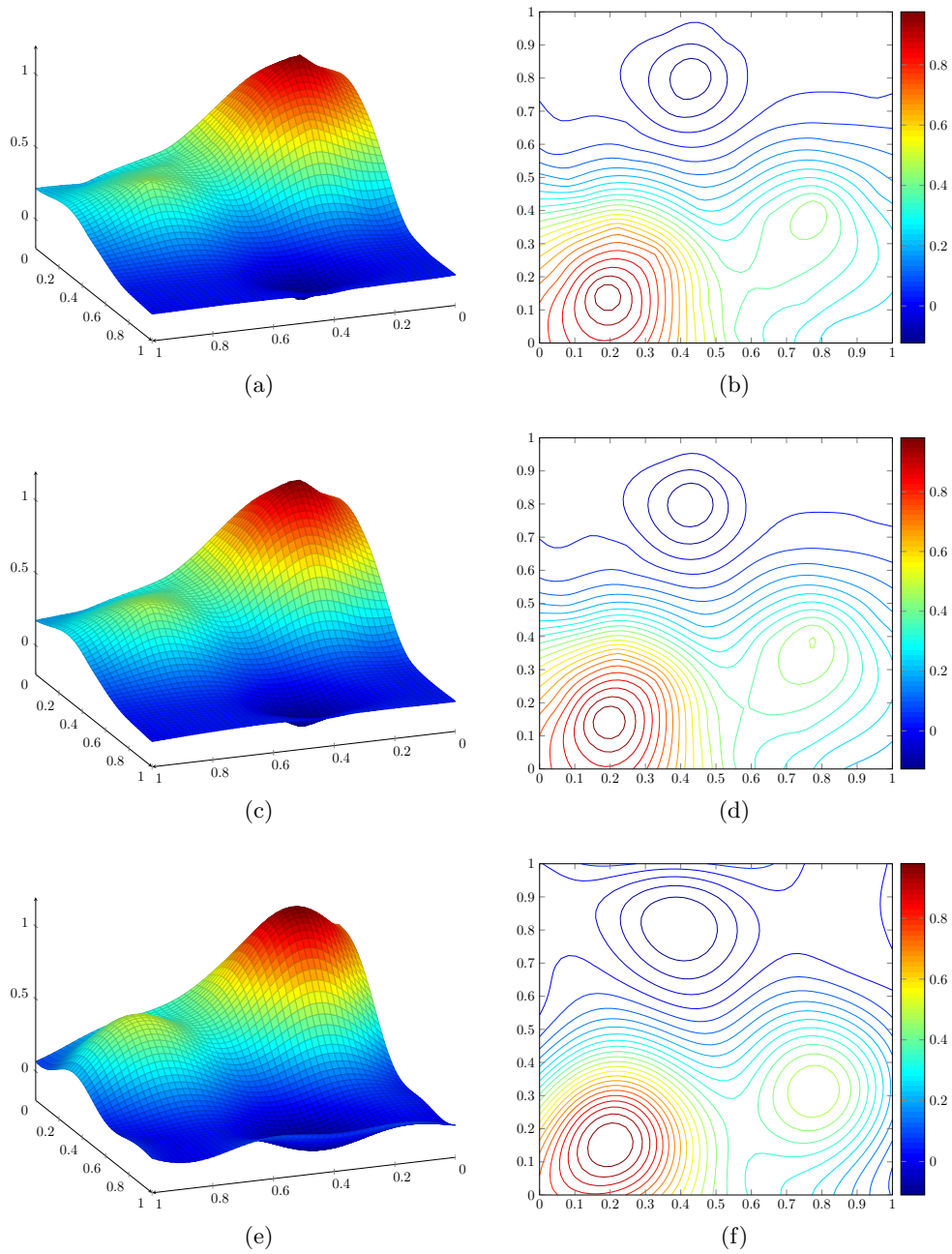


Figura 4.8: Superficies obtenidas mediante kriging ordinario utilizando diferentes modelos admisibles de variograma (a) Superficie obtenida utilizando un modelo lineal. (b) Curvas de nivel. (c) Superficie obtenida utilizando un modelo de función potencia. (d) Curvas de nivel. (e) Superficie obtenida utilizando un modelo gaussiano junto con efecto nugget. (f) Curvas de nivel

La figura 4.8 muestra las superficies obtenidas mediante el uso de kriging ordinario. La figura 4.8a muestra la superficie que se obtuvo utilizando un modelo de variograma lineal con pendiente $\hat{m} = 0.13953$, la figura 4.8c muestra la superficie que se tiene al usar un modelo de función potencia con exponente $\alpha = 1.4115$ y pendiente $\hat{m} = 0.21954$, mientras que la figura 4.8e muestra la superficie obtenida al usar un modelo gaussiano con rango de 0.37202 y umbral de 0.087489, junto con efecto nugget de 1.8356×10^{-4} . Con cualquiera de los tres modelos se lograron buenos resultados de la estimación de la función f_1 . Los modelos de funciones potencia presentan resultados similares, solo que en el caso del modelo lineal se puede apreciar algunos picos cerca de las posiciones de los datos mientras que para el otro modelo no, esto es debido a que el modelo lineal presenta dicho comportamiento cerca del origen, es decir, cuando la distancia es cero. Para el caso del modelo gaussiano, se puede apreciar una superficie más suave a causa de las características de dicho modelo.

Para finalizar, es pertinente hacer un comentario acerca del modelo gaussiano ya que su uso siempre debe ser acompañado del efecto nugget, aunque sea un valor pequeño del mismo, esto porque si se utiliza solo provoca por lo general un mal condicionamiento del sistema kriging y por tanto una estimación errónea de los nuevos valores (Armstrong [1], Chilès & Delfiner [3]).

■

Capítulo 5

B-spline multinivel

Hasta el momento se ha considerado solo el problema de interpolación, pero como se menciono antes, algunas veces se puede tener que los valores de los datos z_i comprendan cierto error. Así, en lugar de considerar el problema de interpolación se plantea el problema de aproximación, donde la única diferencia entre estos dos es que para el problema de aproximación la condición de interpolación se cambia por $F(\mathbf{x}_i) \approx z_i$. De este modo, si los valores de los datos son obtenidos a partir de una función escalar desconocida $f : \mathcal{D} \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ al ser evaluada en un conjunto de posiciones arbitrarias $\mathbf{x}_i = (x_i, y_i)^\top \in \mathbb{R}^2$, $i = 1, \dots, N$, es decir, $f(\mathbf{x}_i) = z_i$, entonces F es una aproximación de f en todo \mathcal{D} .

Para resolver este problema se estudia un algoritmo para construir una función de aproximación C^2 a partir de los datos dados. Este algoritmo aplica una técnica de aproximación B-spline a una jerarquía de mallas de control para generar una secuencia de funciones cuya suma aproxime la función desconocida f . Una mejora se obtiene al usar refinamiento B-spline para representar esta suma de varias funciones como una sola función B-spline. En esta sección se estudiarán las ideas detrás del algoritmo B-spline multinivel y también del refinamiento B-spline. Como antes, aunque el método se puede aplicar a datos multivariados, se centra su aplicación a datos bivariados.

5.1. Preliminares

Se fijarán algunas ideas que serán importantes para que el desarrollo sea inteligible. Sea el intervalo $[0, m]$ para $m \in \mathbb{N}$ y considérese la partición uniforme sobre este dada por $\Xi = \{\xi_i = i \mid i = 0, \dots, m\}$. Se está interesado en encontrar la base B-spline para el espacio lineal de splines cúbicos sobre la partición Ξ , el cual se denotará por $\mathcal{S}_3(\Xi)$.

De acuerdo con Powell [13], si se tiene el intervalo $[a, b]$ y la partición $a = \xi_0 < \xi_1, \dots, \xi_{m-1} < \xi_m = b$, se puede construir un B-spline de grado 3 que sea diferente de cero en (ξ_p, ξ_{p+4}) , $p = 0, \dots, m - 4$ mediante

$$\bar{B}_p(x) = \sum_{j=p}^{p+4} \left[\prod_{\substack{i=p \\ i \neq j}}^{p+4} \frac{1}{(\xi_i - \xi_j)} \right] (x - \xi_j)_+^3, \quad -\infty < x < \infty, \quad (5.1)$$

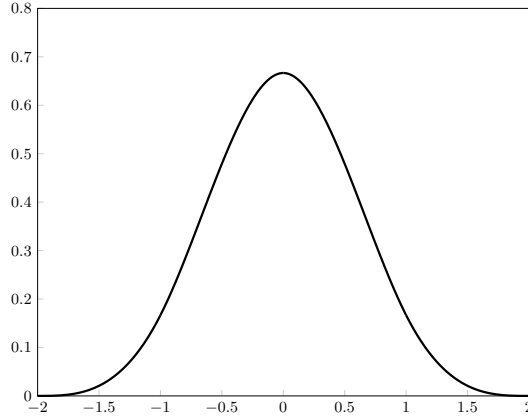


Figura 5.1: B-spline cúbico $B(x)$ sobre la partición $\{-2, -1, 0, 1, 2\}$

donde $(x - \xi_j)_+ = \max(0, x - \xi_j)$.

Ejemplo 5.1. Sea el intervalo $[-2, 2]$, con la partición $\{\xi_0 = -2, \xi_1 = -1, \xi_2 = 0, \xi_3 = 1, \xi_4 = 2\}$. Para $p = 0$, se construye el B-spline $B(x) = 4\overline{B}_0(x)$ sobre el intervalo $(-2, 2)$ utilizando la definición de Powell (5.1), esto es

$$\begin{aligned} B(x) &= 4\overline{B}_0(x) = 4 \left[\frac{1}{24}(x+2)_+^3 - \frac{1}{6}(x+1)_+^3 + \frac{1}{4}(x)_+^3 - \frac{1}{6}(x-1)_+^3 + \frac{1}{24}(x-2)_+^3 \right] \\ &= \frac{1}{6}(x+2)_+^3 - \frac{2}{3}(x+1)_+^3 + (x)_+^3 - \frac{2}{3}(x-1)_+^3 + \frac{1}{6}(x-2)_+^3. \end{aligned}$$

Ya que $B(x)$ es diferente de cero solo en el intervalo $(-2, 2)$, se puede reescribir la última ecuación como

$$B(x) = \begin{cases} \frac{(x+2)^3}{6} & -2 < x \leq -1 \\ \frac{(-3x^3 - 6x^2 + 4)}{6} & -1 \leq x < 0 \\ \frac{(3x^3 - 6x^2 + 4)}{6} & 0 \leq x < 1 \\ \frac{(2-x)^3}{6} & 1 \leq x < 2 \end{cases}, \quad (5.2)$$

de este modo, $B(x)$ es un B-spline cúbico sobre la partición $(-2, -1, 0, 1, 2)$. La figura 5.1 muestra la forma del B-spline $B(x)$. ■

La dimensión del espacio $\mathcal{S}_3(\Xi)$ es $m + 3$, por lo que se necesitan $m + 3$ funciones B-spline cúbicas para la base (véase Powell [13]). Una forma de dar la base es añadiendo

puntos adicionales en los extremos del intervalo $[0, m]$, de tal manera que se pueda utilizar la definición de Powell (ecuación (5.1)) sobre la partición $\{\xi_j = j - 3 \mid j = 0, \dots, m + 6\}$. Otra forma, es utilizar el hecho de que la partición es uniforme, entonces se puede utilizar (5.2) y una traslación de los puntos $t(x) = x - k$, de tal forma, que la base queda dada por

$$\mathcal{B}_k(x) = \begin{cases} \frac{(x + 2 - k)^3}{6} & k - 2 < x \leq k - 1 \\ \frac{[3x^3 + (9k + 6)x^2 + (12k - 9k^2)x + (3k^3 - 6k^2) + 4]}{6} & k - 1 \leq x < k \\ \frac{[3x^3 - (9k + 6)x^2 + (12k + 9k^2)x - (3k^3 + 6k^2) + 4]}{6} & k \leq x < k + 1 \\ \frac{(2 + k - x)^3}{6} & k + 1 \leq x < k + 2 \end{cases} \quad (5.3)$$

con $k = -1, \dots, m + 1$ y $\mathcal{B}_k(x) = B(x - k)$. \mathcal{B}_k denota que el B-spline es distinto de cero en el intervalo $(k - 2, k + 2)$. De ahora en adelante se empleará esta última descripción.

De esta manera, cualquier spline cúbico f en $\mathcal{S}_3(\Xi)$ puede ser expresada por una combinación lineal de los elementos de la base (5.3), *i.e.*

$$f(x) = \sum_{k=-1}^{m+1} \phi_k \mathcal{B}_k(x) = \sum_{k=-1}^{m+1} \phi_k B(x - k). \quad (5.4)$$

Observación 5.1. Si se toma $x \in [i, i + 1)$ con $i = 0, \dots, m - 1$, entonces (5.4) se reduce a

$$f(x) = \sum_{k=i-1}^{i+2} \phi_k \mathcal{B}_k(x) = \sum_{k=i-1}^{i+2} \phi_k B(x - k).$$

ya que son los únicos elementos de la base que son distintos de cero en ese intervalo.

Por otro lado, nótese que cada elemento de la base \mathcal{B}_k esta conformado por cuatro segmentos de curva, correspondientes a los intervalos en la cual es diferente de cero, por lo cual se puede hacer una traslación de manera conveniente para que estos segmentos de curva estén definidos en el intervalo $[0, 1)$.

Se denota por β_k^ℓ , para $\ell = 0, 1, 2, 3$, al segmento de curva para el cual \mathcal{B}_k esta definido en el intervalo $[k + 1 - \ell, k + 2 - \ell)$, de manera que $\mathcal{B}_k(x) = \beta_k^\ell(x)$ sobre este intervalo (véase figura 5.2). Tómesese primero el segmento de curva

$$\beta_k^0(x) = \frac{1}{6}(2 + k - x)^3 \quad \text{para } x \in [k + 1, k + 2)$$

y la traslación $t_0(x) = x - k - 1$, entonces la función se reescribe como,

$$\beta_k^0(x) = \frac{1}{6}(1 - t_0(x))^3 \quad \text{para } 0 \leq t_0(x) < 1.$$

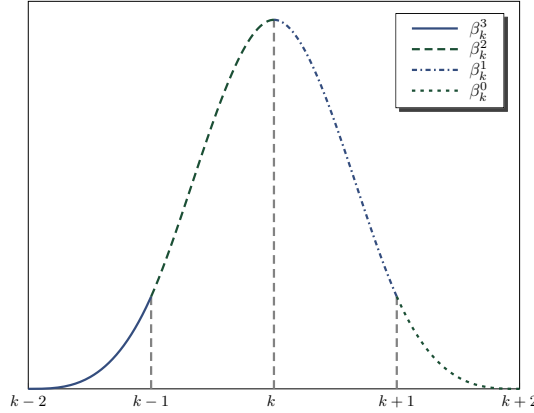


Figura 5.2: Los cuatro segmentos de curva β_k^ℓ del B-spline cúbico \mathcal{B}_k

De forma análoga, para los tres segmentos de curva restantes

$$\beta_k^1(x) = [3x^3 - (9k + 6)x^2 + (12k + 9k^2)x - (3k^3 + 6k^2) + 4] / 6 \quad \text{para } x \in [k, k + 1),$$

$$\beta_k^2(x) = [3x^3 + (9k + 6)x^2 + (12k - 9k^2)x + (3k^3 - 6k^2) + 4] / 6 \quad \text{para } x \in [k - 1, k),$$

$$\beta_k^3(x) = (x + 2 - k)^3 / 6 \quad \text{para } x \in [k - 2, k - 1),$$

se toman las transformaciones

$$\begin{aligned} t_1(x) &= x - k, \\ t_2(x) &= x - k + 1, \\ t_3(x) &= x - k + 2, \end{aligned}$$

respectivamente, de manera tal que estos segmentos de curva se pueden reescribir como

$$\begin{aligned} \beta_k^1(x) &= \frac{1}{6}(3t_1(x)^3 - 6t_1(x)^2 - 4) && \text{para } 0 \leq t_1(x) < 1, \\ \beta_k^2(x) &= \frac{1}{6}(-3t_2(x)^3 + 3t_2(x)^2 + 3t_2(x) - 4) && \text{para } 0 \leq t_2(x) < 1, \\ \beta_k^3(x) &= \frac{1}{6}(t_3(x))^3 && \text{para } 0 \leq t_3(x) < 1. \end{aligned}$$

Dado que las traslaciones t_ℓ cumplen que $0 \leq t_\ell < 1$, entonces, pueden ser reemplazadas convenientemente por una variable t , con $0 \leq t < 1$, de tal manera que los segmentos de curva sean renombrados por

$$\begin{aligned} B_0(t) &= (1 - t)^3 / 6, \\ B_1(t) &= (3t^3 - 6t^2 + 4) / 6, \\ B_2(t) &= (-3t^3 + 3t^2 + 3t + 1) / 6, \\ B_3(t) &= t^3 / 6, \end{aligned} \tag{5.5}$$

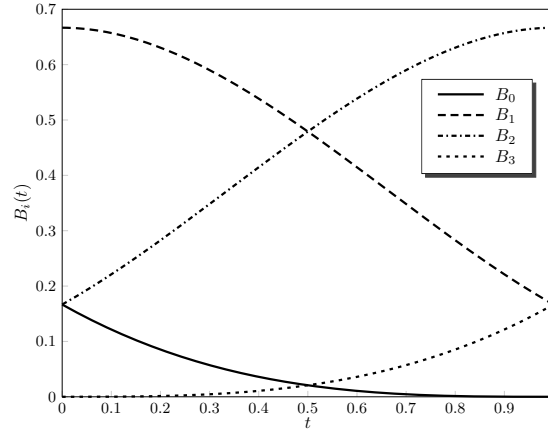


Figura 5.3: Funciones base B-spline cúbicas

donde

$$B_\ell(t) = \beta_k^\ell(x), \quad t = t_\ell(x) = x - k - 1 + \ell,$$

para $x \in [k + 1 - \ell, k + 2 - \ell]$, $k = -1, \dots, m + 1$. La figura 5.3 muestra la gráfica de estas funciones a las cuales se llamará *funciones base B-spline cúbicas*.

Resultado 5.1. Sea f una función en $\mathcal{S}_3(\Xi)$. Entonces para $x \in [i, i + 1)$ ($i = 0, \dots, m - 1$),

$$f(x) = \sum_{\ell=0}^3 \phi_{\ell+i-1} B_\ell(t), \quad t = t(x) = x - i, \quad (5.6)$$

donde B_ℓ son las funciones base B-spline dadas en (5.5).

Demostración. Como $x \in [i, i + 1)$, de la observación 5.1 y recordando que $\mathcal{B}_k(x) = \beta_k^\ell(x)$ para $x \in [k + 1 - \ell, k + 2 - \ell)$ ($\ell = 0, 1, 2, 3$) se tiene que,

$$\begin{aligned} f(x) &= \sum_{k=i-1}^{i+2} \phi_k \mathcal{B}_k(x) = \phi_{i-1} \mathcal{B}_{i-1}(x) + \phi_i \mathcal{B}_i(x) + \phi_{i+1} \mathcal{B}_{i+1}(x) + \phi_{i+2} \mathcal{B}_{i+2}(x) \\ &= \phi_{i-1} \beta_{i-1}^0(x) + \phi_i \beta_i^1(x) + \phi_{i+1} \beta_{i+1}^2(x) + \phi_{i+2} \beta_{i+2}^3(x) = \sum_{\ell=0}^3 \phi_{\ell+i-1} \beta_{\ell+i-1}^\ell(x). \end{aligned}$$

Como

$$B_\ell(t) = \beta_k^\ell(x) = \beta_{\ell+i-1}^\ell(x),$$

con

$$t = t(x) = x - k - 1 + \ell = x - \ell - i + 1 - 1 + \ell = x - i;$$

Sea $\phi_{i,j}$ el valor del i, j -ésimo punto de control en la malla Φ , localizado en (i, j) , para $i = -1, 0, \dots, m + 1$ y $j = -1, 0, \dots, n + 1$. La función de aproximación F es definida en términos de estos puntos de control por,

$$F(\mathbf{x}) = F(x, y) = \sum_{k=0}^3 \sum_{\ell=0}^3 B_k(s) B_\ell(t) \phi_{(i+k)(j+\ell)}, \quad (5.7)$$

donde $i = \lfloor x \rfloor - 1$, $j = \lfloor y \rfloor - 1$, $s = x - \lfloor x \rfloor$ y $t = y - \lfloor y \rfloor$. B_k y B_ℓ son las funciones base B-spline cúbicas uniformes definidas en (5.5) como:

$$\begin{aligned} B_0(r) &= (1 - r)^3/6, \\ B_1(r) &= (3r^3 - 6r^2 + 4)/6, \\ B_2(r) &= (-3r^3 + 3r^2 + 3r + 1)/6, \\ B_3(r) &= r^3/6, \end{aligned}$$

para $0 \leq r < 1$. Estas sirven para ponderar la contribución de cada punto de control a $F(\mathbf{x})$ basado en su distancia a \mathbf{x} .

Observación 5.2. Nótese que la función de aproximación F es obtenida del producto tensorial de dos funciones spline cúbicas, una en la dirección x y otra en la dirección y , las cuales están en términos de las funciones base B-spline cúbicas (véase resultado 5.1), por lo cual, hay dieciséis puntos de control de la malla ya que cada spline en cada dirección cuenta con cuatro coeficientes correspondientes a los elementos de la base que determinan los splines en los intervalos $[i, i + 1)$ para x y $[j, j + 1)$ para y .

Con esta formulación, el problema de obtener la función F se reduce a encontrar los puntos de control en Φ que mejor aproximen los datos dispersos en \mathcal{P} .

Para determinar la malla de control desconocida Φ , primero se considera un punto de los datos (x_c, y_c, z_c) en \mathcal{P} . Si se utiliza la definición de la función de aproximación F dada en (5.7), se tiene que $F(x_c, y_c)$ esta determinada por los dieciséis puntos de control en la vecindad de (x_c, y_c) (véase observación 5.2). Sin pérdida de generalidad, se supondrá que $1 \leq x_c, y_c < 2$ (para hacer $i = j = 0$), de este modo, los puntos de control que determinan el valor de F en (x_c, y_c) son $\phi_{k\ell}$, para $k, \ell = 0, 1, 2, 3$, los cuales se piden que satisfagan

$$F(x_c, y_c) = \sum_{k=0}^3 \sum_{\ell=0}^3 w_{k\ell} \phi_{k\ell} = z_c,$$

donde $w_{k\ell} = B_k(s) B_\ell(t)$ y $s = x_c - 1$, $t = y_c - 1$, ya que al no haber más puntos de los datos la mejor aproximación es que $F(x_c, y_c)$ sea igual al valor conocido z_c . Otras posibilidades pueden ser consideradas, pero seguiremos el desarrollo de Lee *et al.* [12].

Sea $F(\mathbf{x}, \boldsymbol{\varphi}) = \sum_{k=0}^3 \sum_{\ell=0}^3 w_{k\ell} \phi_{k\ell}$, la cual se puede reescribir en forma vectorial como:

$$F(\mathbf{x}, \boldsymbol{\varphi}) = \mathbf{w}^t \boldsymbol{\varphi}$$

donde $\mathbf{x} \in \mathbb{R}^2$ y $\mathbf{w} = (w_{00}, w_{01}, \dots, w_{32}, w_{33})^T$, $\boldsymbol{\varphi} = (\phi_{00}, \phi_{01}, \dots, \phi_{32}, \phi_{33})^T \in \mathbb{R}^{16}$. Sea $S_P \subset \mathbb{R}^{16}$ definido por

$$S_P = \{\boldsymbol{\varphi} \mid F(\mathbf{x}, \boldsymbol{\varphi}) = z_c\},$$

entonces, queremos encontrar $\boldsymbol{\varphi} \in S_P$ tal que,

$$\min_{\boldsymbol{\varphi} \in S_P} \|\boldsymbol{\varphi}\|^2.$$

Dado que se pide la condición de que $F(x_c, y_c) = z_c$, se puede utilizar un multiplicador de Lagrange para convertir el problema anterior en un problema sin restricciones, que nos lleva a minimizar la función

$$L(\boldsymbol{\varphi}, \mathbf{w}) = \frac{1}{2} \|\boldsymbol{\varphi}\|^2 - \lambda(\mathbf{w}^t \boldsymbol{\varphi} - z_c).$$

Derivando parcialmente con respecto a $\boldsymbol{\varphi}$ y λ , esto es

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\varphi}} &= \frac{1}{2} \frac{\partial}{\partial \boldsymbol{\varphi}} \boldsymbol{\varphi}^t \boldsymbol{\varphi} - \lambda \frac{\partial}{\partial \boldsymbol{\varphi}} \mathbf{w}^t \boldsymbol{\varphi} = \boldsymbol{\varphi}^t - \lambda \mathbf{w}^t, \\ \frac{\partial L}{\partial \lambda} &= -(\mathbf{w}^t \boldsymbol{\varphi} - z_c) \frac{\partial}{\partial \lambda} \lambda = z_c - \mathbf{w}^t \boldsymbol{\varphi}, \end{aligned}$$

e igualando a cero las expresiones anteriores, se tiene que

$$\begin{aligned} \boldsymbol{\varphi} &= \lambda \mathbf{w}, \\ \mathbf{w}^t \boldsymbol{\varphi} &= z_c. \end{aligned}$$

Utilizando las dos últimas ecuaciones, se obtiene que

$$\begin{aligned} \mathbf{w}^t \boldsymbol{\varphi} &= z_c \\ \mathbf{w}^t (\lambda \mathbf{w}) &= z_c \\ \lambda \|\mathbf{w}\|^2 &= z_c \\ \lambda &= \frac{z_c}{\|\mathbf{w}\|^2} \end{aligned}$$

por lo tanto,

$$\boldsymbol{\varphi} = \frac{z_c}{\|\mathbf{w}\|^2} \mathbf{w}.$$

Como

$$\|\mathbf{w}\|^2 = \sum_{a=0}^3 \sum_{b=0}^3 w_{ab}^2,$$

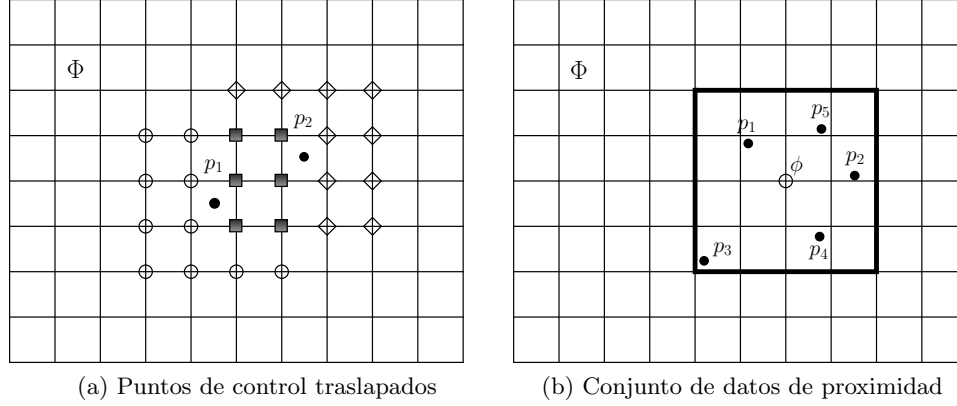


Figura 5.5: Relación posicional entre los datos y los puntos de control

y dado que $\varphi_i = (z_c / \|\mathbf{w}\|^2) \mathbf{w}_i$ para $i = 1, \dots, 16$, entonces

$$\phi_{k\ell} = \frac{w_{k\ell} z_c}{\sum_{a=0}^3 \sum_{b=0}^3 w_{ab}^2}, \quad (5.8)$$

para $k, \ell = 0, 1, 2, 3$. En esta solución, los puntos de control son cercanos a cero si se encuentran lejos de (x_c, y_c) ya que estos están relacionados con pesos $w_{k\ell}$ casi cero.

Ahora se consideran todos los puntos de los datos en \mathcal{P} . Como ya se vio, un punto de los datos define una vecindad de 4×4 puntos de control que son los que participan en el cálculo de la función de aproximación F en dicho punto de los datos. Así, se puede proceder como antes al utilizar la ecuación (5.8) para obtener los puntos de control correspondientes a la vecindad de cada uno de los puntos de los datos en \mathcal{P} . Nótese que estas vecindades pueden superponerse para puntos suficientemente cercanos, como por ejemplo los puntos p_1 y p_2 en la figura 5.5a, por lo que, si utilizamos (5.8) para obtener los puntos de control, estos tendrán diferentes valores si estos pertenecen a diferentes vecindades.

En general, se resuelven asignaciones múltiples a un punto de control ϕ al considerar los puntos de los datos en su vecindad de 4×4 (como se muestra en la figura 5.5b), ya que solo estos puntos pueden influenciar el valor de ϕ al ser obtenido mediante (5.8). Se denominará a estos conjuntos de puntos el conjunto de datos de proximidad de ϕ .

Sea P_{ij} el conjunto de datos de proximidad de un punto de control ϕ_{ij} tal que

$$P_{ij} = \{(x_c, y_c, z_c) \in \mathcal{P} \mid i - 2 \leq x_c < i + 2, j - 2 \leq y_c < j + 2\},$$

luego, para cada $(x_c, y_c, z_c) \in P_{ij}$, (5.8) da a ϕ_{ij} un valor diferente ϕ_{ij}^c

$$\phi_{ij}^c = \frac{w_c z_c}{\sum_{a=0}^3 \sum_{b=0}^3 w_{ab}^2},$$

donde $w_c = w_{k\ell} = B_k(s)B_\ell(t)$, $k = (i+1) - \lfloor x_c \rfloor$, $\ell = (j+1) - \lfloor y_c \rfloor$, $s = x_c - \lfloor x_c \rfloor$, $t = y_c - \lfloor y_c \rfloor$.

Si suponemos que ϕ_{ij} es el valor real del i, j -ésimo punto de control, entonces este es elegido para minimizar el error

$$e(\phi_{ij}) = \sum_c (w_c \phi_{ij} - w_c \phi_{ij}^c)^2$$

donde, el término $(w_c \phi_{ij} - w_c \phi_{ij}^c)$ es la diferencia entre el real y la contribución esperada de ϕ_{ij} a la función F en (x_c, y_c) , es decir, este es el error de aproximación debido a ϕ_{ij} , asumiendo que los otros puntos de control alrededor de (x_c, y_c, z_c) tienen sus valores determinados por (5.8) usando ese punto.

Diferenciando el error $e(\phi_{ij})$ con respecto de ϕ_{ij} ,

$$\begin{aligned} \frac{d}{d\phi_{ij}} e(\phi_{ij}) &= \sum_c w_c^2 \frac{d}{d\phi_{ij}} (\phi_{ij} - \phi_{ij}^c)^2 \\ &= 2 \sum_c w_c^2 (\phi_{ij} - \phi_{ij}^c), \end{aligned}$$

igualando a cero la última expresión

$$\begin{aligned} 2 \sum_c w_c^2 (\phi_{ij} - \phi_{ij}^c) &= 0 \\ \sum_c w_c^2 \phi_{ij} &= \sum_c w_c^2 \phi_{ij}^c, \end{aligned}$$

y por último despejando ϕ_{ij} , se obtiene

$$\phi_{ij} = \frac{\sum_c w_c^2 \phi_{ij}^c}{\sum_c w_c^2}. \quad (5.9)$$

El conjunto de datos de proximidad P_{ij} es el conjunto de puntos en \mathcal{P} en el cual el punto de control ϕ_{ij} tiene una influencia sobre la función de aproximación F . Cuando P_{ij} contiene varios puntos, (5.9) da una solución de mínimos cuadrados a ϕ_{ij} , la cual minimiza un error de aproximación local. Cuando P_{ij} consiste solo de un punto, (5.9) se reduce a (5.8), para el cual no hay error de aproximación ya que se considera el caso de interpolación. Cuando P_{ij} es vacío, sin embargo, ϕ_{ij} no tiene influencia sobre $F(x_c, y_c)$ para cualquier

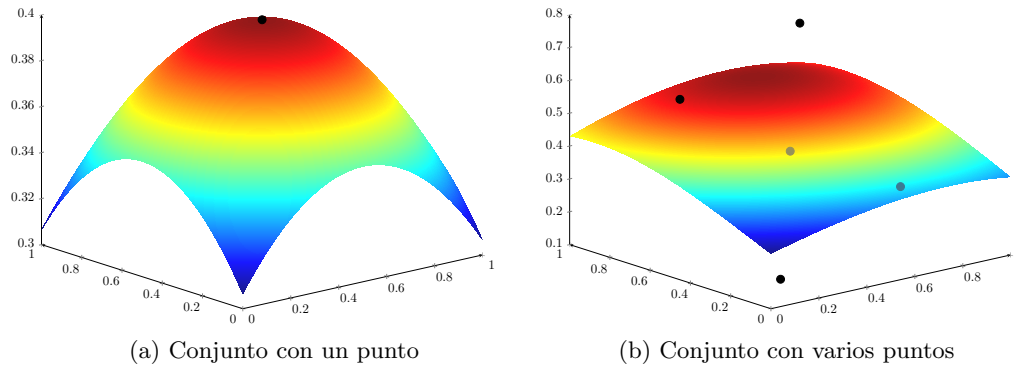


Figura 5.6: Comportamiento de la función de aproximación F definida en (5.7)

punto de los datos (x_c, y_c, z_c) en \mathcal{P} , lo que implica que a ϕ_{ij} se le puede asignar cualquier valor arbitrario, tal como cero o el promedio de los $\{z_c\}$ sin afectar la aproximación del error.

Observación 5.3. Cuando P_{ij} es vacío, se asigna cero a ϕ_{ij} para hacer a la función de aproximación F tender a cero en su vecindad. Esta es una propiedad útil en el algoritmo multinivel.

La figura 5.6 muestra gráficamente el comportamiento de la función de aproximación F . Si todos los ϕ_{ij} que participan en el cálculo de F son determinados mediante (5.8), entonces la función F será una función de interpolación (figura 5.6a). En cambio, si al menos un ϕ_{ij} que participa en el cálculo de F es compartido por varios de los puntos de los datos, entonces estos son determinados mediante (5.9), de esta manera la función F es una función de aproximación (figura 5.6b).

5.2.2. Algoritmo

Para determinar la malla de control Φ a partir de los puntos de los datos, no es necesario identificar de manera explícita el conjunto de datos de proximidad para cada punto de control, ya que, como cada punto de los datos en \mathcal{P} influye solamente un conjunto de 4×4 puntos de control vecinos, este pertenece solo a los conjuntos de datos de proximidad de esos puntos de control. Por lo tanto, se puede acumular eficientemente el numerador y el denominador de (5.9) para cada punto de control al considerar cada punto de los datos en turno. Así, el valor de un punto de control es entonces obtenido por una división si el denominador es distinto de cero. Un denominador nulo ocurre solo en el caso cuando un punto de control tiene un conjunto de datos de proximidad vacío, en tal caso, se asigna el valor cero al punto de control.

El siguiente pseudocódigo describe el método de aproximación B-Spline, el cual se denota como algoritmo BA , y consiste básicamente en la búsqueda de la malla de control Φ .

Algoritmo 5.1: Algoritmo BA

Entrada: datos dispersos $\mathcal{P} = \{(x_c, y_c, z_c)\}$
Salida: malla de control $\Phi = \{\phi_{ij}\}$

- 1 **Para todo** i, j **hacer**
- 2 | $\delta_{ij} = 0$ y $\omega_{ij} = 0$
- 3 **Para cada** (x_c, y_c, z_c) **en** P **hacer**
- 4 | $i = \lfloor x_c \rfloor - 1$ y $j = \lfloor y_c \rfloor - 1$
- 5 | $s = x_c - \lfloor x_c \rfloor$ y $t = y_c - \lfloor y_c \rfloor$
- 6 | calcular $w_{k\ell}$'s y $\sum_{a=0}^3 \sum_{b=0}^3 w_{ab}^2$
- 7 | **Para** $k, \ell = 0, 1, 2, 3$ **hacer**
- 8 | | calcular $\phi_{k\ell}$ con (5.8)
- 9 | | sumar $w_{k\ell}^2 \phi_{k\ell}$ a $\delta_{(i+k)(j+\ell)}$
- 10 | | sumar $w_{k\ell}^2$ a $\omega_{(i+k)(j+\ell)}$
- 11 **Para todo** i, j **hacer**
- 12 | **Si** $\omega_{ij} \neq 0$ **entonces**
- 13 | | calcular $\phi_{ij} = \delta_{ij} / \omega_{ij}$
- 14 | **de lo contrario**
- 15 | | $\phi_{ij} = 0$

Observación 5.4. Dependiendo del tamaño de la malla de control que se utilice, de acuerdo con el algoritmo BA (algoritmo 5), la función de aproximación B-spline tiene algunos de los siguientes comportamientos.

1. Si la malla de control es gruesa, es decir, que hay muy pocos puntos de control, entonces la función de aproximación B-spline (5.7), será una función que suavice los datos originales \mathcal{P} . La figura 5.7b muestra un ejemplo de este caso.
2. Si la malla de control es fina, esto es, que se tenga muchos puntos de control, y los puntos de los datos estén lo suficientemente lejanos unos de otros como para que cada punto de control solo sea influenciado por un punto de los datos, entonces la función de aproximación B-spline (5.7), se vuelve una función de interpolación. Esta función de interpolación es diferente de cero solo en vecindades pequeñas alrededor de los puntos de los datos, por lo cual se tiene la presencia de picos (observe la figura 5.7d). El hecho de que la función de aproximación B-spline se casi cero, se debe a que a los puntos de control se les asigna el valor cero (observación 5.3), cuando estos no son influenciados por puntos de los datos, por lo cual se pierde suavidad, pero esta es una propiedad muy útil para el algoritmo multinivel.
3. Si la malla de control es fina, pero hay puntos de los datos que están lo suficientemente cerca para que un punto de control este influenciado por más de un punto de los datos, entonces se tendrá un comportamiento mixto, como lo muestra la figura 5.7c.

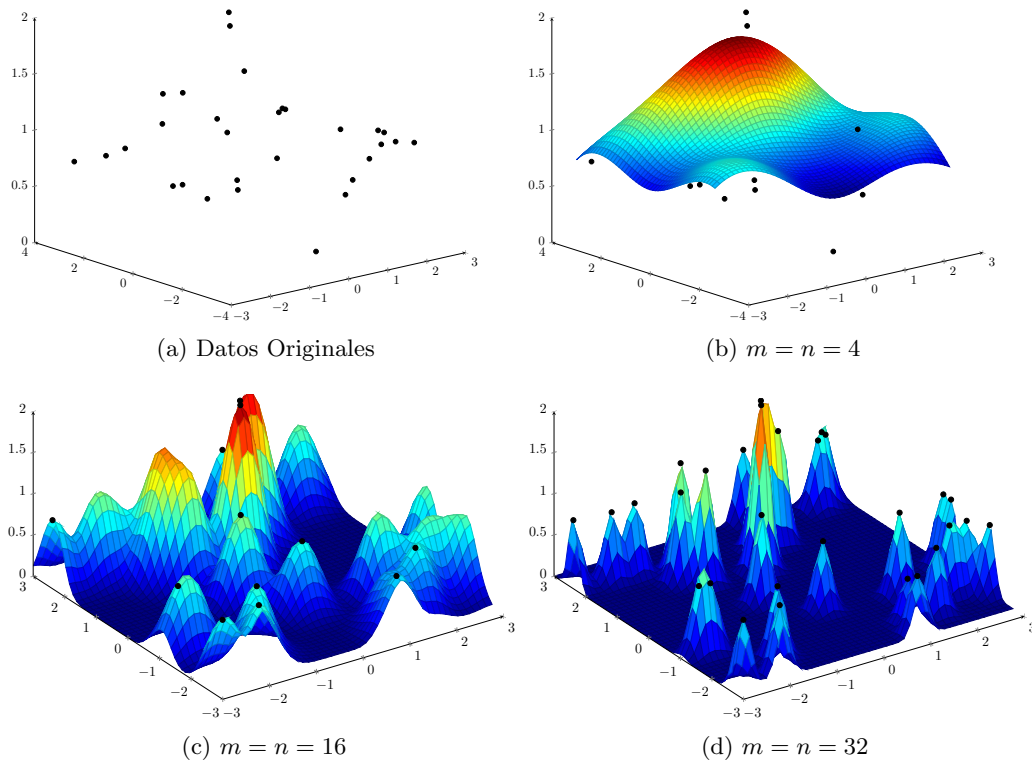


Figura 5.7: Ejemplos de aproximación B-spline para diferentes resoluciones de la malla de control

5.3. Aproximación B-Spline Multinivel

De acuerdo a la observación 5.3, al utilizar el algoritmo BA, dependiendo del tamaño de la malla de control que se utilice, se puede obtener una función suave que aproxime los datos o una función que los interpole pero que presente picos, lo cual puede ser visto como pérdida en la suavidad de la función de interpolación.

En esta sección se presenta un algoritmo de aproximación B-spline tipo multinivel que evite la compensación entre suavidad y precisión de la función de aproximación generada por el algoritmo BA. La función resultante alcanza una forma suave a la par que se aproxima al conjunto de datos dado \mathcal{P} . El algoritmo hace uso de una jerarquía de mallas de control para generar una secuencia de funciones cuya suma se acerca a la función de aproximación deseada. En la secuencia, una función de una malla más gruesa da una aproximación inicial, la cual se va refinando en precisión por funciones obtenidas de mallas más finas. Al final se usa refinamiento B-spline para reducir la suma de estas funciones a una función equivalente.

5.3.1. Algoritmo Básico

Considérese una jerarquía de mallas de control, $\Phi_0, \Phi_1, \dots, \Phi_h$, sobrepuestos en un dominio Ω . Se supone que el espaciamiento entre los puntos de control para Φ_0 está dado y que el espaciamiento es dividido por 2 de una malla a la siguiente (obsérvese la figura 5.8). Bajo esta idea, si Φ_k es una malla de $(m+3) \times (n+3)$, la siguiente malla más fina Φ_{k+1} va a tener $(2m+3) \times (2n+3)$ puntos de control. La posición del punto de control (i, j) en Φ_k coincide con el punto de control $(2i, 2j)$ en Φ_{k+1} .

La aproximación B-spline multinivel comienza por aplicar el algoritmo BA a \mathcal{P} con la malla de control más gruesa Φ_0 . La función resultante F_0 sirve como una aproximación inicial suave que posiblemente deje una gran discrepancia en los puntos de los datos \mathcal{P} . En particular, F_0 deja una desviación $\Delta^1 z_c = z_c - F_0(x_c, y_c)$, para cada punto (x_c, y_c, z_c) en \mathcal{P} . La siguiente malla de control más fina Φ_1 es entonces usada para obtener la función F_1 que aproxime la diferencia $\mathcal{P}_1 = \{(x_c, y_c, \Delta^1 z_c)\}$. De esta forma, la suma $F_0 + F_1$ da una desviación menor $\Delta^2 z_c = z_c - F_0(x_c, y_c) - F_1(x_c, y_c)$ para cada punto (x_c, y_c, z_c) en \mathcal{P} .

En general, para cada nivel k en la jerarquía, se obtiene la función F_k al usar la malla de control Φ_k para aproximar los datos $\mathcal{P}_k = \{(x_c, y_c, \Delta^k z_c)\}$ donde

$$\Delta^k z_c = z_c - \sum_{i=0}^{k-1} F_i(x_c, y_c) = \Delta^{k-1} z_c - F_{k-1}(x_c, y_c),$$

y $\Delta^0 z_c = z_c$. Este proceso comienza desde la malla más gruesa Φ_0 y continúa incrementalmente hacia la malla más fina Φ_h . La función final de aproximación F es definida como la suma de las funciones F_k , *i.e.*

$$F = \sum_{k=0}^h F_k.$$

Observación 5.5. Solo la malla de control Φ_0 es aplicada a los datos originales \mathcal{P} para obtener la forma global de la función F . Todas las mallas finas sucesivas sirven para aproximar y remover el error residual. Esto se logra por la propiedad de que para mallas muy finas, la función de aproximación dada en (5.7) es diferente de cero solo en vecindades pequeñas alrededor de los puntos de los datos (véase observación 5.4 inciso 2), así, mientras k crece, F_k tiende a cero. De esta manera, tenemos una solución incremental para la función f que produce una aproximación suave y cercana a \mathcal{P} .

El siguiente pseudocódigo describe el algoritmo básico para la aproximación multinivel B-spline, el cual se denotará por algoritmo básico MBA. Nótese que una jerarquía de mallas de control es suficiente para representar la función F ya que cada F_k puede ser representada por Φ_k , y F es la suma de los $\{F_k\}$.

Algoritmo 5.2: Algoritmo Básico MBA

Entrada: datos dispersos $\mathcal{P} = \{(x_c, y_c, z_c)\}$

Salida: jerarquía de mallas de control $\Phi_0, \Phi_1, \dots, \Phi_h$

- 1 $k=0$
 - 2 **Mientras** $k \leq h$ **hacer**
 - 3 $\mathcal{P}_k = \{(x_c, y_c, \Delta^k z_c)\}$
 - 4 calcular Φ_k de \mathcal{P}_k mediante el algoritmo BA
 - 5 calcular $\Delta^k z_c = \Delta^{k-1} z_c - F_{k-1}(x_c, y_c)$ para cada punto de los datos
 - 6 $k = k + 1$
-

La figura 5.8 muestra un diagrama esquemático del proceso seguido por el algoritmo.

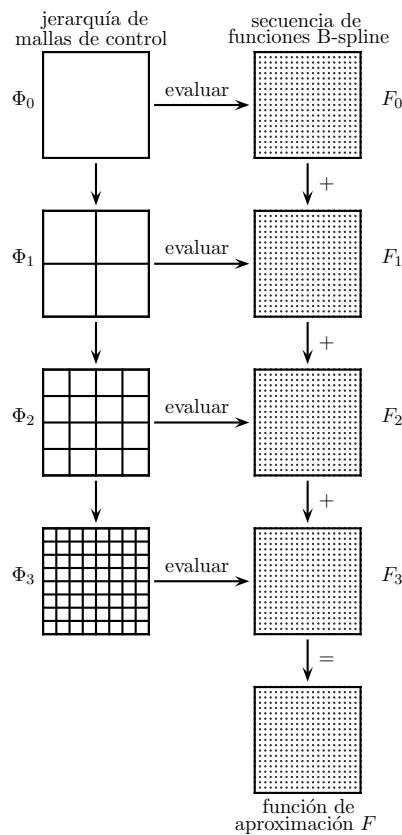


Figura 5.8: Diagrama esquemático del algoritmo básico MBA

5.4. Algoritmo B-spline Multinivel con refinamiento

El algoritmo básico MBA genera una jerarquía de mallas de control que representa la función de aproximación F . Para evaluar F , se debe determinar la función F_k a partir de la malla de control Φ_k para cada nivel k y adicionarlo sobre el dominio Ω .

Esto introduce un aumento significativo en el tiempo si F tiene que ser evaluada en un gran número de puntos en Ω . Por esto, se propone tratar este problema usando progresivamente un refinamiento B-spline a la jerarquía de las mallas de control. Esto permite que F sea representada por una función B-spline en lugar de la suma de varias funciones B-spline. Consecuentemente, el cálculo de F_k es limitado a un pequeño número de puntos de control en Φ_k en lugar de todos los puntos en Ω .

Sea $F(\Phi)$ la función B-spline generada por la malla de control Φ y $|\Phi|$ denote el tamaño de la malla Φ . Con el refinamiento B-spline, se puede construir una malla de control Φ'_0 a partir de la malla más gruesa Φ_0 tal que $F(\Phi'_0) = F_0$ y $|\Phi'_0| = |\Phi_1|$. Entonces la suma de funciones F_0 y F_1 puede ser representada por la malla Ψ_1 la cual sea resultado de sumar a cada par correspondiente de puntos de control en Φ'_0 y Φ_1 , esto es, $F(\Psi_1) = g_1 = f_0 + f_1$, donde $\Psi_1 = \Phi'_0 + \Phi_1$.

En general, sea

$$g_k = \sum_{i=0}^k F_i$$

la suma parcial de funciones F_i hasta el nivel de jerarquía k . Supóngase que la función g_{k-1} es representada por una malla de control Ψ_{k-1} tal que $|\Psi_{k-1}| = |\Phi_{k-1}|$. En la misma forma como se calculo Ψ_1 antes, se puede refinar Ψ_{k-1} para obtener Ψ'_{k-1} , y sumar Ψ'_{k-1} a Φ_k para obtener Ψ_k tal que $F(\Psi_k) = g_k$ y $|\Psi_k| = |\Phi_k|$, esto es, $\Psi_k = \Psi'_{k-1} + \Phi_k$.

Por lo tanto, de $g_0 = F_0$ y $\Psi_0 = \Phi_0$, se puede calcular una sucesión de mallas de control Ψ_k que progresivamente construyan una malla de control Ψ_h para la función de aproximación final $F = g_h$.

Observación 5.6. El proceso descrito por este algoritmo para calcular la función de aproximación $F = g_k$ sigue la idea del método multigrad, ya que, para obtener una mejor aproximación $g_k = F(\Psi_k)$ a partir de una aproximación previa $g_{k-1} = F(\Psi_{k-1})$, es necesario hacer un refinamiento Ψ'_{k-1} de Ψ_{k-1} y hacer una corrección Φ_k para obtener $\Psi_k = \Psi'_{k-1} + \Phi_k$, que caracteriza a la función g_k . Así, este algoritmo reduce considerablemente los costos de calcular g_k , ya que se utilizan los cálculos previos g_{k-1} .

De acuerdo con Lee *et al.* [12], existen varios métodos para refinar una malla de control en otra de tal manera que ambas generan la misma función B-spline. A continuación se describe el proceso de refinamiento que se utiliza en este algoritmo.

5.4.1. Refinamiento del spline bicúbico uniforme

Para mostrar explícitamente el refinamiento que se utiliza en el algoritmo MBA, se iniciará exponiendo las ideas del refinamiento utilizado en el caso univariado y posteriormente usar este para obtener el refinamiento en el caso bivariado.

Refinamiento en una dimensión

Sea el intervalo $[0, m]$ para $m \in \mathbb{N}$ y se considera la partición uniforme sobre este dada por $\Xi = \{\xi_i = i \mid i = 0, \dots, m\}$. Se toma nuevamente el espacio lineal de splines cúbicos definido sobre esta partición, $\mathcal{S} = \mathcal{S}_3(\Xi)$. Sea el refinamiento de la partición anterior, dado por los puntos $\Xi' = \{\xi'_j = \frac{1}{2}j \mid j = 0, \dots, 2m\}$, que se obtiene al insertar puntos a la mitad de los $\{\xi_i\}$. Nótese que $\xi_i = \xi'_j$ para $j = 2i$ ($i = 0, \dots, m$). Entonces sobre esta nueva partición del intervalo $[0, m]$, se toma el espacio lineal de splines cúbicos, $\mathcal{S}' = \mathcal{S}_3(\Xi')$, que cumplen $\mathcal{S} \subset \mathcal{S}'$.

Análogamente a como se procedió antes para el espacio \mathcal{S} , se puede construir una base de B-splines a partir de

$$B'(x) = 2\overline{B}_0(x) = \begin{cases} \frac{(2x+2)^3}{6} & -1 < x \leq -\frac{1}{2} \\ \frac{(-24x^3 - 24x^2 + 4)}{6} & -\frac{1}{2} \leq x < 0 \\ \frac{(24x^3 - 24x^2 + 4)}{6} & 0 \leq x < \frac{1}{2} \\ \frac{(2-2x)^3}{6} & \frac{1}{2} \leq x < 1 \end{cases},$$

donde \overline{B}_0 esta dado por la definición de Powell (5.1) sobre la partición Ξ' . Entonces $B'(x)$ es un B-spline cúbico sobre la partición $\{-1, -1/2, 0, 1/2, 1\}$. Luego la base para el espacio \mathcal{S}' esta dada por:

$$\mathcal{B}'_k(x) = B' \left(x - \frac{1}{2}k \right) \quad \text{para } k = -1, 0, \dots, 2m+1, \quad (5.10)$$

donde \mathcal{B}'_k es distinto de cero para el intervalo $(\frac{1}{2}k - 1, \frac{1}{2}k + 1)$. Así, cualquier spline cúbico f en \mathcal{S}' , puede ser expresado por una combinación lineal de los elementos de la base (5.10), *i.e.*

$$f(x) = \sum_{k=-1}^{2m+1} \phi'_k \mathcal{B}'_k(x) = \sum_{k=-1}^{2m+1} \phi'_k B' \left(x - \frac{1}{2}k \right). \quad (5.11)$$

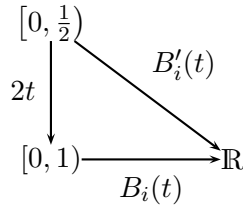
Ahora, si $x \in [-1, 1]$, véase que bajo la transformación $\gamma = \gamma(x) = 2x$, se tiene que $[-1, 1] \xrightarrow{\gamma} [-2, 2]$, por lo cual $B'(x) = B(2x)$, donde $B(x)$ esta dado por (5.2). Dada esta relación entre B' y B , y que las bases de los dos espacios son obtenidos a partir de estas,

entonces es posible construir las funciones base B-spline sobre el espacio \mathcal{S}' , a partir de las funciones base B-spline sobre el espacio \mathcal{S} , dadas en (5.5).

Si se denota por $B'_i(t)$, $i = 0, 1, 2, 3$, las funciones base sobre el espacio refinado \mathcal{S}' , entonces para el intervalo $0 \leq t < \frac{1}{2}$, se tiene que $B'_i(t) = B_i(2t)$ están dadas por,

$$\begin{aligned} B'_0(t) &= \frac{1}{6}(1-2t)^3, \\ B'_1(t) &= 4t^3 - 4t^2 + \frac{2}{3}, \\ B'_2(t) &= -4t^3 + 2t^2 + t + \frac{1}{6}, \\ B'_3(t) &= \frac{4}{3}t^3. \end{aligned} \tag{5.12}$$

donde las funciones B-spline están relacionadas como se muestra en el siguiente diagrama,



De esta manera, de forma análoga al resultado 5.1, se tiene que para $x \in [\frac{1}{2}j, \frac{1}{2}(j+1))$, la función en (5.11), se puede expresar como

$$f(x) = \sum_{k=-1}^{2m+1} \phi'_k \mathcal{B}'_k(x) = \sum_{k=j-1}^{j+2} \phi'_k \mathcal{B}'_k(x) = \sum_{\ell=0}^3 \phi'_{\ell+j-1} B'_\ell(t). \tag{5.13}$$

para $t = t(x) = x - \frac{1}{2}j$.

Luego, para $x \in [\frac{1}{2}j, \frac{1}{2}(j+1)) \subset [i, i+1)$, por las ecuaciones (5.6) y (5.13), se tiene que el spline cúbico $f(x)$ puede ser expresada en ambos espacios, \mathcal{S} y \mathcal{S}' respectivamente, por

$$f(x) = \sum_{\ell=0}^3 \phi_{\ell+i-1} B_\ell(t) = \sum_{\ell=0}^3 \phi'_{\ell+j-1} B'_\ell(t) = \sum_{\ell=0}^3 \phi'_{\ell+2i-1} B'_\ell(t)$$

donde $\{\phi'_{\ell+j-1} = \phi'_{\ell+2i-1}\}$, ya que $j = 2i$, y $\{\phi_{\ell+i-1}\}$ son los coeficientes del spline en \mathcal{S}' y \mathcal{S} respectivamente.

Ahora, tomando las definiciones de las funciones base en ambos espacios (ecuaciones (5.5)

y (5.12)), y ordenando los términos por el grado en t , de la última ecuación se obtiene,

$$\begin{aligned} & \left(-\frac{4}{3}\phi'_{2i-1} + 4\phi'_{2i} - 4\phi'_{2i+1} + \frac{4}{3}\phi'_{2i+2} \right) t^3 + (2\phi'_{2i-1} - 4\phi'_{2i} + 2\phi'_{2i+1}) t^2 \\ & - (\phi'_{2i-1} - \phi'_{2i+1}) t + \left(\frac{1}{6}\phi'_{2i-1} + \frac{2}{3}\phi'_{2i} + \frac{1}{6}\phi'_{2i+1} \right) \\ = & \left(-\frac{1}{6}\phi_{i-1} + \frac{1}{2}\phi_i - \frac{1}{2}\phi_{i+1} + \frac{1}{6}\phi_{i+2} \right) t^3 + \left(\frac{1}{2}\phi_{i-1} - \phi_i + \frac{1}{2}\phi_{i+1} \right) t^2 \\ & - \left(\frac{1}{2}\phi_{i-1} - \frac{1}{2}\phi_{i+1} \right) t + \left(\frac{1}{6}\phi_{i-1} + \frac{2}{3}\phi_i + \frac{1}{6}\phi_{i+1} \right). \end{aligned}$$

Comparando término a término para cada grado en t , se obtiene un sistema lineal de cuatro ecuaciones en las incógnitas $\phi'_{\ell+2i-1}$

$$\begin{pmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 \\ -1 & 0 & 1 & 0 \\ 2 & -4 & 2 & 0 \\ -\frac{4}{3} & 4 & -4 & \frac{4}{3} \end{pmatrix} \begin{pmatrix} \phi'_{2i-1} \\ \phi'_{2i} \\ \phi'_{2i+1} \\ \phi'_{2i+2} \end{pmatrix} = \begin{pmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} & 0 \\ -\frac{1}{6} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{6} \end{pmatrix} \begin{pmatrix} \phi_{i-1} \\ \phi_i \\ \phi_{i+1} \\ \phi_{i+2} \end{pmatrix},$$

donde, la solución esta dada por

$$\begin{pmatrix} \phi'_{2i-1} \\ \phi'_{2i} \\ \phi'_{2i+1} \\ \phi'_{2i+2} \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} & \frac{1}{6} & 0 \\ 1 & 0 & -\frac{1}{12} & 0 \\ 1 & \frac{1}{2} & \frac{1}{6} & 0 \\ 1 & 1 & \frac{11}{12} & \frac{3}{4} \end{pmatrix} \begin{pmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} & 0 \\ -\frac{1}{6} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{6} \end{pmatrix} \begin{pmatrix} \phi_{i-1} \\ \phi_i \\ \phi_{i+1} \\ \phi_{i+2} \end{pmatrix},$$

lo cual nos da los coeficientes en el espacio refinado en términos de los coeficientes conocidos en el espacio original

$$\begin{aligned} \phi'_{2i-1} &= \frac{1}{2}(\phi_{i-1} + \phi_i) \\ \phi'_{2i} &= \frac{1}{8}(\phi_{i-1} + 6\phi_i + \phi_{i+1}) \\ \phi'_{2i+1} &= \frac{1}{2}(\phi_i + \phi_{i+1}) \\ \phi'_{2i+2} &= \frac{1}{8}(\phi_i + 6\phi_{i+1} + \phi_{i+2}) \end{aligned}$$

Véase que ϕ'_{2i-1} es como ϕ'_{2i+1} solo que los índices están trasladados, de la misma manera para ϕ'_{2i} y ϕ'_{2i+2} . Esto se debe a la simetría de las funciones base alrededor del nodo de la malla. Tanto ϕ'_{2i-1} como ϕ'_{2i+1} corresponden a nuevos nodos de la malla en la malla refinada I' a la mitad de los nodos de la malla existentes en la malla original I , y ϕ'_{2i+2} y ϕ'_{2i} corresponden a nuevos nodos en la misma posición que aquellos en la malla original (véase la figura 5.9), así, solo es necesario considerar,

$$\begin{aligned} \phi'_{2i} &= \frac{1}{8}(\phi_{i-1} + 6\phi_i + \phi_{i+1}) \\ \phi'_{2i+1} &= \frac{1}{2}(\phi_i + \phi_{i+1}) \end{aligned} \tag{5.14}$$

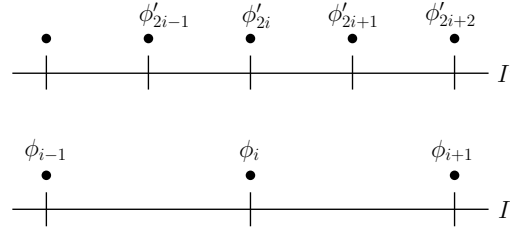


Figura 5.9: Nuevos coeficientes de la malla refinada I' en el caso univariado

Refinamiento en dos dimensiones

Ahora, se quieren encontrar los coeficientes en el caso bivariado para la malla de control refinada Φ' en términos de los coeficientes de Φ . Sea $f(s, t) = \sum_k \sum_\ell \phi_{k\ell} B_k(s) B_\ell(t)$, denote una superficie del producto tensorial B-spline en el espacio definido por Φ_k . Esta ecuación puede ser reescrita como

$$f(s, t) = \sum_k B_k(t) \left[\sum_\ell \phi_{k\ell} B_\ell(s) \right].$$

Nótese que, para cada k en Φ , la suma entre corchetes describe una curva del espacio \mathcal{S} en la variable s . Con esto, cada fila en el producto tensorial tiene una curva de \mathcal{S} correspondiente. Se puede ahora refinar estas curvas para todo k usando el refinamiento univariado dado en (5.14).

$$f(s, t) = \sum_k B_k(t) \left[\sum_\ell \hat{\phi}_{k\ell} B'_\ell(s) \right] = \sum_k \sum_\ell \hat{\phi}_{k\ell} B_k(t) B'_\ell(s).$$

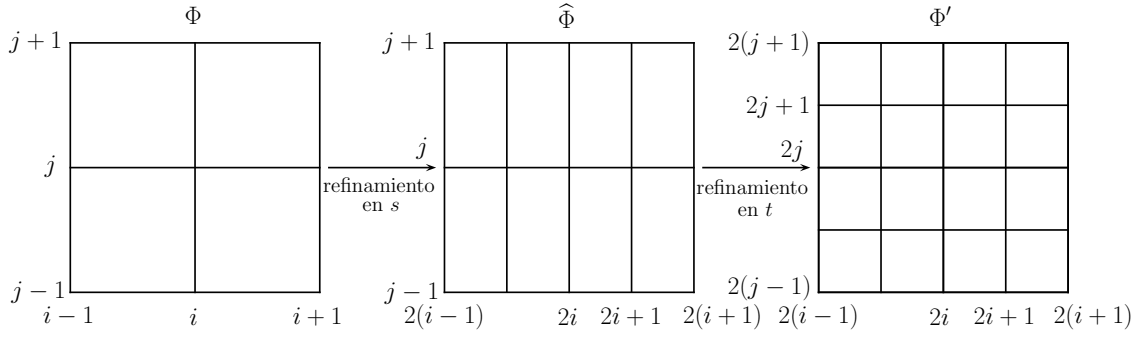
El resultado hasta ahora es la superficie spline original expresada por un nuevo producto tensorial dada por la malla $\hat{\Phi}$ donde las nuevas líneas de la malla en la dirección t han sido insertadas a la mitad de las líneas de la malla Φ . De forma similar a como se ha hecho antes, se puede reescribir la última ecuación como

$$f(s, t) = \sum_\ell B'_\ell(s) \left[\sum_k \hat{\phi}_{k\ell} B_k(t) \right],$$

tal que para cada columna ℓ en $\hat{\Phi}$, la suma entre corchetes describe una curva del espacio \mathcal{S} en la variable t . De nuevo, esta curva puede ser refinada mediante (5.14), así,

$$f(s, t) = \sum_\ell B'_\ell(s) \left[\sum_k \phi'_{k\ell} B'_k(t) \right] = \sum_k \sum_\ell \phi'_{k\ell} B'_k(t) B'_\ell(s).$$

Bajo estas ideas, es ahora fácil encontrar el refinamiento para el caso bivariado a partir del refinamiento dado por el caso univariado (5.14). El diagrama esquemático del proceso


 Figura 5.10: Nuevos coeficientes de la malla refinada Φ' en el caso bivariado

esta dado en la figura 5.10.

Como primero se hace el refinamiento en la dirección s entonces se obtiene una malla de control $\hat{\Phi}$, el cual, sus puntos de control están dados por,

$$\begin{aligned}\hat{\phi}_{2i,j} &= \frac{1}{8}(\phi_{i-1,j} + 6\phi_{i,j} + \phi_{i+1,j}) \\ \hat{\phi}_{2i+1,j} &= \frac{1}{2}(\phi_{i,j} + \phi_{i+1,j}).\end{aligned}$$

Utilizando nuevamente el refinamiento unidimensional, pero ahora sobre la malla de control $\hat{\Phi}$ y en la dirección t , se tiene que:

$$\begin{aligned}\phi'_{2i,2j} &= \frac{1}{8}(\hat{\phi}_{2i,j-1} + 6\hat{\phi}_{2i,j} + \hat{\phi}_{2i,j+1}) \\ \phi'_{2i,2j+1} &= \frac{1}{2}(\hat{\phi}_{2i,j} + \hat{\phi}_{2i,j+1}) \\ \phi'_{2i+1,2j} &= \frac{1}{8}(\hat{\phi}_{2i+1,j-1} + 6\hat{\phi}_{2i+1,j} + \hat{\phi}_{2i+1,j+1}) \\ \phi'_{2i+1,2j+1} &= \frac{1}{2}(\hat{\phi}_{2i+1,j} + \hat{\phi}_{2i+1,j+1})\end{aligned}$$

que finalmente nos da el refinamiento bidimensional,

$$\begin{aligned}\phi'_{2i,2j} &= \frac{1}{64}[\phi_{i-1,j-1} + \phi_{i-1,j+1} + \phi_{i+1,j-1} + \phi_{i+1,j+1} \\ &\quad + 6(\phi_{i-1,j} + \phi_{i,j-1} + \phi_{i,j+1} + \phi_{i+1,j}) + 36\phi_{i,j}] \\ \phi'_{2i,2j+1} &= \frac{1}{16}[\phi_{i-1,j} + \phi_{i-1,j+1} + \phi_{i+1,j} + \phi_{i+1,j+1} + 6(\phi_{i,j} + \phi_{i,j+1})] \\ \phi'_{2i+1,2j} &= \frac{1}{16}[\phi_{i,j-1} + \phi_{i,j+1} + \phi_{i+1,j-1} + \phi_{i+1,j+1} + 6(\phi_{i,j} + \phi_{i+1,j})] \\ \phi'_{2i+1,2j+1} &= \frac{1}{4}[\phi_{i,j} + \phi_{i,j+1} + \phi_{i+1,j} + \phi_{i+1,j+1}]\end{aligned}\quad (5.15)$$

5.4.2. Algoritmo con refinamiento B-spline

Para generar la malla de control Ψ_h a partir de los datos \mathcal{P} , no se tiene que conservar los elementos Φ_k , Ψ_k , Ψ'_k y \mathcal{P}_k para todo k . Si se aplica la aproximación B-spline y el refinamiento juntos en cada nivel, Ψ_h puede ser obtenida recorriendo la jerarquía de mallas de control desde la más gruesa hasta la más fina. En el recorrido, solo una variable por cada uno de los datos y secuencia de mallas de control es suficiente para manejar el cálculo y los resultados intermedios.

Esta técnica a la cual se le da el nombre de algoritmo MBA, se describe en el siguiente pseudocódigo. En el algoritmo, $\mathcal{P} - F(\Phi)$ denota la actualización de los datos $\{x_c, y_c, \Delta^{k+1}z_c\}$, donde $\mathcal{P} = \{x_c, y_c, \Delta^k z_c\}$ y $F_k = F(\Phi)$.

Algoritmo 5.3: Algoritmo MBA

Entrada: datos dispersos $\mathcal{P} = \{(x_c, y_c, z_c)\}$
Salida: malla de control Ψ

- 1 sea Φ la malla de control más gruesa
- 2 sea $\Psi' = 0$
- 3 **Mientras** Φ no exceda la malla de control más fina **hacer**
- 4 calcular Φ a partir de P por el algoritmo BA
- 5 calcular $\mathcal{P} = \mathcal{P} - F(\Phi)$
- 6 calcular $\Psi = \Psi' + \Phi$
- 7 sea Φ la siguiente malla de control fina
- 8 refinar Ψ en Ψ' mediante el cual $F(\Psi') = F(\Psi)$ y $|\Psi'| = |\Phi|$

Algunas notas finales sobre el algoritmo son dadas en la siguiente:

Observación 5.7. Las siguientes notas son aspectos prácticos del algoritmo MBA.

- Los datos de entrada $\mathcal{P} = \{(x_c, y_c, z_c)\}$ deben ser transformados constantemente para poder utilizar mallas de control de diferentes tamaños; esto es, que cumplan la suposición de que $(x_c, y_c) \in \Omega$ para poder utilizar el algoritmo BA.
- Como se ha mencionado antes, si la malla Φ_k tiene un tamaño $(m+3) \times (n+3)$ entonces la mallas más fina siguiente tendrá un tamaño $(2m+3) \times (2n+3)$, donde m y n son especificados.
- El cambio del tamaño de las mallas mencionado en el punto anterior, permite que el refinamiento sea llevado a cabo mediante el refinamiento bidimensional dado en (5.15).
- La condición de paro del algoritmo MBA, consiste en no pasar un tamaño máximo de la malla de control especificado. Se puede utilizar una condición de paro en forma conjunta o separada de esta, al considerar una tolerancia sobre el error residual, esto es, que $\|\Delta^{k+1}\mathbf{z}\| < tol$, donde tol es un valor especificado.

La figura 5.11 muestra un diagrama esquemático del proceso seguido por el algoritmo.

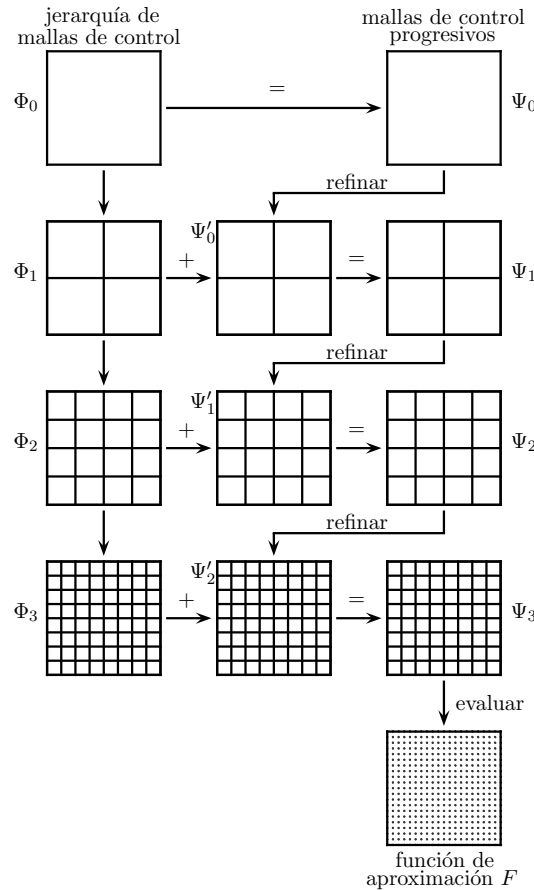


Figura 5.11: Diagrama esquemático del algoritmo MBA

Ejemplo 5.2. Considérese nuevamente, como en el ejemplo 2.4, la distribución aleatoria de 50 puntos, para los cuales se evalúa la función de prueba $f_1(\mathbf{x})$ de Franke dada en [6] por

$$f_1(x, y) = 0.75 \exp\left(-\frac{(9x - 2)^2 + (9y - 2)^2}{4}\right) + 0.75 \exp\left(-\frac{(9x + 1)^2}{49} - \frac{9y + 1}{10}\right) + 0.5 \exp\left(-\frac{(9x - 7)^2 + (9y - 3)^2}{4}\right) - 0.2 \exp(-(9x - 4)^2 - (9y - 7)^2).$$

La malla de control inicial se construyó tomando $m = n = 1$ y la malla final de control tomando $m = n = 2^9$. En la figura 5.12 se comparan los resultados obtenidos con la función de aproximación obtenida mediante el algoritmo MBA contra la función original f_1 . Se puede ver la discrepancia entre ambas, aunque en general tienen una forma parecida. Esta discrepancia está relacionada con el número de puntos de los datos.

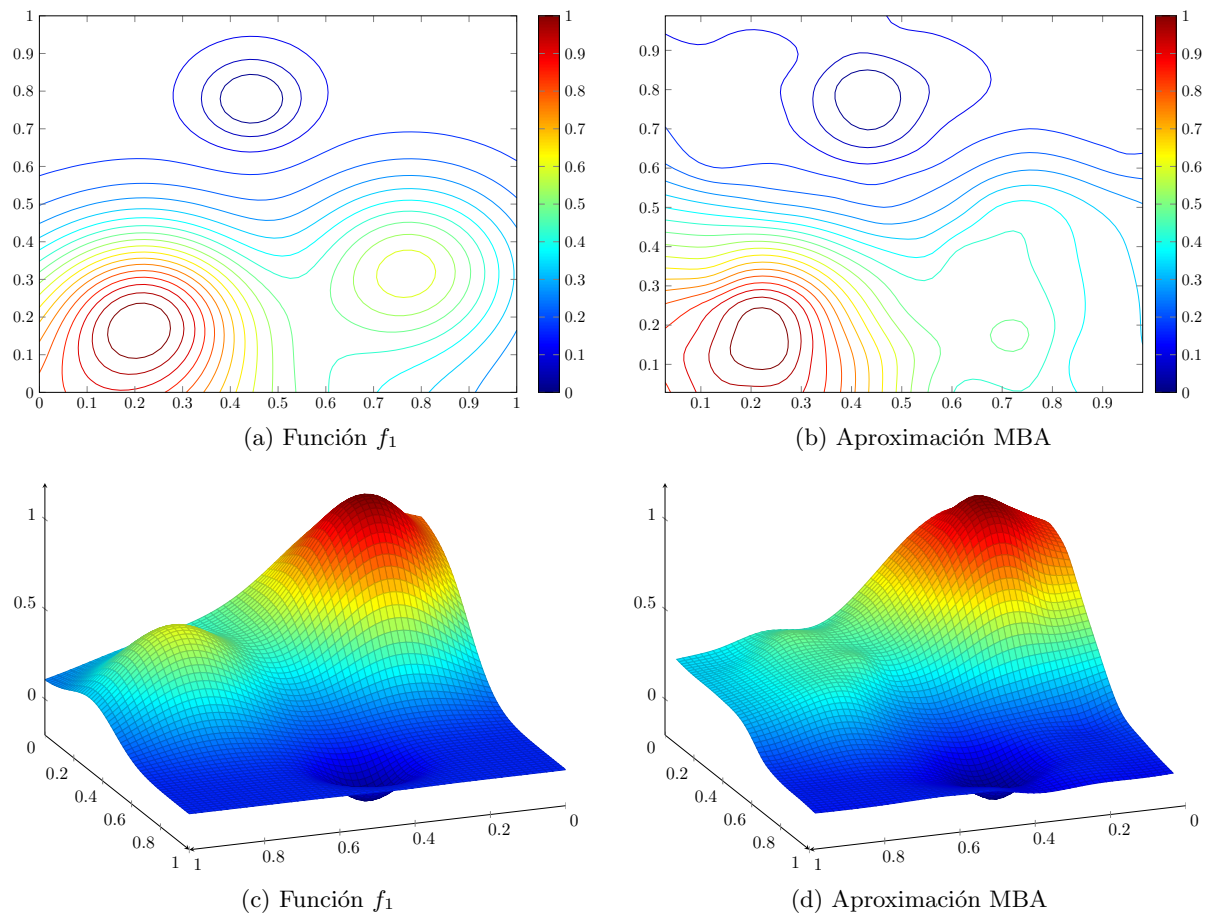


Figura 5.12: (a)-(b) Comparación de las curvas de nivel. (c)-(d) Comparación de las superficies.

La figura 5.12d muestra que la forma global de la función de aproximación obtenida mediante el algoritmo BA es bastante similar a la función f_1 , a pesar de que el número de puntos es muy pequeño.

■

Capítulo 6

Sistema y aplicaciones.

Como se ha mencionado con anterioridad, el objetivo del presente trabajo es utilizar con un enfoque práctico los métodos de interpolación de datos dispersos expuestos en los capítulos previos. Con este fin se ha creado el sistema *UNAMSI v 1.0*, una herramienta en *Matlab R2010a* para aplicar de manera sencilla dichos métodos a un conjunto de datos en el espacio. Adicionalmente, dentro del sistema se encuentra el módulo de construcción de mallas 3D mediante copias verticales, una de las primeras aplicaciones que se pretende dar al sistema.

6.1. Sistema *UNAMSI v1.0*.

En la presente sección se describen las partes esenciales del sistema además de mostrarse algunos ejemplos ilustrativos de su uso. El enfoque general del sistema es trabajar sobre un conjunto específico de datos en el espacio al cual se le pueda aplicar el método de interpolación que más se ajuste a las necesidades del usuario. Por ello se proveen varias herramientas visuales y de edición que serán descritas en su momento. La figura 6.1 muestra la ventana principal del sistema desde donde es posible elegir alguna opción entre los siguientes menús:

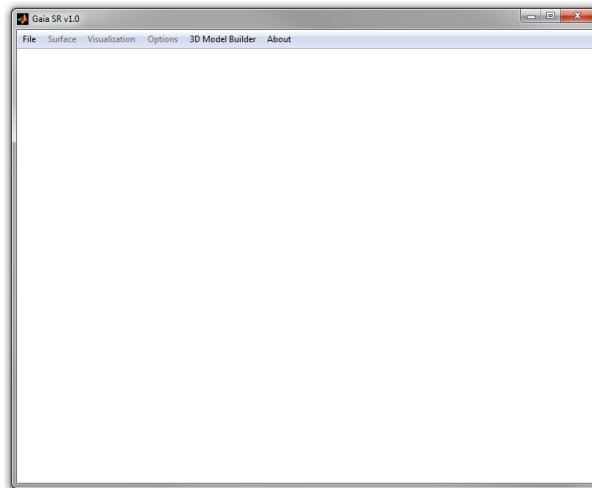


Figura 6.1: Pantalla principal del sistema *UNAMSI v1.0*.

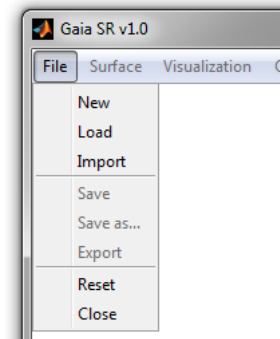


Figura 6.2: Opciones en el menú *File*.

- *File*, incluye las instrucciones para abrir, crear y guardar los archivos de datos.
- *Surface*, contiene las instrucciones para construir una superficie mediante algún método seleccionado o generar un variograma a partir de los datos disponibles.
- *Visualization*, permite visualizar los diferentes elementos generados (datos, superficie y variograma).
- *Options*, da una gama amplia de opciones para cada uno de los elementos generados.
- *3D Model Builder*, es un acceso al modulo de construcción de modelos geométricos 3D.
- *About*, despliega información del sistema.

Los menús *Surface*, *Visualization* y *Options* se encuentran deshabilitados hasta que algún archivo es cargado o se ha creado un conjunto de datos.

6.1.1. Abrir, crear o importar un conjunto de datos.

Para empezar a trabajar con el sistema lo primero que se debe hacer es contar con un conjunto de datos cargado en el mismo. Esta acción puede realizarse de tres maneras distintas desde el menú *File* (véase figura 6.2):

- 1) *New*, crea un conjunto de datos.
- 2) *Load*, carga un archivo que contenga un conjunto de datos.
- 3) *Import*, permite importar algunos formatos de archivos.

Crear un conjunto de datos con *New*.

Con el sistema se puede crear un conjunto de datos manualmente para trabajar mediante la opción *New* que se encuentra dentro *File*. Cuando esta opción es seleccionada aparecerá la

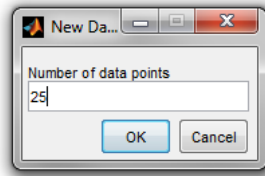


Figura 6.3: Ventana para especificar el tamaño del conjunto de datos a crear.

ventana mostrada en la figura 6.3 para solicitar el tamaño del conjunto de datos. Una vez que se ha confirmado el valor correspondiente se mostrará dentro de la ventana principal las herramientas para visualizar y editar los datos (figura 6.4) que son, una tabla que contiene los valores ¹ de los datos y una gráfica de la nube de puntos en el espacio.

Por defecto la tabla se llena con valores en cero en todas las posiciones. Para modificar las entradas de la tabla solo basta con seleccionar la casilla correspondiente e introducir el nuevo valor. Cada vez que se modifica alguna entrada en la tabla también se actualiza la gráfica de la nube de puntos.

Abrir un conjunto de datos con *Load*.

Para cargar un fichero de datos dentro del sistema se debe utilizar la opción *Load*. Los archivos por defecto que se utilizan para el sistema son los archivos `mat` de `Matlab`. Para que estos archivos funcionen en el sistema debe haber dos variables dentro del fichero `mat` nombradas como:

- **S**, matriz de tamaño $N \times 2$ que contiene las posiciones en el plano (coordenadas (x_i, y_i)).
- **Y**, vector de tamaño N que contienen las respuestas o salidas z_i de una función escalar desconocida.

Si alguna de las variables no se encuentra dentro del archivo o no corresponde al tipo especificado se desplegará un mensaje de error. Los archivos pueden tener otras variables definidas dentro de él pero estas no serán utilizadas ya que solo se busca dentro del archivo las variables antes mencionadas. Una vez que se ha cargado el archivo correctamente se deben mostrar las herramientas de visualización y modificación de datos (figura 6.4) con la información contenida.

Importar un conjunto de datos con *Import*.

El sistema cuenta con la opción de importar algunos tipos de archivos. Los archivos soportados por el sistema son `txt` (texto plano), `csv` (valores separados por comas) y `dig`

¹Los valores corresponden a las coordenadas x , y y z , donde el punto (x, y) es la posición en el plano y z la respuesta o salida de una función escalar desconocida f .

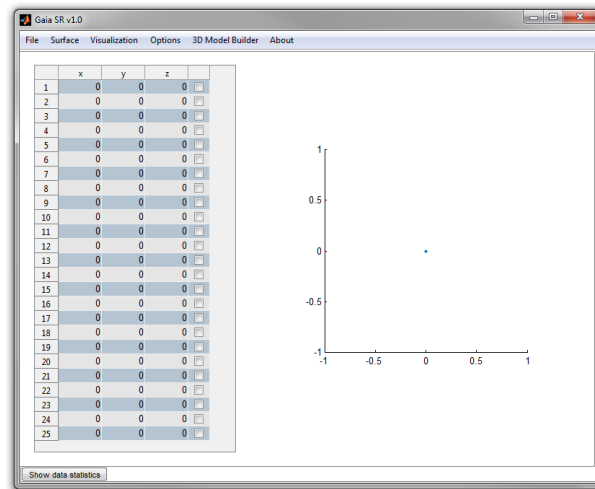


Figura 6.4: Herramientas para la visualización y modificación de datos.

(curvas de nivel). En caso de algún error en el formato del archivo se mostrará un mensaje de error. Una vez que se ha cargado el archivo correctamente se deben mostrar las herramientas de visualización y modificación de datos (figura 6.4) con la información contenida.

6.1.2. Guardar un conjunto de datos.

Cuando los datos son modificados los cambios realizados pueden ser guardados con las opciones *Save*, *Save as* y *Export* dentro del menú *File*. Estas opciones permanecen desactivadas mientras no estén visibles las herramientas de visualización y modificación de los datos. Las opciones para guardar tienen diferentes utilidades:

- 1) *Save*. Esta opción guarda los cambios realizados a los datos directamente sobre el archivo por lo que cada vez que se utiliza la opción un mensaje de advertencia es mostrado. Se recomienda hacer una copia del archivo utilizando *Save as* para no perder la información original. La opción solo se activa cuando algún cambio es realizado a los datos.
- 2) *Save as*. La opción permite guardar los datos en todo momento y es útil para hacer una copia del archivo original o guardar los cambios con otro nombre de archivo.
- 3) *Export*. Sirve para exportar los datos a archivos `txt` (texto plano) y `csv` (valores separados por comas), para que puedan ser utilizados en otros programas y editores.

El sistema utiliza los datos que están guardados en el archivo por lo que las modificaciones hechas a estos solo pueden ser utilizadas hasta que sean guardadas con *Save* y *Save as*. El formato en que son guardados los datos es el de los archivos `mat`.

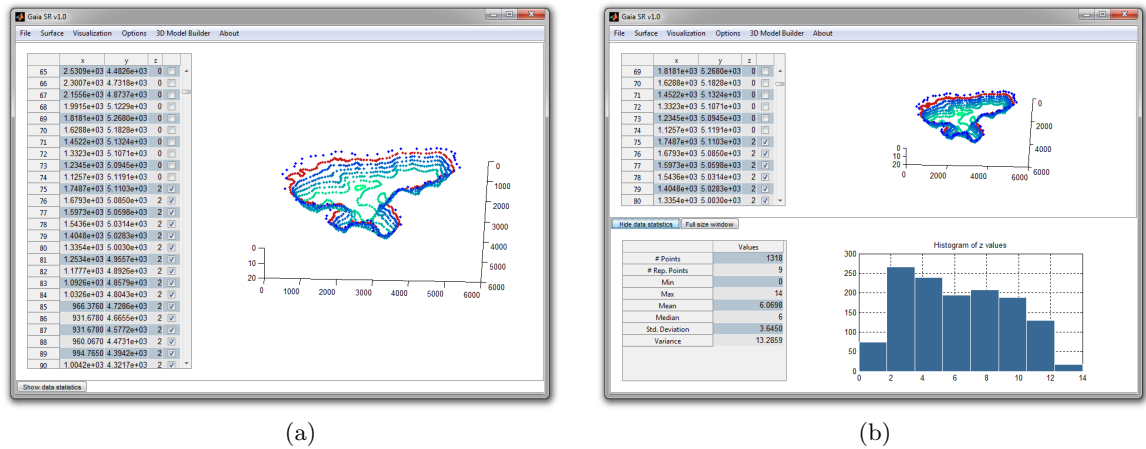


Figura 6.5: (a) Herramientas principales de visualización y modificación de los datos. (b) Herramientas principales junto con información estadística de los datos.

6.1.3. Herramientas y opciones de visualización y modificación de los datos.

El sistema cuenta con varias opciones para visualizar y modificar los datos que se hayan cargado. La figura 6.4 muestra las herramientas principales que son la tabla y la gráfica de la nube de puntos. La tabla tiene la función de mostrar los valores en las tres columnas x , y y z , correspondientes las dos primeras a la posición y la tercera a la respuesta, y la selección de los puntos, mediante el uso de la cuarta columna, tanto para su edición así como para ser usados de manera interactiva con la gráfica. La gráfica permite observar la nube de puntos desde diferentes perspectivas y diferentes conjuntos de puntos seleccionados en la tabla (figura 6.5a).

En adición a las herramientas de visualización y modificación el sistema cuenta con el cálculo de algunos estadísticos unidimensionales usuales sobre las respuestas (datos de la columna z de la tabla) y también de su correspondiente histograma (figura 6.5b) con el objetivo de dar información adicional sobre los datos. Para acceder a dicha información estadística solo es necesario presionar el botón *Show data statistics* situado en la parte inferior izquierda de la ventana.

Opciones para la manipulación de los datos.

Para acceder a las opciones para la manipulación de los datos es necesario dirigirse al submenú *Data* dentro del menú *Options*. El siguiente listado da una descripción de las opciones disponibles.

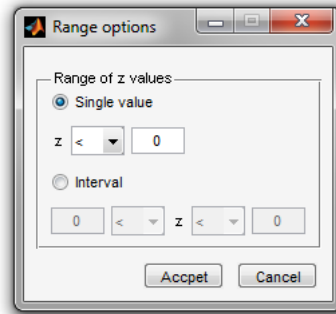


Figura 6.6: Ventana para especificar los intervalos de selección.

1. *Edit.* Activa o desactiva la opción de edición de la tabla de datos. Cuando la opción está activa esta aparece marcada en el submenú. Por defecto esta desactivada cuando se abre o importa un archivo de datos y está activa cuando se crea un conjunto de datos.
2. *Restore.* Permite regresar a la versión de los datos guardada en el archivo. Cada vez que se guardan las modificaciones la opción se desactiva.
3. *Resize.* Proporciona un grupo de opciones para borrar y agregar datos.
 - *Change number of data*, permite cambiar el número de puntos de los datos en la tabla. Cuando se selecciona esta opción aparece una ventana como la de la figura 6.3, donde se debe poner el número de puntos de los datos que se quiera. Si el valor puesto es menor que el número de datos en la tabla actual, se mostrará una advertencia de que solo se conservarán un número de elementos de la tabla igual al valor dado (desde el primero hasta el elemento que ocupa el lugar del valor). Si el valor es mayor, se agregaran nuevos espacios en la tabla rellenos por defecto con ceros.
 - *Remove selected*, remueve los elementos seleccionados en la tabla.
 - *Remove not selected*, remueve los elementos de la tabla que no estén seleccionados.
 - *Remove repeated data*, busca y quita los puntos repetidos de la tabla.
4. *Select.* Da un grupo de opciones para la selección y visualización de los datos en la tabla.
 - *All*, selecciona todos los datos de la tabla.
 - *None*, quita la selección de los elementos previamente marcados.
 - *Invert*, invierte la selección.
 - *By a range of values*, permite la selección de los elementos de la tabla para intervalos de las respuestas (columna *z*). Cuando se elige esta opción aparece una

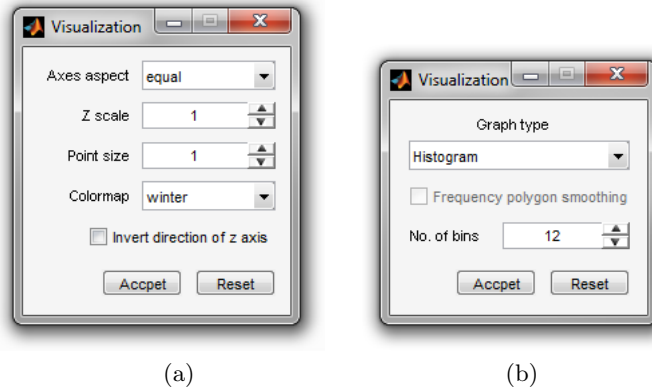


Figura 6.7: (a) Ventana de herramientas de visualización de la gráfica de la nube de puntos. (b) Ventana de herramientas de visualización del histograma.

ventana con dos opciones para los intervalos de selección (figura 6.6). *Single value*, da la alternativa de los intervalos $(-\infty, v)$, $(-\infty, v]$, $[v, \infty)$, (v, ∞) y del conjunto de todos los elementos de la tabla cuyo valor sea v . *Interval*, permite seleccionar elementos con respecto a los intervalos (v_1, v_2) , $[v_1, v_2)$, $(v_1, v_2]$ y $[v_1, v_2]$. La última opción junto con *Invert* permite la selección de intervalos, como por ejemplo, $(-\infty, v_1] \cup [v_2, \infty)$ si se selecciona previamente el intervalo (v_1, v_2) .

- *Repeated data*, busca y selecciona los elementos repetidos para que puedan ser vistos en la gráfica de la nube de puntos.

5. *Graph*. Ofrece un conjunto de opciones para la gráfica de la nube de puntos.

- *Zoom on/off*, activa y desactiva la opción de zoom de la gráfica de la nube de puntos. Cuando la opción se activa esta aparece marcada en el submenú, la gráfica se cambia a a vista 2D y se desactiva la rotación. Una vez que se desactiva la opción de zoom, se vuelve a activar la rotación.
- *Reset view*, regresa la gráfica a la vista por defecto (vista 2D, el zoom desactivado y la rotación activada).
- *Visualization*, permite el acceso al control de visualización de la gráfica de la nube de puntos (figura 6.7a). Con el control se puede modificar el aspecto de los ejes, cambiar la escala en el eje z (cuando el aspecto de los ejes se define igual para los ejes x y y), el tamaño de los puntos en la nube, el mapa de colores y la orientación del eje z .
- *Save as image*, ofrece la opción de guardar la gráfica de la nube de puntos como un archivo de imagen (bmp, eps, jpg, pdf, png, tif).

6) *Statistics*. Provee un conjunto de opciones para el cálculo de la información estadística. La opción se encuentra activa solo cuando la información estadística esta visible.

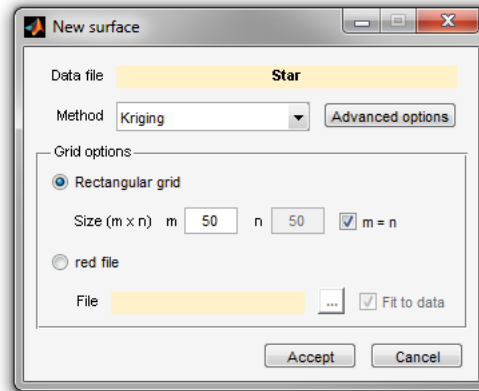


Figura 6.8: Ventana para especificar los elementos necesarios para la reconstrucción de una superficie.

- *Update*, vuelve a calcular la información estadística y el histograma con los datos que se encuentran en la tabla.
- *Save*, permite guardar la información estadística en formato *txt* y *csv*.
- *Histogram*, da la opción de guardar el histograma como una imagen (*bmp*, *eps*, *jpg*, *pdf*, *png*, *tif*) mediante el submenú *Save as image* o dar acceso al control de visualización del histograma mediante el submenú *Visualization* (figura 6.7b). Con el control es posible cambiar la forma como se muestra el histograma con tres opciones diferentes que son: a) Solo el histograma, b) el histograma junto con el polígono de frecuencias y c) solo el polígono de frecuencias. También, con el control, se puede activar el suavizamiento del polígono de frecuencias y cambiar el número de clases usadas para el cálculo de histograma.

6.1.4. Reconstrucción de superficies

Para reconstruir una superficie mediante alguno de los métodos expuestos en los primeros capítulos solo basta abrir o crear algún conjunto de datos y después elegir el submenú *Smooth surface* dentro del menú *Surface* para que se abra la ventana de opciones que se muestra en la figura 6.8. En esta ventana se fijan tanto el método de interpolación a usar así como la malla sobre la que se proyecta la superficie reconstruida. Esta opción utiliza los datos guardados en el archivo, por lo que, los cambios sobre los datos sin guardar no tienen efecto sobre la reconstrucción hasta que sean guardados como un archivo del sistema.

Como ya se ha mencionado de forma breve, para reconstruir una superficie es necesario definir el método de interpolación y la malla sobre la que se proyecta la superficie dentro de la ventana de opciones que se detalla a continuación.

1. *Method*. Permite elegir el método de interpolación a usar con valores por defecto para sus parámetros. Los métodos a elegir son: Distancia inversa, kriging, B-spline multini-vel, funciones de base radial y Shepard modificado.
2. *Advanced options*. Proporciona acceso a los controles para establecer los valores de los parámetros de cada método que el usuario desee. Esta es una opción para usuarios expertos.
3. *Grid options*. Define el tipo de malla a utilizar mediante dos opciones:
 - *Rectangular grid*, genera una malla rectangular² de tamaño $m \times n$ que cubre la extensión de los datos.
 - *red File*, utiliza una malla generada por el sistema UNAMalla . Si las magnitudes de las coordenadas de la malla no corresponden con la de los datos se puede utilizar la casilla `Fit to Data` para forzarlos a que se ajusten a la extensión de los datos. Se recomienda que las magnitudes de las coordenadas correspondan para obtener mejores resultados. Para mayores referencias sobre el formato de los archivos `red` consulte [20].

Para iniciar el proceso de reconstrucción solo es necesario especificar el método y la malla en la ventana de la figura 6.8. De esta forma el proceso se iniciará con los valores por defecto para el método seleccionado. Si se quiere iniciar con valores específicos de los parámetros del método seleccionado es necesario fijar primero las opciones de la malla y después entrar a la opción *Advanced options* para dar los valores de los parámetros e iniciar el proceso desde ahí.

Al iniciar el proceso de reconstrucción se verifica que el número de puntos distintos sean suficientes para el método seleccionado. Si no se encuentra ningún problema aparecerá la ventana de visualización de la superficie (figura 6.9). Si se quiere cambiar el método a utilizar o la malla solo basta entrar nuevamente a la opción *Smooth surface*, aunque esto borrará la superficie previamente reconstruida. Automáticamente se cargan las opciones y valores elegidos en la reconstrucción previa, pero, si se han utilizado las opciones avanzadas de los métodos y se quieren volver a utilizar, es necesario volver a entrar en esta opción y comenzar el proceso de reconstrucción desde ahí.

La ventana de visualización de la superficie cuenta con las siguientes opciones dentro del panel que se encuentra a la derecha:

1. *3D View*, permite observar la superficie reconstruida junto con otros elementos como la malla, curvas de nivel y los puntos de los datos (figura 6.9a).
2. *2D View*, cambia la vista a una bidimensional donde se muestran las curvas de nivel y también los puntos de los datos (figura 6.9b).

²Cuadrada si la casilla `m = n` (véase figura 6.8) se encuentra activa, donde m define las particiones sobre el eje x y n las particiones sobre el eje y .

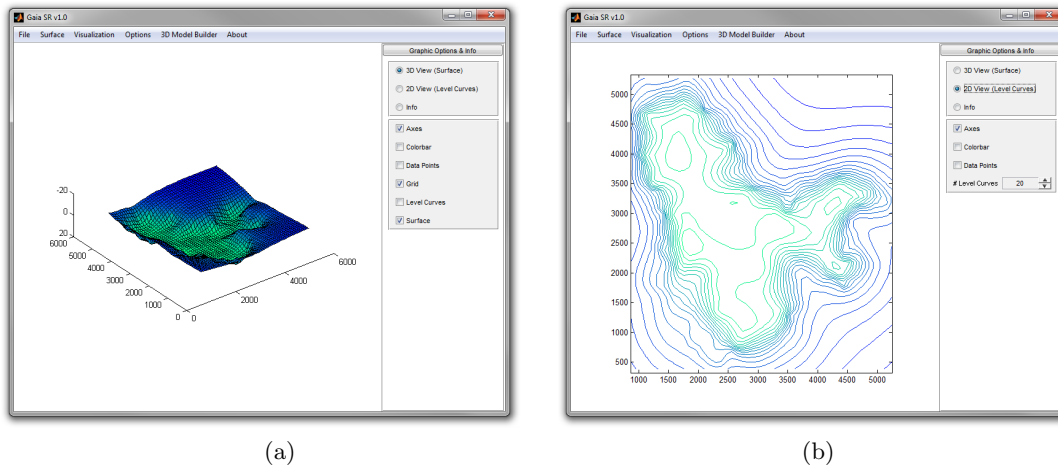


Figura 6.9: Ventana de visualización de la superficie. (a) Vista 3D (superficie). (b) Vista 2D (curvas de nivel).

3. *Info*, muestra en el panel derecho información sobre la superficie reconstruida. La información que se da es: El nombre del archivo de datos, el número de puntos original, el número de puntos repetidos, el número de puntos usados, el tamaño de la malla y el método utilizado junto con sus parámetros.

Opciones avanzadas para los métodos de interpolación.

Cada uno de los métodos utilizados tienen parámetros específicos que necesitan ser definidos. Usualmente el sistema utiliza valores por defecto para ellos pero estos pueden ser especificados por el usuario mediante la opción *Advanced options* dentro de la ventana de opciones para la reconstrucción de la superficie. Para cada método se despliega un control para fijar el valor de los parámetros, cuyo listado por método es el siguiente:

1. *Inverse Distance*. Para el método solo se define un solo parámetro.
 - *Power*, define la potencia a la que se elevan los pesos de la ponderación. Su valor debe ser un valor no negativo. El valor por defecto es 2.
2. *Kriging*. Los parámetros que se definen aquí están determinados por el modelo de variograma usado. Es recomendable hacer un análisis de variograma con la opción del menú *Variogram* dentro de *Surface* para cargarlos mediante la opción del control *Load parameters*.
 - *Variogram model*, es el modelo admisible de variograma que se utiliza. Las posibilidades son: Lineal, potencia, esférico, exponencial y gaussiano. La opción por defecto es lineal.

- *+Nugget Effect*, utiliza un modelo admisible de variograma (definido por el parámetro anterior) junto con el modelo de efecto nugget. Por defecto se encuentra desactivado excepto para el modelo gaussiano. El valor del efecto nugget debe ser positivo.
- *Sill/Slope*, define el valor de la pendiente para el caso de los modelos lineal y potencia y el “umbral” (*sill*) para los otros modelos. El valor por defecto es 1 y este debe ser un valor no negativo.
- *Exponent/Range*, establece el valor del exponente para el caso del modelo potencia y el “rango” para los demás modelos excepto el lineal que no utiliza este parámetro. Por defecto se encuentra desactivada la opción por ser el modelo por omisión el lineal. Cuando se define el exponente los valores deben estar entre 0 y 2 y en el caso del “rango” debe ser positivo.
- *Anisotropy*, define el ángulo y la razón de anisotropía. Los valores por defecto son 0 y 1 (caso isótropo) y van de -179° a 179° y de 0.05 a 20, respectivamente. Los ángulos deben ser valores enteros y para un valor diferente de 1 para la razón de anisotropía se considera anisotropía geométrica.
- *Mean function*, establece la forma de kriging utilizado, ordinario (constante) o universal (lineal y cuadrático). La opción por defecto es constante.

3. *Multilevel B-spline*. En este método solo es necesario definir dos parámetros.

- *Maximum number of levels*, fija el número máximo de niveles utilizados para la aproximación multinivel. El valor por defecto es 12 y debe ser un valor entero entre 1 y 14 por el uso de memoria.
- *Start with linear approximation*, inicia la aproximación con el plano que mejor ajusta los datos en el sentido de mínimos cuadrados. Por defecto esta opción se encuentra activa.

4. *Radial Basis Functions*. De manera similar a kriging los parámetros dependen de la base de funciones de base radial seleccionada.

- *Basis function*, define la base de funciones de base radial que se utiliza. Las posibilidades son: lineales, cúbicas, splines de placa delgada, gaussianas, multicuadráticas y multicuadráticas inversas. La opción por defecto es splines de placa delgada.
- *Shape parameter*, establece el parámetro de forma de las funciones que lo utilizan (gaussianas, multicuadráticas y multicuadráticas inversas). Por defecto la opción se encuentra desactivada.
- *Polynomial part*, permite elegir si se utiliza el método con una parte polinomial o no. Las opciones son: Ninguna, contante, lineal y cuadrática. La opción por defecto es lineal debido a la base utilizada por omisión.

5. *Shepard*. El método solo requiere de dos parámetros.

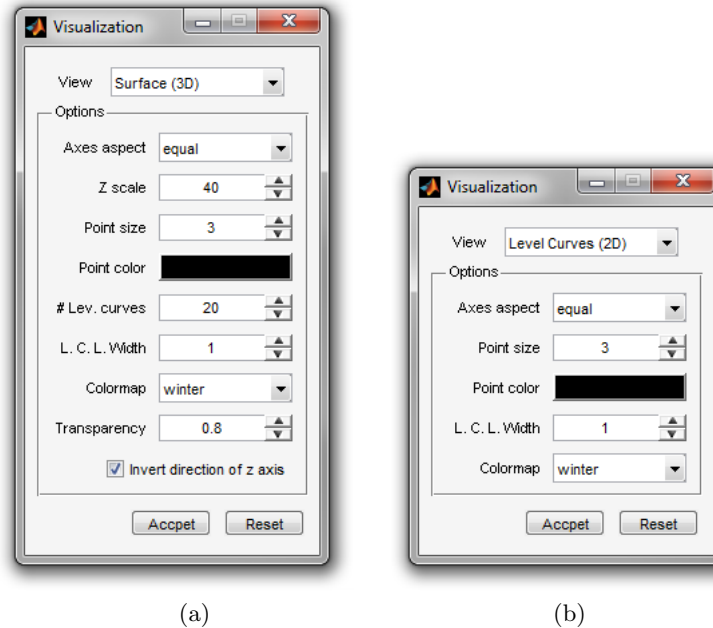


Figura 6.10: Ventana de herramientas de visualización de la superficie y de las curvas de nivel. (a) Controles para la vista 3D. (b) Controles para la vista 2D.

- *Quadratic neighbours*, define el número de puntos que participan en el ajuste de las funciones cuadráticas. El valor por defecto es 13.
- *Weighting neighbours*, define el número de funciones cuadráticas que se incluyen en la ponderación de un nuevo punto. El valor por defecto es 19.

Opciones de la visualización de la superficie.

Para acceder a las opciones de visualización de la superficie es necesario dirigirse al submenú *Surface* dentro del menú *Options*. El siguiente listado da una descripción de las opciones disponibles.

1. *Save*. Sirve para guardar la superficie reconstruida en el formato **gsf** que en esencia es el formato **mat** con una extensión diferente para no mezclarse con los archivos de datos utilizados por el sistema. En este archivo se guardan tres matrices con las coordenadas (x, y, z) para ser graficadas con **surf** y **mesh**. Los archivos se cargan en **Matlab** de la siguiente forma: `load(nombre_del_archivo.gsf, '-mat')`.
2. *Save as image*. Ofrece la opción de guardar la gráfica de la superficie o de las curvas de nivel, dependiendo de cual se este visualizando en el momento de elegir la opción, como un archivo de imagen (**bmp**, **eps**, **jpg**, **pdf**, **png**, **tif**).

3. *Reset view*. Regresa a la configuración inicial los controles de visualización. Tiene dos opciones que son: *Current view*, que reestablece los controles de la vista actual y *All views*, que hace lo mismo para todas las vistas. Ambas opciones tienen otras subopciones que permiten reestablecer los controles del panel derecho (*Principal settings*) o de toda la ventana de visualización (*All settings*).
4. *Visualization*. Permite el acceso al control de visualización de las gráficas de la superficie y de las curvas de nivel (figura 6.10). Con el control se puede modificar para ambas vistas el aspecto de los ejes, el tamaño de los puntos de los datos junto con el color, el ancho de línea de las curvas de nivel y el mapa de colores. Adicionalmente, solo para la superficie, se puede cambiar la escala en el eje z (cuando el aspecto de los ejes se define igual para los ejes x y y), el número de curvas de nivel, la transparencia de la superficie y la orientación del eje z .
5. *Export to 3DMB*. La opción sirve para utilizar la superficie reconstruida dentro del módulo de construcción de mallas 3D.

6.1.5. Cálculo y ajuste de variogramas

Dentro de las herramientas de análisis de los datos para el uso del método kriging se encuentra el cálculo del variograma experimental y su posterior ajuste mediante un modelo admisible. Se puede acceder a esta herramienta desde el menú *Variogram* dentro de *Surface*. Si se cuenta con pocos puntos el cálculo de variograma puede ser poco confiable. La herramienta está pensada para el análisis de la correlación espacial de los datos así como para detectar la presencia de anisotropía, ya que permite calcular el variograma experimental en diferentes direcciones al igual que con diferentes tolerancias en los ángulos. También permite realizar el ajuste de los modelos admisibles de forma manual además de contar con ajuste automático.

Para calcular el variograma experimental solo basta tener cargado un conjunto de datos y elegir la opción *Variogram* dentro del menú *Surface*. El cálculo depende de los datos guardados en el archivo, por lo que, los cambios a los datos sin guardar no tienen efecto sobre el cálculo hasta que sean guardados como un archivo del sistema. Por defecto se calcula un variograma omnidireccional (en todas las direcciones) ajustado por un modelo lineal.

Panel de opciones para el cálculo del variograma experimental y ajuste de modelos.

El panel de opciones constan de dos pestañas de opciones *Experimental* (figura 6.11a) y *Model* (figura 6.11b) que proveen un conjunto de opciones para el cálculo del variograma experimental al igual que para el ajuste de los modelos admisibles.

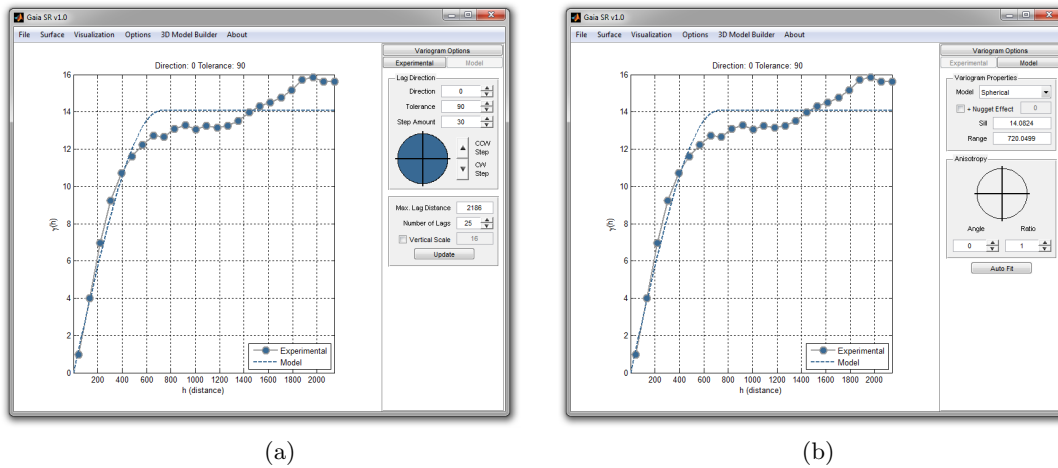


Figura 6.11: Ventana de visualización del variograma. (a) Panel de herramientas para el cálculo del variograma experimental. (b) Panel de herramientas para el ajuste de los modelos admisibles.

Opciones para el cálculo del variograma experimental.

La pestaña *Experimental* en el panel de opciones contiene parámetros para el cálculo del variograma experimental como son la dirección de los intervalos, la distancia máxima considerada, el número de intervalos y la escala vertical. Las diferentes opciones se describen a continuación:

1. *Lag Direction*. El grupo de opciones *Lag Direction* especifica las opciones del cálculo y la visualización del variograma experimental. El gráfico en esta parte muestra la región que se considera en el cálculo. El cambio en cualquiera de estas opciones se refleja automáticamente en las graficas.
 - *Direction*. Esta opción define la dirección o ángulo en la que se calcula el variograma experimental. Esta dirección se mide en grados y toma valores enteros entre -179° y 179° .
 - *Tolerance*. La opción especifica la tolerancia angular que se considera en el cálculo del variograma experimental y toma valores enteros entre 0° y 90° . Entonces solo los pares de observaciones cuya distancia tenga un ángulo entre $Direction - Tolerance$ y $Direction + Tolerance$ son considerados en el cálculo. El valor por defecto es de 90° y con este se calcula el variograma omnidireccional (que considera todas las direcciones) que es válido para el caso isótropo.
 - *Step Amount*, *Step CW* y *Step CCW*. La opción *Step Amount* controla el cambio en la dirección de cálculo activada por los botones *Step CW* y *Step CCW*, esto permite hacer una revisión visual en varias direcciones de el ajuste cuando se

modela el caso anisótropo. *Step CW* decrementa el valor en *Direction* la cantidad especificada en *Step Amount*, esto da un cambio en la dirección de cálculo en el sentido horario. *Step CCW* incrementa el valor en *Direction* la cantidad especificada en *Step Amount*, esto da un cambio en la dirección de cálculo en el sentido antihorario.

2. *Max Lag Distance*. Especifica la distancia máxima que se considera en el cálculo del variograma experimental. Este valor se toma por defecto como un tercio de la longitud de la diagonal de la extensión máxima de los datos. Los valores para intervalos arriba de este valor no son considerados en el cálculo del variograma experimental.
3. *Number of Lags*. Define el número de intervalos considerados en el cálculo del variograma experimental y estos son representados como puntos en la gráfica. El valor por defecto es 25 y este debe ser un valor entero mayor o igual a 1.
4. *Vertical Scale*. Controla la escala vertical del gráfico. Cuando la casilla está activa, el gráfico se vuelve a recalcular de tal manera que las gráficas de los variogramas se ajusten a este, y cuando está inactiva los ejes se quedan fijos, facilitando la verificación visual del ajuste en varias direcciones.
5. *Update*. El botón *Update* aplica los cambios realizados en las opciones *Max Lag Distance*, *Number Lags* y *Vertical Scale*.

Opciones para el ajuste de los modelos admisibles.

La pestaña *Model* en la ventana de opciones avanzadas permite elegir entre los diferentes modelos admisibles de variogramas que se pueden ajustar al variograma experimental, cambiar los valores de los parámetros de los modelos, definir la elipse de anisotropía, además cuenta con la opción de autoajuste. Las diferentes opciones se describen a continuación:

1. *Model properties*. El grupo de opciones *Model properties* controlan el ajuste del modelo seleccionado y los parámetros del mismo. Los cambios en estas opciones se realizan automáticamente y pueden ser modificados manualmente cuando se requiera.
 - *Variogram models*. Permite elegir entre los diferentes modelos admisibles de variograma que se pueden utilizar. Los modelos que se consideran son: Lineal, potencia, esférico, exponencial y gaussiano. Al cambiar de modelo, la gráfica de este se ajusta al variograma experimental que está en la ventana de visualización y los valores de los parámetros se calculan automáticamente. Esto permite el ajuste del variograma omnidireccional (caso isotrópico), cuando la opción *Tolerance* en la pestaña *Experimental* tiene el valor de 90° .
 - *Nugget Effect*. Utiliza un modelo admisible junto con el modelo de efecto nugget. Esta casilla se activa automáticamente cuando el modelo gaussiano es seleccionado. Para los demás modelos esta casilla es opcional. El modelo gaussiano puede ser ajustado sin efecto nugget, pero esto puede provocar un resultado no deseado

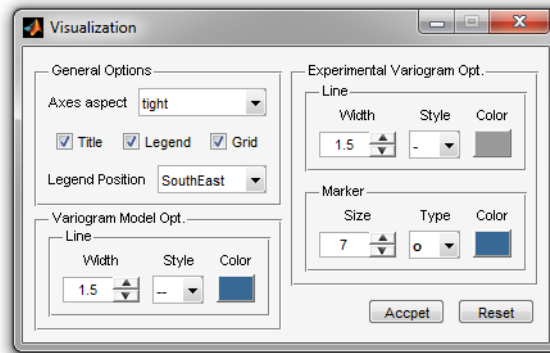


Figura 6.12: Ventana para especificar las opciones de visualización del variograma.

en la interpolación mediante kriging. El valor del efecto nugget debe ser un valor no negativo.

- *Sill/Slope*. La opción especifica el valor de la pendiente para el caso del modelo lineal y potencia y el “umbral” (*sill*) para los demás modelos.
 - *Exponent/Range*. Especifica el valor del exponente para el modelo potencia y el valor del “rango” para todos los demás modelos a excepción del lineal.
2. *Anisotropy*. El grupo de opciones *Anisotropy* definen y permiten la visualización de la elipse de anisotropía, cuando esta sea requerida.
 - *Angle*. Permite establecer la dirección del eje mayor de la elipse de anisotropía. Su valor está medido en grados enteros en un rango de -179° a 179° .
 - *Ratio*. Establece la razón de anisotropía. Debe ser un valor entre 0.05 a 20.
 3. *Autofit*. El botón *Autofit* es una herramienta que permite dar una aproximación del ajuste en el caso anisótropo, el cual debe ser verificado con las herramientas de la pestaña *Experimental* antes de ser utilizado en la interpolación mediante kriging.

Opciones de la visualización del variograma.

La ventana de visualización del variograma cuenta con una serie de opciones a las cuales se puede acceder mediante el submenú *Variogram* dentro del menú *Options*. A continuación se da el listado de las opciones.

1. *Save as image*. Ofrece la opción de guardar la gráfica del variograma, como un archivo de imagen (*bmp*, *eps*, *jpg*, *pdf*, *png*, *tif*).
2. *Save parameters*. Permite guardar los parámetros del modelo ajustado al variograma experimental para que sean utilizados en la reconstrucción de la superficie cuando se usan las opciones avanzadas de kriging. El formato de archivo es *var* que simplemente es texto plano.

3. *Reset*. Permite reestablecer los controles del panel mediante tres opciones: *Experimental settings*, aplica la acción a los controles que se encuentran en la pestaña *Experimental*, *Model settings*, aplica la acción a los controles que se encuentran en la pestaña *Model* y *All settings* a ambas pestañas.
4. *Visualization*. Da acceso al control de visualización del variograma experimental y del modelo ajustado (figura 6.12). Se divide en tres partes principales.
 - *General options*. Aquí se definen el aspecto de los ejes, la activación del título, la malla de referencia y la leyenda junto con su posición.
 - *Variogram model options*. Establece la visualización del modelo de variograma, específicamente las propiedades de la línea con que se dibuja la gráfica. Las opciones son el ancho, estilo y color de la línea.
 - *Experimental variogram options*. Define la visualización del variograma experimental, para ser exactos, las propiedades de la línea y del marcador con que se dibuja la gráfica. Las opciones son el ancho, estilo y color de la línea y tamaño, tipo y color del marcador.

6.2. Módulo de construcción de mallas 3D.

Una de las primeras aplicaciones del módulo ha sido la generación de mallas 3D. La idea de la construcción de las mallas se divide esencialmente en dos pasos:

- Realizar la reconstrucción de una superficie en una región de interés mediante los métodos de interpolación de datos dispersos utilizando mallas estructuradas en dos dimensiones para generar la superficie base (véase figura 6.13a).
- Construir la malla 3D a partir de copias verticales de la superficie base, como se observa en la figura 6.13b.

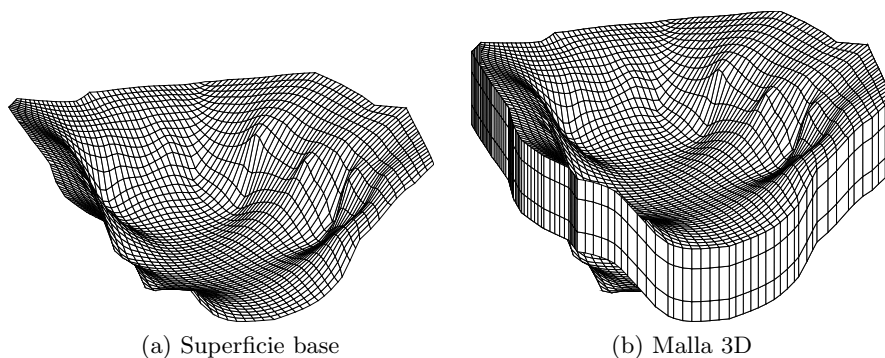


Figura 6.13: Proceso de la construcción de la malla 3D

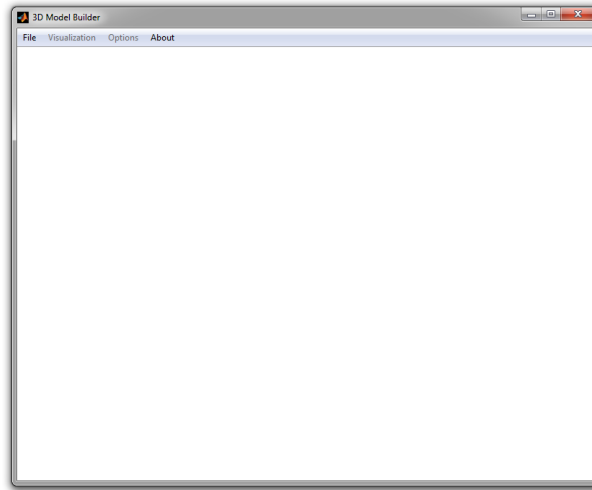


Figura 6.14: Pantalla principal del módulo constructor de mallas 3D

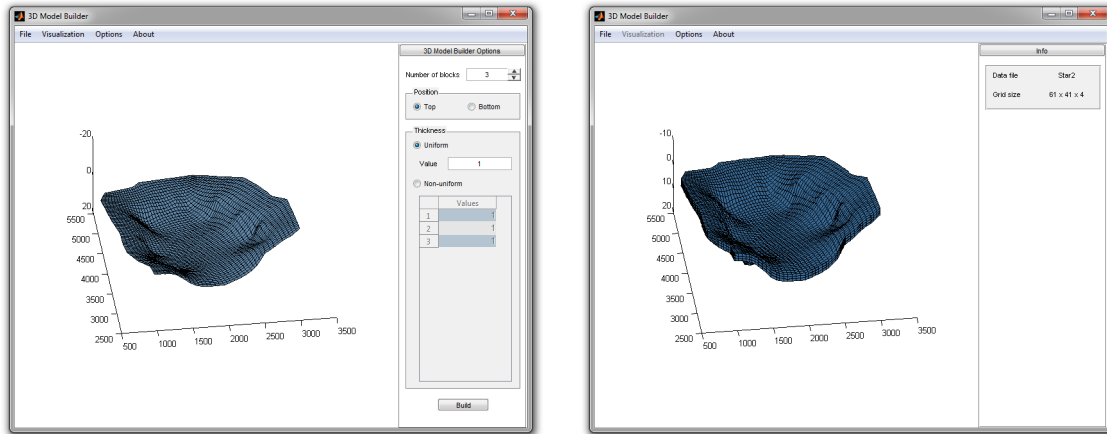
La herramienta, que se muestra en la figura 6.14, está pensada para aprovechar la estructura de las mallas utilizadas para la reconstrucción de la superficie. De este modo cuando se generan las copias verticales para construir la malla 3D es fácil extender y conservar esta propiedad de las mallas en el plano utilizadas (Barrera *et al.* [2]). La herramienta de construcción permite visualizar la superficie base y definir el número de bloques (copias) que se deseen utilizar. Para ello permite especificar el grosor de los bloques de manera uniforme o con diferentes valores para las separaciones entre capa y capa, dando flexibilidad para ciertas aplicaciones, como la generación de mallas hexaedrales.

6.2.1. Descripción del módulo.

El módulo desarrollado dentro del sistema se compone de un conjunto de herramientas para la construcción de mallas 3D a partir de copias verticales de las superficies reconstruidas. La ventana principal del módulo se muestra en la figura 6.14 y se puede acceder a ella mediante el menú *3D Model Builder* que se encuentra entre las opciones principales del sistema.

Este módulo está pensado para trabajar con una superficie reconstruida con el sistema cargada mediante un fichero de datos o directamente a través del submenú *Export to 3DMB* que se encuentra dentro de *Options* → *Surface*. También permite la visualización de mallas construidas con el módulo. Las opciones principales del módulo se listan a continuación:

- *File*, comprende las instrucciones para crear, abrir y guardar los archivos de datos.
- *Visualization*, permite cambiar entre las barras de herramientas laterales que se pueden usar.



(a) Barra lateral de herramientas de construcción

(b) Barra lateral de información de la malla

Figura 6.15: Visualización de la superficie y la malla 3D construida.

- *Options*, proporciona una variedad de opciones para la visualización de la malla construida.
- *About*, despliega información del módulo.

Los menús *Visualization* y *Options* se encuentran deshabilitados hasta que una superficie o una malla 3D hayan sido cargadas al módulo.

Crear y cargar una malla 3D.

Para comenzar a trabajar con el módulo se puede utilizar alguna de las dos siguientes opciones que se encuentran en el menú *File*:

1. *New*, permite cargar una superficie mediante un archivo de datos para utilizarla como base en la construcción de la malla 3D. Los archivos de datos que se utilizan son **gsf** y **dat** que pueden ser generados con el sistema. Una vez que se ha cargado la malla correctamente se muestra la gráfica de la superficie junto con la barra de herramientas lateral con las opciones de construcción de la malla 3D (véase figura 6.15a).
2. *Load*, da la posibilidad de cargar una malla 3D construida previamente. Estas mallas son guardadas en archivos **mesh3d**. Cuando la malla guardada es cargada dentro del sistema se abre la barra lateral de información (figura 6.15b). Cuando se utiliza esta opción el menú *Visualization* queda desactivado.

Guardar y exportar una malla 3D.

Cuando se ha construido una malla 3D esta puede ser guardada o exportada en algunos formatos. Las opciones para guardar y exportar se describen a continuación.

1. *Save*, posibilita guardar la malla construida en el formato `mesh3d`. Cuando se carga una malla previamente construida esta opción permite hacer una copia del archivo original.
2. *Export*, da la opción de exportar la malla al formato de *Tecplot*.

Barra lateral de herramientas de construcción y barra lateral de información de la malla.

El módulo cuenta con dos barras laterales, una con herramientas para la construcción de la malla y otra con información de la misma y que a continuación se describen:

1. **Barra de herramientas de construcción.** La barra proporciona varias herramientas para la construcción de la malla mediante copias verticales. Las herramientas comprendidas en la barra son:
 - *Number of blocks*. Define el número de bloques que tendrá la malla. Este valor está relacionado con el número de copias verticales que se hace de la superficie original. El valor por defecto es de tres bloques.
 - *Position*. Define la posición que tendrá la superficie original. Las opciones son: *Top* cuando se toma como la cima y las copias se generan hacia abajo y *Bottom* cuando se toma como base y las copias se generan hacia arriba. La opción por defecto es *Top*.
 - *Thickness*. Establece los grosores que tendrá cada bloque. Cuenta con dos opciones que son: *Uniform* cuando todos los bloques tiene el mismo grosor y *Non-uniform* cuando se quieren definir grosores distintos. La opción por defecto es *Uniform* con grosor de los bloques uno.
2. **Barra de información.** Muestra la información de la malla construida. La información que se muestra es el nombre del archivo y la dimensión de la malla.

Opciones de visualización.

Se cuentan con varias opciones de visualización de la gráfica. Estas están englobadas dentro del menú *options* y son:

1. *Zoom on/off*. Activa y desactiva la opción de zoom de la gráfica. Cuando la opción se activa esta aparece marcada en el submenú, la gráfica se cambia a la vista 2D y se desactiva la rotación. Una vez que se desactiva la opción de zoom, se vuelve a activar la rotación.
2. *Grid*. Activa y desactiva la malla de referencia.
3. *Reset*. Hay tres opciones que son: *Graph view*, que regresa la gráfica a la vista por defecto (vista 3D, el zoom desactivado y la rotación activada), *Toolbar*, la cual regresa la barra de herramientas de construcción a sus valores por omisión y *All* que aplica las dos acciones anteriores.

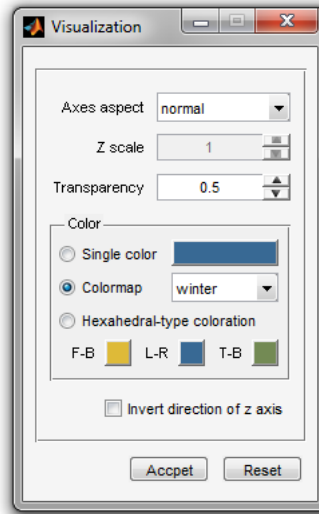


Figura 6.16: Ventana de opciones de visualización.

4. *Visualization*. Permite el acceso al control de visualización de la gráfica (figura 6.16). Con el control se puede modificar el aspecto de los ejes, cambiar la escala en el eje z (cuando el aspecto de los ejes se define igual para los ejes x y y), la transparencia, la orientación del eje z y la forma en que se asigna el color a la malla. El color se asigna en tres formas distintas: *Single color*, que aplica un solo color a todas las capas, *Colormap* asigna un mapa de colores y *Hexahedral-type coloration* que colorea las caras del hexaedro que se forma por la elección de las mallas estructuradas.
5. *Save as image*. Ofrece la opción de guardar la gráfica del variograma, como un archivo de imagen (**bmp**, **eps**, **jpg**, **pdf**, **png**, **tif**).

Capítulo 7

Conclusiones

En este trabajo se hizo una revisión de varios métodos usados en diversas aplicaciones para tratar con el problema de interpolación de datos dispersos en dos dimensiones. En especial se describió el método kriging que es posiblemente el método más utilizado dentro de varios campos de la industria e ingeniería por su enfoque estadístico que ha dado mejores resultados que otros métodos. Aparte de la descripción de la teoría que da sustento a los métodos, como uno de los objetivos principales de este trabajo, estos fueron implementados dentro del sistema *UNAMSI v1.0* para su aplicación en la generación de mallas 3D que son de gran utilidad en la modelación de una gran cantidad de aplicaciones.

El sistema desarrollado es una herramienta de propósito general, por esta razón, permite la resolución del problema de interpolación mediante diferentes métodos, esto para que el usuario tenga la posibilidad de elegir el que mejor se adecue a sus necesidades. Dentro de las cualidades con las que cuenta el sistema es que proporciona herramientas visuales y de edición para el procesamiento de los datos, además de que permite efectuar el cálculo de algunos estadísticos univariados y de histogramas para el análisis de los mismos. También cuenta con herramientas para el cálculo y ajuste del variograma, parte indispensable dentro del método kriging. El módulo realizado dentro del sistema sirve para construir de forma sencilla mallas 3D sobre algunas regiones complicadas aprovechando las cualidades de las mallas obtenidas por el sistema UNAMalla y que puede ser usadas para construir una malla hexaedral estructurada útiles en una gama amplia de aplicaciones.

Dentro del trabajo a futuro se pretende hacer la migración del sistema fuera de **Matlab** para incorporarse a las herramientas de aplicación del sistema UNAMalla. También, ampliar las capacidades de la herramienta de construcción de mallas 3D. Se pretende subir un binario del sistema en la página del grupo UNAMALLA [20] para que sea probado.

Apéndice A

Resultados auxiliares

A.1. Algunos resultados sobre el interpolador $F_{S_0}^*$

En esta parte se consideran algunos resultados sobre el interpolador $F_{S_0}^*$, que también se cumplen para F_{S_0} , al ser este un caso especial de $F_{S_0}^*$ cuando se toman todas las potencias p_i iguales ($i = 1, \dots, N$), de este modo se formaliza la observación 2.2.

Teorema A.1. Sean los nodos denotados por $\mathbf{x}_i = (x_i, y_i)^T \in \mathbb{R}^2$ para $i = 1, \dots, N$, distintos entre sí y $p_i > 0$ para toda i . La función interpolante $F_{S_0}^*$ definida en (2.5) tiene las siguientes propiedades:

a) Si $p_i > 1$ entonces

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_i} \frac{\partial F_{S_0}^*}{\partial x} = \lim_{\mathbf{x} \rightarrow \mathbf{x}_i} \frac{\partial F_{S_0}^*}{\partial y} = 0.$$

b) Si $0 < p_i \leq 1$, entonces, en general, la primera derivada parcial no existe en \mathbf{x}_i .

Demostración. Para $p_i > 1$, recordando que $d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$ entonces

$$\frac{\partial d_i^{p_i}}{\partial x} = p_i d_i^{p_i - 2} (x - x_i),$$

por lo tanto,

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_i} \frac{\partial d_i^{p_i}}{\partial x} = \lim_{\mathbf{x} \rightarrow \mathbf{x}_i} p_i d_i^{p_i - 2} (x - x_i) = 0. \quad (\text{A.1})$$

Sea

$$B_k(\mathbf{x}) = \prod_{\substack{j=1 \\ j \neq k}}^N d_j^{p_j}, \quad \text{con} \quad B_k(\mathbf{x}_i) = 0 \quad \text{si} \quad k \neq i,$$

para $k = 1, \dots, N$, y además se tiene que

$$\frac{\partial B_k}{\partial x} = \sum_{\substack{m=1 \\ m \neq k}}^N \frac{\partial d_m^{p_m}}{\partial x} \prod_{\substack{\ell=1 \\ \ell \neq m \\ \ell \neq k}}^N d_\ell^{p_\ell}.$$

De esta forma,

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_i} \frac{\partial B_k}{\partial x} = 0 \quad \text{si} \quad i \neq k. \quad (\text{A.2})$$

Entonces se tiene que

$$\frac{\partial F_{S_0}}{\partial x} = \frac{\left(\sum_{k=1}^N B_k(\mathbf{x}) \right) \left(\sum_{\ell=1}^N z_\ell \cdot \frac{\partial B_\ell}{\partial x} \right) - \left(\sum_{\ell=1}^N z_\ell B_\ell(\mathbf{x}) \right) \left(\sum_{k=1}^N \frac{\partial B_k}{\partial x} \right)}{\left(\sum_{k=1}^N B_k(\mathbf{x}) \right)^2}, \quad (\text{A.3})$$

combinando las ecuaciones (A.1) y (A.2) con (A.3), se tiene

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_i} \frac{\partial F_{S_0}}{\partial x} = \frac{B_i(\mathbf{x}_i) z_i \frac{\partial B_i}{\partial x} \Big|_{\mathbf{x}_i} - z_i B_i(\mathbf{x}_i) \frac{\partial B_i}{\partial x} \Big|_{\mathbf{x}_i}}{B_i^2(\mathbf{x}_i)} = 0.$$

De manera análoga se obtiene que

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_i} \frac{\partial F_{S_0}}{\partial y} = 0,$$

por tanto, se cumple el inciso a).

Si $0 < p_i \leq 1$, entonces los límites (A.1) y (A.2) no existen, así, en general, de (A.3), las primeras derivadas no existen en \mathbf{x}_i . \square

Teorema A.2. *Sea $D \subset \mathbb{R}^2$ una región acotada tal que los nodos denotados por $\mathbf{x}_i = (x_i, y_i)^T \in D$ para $i = 1, \dots, N$, distintos entre sí. Supóngase también que los índices i son elegidos de manera tal que se cumple que*

$$p_1 = p_2 = \dots = p_m < p_{m+1}, \quad y \quad p_{m+1} \leq p_{m+2} \leq \dots \leq p_N,$$

i.e., las primeras m potencias son iguales y minimales. Si $\gamma = \min_i d_i$, entonces las funciones

$$v_i^*(\mathbf{x}) = \frac{\prod_{\substack{j=1 \\ j \neq i}}^N d_j^{p_j}}{\sum_{k=1}^N \prod_{\substack{j=1 \\ j \neq k}}^N d_j^{p_j}}. \quad (\text{A.4})$$

satisfacen

$$\lim_{\gamma \rightarrow \infty} v_i^*(\mathbf{x}) = \frac{1}{m} \quad \text{si} \quad i \leq m \quad y \quad (\text{A.5})$$

$$\lim_{\gamma \rightarrow \infty} v_i^*(\mathbf{x}) = 0 \quad \text{si} \quad i > m \quad (\text{A.6})$$

Demostración. Defínase

$$\rho = \max_{i,j} d_{i,j} = \max_{i,j} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad \text{y} \quad \beta = \sum_{j=1}^n p_j.$$

Por una parte se tiene que $\gamma^{p_j} \leq d_j^{p_j}$, $j = 1, \dots, N$, por la definición de γ . Por otra parte, utilizando la desigualdad del triángulo, se tiene que $d_j^{p_j} \leq (\gamma + \rho)^{p_j}$, entonces

$$\begin{aligned} \prod_{\substack{j=1 \\ j \neq i}}^N \gamma^{p_j} &\leq \prod_{\substack{j=1 \\ j \neq i}}^N d_j^{p_j} \leq \prod_{\substack{j=1 \\ j \neq i}}^N (\gamma + \rho)^{p_j} \\ \gamma^{\beta - p_i} &\leq \prod_{\substack{j=1 \\ j \neq i}}^N d_j^{p_j} \leq (\gamma + \rho)^{\beta - p_i}, \end{aligned} \quad (\text{A.7})$$

esto para $i = 1, \dots, m$. Usando (A.7), se obtienen las siguientes cotas para cada v_i en (A.4)

$$\frac{\gamma^{\beta - p_i}}{m(\gamma + \rho)^{\beta - p_1} + \sum_{k=m+1}^N (\gamma + \rho)^{\beta - p_k}} \leq v_i^*(\mathbf{x}) \leq \frac{(\gamma + \rho)^{\beta - p_i}}{m\gamma^{\beta - p_1} + \sum_{k=m+1}^N \gamma^{\beta - p_k}} \quad (\text{A.8})$$

Las ecuaciones (A.5) y (A.6) se siguen de la ecuación (A.8) al calcular los límites apropiados. \square

Corolario A.3. *Bajo las hipótesis del teorema A.2, se tiene que para el interpolador $F_{S_0}^*$, se cumple que,*

$$\lim_{\gamma \rightarrow \infty} F_{S_0}^* = \frac{1}{m} \sum_{i=1}^m z_i.$$

Demostración. Del teorema A.2, bajo las hipótesis se cumplen (A.5) y (A.6), entonces se tiene que,

$$\lim_{\gamma \rightarrow \infty} F_{S_0} = \lim_{\gamma \rightarrow \infty} \sum_{i=1}^N z_i v_i^*(\mathbf{x}) = \lim_{\gamma \rightarrow \infty} \sum_{i=1}^m z_i v_i^*(\mathbf{x}) + \lim_{\gamma \rightarrow \infty} \sum_{i=m+1}^N z_i v_i^*(\mathbf{x}) = \frac{1}{m} \sum_{i=1}^m z_i.$$

\square

A.2. Espacios de funciones y optimalidad de las funciones de base radial

En esta sección se verá como a cada función radial básica se le puede asociar un espacio de funciones que usualmente se le conoce como su espacio nativo. Entonces se puede establecer una conexión a los espacios de Hilbert con kernel reproductor, los cuales serán útiles para obtener algunos resultados de optimalidad para la interpolación mediante funciones de base radial.

A.2.1. Espacios de Hilbert con kernel reproductor y espacios nativos para funciones estrictamente definidas positivas

Se comenzara con algunos conceptos sobre espacios de Hilbert con kernel reproductor.

Definición A.1. Sea \mathcal{H} un espacio de Hilbert de funciones reales $f : \Omega(\subseteq \mathbb{R}^d) \rightarrow \mathbb{R}$ con producto interior $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. Una función $K : \Omega \times \Omega \rightarrow \mathbb{R}$ es llamada kernel reproductor para \mathcal{H} si

- 1) $K(\cdot, \mathbf{x}) \in \mathcal{H}$ para toda $\mathbf{x} \in \Omega$.
- 2) $f(\mathbf{x}) = \langle f, K(\cdot, \mathbf{x}) \rangle_{\mathcal{H}}$ para toda $f \in \mathcal{H}$ y toda $\mathbf{x} \in \Omega$.

El nombre de kernel reproductor es tomado por la propiedad reproductora 2) en la definición A.1. Se puede ver que el kernel reproductor en un espacio de Hilbert es único, y que la existencia de un kernel reproductor es equivalente al hecho de que los funcionales de evaluación puntual $\delta_{\mathbf{x}}$ son funcionales lineales acotados sobre Ω , i.e., existe una constante positiva $M = M_{\mathbf{x}}$ tal que

$$|\delta_{\mathbf{x}} f| = |f(\mathbf{x})| \leq M \|f\|_{\mathcal{H}},$$

para todo $f \in \mathcal{H}$ y todo $\mathbf{x} \in \Omega$. Este último hecho se debe al teorema de representación de Riesz. Otras propiedades de los kernels reproductores están dadas por,

Teorema A.4. *Supóngase que \mathcal{H} es un espacio de Hilbert de funciones $f : \Omega \rightarrow \mathbb{R}$ con kernel reproductor K . Entonces se tiene que*

- 1) $K(\mathbf{x}, \mathbf{y}) = \langle K(\cdot, \mathbf{y}), K(\cdot, \mathbf{x}) \rangle_{\mathcal{H}}$ para $\mathbf{x}, \mathbf{y} \in \Omega$.
- 2) $K(\mathbf{x}, \mathbf{y}) = K(\mathbf{y}, \mathbf{x})$ para $\mathbf{x}, \mathbf{y} \in \Omega$.
- 3) *Convergencia en norma del espacio de Hilbert implica convergencia puntual, i.e., si se tiene que $\|f - f_n\|_{\mathcal{H}} \rightarrow 0$ para $n \rightarrow \infty$ entonces $|f(\mathbf{x}) - f_n(\mathbf{x})| \rightarrow 0$ para todo $\mathbf{x} \in \Omega$.*

Ahora es interesante que el kernel reproductor K es definido positivo. Aquí se usa una leve generalización de la definición de función definida positiva a kernel definido positivo. Esencialmente, solo hay que remplazar $\Phi(\mathbf{x}_j - \mathbf{x}_k)$ en la definición 3.6 por $K(\mathbf{x}_j, \mathbf{x}_k)$. Recuérdese que el espacio de funcionales lineales acotados sobre \mathcal{H} es conocido como su dual, y se denota por \mathcal{H}^* .

Teorema A.5. *Supóngase que \mathcal{H} es un espacio de funciones de Hilbert con kernel reproductor $K : \Omega \times \Omega \rightarrow \mathbb{R}$. Entonces K es definido positivo. Más aún, K es estrictamente definido positivo si y sólo si los funcionales de evaluación puntual $\delta_{\mathbf{x}}$ son linealmente independientes en el dual \mathcal{H}^* .*

Este teorema da una dirección de la conexión entre funciones estrictamente definidas positivas y kernels reproductores. Sin embargo, se está también interesado en la otra dirección. Ya que las funciones de base radial que se han obtenido para el problema de interpolación son funciones estrictamente positivas definidas, se quiere encontrar como construir un espacio de

Hilbert con kernel reproductor asociado con esas funciones básicas estrictamente definidas positivas. Entonces se mostrará que toda función básica radial estrictamente definida positiva puede ser de hecho asociada con un espacio de Hilbert con kernel reproductor el cual será conocido como su *espacio nativo*.

Primero nótese que la definición A.1 dice que \mathcal{H} contiene todas las funciones de la forma

$$f = \sum_{j=1}^N c_j K(\cdot, \mathbf{x}_j)$$

suponiendo que $\mathbf{x}_j \in \Omega$. Como consecuencia del teorema A.4 se tiene que

$$\begin{aligned} \|f\|_{\mathcal{H}}^2 &= \left\langle \sum_{j=1}^N c_j K(\cdot, \mathbf{x}_j), \sum_{k=1}^N c_k K(\cdot, \mathbf{x}_k) \right\rangle_{\mathcal{H}} \\ &= \sum_{j=1}^N \sum_{k=1}^N c_j c_k \langle K(\cdot, \mathbf{x}_j), K(\cdot, \mathbf{x}_k) \rangle_{\mathcal{H}} \\ &= \sum_{j=1}^N \sum_{k=1}^N c_j c_k K(\mathbf{x}_j, \mathbf{x}_k). \end{aligned}$$

Por lo tanto, se puede definir el espacio

$$H_K(\Omega) = \text{span} \{K(\cdot, \mathbf{y}) \mid \mathbf{y} \in \Omega\} \tag{A.9}$$

con la forma bilineal asociada $\langle \cdot, \cdot \rangle_K$ dada por

$$\left\langle \sum_{j=1}^{N_K} c_j K(\cdot, \mathbf{x}_j), \sum_{k=1}^{N_K} d_k K(\cdot, \mathbf{y}_k) \right\rangle_K = \sum_{j=1}^{N_K} \sum_{k=1}^{N_K} c_j d_k K(\mathbf{x}_j, \mathbf{y}_k),$$

donde es posible que $N_K = \infty$.

Teorema A.6. Si $K : \Omega \times \Omega \rightarrow \mathbb{R}$ es un kernel simétrico estrictamente definido positivo, entonces la forma bilineal $\langle \cdot, \cdot \rangle_K$ define un producto interior sobre $H_K(\Omega)$ dado por (A.9). Más aún, $H_K(\Omega)$ es un espacio pre-Hilbert con kernel reproductor K .

Observación A.2. Ya que el teorema anterior dice que $H_K(\Omega)$ es un espacio pre-Hilbert, i.e., no es necesario que sea completo, ahora se define el *espacio nativo* $\mathcal{N}_K(\Omega)$ de K al espacio que se obtiene al completar $H_K(\Omega)$ con respecto a la K -norma $\|\cdot\|_K$ tal que $\|f\|_K = \|f\|_{\mathcal{N}_K(\Omega)}$ para todo $f \in H_K(\Omega)$. Los detalles pueden ser vistos en Wendland [22].

En el caso especial cuando se está tratando con funciones estrictamente definidas positivas $\Phi(\mathbf{x} - \mathbf{y}) = K(\mathbf{x}, \mathbf{y})$ y cuando $\Omega = \mathbb{R}^d$ se obtiene una caracterización de los espacios nativos en términos de transformadas de Fourier.

Teorema A.7. *Supóngase que $\Phi \in C(\mathbb{R}^d) \cap L_1(\mathbb{R}^d)$ es una función estrictamente definida positiva de variable real. Defínase*

$$\mathcal{G} = \left\{ f \in C(\mathbb{R}^d) \cap L_2(\mathbb{R}^d) \mid \frac{\hat{f}}{\sqrt{\hat{\Phi}}} \in L_2(\mathbb{R}^d) \right\}$$

y supóngase que este espacio cuenta con la forma bilineal

$$\langle f, g \rangle_{\mathcal{G}} = \frac{1}{\sqrt{(2\pi)^d}} \left\langle \frac{\hat{f}}{\sqrt{\hat{\Phi}}}, \frac{\hat{g}}{\sqrt{\hat{\Phi}}} \right\rangle_{L_2(\mathbb{R}^d)} = \frac{1}{\sqrt{(2\pi)^d}} \int_{\mathbb{R}^d} \frac{\hat{f}(\boldsymbol{\omega}) \overline{\hat{g}(\boldsymbol{\omega})}}{\hat{\Phi}(\boldsymbol{\omega})} d\boldsymbol{\omega}.$$

Entonces \mathcal{G} es un espacio de Hilbert real con producto interior $\langle \cdot, \cdot \rangle_{\mathcal{G}}$ y kernel reproductor $\Phi(\cdot - \cdot)$. Por lo tanto, \mathcal{G} es el espacio nativo de Φ sobre \mathbb{R}^d , i.e., $\mathcal{G} = \mathcal{N}_{\Phi}(\mathbb{R}^d)$ y ambos productos interiores coinciden. En particular, cada $f \in \mathcal{N}_{\Phi}(\mathbb{R}^d)$ puede ser recuperado de su transformada de Fourier $\hat{f} \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$.

Con este resultado se puede encontrar los espacios nativos para las funciones estrictamente definidas positivas, en particular para las funciones gaussianas y las multicuadráticas inversas vistas en los ejemplos de la subsección 3.3.1.

Algunos de estos resultados pueden ser extendidos a funciones condicionalmente estrictamente definidas positivas, aunque no se mostrarán aquí (véase Wendland [22]), con estos se pueden encontrar los espacios nativos para las potencias radiales y los splines de placa delgada.

Ejemplo A.1. Se puede mostrar que los espacios de Beppo-Levi de orden k definidos como

$$BL_k(\mathbb{R}^d) = \{f \in C(\mathbb{R}^d) \mid D^{\boldsymbol{\alpha}} f \in L_2(\mathbb{R}^d) \text{ para todo } |\boldsymbol{\alpha}| = k, \boldsymbol{\alpha} \in \mathbb{N}^d\}$$

donde $D^{\boldsymbol{\alpha}}$ denota una derivada generalizada de orden $\boldsymbol{\alpha}$ son los espacios nativos de las potencias radiales y los splines de placa delgada. Alternativamente, los espacios de Beppo-Levi sobre \mathbb{R}^d son definidos como

$$BL_k(\mathbb{R}^d) = \{f \in C(\mathbb{R}^d) \mid \hat{f}(\cdot) \|\cdot\|^m \in L_2(\mathbb{R}^d)\},$$

entonces las ecuaciones de las transformadas de Fourier de las potencias radiales y los splines de placa delgada muestran inmediatamente que sus espacios nativos son los espacios de Beppo-Levi. ■

A.2.2. La optimalidad de la interpolación con funciones de base radial

Para terminar se verá que dentro de los espacios nativos de Hilbert asociados con funciones radiales condicionalmente estrictamente definidas positivas el interpolante de función de base radial da la *mejor aproximación* a una función de los datos dada.

En el contexto de espacios de Hilbert a lo que se refiere con la mejor aproximación es a encontrar el interpolante de norma mínima. De forma más precisa, dado un espacio de Hilbert \mathcal{H} y los datos $f(\mathbf{x}_1) = \lambda_1(f), \dots, f(\mathbf{x}_N) = \lambda_N(f) \in \mathbb{R}$ con $\{\lambda_1, \dots, \lambda_N\} \subseteq \mathcal{H}^*$, el interpolante de norma mínima es aquella función $g^* \in \mathcal{H}$ que satisface

$$\lambda_j(g^*) = f(\mathbf{x}_j), \quad j = 1, \dots, N$$

y para el cual

$$\|g^*\|_{\mathcal{H}} = \min_{\substack{g \in \mathcal{H} \\ \lambda_j(g) = f(\mathbf{x}_j), j=1, \dots, N}} \|g\|_{\mathcal{H}}.$$

Resulta que el interpolante de función de base radial con función básica Φ satisface este criterio si se toma \mathcal{H} como el espacio nativo asociado $\mathcal{N}_{\Phi}(\Omega)$. Entonces el siguiente resultado dice que el interpolante de base radial para cualquier función condicionalmente estrictamente definida positiva Φ es el interpolante de norma mínima de $\mathcal{N}_{\Phi}(\Omega)$.

Teorema A.8. *Supóngase que $\Phi \in C(\Omega \times \Omega)$ es un kernel estrictamente condicionalmente definido positivo con respecto al espacio de dimensión finita $P \subset C(\Omega)$ y que X es P -unisolvante. Si los valores $f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)$ son dados, entonces el interpolante s es un interpolante de (semi) norma mínima de $\{f(\mathbf{x}_j)\}_{j=1}^N$, i.e.,*

$$|s|_{\mathcal{N}_{\Phi}(\Omega)}^2 = \min_{\substack{g \in \mathcal{N}_{\Phi}(\Omega) \\ g(\mathbf{x}_j) = f(\mathbf{x}_j), j=1, \dots, N}} |g|_{\mathcal{N}_{\Phi}(\Omega)}.$$

El espacio P mencionado en el teorema A.8 se toma usualmente como el espacio $\Pi_{m-1}(\mathbb{R}^d)$ de polinomios multivariados. Incluso, si Φ es estrictamente definida positiva entonces las seminormas en el teorema A.8 se convierten en normas.

Ejemplo A.2. Ya se ha comentado que el espacio nativo de los spline de placa delgada $\phi(r) = r^2 \log r$, $r = \|\mathbf{x}\|$ con $\mathbf{x} = (x, y) \in \mathbb{R}^2$ esta dado por el espacio de Beppo-Levi $BL_2(\mathbb{R}^2)$. Ahora, la correspondiente seminorma en el espacio de Beppo-Levi $BL_2(\mathbb{R}^2)$ es

$$|f|_{BL_2(\mathbb{R}^2)}^2 = \int_{\mathbb{R}^2} \left(\left| \frac{\partial^2}{\partial x^2} f(\mathbf{x}) \right|^2 + 2 \left| \frac{\partial^2}{\partial x \partial y} f(\mathbf{x}) \right|^2 + \left| \frac{\partial^2}{\partial y^2} f(\mathbf{x}) \right|^2 \right) d\mathbf{x}.$$

Por el teorema A.4 el spline de placa delgada minimiza esta seminorma, la cual se puede ver como la energía de torsión de una placa delgada, dando el nombre a estas funciones. ■

Apéndice B

Conceptos básicos de la teoría de la probabilidad.

En este apéndice se introducirán algunas nociones básicas de la teoría de la probabilidad útiles para la comprensión de términos utilizados en la parte de los métodos geostadísticos.

B.1. Variables aleatorias

En la práctica ocurren dos tipos de fenómenos o experimentos: Los primeros que se pueden nombrar son los fenómenos deterministas, los cuales son ciertas acciones en las que se puede predecir lo que va a ocurrir con exactitud. A diferencia de estos, están los fenómenos aleatorios los cuales son acciones en las que no se sabe lo que ocurrirá al ser realizados, ya que pueden obtenerse cualquiera de varias alternativas, aún bajo las mismas condiciones en que se realicen estas acciones. Con ayuda de la teoría de la probabilidad, se podrá realizar el estudio de los fenómenos aleatorios. Se darán algunas definiciones que permitirán adentrarnos en el desarrollo de esta teoría.

Definición B.1. Dado un fenómeno aleatorio, al conjunto de todos sus posibles resultados se le llama **espacio muestral** y se denotará por Ω .

Definición B.2. Sea un fenómeno aleatorio con espacio muestral Ω , una **variable aleatoria** es una función X cuyo dominio es Ω y contradominio es \mathbb{R} , *i.e.*, $X : \Omega \rightarrow \mathbb{R}$.

En adición a estas dos definiciones, se llamará evento a cualquier subconjunto del espacio muestral y se define a la clase \mathcal{F} como la colección de todos los eventos de Ω . Con estas definiciones, la medida de probabilidad se puede ver como sigue. Sea \mathcal{F} la clase de eventos de Ω , entonces para cada evento $A \in \mathcal{F}$ se le asigna un número $P(A) \in [0, 1]$. Este número es la fracción esperada de ocurrencia del evento A , a lo largo de una serie de experimentos donde A o A^c son observados. Algunas propiedades elementales de las medidas de probabilidad son resumidas a continuación.

i) Se cumple que:

$$P(\Omega) = 1 \text{ y } P(\emptyset) = 0.$$

ii) Para eventos $A, B \in \mathcal{F}$,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

iii) Si A y B son disjuntos, es decir $A \cap B = \emptyset$, entonces

$$P(A \cup B) = P(A) + P(B).$$

Obsérvese además que de la propiedad (iii) se sigue que $P(A^c) = 1 - P(A)$. La relación entre variables aleatorias y probabilidad puede ser caracterizada por ciertas cantidades numéricas. Ahora se considerarán algunas de estas.

Definición B.3. La colección de las probabilidades

$$F_X(x) = P(X \leq x) = P(\{\omega \mid X(\omega) \leq x\}), \quad x \in \mathbb{R}, \quad \omega \in \Omega$$

es la **función de distribución de probabilidad acumulada** F_X de X .

Ejemplo B.1. Utilizando la definición B.3 se pueden determinar probabilidades en términos de la función de distribución de probabilidad acumulada, como:

- 1) $P(\{\omega \mid a < X(\omega) \leq b\}) = F_X(b) - F_X(a)$ con $a < b$.
- 2) $P(X = x) = F_X(x) - \lim_{h \rightarrow 0} F_X(x - h)$.

Con estas probabilidades se puede aproximar la probabilidad del evento $\{\omega \mid X(\omega) \in B\}$ para cualquier conjunto complicado $B \subset \mathbb{R}$. ■

Definición B.4. La **distribución** de X para $\mathcal{B} \subset \mathbb{R}$ adecuado esta dada por

$$P_X(\mathcal{B}) = P(X \in \mathcal{B}) = P(\{\omega \mid X(\omega) \in \mathcal{B}\}).$$

Observación B.5. Se denotará por \mathcal{B} a los subconjuntos adecuados de \mathbb{R}^n . Estos subconjuntos son los conjuntos de Borel, los cuales son obtenidos por un número numerable de operaciones \cap , \cup ó c actuando sobre los intervalos de \mathbb{R}^n .

La distribución P_X y la función de distribución de probabilidad acumulada F_X son nociones equivalentes en el sentido de que ambas pueden ser usadas para calcular la probabilidad de cualquier evento $\{X \in \mathcal{B}\}$. Una función de distribución de probabilidad es continua o discreta (tiene saltos). Considérese primero el caso especial cuando la función de distribución F_X es una función discreta.

Definición B.6.

$$F_X(x) = \sum_{k: x_k \leq x} p_k, \quad x \in \mathbb{R}, \quad (\text{B.1})$$

donde $0 \leq p_k \leq 1$ para toda k y $\sum_{k=1}^{\infty} p_k = 1$, con $p_k = P(X = x_k)$.

La función de distribución de probabilidad acumulada (B.1) y la correspondiente distribución se dice que son discretas. Una variable aleatoria con función de distribución de probabilidad acumulada (B.1) es una **variable aleatoria discreta**.

En contraste a las distribuciones y variables aleatorias discretas, la función de distribución de una variable aleatoria continua no tiene saltos, por lo tanto $P(x) = 0$ para toda x , o equivalentemente,

$$\lim_{h \rightarrow 0} F_X(x+h) = F_X(x) \quad \forall x, \quad (\text{B.2})$$

es decir, tal variable aleatoria asume cualquier valor con probabilidad 0. Una variable aleatoria continua obtiene su nombre de la propiedad de continuidad (B.2) de la función de distribución de probabilidad acumulada F_X .

Definición B.7. Sea X una variable aleatoria continua con función de densidad de probabilidad f_X , se define su función de distribución de probabilidad acumulada por:

$$F_X(x) = \int_{-\infty}^x f_X(y) \, dy, \quad x \in \mathbb{R},$$

donde, la función de densidad de probabilidad cumple con:

$$f_X(x) \geq 0 \quad \forall x \in \mathbb{R}, \quad \int_{-\infty}^{\infty} f_X(x) \, dx = 1.$$

Ejemplo B.2 (La distribución normal). Una importante distribución continua es la *distribución normal o gaussiana* $N(\mu, \sigma^2)$ con parámetros $\mu \in \mathbb{R}$, $\sigma > 0$. Esta tiene la función de densidad de probabilidad siguiente:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

Si $X \sim N(0, 1)$, se dirá que la variable aleatoria X tiene una distribución normal estándar, y se denota usualmente con φ su función de densidad de probabilidad f_X y con Φ su función de distribución de probabilidad acumulada F_X . ■

Por otro lado, algunas características interesantes de una variable aleatoria X son la esperanza $E[X]$, la varianza $\text{Var}(X)$ y los momentos $E[X^l]$. A continuación se definirán estas características tanto para variables aleatorias discretas como continuas.

Definición B.8. Sea X una variable aleatoria continua con función de densidad de probabilidad f_X .

i) La **esperanza o valor medio** de X esta dada por, $\mu_X = E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx$.

ii) La **varianza** de X esta definida como, $\sigma_X^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) \, dx$.

iii) El l -ésimo momento de X para $l \in \mathbb{N}$ esta definido como, $E[X^l] = \int_{-\infty}^{\infty} x^l f_X(x) dx$.

iv) Para una función real g la esperanza de $g(X)$ esta dada por, $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$.

Similarmente se define:

Definición B.9. Sea X una variable aleatoria discreta con probabilidades $p_k = P(X = x_k)$.

i) La **esperanza o valor medio** de X esta dada por, $\mu_X = E[X] = \sum_{k=1}^{\infty} x_k p_k$.

ii) La **varianza** de X esta definida como, $\sigma_X^2 = \text{Var}(X) = \sum_{k=1}^{\infty} (x_k - \mu_X)^2 p_k$.

iii) El l -ésimo momento de X para $l \in \mathbb{N}$ esta definido como, $E[X^l] = \sum_{k=1}^{\infty} x_k^l p_k$.

iv) Para una función real g la esperanza de $g(X)$ esta dada por, $E[g(X)] = \sum_{k=1}^{\infty} g(x_k) p_k$.

Se puede considerar la esperanza μ_X como el *centro de gravedad* de la variable aleatoria X , es decir, los valores aleatorios $X(\omega)$, con $\omega \in \Omega$, que están concentrados alrededor del valor no aleatorio μ_X . La extensión o dispersión de los valores aleatorios $X(\omega)$ alrededor de la esperanza μ_X es descrita por la varianza σ_X^2 y la desviación estándar σ_X .

B.2. Vectores aleatorios.

Definición B.10. Se dice que $\mathbf{X} = (X_1, \dots, X_n)$ es un **vector aleatorio** de dimensión n si sus componentes X_1, \dots, X_n son variables aleatorias reales de dimensión uno.

Análogamente a las variables aleatorias de una dimensión se puede introducir la función de distribución de probabilidad, la esperanza, los momentos, la matriz de varianza-covarianza de un vector aleatorio, con el objetivo de describir su distribución y su estructura de dependencia. Como se hizo anteriormente, se considera una colección \mathcal{F} de subconjuntos de Ω y se define una medida de probabilidad en este, es decir, se asigna un número $P(A) \in [0, 1]$ para cada $A \in \mathcal{F}$.

Definición B.11. La colección de las probabilidades

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= P(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= P(\{\omega \mid X_1(\omega) \leq x_1, \dots, X_n(\omega) \leq x_n\}), \quad \mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n, \end{aligned}$$

es la **función de distribución de probabilidad conjunta** $F_{\mathbf{X}}$ de \mathbf{X} .

Esto provee la probabilidad del evento que \mathbf{X} asume valores en el rectángulo

$$(\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \mid a_i < x_i \leq b_i, i = 1, \dots, n\}.$$

Como en el caso de variables aleatorias de una dimensión, estas probabilidades aproximan $P(\mathbf{X} \in B)$ para conjuntos B muy generales.

Definición B.12. La colección de probabilidades

$$P_{\mathbf{X}}(\mathcal{B}) = P(\mathbf{X} \in \mathcal{B}) = P(\{\omega \mid \mathbf{X}(\omega) \in \mathcal{B}\}) \quad \mathcal{B} \subset \mathbb{R}^n,$$

constituyen la **distribución** de \mathbf{X} .

En un sentido matemático, la distribución y la función de distribución de probabilidad conjunta de un vector aleatorio \mathbf{X} son nociones equivalentes. Ambas, $F_{\mathbf{X}}$ y $P_{\mathbf{X}}$, pueden ser usadas para calcular la probabilidad de cualquier evento $\{\mathbf{X} \in \mathcal{B}\}$. Análogamente a las variables aleatorias se puede introducir vectores aleatorios y distribuciones de probabilidad discretas y continuas. Para estos propósitos, los vectores aleatorios continuos con densidad van a ser relevantes, y por tanto se restringirá la atención a estos.

Definición B.13. Si la distribución de un vector \mathbf{X} tiene función de densidad de probabilidad $f_{\mathbf{X}}$, se puede representar la función de distribución de probabilidad conjunta $F_{\mathbf{X}}$ de \mathbf{X} como,

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(y_1, \dots, y_n) dy_1 \cdots dy_n, \quad (x_1, \dots, x_n)^T \in \mathbb{R}^n,$$

donde la densidad es una función que satisface

$$f_{\mathbf{X}}(\mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n, \quad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \cdots dx_n = 1.$$

Si un vector \mathbf{X} tiene función de densidad de probabilidad $f_{\mathbf{X}}$, todas sus componentes X_i , los vectores de los pares (X_i, X_j) , tripletas (X_i, X_j, X_k) , etc., tienen una función de densidad de probabilidad. Estas son llamadas *funciones de densidad de probabilidad marginales*. Además, la esperanza de un vector aleatorio tiene una función similar como el valor medio de una variable aleatoria. Los valores $\mathbf{X}(\omega)$ están concentrados alrededor de este. Así se consideran las siguientes definiciones.

Definición B.14. Sea \mathbf{X} un vector aleatorio.

i) La **esperanza o media** de \mathbf{X} esta dado por:

$$\mu_{\mathbf{X}} = E[\mathbf{X}] = (E[X_1], \dots, E[X_n]).$$

ii) La **matriz de varianza-covarianza** de \mathbf{X} esta definida por:

$$\Sigma_{\mathbf{X}} = (\text{Cov}(X_i, X_j); i, j = 1, \dots, n),$$

donde

$$\text{Cov}(X_i, X_j) = \text{E} [(X_i - \mu_{X_i})(X_j - \mu_{X_j})] = \text{E}[X_i X_j] - \mu_{X_i} \mu_{X_j},$$

es la **covarianza de** X_i y X_j . Nótese que $\text{Cov}(X_i, X_i) = \sigma_{X_i}^2$.

Es conveniente el estandarizar covarianzas de las variables aleatorias mediante la división por sus desviaciones estándar correspondientes. La cantidad resultante

$$\text{corr}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}} = \frac{\text{E} [(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})]}{\sigma_{X_1} \sigma_{X_2}},$$

es la **correlación** de X_1 y X_2 . Como resultado de esta estandarización, la correlación de dos variables aleatorias esta siempre entre -1 y 1.

Ejemplo B.3 (Vector aleatorio gaussiano). Un vector aleatorio normal o gaussiano tiene función de distribución de probabilidad acumulada normal o gaussiana. La función de distribución de probabilidad normal o gaussiana de dimensión n esta dada por su función de densidad de probabilidad

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})^{\top}\right), \quad \mathbf{x} \in \mathbb{R}^n, \quad (\text{B.3})$$

con parámetros $\boldsymbol{\mu} \in \mathbb{R}^n$ y Σ , una matriz simétrica definida positiva de tamaño $n \times n$, donde Σ^{-1} es su inversa y $\det \Sigma$ su determinante. El parámetro $\boldsymbol{\mu}$ es la esperanza $\boldsymbol{\mu}_{\mathbf{X}}$ de \mathbf{X} y Σ es la matriz de varianza-covarianza $\Sigma_{\mathbf{X}}$. Así la función de densidad de probabilidad del vector gaussiano, y por tanto su función de distribución de probabilidad, están completamente determinadas a través de su esperanza y matriz de varianza-covarianza. ■

B.3. Procesos estocásticos.

Definición B.15. Un **proceso estocástico** X es una colección de variables aleatorias

$$(X_t, t \in T) = (X_t(\omega), t \in T, \omega \in \Omega),$$

definida en algún espacio Ω .

En general, T es con frecuencia un intervalo, por ejemplo $T = [a, b]$, $[a, b)$ o $[a, \infty)$ para $a < b$. Entonces se llamará a X un proceso de tiempo continuo en contraste a un proceso de tiempo discreto. En el último caso, T es un conjunto finito o infinito numerable. Por obvias razones, el índice t de una variable aleatoria X_t es frecuentemente referida al tiempo.

Definición B.16. Un proceso estocástico X es una función de dos variables. Para un instante fijo de tiempo t , esta es una variable aleatoria:

$$X_t = X_t(\omega), \quad \omega \in \Omega$$

Para una salida aleatoria fija $\omega \in \Omega$, esta es una función del tiempo:

$$X_t = X_t(\omega), \quad t \in T.$$

Esta función es llamada una **realización**, una **trayectoria** o una **trayectoria de la muestra** del proceso X .

Se tiene que los conceptos de variable aleatoria X y de proceso estocástico ($X_t, t \in T$) no son muy diferentes. Ambos tienen trayectorias, pero la trayectoria $X(\omega)$ con $\omega \in \Omega$ de una variable aleatoria es un número, mientras que la trayectoria $X_t(\omega), t \in T$, de un proceso estocástico es una función en T . Así, sería correcto entender un proceso estocástico como un elemento aleatorio tomando funciones como valores. Más aún, se puede interpretar una variable aleatoria y un vector aleatorio como un proceso estocástico especial con un conjunto finito de índices T .

En analogía a las variables aleatorias y los vectores aleatorios se desea introducir características no aleatorias de un proceso estocástico tales como su distribución, esperanza, etc. y describir su estructura de dependencia. Esta es una tarea mucho más complicada que la descripción de un vector aleatorio. En efecto, un proceso estocástico no trivial $X = (X_t, t \in T)$ con un conjunto infinito de índices T es un objeto de dimensión infinita; esto puede ser entendido como una colección infinita de variables aleatorias $X_t, t \in T$. Ya que los valores de X son funciones sobre T , la distribución de X puede ser definida sobre subconjuntos de un cierto *espacio de funciones*, es decir

$$P(X \in A), \quad A \in \mathcal{F}, \tag{B.4}$$

donde \mathcal{F} es una colección de subconjuntos adecuados de este espacio de funciones. La observación clave es que un proceso estocástico puede ser interpretado como una colección de vectores aleatorios.

Definición B.17. Las **distribuciones finito-dimensionales** (fidis) de un proceso estocástico X son las distribuciones de vectores de dimensión finita

$$(X_{t_1}, \dots, X_{t_n}), \quad t_1, \dots, t_n \in T,$$

para todas las posibles elecciones de tiempos $t_1, \dots, t_n \in T$ y cada $n \geq 1$.

Se puede conceptualizar las fidis mucho más fácilmente que la complicada distribución (B.4) de un proceso estocástico. Se puede mostrar que las fidis determinan la distribución de X . En este sentido, se refiere a la colección de las fidis como la *distribución de los procesos estocásticos*. Los procesos estocásticos pueden ser clasificados de acuerdo a diferentes criterios. Uno de ellos es por el tipo de sus fidis.

Ejemplo B.4. (Proceso gaussiano.) Un proceso estocástico es llamado gaussiano si todas sus fidis son gaussianas multivariadas con función de densidad de probabilidad (B.3). Del ejemplo B.3 se tiene que los parámetros μ y Σ de un vector gaussiano son su esperanza y matriz de varianza-covarianza, respectivamente. Por lo tanto, la distribución de un proceso estocástico gaussiano es determinada solamente por la colección de las esperanzas y matrices de varianza-covarianza de sus fidis.

Un proceso gaussiano simple sobre $T = [0, 1]$ consiste de variables aleatorias $N(0, 1)$ iid. En este caso las fidis estan caracterizadas por las funciones de distribución

$$\begin{aligned} P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n) &= P(X_{t_1} \leq x_1) \cdots P(X_{t_n} \leq x_n) \\ &= \Phi(x_1) \cdots \Phi(x_n), \\ 0 \leq t_1 \leq \dots \leq t_n \leq 1, \quad (x_1, \dots, x_n) \in \mathbb{R}^n. \end{aligned}$$

Para estos procesos, las trayectorias son muy irregulares. ■

Para un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ se definió la esperanza $\mu_{\mathbf{X}} = (E(X_1), \dots, E(X_n))$ y la matriz de varianza-covarianza $\Sigma_{\mathbf{X}} = (\text{Cov}(X_i, X_j), i, j = 1, \dots, n)$. Un proceso estocástico $X = (X_t, t \in T)$ puede ser considerado como la colección de todos los vectores aleatorios $(X_{t_1}, \dots, X_{t_n})$ para $t_1, \dots, t_n \in T$ y $n \geq 1$. Para cada uno de ellos se puede determinar la esperanza y la matriz de varianza-covarianza. Alternativamente se puede considerar estas cantidades como funciones de $t \in T$:

Definición B.18. Sea $X = (X_t, t \in T)$ un proceso estocástico.

i) La **función de esperanza** de X esta dada por:

$$\mu_X(t) = \mu_{X_t} = E(X_t), \quad t \in T.$$

ii) La **función de covarianza** de X se define como:

$$c_X(t, s) = \text{Cov}(X_t, X_s) = E[(X_t - \mu_X(t))(X_s - \mu_X(s))], \quad t, s \in T.$$

iii) La **función de varianza** de X esta definida por:

$$\sigma_x^2(t) = c_X(t, t) = \text{Var}(X_t), \quad t \in T.$$

Como para un vector aleatorio, la función de esperanza $\mu_X(t)$ es una cantidad determinística alrededor de la cual las trayectorias de X están concentradas. La función de covarianza $c_X(t, s)$ es una medida de dependencia en el proceso X . La función de varianza $\sigma_X^2(t)$ puede ser considerada como la medida de dispersión de una trayectoria de X alrededor de $\mu_X(t)$.

Otra forma de clasificación de procesos estocásticos consiste de la imposición de una estructura especial de dependencia.

Definición B.19. El proceso $X = (X_t, t \in T)$, $T \subset \mathbb{R}$, es **estrictamente estacionario** si las fids son invariantes bajo cambios de los índices t

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_n+h}), \quad (\text{B.5})$$

para todas las posibles elecciones de índices $t_1, \dots, t_n \in T$, $n \geq 1$ y h tal que $t_1+h, \dots, t_n+h \in T$.

Observación B.20. Aquí $\stackrel{d}{=}$ significa la identidad de las distribuciones. Para los vectores aleatorios en (B.5), esto significa que sus funciones de distribución son idénticas.

Si se describe un proceso de la vida real (estrictamente o en un sentido amplio) mediante un proceso estocástico estacionario, entonces se creería que las propiedades características de este proceso no cambian cuando el tiempo pasa. La estructura de dependencia descrita por las fids o la función de covarianza es invariante mediante cambios en el tiempo, la cual es una restricción relativamente más fuerte sobre el proceso subyacente, sin embargo, esta es una suposición estándar en muchos campos relacionados con la probabilidad, tales como la estadística y el análisis de series de tiempo.

La propiedad estacionaria puede también ser impuesta sobre los incrementos del proceso. El proceso en si mismo es entonces, no necesariamente estacionario.

Definición B.21. Sean $X = (X_t, t \in T)$ un proceso estocástico y $T \subset \mathbb{R}$ un intervalo. Se dice que X tiene **incrementos estacionarios** si

$$X_t - X_s \stackrel{d}{=} X_{t+h} - X_{s+h} \text{ para todo } s, t \in T \text{ y } h \text{ con } t+h, s+h \in T.$$

Bibliografía

- [1] Armstrong M.; *Basic Linear Geostatistics*, Springer, 1998.
- [2] Barrera P., Cortés, J.J., González G., Domínguez Mota F.J., Tinoco J.G.; *Smoothness and Convex Area Functionals-Revisited*, Journal on Scientific Computing, SIAM, 2010.
- [3] Chilès J. P., Delfiner P.; *Geostatistics. Modeling Spatial Uncertainty*, John Wiley & Sons, 1999.
- [4] Deutsch C. V., Journel A. G.; *Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 1998.
- [5] Fasshauer G. F.; *Meshfree Approximation Methods with MATLAB*, World Scientific Publishing Company, 2007.
- [6] Franke R.; *A Critical Comparison of Some Methods for Interpolation of Scattered Data*, Naval Postgrad. Sch. Tech. Rep. NPS53- 79-003, 1979.
- [7] Franke R.; *Scattered Data Interpolation: Test of Some Methods*, Mathematics of Computations, v. 33, n. 157, p. 181-200, 1982.
- [8] Franke R., Nielson, G.; *Smooth Interpolation of Large Sets of Scattered Data*, International Journal for Numerical Methods in Engineering, v. 15, p. 1691-1704, 1980.
- [9] Gordon W. J., Wixom J. A.; *Shepard's Method of "Metric Interpolation" to Bivariate and Multivariate Interpolation*, Mathematics of Computations, Vol. 32, No. 141, pp. 253-264, 1978.
- [10] Hjelle Ø.; *Approximation of Scattered Data with Multilevel B-Splines*, SINTEF report, 2001.
- [11] Isaaks E. H., Srivastava R. M.; *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 1989.
- [12] Lee S., Wolberg G., Shin S. Y.; *Scattered Data Interpolation with Multilevel B-Splines*, IEEE Transactions on Visualization and Computer Graphics, Vol. 3, No. 3, July-September 1997.
- [13] Powell M. J. D.; *Approximation Theory and Methods*, Cambridge University Press, 1981.

-
- [14] Powell M. J. D.; *The Theory of Radial Basis Functions in 1990*, in Advances in Numerical Analysis II: Wavelets, Subdivision, and Radial Basis Functions, W. Light (ed.) Oxford University Press, págs. 105-210, 1992.
- [15] Renka R. J.; *Multivariate Interpolation of Large Sets of Scattered Data*, ACM Transaction on Mathematical Software, vol. 14, no. 2, p. 139-148, 1988.
- [16] Renka R. J.; *Algorithm 660. QSHEP2D: Quadratic Shepard Method for Bivariate Interpolation of Scattered Data*, ACM Transaction on Mathematical Software, vol. 14, no. 2, p. 149-150, 1988.
- [17] Renka R. J.; *Algorithm 790. CSHEP2D: Cubic Shepard Method for Bivariate Interpolation of Scattered Data*, ACM Transactions on Mathematical Software, vol. 25, no. 1, p. 70-73, 1999.
- [18] Shepard D.; *A Two Dimensional Interpolation Function for Irregularly Spaced Data*, Proc. 23rd Nat. Conf. ACM, p. 517-523, 1968.
- [19] Spath H.; *Two Dimensional Spline Interpolation Algorithms*, A. K. Peters, 1995.
- [20] UNAMALLA GROUP.; *Página web del grupo UNAMALLA*, <http://www.matematicas.unam.mx/unamalla>.
- [21] Wackernagel H.; *Multivariate Geostatistics: An Introduction with Applications*, Springer, Berlin, 2003.
- [22] Wendland H.; *Scattered Data Approximation*, Cambridge University Press, 2005.